

To: UTC  
 Title: Regarding the Behdini Kurdish Heh  
 From: Lorna Priest Evans (SIL International)  
 Date: 20 April 2021


## Northern/Behdini Kurdish

Behdini Kurdish [kmr] has a glyph variant for the *heh*. In order for font developers to better support these languages, these variants should be documented in the Unicode standard.

See Figure 1, 2, 3.

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.

For close to 20 years, Behdini Kurdish has used both U+0647 ARABIC LETTER HEH (ه ههه) 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.

The Behdini Kurdish *heh* looks like this: 

Three possibilities for a solution should be considered. These would be to:

1. consider the Behdini Kurdish *heh* a glyph variant of U+0647 (preferred so that documents do not need changing)
2. consider the Behdini Kurdish *heh* a glyph variant of U+06BE
3. encode a new Kurdish Behdini *heh*

## Information in the Unicode Standard regarding *heh* and *heh doachashmee*

The issue of which HEH to use in Kurdish is not discussed in the Unicode chapter on Arabic. However, this information is on pages 380-381:

**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.

U+06BE ARABIC LETTER HEH DOACHASHMEE is used to represent any heh-like letter that appears with stems at both sides in all contextual forms. The exact contextual shapes of the letter depend on the language and the style of writing. The forms shown in Table 9-8 for KNOTTED HEH are used in certain styles of writing in South Asia. Other South Asian styles may use different medial and final forms. The style used in China and Central Asia for languages such as Uyghur uses medial and final forms for HEH DOACHASHMEE that are visually similar to the medial form of HEH shown in Table 9-8.

Table 9-8 includes the following glyphs for U+0647 HEH and U+06BE HEH DOACHASHMEE

HEH				
KNOTTED HEH				

The table below also takes into account the compatibility block:

USV	Source for shapes	Isolate	Final	Medial	Initial	All forms
<b>Dual-joining</b>						
U+0647 HEH	from Arabic compatibility block	ه	ه	ه	ه	ه ه ه
U+06BE HEH DOACHASHMEE	common shape	ه	ه	ه	ه	ه ه ه ه
	from Arabic compatibility block	ه	ه	ه	ه	ه ه ه ه
KURDISH HEH		ه	ه	ه	ه	ه ه ه ه

The "knotted heh", or *heh doachashmee* is visually most similar to the Kurdish heh. However, this graphic does not show the complete picture. U+06BE is generally used for aspirated consonants, and the Kurdish heh is used as a normal heh, not for aspiration.

### Consider the Behdini Kurdish heh to be a glyph variant of U+0647 HEH

In order to distinguish the final form of U+0647 and U+06D5, Behdini Kurdish has long (20+ years) used a glyph variant for *heh* (U+0647), 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	ه ه ه	ه ه ه ه

Pournader [4] also mentions Behdini Kurdish and Kew’s [2] document in his paper “The right hehs for Arabic script orthographies of Sorani Kurdish and Uighur.” 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.

This solution would be attractive to the current users, because existing documents would not need converting to a new codepoint. Since the *sound* for the character is equivalent to the normal usage for U+0647, it makes perfect sense to use U+0647 as the codepoint.

These forms that Kurdish uses (for U+0647) do exist in Arabic calligraphy, and people would recognize it, although they might think it is odd.



### Outcome with this option:

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 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.

Table 9.x Behdini Kurdish heh design

Common Shape	Behdini Kurdish Shape
	

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.

### Consider the Behdini Kurdish heh to be a glyph variant of U+06BE HEH DOACHASHMEE

This second option might be an attractive solution because the shapes for the Kurdish heh may be most similar to *heh doachashemee*. Additionally, the Unicode spec does say “The exact contextual shapes of the letter depend on the language and the style of writing.” However, U+06BE is generally used for aspirated consonants, and the Kurdish *heh* is used as a normal *heh*, not for aspiration. Additionally, this solution would require all Behdini Kurdish documents to change the existing codepoints from U+0647 to U+06BE.



This solution has no advantage over solution one. Both require language context sensitive font features, which have very little economic incentive for implementation.

