Complex-script shaping using `luaotfload` and HarfBuzz

Graham Douglas

July 30, 2021

Introduction

This project is processed by LuaHBTEX to demonstrate the text-shaping capability of the `loaotfload` Lua library, comparing results from using `mode=harf` vs `mode=node`. We use three different Arabic typefaces, defining various `\font` "versions" by using different settings for `loaoftload`. In these examples we make use of the primitive commands `\pardir` and `\textdir` to set the correct paragraph and text directions. We use the Amiri and Aref Ruqqa fonts to demonstrate improved shaping results when using the HarfBuzz-based shaper via `mode=harf`.

The *unshaped* input: Unicode text

Firstly, here are the input characters: *unshaped* Unicode Arabic characters displayed in the order they are present in the Overleaf `main.tex` file—this is the so-called *logical order*—the sequence in which the characters are read-in by the LuaHBTEX engine:

<div dir="rtl">

مِيحِّرلَاِ نْأَمْحِّرلَاِ هّلِلاِ مْسِب

</div>

Because the Arabic script is read right-to-left, the reading order—also called visual or display order—of the typeset glyphs is the reverse of the sequence in which the unshaped text (characters) are stored in the text file and read-in by the TEX engine (or any other software).

The *shaped* result

Here, we'll use the `Scheherazade` font, which is included with TEX Live. We'll create two instances of Scheherazade:

- \ScheherazadeHarf which uses luaotfload's node-based shaping (mode=node):

  \font\ScheherazadeNode =
  {name:Scheherazade:mode=node;script=arab} at 40pt

- \ScheherazadeNode which uses luaotfload's HarfBuzz-based shaping (mode=harf):

  \font\ScheherazadeHarf =
  {name:Scheherazade:mode=harf;script=arab} at 40pt

Using \ScheherazadeHarf (mode=harf)

Here, the Unicode text has been *shaped* using \ScheherazadeHarf—defined using the Scheherazade font with shaping applied by HarfBuzz:

$$\text{بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ}$$

Using \ScheherazadeNode (mode=node)

The next example uses the same Unicode text but it is *shaped* using the font \ScheherazadeNode—which was defined using the Scheherazade font with shaping applied by luaotfload's built-in shaper written in Lua (node-based shaping):

$$\text{بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ}$$

Note how HarfBuzz-based shaping and luaotfload's *built-in* shaping produce equally good results—using the Scheherazade font and these examples.

However, for other fonts, and/or scripts, HarfBuzz-based shaping can produce superior results. As of March 2021, page 13 of the luaotfload documentation (as of version 3.17) advises "harf *mode improves greatly the rendering of indic and arabic scripts and is highly recommended for such scripts*".

Forgetting to set the script

Here, we defined the font \ScheherazadeNoShape but deliberately omitted to set the correct script:

```
\font\ScheherazadeNoShape =
{name:Scheherazade:mode=node} at 40pt
```

In this case, the text is unshaped because we did not provide `luaotfload` with the correct `script` setting—we omitted `script=arab` in the options provided to `luaoftload`:

بِسْمِ ٱللّٰهِ ٱلرّحْمَٰنِ ٱلرّحِيمِ

Note that the unshaped characters are shown in their visual (display) order.

Examples where HarfBuzz gives better results

The following examples demonstrate fonts for which HarfBuzz, via `luaotfload`'s `mode=harf`, improves the shaping.

Amiri: `luaotfload`'s `mode=harf` vs `mode=node`

The following example uses the Amiri font.

`luaoftload` with `mode=harf`:   بِسْمِ ٱللّٰهِ ٱلرّحْمٰنِ ٱلرّحِيمِ

`luaoftload` with `mode=node`:   بِسْمِ ٱللّٰهِ ٱلرّحْمٰنِ ٱلرّحِيمِ

With `mode=node` you can see "collisions" in some diacritics—perhaps easier to see if we increase the font size: HarfBuzz (`mode=harf`) ح vs `luaotfload`'s built-in shaper (`mode=node`) ح. With `mode=node`, the "fatha" and "shadda" characters have overlapped—like this rather than being positioned atop each other like this: .
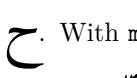
Aref Ruqqa: `luaotfload`'s `mode=harf` vs `mode=node`

The following text uses the Aref Rugga font, and is based on the example shown in the TUGBoat article Bringing world scripts to LuaTEX: The HarfBuzz experiment, written by Khaled Hosny.

luaoftload with mode=harf: بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

luaoftload with mode=node: بِسۡمِ ٱللَّهِ ٱلرِّحۡمَٰنِ ٱلرِّحِيمِ

Here, HarfBuzz produces clearly superior results because the glyphs are stacked correctly, in accordance with the [Ruqqa style](#).