

ROMANIZATION OF KURDISH

BGN/PCGN 2007

The tabulation below is applicable to the Kurdish language as a whole. It is based for the most part on the Hawar Roman alphabet used in the Library of Congress *Standard Kurdish Orthography Table*, but it also incorporates certain non-Hawar elements found in *A Kurdish-English Dictionary* (Taufiq Wahby & C J Edmonds, OUP, 1966)¹. The tabulation covers both major varieties of the Kurdish language: Kurmanji and Sorani. *Kurmanji* is spoken principally in Turkey and in Iraq, north of the Great Zab River (Dahūk/Dihok Governorate). It is generally written in Roman script (but sometimes in Perso-Arabic script in Iraq), and usually employs the Roman orthography as shown in the ‘Romanization’ column in the table below. *Sorani* is spoken principally in Iraq, south of the Great Zab River (Arbīl/Hewlêr and As Sulaymānīyah/Sîlêmanî governorates). It is generally written in Perso-Arabic script, and usually employs the Perso-Arabic script orthography shown in the tables below. As the grammar and script differ between different varieties of Kurdish, some of the notes and rules in this system may be applicable only to certain varieties of Kurdish.

Kurdish-language geographical names in Turkey will usually be found in Roman script, and so no romanization process will be required. The digraph options for consonant letters 6, 15, and 20 will not be encountered for such names. In Iraq, Syria, and Iran, Kurdish will usually be encountered in Perso-Arabic script as shown in the table below, in which case it may be romanized into the corresponding Roman-script form as shown in the Romanization column below, here employing the digraph options for consonant letters 6, 15, and 20, and taking into account the relevant notes and special rules. Kurdish geographical names for places and features outside Turkey, found in Roman-script form, should, where necessary and if possible, be tailored to fit the orthography of the ‘Romanization’ column below and should employ the digraph options for consonant letters 6, 15, and 20.

¹ 2022 Revision also makes use of W. M. Thackston’s Sorani Kurdish Reference Grammar, A Basic Guide to Kurdish Grammar by D.H. Kim and M.L. Chyet’s Kurdish-English Dictionary.

CONSONANT CHARACTERS

	Word-final Form	Word-medial Form	Word-initial Form	Independent	Unicode value (Independent)	Romanization ²	Roman Unicode Value ³	Notes	Example - Kurdish	Example - Romanization
1.	ء	ء	ء	ء	0621	'	2019	See notes 1 & 7	-	No example available
2.	ب	ب	ب	ب	0628	b	0062		خەبات	Xebat
3.	پ	پ	پ	پ	067E	p	0070		پاریزگای هەولێر	Parêzga-î Hewlêr
4.	ت	ت	ت	ت	062A	t	0074	See note 2	ماوەت	Mawet
5.	ج	ج	ج	ج	062C	c	0063		پێنجوێن	Pênciwên
6.	چ	چ	چ	چ	0686	ch / ç	0063+0068 / 00E7	See notes 3 & 7	چەمچەمەل	Chemchemat
7.	ح	ح	ح	ح	062D	h̄	1E27		ناحیەى بەحرکە	Naḥiye-î Beḥirke
8.	خ	خ	خ	خ	062E	x	0078		زاخۆ	Zaxo
9.	د	د	د	د	062F	d	0064		دەهۆک	Dihok
10.	ر	ر	ر	ر	0631	r	0072		کەلار	Kelar
11.	ڕ	ڕ	ڕ	ڕ	0695	r̄	1E5F		ڕانیە	Raniye
12.	ز	ز	ز	ز	0632	z	007A		زایى گەورە	Zab-î Gewre
13.	ژ	ژ	ژ	ژ	0698	j	006A		شاربازێر	Sharbajêr
14.	س	س	س	س	0633	s	0073		سڵیمانى	Slêmanî

² The Romanization column shows only lowercase forms but, when applying the table, uppercase and lowercase Roman letters as appropriate should be used. For upper-case Unicode values see note 7.

³ Unicode values for lower case characters provided in this table. For upper-case Unicode values see note 7.

