Notes on some Unicode Arabic characters: recommendations for usage

Jonathan Kew

Draft 2 — April 21, 2005

Contents

1 Introduction

In certain cases, the Unicode standard encodes separate characters for forms that would be considered glyph variants of a single character in Arabic. While this is sometimes necessary, in order to support writing systems where the shapes are used contrastively, it also raises sometimes raises questions of which character to use, among several possibilities. This document discusses some of these situations, and attempts to offer guidance for implementers and users of the Standard.

To an Arabic reader, the glyphs ك, ک, and ڪ are all clearly recognizable as forms of the same letter, *kaf*. The first, ك, is typical of the designs seen in common text typefaces based on a simplified Naskh style of writing. ک is an alternate form that seems to be based on Nastaliq style, and ڪ is a swash form sometimes used, normally in initial or medial position, for stylistic effect or as part of line justification. Similarly, ي and ى are both *yeh*, the dots being optional.

However, as the Arabic script has been adopted and adapted for writing many other languages, these different shapes have sometimes been taken and used as distinct letters in such writing systems. Even where the alternate forms of a single Arabic letter are not used contrastively within a single writing system, the range of shapes that are recognized and accepted may be much more restricted than was the case with the original Arabic letter.

Note that this document does not discuss the "presentation forms" of Arabic letters. These are not recommended for encoding data; they exist only for legacy compatibility reasons. Thus, except where the context specifically refers to joining forms, references here to different "shapes", "forms", or "glyphs" for a given Unicode character are not referring to the initial, medial, and final linking forms, or to ligatures, but to different designs of the basic unjoined letter (and correspondingly different linked forms).

Not every character nor every language is discussed here (far from it); however, it is hoped that the principles used can be applied where similar encoding choices need to be made for other writing systems and additional letters.

Some of the recommendations given here are based in part on the presentation *Guidelines to Use of Arabic Characters* by Kamal Mansour at the 24th Internationalization and Unicode Conference (September 2003) in Atlanta, GA. Others are based on discussions with specialists studying various of the languages concerned, and on experience gained in implementing a variety of fonts and software systems.

2 KAF-based letters

Here, we consider the Unicode characters U+0643 ك, U+06A9 ک, and U+06AA ڪ, and other characters based on these forms. These are all forms of the Arabic letter *kaf*, written in different styles.

I am not aware of any language whose writing system uses both ك and ک contrastively; indeed, this seems highly unlikely, as in both initial and medial positions, their linked forms are the same: ك joins as ـكـكـك, while ک joins as ـکـکـک. On the other hand, ک and ڪ do occur together and must be distinguished; and in some writing systems, the default shape of U+0643 ك is not considered correct for *kaf*. Similarly, where the alphabet has been extended by the addition of dots or other marks to *kaf*, this may apply only to one specific shape of the letter.

2.1 Arabic

The Arabic letter *kaf* is encoded as U+0643 ك. Depending on the type design, and possibly other stylistic factors, this character might be rendered with forms more like ک or ڪ, but *kaf* in Arabic

should nevertheless always be encoded with U+0643. The selection of alternate glyphs would occur as a result of typeface choice, formatting processes, and higher-level protocols, without altering the encoded text.

In the absence of specific reasons to use a different *kaf* character, U+0643 should also be considered the default choice to encode the corresponding /k/ letter in other languages where the Arabic script is used. However, if the script has been adopted not directly from Arabic, but from another source such as Persian or Sindhi, the practices of that more immediate source should generally be considered first.

- use U+0643 ك for *kaf*
- U+06A9 ک and U+06AA ڪ should *not* be used for stylistic effect

2.2 Persian

In Persian (Farsi), the typical Arabic shape ك is not considered an acceptable form for *kaf*. The standard *Information Technology – Persian Information Interchange and Display Mechanism, using Unicode* (ISIRI 6219)[1] recommends the use of U+06A9 ک for Persian *kaf*, permitting both Arabic and Persian forms to co-occur in plain text without needing markup or other higher-level protocols to distinguish the two.

While the recommendation is to use U+06A9 ک for *kaf* when encoding Persian text in Unicode, users should be aware that there is likely to be a considerable amount of Persian text where U+0643 ك is used, making no distinction from Arabic *kaf*. In many cases, Arabic fonts have been "adapted" for Persian by simply changing the glyph at U+0643 (and its corresponding final form), to obtain the correct Persian appearance with software systems (keyboards, mappings from legacy codepages, etc.) that were designed for Arabic.

