

To: UTC

Title: Request for annotations for Sindhi and Behdini Kurdish

From: Lorna Priest Evans (SIL International)

Date: 7 December 2020

## 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: ٴ ٴ  
 Sindhi shape: ٴ ٴ

### 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 م م م  
 Sindhi shape: م م م 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**


Default Shape	Sindhi Shape
	


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 هه) and U+06D5 ARABIC LETTER AE (ا ا). 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.

	Default	Behdini Kurdish
U+0647		

## Desired outcome

**Letter heh.** In the case of U+0647 ARABIC LETTER HEH, the glyph ه 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.

Common Shape	Behdini Kurdish Shape
	

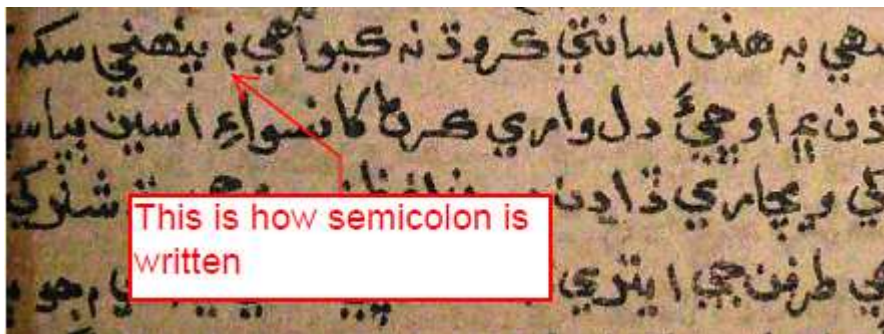
## Samples



جي هٿن ۾ موجود ڊڪشنريءَ تي لڳ ڀڳ نون سالن تائين سنڌي ٻوليءَ ۽ لسانيات جا عالم ۽ ماهر جناب سراج الحق ميمڻ، ڊاڪٽر غلام علي الانا، پروفيسر قلندر شاهه لڪياري، جناب سيد حاڪم علي شاهه بخاري ۽ ڊاڪٽر فهميده حسين ترجمي ۽ ادارت جي مرحلن ۾ شامل رهيا ۽ رهنمائي ڪندا رهيا. سنڌيڪارن جي چونڊ جي مرحلي ۾ ممتاز دانشورن ڊاڪٽر نبي بخش خان بلوچ، جناب محمد ابراهيم جويو، ڊاڪٽر غلام علي الانا ۽ جناب سراج الحق ميمڻ صاحب جن مدد فرمائي.

ٻين علمي ۽ تعليمي اشاعتن سان گڏوگڏ اوڪسفرڊ جو تمام اهم ڪم لغت سازي آهي. دنيا ۾ انگريزي لغتن جي شروعات سورھين صدي عيسويءَ جي پڇاڙڪن سالن کان ٿئي ٿي. انگريزي لغت نويسيءَ جي تاريخ ۾ سيموئل جانسن (1755ع) جو نالو چوٽيءَ تي آهي. بعد ۾ ٻه لساني لغتن جو سلسلو شروع ٿئي ٿو، جيڪو سدائين کان ئي نهايت ڪنن پئي رهيو آهي. ڇاڪاڻ ته اهو دراصل هڪ مڪمل تهذيب ۽ تاريخ کي ٻي تهذيب جي تجربن ۾ بيان

**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]



خال ، خال دى بۆتە گەلەك باشتر چىكەم و تۆزا  
كەتتە سەر.....

-باشە بەس يى ھىيار بى تېلېن تە يىن پيس ب  
گۆرا من نەكەفن.

-بلا خال ل سەر ھەردوو چاقىن خۆ..

-ھا ھا چاقىن تە ئەوین ھەكى يىن سقوران، تو  
دچە قوتابخانى؟

- بەلى خال ، ل ريزا سىي .

-ئەفە چو كويسەلەيە دگەل تە ؟

-ئەو نە كويسەلەيە خال ، ئەو براين من يى  
بچويكە، پىچەك دفنا خۆ خوراند، پى من يى چەپى  
كورتە زىي راستى ودختى دزقەمەفە مال نەشيم  
سندوفكا خۆ بەلگرم ئەو ھارى من دكەت.

