

Khmer

The United Nations recommended system was approved in 1972 (resolution II/10), based on the system used by the BGN/PCGN (1972), this being a modified version of the Service Géographique Khmère (SGK) 1959 system. The table and notes on its usage were published in volume II of the conference report¹.

The system is used in many international cartographic products. In 1994-1995 the Gazetteer of Cambodia was produced using this system with some proposed modifications. However, since 1995 the Geography Department of the Ministry of Land Management and Urban Planning of Cambodia has been developing a new romanization system which was subsequently used in the second edition of the Gazetteer of Cambodia in 1996. This provisional system which does not contain any diacritical marks was further modified in 1997 and continues to be used in the country.

Khmer uses an alphasyllabic script whereby each character represents a syllable rather than one sound. Vowels and diphthongs are marked in two ways: as independent characters (used syllable-initially) and in an abbreviated form, to denote vowels after consonants. The romanization system is complicated by many additional rules. In Khmer writing word division is not ordinarily indicated and Khmer diacritical marks are often omitted. The romanization is generally not reversible to its original script form.

Romanization

I. Consonant characters

1	ក	kâ	8	ជ	chô	15	ណ	nâ
2	ខ	khâ	9	ឃ	chhô	16	ត	tâ
3	ក	kô	10	ញ	nhô ^A	17	ថ	thâ
4	ឃ	khô	11	ដ	dâ	18	ទ	tô
5	ង	ngô	12	ប	thâ	19	ធ	thô
6	ច	châ	13	ឌ	dô	20	ន	nô
7	ឆ	chhâ	14	ត	thô	21	ប	bâ ^B

22	ផ	phâ	27	វ	rô	32	ឡ	lâ
23	ព	pô	28	ល	lô	33	អ	'â ^C
24	ភ	phô	29	វ	vô			
25	ម	mô	30	ស	sâ			
26	យ	yô	31	ហ	hâ			

^A If used with a subscript character the lower element of the character is omitted: ញ *nhchô*.

^B If used with a subscript character the romanization of the character will be *p* (see note 4).

^C See note 5.

II. Subscript consonant characters (្ក stands for any consonant, see also note 3)

Character numbers correspond to those in Table I.

1	្ក	k	12	្ទ	th	23	្ថ	p
2	្ខ	kh	13	្ធ	d	24	្ត	ph
3	្គ	k	14	្ន	th	25	្ផ	m
4	្ឃ	kh	15	្ប	n	26	្ព	y
5	្ង	ng	16	្ផ	t	27	្ភ	r
6	្ច	ch	17	្ព	th	28	្ម	l
7	្ឆ	chh	18	្ភ	t	29	្យ	v
8	្ជ	ch	19	្ម	th	30	្រ	s
9	្ឈ	chh	20	្យ	n	31	្ល	h
10	្ញ	nh ^A	21	្រ	b	33	្វ	'
11	្ដ	d	22	្ល	ph			

^A With the same base character, full form is used (𑄛), in all other cases only the lower element is used as subscript, e.g. 𑄛 *khnhâ*.

III. Independent vowel characters

1	𑄛	ě	6	𑄛	roë	11	𑄛	aô ^B
2	𑄛	ei	7	𑄛	lœ	12	𑄛	au
3	𑄛	õ, ǔ ^A	8	𑄛	lœ			
4	𑄛	âu	9	𑄛	ê			
5	𑄛	rœ	10	𑄛	ai			

^A See note 10.

^B Variation: 𑄛.

IV. Vocalic nuclei (𑄛 stands for any consonant character)

Where variants in romanization separated by a dash are given the first is to be used in the *â*-series and the second in the *ô*-series.

1	𑄛	a-éa ^A	7	𑄛	o-u	13	𑄛	ê
2	𑄛	ě-ǐ	8	𑄛	uõ	14	𑄛	ai-ey
3	𑄛	ei-i	9	𑄛	aeu-eu	15	𑄛	aô-oũ
4	𑄛	œ	10	𑄛	œă	16	𑄛	au-õu
5	𑄛	œ	11	𑄛	iě			
6	𑄛	õ-ǔ	12	𑄛	é			

^A With the character 𑄛 a ligature 𑄛 *ba* is used (see note 4).

