

Pushto

Letters of the Alphabet

Initial	Medial	Final	Alone	Romanization
ا	ل	ل	ا	omit (see Note 1)
ب	ب	ب	ب	b
پ	پ	پ	پ	p
ت	ت	ت	ت	t
ټ	ټ	ټ	ټ	ṭ
ث	ث	ث	ث	ṯ
ج	ج	ج	ج	j
چ	چ	چ	چ	ch
ه	ه	ه	ه	h
ځ	ځ	ځ	ځ	ś
ښ	ښ	ښ	ښ	z
څ	څ	څ	څ	kh
د	د	د	د	d
ډ	ډ	ډ	ډ	ḍ
ز	ز	ز	ز	z
ر	ر	ر	ر	r
ړ	ړ	ړ	ړ	r̄
س	س	س	س	s
ش	ش	ش	ش	sh
ښ	ښ	ښ	ښ	zh
ښ	ښ	ښ	ښ	zh
س	س	س	س	s
ش	ش	ش	ش	sh
ښ	ښ	ښ	ښ	śh
ص	ص	ص	ص	ṣ
ط	ط	ط	ط	ṭ
ظ	ظ	ظ	ظ	ṭ̣
ع	ع	ع	ع	ẓ
غ	غ	غ	غ	' (ayn)
ف	ف	ف	ف	gh
ق	ق	ق	ق	f
ک	ک	ک	ک	q
ګ	ګ	ګ	ګ	k
ګ	ګ	ګ	ګ	g
ل	ل	ل	ل	l
م	م	م	م	m
ن	ن	ن	ن	n
ښ	ښ	ښ	ښ	ṇ

نږ	نږ	نږ	نږ	ŋ
و	و	و	و	w (see Note 2)
ه	ه	ه ، ء	ه ، ء	h
ي	ي	ي ، ي	ي ، ي	y (see Note 3)

Vowels and Diphthongs (see Note 4)

Romanization	Initial	Medial	Final
a (see Note 5)	آ	ا	ا
u	أ	و	و
i	ا	ا	ا
ā	آ	آ	آ
á	-	-	(see Note 6) ا
ū	أ	و	و
ī	ا	ا	ا ، ا
o	او	و	و
e	ا ، ا	ا ، ا	ا ، ا ، ا ، ا
aw	او	و	و
ay	ا	ا	ا ، ا
ay	-	-	ا ، ا ، ا

Notes

1. For the use of ا (alif) see the table of vowels and diphthongs, and rules 1-4.
2. For other uses of و see the table of vowels and diphthongs, and rule 5.
3. For other forms and used of ا see the table of vowels and diphthongs, and rules 6 and 7.
4. Vowel points are rarely found in Pushto writing and printing. Vowels not indicated in the script are supplied in romanization.
5. In addition to the three short vowels recognized by the orthography of all languages using Arabic script, Pushto possesses a short, central vowel for which the sign ا (differentiated from ا) has sometimes been used. Both signs, whether written or inferred, may be romanized a. When this central vowel is indicated by ء (hamzah), the vowel is romanized a; the hamzah in this case is not represented in romanization. See rule 8(c) and (d). When it is desired for any reason to show the presence of a central vowel in a particular word, the romanization a may be used.
6. See rule 7(a).
7. See rule 7(b).

RULES OF APPLICATION

Letters Which May Be Romanized in Different Ways Depending on Their Context

1. **ا** (*alif*) is used:
 - (a) To indicate the presence of any of the short vowels *a, u, i*.

atāh	اتة
uḷas	الس
in'ām	انعام
 - (b) To represent the long vowel romanized *ā*. When so used, **ا** is usually written with **آ** (*maddah*).

ādam	آدم
------	-----
 - (c) To represent, in combination with **و** or **ى** :
 - (1) The long vowels romanized *ū* and *ī* respectively.

