

ROMANIZATION OF UYGHUR (Uighur)

BGN/PCGN 2023 Romanization Agreement

The BGN/PCGN system for Uyghur¹ is designed for the romanization of geographic names written in Uyghur Perso-Arabic script.²

Uyghur is a Turkic language, spoken by ethnic Uyghurs in parts of China, as well as in Kazakhstan, Kyrgyzstan, Tajikistan, and Uzbekistan.³ Uyghur has official status in China's Xinjiang Uyghur Autonomous Region.

The Uyghur alphabet utilizes the Perso-Arabic script and displays characters from the Arabic alphabet with the addition of four consonant characters (پ، چ، گ، ئ). Unlike in Arabic, all vowels are written in Uyghur, meaning that it is a fully alphabetic system of writing, and not an abjad. Uyghur script is written from right to left and does not make a distinction between upper and lower case.

This BGN/PCGN Agreement reflects the New Uyghur Latin Romanization System (NUL), designed and published by the Xinjiang Language and Writing Systems Committee on January 11 2008.⁴ NUL is based on, and is largely consistent with, the Latin Script for Uyghur (LSU)⁵ developed in 2000-2001 at Xinjiang University in Ürümchi. The NUL system, however, has not been officially utilized in Xinjiang, despite the Xinjiang government's 2008 official announcement of its approval. Nonetheless, the system has been widely adopted by Uyghurs both inside and outside China. The NUL possesses key qualities of a romanization system: its prevalence, its ease of digital representation, its faithfulness of phonemic principles of the Uyghur language, and its reversibility.

¹ There is inconsistent spelling of the language name, for clarity, this document will use the spelling 'Uyghur' exclusively in reference to the language and alphabet.

² The Uyghur Perso-Arabic script is the official writing system for Uyghur within Xinjiang, China.

³ Outside Xinjiang, Cyrillic is the writing system used by roughly half a million Uyghurs living in Central Asia. Source: Center for Languages of the Central Asian Region, Indiana University: <https://celcar.indiana.edu/materials/language-portal/uyghur/index.html>.

⁴ The Design on the Romanization of Uyghur Characters – 2008, the Minority Language and Writing Systems Committee of the Xinjiang Uyghur Autonomous Region; the original link (<http://www.xjyw.gov.cn/Web/article.aspx?ArticleID=1088>) has been removed, but it is still accessible via WayBack Machine: <https://web.archive.org/web/20131201102602/http://www.xjyw.gov.cn/Web/article.aspx?ArticleID=1088> [date accessed: November 22, 2021].

⁵ The LSU (Latin Script for Uyghur) is a Latin script based on the romanization of the Perso-Arabic script.

Table 1: Consonant Characters

	Script				Unicode value (Independent)	Romanization	Roman Unicode value (lower case)	Examples of Uyghur Toponyms	
	Final	Medial	Initial	Independent				Uyghur Script	Romanization
1	ب	ب	ب	ب	0628	b	0062	باركۆل، بۇلاقتاغ	Barköl, Bulaqtagh
2	پ	پ	پ	پ	067E	p	0070	پىچان، پوسكام	Pichan, Poskam
3	ت	ت	ت	ت	062A	t	0074	تۇرپان، تەڭرىتاغ	Turpan, Tengritagh
4	ج	ج	ج	ج	062C	j	006A	جەرنىبۇلاق، جىمسار	Jerenbulaq, Jimsar
5	چ	چ	چ	چ	0686	ch	0063+0068	چەرچەن، چاقىلىق	Cherchen, Chaqiliq
6	خ	خ	خ	خ	062E	x	0078	خوتەن، خوشۇت	Xoten, Xoshut
7	د	د	د	د	062F	d	0064	دۆربىلجىن، داۋانكارىز	Dörbiljin, Dawankariz
8	ر	ر	ر	ر	0631	r	0072	روشەنباغ، يارغۇل	Roshenbagh, Yarghul
9	ز	ز	ز	ز	0632	z	007A	زەرەپشان، زارات	Zerepshan, Zarat
10	ژ	ژ	ژ	ژ	0698	zh	007A+0068	ژورنال، ژانىر	Zhornal, Zhanir ⁶
11	س	س	س	س	0633	s	0073	ساغان، سايباغ	Saghan, Saybagh
12	ش	ش	ش	ش	0634	sh	0073+0068	شىخەنزە، شايار	Shixenze, Shayar
13	غ	غ	غ	غ	063A	gh	0067+0068	غۇلجا، مایتاغ	Ghulja, Maytagh
14	ف	ف	ف	ف	0641	f	0066	فەڭقاشان، فۇكاڭ	Fegashan, Fukang

⁶ BGN and PCGN are not aware of any Uyghur place names containing the letter ژ (zh). This letter is primarily used in spellings of non-native loan words, particularly of Russian origin. E.g. ژانىر → zhanir “genre”, ژورنال → zhornal “journal/magazine”.

