

BGN/PCGN NATIONAL ROMANIZATION SYSTEM FOR AFGHANISTAN

BGN/PCGN 2007 System

This romanization system agreed by BGN and PCGN in November 2007, accommodates the linguistic complexity of Afghanistan as manifest in its geographical names. The following tabulation shows the original Perso-Arabic script with accompanying Unicode value (columns 1a and b), the Yaqubi romanization¹ (column 2), the BGN/PCGN romanization with accompanying Unicode value (columns 3a and b), an English phonetic example (column 4), and an example toponym (columns 5a and b).

The tables function as both a romanization system for Afghanistan (i.e. with access to the original script, these tables can be applied to get a standardized Roman result - moving from columns 1 to 3) and as a means of converting the available Yaqubi Roman-script spellings, as appear on the Fairchild Aerial Surveys map series, to standard BGN/PCGN spellings (moving from columns 2 to 3).

¹ The “Yaqubi” romanization system was developed in 1959 by Muzaffarud Din Yaqubi. It is a native official system designed to reflect Afghan names, both Dari and Pashto, and both pronunciation and genuine linguistic truth.

Consonant Characters

	Script	Unicode value	Yaghoubi	BGN/PCGN	BGN/PCGN Unicode value	English Phonetic	Example	
1	ب	0628	b	b	0062	\b\ as in <i>bob</i>	بَعْلَان	Baghlān
2	پ	067E	p	p	0070	\p\ as in <i>pep</i>	پُوٹَکَی	Pōṭakay
3	ت	062A	t	t	0074	\t\ as in <i>tall</i>	شیرین تَغَاب	Shīrīn Tagāb
4	ٹ	067C	t	ṭ	1E6F	retroflexive \t\	کُوٹ	Kōṭ
5	س	062B	s	s̄	0073+0304	\s\ as in <i>sand</i>	سَابِر	Sābir
6	ج	062C	j	j	006A	\j\ as in <i>jazz</i>	جَلَالِآبَاد	Jalālābād
7	چ	0686	č	ch	0063/0068	\ch\ as in <i>chat</i>	چَارِکَار	Chārīkār
8	ح	062D	ḥ	ḥ	1E29	\h\ as in <i>hot</i>	حَضْرَتِ سُلْطَان	Ḥazrat-e Sulṭān
9	خ	062E	kh	kh	006B/0068	\ch\ as in German <i>ach</i>	خُوَسْت	Khōst
10	ج ²	0681	j	dz	0064+007A	\dz\ as in <i>adze</i>	حَدْرَانِ	Dzadrāṇ
11	س	0685	c	ts	0074/0073	ts\ as in <i>gnats</i>	سُوکَی	Tsowkêy
12	د	062F	d	d	0064	\d\ as in <i>dog</i>	سَپِن بُولْدَک	Spīn Bōldak

² A variant form ج² is seen infrequently; it does not have a single Unicode encoding.

	Script	Unicode value	Yaghoubi	BGN/PCGN	BGN/PCGN Unicode value	English Phonetic	Example
13	پ	0689	d	d	1E0F	retroflexive \d\	دَند و پَتَان Dand wa Patān
14	ذ	0630	z̄	z̄	007A+0304	\z\ as in <i>zero</i>	گُذَرگَاهُ نُور Gužargāh-e Nūr
15	ر	0631	r̄	r	0072	\r\ as in <i>run</i>	کَندهَار Kandahār
16	ړ	0693	r	r̄	1E5F	retroflexive \r\	اَندرِ Andar̄
17	ز	0632	z	z	007A	\z\ as in <i>zero</i>	کُنْدُز Kunduz
18	ژ	0698	ž	zh	007A+0068	\zh\ as in <i>azure</i>	میرِ اسَلَم ژَرَنده Mīr Aslam Zhrandah
19	ږ	0696	ž̄	zh̄	007A+0332 0068+0332	retroflexive \zh\	ږیره Zhīrah
20	س	0633	s	s	0073	\s\ as in <i>sand</i>	سَمَنگان Samangān
21	ش	0634	š	sh	0073+0068	\sh\ as in <i>ship</i>	مَزَارِ شَرِيف Mazār-e Sharīf
22	ښ	069A	š̄	sh̄	0073+0332 0068+0332	retroflexive \sh\	کَبِهته کَلَا Kshêtah Kalā
23	ص	0635	s̄	ş	015F	\s\ as in <i>sand</i>	قَيْصَار Qayşār

