Fixing the Kashmiri Yeh

To: Unicode Technical Committee

Author: Manish Goregaokar (मनीष गोरेगांवकर) and Roozbeh Pournader (روزبه پورنادر)

Date: 2024-06-23

This document proposes solutions to the issues around Kashmiri Yeh described in L2/23-273R.

Proposal

1. Change the representative glyph for U+0620 ARABIC LETTER KASHMIRI YEH from ي to ی.
2. Change the Joining_Group for U+0620 from YEH to a new joining group KASHMIRI YEH, and its schematic name from DOTLESS YEH WITH SEPARATE RING BELOW to KASHMIRI YEH. Here is the proposed ArabicShaping.txt line:

   ```
   0620; KASHMIRI YEH; D; KASHMIRI YEH
   ```

3. Change the Core Specification as follows:

   a. In Table 9-7, add a row with the following information:

KASHMIRI YEH	ی	ی	ہ	ہ	

   b. Change the row for Kashmiri Yeh in Table 9-10 as follows:

U+0620 KASHMIRI YEH	KASHMIRI YEH	ی	ی	ہ	ہ

   c. Change the existing paragraph in Section 9.2.4 about the Kashmiri Yeh from:

   > U+0620 ARABIC LETTER KASHMIRI YEH is used in Kashmiri text to indicate that the preceding consonantal sound is palatalized. The letter has the form of a *yeh* with a diacritic small circle below. It has the YEH joining group, with the shapes as shown in the fifth row of *Table 9-10*. However, when Kashmiri is written in Nastaliq style, the final and isolated forms of *kashmiri yeh*
   >
   > usually appear as truncated *yeh* shapes (ی) without the diacritic ring.

   to:

U+0620 ARABIC LETTER KASHMIRI YEH is used in Kashmiri text to indicate that the preceding consonantal sound is palatalized. The letter has the form of a *yeh* with a diacritic small circle below in initial and medial forms, but its final and isolated forms appear as truncated *yeh* shapes without the diacritic ring. It has a joining group of its own, KASHMIRI YEH, with the shapes as shown in the fifth row of *Table 9-10*, as well as *Table 9-7*. (The Unicode Standard, prior to version 16.0, had specified different shapes for this letter when written in Naskh and Nastaliq styles, which was a mistake.)

Background

A more complete background may be found in [L2/23-273R](#).

Kashmiri is an Indo-Aryan language that has traditional orthographies in multiple scripts, including Perso-Arabic. The Perso-Arabic orthography for Kashmiri is structured similarly to that of Urdu with some additional characters, including an additional "letter" to mark palatalization.

This character was proposed in [L2/09-215](#), subsequently encoded as [U+0620](#) ARABIC LETTER KASHMIRI YEH in Unicode 6.0.

Kashmiri uses a number of code points called YEH, all considered distinct with distinct purposes:

	Final	Medial	Initial	Isolated
[U+06CC](#) ARABIC LETTER FARSI YEH	ى	یـ	یـ	ی
[U+0620](#) ARABIC LETTER KASHMIRI YEH	ؠ	ـؠ	ؠـ	ؠ
[U+06D2](#) ARABIC LETTER YEH BARREE	ے			ے

Note that the final and isolated forms of FARSI YEH and KASHMIRI YEH are similar but KASHMIRI YEH has a much shorter swash.

[U+0620](#) ARABIC LETTER KASHMIRI YEH currently looks like this in various fonts:

U+0620	Final	Medial	Initial	Isolated
Noto Nastaliq (Google)	و	ﯛ	ﯩ	ﯼ
Scheherazade New (SIL)	ى	ﯛ	ﯛ	ى
Noto Sans (Google)	ﯤ	ﯣ	ﯢ	ﯡ
Amiri (Khaled Hosny)	ﯤ	ﯣ	ﯢ	ﯡ

The form of KASHMIRI YEH with the ring is usually called "Gol Ye" ("round ye"), whilst the form with the shorter swoosh and no ring is called "Taler Ye" ("half ye").

Most Naskh fonts use the Gol Ye form in final position. Kashmiri is typically written in the Nastaliq style where it seems like Taler Ye is consistently used at the end of the word. This is in line with the proposal in L2/09-215, which suggests that Naskh and Nastaliq differ on how they ought to render this character in final form. This recommendation has made it into Chapter 9 of the Unicode Standard:

Table 9-10. Forms of the Arabic Letter *yeh* (Continued)

Character	Joining Group	X_n	X_r	X_m	X_l
U+0620 KASHMIRI YEH	YEH	ﯼ	ﯤ	ﯛ	ﯛ

Unicode also has a character U+06CD ARABIC LETTER YEH WITH TAIL for use in Pashto (since 1.1), only found at the end of words for the feminine marker. This was not encoded for Kashmiri.

