

Subject **Handling of the Heh in Sindhi Text**
From Kamal Mansour
Date 10 May 2023
 Supercedes document L2/22-053

Summary

This paper continues the discourse begun by Lorna Evans in July of 2021 (L2/22-052) regarding the use of the various Heh letters in Sindhi. After much research and consultation of manuscripts and printed documents in the Sindhi language, it has become apparent that its Perso-Arabic orthography has never reached a formal standardization for the representation of its various *h*-sounds. During the era of metal type, most Sindhi type was adapted from typefaces originally designed for the Urdu or Arabic languages. Consequently, the broader character set for Sindhi was always presented with certain inconsistencies and compromises, some of which have been carried forward into the digital domain. Nowadays, even public authorities in Pakistan publish documents with inconsistent spelling for the sizable population of around 30 million Sindhi speakers. This paper proposes a standard representation for the various Heh letters that is consistent with the phonetics of Sindhi. The Appendix provides a summary of the methodology that led to the final recommendations.

Identifying Requirements

In contemporary practice, many Sindhi texts represent the *h*-sounds by using the closest visual match of the Heh-letter forms in a particular font. Unfortunately, the Heh family of letters in Perso-Arabic script is so rich in variant forms so that once text is set in different fonts, there is no guarantee that the originally intended form will be preserved. This leads me to conclude that, first of all, a consistent encoding in plain text must be established. Secondly, conformant fonts must be built so that the default letter shapes satisfy the needs of the Sindhi language, with support for other

regional languages as a secondary consideration. Finally, Sindhi is typically set in the Naskh style in spite of the prevalence of Nastaliq style for other languages in the South Asian region. Hence, shape variants in Sindhi fonts should be consistent with Naskh style.

As confirmed and named by Jokhio, 3 distinct types of *h*-sounds occur in Sindhi [Jokh20]:

1. *Normal or pronounced* (ملفوظي) **h**, as in شھرُ [ʃəhərʊ]; similar to the first sound of English *hat* or Arabic هام [ha:m]
2. *Aspirated or embedded* (وسرگي) **h**, as in تانگھو [t̪aŋgʰo]; same as the aspiration in Urdu گھی [gʰi]
3. *Weak or hidden* (مختفي) **h**, as in the final sound of بہ [bɪh]; similar to the final sound in Arabic فقه [fiqh]

This **h** typically occurs at the end of a word and sounds like a waning release of air.

Since these 3 *h*-sounds can appear in similar contextual positions, each of them must be encoded uniquely to prevent ambiguity. To satisfy this essential requirement, I recommend assigning each *h*-sound to a unique Unicode character:

Normal or pronounced h: U+0647 (Arabic Letter Heh)

Aspirated or embedded h: U+06BE (Arabic Letter Heh Doachashmee)

Weak or hidden h: U+06D5 (Arabic Letter Heh Goal)

In typography, the contextual shapes for each Heh-letter should conform to the following patterns (*i.s.* stands for “isolated shape”):

Normal Heh (U+0647) هي کنھن بہ شاھ (*h i.s.*)

Aspirated Heh (U+06BE) ماڳھي پڙھن ساڙھ ٻاجھ (*h i.s.*)

Weak Heh (U+06D5) نہ (*h i.s.*)

The letter shapes in the above examples conform to the recommendations by Evans in L2/22-052. Based on these samples, we observe that the initial and isolated variants of the Normal and Aspirated Heh share the same shapes, leading to ambiguity in certain contexts. Some would argue that the context, in many cases, can resolve the ambiguity because the Aspirated Heh takes those shapes only in combination with the less common set of aspirated consonants such as **ڙ ڙو ڙڙو**. To improve the clarity of Sindhi orthography in the long run, I would like to propose two alternative ways of eliminating this ambiguity.

