To: Unicode Technical Committee
From: Azzeddine Lazrek (via Deborah Anderson)
Contact: lazrek@uca.ac.ma
Subject: More precisions about Hamza
Date: 27 December 2017

Some remarks about comments on page 4-6 of the script ad hoc recommendations:
Recommendations to UTC #153 November 2017 on Script Proposals L2/17-384
http://www.unicode.org/L2/L2017/17384-script-ad-hoc-recs.pdf
against L2/17-252 Proposal to encode some Hamza Quranic marks.

Ad hoc recommendations:
- Character #1 ٰ ARABIC HIGH WASLA

The examples show this combining character only over alef.

The proposal mentions "Moreover, we could need to use Wasla alone, to explain its use for example." This indeed happens, and is used in pedagogical text. We do not recommend encoding a separate combining mark due to destabilization of encoding, but a spacing mark for wasla, similar to the characters in the U+FBB2..U+FBC1 range could be encoded.

Note that a visually similar character placed over other characters was earlier proposed by Mussa Abudena (figure 1 in L2/16-056), but that is a waqf sign, which is graphically much larger. The script ad hoc had earlier recommended this character for encoding in L2/16-342. After discussion with Arabic experts, the suggested name put forward was ARABIC SMALL HIGH WORD SAH (with an annotation “sign of waqf”).

Comments:
This proposed character has no semantic relation with the earlier proposed one with the suggested name ARABIC SMALL HIGH WORD SAH (with an annotation “sign of waqf”); even if they have some visually similarities.

Effectively, the first one could be only over Alef. However, the second one is over the end of a sentence.

We don’t need to encode this character only to be able to use Wasla alone for pedagogical uses. Mainly, we need this character to be able to add it if it appears, as it have already explained at L2/17-068, (see Table 1 – Variations of an encoding word Page 3).
For example, how to codify the following word in different situations:

<table>
<thead>
<tr>
<th></th>
<th>ﺍ.vector</th>
<th>ﺍ</th>
<th>ب</th>
<th>د</th>
<th>و</th>
<th>ا</th>
</tr>
</thead>
<tbody>
<tr>
<td>0627</td>
<td>0639</td>
<td>0628</td>
<td>062F</td>
<td>0648</td>
<td>0627</td>
<td></td>
</tr>
</tbody>
</table>

Where X could be:

<table>
<thead>
<tr>
<th></th>
<th>ا</th>
<th>عبدال</th>
<th>ﺍ</th>
<th>ﺍ</th>
<th>ﺍ</th>
<th>ﺍ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0627</td>
<td></td>
<td>ﻉٰ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>Nothing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and
• The variant ٰ must be encoded, by the sequence: ٰ U+0627 and ٩ 0654, but not by: ٰ U+0623, even if it exits. It will be recommended to not use ٰ U+0623 anymore.

• The variant ِ must be encoded, by the sequence: ِ U+0627 and ِ, but not by: ِ U+0671, even if it exits. It will be recommended to not use ِ U+0671 anymore.

• The variant ُ ُ must be encoded by the sequence: ُ ُ U+0627 and ُُ and ُُ, but not only by one code.

• The variant ُ ُ must be encoded by the sequence: ُ ُ U+0627 and ُُ and ُُ, but not only by one code, as it is proposed by Mussa Abudena L2/15-329 and L2/16-044.

By ن, any research engine will be able to find any occurrence of the preceding word at all its situations.

Regarding the ad hoc recommendations’ comment:

We do not recommend encoding a separate combining mark due to destabilization of encoding, but a spacing mark for wasla, similar to the characters in the U+FBB2..U+FBC1 range could be encoded.

... We recommend … a revised proposal should be prepared, with references to the appropriate examples:

a. ARABIC SYMBOL WASLA ABOVE (at U+FBC2)

Ad hoc recommendations:

We recommend … a revised proposal should be prepared, with references to the appropriate examples:

[...]

c. If the size of the stroke can be resolved (and code points to be determined)

ALEF WITH ATTACHED FATHA (whose glyph has a stroke at the top)
ALEF WITH MIDDLE STROKE
ALEF WITH ATTACHED KASRA (whose glyph has a stroke on the bottom)

Comments:
In our opinion, for the same reason, I do not agree to encode ALEF WITH some symbols. However, I have proposed to encode those symbols ALONE.

(The numbers in the far left column below refer to proposed characters from L2/17-252.)

• Characters #3 and #5 have same semantic with different shapes, we find both in practice.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 3 | 08C2 | ARABIC SMALL DOT LEFT  
• Starting Hamza Wasl with Damma  
→ 302E • 〇 Hangul single dot tone mark |
| 4 | 08C3 | ARABIC SMALL DOT BELOW  
• Starting Hamza Wasl with Kasra |

• Characters #4 and #6 have same semantic with different shapes, we find both in practice.
• Characters #11 and #12 have same semantic with different positions, we find both in practice.

• Characters #2 and #10 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

• Characters #3 and #11 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

• Characters #4 and #13 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

• Characters #7 and #16 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

• Characters #8 and #18 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.
- Characters #9 and #19 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

![Character #9 and #19](image)

- Characters #17 and #18 have same semantic with different positions, we find both in practice.

![Character #17 and #18](image)

**Ad hoc recommendations:**

Character #5 ⊘ ARABIC SMALL RING LEFT and Character #6 ⊘ ARABIC SMALL RING BELOW

What is the character in Figure 36.b (see below, from L2/17-252)? Is it #6?

![Character in Figure 36.b](image)

**Comments:**
Characters #6 and one used in Figure 36.b have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

The first one is used for Starting Hamza Wasl with Kasra (find at some Mushaf Qaloon). However, the second one is used for Al-Imala (find at some Mushaf).

The first one is used only under Alef. However, the second one is used under some other letters.