	Final	Isolated

Noto Nastaliq	ک	کی
Noto Sans	ے	ی

Unicode also has a character [U+06C1](#) ARABIC LETTER HEH GOAL, used in Kashmiri (and Urdu, and Sindhi, and many other languages), used to represent an /h/ sound. It is not a type of YEH; it is a different letter, however the final form is *somewhat* similar-looking to Taler Ye.

	Final	**Isolated**
Noto Nastaliq	ہ	ہ
Noto Sans	ہ	ہ

As documented in [L2/23-273R](#), the situation appears to be that:
- Native users of the orthography consistently recognize the Gol (with circle) form in word-final position as a mistake, regardless of whether the font is Naskh or Nastaliq. This usage is mostly only attested in situations where people are using Unicode, or in Unicode documents.
- A lot of extant digital Kashmiri text uses [U+06CD](#) ARABIC LETTER YEH WITH TAIL or sometimes [U+06C1](#) ARABIC LETTER HEH GOAL to represent Taler Ye at the end of words. Native users consistently recognize this as a hack.
- Users are split on whether Taler Ye is allowed in word-medial position, a situation that is attested in some documents. Academicians seem to be convinced that these cases are nonstandard orthography or mistakes, and unnecessary from a phonetic or phonemic point of view. It is definitely not a *consistently* found orthographic feature, and even when found it is relatively rare.
- Isolated forms of Gol Ye are found only in nominative and pedagogical contexts. Quite often this is

  done in a modified YEH BARREE form , often paired with an alef (اے).

This evidence strongly points to a reason to stop recommending Gol Ye ever being rendered word-finally. There are weak reasons to disunify the characters due to the attestation of word-medial Taler Ye, but its disputed status and generally marginal occurrence makes it relatively okay to require a ZWNJ to force.

This evidence does not provide a conclusive direction on which isolated form to pick, however since an isolated KASHMIRI YEH may occur word-medially after a right-joining letter, the isolated form *must* be a Taler Ye.

Recommendation

The current situation is clearly suboptimal for Kashmiri users. Most fonts show an incorrect form of the palatalization mark at the end of words, and Kashmiri users have been reaching for a different also-incorrect-but-not-as-bad encoding to remedy this.

The Unicode Standard and software across the board should be updated to attempt to remedy this.

The Unicode standard

(A precise recommendation for the standard can be found in the Proposal section of this document)

Table 9-10 in the Unicode standard should be updated to use the Taler Ye form in isolated/final positions. The text in Chapter 9 of the Unicode Standard should similarly be updated to say that the Taler form is consistently found word-finally regardless of font, and potentially mention the previous mistaken recommendation.

The code chart should also be updated with a new glyph that is a Taler form.

L2/23-241 already proposed annotations for U+0620 ARABIC LETTER KASHMIRI YEH and U+06CD ARABIC LETTER YEH WITH TAIL, accepted by 177-A50.

A new Joining_Group for KASHMIRI YEH will need to be introduced, and U+0620 should be updated to use it.

It may also make sense to introduce a variation selector combination that makes KASHMIRI YEH render as the attested pedagogical form for Gol Ye (ﴰ). But we are not proposing that in this document. Further research and discussions are needed to determine the best course of action for representing the form.

Fonts

Fonts will need to be updated to align with the new core spec recommendations: They should consistently produce final and isolated KASHMIRI YEH as Taler Ye.

Keyboards

Keyboards should remove any YEH WITH TAIL key and instead maintain a single KASHMIRI YEH key, rendered in a font that shows it as Taler Ye (Showing Gol Ye leads to people reaching for other characters to render Taler Ye).

Other potential UI improvements would be to show both forms on the key (و/ؠ), or have a separate key (with identical input) that has a Gol form.

Data

Systems containing Kashmiri text should replace YEH WITH TAIL with KASHMIRI YEH.

Word-medial Taler Ye users

People wishing to use Taler Ye word-medially may use ZWNJ to force the form.

Further reading

- [L2/23-273R](#) Report on Kashmiri Yeh
- [CLDR-16595](#) Multiple issues in CLDR ks.xml
- [L2/23-241](#) Annotations for Kashmiri Yeh
- [L2/09-215](#) Proposal to add two Kashmiri characters

- [L2/10-169](#) Comments on the Proposed ى Arabic Letter Kashmiri Yeh

- [Kashmiri orthography notes](#)
- [W3C alreq #249](#) Lack of font support for Kashmiri characters makes text deviate from true semantics