First Approach

To distinguish between the two Hehs, the initial and isolated glyph shapes of the Aspirated Heh could be supplemented with a mark (e.g. U+0658 or U+0307) to distinguish it from the Normal Heh:

ساڙھ پڙھڻ ساڙھ پڙھڻ

Second Approach

Alternatively, a unique shape could be assigned to the initial and isolated glyphs of the Aspirated Heh. In the Thuluth style of Arabic script, the initial shape of the Heh can resemble the medial shape (see figure 1), as in **گهي**. This shape can be adopted for the initial glyph as in **هي**, and in a similar manner for the isolated glyph as follows **ه**.

While either of these approaches can effectively eliminate the inherent ambiguity, the first one requires less effort in design, while the second is clearly distinct.

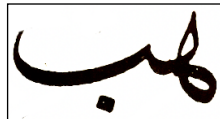


Figure 1

The recommendations I present below are based on a lengthy investigation that included many exchanges and discussions with Abdul-Majid Bhurgri, who supplied invaluable information about the historical background of Sindhi typography. According to Bhurgri, the main reason for the current lack of standardization is that metal type in the 20th century was never sufficiently customized for the requirements of the Sindhi language; instead, sorts prepared for Arabic and Urdu were utilized for Sindhi with the consequent ambiguities. Even after the transition to digital type, no standard has been issued about the unique mapping of the *h*-sounds to specific characters. Part of the problem of seeking guidance in historical documents is that conventions have changed over time.

Shapes for Heh characters in Sindhi

I am proposing the following character mapping, along with its corresponding contextual shapes:

1. **Normal h**

u+0647

Shapes (from right to left: isolated, initial, medial, final)



2. **Aspirated h**

u+06BE

Shapes (from right to left: isolated, initial, medial, final)



Any of the following letters can be followed by an aspirated h (Heh):

ل م ن گ ژ ڈ و ر ٹ ج

Since most of the letters above connect to the aspirate Heh, the initial and isolated shapes will only appear following letters ڈ و ر ٹ

When combined with the aspirated Heh, the above letters appear as follows:

ڈھ وھ رھ ژھ ژھا
لھت مھت نہت گھت ٹھت جھت
لھ مھ نہ گھ ٹھ جھ

Alternatively, the proposed Tuluth-style Heh can be adopted for the initial and isolated shapes:

ه

3. Weak or silent h

u+06D5

Shapes (from right to left: isolated, final)

ه ه

As this letter appears only at the end of words, there is no need for initial and medial shapes. Appropriately, the letter u+06D5 links exclusively to its right-hand neighbor.

Arabic citations embedded in Sindhi text

Finally, in addition to the Heh characters needed for the Sindhi language, any Arabic-language text cited in Sindhi text is also expected to be stylistically distinct. Traditionally, the Heh (u+0647) in an Arabic quotation is rendered in a pattern distinct from that of Sindhi. Typically, the shapes appear as follows:

Shapes (from right to left: isolated, initial, medial, final)



According to Bhurghri, the following contextual shapes are also acceptable alternatives for Arabic-language text:



Since the above shapes for Arabic Heh differ from those for Sindhi, future fonts must enable the user to switch readily between the two sets. More details follow below under *Font Considerations*.

Confusable Shapes

It is quite common to see the final form of the Sindhi aspirated Heh rendered as:



This visual presentation can be misleading because it appears to show a medial form of the aspirated Heh followed by a final form of the silent Heh. According to Bhurghri, this practice stems from the lack of Sindhi fonts that

include a tapered final form of the aspirated Heh (as in Figure 2). To avoid using an abruptly truncated medial form (Figure 3) in final position, the final form of the silent Heh was simply appended to simulate a smoother terminal. While this brute-force approach might satisfy a visual need, it also introduces a stray character into the text, resulting in unreliable searches.

Final Aspirated Heh



Figure 2

Medial Aspirated Heh



Figure 3

A font designed to meet the requirements of Sindhi must automatically render all the necessary shapes of the aspirated Heh.