Therefore, while *producers* of Persian text should use U+06A9 ک for *kaf*, it may be advisable for *consumers* of Persian text data, especially if accepting input data from arbitrary sources, to recognize U+0643 as well, perhaps offering an option to remap this code to U+06A9 if appropriate.

- use U+06A9 ک for *kaf*
- U+0643 ك for *kaf* may be encountered in data

2.3 Urdu

Urdu tends to follow Persian writing conventions more closely than Arabic, and in particular the shape ک is clearly the preferred *kaf*, with ك being viewed as Arabic and "foreign". This preference probably arises because Urdu is almost universally written in Nastaliq style script, where the form of *kaf* resembles ک (even when the language is Arabic); however, in Urdu the preference is so strongly established that ك would be considered incorrect even in non-Nastaliq styles, rather than being seen as dependent on the style in use. (The history is probably similar for Persian, which also has a long tradition of Nastaliq calligraphy, even though that style is less widely used now.)

The same encoding recommendation therefore applies for Urdu as for Persian:

- use U+06A9 ک for *kaf*
- U+0643 ك for *kaf* may be encountered in data

[1] See http://www.farsiweb.info/standard/; note that the document is in Persian.

2.4 Sindhi

The Sindhi language has a contrast between unaspirated and aspirated consonants. When the Arabic script was adopted and extended to write Sindhi, the form ک was used to represent an aspirated velar consonant /kh/, while the form ڪ was used for the unaspirated /k/. The form ك is not used in writing Sindhi.

To encode Sindhi, then, the two Unicode characters U+06AA ڪ and U+06A9 ک should be used for /k/ and /kh/ respectively. It is probably less likely that U+0643 will be found in Sindhi data than in Persian or Urdu, as Sindhi does not have the same history as Persian and Urdu of legacy implementations based on slightly-extended Arabic systems with a few glyph changes. If it does occur in Sindhi text, it will most likely be representing /kh/ (properly encoded as U+06A9), as in some positions these share similar glyph shapes.

(It may be interesting to note that the Unicode character name of U+06A9 ک ARABIC LETTER KEHEH looks like an attempt to indicate in transcription the *aspirated kaf* sound of Sindhi. This supports the view that this character was encoded, perhaps originally in a legacy codepage, specifically for the contrastive Sindhi /kh/ usage where ك is not a recognized form.)

- use U+06A9 ک for aspirated *kaf* /kh/
- use U+06AA ڪ for unaspirated *kaf* /k/
- U+0643 ك should not occur, but probably represents /kh/ if encountered in data

2.5 Jawi (Malay) *gaf*

Malay written in Arabic script (known as Jawi) uses a *kaf* modified by the addition of a dot above to represent a voiced consonant /g/. This could be encoded using U+06AC ڬ, and indeed the Names List annotation found in Unicode versions up to 4.0 suggests this. However, old Malay sources consistently write this character as ک, using the Persian *kaf* as a base and not the Arabic *kaf*. This is true even where the Malay sources use ك for *kaf*, and applies to both printed and hand-written materials. The form ڬ does not appear to be a legitimate rendering of Jawi *gaf*.

The strength of the preference for the shape ک rather than ڬ may be gauged from the fact that some writers, faced with computer systems that only provided U+06AC ڬ, have used this character but added a *kashida* (extender) character after it in final or isolated position, in order to get a printed result such as ڬـ. Although this is typographically quite unsatisfactory, it has been preferred over the ڬ shape.

It is therefore recommended that Jawi *gaf* be encoded as U+0762 گ (newly added in Unicode version 4.1); the use of U+06AC ڬ is *not* recommended, though it may be found in some existing text data, especially in view of the fact that in Unicode versions prior to 4.1, U+0762 گ was not encoded. The character U+06AC should be used only for languages where its nominal form ڬ would be an acceptable, recognized way to write the relevant letter.

- use U+0643 ك for *kaf*
- use U+0762 گ for *gaf*
- U+06AC ڬ for *gaf* may be encountered in existing data

2.6 Moroccan Arabic *gaf*

Like Malay, Moroccan Arabic adds a *gaf* letter to the standard Arabic alphabet. In this case, it is written as a *kaf* with three dots above. However, like the Jawi (Malay) case, the base form used is consistently ک and not ك, even though the ك shape is used for *kaf*. Just as with Malay, there are

sources that show authors deliberately using a final *kashida* to force a ك-shaped character to take on the shape ک, so strong is the feeling that ڭ rather than ڬ is the correct form for *gaf*.