ūṣh	اوش
īrān	ايران
 - (2) The long vowels romanized *o* and *e* respectively.

os	اوس
eṣhal	اينسل
 - (3) The diphthongs romanized *aw* and *ay* respectively.

awwal	اول
ayyām	ايام
2. **ا** (*alif*), when medial or final, is used to represent the long vowel romanized *ā*.

bābā	بابا
------	------
3. **ا** (*alif*), when final, sometimes represents the combination romanized *an*. This value is confined to a few words of Arabic origin. See rule 10.
4. **ا** (*alif*), when used to support **ء** (*hamzah*), is not represented in romanization; see rule 8.
5. **و** is used to represent:
 - (a) The consonant romanized *w*.

wror	ورور
------	------

In Persian words, silent **و** following **خ** is retained in romanization.

khwushhāl	خوشحال
-----------	--------
 - (b) The long vowel romanized *ū*. See also rule 1(c1).

nūm	نوم
-----	-----
 - (c) The long vowel romanized *o*. See also rule 1(c2).

kor	کور
-----	-----

Push

- (c) When written over final ی or ے, ء indicates the diphthong romanized *ay*. See rule 7(b).
ghwārāy غوارای
- (d) In the orthography of Peshawar, ء when written over final ۛ indicates the short vowel romanized *a*. See the table of vowels and diphthongs, footnote 7.
naḥ نه
9. ̄ (maddah) is written over ۛ (alif) to indicate the long vowel romanized *ā*.
ākhir آخر
10. *Tanwīn* (written ̄) occurs in words of Arabic origin and represents a final syllable which is romanized *an*. *Tanwīn* is usually associate with ^l (alif), which is not represented in romanization. See rule 3.
fiṭṭatan فطرتا
11. ̄ (tashdīd) is represented by doubling the letter or digraph concerned.
awal اول

Grammatical Structure as It Affects Romanization

12. *Izāfah*. When two words are associated in the Persian construction known as *izāfah*, the first (the *muzāf*) is followed by an additional letter or syllable in romanization. This is added according to the following rules:
- (a) When the *muzāf* bears no special mark of *izāfah*, it is followed by *-i*.
dars-i 'ibrat درس عبرت
ṣāḥib-i mubārak صاحب مبارک
- (b) When the *muzāf* is marked by the addition of ء (*hamzah*), it is followed by *-i*.
- (c) When the *muzāf* is marked by the addition of ی, it is followed by *-yi*.

Affixes and Compounds

13. Affixes.

When the affix and the word with which it is connected grammatically are written separately in Pushto, the two are separated in romanization by a single prime (').

nā' ashnā نلشنا
war' astawī وراستوی

Note 1: The particle ډ (*da*), meaning "of", is always written as a separate word in romanization.
da daḥ ده د

Note 2: The Arabic article *al* is separated from what follows by a hyphen. The *l* of the article is retained in romanization, regardless of whether or not this *l* is assimilated in pronunciation to the sound of the following letter.

'Abd al-Rashīd عبد الرشید
Faḏl al-Ḥaqq فضل الحق
Niḏām al-Dīn نظام الدین

14. Compounds.

The elements of a compound (except a compound personal name) are separated in romanization by a single prime (').

	kitāb'khānah	کتابخانه
<i>but</i>	marīẓ'khānah	کتابخانه

Orthography of Pushto in Romanization

15. Capitalization.

- (a) Rules for the capitalization of English are followed, except that the Arabic article *al* is lower cased in all positions.
- (b) Diacritics are used with both capital and lower case letters.

16. Foreign words in a Pushto context, including Arabic, Persian, and Urdu words, are romanized according to the rules for Pushto. For vowels not indicated in the script, the Pushto vowels nearest those in the original pronunciation of the word are supplied in romanization.

Yūnīwarsiṭī Buk Ejaṅsī	یونیورسٹی بک ایجنسی
(<i>not</i> University Book Agency)	
Dārmastatar	دارمستتر
(<i>not</i> Darmesteter)	

SPECIAL CHARACTERS AND CHARACTER MODIFIERS IN ROMANIZATION

<i>Special Characters</i>	<i>Name</i>	<i>USMARC hexadecimal code</i>
'	soft sign (prime)	A7
ʾ	alif	AE
ʿ	ayn	B0
<i>Character Modifiers</i>	<i>Name</i>	<i>USMARC hexadecimal code</i>
◌́	acute	E2
◌̄	macron	E5
◌̇	dot above	E7
◌̈	dot below	F2
◌̉	double dot below	F3
◌̑	underscore	F6