	Word-final Form	Word-medial Form	Word-initial Form	Independent	Unicode value (Independent)	Romanization ²	Roman Unicode Value ³	Notes	Example - Kurdish	Example - Romanization
15.	ش	شد	شد	ش	0634	sh / ş	0073+0068 / 015F	See note 3	شارمزر	Sharezur
16.	ص	صد	صد	ص	0635	ş	1E63	See notes 2 & 7	-	No example available
17.	ض	خذ	خذ	ض	0636	đ	1E0D	See notes 2 & 7	-	No example available
18.	ط	ط	ط	ط	0637	ţ	1E6D	See notes 2 & 7	-	No example available
19.	ع	ع	ع	ع	0639	‘	2018	See note 7	عربەت	‘Erbet
20.	غ	غ	غ	غ	063A	gh / ħ	0067+0068 / 1E8D	See notes 3 & 7	قەرەداغ	Qeredagh
21.	ف	ف	ف	ف	0641	f	0066		فەقییان	Feqêyan
22.	ڤ	ڤ	ڤ	ڤ	06A4	v	0076	See note 4	ڤاکسین	vaksîn (=vaccine)
23.	ق	ق	ق	ق	0642	q	0071		شەڤلاره	Sheqławe
24.	ک	ک	ک	ک	06A9	k	006B		دوکان	Dukan
25.	گ	گ	گ	گ	06AF	g	0067		مێرگەسۆر	Mêrgesor
26.	ل	ل	ل	ل	0644	l	006C		سێمێل	Sêmêl
27.	ڵ	ڵ	ڵ	ڵ	06B5	ł	0142	Formerly written ڵ according to type available; may vary on older sources. See note 7	هەڵبەجە	Hełebce
28.	م	م	م	م	0645	m	006D		چەمچەمال	Chemchemaf
29.	ن	ن	ن	ن	0646	n	006E		ناوەندی هەولێر	Nawend-î Hewlêr

	Word-final Form	Word-medial Form	Word-initial Form	Independent	Unicode value (Independent)	Romanization ²	Roman Unicode Value ³	Notes	Example - Kurdish	Example - Romanization
30.	و	و	و	و	0648	w	0077	See note 4; see special rules 4 & 5	ماوەت	Mawet
31.	ه	هه	ه	ه	0647	h	0068	See note 5	ههولێر	Hewlêr
32.	ی	ی	ی	ی	06CC	y	0079	See special rules 4 & 5	کۆیه	Koye

VOWELS

	Kurdish	Unicode Value	Romanization ⁴	Roman Unicode Value ⁵	Notes	Example - Kurdish	Example - Romanization
1.	ئە/ه	06D5/0626+06D5	e	0065	See notes 1 & 5	عەربەت	'Erbet
2.	ئا/ا	0627/0626+0627	a	0061	See note 1	قەزای ئامێدی	Qeza-î Amêdî
3.	ئى/ی	06CC/0626+06CC	î	00EE	See notes 1, 6 & 7; see special rules 4 & 5	قەزای دەربەندیخان	Qeza-î Derbendîxan
4.	(no mark) ⁶	No value	i	0069	See special rule 2	پشدر	Pishder
5.	ئى/ئى	06CE/0626+06CE	ê	00EA	See note 7	سێمێل	Sêmêl
6.	ئو/و	0648/0626+0648	u	0075	See special rules 4 & 5	خورمأل	Xurmał
7.	ئوو/وو	0648+0648/ 0626+0648+0648	û	00FB	See note 7	مەخمور	Mexmûr
8.	ئو/ۆ	06C6/0626+06C6	o	006F		کۆیه	Koye
9.	ۆ	0648+0304	ö	00F6	Rare; previously written <i>وي</i> . See note 7	-	No example available
10.	ۆ	06CA	ü	00FC	Only appearing in some dialects and only in old sources. Often equated to /û/ (row 7 above). Sometimes written <i>يو</i> . See note 7.	-	No example available

⁴ See footnote 2.

⁵ See footnote 3.

⁶ A reference source may sometimes be required to ensure correct identification of the presence of this unwritten vowel.

NOTES

1. In pure Kurdish words, *hamza* is borne by *yā'* (ڤ) and occurs only before initial vowels; it is not romanized. Medial and final *hamza* in Arabic borrowings are romanized by ' (apostrophe – Unicode encoding 2019).
2. The letters ط ظ ض ص ز ت ز س ز س respectively do not occur in pure Kurdish words. In Arabic borrowings some writers retain these letters, others substitute ز ت ز س respectively. Only the letters ط ظ and ص are catered for in the Library of Congress tabulation, as reflected in lines 16-18 of the above Consonant table. Words of obvious Arabic origin occurring in a Kurdish toponymic environment will be treated as Kurdish rather than Arabic, as will words of other non-Kurdish origins.
3. The digraph options appearing in rows 6, 15, and 20 of the Consonants table should be used for Kurdish geographical names in Iraq, Iran, and Syria. The single-character options should be used for Kurdish geographical names in Turkey.
4. هـ is used to represent **v** in foreign words. Some southern Kurdish writers use it to represent the **v** in borrowings from northern Kurdish dialects. و is pronounced as a **v** in the north and as a **w** elsewhere.
5. *Hā'* can be used as a vowel or a consonant. The initial (هـ) and medial (هـ) forms are used for the consonant **h** (Consonant table, row 31), while the final (هـ) and independent (هـ) forms are used to represent the vowel **e** (Vowel table, row 1). Therefore, when used as a consonant, the final and independent forms of *hā'* will be seen as 'هـ' instead of 'هـ' and 'هـ', respectively. For example, مهه → *meh*, (= month)⁷. When used as 'e', the *hā'* behaves like the letters *alif* (ا), *wāw* (و), *dāl* (د), and *rā* (ر), in that it never joins to the following letter (*i.e.* it has no medial form). Consequently, the following letter will display the initial form, e.g. ههولێر → *Hewlêr* (unless there is only one following letter, in which case it will be written in the independent form, e.g. ماوهت → *Mawet*). As with other vowels (see special rules 2 and 3), initial **e** is preceded by the *kursî hamza*, yielding initial هـ, e.g. ئەهني → *enî* (= forehead).
6. In pure Kurdish words, the vowel ی is always long î, e.g. کانی ماسی → *Kanî Masê*. When it represents the *îzafe* in Sorani Kurdish⁸, it is also romanized î and joined by means of a hyphen to its preceding word e.g. پارێزگای دهۆک → *Parêzga-î Dihok*. In Sorani, the *îzafe* is used:
 - a. To link two parts of a possessive construction, e.g. پارێزگای دهۆک → *Parêzga-î Dihok* (= Dahuk Province, lit. the Province of Dahuk)
 - b. To link nouns and adjectives, e.g. هۆتیلی باش → *hotêl-î bash* (= good hotel)
 - c. With ordinal numbers, e.g. کەسی دووهم → *kes-î duwem* (= the second person)