	Script				Unicode value (Independent)	Romanization	Roman Unicode value (lower case)	Examples of Uyghur Toponyms	
	Final	Medial	Initial	Independent				Uyghur Script	Romanization
15	ق	ق	ق	ق	0642	q	0071	قاغلىق، قورغاس	Qaghiliq, Qorghas
16	ك	ك	ك	ك	0643	k	006B	كورلا، كەلپىن	Korla, Kelpin
17	گ	گ	گ	گ	06AF	g	0067	گۇما، گۇچۇڭ	Guma, Guchung
18	ڭ	ڭ	-	ڭ	06AD	ng ⁷	006E+0067	يېڭىسار، داۋانچىڭ	Yëngisar, Dawanching
19	ل	ل	ل	ل	0644	l	006C	لوپنۇر، لەنگەر	Lopnur, Len'ger ⁸
20	م	م	م	م	0645	m	006D	ماناس، مارالبېشى	Manas, Maralbëshi
21	ن	ن	ن	ن	0646	n	006E	نىيە، نىلقا	Niye, Nilqa
22	ھ	ھ	ھ	ھ	FBAA	h	0068	ئىدكاه، قاراهوجا	Idkah, Qarahoja
23	ۋ	ۋ	ۋ	ۋ	06CB	w	0077	ئالاتاۋ، پەيزىۋات	Alataw, Peyziwat
24 ⁹	ي	ي	ي	ي	064A	y	0079	يەكەن، يېڭىشەھەر	Yeken, Yëngisheher

⁷ This character/sound 'ng' does not occur word-initially. It is only written in independent, medial, and final forms.

⁸ Apostrophe distinguishes between n+g and the digraph 'ng'. See Note 2.

⁹ The consonant character ي (y) always includes two horizontal subdots; the vowel character ى (table 2, row 4) looks similar but has no dots; the vowel character ې (table 2, row 2) has two vertical subdots.

Table 2: Vowel Characters¹⁰

	Script				Unicode value (Independent)	Romanization	Roman Unicode value (lower case)	Examples of Uyghur Toponyms	
	Final	Medial	Initial	Independent (see Note 2)				Uyghur Script	Romanization
1	ا، ئا	ا	ئا	ئا، ا	0627, 0626+0627	a	0061	ئارال، ئالتاي	Aral, Altay
2	ئې، ئېي	ئې، پ	ئې، پ	ئې، ي	06D0, 0626+06D0	ë	00EB	ئېرىقبويى، يېڭىيەر	Ėriqboyi, Yëngiyer
3	ئە، ئەه	ئە، ئەه	ئە	ئە، ە	0647, 0626+0647	e	0065	ئەينىۋات، مەكىت	Eyniwat, Mekit
4	ئى، ئىي	ئى، د	ئى، د	ئى، ى	0649, 0626+0649	i	0069	ئىۋىرغول، تارىم	Iwirghol, Tarim
5	ئو، ئو	و	ئو	ئو، و	0648, 0626+0648	o	006F	ئونسۇ، ئورقۇ	Onsu, Orqu
6	ئۆ، ئۆ	ۆ	ئۆ	ئۆ، ۆ	06C6, 0626+06C6	ö	00F6	كۆكدالا، چۆچەك	Kökdala, Chöchek
7	ئۇ، ئۇ	ۇ	ئۇ	ئۇ، ۇ	06C7, 0626+06C7	u	0075	ئۇچتۇرپان، ئۇلۇغچات	Uchturpan, Ulughchat
8	ئۈ، ئۈ	ۈ	ئۈ	ئۈ، ۈ	06C8, 0626+06C8	ü	00FC	ئۈرۈمچى، بۈگۈر	Ürümchi, Bügür

¹⁰ See Appendix for Unicode table of Latin extended characters.