	Script	Unicode value	Yaghoubi	BGN/PCGN	BGN/PCGN Unicode value	English Phonetic	Example	
24	ض	0636	dz	z̤	1E95	\z\ as in <i>zero</i>	فَيْضْ أَبَاد	Faīzābād
25	ط	0637	t̤	ṭ	0163	\t\ as in <i>tall</i>	حَضْرَتِ سُلْطَان	Ḥazrat-e Sultān
26	ظ	0638	z̤	z̤	007A+0327	\z\ as in <i>zero</i>	ظَاهِرْ كَلَا	Ẓāhir Kalā
27	ع	0639	'	'	2018	no sound	پُلْ عَلم	Pul-e 'Alam
28	غ	063A	gh	gh	0067+0068	no English equivalent / like \r\ in French <i>mer</i>	غَزْنِي	Ghaznī
29	ف	0641	f	f	0066	\f\ as in <i>fat</i>	مَزَارِ شَرِيف	Mazār-e Sharīf
30	ق	0642	q	q	0071	no English equivalent / closest sound is guttural \k\	قَيْصَار	Qayṣār
31	ك	06A9	k	k	006B	\k\ as in <i>kitten</i>	كَنْدَهَار	Kandahār
32	گ	06AF	g	g	0067	\g\ as in <i>gain</i>	گَرْدِيز	Gardēz
33	ل	0644	l	l	006C	\l\ as in <i>land</i>	كَابُل	Kābul
34	م	0645	m	m	006D	\m\ as in <i>mint</i>	مَيْمَنَه	Maīmanah
35	ن	0646	n	n	006E	\n\ as in <i>night</i>	خَانَ أَبَاد	Khānābād
36	ښ	06BC	ɳ	ɳ	1E49	retroflexive \n\	مَانِي	Māṇēy

	Script	Unicode value	Yaghoubi	BGN/PCGN	BGN/PCGN Unicode value	English Phonetic	Example	
37	و	0648	w	w	0077	\w\ as in <i>wait</i>	وَآخَانَ	Wākhān
38	ه	0647	h	h	0068	\h\ as in <i>hot</i>	هَرَات	Herāt
39	ی	0649	y	y	0079	\y\ as in <i>yard</i>	یَنگی قَلَعَه	Yangī Qal'ah

Vowel, Diphthong and Diacritical Characters

1	unwritten	064E	a, â	a	0061	short \a\ as in <i>fat, man</i>	جَلال آباد	Jalālābād
2	unwritten	0650	e, i	e, i ³	0065+0069	short \i\ or \e\ as in <i>sit</i> or <i>ten</i>	هَرَات پُلِ حِصَار	Herāt, Pul-e Hişār
3	unwritten	064F	o, u	o, u ⁴	006F, 0075	short \u\ or \o\ as in <i>put</i> or <i>no</i>	مُرغَاب کَابُل	Murghāb, Kābul
4	unwritten	0659	ə, ě	ê	00EA	Pashto vowel \ə\ as in <i>up, jug</i>	گَرْدُون	Gêrdōn

³ Both *e* and *i* are available to romanize this short vowel, depending on local usage and/or root language. In cases where the sound is uncertain, *i* is the default romanization in BGN/PCGN standardization procedures.

⁴ Both *o* and *u* are available to romanize this short vowel, depending on local usage and/or root language. In cases where the sound is uncertain, *u* is the default romanization in BGN/PCGN standardization procedures.

	Script	Unicode value	Yaghoubi	BGN/PCGN	BGN/PCGN Unicode value	English Phonetic	Example
5	آ ⁵	0622	ā	ā	0101	long \aa\ as in <i>car, fall</i>	آب بَند Āb Band
6	و	0648	ō	ō	014D	long \o\ as in <i>boat, toe</i>	سپين بُولدَاک Spīn Bōldak
7	و	0648	u, ū	ū	016B	long \oo\ as in <i>moon, rule</i>	بَلا بُلُوک Bālā Bulūk
8	و	0648	aw or āw	ow	006F+0077	diph. \ow\ as in <i>now, town</i>	جوزجَان Jowzjān
9	ی	06CC (064A)	i or ī	ī	012B	long \ee\ as in <i>meet, see</i>	غَزَنِي، سپين Ghaznī, Spīn
10	ی	06CC	ay or ē	ay or ē ⁶	0113, 0061+0079	diph. \ay\ as in <i>date, maid</i>	مَيَوَند، رِيگ Maywand, Rēg
11	ې	06D0	ē or ay	ē	0113	Pashto vowel \ā\ as in <i>rain, tail</i>	گَرْدېز Gardēz
12	ی	06CC	ay or āy	ay or aī ⁷	0113, 0061+0079, 0061+012B	diph. \ā\ as in <i>my, eye</i>	مَيَدَان شَهْر Maīdān Shahr

