

Arabic / العربية al-‘Arabīyah

romanization: UN 1972 + KNAB modifications

Arabic is written from right to left. The Arabic script usually omits vowel points and diacritical marks from writing which makes it difficult to obtain uniform results in the romanization of Arabic. It is essential to identify correctly the words which appear in any particular name and to know the standard Arabic-script spelling including proper pointing. One must also take into account dialectal and idiosyncratic deviations.

In the romanization system below column 1 denotes an independent consonant character, column 2 the initial, column 3 the medial and column 4 the final form of a character. Column 5 gives the romanization equivalent.

I. Consonant characters

1	2	3	4	5	11	ر	ر	r	22	ق	ق	ق	ق	q
1	ء			ء ^{-A}	12	ز	ز	z	23	ك	ك	ك	ك	k
2	ا		ا	ا ^{-B}	13	س	س	s	24	ل	ل	ل	ل	l
3	ب	ب	ب	b	14	ش	ش	sh	25	م	م	م	م	m
4	ت	ت	ت	t ^C	15	ص	ص	ṣ	26	ن	ن	ن	ن	n
5	ث	ث	ث	th	16	ض	ض	ḍ	27	ه	ه	ه	ه	h ^C
6	ج	ج	ج	j	17	ط	ط	ṭ	28	و			و	w
7	ح	ح	ح	ḥ	18	ظ	ظ	ẓ	29	ي	ي	ي	ي	y
8	خ	خ	خ	kh	19	ع	ع	‘						
9	د		د	d	20	غ	غ	gh						
10	ذ		ذ	dh	21	ف	ف	f						

^A Not romanized word-initially.

^B Not romanized, but see romanizations accompanying *alif* (ا) in Table II.

^C In certain endings, an original *tā’* (ت) is written ة, i.e., like *hā’* (ه) with two dots, and is known as *tā’ marbūṭah*. It is romanized *h*, except in the construct form of feminine nouns, where it is romanized *t*, instead, e.g. همزة *hamzah* but همزة القتع *hamzat al-qat’*.

II. Vowels, diphthongs and diacritical marks (ب stands for any consonant character)

1	ب	a	4	ب	i	7	بَا	ā
2	بَو	aw	5	بُ	u	8	آ	ā
3	بَي	ay	6	بْ ^(A)		9	بِي	ī

10	بُو	ū	13	بِ	i ⁿ
11	بَى	á	14	بُ	u ⁿ
12	بَّ	a ⁿ	15	بْ	(B)

^A Marks absence of the vowel.

^B Marks doubling of the consonant.

III. Numbers

.	١	٢	٣	٤	٥	٦	٧	٨	٩
0	1	2	3	4	5	6	7	8	9

Numbers are written from left to right: ١٩٩٨ = 1998.

IV. Additional characters in some Arabic countries

(used in names and elsewhere for some missing sounds)

1	ب	f (character variant in Maghrib countries)
2	ق, ق	q (character variant in Maghrib countries)
3	ف	v (in Iraq, etc.)
4	ڤ	v (in Tunisia and elsewhere in Maghrib)
5	ق	g (in Tunisia)
6	ڇ	ch [tʃ] (in Maghrib countries)

Notes

1. *Hamzah* (ء) is always written on a word-initial *alif* (ا), except the article *al* and the form with *maddah* (آ). The bearer of *hamzah* is determined as follows:
 - a. word-initially always on *alif*, e.g. أَخَذَ *akhadha*;
 - b. word-medially:
 - i. vowelless *hamzah* – by the previous vowel, e.g. بئر *bi'r*;
 - ii. vowelless *hamzah* after a vowelless consonant – by the following vowel, e.g. تَسْأَلُ *tas'alu* (exception: if preceded by y, then always on y, e.g. هَيْئَةً *hay'ah*);
 - iii. if the *hamzah* is preceded and followed by a vowel, then the bearer in different combinations is:
 1. a'a – *alif*, e.g. سَأَلَ *sa'ala*,
 2. V'u, u'V (V = any vowel) – w, e.g. سُؤَالَ *su'āl*,
 3. V'i, i'V – y, e.g. رَأَيْسَ *ra'īs*,
 4. u'i, i'u – y, e.g. سُؤِيلًا *su'ila*,
 5. ā'a, ā'ā, ū'a – without a bearer, e.g. بِنَاءَاتٍ *binā'āt*;
 - c. word-finally:
 - i. after a short vowel – by that vowel, e.g. قَرَأَ *qara'a*, قُرِيَ *quri'a*;
 - ii. after a long vowel or a vowelless consonant – without a bearer, e.g. أَصْدِقَاءُ *aṣḍiqā'u*;

- d. long *ā* following the *hamzah* is written with a *maddah* (except after *ā*), e.g. *آب āb*.
2. The first definite article *al* starting a name is written with a capital letter, all other articles within a name are written with a small letter. A hyphen is used between the article and the name, e.g. الوادي *Al-Wādī*.
 3. If the article *al* precedes a word starting with „sun letters“ (*d, dh, ḍ, l, n, r, s, sh, ṣ, t, th, ṭ, z, ḏ*), the *l* of the article is assimilated to the following sound, e.g. الشارقة *Ash-Shāriqah*, not *Al-Shāriqah*.
 4. The preferred romanization of adjectival endings *ي* is *īy* (not *iyy*).
 5. *Hamzat al waṣl* (أ) is romanized ' ; it is used only in the romanization of Classical Arabic (cf. the name of the character in Classical Arabic *hamzatu 'l waṣli*).

KNAB modifications to ensure reversibility (not shown in public version)

1. A middle dot (·) is used to distinguish ambiguous combinations, e.g. *ته = t·h, كه = k·h*, etc.
2. Word-final *-ah* and *-at* correspond to Arabic ة; if *ه* or *ت* occur word-finally, then these are romanized *-a·h* and *-a·t*. If ة is used after a vowel other than *a*, also a middle dot is used: *قناة qanā·t*.
3. In order to distinguish some other irregularities, occurring mainly in non-Arabic languages, the following modifications are used:
 - a. a final form of a character occurring in the middle of a word is indicated by " , e.g. *بھاولپور Bhāval"pur*;
 - b. " word-finally indicates that in the original script the words are written into one word, e.g. *براکي برک Barākī" Barak* (if the final form of a character is missing, then a middle dot is used, e.g. *پلخمری Pul-i· Khumrī*);
 - c. _ indicates that a name romanized as one word, is written separately in the original script, e.g. *جلال آباد Jalāl_ābād*;
 - d. in languages where *ی* is usually dotless but in some cases dotted characters (ي) are used, the latter are denoted by *ÿ* or *ï*.

Pronunciation

<i>ā</i>	[a:]	<i>ī</i>	[i:]	<i>ṭ</i>	[t ^s] ^A
<i>á</i>	[a:]	<i>j</i>	[dʒ]~[ʒ]	<i>ū</i>	[u:]
<i>dh</i>	[ð]	<i>kh</i>	[x]~[χ]	<i>y</i>	[j]
<i>ḍ</i>	[d ^s] ^A	<i>q</i>	[q]	<i>z</i>	[ð ^s]~[z ^s] ^A
<i>gh</i>	[ɣ]~[ʁ]	<i>sh</i>	[ʃ]	'	[ʔ]
<i>h</i>	[h], [-] ^B	<i>ṣ</i>	[s ^s] ^A	‘	[ʕ]
<i>ḥ</i>	[ħ]	<i>th</i>	[θ]		

^A So-called emphatic sounds that change the colour of neighbouring sounds.

^B Word-finally mostly not pronounced.