Font Considerations

Most Arabic-script fonts today are designed primarily for the linguistic triad of Arabic-Farsi-Urdu, with secondary accommodations for other languages. In current computing environments (applications or operating systems) it is rare to find specific support for Sindhi, whether through locale setting or language-tagging. Because of Sindhi's particular typographic requirements, it is imperative to make available some fonts whose native state supports the required Heh-shapes for Sindhi orthography. Any accommodation for other languages, including Arabic-Farsi-Urdu, should be carried out by means of secondary selection of locale, language-tagging, or stylistic variation. Because Sindhi's requirements differ from those of other languages, the *neutral* (native) state of the font should serve the needs of the Sindhi language without the need for any additional choices on the part of the user. To accommodate Arabic citations embedded in Sindhi-language text, the text of the citation can be tagged as Arabic, or set in a stylistic style specifically designed for this purpose.

Conclusion

The consistent encoding of Sindhi-language text will encourage the growth of its digital patrimony and all the fields of study that it will subsequently enable. The current Unicode repertoire of Perso-Arabic characters has proved sufficient for Sindhi, but uniform usage will further improve the quality of Sindhi text. A successful outcome will eventually depend on the following steps and actions:

- a. Creation and adoption of a Unicode-based orthographic standard
- b. Conformant fonts that support Sindhi
- c. Digital keyboards, both web-based and for all digital devices
- d. Revision of archived data to conform to standard orthography

Appendix

The contemporary orthography of the Sindhi language is largely standardized except for the group of Heh letters. In order to distinguish between the various types of Heh as represented in text, I have referred to the Sindhi-English Dictionary [SED] as a source for sample words since it offers a phonetic transcription for each word in its collection. Using the transcription as a primary guide, one can collect a list of words whose Heh letters can be distinguished with a good degree of certainty. Occasionally, the degree of certainty can be even higher when two words differ only in an underlying *h*-sound such as when one is the *normal* (or pronounced) one, while the other is *aspirated*. In some cases, it was helpful to corroborate simple words by consulting a children's primer [PRI] in spite of its inconsistent orthography.

Once I had collected a set of words that represent various types of *h*-sounds, Abdul-Majid Bhurghri wrote them out by hand. We then reviewed the list together to determine the best typographic representation for each instance of a Heh letter. Such a review was needed because handwriting and typographic forms are not always consistent with one another.

The above recommendations are an amalgamation of the discussions with Bhurghri, the phonetic analysis by Jokhio [Jokh20], the Sindhi English Dictionary, as well as previous work by Evans and Kew [KWJ].

References

[Jokh20]

ڊاڪٽر الطاف جوڪيو. صوتيات جي اصطلاحن 'وسرڳ ۽ اوسرڳ' جو تنقيدي جائزو. جلد: ۱۲. شمارو: ۲۳ ڊسمبر ۲۰۲۰ ص: ۹-۲۱. وفاقي اردو يونيورسٽي آف آرٽس، سائنس ۽ ٽيڪنالاجي، عبدالحق ڪيمپس، ڪراچي.

(Translated) Jokhio, Altaf. "A Critical Review of Aspirated and Unaspirated Terms of Phonetics". *Biannual Research Journal*, Vol 12 (23). December 2020, pp. 9-21. Sindhi Department, Federal Urdu University of Arts, Science & Technology. Karachi, Pakistan.

[SED]

The Sindhi English Dictionary project was supervised by Dr. Sarmad Hussain, who heads the Center for Research in Urdu Language Processing (CRULP) <http://sed.cle.org.pk/>

[PRI]

Sindhi Primer

Retrieved from <https://www.taleem360.com/class-1-sindhi-primer-text-book-in-sindhi-by-stbb-pdf-format-fii>

[KWJ]

Kew, Jonathan. 2005. Notes on some Unicode Arabic characters: recommendations for usage (draft 2). SIL International.