⁵ An *alif* with *mad* (آ) is written only in the initial position by BGN/PCGN standardization procedures, in keeping with Persian language family standards of use of the Arabic alphabet. The same letter written in a medial or final position is written ا . . .

⁶ The character ی should be romanized *ay* or *ē* according to its root language or local pronunciation. In case of uncertainty a reference source (such as the Fairchild Aerial Surveys map series, or a BGN/PCGN approved policy document/list of recommended spellings) should be consulted.

⁷ Both the combination *ay* and *aī* are available to romanize this character according to its root language or local pronunciation. In cases where the sound is uncertain *ay* is the default romanization in BGN/PCGN standardization procedures

	Script	Unicode value	Yaghoubi	BGN/PCGN	BGN/PCGN Unicode value	English Phonetic	Example	
13	اِ	06CC	ā	á	00E1	long \aa\ as in <i>car, fall</i>	دَنَدُو سَفَلِيْ	Dand-e Suflá
14	ی	06CD	e y or ay	êy	00EA+0079	Pashto vowel \əy\ word ending	خَوَكِي	Tsowkêy
15	unwritten	0621	'	'	2019	as in <i>day</i> or <i>yea</i>	هَوَائِي دَگَر	Hawāī Dagar
	unwritten	0674	-i-, e, ī	-e, -ye	012B	as in <i>day</i> or <i>yea</i> (glottal stop)	مَزَار شَرِيْف	Mazār-e Sharīf

Other Diacritical Marks and Language Conventions

1	ای	0627, 06CC	ay or āy	āy	0101	\ā\ as in <i>my, eye</i>	دایکندی	Dāykundī
2	یا	06CC, 0627	ya or yā	īā, iyā or yā ⁸	0101, 012B	\īa\ as in <i>Alexandria</i>	زیارت	Zīarat
					OR		غوریان	Ghōriyān
					OR		میا	Myā

⁸ The character combination ی should be romanized *īā, iyā or yā* according to its root language or local pronunciation. In case of uncertainty a reference source (such as the Fairchild Aerial Surveys map series, or a BGN/PCGN approved policy document or list of recommended spellings) should be consulted.

	Script	Unicode value	Yaghoubi	BGN/PCGN	BGN/PCGN Unicode value	English Phonetic	Example	
3	ا ^۰	0627				(no sound)	ارزگان	Uruzgān
4	و	0648	w	w	0648	(no sound)	خواجه	Khwājah
5	ئ	0626		'	2019	(glottal stop)	درواز پائین	Darwāz-e Pā'īn
6	ه	06C0	. . .h-e			\heh\ as in <i>hay</i>	کوه صافی	Kōh-e Şāfī
7	س	0651	(Consonant doubler)			(no sound)	محمد آغا	Muḥammad Āghā
8	الله ^۰ & see Note 5	FDF2		Allāh			عبدالله شيله	'Abdullāh Shēlah

⁹ This convention applies only to the initial *alif* which has no sound of its own, serving as a seat for an unwritten short vowel. This is in keeping with Persian language family standards of use of the Arabic alphabet whereby words that begin or end with an unwritten short vowel must utilize a silent consonant to carry the sound in written form.

