ROMANIZATION OF ARABIC

BGN/PCGN 1956 System
(Revised Presentation 2019)

This System was adopted by the BGN in 1946 and by the PCGN in 1956 and is applied by BGN and PCGN in the systematic romanization of Arabic geographical names in Bahrain, Egypt, Iraq, Jordan, Kuwait, Libya, Oman, Qatar, Saudi Arabia, Syria, the United Arab Emirates, Yemen, the West Bank and Gaza Strip[1].

Uniform results in the romanization of Arabic are difficult to obtain, since vowel points and diacritical marks are generally omitted from both handwriting and printed script. It follows that for correct identification of the words which appear in any particular name, knowledge of its standard Arabic-script spelling including proper pointing, and recognition of dialectal and idiosyncratic deviations are essential.

In order to bring about uniformity in the Roman-script spelling of geographical names in Arabic-language areas, the system is based insofar as possible on fully pointed Modern Standard Arabic (MSA). In the interest of clarity, vowel pointing to indicate short vowels has been applied to the examples given below, and examples of the, more usual, unpointed script have also been provided; it should also be noted that the dots which occur on some characters of the Arabic script are not vowels but rather are an integral part of the base consonant.

Arabic script is written from right to left, and does not make a distinction between upper and lower case.

Table 1 below shows the written consonant characters used in standard Arabic script; Table 2 shows vowel characters and other diacritical marks, some of which are usually not written; and Table 3 shows the Perso-Arabic script characters which, although not part of the standard Arabic alphabet, are sometimes used to represent sounds not present in Standard Arabic. These tables should be used in conjunction with the Notes and Special Rules which follow. These explain elements of Arabic grammar and orthography which affect romanization and should be treated as an integral part of the Romanization System for Arabic.

---

[1] It is not used to represent names in Algeria, Chad, Comoros, Djibouti, Lebanon, Mauritania, Morocco, Sudan or Tunisia, where the spellings used on official Roman-script sources are used.

Table 1: Standard Arabic Consonant Characters

	Script				Unicode value[2] (Independent)	Romanization	Roman Unicode value (lower case)	Example		
	Final	Medial	Initial	Independent				Pointed Script	Unpointed Script	Roman Script
1.		ء			0621	*not romanized in word-initial position* [Note 2]	-	أَبُو ظَبْي	أبو ظبي	Abū Ẓaby
						' in all other positions [Note 2]	2019	بِئْر زَيْت	بئر زيت	Bi'r Zayt
2.	ـا		ا		0627	*See Notes 3 & 10*	-	أمِّ العَمَد	أم العمد	Umm al 'Amad
3.	ب	ـبـ	بـ	ب	0628	b	0062	البَحرَين	البحرين	Al Baḥrayn
4.	ت	ـتـ	تـ	ت	062A	t	0074	الكُوت	الكوت	Al Kūt
5.	ث	ـثـ	ثـ	ث	062B	th	0073+0304	الثُّلَيْثُوَات	الثليثوات	Ath Thulaythuwāt
6.	ج	ـجـ	جـ	ج	062C	j	006A	الجَزِيرَة	الجزيرة	Al Jazīrah
7.	ح	ـحـ	حـ	ح	062D	ḥ	1E29	المَحْمُودِيَّة	المحمودية	Al Maḥmūdīyah
8.	خ	ـخـ	خـ	خ	062E	kh	006B+0068	خَيْبَر	خيبر	Khaybar
9.	ـد	ـد	د	د	062F	d	0064	دَمَنْهُور	دمنهور	Damanhūr
10.	ـذ	ـذ	ذ	ذ	0630	dh	007A+0304	ذَهَب	ذهب	Dhahab
11.	ـر	ـر	ر	ر	0631	r	0072	الرَّوْضَة	الروضة	Ar Rawḍah
12.	ـز	ـز	ز	ز	0632	z	007A	زُوَارَة	زوارة	Zuwārah
13.	س	ـسـ	سـ	س	0633	s	0073	السُّلَيْمَانِيَّة	السليمانية	As Sulaymānīyah