Therefore, it is recommended that Moroccan *gaf* should be represented as U+0763 ڭ (new in Unicode version 4.1), and not as U+06AD ڬ. The latter should be used for languages where a ك-based shape is accepted, as appears to be the case for Uighur *eng*, for example.

- use U+0643 ك for *kaf*
- use U+0763 ڭ for *gaf*

2.7 Uighur, Kirghiz and Kazakh *eng*

In a number of Central Asian languages, the letter *eng* (a velar nasal) appears to be an example where an extended Arabic letter, constructed as a *kaf* with three dots above, has been based directly on the traditional ك shape. Although these writing systems use the Persian and Urdu form U+06AF گ for *gaf*, which is based on a ک shape, the *eng* is typically written as ڭ and should therefore be encoded as U+06AD.

- use U+0643 ك for *kaf*
- use U+06AF گ for *gaf*
- use U+06AD ڭ for *eng*

3 HEH-based letters

Like *kaf*, the letter *heh* also has several variant glyphs. The nominal character may be written as either ه or ھ. In this case, there are also variants of the associated joining forms; thus, a document where isolated *heh* is written as ه might use either ﻫ or ﻬ as the medial form, and any of ﻪ, ﻩ or ﻬ as the corresponding final form. In some cases, these variants may be freely interchangeable at the discretion of the writer or type designer, but in some languages and for some extended Arabic letters based on a *heh* form, there are specific requirements concerning the shapes used.

Unicode encodes ARABIC LETTER HEH as U+0647 ه, with alternate possibilities including HEH DOACHASHMEE at U+06BE ھ, HEH GOAL at U+06C1 ہ, and LETTER AE at U+06D5 ە. In addition, there are various modified forms of some of these, with dots or other marks added.

To better distinguish among these *heh* letters, we also consider typical linked glyphs used for each character (remembering that there may be variation in these):

USV	*Name*	*Isolate*	*Final*	*Medial*	*Initial*
U+0647	ARABIC LETTER HEH	ه	ﻪ	ﻫ	ﻫ
U+06BE	ARABIC LETTER HEH DOACHASHMEE	ھ	ﻬ	ﻬ	ﻬ
U+06C1	ARABIC LETTER HEH GOAL	ہ	ﻩ	ﻩ	ﻩ
U+06D5	ARABIC LETTER AE	ە	ﻪ		

Note that U+06D5 ە differs from the others in that it is right-linking, not dual-linking. It is thus similar to a sequence <*heh*, *zwnj*>, although there is no such formal equivalence defined via a canonical (or even compatibility) decomposition.

3.1 Arabic

Arabic has just one *heh* character, encoded with the basic Arabic block code U+0647 ه. This character is sometimes written as ھ, particularly when used as part of the *abjad-hawaz* numbering system using Arabic letters; and in medial position it may appear as either ﻫ or ﻬ depending on the style of

script. Both forms may occur even within the same line of text, at the discretion of the calligrapher. Nevertheless, these are not regarded as different characters, but merely as glyph variants of the same character, U+0647.

There is also the feminine ending written as ة, a form based on *heh* despite its name TEH MARBUTA. This has its own Unicode value, U+0629. Although it could be considered to be a form of *teh*, this is not reflected in the encoding, as the written form is different, and Unicode encodes the script, not the language.

- use U+0647 ه for *heh*
- other *heh*-like characters should not occur

3.2 Persian

Persian uses the same letter *heh* as Arabic, and has no reason to encode it differently.

A uniquely Persian usage is the form ۀ, a letter *heh* with *hamza* above, used in compound words (*izafet*). There is a Unicode character U+06C0 ۀ that has this appearance. However, this is defined as canonically equivalent to the sequence <U+06D5 ه, U+0654 ◌ٔ>, and *not* to a sequence involving the normal *heh*, U+0647 ه. As Persian is considered to have only one *heh*, the ISIRI 6219 standard specifies that U+06C0 ۀ is *not* to be used in Persian text; rather, *izafet* is encoded as <U+0647 ه, U+0654 ◌ٔ>, with a zero width non-joiner if necessary to prevent linking to a following letter.