V. Shortened syllables and vocalic nuclei with *anuswara* or *visarga*

Where variants in romanization separated by a dash are given the one before the dash is to be used in the *â*-series and the one(s) after the dash in the *ô*-series. (ñ stands for any consonant character.)

1	ññ á-ó	5	ṇ̃ om-üm	9	ññ: éh
2	ñṇ̃ ä-oă,eă ^A	6	ṇ̣̃ äm-ôâm	10	ñṇ̃: aôh-ôăh
3	ṇ̃ṇ̃ ä-oă,eă ^A	7	ṇ̃: äh-eăh	11	ṇ̣̣̃ äng-eăng
4	ṇ̣̃ âm-um	8	ṇ̣̃: ôh-Ûh		

^A eă if followed by *k*, *ng*, *h*; otherwise oă.

Notes

1. Khmer consonants are divided into two series, the *â*-series and the *ô*-series, as indicated in the consonant table. With vocalic nuclei these consonants may produce different romanizations, as shown in the abbreviated vowel table: ñ *kâ*, ្រñ *krâ*, ñ *kô*, ្រñ *krô*. A Khmer consonant in syllable-final position, not accompanied by a vowel marker or by ̣, should generally be romanized without a vowel letter following: ññ *kâk*, អង្គ *'ângk* (exception: ព្រង្រ *pôngrô*, also written ព្រង្រ *pôngrôr* and ព្រង្រហ្រ *pôngrôh*).
2. The Khmer diacritical mark ̣ or ̣ written above an *â*-series consonant (except ឃ and ហ; see note 4) changes it to the *ô*-series: ឃ̣ *héang*. The diacritical mark ̣ written above an *ô*-series consonant changes it to the *â*-series: ង̣ *nhâng*. When either of these marks would conflict with another symbol written above a character, the mark ̣ or ̣ may be written in its place: ឃ̣ *hi*, ង̣ *dâmrei*. (These marks are frequently omitted in Khmer writing, particularly in words of Indic provenance.)
3. The second consonant of a Khmer graphic cluster is generally written below the base consonant in the special form called a “foot”: ង *khnâng*. There is no foot for the character ង្រ. The “feet” ង and ង usually represent the characters ង and ង, respectively, rather than ង and ង: ង *kdei*, ង្រ *kânthéay*, but ង្រ *kântráb*. A “foot” determines

the series of the following vocalic nucleus unless it is a nasal (ង *ng*, ញ *nh*, ណ *n*, ន *n*, ម *m*) or យ *y*, រ *r*, ល *l*, វ *v*, ស *s*. In that case, the base consonant determines the vocalic series, e.g. ខ្យង *khpông*, ល្អ *l'â*, ថ្ម *thmâ*, ស្វាយ *svay*. Syllable-final យ and ង sometimes appear as “feet”: ស្វាយ or ស្វាយ្យ *svay*, ទាំង or ទាំង្យ *teäng*. This practice appears to be optional and such irregular Khmer spellings are not reflected as such in romanization.

4. The combination ប plus ា is written បា *ba*. The latter character is a graphic device designed to prevent confusion with ហា *hâ*. The characters ប and បា with the diacritical mark ̂ are romanized *p* in the *â*-series, rather than as *b* in the *ô*-series: បង *pâng*, ប៉ាតា *patau*. The diacritical mark ̣ or ̤ is substituted where a conflict with another symbol written above a character would occur: បី *pei*. The characters ប and បា when accompanied by a “foot” are also romanized as *p* in the *â*-series, although the Khmer diacritical mark is generally omitted: ប្រែង *plêng*, ប្អ *p'â*, ប្រាប់ *prăb*.
5. The *â*-series consonant អ is romanized by means of an apostrophe ('): ក្រែក *k'êk*, ចង្កៀត *châng'iet*, រមែល *rô'el*, រឿ *'vei*, អង *'ang*. In word-initial position before a vowel, ' may be omitted: អង *ang*.
6. The Khmer diacritical mark ̇ appears only in two combinations (• exemplifies any consonant): •̇ (examples: បាត *bát*, ខ្យង *khpós*) and ា̇. The diacritical mark ̈ appears only in the combination ̈•. In the *â*-series both ា̇ and ̈• are romanized *ă*: ចាក *chăk* and ច័ក *chăk*. In the *ô*-series both ា̇ and ̈• are romanized *eă* when followed by *k*, *ng* or *h*; otherwise, they are romanized *oă*: រោក *rôpeăk*, មាត *moăt*, វ័ង្ស *veăngk*, ភ័រ្យ *phoăpv*.
7. The combination រ̈ is romanized *r* before a consonant: ថម *thôrm*. The combination រ̈̇ is romanized as *r* before a consonant preceded by a shortened vocalic nucleus: កាវ *koărr*.

