

1. Pattern One

Perceptually identical letters Kaf, (K) Heh (H), Yeh (Y) have regionally preferred but typographically unstable shaping differences. Urdu Nastaleeq was the reason to introduce new code points for differences that are conditioned by script style. This KHY group is a monument of cultural tribalism and a breeding ground for spoofing.


B. Form: typically in nastaleeq and nastaleeq-derived fonts, the artificial regional difference between final Heh variants cannot be expressed, nor can those of Kaf be distinguished.

Arabic:

مكة

مکه In Farsi style (Arab perspective):

کے

 In Urdu style (Pakistani perspective):

کتابخانه

M		K		Y		H		total variants
م	0645	ك	0643	ي	064a	ة	0629	
		ك	06A9	ي	06CC	ة	06C3	
				ي	06Do	ه	06D5	
						ه	0647	
	1		2		3		4	24

[illegible]

2. Pattern two

Distinct but closely similar letters that are vulnerable to the combination of simplified typography and small sizes.

- a. Unfamiliar characters that look unsuspected to a specific audience in all positions.

example - GRG: خ vs. خ; in context: خرج vs. خرج - خرج vs. خرج

- b. Unfamiliar characters that look unsuspected to a specific audience in limited positions.

3. Pattern Three

The chief system fonts are so poorly designed, that the distinctive gap following discontinuous letters is hardly visible. In combination with a tall tooth (derived from Noon in Syriac) in such typefaces, the difference is dangerously low in the typical small size of a URL.

example - BS/D S: ذس vs. ذس - ذس vs. ذس

4. Pattern Four

Orthographic variation between spellings with and without Hamza that are difficult to perceive by native readers because the meaning is not affected (the omitting of dots on final Heh in the Kef, Heh, Yeh examples is also an instance of this type).

example 1 - AGRA: أجزاء vs. أجزاء , اجزا , اجزاء

A		G		R		A		?		total variants
أ	0623	ج	062c	ز	0632	ا	0627	ء	0621	
ا	0627					ا	0627			
2		1		1		1		2		4

example 2 - ZAYD: زايد vs. زائد

Koranic Arabic reflects a variety of Arabic that had lost glottal stop (Hamz) in non-initial position. The glottal stop survived only in initial position, written with Alef. In non-initial position traces of hamza are their replacement sounds W and Y - but following a consonant it had disappeared without a trace. Modern Arabic continues this pattern: no hamza is ever spoken in middle and final position.

For reasons that have not yet been clarified, hamza was introduced for Classical Arabic. Orthography had however already frozen in conventions that replace (W,Y) or omit hamza altogether. Hence the inline and superscript hamzas, on top of their replacements or between letters that cannot really handle a newcomer that was supposed to have disappeared forever.

Non-Arabic orthographies often borrowed words in the pre-classical or Koranic form. Persian has zaayid where Classical Arabic has zaa'id. In The Arabic world EVERYBODY says zaayid in casual speech. Even though hamza may words look different, its absence or presence goes unnoticed, because it's redundant. So there's a special case where clearly distinct unicodes generate clearly distinct shapes, and yet the reader has difficulty to perceive any functional difference. It's vaguely similar to non-functional differences between US and UK spellings of words like honourable neighbour and encyclopaedia.

5. Pattern Five

This is the *häägen dazš* pattern. It can be seen in Arabic messaging, using "cool" unicode letters for normal text. For certain audiences the diacritics are so absurd, that they do not interfere with legibility at all, while the underlying unicode points are totally different.

example 2: محمد vs. مُحَمَّد