- use U+0647 ه for *heh*
- use <U+0647 ه, U+0654 ◌ٔ, U+200C ZERO WIDTH NON-JOINER> for *izafet*
- other *heh*-like characters should not occur
- in particular, do *not* use U+06C0 ۀ for *izafet*

3.3 Urdu

In Urdu writing, different forms of *heh* are used to show the distinction between an /h/ consonant and aspiration of a plosive. Aspiration is written with the double-looped *doachashmee* form ھ (sometimes also referred to as a *butterfly* or *knotted heh*), while the separate /h/ consonant is most commonly written with non-looped forms such as ہ ہـ ـہ. As this semantic distinction is required as part of the Urdu writing system, two separate *heh* characters must be encoded.

For aspiration, Unicode provides U+06BE ھ, encoded for this exact purpose.

The /h/ consonant is more problematic. The Unicode character U+06C1 ہ provides by default the typical shapes expected for Urdu *heh*, and was apparently added to the standard for this purpose. (The name HEH GOAL is a rough transcription of the Urdu expression ہے گول (round *heh*), indicating its origin as a character intended for Urdu usage.) However, the Urdu *heh* is not really a different letter from the Arabic or Persian one; it is the same letter. It is merely written (usually) with a different choice among the possible glyph variants of that letter, the "Urdu shapes" being more typical of Nastaliq script, the calligraphic style normally used for Urdu. For consistency across the languages that use Arabic script, therefore, it is preferable to use the standard *heh* character, U+0647 ه, and treat the preferred forms for Urdu as a language-specific font variant.

- use U+06BE ھ to form aspirate digraphs (بھ, پھ, دھ, etc.)
- use U+0647 ه for *heh*, with joining glyphs similar to those seen for U+06C1: ہ ہـ ـہ
- do not use U+06C1 ہ, which is a glyph variant of U+0647
- treat U+06C1 as equivalent to U+0647 if found in Urdu data

3.4 Sindhi

Like Urdu, Sindhi also uses a *doachashmee* form of *heh* to write some aspirated consonants (specifically جھ and گھ; there are unique letters for other aspirates). Unlike Urdu, the normal /h/ consonant is also commonly written with a ھ form; however, some writers, at least, make a distinction between the letter *heh* /h/ and aspiration by writing the medial form of *heh* as ھـ while using ھـ for the aspirate digraphs. Thus, جھ (initial aspirated /jh/) contrasts with جھ (initial /j/ followed by medial /h/). Some writers of Sindhi script also prefer to use double-looped forms ھـ ھ rather than ہـ ہ in isolate and (more rarely) final positions.

- use U+06BE ھ to form aspirate digraphs جھ and گھ
- use U+0647 ہ for *heh*, preferring double-looped glyphs in most cases

3.5 Parkari

The Parkari language (and other related languages in the southern Sindh) uses a writing system based on Sindhi (see above); however, it has an additional *heh*-based letter, encoded as U+06FF ۿ. Font designers should note that the linked glyphs associated with U+06FF should all be based on a double-looped *heh*, thus: ۿ ۿ ۿ. For consistency, it is common to use the forms ھ ھ ھ ھ for the normal *heh*, U+0647.

Parkari also uses a final *heh* with the form ۾ when writing certain vowel sounds. This can be encoded using U+06C1 ہ *heh goal*, which is an appropriate choice here as this shape needs to contrast with the normal *heh*.

- use U+06BE ھ to form aspirate digraphs جھ and گھ
- use U+0647 ہ for *heh*, preferring double-looped glyphs ھ ھ ھ ھ
- use U+06C1 ہ in final position for the vowel written as ئہ <U+0626, U+06C1>
- Parkari also uses U+06FF ۿ

3.6 Kurdish

In Kurdish, there is a normal /h/ consonant written with the double-looped *heh* ھ, and in addition the /e/ vowel is written with the *heh*-based forms ہ ہ (similar to Parkari usage of the ۾ form). In Sorani, probably the most widely-written Kurdish dialect, /h/ does not occur in final position, and therefore there is no confusion between /h/ and this /e/ vowel, but in the Behdini dialect, it is important that final /h/ is distinguished from /e/. For this reason, writers use a double-looped final form of *heh* ھـ for the /h/ consonant.