NOTES

- 1 Unicode values listed in this table are required for all BGN/PCGN production, to ensure standardization and to minimize confusion from competing representations of a given letter. Notably, the Persian ک (Unicode value 06A9) is prescribed rather than the Arabic ك (Unicode value 0643 or FEDA or FED9), the Persian گ (Unicode value 06AF) is prescribed rather than گ (Unicode value 06AB) or گ (Unicode value 06B0) or ک (Unicode value 0643 or FEDA or FED9), and the Pashto letter خ (Unicode value 0681) is prescribed rather than the *heh* with a dot above and a dot below (Unicode value unknown). For the letter ی in its many variations, care must be exercised to follow this romanization guide's prescriptions to eliminate confusion for search engines and software. BGN/PCGN does not use the letter ی (Unicode value FEEF) in any Afghan word.
- 2 Occasionally the letter sequences كه, زه, سه, and گه occur without intervening vowels. They may be romanized *k·h*, *z·h*, *s·h*, and *g·h* in order to differentiate these romanizations from the digraphs *kh*, *zh*, *sh*, and *gh*, which are used to represent the letters خ, ژ, ش, and غ. Additionally, the Pashto letters خ and ح, routinely romanized *ts* and *dz*, may be alternatively romanized *tš* and *dž* when for special reasons it is desired that confusion be avoided with the character sequences تس (*ts*) and دز (*dz*), respectively.
- 3 The vagaries of written Afghan languages, as pertains to spacing and word division, are addressed as follows. Spaces may be added to or subtracted from Afghan words written in Arabic script, for the purposes of standardization. This is particularly relevant when the words are hand-written, are rendered “artistically”, or express other such non-standard flourishes, as long as the sense of the toponym, word, or phrase is not compromised. Analysts proceed in this regard conservatively.

Romanized toponyms are typically divided into constituent words (spaces and other grammatical rules applied) when those words can stand independently, for purposes of standardization and minimization of confusion, particularly in situations where Afghan writers are inconsistent in their application of spacing and word breaks. When the Afghan word or suffix is only used in combination with other nouns or adjectives, then it should be appended to the preceding word in its romanization. This includes (but is not limited to) *-ābād*, *-zāī*, *-zādah*, *-tū*, *-wand*, *-gaī*, *-kaī*, -

pūr, *-dēsh*, *-lar*, *-lī*, *-lū* and *ullāh*, as, for example, seen in *Raḥmatābād* (رحمت آباد) and *Raḥmatullāh* (رحمت الله), but *Raḥmat Khēl* (رحمتخیل) and *Raḥmat Shahr* (رحمتشهر).

- 4 The one-letter words د (Pashto) and و (Dari) are romanized *dê* and *wa*, respectively.
- 5 The word الله, meaning God, should always be romanized *Allāh*, except as specified in note 3. Note that the Unicode value FDF2 spells *Allāh*, but omits the *alif* in some common fonts, including Times New Roman. If in doubt, try in Arial Unicode MS to verify. Also note that the “dagger alif” (َ) above the second ل (lām) in the word الله, is not written but should be romanized *ā*, like a full-size alif.
- 6 In names of Arabic origin, the *l* of the definite article *al* is assimilated before the ‘sun letters’ *t, s̄, d, z̄, r, z, s, sh, ṣ, z̄, ṭ, ṣ̄*, and *n*. In its romanization, the article should be separated from the name it precedes and should not be capitalized except at the beginning of a name, e.g. جبل السراج → *Jabal as Sarāj*
- 7 Specifically Arabic names, particularly those of Islamic significance, often include a consonant which would ordinarily be written with a *tashdīd* (*shaddah*) in Arabic, which may be romanized by writing the equivalent consonant twice, e.g., مُحَمَّد → *Muḥammad*. This rule is particularly in force if the word is commonly romanized with a doubled consonant by Afghan naming authorities. For example, *ḥājī* is rarely seen in Afghan-produced romanizations with a doubled consonant, but *‘Abbās* is. The doubled *y* consonant is almost always seen thusly, as in *Sayyid* or *Qayyūm*.
- 8 In Afghan names which contain an *izāfah*, it should be romanized as *-e* or *-ye* according to common pronunciations, but generally, *-e* is used if the preceding word ends with a consonant other than silent *heh*, and *-ye* if the preceding word ends with a vowel sound. Scholarly sources indicate that *heh* is silent in *darah* and *qal‘ah* (thus *darah-ye*, *qal‘ah-ye*), but lightly spoken in *kōh* and *chāh* (thus *kōh-e*, *chāh-e*).
- 9 The character sequence خو, where followed by ِ or ى should be Romanized *khwā* or *khwī*, where the *w* is not pronounced, or only weakly so, as in *khwājah*.

10 Plural nouns ending in *-hā* or *-ān* should always be romanized as a single word, regardless of whether a space appears in a Perso-Arabic script source.

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

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