Definition of terms

Definition of terms

- **Arabic:** By itself, this term can refer to the *script*, which is used by many languages including Arabic, or specifically the *language* as distinct from other languages that use the script. In this document where the distinction is important we'll try to qualify the term by saying "Arabic script" or "Arabic language".
- **shadda:** U+0651 ARABIC SHADDA, a combining mark used to represent gemination (elongation) of a consonant, drawn above the consonant it modifies.
- **vowel mark:** generic term to identify the combining marks that are most often used to represent vowels, such marks are drawn above or below the consonant that the vowel logically follows.
- **kasra:** U+0650 ARABIC KASRA, a particularly common vowel mark below. Unless otherwise indicated rules related to *kasra* also apply to U+064D ARABIC KASRATAN.
- **hamza:** U+0654 ARABIC HAMZA ABOVE or U+0655 ARABIC HAMZA BELOW and also used as part of existing compositions such as U+0623 ARABIC LETTER ALEF WITH HAMZA ABOVE.
- **maddah:** U+0653 ARABIC MADDAH ABOVE and also part of existing composition U+0622 ARABIC LETTER ALEF WITH MADDA ABOVE. Unless otherwise indicated, rules related to *maddah* also apply to U+06E4 ARABIC SMALL HIGH MADDA.
- **ligature:** the special form that specific sequences of letters can take that is different to how the letters might otherwise appear. The *lam-alef* ligature is a well known example. Note that within a font, ligature forms can be *implemented* as one glyph (often called a *ligature glyph*) or as multiple contextually-shaped glyphs that result in the desired display.
- **CGJ:** U+034F COMBINING GRAPHEME JOINER

Caveats

- We recognize that different calligraphic or language traditions may disagree about some things, e.g., Persian and Arabic rules may be different, or things that are semantically identical in Arabic may not be so in another language.
- We also recognize that a lot of the dictates regarding correctness are based on existing renderings of the Qur'an and thus may be more particular or stringent than required for practical orthographies.
- Finally it is clear that the Unicode-normalized ordering of diacritics, while helpful for its intended purpose (string equality testing), is not particularly helpful for rendering. Specifically, the Unicode normal order does not put marks in an order that is "logical" to the reader or is graphically appropriate. We might hope, for example, that marks that are rendered "closer" to the base would appear early in a normalized string, but they don't. Shadda and hamza are particularly problematic in this regard.

Unicode Normal Order

For reference, here are the Arabic combining marks shown in ascending order by their Unicode Combining Class values. If text is presented to the rendering engine in NFD, this is the ordering of the marks:

Character	Combining Class
064B ARABIC FATHATAN 08F0 ARABIC OPEN FATHATAN	27
064C ARABIC DAMMATAN 08F1 ARABIC OPEN DAMMATAN	28
064D ARABIC KASRATAN 08F2 ARABIC OPEN KASRATAN	29
0618 ARABIC SMALL FATHA 064E ARABIC FATHA	30
0619 ARABIC SMALL DAMMA 064F ARABIC DAMMA	31
061A ARABIC SMALL KASRA 0650 ARABIC KASRA	32
0651 ARABIC SHADDA	33
0652 ARABIC SUKUN	34
0670 ARABIC LETTER SUPERSCRIPT ALEF	35
0655 ARABIC HAMZA BELOW 0656 ARABIC SUBSCRIPT ALEF 065C ARABIC VOWEL SIGN DOT BELOW 065F ARABIC WAVY HAMZA BELOW 06E3 ARABIC SMALL LOW SEEN 06EA ARABIC EMPTY CENTRE LOW STOP 06ED ARABIC SMALL LOW MEEM 08CF ARABIC LARGE ROUND DOT BELOW 08D0 ARABIC SUKUN BELOW 08D1 ARABIC LARGE CIRCLE BELOW 08D2 ARABIC LARGE ROUND DOT INSIDE CIRCLE BELOW 08D3 ARABIC SMALL LOW WAW 08E3 ARABIC TURNED DAMMA BELOW 08E6 ARABIC CURLY KASRA 08E9 ARABIC CURLY KASRATAN 08ED ARABIC TONE ONE DOT BELOW 08EE ARABIC TONE TWO DOTS BELOW 08EF ARABIC TONE LOOP BELOW 08F6 ARABIC KASRA WITH DOT BELOW 08F9 ARABIC LEFT ARROWHEAD BELOW 08FA ARABIC RIGHT ARROWHEAD BELOW	220
0610 ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM 0611 ARABIC SIGN ALAYHE ASSALLAM 0612 ARABIC SIGN RAHMATULLAH ALAYHE 0613 ARABIC SIGN RADI ALLAHOU ANHU 0614 ARABIC SIGN TAKHALLUS 0615 ARABIC SMALL HIGH TAH 0616 ARABIC SMALL HIGH LIGATURE ALEF WITH LAM WITH YEH	230