Some implementations appear to have encouraged the use of U+06BE ھ for *heh* in Kurdish, because typical glyph shapes used for this character more nearly correspond to Kurdish preferences. However, this is not recommended, as it obscures the identity of the letter as the "normal" *heh* of Arabic script, and the shape of the final form of U+06BE (typically ھ) is still not in accordance with Behdini usage. It is therefore recommended that Kurdish *heh* should be encoded as U+0647 ہ, just as for other languages, with the preferred glyph shapes being provided as a language-specific font variant.

- use U+0647 ہ for *heh*, with preferred glyph shapes ھ ھ ھ ھ
- use U+06D5 ہ for the final /e/ vowel written as ئہ <U+0626, U+06D5>
- do not use U+06BE ھ for *heh*, but recognize that some data may include it

4 YEH-based letters

The letter *yeh*, like other Arabic letters, was originally written without dots, ى. Over time, the dotted form ي developed, representing the same letter. There is also an alternate form ے seen in some styles of writing. (Note that although the use of dots under the isolated and final forms of *yeh* may be optional, they are not optional in modern writing under the initial يـ and medial ـيـ forms.

The *yeh* shape ى is also used, optionally with a superscript *alef* (encoded as U+0670 ◌ٰ), to write a final *alef* in some instances. This usage, *alef maksura*, omits the dots and is regarded as a different letter than normal *yeh*.

4.1 Arabic

Although it may be written in various ways, Arabic has just one *yeh* letter and should encode it consistently using U+064A ي. Although the representative glyph shown for this Unicode character has dots below, it may be rendered as a dotless ى shape in some styles and locales (this would be expected in a Nastaliq font, for example); it could even appear as a ے shape in appropriate calligraphic contexts.

- use U+064A ي for *yeh*
- use U+0649 ى for *alef maksura* (only occurs word finally)
- do not use U+06CC ى or U+06D2 ے for *yeh*

4.2 Persian

In Persian, *yeh* is consistently written without dots, ى, in isolated and final positions. The dotted form that is the typical rendering of U+064A ي would be considered incorrect. To allow the Persian form ى to be reliably rendered, and to co-occur with the typical Arabic ي in the same text, Unicode provides a separate code U+06CC ى FARSI YEH. Although the Persian *yeh* is clearly related to the Arabic letter, and could have been considered a glyph variant of it, the difference in typical appearance is not an optional feature; it is a requirement for proper rendering of Persian text.

Note that it is possible to produce the appropriate *appearance* for Persian by the use of Arabic *yeh* U+064A ي in initial and medial positions, and *alef maksura* U+0649 ى in final and isolated positions. This represents an inconsistent encoding of the Persian letter, and should not be done; however, users working with systems designed for Arabic may have used such "hacks" to achieve the desired rendering.

There have also been cases where Arabic fonts have been altered to render U+064A with a dotless glyph, and U+0649 with a dotted one (in order to support the Persian preference for *yeh*, while still providing the possibility of rendering the Arabic form where needed). This swapping of glyphs represents a deviation from the Unicode standard, and leads to data interchange problems; it is not a correct way to encode Persian text.

When Arabic words spelled with *alef maksura* are used in Persian, it could be considered most logical to encode these with U+0649 ى, for consistency with Arabic. However, as U+0649 ى is indistinguishable from the Persian *yeh* U+06CC ى in word-final position (the only position *alef maksura* should occur, it is unlikely that users will make such a distinction.

- use U+06CC ى for *yeh*
- be aware that some Persian text may be encoded with U+064A and/or U+0649 for *yeh*

4.3 Urdu

In Urdu, a distinction is made between the form ی, representing an /i/ vowel, and ے, representing /e/. The two forms are known as چھوٹی یے *small yeh* and بڑی یے *large yeh* respectively, and are considered separate letters. Thus, unlike in Arabic, the form ے must be encoded separately from ی, not treated as an optional calligraphic glyph variant.

For چھوٹی یے *small yeh*, Urdu follows the Persian convention of writing without dots; the form ي would be considered incorrect. As with Persian, there is the possibility of encountering data where the Arabic *yeh* and/or *alef maksura* has been used to stand in for the correct form, possibly with modified fonts, so processes may need to be prepared for some inconsistencies in encoding.

- use U+06CC ی for *small yeh* /i/
- use U+06D2 ے for *large yeh* /e/
- be aware that U+064A ي and/or U+0649 ى may also be found

4.4 Sindhi

Although Sindhi is a neighboring language to Urdu, it follows the Arabic convention of writing *yeh* with dots, ي, rather than in the Persian and Urdu way. A dotless form is only seen when writing Arabic-derived words with *alef maksura*, not for the normal *yeh* letter. It therefore uses the same character codes as Arabic for *yeh*.