---

[2] https://unicode.org/charts/PDF/U0600.pdf

	Script				Unicode value² (Independent)	Romanization	Roman Unicode value (lower case)	Example		
	Final	Medial	Initial	Independent				Pointed Script	Unpointed Script	Roman Script
14.	ش	ﺸ	ﺷ	ش	0634	sh	0073+0068	الشَّام	الشام	Ash Shām
15.	ص	ﺼ	ﺻ	ص	0635	ş	015F	قَيْصُومَة	قيصومة	Qayşūmah
16.	ض	ﻀ	ﺿ	ض	0636	ḑ	1E11	ضَوْر	ضور	Ḑawr
17.	ط	ﻂ	ﻃ	ط	0637	ţ	0163	القُنَيْطِرَة	القنيطرة	Al Qunayţirah
18.	ظ	ﻆ	ﻇ	ظ	0638	ẓ	007A+0327	ظُفَار	ظفار	Ẓufār
19.	ع	ﻌ	ﻋ	ع	0639	'	2018	أَبُو عَرِيش	أبو عريش	Abū 'Arīsh
20.	غ	ﻐ	ﻏ	غ	063A	gh	0067+0068	بَغْدَاد	بغداد	Baghdād
21.	ف	ﻔ	ﻓ	ف	0641	f	0066	الفُرَات	الفرات	Al Furāt
22.	ق	ﻘ	ﻗ	ق	0642	q	0071	قَطَر	قطر	Qaţar
23.	ك	ﻜ	ﻛ	ک	06A9	k	006B	الكُوَيْت	الكويت	Al Kuwayt
24.	ل	ﻠ	ﻟ	ل	0644	l *Note 10*	006C	حَلَب	حلب	Ḩalab
25.	م	ﻤ	ﻣ	م	0645	m	006D	مَكَّة	مكة	Makkah
26.	ن	ﻨ	ﻧ	ن	0646	n	006E	نَخْل	نخل	Nakhl
27.	ه	ﻬ ,ﻪ	ﻫ	ه	0647	h	0068	جَبَل هَارُون	جبل هارون	Jabal Hārūn
28.	و	و	و	و	0648	w	0077	وَادِي غَضَا	وادي غضا	Wādī Ghaḑā
29.	ي	ﻴ	ﻳ	ي	064A	y	0079	اليَمَن	اليمن	Al Yaman

	Script				Unicode value² (Independent)	Romanization	Roman Unicode value (lower case)	Example		
	Final	Medial	Initial	Independent				Pointed Script	Unpointed Script	Roman Script
30.	ة	-	-	ة	FE93	ah *or* at [Note 4]	0061+0068, 0061+0074	القَاهِرَة	القاهرة	Al Qāhirah
								المَدِينَة المُنَوَّرَة	المدينة المنورة	Al Madīnah al Munawwarah
								مُحَافَظَة دِمَشْق	محافظة دمشق	Muḥāfaẓat Dimashq

Table 2: Vowel Characters and Diacritical Marks[3]

	Script	Unicode value	Romanization	Roman Unicode value (lower case)	Example		
					Pointed Script	Unpointed Script	Roman Script
1.	◌َ	064E	a	0061	البَصْرَة	البصرة	Al Baṣrah
2.	◌ِ	0650	i[4]	0069	الرِّيَاض	الرياض	Ar Riyāḍ
3.	◌ُ	064F	u	0075	القُدْس	القدس	Al Quds
4.	اَ	0627	ā *Notes 3 & 10*	0101	بَاب المَنْدَب	باب المندب	Bāb al Mandab
5.	◌ِ ي	0649	ī	012B	المَدِينَة	المدينة	Al Madīnah
6.	◌ُو	0648	ū	016B	صُور	صور	Ṣūr
7.	ىٰ , ◌َى	0649, 0649+0670	á *Note5*	00E1	مَرْسَىٰ مَطْرُوح	مرسى مطروح	Marsá Maṭrūḥ
8.	◌ْ	0652	*not romanized*	-	*Indicates absence of short vowel*		
9.	◌َيْ	064A	ay	0061+0079, 0061+012B	صَيْدَا	صيدا	Ṣaydā
10.	◌َوْ	0648	aw	006F+0077	الدَّوْحَة	الدوحة	Ad Dawḥah
11.	◌ً	064B	aⁿ		*See note 6*		
12.	◌ٍ	064D	iⁿ		*See note 6*		
13.	◌ٌ	064C	uⁿ		*See note 6*		
14.	◌ّ	0651	*doubling of consonant* *Note7*	-	مُحَمَّد	محمد	Muḥammad

---

[3] Numbers 1, 2, 3, 8, 11, 12, 13 and 14 are not usually written.