پرى قەرەبالغە ، ھندەك زەلامىن ب چەك ژ وى  
ترۆمبىلى ھاتنە خار بەرى خودانا بازنگان پەييا  
ببیت زەلامەكى دەرگەھ بۆ قەكر، دەنگى شىلكەكا  
گۆللەيان ژدوویر دەھات، ھندەك گەنجا پشتا خۆ يا  
دايە دیوارەكى و ئىكى زنجیرەكا د دەستی داو ھەر  
ییل دۆر تېلەكا خۆ دزقەرنیت و بەر ددەت ، سەرى  
وى ھەكى ئامیرەكى رادارى ھەرى دزقەرت چەپ و  
راست ، جار دەنگى كاویس ئاغای و ئافزا لو خالۆ لو  
خالۆ، جار يى داخل ھەسەن ، جار ماڤى ماڤى و  
جار گولەشەنگەم، جار ئەياز يى ھاوار دكەت  
چىكەم ئەز ؟ ژ وى تۆمارگەھا ل دووماھىكا پرى  
يى بلند دبیت.

سندوفكا وى يا بویاغى يا ل ئاخى دخشییت و ملۆ  
وى يى چەپى يى خار كرى گەھشەتە بەرانبەر  
قۆركىن چىچۆیى چایچى.

خودانى زكى مەزن ھىشتا يى قورچكا ل جایى د

Figure 4: ARABIC LETTER HEH (isolate, initial and medial). Behdini Kurdish language. [kmr-1].

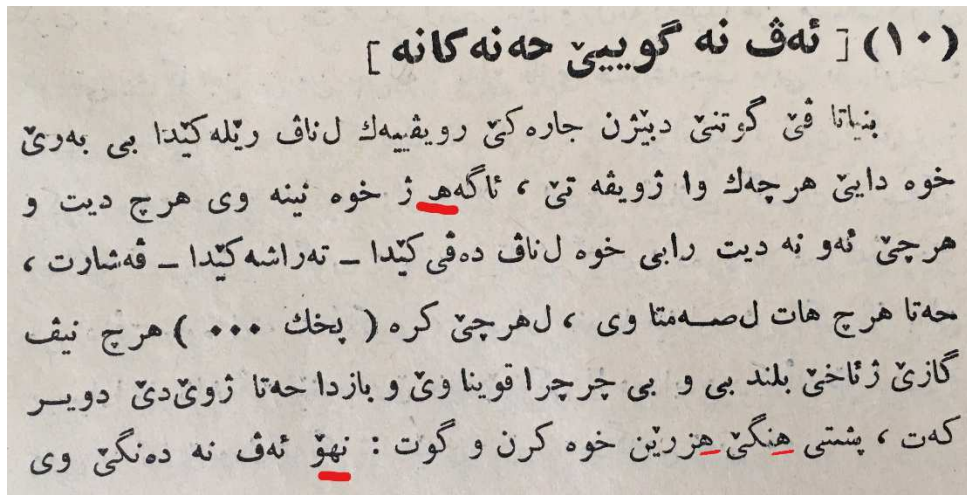


Figure 5: ARABIC LETTER HEH (isolate, initial and medial). Behdini Kurdish language. [kmr-2].

سیه سالی لگوندا

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

## References

- [1] Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the World*. Twenty-third edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>. (accessed 14-Aug-2020).
- [2] Kew, Jonathan. 2005. Notes on some Unicode Arabic characters: recommendations for usage. [https://scripts.sil.org/cms/scripts/render\\_download.php?format=file&media\\_id=arabicletterusagenotes&filename=ArabicLetterUsageNotes.pdf](https://scripts.sil.org/cms/scripts/render_download.php?format=file&media_id=arabicletterusagenotes&filename=ArabicLetterUsageNotes.pdf)
- [3] Pournader, Roozbeh. 2014. The right hehs for Arabic script orthographies of Sorani Kurdish and Uighur <http://www.unicode.org/cgi-bin/GetMatchingDocs.pl?L2/14-136>. (Accessed 6 Nov 2020) Minutes: <https://www.unicode.org/L2/L2014/14100.htm> Still an open action item for Roozbeh: <https://www.unicode.org/L2/L-SD2.htm>
- [4] The Unicode Consortium. The Unicode Standard, Version 13.0.0, (Mountain View, CA: The Unicode Consortium, 2020. ISBN 978-1-936213-26-9) <http://www.unicode.org/versions/Unicode13.0.0/>

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

- [sd-1] Swami Vivekanand published by Sahtia Academy, New Delhi in 1965.
- [sd-2] Oxford Concise English Sindhi Dictionary. Publishing in 2010.
- [sd-3] This lithographed book is dated 22 December 1911 and was printed in 1913 by Nawalkishore press Lahore.

### **Behdini Kurdish References**

- [kmr-1]. Peyv – Journal of Duhok Writers Union no 82. Page 106.
- [kmr-2] Miştaxa Çiye- M. M. Dêrşewî. 1979. Page 7.
- [kmr-3] کرته لیڤه بکه بو دابه‌زاندن