- use U+064A ي for *yeh*
- use U+0649 ى for *alef maksura* in Arabic-derived words

4.5 Kurdish

Kurdish practice appears to be to write *yeh* in the Persian style, without dots in final/isolate position, indicating that U+06CC ی is the preferred character. In addition, there is a version of *yeh* with a 'small v' mark above, encoded as U+06CE ۍ. Like the regular *yeh*, this character also acquires two dots below when it links to a following letter, so the joining forms are ۑ ۑ ۍ.

It should be noted that the 'small v' here is an integral part of the letter, not a vowel mark akin to *fatha*, etc. The 'small v' is analogous to the dots and other marks used to create new letters as modifications of a basic Arabic letter shape. Therefore, it would be inappropriate to represent this Kurdish letter as U+06CC ی *yeh* with the vowel mark U+065A ◌ۘ added.

I have also seen Kurdish texts that use a *yeh* with a horizontal bar above, ۓ, but do not have adequate information concerning this at present.

- use U+06CC ی for *yeh*
- use U+06CE ۍ for *yeh with small v*
- do not use <U+06CC ی, U+065A ◌ۘ> as alternative to U+06CE ۍ

4.6 Uighur

The Uighur writing system includes several unusual innovations. In addition to a standard Arabic *yeh*, written with dots as ي, it uses a form with two dots placed vertically below ې, and a dotless form ى. Unlike *alef maksura* in Arabic, the dotless form also occurs within words (not only in final position); and unlike the Persian *yeh*, it does not acquire dots when linked (as it must be distinguished from the normal *yeh*). Thus, there are three distinct *yeh*-shaped letters, each with a full set of joining forms.

The dotless and vertical-dotted *yeh* letters are typically preceded with a *hamza* when they occur in initial or final position, or at a syllable break, and may be shown in this way in sources that discuss the writing system. For example, the vertical-dotted *yeh* may be shown as ئې ـئې ـئې ې. This is, however, a spelling convention and not an inherent aspect of the shaping of the letter.

- use U+0649 ى (joining forms ـد ـى) for the dotless *yeh*
- use U+064A ي (joining forms ـيـ ـي) for the dotted *yeh*
- use U+06D0 ې (joining forms ـېـ ـې) for the *yeh* with vertical dots
- insert U+0626 ئ where a form with preceding *hamza* is required, e.g., ئدى is encoded <U+0626 ئ, U+0649 ى, U+062F د, U+0649 ى>
- do not use U+06CC ى (Persian *yeh*) for the dotless *yeh*

5 For font designers: summary of *heh* glyph variants

This document recommends that the "normal" Arabic *heh* character U+0647 ه should be rendered with different glyph shapes to match the requirements and expectations of particular languages and locales. To provide a convenient reference for developers of general-purpose Unicode fonts that cover the Arabic block and are intended to be widely useful, some appropriate language-specific renderings are summarized here.

The glyphs shown here are a typical "simplified Naskh" style, which forms the basis for most general-purpose fonts. In other styles such as Nastaliq, Ruqa, Kufi, etc., there may be significant differences to consider. This summary does not claim to be a substitute for in-depth knowledge of the particular writing traditions of different user communities; it is merely a guidepost intended to promote better consistency in type design and encoding practices.

Typical default shapes for…		*Isolate*	*Final*	*Medial*	*Initial*
U+0647	ARABIC LETTER HEH	ه	ـه	ـهـ or ـھـ	هـ
U+06BE	… DOACHASHMEE	ھ	ـھ	ـھـ	ھـ
U+06C1	… GOAL	ہ	ـہ	ـہـ	ہـ
U+06FF	… WITH INVERTED SMALL V ABOVE	ۿ	ـۿ	ـۿـ	ۿـ
Urdu					
U+0647	ARABIC LETTER HEH	ہ	ـہ	ـہـ	ہـ
Sindhi					
U+0647	ARABIC LETTER HEH	ھ	ـھ or ـه	ـھـ or ـهـ	ھـ
Parkari					
U+0647	ARABIC LETTER HEH	ھ	ـھ	ـھـ	ھـ
Kurdish					
U+0647	ARABIC LETTER HEH	ھ	ـھ	ـھـ	ھـ