Character	Combining Class
0617 ARABIC SMALL HIGH ZAIN	
0653 ARABIC MADDAH ABOVE	
0654 ARABIC HAMZA ABOVE	
0657 ARABIC INVERTED DAMMA	
0658 ARABIC MARK NOON GHUNNA	
0659 ARABIC ZWARAKAY	
065A ARABIC VOWEL SIGN SMALL V ABOVE	
065B ARABIC VOWEL SIGN INVERTED SMALL V ABOVE	
065D ARABIC REVERSED DAMMA	
065E ARABIC FATHA WITH TWO DOTS	
06D6 ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA	
06D7 ARABIC SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA	
06D8 ARABIC SMALL HIGH MEEM INITIAL FORM	
06D9 ARABIC SMALL HIGH LAM ALEF	
06DA ARABIC SMALL HIGH JEEM	
06DB ARABIC SMALL HIGH THREE DOTS	
06DC ARABIC SMALL HIGH SEEN	
06DF ARABIC SMALL HIGH ROUNDED ZERO	
06E0 ARABIC SMALL HIGH UPRIGHT RECTANGULAR ZERO	
06E1 ARABIC SMALL HIGH DOTLESS HEAD OF KHAH	
06E2 ARABIC SMALL HIGH MEEM ISOLATED FORM	
06E4 ARABIC SMALL HIGH MADDA	
06E7 ARABIC SMALL HIGH YEH	
06E8 ARABIC SMALL HIGH NOON	
06EB ARABIC EMPTY CENTRE HIGH STOP	
06EC ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE	
08CA ARABIC SMALL HIGH FARSI YEH	
08CB ARABIC SMALL HIGH YEH BARREE WITH TWO DOTS BELOW	
08CC ARABIC SMALL HIGH WORD SAH	
08CD ARABIC SMALL HIGH ZAH	
08CE ARABIC LARGE ROUND DOT ABOVE	
08D4 ARABIC SMALL HIGH WORD AR-RUB	
08D5 ARABIC SMALL HIGH SAD	
08D6 ARABIC SMALL HIGH AIN	
08D7 ARABIC SMALL HIGH QAF	
08D8 ARABIC SMALL HIGH NOON WITH KASRA	
08D9 ARABIC SMALL LOW NOON WITH KASRA	
08DA ARABIC SMALL HIGH WORD ATH-THALATHA	
08DB ARABIC SMALL HIGH WORD AS-SAJDA	
08DC ARABIC SMALL HIGH WORD AN-NISF	
08DD ARABIC SMALL HIGH WORD SAKTA	
08DE ARABIC SMALL HIGH WORD QIF	
08DF ARABIC SMALL HIGH WORD WAQFA	
08E0 ARABIC SMALL HIGH FOOTNOTE MARKER	
08E1 ARABIC SMALL HIGH SIGN SAFHA	
08E4 ARABIC CURLY FATHA	
08E5 ARABIC CURLY DAMMA	

Character	Combining Class
08E7 ARABIC CURLY FATHATAN	
08E8 ARABIC CURLY DAMMATAN	
08EA ARABIC TONE ONE DOT ABOVE	
08EB ARABIC TONE TWO DOTS ABOVE	
08EC ARABIC TONE LOOP ABOVE	
08F3 ARABIC SMALL HIGH WAW	
08F4 ARABIC FATHA WITH RING	
08F5 ARABIC FATHA WITH DOT ABOVE	
08F7 ARABIC LEFT ARROWHEAD ABOVE	
08F8 ARABIC RIGHT ARROWHEAD ABOVE	
08FB ARABIC DOUBLE RIGHT ARROWHEAD ABOVE	
08FC ARABIC DOUBLE RIGHT ARROWHEAD ABOVE WITH DOT	
08FD ARABIC RIGHT ARROWHEAD ABOVE WITH DOT	
08FE ARABIC DAMMA WITH DOT	
08FF ARABIC MARK SIDEWAYS NOON GHUNNA	

Red characters are part of the Modifier Combining Marks (MCM) as discussed in Unicode Technical Report #53. The blue character *shadda* is also discussed in UTR #53.

This guide is from the font-arab-tools project and is copyright © 2022 SIL International.