

Definition of terms

Definition of terms

- **Arabic:** since this whole document is about Arabic script, from here on we'll use the term "Arabic" to mean the *Arabic language*. If we need to identify the script, we'll say "ABS".
- **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 MADDAH ABOVE. Unless otherwise indicated, rules related to *maddah* also apply to U+06E4 ARABIC SMALL HIGH MADDAH.
- **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:

U+064B ARABIC FATHATAN U+08F0 ARABIC OPEN FATHATAN	27
U+064C ARABIC DAMMATAN U+08F1 ARABIC OPEN DAMMATAN	28
U+064D ARABIC KASRATAN U+08F2 ARABIC OPEN KASRATAN	29
U+0618 ARABIC SMALL FATHA U+064E ARABIC FATHA	30
U+0619 ARABIC SMALL DAMMA U+064F ARABIC DAMMA	31
U+061A ARABIC SMALL KASRA U+0650 ARABIC KASRA	32
U+0651 ARABIC SHADDA	33
U+0652 ARABIC SUKUN	34
U+0670 ARABIC LETTER SUPERScript ALEF	35
U+0655 ARABIC HAMZA BELOW U+0656 ARABIC SUBSCRIPT ALEF U+065C ARABIC VOWEL SIGN DOT BELOW U+065F ARABIC WAVY HAMZA BELOW U+06E3 ARABIC SMALL LOW SEEN U+06EA ARABIC EMPTY CENTRE LOW STOP U+06ED ARABIC SMALL LOW MEEM U+08CF ARABIC LARGE ROUND DOT BELOW U+08D0 ARABIC SUKUN BELOW U+08D1 ARABIC LARGE CIRCLE BELOW U+08D2 ARABIC LARGE ROUND DOT INSIDE CIRCLE BELOW U+08D3 ARABIC SMALL LOW WAW U+08E3 ARABIC TURNED DAMMA BELOW U+08E6 ARABIC CURLY KASRA U+08E9 ARABIC CURLY KASRATAN U+08ED ARABIC TONE ONE DOT BELOW U+08EE ARABIC TONE TWO DOTS BELOW U+08EF ARABIC TONE LOOP BELOW U+08F6 ARABIC KASRA WITH DOT BELOW U+08F9 ARABIC LEFT ARROWHEAD BELOW U+08FA ARABIC RIGHT ARROWHEAD BELOW	220

U+0610 ARABIC SIGN SALLALLAHOU ALAYHE WASSALLAM U+0611 ARABIC SIGN ALAYHE ASSALLAM U+0612 ARABIC SIGN RAHMATULLAH ALAYHE U+0613 ARABIC SIGN RADI ALLAHOU ANHU U+0614 ARABIC SIGN TAKHALLUS U+0615 ARABIC SMALL HIGH TAH U+0616 ARABIC SMALL HIGH LIGATURE ALEF WITH LAM WITH YEH U+0617 ARABIC SMALL HIGH ZAIN U+0653 ARABIC MADDAH ABOVE U+0654 ARABIC HAMZA ABOVE U+0657 ARABIC INVERTED DAMMA U+0658 ARABIC MARK NOON GHUNNA U+0659 ARABIC ZWARAKAY U+065A ARABIC VOWEL SIGN SMALL V ABOVE U+065B ARABIC VOWEL SIGN INVERTED SMALL V ABOVE U+065D ARABIC REVERSED DAMMA U+065E ARABIC FATHA WITH TWO DOTS U+06D6 ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA U+06D7 ARABIC SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA U+06D8 ARABIC SMALL HIGH MEEM INITIAL FORM U+06D9 ARABIC SMALL HIGH LAM ALEF U+06DA ARABIC SMALL HIGH JEEM U+06DB ARABIC SMALL HIGH THREE DOTS U+06DC ARABIC SMALL HIGH SEEN U+06DF ARABIC SMALL HIGH ROUNDED ZERO U+06E0 ARABIC SMALL HIGH UPRIGHT RECTANGULAR ZERO U+06E1 ARABIC SMALL HIGH DOTLESS HEAD OF KHAH U+06E2 ARABIC SMALL HIGH MEEM ISOLATED FORM U+06E4 ARABIC SMALL HIGH MADDA U+06E7 ARABIC SMALL HIGH YEH U+06E8 ARABIC SMALL HIGH NOON U+06EB ARABIC EMPTY CENTRE HIGH STOP U+06EC ARABIC ROUNDED HIGH	230
--	-----

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.