Introduction

Both Sindhi [snd] and Behdini Kurdish [kmr] have glyph variants. In order for font developers to better support these languages, these variants should be documented in the Unicode standard.

Sindhi

The Sindhi [snd] language is spoken by over 31+ million people in Pakistan and India. Sindhi can be written with the Arabic, Devanagari, Gurmukhi, and Khojki scripts (script use is generally based on the country or region of use although the Arabic script is the primary script in use) [1].

U+204F REVERSED SEMICOLON and U+2E41 REVERSED COMMA

See Figures 1, 2, 3.

The Sindhi comma and semicolon are vertical flips of the standard design:

- Default shape: ٍ،
- Sindhi shape: ٠،

Desired outcome

In Chapter 9: Middle East-I of the Unicode Core Spec [4], in a section on “Punctuation”, it says “Sindhi uses U+2E41 REVERSED COMMA and U+204F REVERSED SEMICOLON.” Because some fonts have used features to offer an alternate design of U+060C and U+061B, this should instead say “Sindhi uses U+2E41 REVERSED COMMA and U+204F REVERSED SEMICOLON. Some fonts have used glyph variants of U+060C ARABIC COMMA and U+061B ARABIC SEMICOLON, although this is not recommended.”

U+0645 ARABIC LETTER MEEM

See Figure 2.

The Sindhi meem has a short tail in isolate and final positions:

- Default shape: م،
- Sindhi shape: م،

Desired outcome

An annotation should be added to this character.

0645 م ARABIC LETTER MEEM
- Sindhi uses a short tail

Additionally, Chapter 9 in the Core Spec should be updated to give an image of the short tail. After Table 9-2 some text and another table could be added:
In general, the distinction between a long tail and a short tail is stylistic. However, Sindhi specifically prefers the *meem* to have a short tail in isolate and final positions:

Table 9.x Sindhi Meem design

<table>
<thead>
<tr>
<th>Default Shape</th>
<th>Sindhi Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>م م م م</td>
</tr>
</tbody>
</table>

No properties need adjusting for this character. Since we are not asking for a change in the charts, the Presentation forms block should not need adjusting.

**U+06FE ARABIC SIGN SINDHI POSTPOSITION MEN**

*See Figure 2.*

U+06FE ARABIC SIGN SINDHI POSTPOSITION MEN was added to Unicode specifically for the Sindhi language. In Figure 2 we can see that this has a short tail (matching the short tail used for U+0645). In the Unicode charts, U+06FE currently has a glyph matching the standard meem.

Chart shape: 

Sindhi shape: 

**Desired outcome**

We believe the glyph for U+06FE should be changed in the charts to have a short tail. If this is done, no annotations are required.

No properties need adjusting for this character. This character is not present in the Presentation forms block.

**Sindhi heh**

Sindhi also has a different form of *heh*. Another document will be submitted regarding the Sindhi *heh*.

**Northern/Behdini Kurdish**

*See Figure 4, 5, 6.*

Behdini is a local variant of Northern Kurdish [kmr] spoken by 1.5-2.0 million people in northern Iraq. This variant uses the Arabic script. Most of the [kmr] speakers are in Turkey; if they read and write Kurdish, they use a Latin-based script. There are [kmr] speakers in Iran, but we are not familiar with their writing conventions. There is a small enclave of [kmr] speakers in Armenia who use Cyrillic script.

**U+0647 ARABIC LETTER HEH**

For close to 20 years, Behdini Kurdish has used both U+0647 ARABIC LETTER HEH (ه ههه) and U+06D5 ARABIC LETTER AE (اء). Since the standard initial and final shapes of *heh* match the shapes for *ae*, they must be distinguished. In order to distinguish them, Behdini Kurdish has a glyph variant for *heh* which should be documented so that font designers will begin to support this variant. This is also documented in Kew, section 3.6, page 7 [2]. The Behdini *heh* represents the same sound as it does in the Arabic language.
Pournader [3] also mentions Behdini Kurdish and Kew’s [2] document in his paper "The right hehs for Arabic script orthographies of Sorani Kurdish and Uyghur." He simply says: "Further research and clarification is needed for various hehs used in the diverse orthographies using the Arabic script. Among those with similar challenges for heh are Sindhi, Kazak, Kirgiz, and Behdini Kurdish. Kew 2005 addresses some of them issues, but the author of the present proposal does not agree with all of its recommendations. Future proposals will address those orthographies." There is no mention of an alternative solution for Behdini Kurdish. While this document contains a great deal of information on Sorani Kurdish and Uyghur a separate document should be submitted for those languages.

**Desired outcome**

Chapter 9 in the Core Spec should be updated to give an image of the Behdini heh. Suggested additional text in blue below:

**Letter heh.** In the case of U+0647 ARABIC LETTER HEH, the glyph ۫ is shown in the code charts. This form is often used to reduce the chance of misidentifying heh as U+0665 ARABIC-INDIC DIGIT FIVE, which has a very similar shape. The isolated forms of U+0647 ARABIC LETTER HEH and U+06C1 ARABIC LETTER HEH GOAL both look like U+06D5 ARABIC LETTER AE. However, for Behdini Kurdish the isolate, initial, and final forms are knotted.

<table>
<thead>
<tr>
<th>Table 9.x Behdini Kurdish heh design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Shape</td>
</tr>
<tr>
<td>📲</td>
</tr>
</tbody>
</table>

No properties need adjusting for these characters. Since we are not asking for a change in the charts, the Presentation forms block should not need adjusting.

**Samples**

![Figure 1: U+060C ARABIC COMMA and U+061B ARABIC SEMICOLON in Sindhi language. [sd-1]](image)
Figure 2: U+060C ARABIC COMMA, U+0645 ARABIC LETTER MEEM, and U+06FE ARABIC SIGN SINDHI POSTPOSITION MEN. Sindhi language. [sd-2]

Figure 3: U+061B ARABIC SEMICOLON. Sindhi language. [sd-3]
Figure 4: ARABIC LETTER HEH (isolate, initial and medial). Behdini Kurdish language. [kmr-1].

Figure 5: ARABIC LETTER HEH (isolate, initial and medial). Behdini Kurdish language. [kmr-2].

Figure 6: ARABIC LETTER HEH (final). Behdini Kurdish language. [kmr-3].
References


**Sindhi References** (with thanks to Abdul Majid Bhurgri for these examples)


[sd-3] This lithographed book is dated 22 December 1911 and was printed in 1913 by Nawalkishore press Lahore.

**Behdini Kurdish References**


[kmr-3] کرته لئهو بكه پۆ دابهۆاندن