Outcome with this option:

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

U+06BE ARABIC LETTER HEH DOACHASHMEE is used to represent any *heh*-like letter that appears with stems at both sides in all contextual forms. The exact contextual shapes of the letter depend on the language and the style of writing. The forms shown in *Table 9-8* for KNOTTED HEH are used in certain styles of writing in South Asia. Other South Asian styles may use different medial and final forms. The style used in China and Central Asia for languages such as Uyghur uses medial and final forms for HEH DOACHASHMEE that are visually similar to the medial form of HEH shown in *Table 9-8*. However, for Behdini Kurdish the final form is knotted.

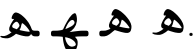
Table 9.x Behdini Kurdish HEH DOACHASHMEE design

Common Shape	Behdini Kurdish Shape
	

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.

### Encode a Behdini Kurdish heh

The third option is to encode a new *heh* for Behdini Kurdish. Pournader [3] says “The Unicode Arabic model has been following a semi-visual model, where a combination of the written shape and semantics of a letter or word is encoded, as opposed to its pure phonemics or semantics.”

If we are to follow this, then a new *heh* should be encoded with the following shapes: .

Outcome with this option:

### Unicode character Properties

```
08xx;ARABIC LETTER KURDISH HEH;Lo;0;AL;;;;;N;;;;;
```

Or

```
08xx;ARABIC LETTER BEHDINI KURDISH HEH;Lo;0;AL;;;;;N;;;;;
```

### Joining type and group for ArabicShaping.txt

```
08xx; KURDISH HEH; D; KURDISH HEH
```

### Annotations

The following annotations are recommended for NamesList.txt:

```
08x ARABIC LETTER KURDISH HEH
```

```
* Behdini Kurdish
```

### Normalization and Confusability Issues

Since the *kurdish heh* is not considered the same character as U+0647, there should not be any decomposition nor normalization issues.

ARABIC LETTER KURDISH HEH is confusable with U+0647 (initial and medial) and with U+06BE (isolate, initial and final) because the shapes in the Kurdish Heh are the same as for common variants for U+0647 and U+06BE.

### Suggested Collation

Suggested collation is for the KURDISH HEH to come after HEH (U+0647).

### Core Specification

The KURDISH HEH would need to be added to “Table 9-8. Dual-Joining Arabic Characters”.

### Opinion on above options

The author is grateful to Jonathan Kew for a review of this document. Kew was quite active between 2001-2003 in encoding Arabic script characters into Unicode. He was either the author of all or most of the proposals for the characters encoded in the **Arabic Supplement** block. These are his comments.

Regarding the Kurdish */heh/*, I would stand by my earlier position that it should be considered a stylistic preference for U+0647. I think this makes the most sense because it is clearly the equivalent letter/phoneme to the standard */heh/* in other languages, and the desired shapes are recognizable as known glyph variants of that letter. It is well understood that */heh/* takes on many forms depending on the specific calligraphic tradition (and even the author’s individual whim); what we’re seeing here is simply a preference for one particular style of writing this letter.

I don’t think disunifying Kurdish */heh/* from Arabic/Persian/etc */heh/* (whether by adopting U+06BE instead, or by encoding a new character) is justified unless we were to decide to disunify the various calligraphic styles of Arabic (Naskh vs Nastaliq vs Ruq’ah vs Kufi etc etc) – and that way lies madness. We haven’t disunified Fraktur from Latin, last I checked; neither should we disunify different stylistic variants of Arabic letters. But we can advise that certain glyph variants are preferred in certain languages/locales.

Besides the fact that I don’t think there is a strong case **for** disunifying Kurdish */heh/* from Arabic */heh/*, I would suggest there is a definite case **against** encoding a new letter. This would just add to the confusion facing every additional language community that has a */heh/* letter, and has developed a convention or preference for a particular glyph variant. And it would be unavoidably confusable in many contexts with one or other (or several!) of the existing characters.

The author is grateful for input from Roozbeh Pournader, Kamal Mansour, and Jonathan Kew into reviewing different versions of this document and their assistance is thinking through these complicated issues.