[4] This vowel mark should be romanized as 'i' when it occurs in the following positions:

   a. Below the consonant character

   b. Below the *shaddah* (see row 14 and note 7), which itself occurs above a consonant character

   c. Below a *hamzah* occurring above an *alif*

	Script	Unicode value	Romanization	Roman Unicode value (lower case)	Example		
					Pointed Script	Unpointed Script	Roman Script
15.	او	0627 + 0648	Ū or Aw *in word initial position*[Note11]	016A or 0041+0077	أُوزُونْلَار	اوزونلار	Ūzūnlār
					أَوْسَط	اوسط	Awsaţ
			āw *in word medial or final position* [Note11]	0101+0077	سَنَاو	سناو	Sanāw
16.	اي	0627+064A	Ī *in word initial position* [Note11]	012B	إِيرَان	ايران	Īrān
			āy *in word medial or final position*	0101+0079	تَلّ السَّرَاي	تل السراي	Tall as Sarāy
17.	ٱ	0671	ʼ *Note 8*	2019	*See note 8*		
18.	آ	0622	Ā *in word initial position* [Notes 3 & 9]	0101	آلْبُو مُعَيْط	البو معيط	Ālbū Muʻayţ
			ʼā *in word medial position* [Notes 3 & 9]	2019+0101	قُرآن	قرآن	Qurʼān

Table 3: Modified/Non-Standard Arabic Script Characters

	Script				Unicode value (Independent)	Romanization	Roman Unicode value (lower case)	Example		
	Final	Medial	Initial	Independent				Pointed Script	Unpointed Script	Roman Script
1.	پ	ـپـ	پـ	پ	067E	p	0070	سَلْمان پَاك	سلمان پاك	Salmān Pāk
2.	ـچ	ـچـ	چـ	چ	0686	ch	0063+0068	تَلّ كوچِك الصَّغِير	تل كوچك الصغير	Tall Kūchik aş Şaghīr
3.	ـڤ	ـڤـ	ڤـ	ڤ	06A4	v	0076	مَزَّة ڤِيلَّات غَرْبِيَّة	مزة ڤيلات غربية	Mazzah Vīllāt Gharbīyah
4.	ـڨ	ـڨـ	ڨـ	ڨ [5]	06A8	g	0067	ڤَفْصَة	ڤفصة	Gafşah (Gafsa)
5.	ـگ	ـگـ	گـ	گ [6]	06AF	g	0067	تَلّ گَمْر	تل گَمر	Tall Gamr
6.	ـڭ	ـڭـ	ڭـ	ڭ [7]	06B4	g	0067	زَاڭُورَة	زاڭورة	Zāgūrah (Zagora)

[5] Used in Tunisian Arabic Script.
[6] Used principally in Iraq, but also sometimes used in other Arabic speaking countries to represent the 'g' sound.
[7] Used in Moroccan Arabic Script.

Numerals									
٠	١	٢	٣	٤	٥	٦	٧	٨	٩
0	1	2	3	4	5	6	7	8	9

Although Perso-Arabic script is written from right to left, numerical expressions, e.g. ١٩٦٨ → 1968, are written from left to right.

Notes

1. The symbol ○ is used in this system to symbolise any Arabic consonant character. It is not itself an Arabic letter.

2. *Hamzah* (ء) is written in Arabic in association with most instances of initial *alif*, except those which belong to the definite article *al* or which bear a *maddah* (see note 9). *Hamzah* is written above the *alif* (أ) if the accompanying short vowel is a *fatḥah* or *ḍammah* and usually below the *alif* (إ) if the accompanying short vowel is a *kasrah*. When the purpose is to indicate the presence of a glottal stop, *hamzah* is written over medial *alif* (أ), *wāw* (ؤ) and *yā'*, typically without dots (ئ); or following final *alif* (اء), these characters serving only to "bear" the hamzah. *Hamzah* following *kasrah* (ِ) is written (ئ); the *yā'* is usually in the initial or medial form and the dots are omitted e.g. *bi'r* (بِئر). *Hamzah* following *ḍammah* (ُ) is written (ؤ). *Hamzah* following a long vowel is written without a bearer and is positioned on the line of print like a regular character, e.g. صنعاء *Şan'ā'*. The romanization of *hamzah* (' – Unicode encoding 2019) should always be carefully distinguished from that of *'ayn* (' - Unicode encoding 2018).

3. *Alif* (ا) occurs with the following uses:
    a. Initially, it indicates that the word begins with a vowel or diphthong; the *alif* itself is not romanized, but rather "carries" the short vowel, which is romanized; e.g., أبو ظبي → *Abū Ẓaby*.
    b. With *maddah* (آ – row 18 in the vowel table), it is represented *ā*; e.g., آلبو مُعَيط → *Ālbū Mu'ayṯ*. See also note 9.
    c. Medially and finally it is represented *ā*; e.g., باب → *Bāb*, صيدا → *Şaydā*.
    d. Medially and finally, *alif* may serve as the bearer of *hamzah*, e.g. رأس → *ra's.* See also note 2.

