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 report[1].

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.

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	6	ច châ	11	ឌ dâ	16	ត tâ	21	ប bâ[B]	26	យ yô	30	ស sâ
2	ខ khâ	7	ឆ chhâ	12	ឋ thâ	17	ថ thâ	22	ផ phâ	27	រ rô	31	ហ hâ
3	គ kô	8	ជ chô	13	ឍ dô	18	ទ tô	23	ព pô	28	ល lô	32	ឡ lâ
4	ឃ khô	9	ឈ chhô	14	ឍ thô	19	ធ thô	24	ភ phô	29	វ vô	33	អ ’â[C]
5	ង ngô	10	ញ nhô[A]	15	ណ nâ	20	ន nô	25	ម mô				

[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.

---

[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.

II. Subscript consonant characters (see also note 3)
Character numbers correspond to those in Table I.

1	k	6	ch	11	d	16	t	21	b	26	y	30	s
2	kh	7	chh	12	th	17	th	22	ph	27	r	31	h
3	k	8	ch	13	d	18	t	23	p	28	l	33	'
4	kh	9	chh	14	th	19	th	24	ph	29	v		
5	ng	10	nh[A]	15	n	20	n	25	m				

[A] With the same base character ⟨ ⟩ is used (ញ្ញ), in all other cases only the lower element is used as subscript, e.g. ⟨ ⟩ *khnhâ*.

III. Independent vowel characters

1	ĕ	3	ŏ,ŭ[A]	5	rœ̆	7	lœ̆	9	ê	11	aô[B]
2	ei	4	âu	6	rœ	8	lœ	10	ai	12	au

[A] See note 10.
[B] Variation: ឰ.

IV. Vocalic nuclei (• 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	a—éa[A]	5	œ	9	aeu—eu	13	ê
2	ĕ—ĭ	6	ŏ—ŭ	10	œă	14	ai—ey
3	ei—i	7	o—u	11	iĕ	15	aô—oŭ
4	œ̆	8	uŏ	12	é	16	au—ŏu

[A] With the character ប *p* a ligature ប៉ 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.

1	á—ó	4	âm—um	7	ăh—eăh	10	aôh—ŏăh
2	ă—oă,eă[A]	5	om—ŭm	8	ŏh—ŭh	11	ăng—eăng
3	ă—oă,eă[A]	6	ăm—ŏâm	9	éh		

[A] *eă* if followed by *k, ng, h*; otherwise *oă* (see note 6).

Notes

1. Khmer consonants are divided into 2 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: កិ *kâ*, ក្រិ *krâ*, គិ *kô*, គ្រិ *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: កក *kâk*, អង្ក *'ângk* (exception: ពង្រៅ *pôngrô*, also written ពង្រៅរ *pôngrôr* and ពង្រៅហ *pôngrôh*).

2. The Khmer diacritical mark ៑ or ៎ written above an *â*-series consonant (except ប and ប៉; see note 4) changes it to the *ô*-series: ហ្គាង *héang*. The diacritical mark ៎ written above an *ô*-series consonant changes it to the *â*-series: ញ្ហាង *nhâng*. When either of these marks would conflict with another symbol written above a character, the mark ់ or ុ may be written in its place: ហ្ហឺ *hi*, ដ្ម្រី *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": ខ្នាង *khnâng*. There is no foot for the character ឡ *lâ*. The "feet" ្ដ and ្ថ usually represent the characters ដ and ថ respectively, rather than គ and ប៉: ក្ដី *kdei*, កន្ថាយ *kânthéay*, but កន្ត្រប់ *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: ខ្ពង *khpông*, ល្អ *l'â*, ថ្ម *thmâ*, ស្វាយ *svay*. Syllable-final យ and ង sometimes appear as "feet": ស្វាយ or ស្វា្យ *svay*, ទាំង or ទា្ង *teăng*. 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: ប៉ង *pâng*, ប៉ាតៅ *patau*. The diacritical mark ់ or ុ is substituted where a conflict with another symbol written above a character would occur: ប៉ី *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: ភ្លេង *plêng*, ប្អ *p'â*, ប្រាប់ *prăb*.

5. The *â*-series consonant អ is romanized by means of an apostrophe ('): ក្អែក *k'êk*, ចង្អៀត *châng'iĕt*, រអិល *rô'ĕl*, អ្វី *'vei*, អាង *'ang*. In word-initial position before a vowel, ' may be omitted: អាង *ang*.

6. The Khmer diacritical mark ˋ appears only in two combinations: ◌ˋ (examples: បត់ *bát*, ខ្ពស់ *khpós*) and ◌ាˋ. The diacritical mark ˊ appears only in the combination ◌ˊ. In the *â*-series both ◌ាˋ and ◌ˊ are romanized *ă*: ចាត់ and ចត់ are both romanized *chăk*. In the *ô*-series both ◌ាˋ and ◌ˊ are romanized *eă* when followed by *k, ng* or *h*; otherwise, they are romanized *oă*: រពាក់ *rôpeăk*, មាត់ *moăt*, រ័ង្គ *veăngk*, ភព្វ *phoăpv*.

7. The combination ◌៌ is romanized *r* before a consonant: ធម៌ *thôrm*. The combination ◌៏ is romanized as *r* before a consonant preceded by a shortened vocalic nucleus: គារ៏ *koărr*.

8. The symbol ◌៊ in syllable-initial position is ignored in romanization: ស៊ *sâ*, ស៊សស *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: តំណ៊ *tâmnâ*, ពម៊ *pômô*.

9. The diacritical mark ◌៍ (which appears above characters and/or vowel markers which are not vocalized) is ignored in romanization: បុណ្យ *bŏny*, ពោធិ៍ *poŭthĭ*, ភូមិ៍ *phumĭ*.

10. The independent character ឩ is romanized either *ŏ* or *ŭ*. A reference source should be consulted where doubt arises.

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	No.	Char.	UN system	Provisional
III.3	ឧ	ŏ, ŭ	o	III.12	ឳ	au	ov
III.-	ឧៗ	(*not given*)	ou	IV.4	◌ើ	œ̆	oe—ue
III.5	ឫ	rœ̆	rue	IV.5	◌ឺ	œ	eu—ueu
III.6	ឬ	rœ	rueu	IV.7	◌ុ	o—u	ou—u
III.7	ឭ	lœ̆	lue	IV.10	◌ៀ	œă	oea
III.8	ឮ	lœ	lueu	IV.13	◌ែ	ê	ae—eae
III.9	ឯ	ê	ae	IV.16	◌ៅ	au—ŏu	au—ov

[2] *Geographical Names of the Kingdom of Cambodia.* Submitted by Cambodia. Eighth United Nations Conference on the Standardization of Geographical Names. Berlin, 27 August – 5 September 2002. Document E/CONF.94/INF.30

V.2	ាំ	ă—oă,eă	(a—ea?)	V.10	ោះ	aôh—ŏăh	aoh—uoh
V.(2)	ាក់	(ăk—eăk)	ak—eak	V.-	ៈ	(*not given*)	ak—eak
V.3	៉	ă—oă,eă	oa(?)—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. Uncertain romanization equivalents are indicated by a question mark.

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*.