Table 3: Numerals

Numerals											
Arabic	0	1	2	3	4	5	6	7	8	9	10
Uyghur	نۆل	بىر	ئىككى	ئۈچ	تۆت	بەش	ئالتە	يەتتە	سەككىز	توققۇز	ئون
Romanization	nöl	bir	ikki	üch	töt	besh	alte	yette	sekkiz	toqquz	on

Notes on Numerals:

- Uyghur uses the Arabic numeral system, written from left to right, as in English. In the table above, numbers are spelled in Uyghur script.
- Uyghur cardinal numbers, when smaller than ten are usually spelled out and numbers larger than ten are written with numerals.
- Ordinal numbers may be spelled out in Uyghur, or written with numeral(s) followed by a hyphen (Unicode 002D) when occurring in the names of streets, highways, or buildings (e.g. schools). In the romanized name the hyphen should be retained but can also be translated as per the examples in parentheses. E.g., 7-مەكتەپ → 7-Mektep (7th Primary School), ئىككىنچى كوچا → Yigirminchi Kocha (20th Street).
- The low line also follows the numeral in dates written in Uyghur Perso-Arabic script. E.g. 11-ئاپرېل → 11-April (11th April), 1996-يىل → yil-1996 (the year 1996).

NOTES

- Some Uyghur script characters join to the following character and some do not. The medial forms of Uyghur script characters behave as in Arabic and Persian, with the exception of ئە and ە. Any character following either ئە or ە will be written in its initial form.

2. An apostrophe (‘ - Unicode 2019) is used in Uyghur in order to mark phonological breaks in sounds and words if:
- a syllable ends with a consonant and is immediately followed by another syllable that begins with a vowel, then these two syllables should be separated with an apostrophe.
For example: قىزىلئاۋات → Qizil’awat; باشئەگىم → Bash’egim; قوشئۆگۈز → Qosh’ögüz
 - Certain combinations of consonants - “n” (ن) + “g” (گ), “n” (ن) + “gh” (غ), and “z” (ز) + “h” (ھ) - may occur side by side but belong to two separate syllables. In which case an apostrophe should be placed between them to distinguish them from diagraphs such as “ng” (نگ) or “zh” (ژ).
For example: لەنگەر → Len’ger; قامچىلانغان → Qamchilan’ghan; ئەزھەر → Ez’her
3. a) Initial vowels are always preceded by *hamzah* (ء). Word-initially, the *hamzah* is not romanized: e.g., ئۇيغۇر → Uyghur, ئالتاي → Altay
- b) In Arabic or Persian loanwords, the *hamzah* is used to separate two vowels written in succession, indicating that the vowels belong to two separate syllables. The first vowel is written in the final form, followed by the second vowel written in its initial form (preceded by *hamzah*). In these cases, the *hamzah* should be romanized with an apostrophe (‘ - Unicode 2019): e.g., مۇئەللىم → mu’ellim, سائەت → sa’et
4. a) In Uyghur, six consonants ج، ڭ، ڭ، ڭ، ڭ، ڭ and ش mostly appear in words borrowed from foreign languages, including place names or personal names.
b) The doubling of a consonant is also an indicator of loanwords of Arabic, Persian, and Russian origins. E.g., نۇرمۇھەممەت → Nurmuhemmet
5. The ligature لا represents ل followed by ئا, and should be romanized **la**, e.g., كورلا → Korla

6. An inventory of symbols and letter-diacritic combinations, with their Unicode encoding, in addition to the unmodified letters of the basic Roman script is:

	Latin Extended	Unicode
1.	Ë	00CB
2.	ë	00EB
3.	Ö	00D6
4.	ö	00F6
5.	Ü	00DC
6.	ü	00FC
7.	'	2019
8.	-	002D

SPECIAL RULES

1. Word division in approved romanized forms shall follow word divisions in the original Uyghur script, without exception. It should be noted that the spelling of the generic name elements may vary according to the rules of vowel harmony and consonant assimilation in the Uyghur language. For example, the word for region, رايون (rayon) changes to رايونى (rayoni) in شىنجاڭ ئۇيغۇر ئاپتونوم رايونى → Shinjang Uyghur Aptonom Rayoni (Xinjiang Uyghur Autonomous Region).

APPENDIX:

Romanization of Uyghur Characters

Published by the Minority Languages and Writing Systems Committee of the Xinjiang Uyghur Autonomous Region in 2008

ئىككىنچى خىل لايىھە: « ئۇيغۇر يېزىقىدىكى ھەرپلەرنى لاتىن ھەرپلىرى بىلەن ئىپادىلەش لايىھىسى »
(تىل-يېزىق كومىتېتى 2008 - يىلى ئېلان قىلغان).

خ	چ	ج	ت	پ	ب	ئە	ئا
x	ch	j	t	p	b	e	a
ف	غ	ش	س	ژ	ز	ر	د
f	gh	sh	s	zh	z	r	d
ھ	ن	م	ل	ڭ	گ	ك	ق
h	n	m	l	ng	g	k	q
ي	ئى	ئې	ۋ	ئۆ	ئو	ئۇ	ئو
y	i	ë	v/w	ü	ö	u	o