8

4. The *tā' marbūţah* character (ة), which looks like *hā'* with two dots above and occurs only at the end of words, is romanized **h**, except in an *iḍāfah*[8] noun phrase construction, where it is romanized **t**, in accordance with pronunciation. e.g. *Muḥāfaẓah* (as an isolated word) but *Muḥāfaẓat Baghdād*. In exceptional cases, when it is necessary to distinguish it from the *tā' marbūţah*, the ending *fatḥah + hā'* (ه ◌) may be romanized *a·h* when the character *hā'* (ه) is pronounced as such. Example: *Muntaza·h*. (See also special rule 12). The *tā marbūţah* is always preceded by the short vowel *fatḥah* (◌) and is therefore romanized as **ah** or **at**, except when it is preceded by *alif* when it is romanized **āh** (not āah), e.g. *Ḥamāh* (حماة), and as **āt** within an *iḍāfah* construction.

5. The character *yā'* (in final form but without dots) preceded by the vowel point *fatḥah* is known as *alif maqşūrah*. This character may also be pointed ىٰ and should be romanized *á*. See character 7 in the vowel table.

6. The classical Arabic grammatical endings written with the nunation symbols (*tanwīn*) may be romanized, when necessary, by aⁿ, iⁿ, uⁿ. In modern spoken Arabic, these endings have become silent and should not be romanized: e.g. classical *alifuⁿ*; modern *alif*.

7. Doubled consonant sounds are represented in Arabic script by placing a *shaddah* (◌ّ) over a consonant character, although like the short vowels the *shaddah* may not always be written. In romanization the letter should be doubled, e.g. *Quwwah*, *'Abbās*. However, the combination of the consonant character *yā'* with a *shaddah* preceded by a *kasrah* (يِّ) at the end of a word is romanized **ī**, e.g. *Gharbī*; a word ending *kasrah + yā'* with a *shaddah + tā' marbūţah* is romanized **īyah** (rather than *iyyah)*, e.g. السُّلَيمانِيَّة is romanized *As Sulaymānīyah* and not *As Sulaymāniyyah*; and when the *kasrah* + yā' + shaddah combination is followed by the sound masculine plural ending (يِّين◌ or يُّون◌) it should be romanized as –īyīn/īyūn, e.g. العباسيين ساحة should be romanized as *Sāḥat al 'Abbāsīyīn*.

8. *Hamzat al waşl* (أ), which is utilized only in the pointing of classical Arabic, is romanized ' as illustrated in the classical form of its name *hamzatu'l waşli*.

9. Since *maddah* (آ), which is placed over *alif* (ا), often occurs in word-initial position, no confusion results from the use of *ā* for *alif maddah* (آ) as well as for *fatḥah* followed by *alif* (اَ).

10. The ligature ﻻ represents *lām-alif*, and should be romanized *lā*.

---

[8] The *iḍāfah* is a grammatical construction in Arabic consisting of two or more nouns, which indicates a relationship of possession between the words. For example, *Muḥāfaẓat Baghdād*, which could be translated in English as the Governorate **of** Baghdad.

11. In word initial position the combination *Alif +Wāw* (او) is sometimes used to render an initial long vowel sound in words of non-Arabic origin. Where this is clearly the case it should be romanized Ū. In words of Arabic or uncertain origin it should be romanized Aw. In word-medial or word-final position it should always be romanized āw. Similarly, the combination *Alif +Yā'* (اي) is romanized Ī to render an initial long vowel sound but as āy in word-medial or word-final position.

Special Rules

1. The Arabic definite article *al* (ال) should be treated as follows:
   a. Initial definite articles should be capitalized and hyphens should not be used to connect parts of names, e.g. *Ash Shāriqah*. When appearing medially in a name the initial 'a' should be lower case, e.g. *Tall al Laḥm*.
   b. When the definite article precedes a word beginning with one of the "sun letters" t, th, d, dh, r, z, s, sh, ṣ, ḍ, ṭ, ẓ, l, or n – the l is assimilated in pronunciation and romanization, thus yielding, for example, the romanization *Ar Riyāḍ*, rather than *Al Riyāḍ* for الرّياض.
   c. If sources contradict over the inclusion or non-inclusion of the definite article in a name, preference should be given to the form with the article.