## Samples

خال ، خال دئ بوته گهلهك باشتر چيكم و تۇزا  
كهتیه سهر.....

-باشه بهس یی هشیار بی تبلین ته یین پیس ب  
گۆرا من نهكهفن.

-بلا خال ل سهر ههردوو چاقین خو..

-ها ها چاقین ته ئهوین وهکی یین سفوران، تو  
دچییه فوتابخانی؟

- بهلی خال ، ل ریزا سیی .

-ئهفه جو کویسهلهیه دگهل ته ؟

-ئهو نه کویسهلهیه خال ، ئهو برایی من یی  
بجویکه، بیچهك دفنا خو خوراند، بی من یی چهپی  
کورتیه ژیی راستی وهختی دزقرمهفه مال نهشیم  
سندوقکا خو بهلگرم ئهو هاری من دکته.

پری قهرهبالغه ، هندهك زهلامین ب چهك ژ وئ  
ترؤمبیلی هاتنه خار بهری خودانا بازنگان پهییا  
ببیت زهلامهکی دهرگهه بو فهكر، دهنگی شیلکهكا  
گۆللهیان ژدوویر دهات، هندهك گهنجا پشتا خو یا  
دایه دیوارهکی و ئیکی زنجیرهكا د دهستی داو ههر  
ییل دۆر تبهلهكا خو دزقرینیت و بهر ددهت ، سهری  
وی وهکی ئامیرهکی راداری ههری دزقریت چهپ و  
راست ، جار دهنگی کاویس ئاغای و ئافزا لو خالو لو  
خالو، جار یی داخل هسهن ، جار مافی مافی و  
جار گولهشهنگهه، جار ئه یاز یی هاوار دکته  
چیکهه ئهز ؟ ژ وئ تۆمارگهها ل دووماهیكا پری  
یی بلند دبیت.

سندوقکا وی یا بویاغی یا ل ئاخى دخشییت و ملی  
وی یی چهپی یی خار کری گههسته بهرانبهر  
قۆرکین چیچویی جایی.

خودانی زکی مهزن هئشتا یی قورجکا ل جایی د

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

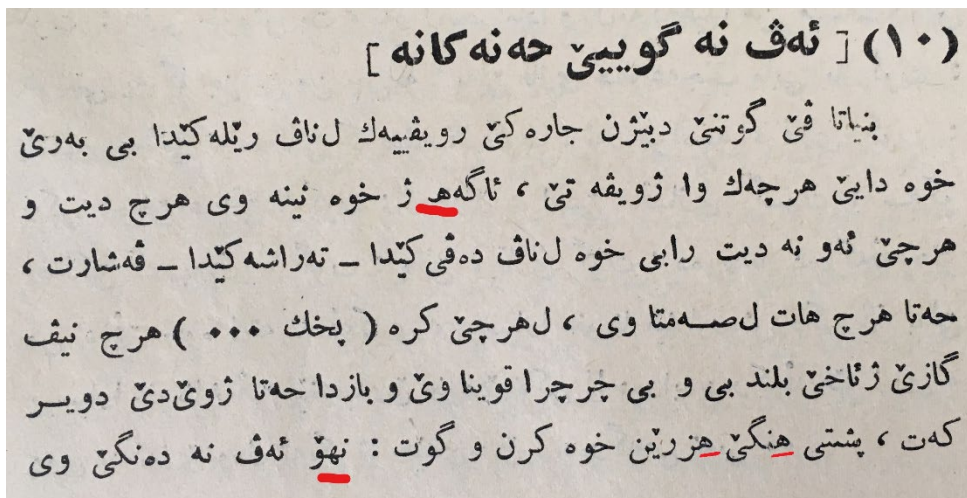


Figure 2: ARABIC LETTER HEH (isolate, initial and medial). Behdini Kurdish language. [kmr-2].

سیه سالی لگوندا

Figure 3: 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. 2012. Initial and medial forms of Arabic Letter Noon Ghunna (L2/12-381 <http://www.unicode.org/L2/L2012/12381-dotless-noon.pdf>)
- [4] 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>
- [5] 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/>

## 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