⁷ This occurs in Kurmanji, when written in Perso-Arabic script so would rarely be encountered.

⁸ The *îzafe* is written differently in Kurmanji, and names should be reproduced as they are written in the Roman script.

7. An inventory of letter-diacritic combinations, used in addition to the unmodified letters of the basic Roman script in the Romanization of Kurdish, with their Unicode encoding, is:

	Latin Extended	Unicode		Latin Extended	Unicode
1.	Ê	00CA	18.	ê	00EA
2.	Î	00CE	19.	î	00EE
3.	Ö	00D6	20.	ö	00F6
4.	Û	00DB	21.	û	00FB
5.	Ü	00DC	22.	ü	00FC
6.	Ç	00C7	23.	ç	00E7
7.	Ɖ	1E0C	24.	ɖ	1E0D
8.	Ĥ	1E26	25.	ĥ	1E27
9.	Ŕ	1E5E	26.	ŗ	1E5F
10.	Ş	015E	27.	ş	015F
11.	Ș	1E62	28.	ș	1E63
12.	Ț	1E6C	29.	ț	1E6D
13.	Ħ	1E26	30.	ħ	1E27
14.	Ł	0141	31.	ł	0142
15.	Ǧ	1E8C	32.	ǧ	1E8D
16.	‘	2018	33.	’	2019
17.	·	00B7			

8. The Romanization column shows only lowercase forms but, when romanizing, uppercase and lowercase Roman letters as appropriate should be used.

SPECIAL RULES

1. The conjunction ‘and’ (و) should be rendered **u** if the preceding word ends in a consonant, and **w** if the preceding word ends in a vowel. It should be separated by spaces from the preceding and following words.
2. In the Perso-Arabic orthography for Kurdish, all vowels are written, with the exception of the short **i**, which is expressed with a *kasrah* under the preceding consonant (◌ِ). In Perso-Arabic script, the *kasrah* will rarely be written, e.g., كرن → *kirin* (= to do). Like all Kurdish vowels, the short **i** will be preceded by a *kursî hamza* (أ) if it appears at the beginning of a word (see 3 below; see row 4 of vowel table).
3. In the Perso-Arabic orthography for Kurdish, when a vowel comes at the beginning of a word, or when a vowel directly follows another vowel, a *kursî hamza* (أ) precedes it, e.g., ئاگر → *agir* (= fire).

4. A Kurdish word will never start with *alif* (ا). A Kurdish word may begin with a *yā'* (ی) or *wāw* (و), but only when they are used as a consonant, when they will be romanized as **y** and **w**, respectively.
5. When preceded by a consonant, *yā'* (ی) and *wāw* (و) should be romanized **î** and **u**, respectively. When preceded by a vowel (including short **i**, which is not written), *yā'* (ی) and *wāw* (و) should be romanized **y** and **w**, respectively.
6. The Arabic sign *shaddah* (ّ) denoting a doubled consonant is not used in Kurdish; doubled consonants, which are rare, are written twice, e.g. *موحهممه* → Muḥemmed; *موسه ننا* → Musanna. *Shaddah* might be used in Arabic borrowings but, as in unpointed Arabic, would generally be omitted.
7. Particles such as *له* 'le' (= at, in, on) and *به* 'be' (= to, for, by, with) should be written separately from their following word, e.g. *له کوردستان* → Le Kurdistanê "in Kurdistan".
8. Occasionally, the character sequences *چه*, *شه* and *گه* occur. They may be romanized c·h, s·h, and g·h in order to differentiate those romanizations from the digraphs ch, sh, and gh. (Middle dot, U+00B7)