2. Conjunctions and prepositions should be romanized according to their written form in Arabic script and should be lower case. In cases where the conjunction or preposition ends in a long or short vowel any assimilated pronunciation should not be shown in the romanized form. e.g. *Khabb **wa** ash Sha'f* (خب والشعف).

   There are two exceptions to this rule:
   a. In the case of the preposition *li* (ل), where the *alif* of the definite article is assimilated in the written form as well as pronunciation, the written form should be shown in romanization as follows: *Miṣr **liṭ** Ṭayarān* (مصر للطيران); *Ash Sharikah al 'Āmmah **lil** Maghāzil* (الشركة العامة للمغازل).
   b. In the case of the preposition *bi* (ب), the *alif* of the definite article is assimilated in pronunciation and, although the *alif* remains in the written form the short vowel it carries changes from 'a' to 'i'. For example: *Al Qaryah **bid** Duwayr* (القرية بالدوير) but ***Ad** Duwayr* (الدوير); and *Al Ḥarajah bil Qur'ān* (الحرجة بالقرآن) but ***Al** Qur'ān* (القرآن).

3. The Arabic word for God (الله) should be written *Allāh*. The *alif khanjarīyah* (*dagger alif*) (ٰ) above the second ل (*lām*) in the word الله, like the short vowels, is not usually written but should be romanized ā, like a full-size *alif*. This diacritical mark appears in a few other Arabic words, for instance on the *alif maqṣūrah* as described in note 5.

4. Names which consist of noun phrases[9] should be written as separate words. The definite article within such names should be romanized *al*, not *ul*, e.g., *'Abd Allāh*, *'Abd ar Raḥmān*, *Dhū al Faqār*, and as noted in special rule 1, the medial *al* should be lower case .

5. The Arabic word بن should be romanized *Bin* rather than *Ibn* whenever written without *alif*, that is between two proper nouns, e.g., *'Umar Bin al Khaṭṭāb*. Where it appears with *alif* (ابن), it should be romanized *Ibn*.

6. The Turkish word *Paşa* should be romanized from Arabic script as *Bāshā*. The Turkish word *Bey* should be romanized as *Bey* in Egyptian names, no matter how it is written in Arabic-language sources, but in other Arabic areas it should be romanized as *Bak* where written بك and as *Bayk* when written بيك.

7. The modern colloquial word *Sīdī* (سيدي) should be give precedence over the classical form *Sayyidī*. This does not preclude the spelling *Sayyidī* if the latter is indicated by the Arabic script or other evidence – for instance, if the *yā'* is written with a *shaddah* (ّ).

8. When colloquial forms of words are found as part of place names on authoritative sources, these spellings should be retained and reproduced in Romanization. Examples of this include:
   a. The colloquial word *Bū* should not be changed to the standard form *Abū*.
   b. The colloquial word for water, written مية on Arabic maps, should be romanized *Mayyat*.
   c. Place names where the initial alif of the Arabic definite article (Al) is omitted, e.g. *Lūsayl* (a truncation of the spelling *Al Wusayl*, which might be expected in MSA).

9. Place names of Aramaic origin in Syria often contain initial consonant clusters consisting of *b* plus another consonant such as *l* or *h*. In romanization, the clusters *bl*, *bh*, etc., should be so represented.

10. In names containing the Arabic word for back, ridge, or hill, appearing as either ظهر (*Ẓahr*) or ضهر (*Ḍahr*) in Arabic sources, the word should be romanized to reflect the particular Arabic spelling shown. Where sources differ, preference should be given to the form found on the most authoritative source.

11. In formal Arabic, the spelling of some words ending in a long vowel character may change according to that word's grammatical function in a sentence. For example, the personal name *Abū Bakr* (ابو بكر) would become *Abī Bakr* (ابي بكر) when preceded by a generic in

---

[9] See also note 4

an *iḍāfah* construction[10] e.g. *Shāriʿ Abī Bakr* (شارع ابي بكر – Abu Bakr Street). The spelling of such words as found on the most authoritative source should be used in the romanized form of the name. Other common words affected by this rule are *Banū/Banī* (sons of…) and *Dhū/Dhī* (owner of…).

Examples of names in this category include *Jabal Abā aṣ Şabbān* (جبل ابا الصبان) and *Muḥāfaẓat Dhī Qār* (محافظة ذي قار).

12. Occasionally the character sequences كﻫ, دْه, سْه, and ثْه occur. They may be romanized *k·h, d·h, s·h,* and *t·h* in order to differentiate these romanizations from the digraphs *kh, dh, sh,* and *th*, which are used to represent the characters خ, ذ, ش, and ث respectively. See also note 4.

---

[10] See footnote 7