8. The symbol ័ in syllable-initial position is ignored in romanization: ស័ *sâ*, ស័ស័ស័ស័ *sâsâs*.
In syllable-final position ័ indicates that the consonant is vowelled, i.e. followed by *â* in the *â*-series, by *ô* in the *ô*-series: តំណ័ *tâmnâ*, ពម័ *pômô*.
9. The diacritical mark ័ (which appears above characters and/or vowel markers which are not vocalized) is ignored in romanization: បុណ្យ័ *bôny*, ពោធិ៍ *pouthi*, ភូមិ័ *phumi*.
10. The independent character ឌី is romanized either *õ* or *ũ*. A reference source should be consulted where doubt arises.
11. Khmer characters may appear in different styles depending on the font used. In a frequently used *Âksâr Mul* (អក្សរមូល) style some of the consonant characters in Table I look quite different, e.g. 5 ឃ / ១ *ngô*, 6 ច / ២ *châ*, 8 ជ / ៣ *chô*, 11 ដ / ៤ *dâ*, 18 ឝ / ៥ *dô*, 20 ស / ៦ *nô*, 29 វ / ៧ *vô*.

Other systems of romanization

The **provisional romanization system by the Geography Department**² of the Ministry of Land Management and Urban Planning of Cambodia (1995, modified 1997) renders the consonants in the same way as described above but the presentation of vowels is somewhat different. As a rule, diacritical signs used in the UN system are omitted but the following equivalents are different. (Numbers refer to the tables and characters in the UN system. Some of the provided romanizations have no explicit counterparts in the UN system.)

No.	Char.	UN system	Provisional			
				III.12 ឌី	au	ov
III.3	ឌ	õ, ù	o	IV.4 ័	œ	oe-ue
III.-	ឌ	(not given)	ou	IV.5 ័	œ	eu-ueu
III.5	ឌ	rœ	rue	IV.7 ័	o-u	ou-u
III.6	ឌ	rœ	rueu	IV.10 ័	œă	oea
III.7	ឌ	lœ	lue	IV.13 ័	ê	ae-eae
III.8	ឌ	lœ	lueu	IV.16 ័	au-đu	au-ov
III.9	ឃ	ê	ae			

V.2 ័័ ă-oă,eă (not given)

V.10 ័័័ aôh-ôăh aoh-uoh

V.(2) ័័័ (ăk-eăk) ak-eak

V.- ័័ (not given) ak-eak

V.3 ័• ă-oă,eă a-oa

V.- ័័័ (not given) eh-is

V.(3) ័័័ (not given) ai-ey

Where variants in romanization separated by a dash are given the one before the dash is to be used in the *â*-series and the one(s) after the dash in the *ô*-series.

Before the last modification in 1997 to the system was made, the vowel *a* / *ă* in combinations ័័, ័័ and ័• (see UN system, Table IV, line 1, Table V, lines 2 and 3) was romanized as *aa*, and the vowel *é* in the combination ័័ (Table IV, line 12) was romanized as *ee*.

References

1. *Second United Nations Conference on the Standardization of Geographical Names*. London, 10-31 May 1972. Vol. II. *Technical papers*. United Nations. New York 1974, pp. 163-164.
2. *Geographical Names of the Kingdom of Cambodia*. Submitted by Cambodia. — 8th United Nations Conference on the Standardization of Geographical Names. Berlin, 27 August – 5 September 2002. Document E/CONF.94/INF.30/Corr.1.