

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

^A If used with a subscript character the lower element of the character is omitted: ក្រ 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 (ក្រ stands for any consonant, see also note 3)

Character numbers correspond to those in Table I.

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

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

III. Independent vowel characters

| | | | | | | | | |
|---|-----|-------------------|----|-----|----|----|-----|-----------------|
| 1 | ក្រ | ě | 6 | ក្រ | rœ | 11 | ក្រ | aô ^B |
| 2 | ក្រ | ei | 7 | ក្រ | lœ | 12 | ក្រ | au |
| 3 | ក្រ | ô, ŭ ^A | 8 | ក្រ | lœ | | | |
| 4 | ក្រ | âu | 9 | ក្រ | ê | | | |
| 5 | ក្រ | rœ | 10 | ក្រ | ai | | | |

^A See note 10.

^B Variation: ក្រ.

IV. Vocalic nuclei (ក្រ 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 | ក្រ | a-éa ^A | 2 | ក្រ | ě-î | 3 | ក្រ | ei-î |
|---|-----|-------------------|---|-----|-----|---|-----|------|

| | | | | | | | | |
|---|----|-----|----|------|--------|----|----|--------|
| 4 | ្ក | œ̃ | 9 | កើ | aeu–eu | 14 | កើ | ai–ey |
| 5 | ្ខ | œ | 10 | កើ្ក | œă | 15 | កើ | aô–oũ |
| 6 | ្គ | ö–ü | 11 | កើ្គ | ië | 16 | កើ | au–öu |
| 7 | ្ឃ | o–u | 12 | កើ | é | | | |
| 8 | ្ង | uö | 13 | កើ | ê | | | |

^A With the character ្ក a ligature ្ក្ក *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. (ក stands for any consonant character.)

| | | | | | |
|---|---------------------------|---|------------|----|---------------|
| 1 | កក់ á-ó | 5 | កំ om-ŭm | 9 | កេៈ éh |
| 2 | កាក់ ă-oă,eă ^A | 6 | កាំ ăm-ôăm | 10 | កេៈៈ aôh-ôăh |
| 3 | ក័ក ă-oă,eă ^A | 7 | កះ ăh-eăh | 11 | កាំង ăng-eăng |
| 4 | កំ âm-um | 8 | កុះ ôh-ŭh | | |

^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: ក *kâ*, ក្រ *krâ*, ក *kô*, ក្រ *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*, អង្គ *'ângk* (exception: ព្រង *pôngrô*, also written ព្រង្រ *pôngrôr* and ព្រងហ *pôngrôh*).
2. The Khmer diacritical mark ័ or ៑ written above an *â*-series consonant (except ឃ and ឃៈ; see note 4) changes it to the *ô*-series: ហ័ង *héang*. The diacritical mark ៑ written above an *ô*-series consonant changes it to the *â*-series: ហ្មង *nhâng*. When either of these marks would conflict with another symbol written above a character, the mark ្ក or ្ខ may be written in its place: ហ្មី *hi*, ដំរី *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”: ខ្មង *khnâng*. There is no foot for the character ឡ. The “feet” ្ក and ្ខ usually represent the characters ង and ង្រ, respectively, rather than ក and ឃ: ក្តី *kdei*, កន្ទាយ *kânthéay*, but កន្ទាប់ *kântráb*. A “foot” determines the series of the following vocalic nucleus unless it is a nasal (ង *ng*, ង្រ *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. ក្រង *khpông*, ក្រង្រ *l'â*, ក្រង្រ *thmâ*, ស្រយ *svay*. Syllable-final យ and ង sometimes appear as “feet”: ស្រយ or ស្រយ្រ *svay*, ទាំង or ទាំង្រ *teăng*. This practice appears to be optional and such irregular Khmer spellings are not reflected as such in romanization.
4. The combination ឃ plus ័ is written ឃ័ *ba*. The latter character is a graphic device designed to prevent confusion with ហ *hâ*. The characters ឃ and ឃ័ with the diacritical mark ័ are romanized *p* in the *â*-series, rather than as *b* in the *ô*-series: ឃង *pâng*, ឃ័កៅ *patau*. The diacritical mark ្ក or ្ខ is substituted where a conflict with another symbol written above a character would occur: ឃ្តី *pei*. The characters ឃ and ឃ័ when accompanied by a “foot” are also romanized as *p* in the *â*-series, although the Khmer diacritical mark is generally omitted: ឃ្តង *plêng*, ឃ្ត្រ *p'â*, ឃ្ត្រាប់ *prăb*.

5. The *â*-series consonant អ is romanized by means of an apostrophe ('): ក្រែក *k'ék*, ចង្កៀត *châng'iet*, រវើល *rô'el*, វើ *'vei*, អាង *'ang*. In word-initial position before a vowel, ' may be omitted: អាង *ang*.
6. The Khmer diacritical mark ៉ appears only in two combinations (• exemplifies any consonant): •៉ (examples: ប៉ាត *bát*, ខ្ពស់ *khpós*) and ា៉. The diacritical mark ៊ appears only in the combination ៊•. In the *â*-series both ា៉ and ៊• are romanized *ă*: ចាក់ and ច័ក are both romanized *chăk*. In the *ô*-series both ា៉ and ៊• are romanized *eă* when followed by *k*, *ng* or *h*; otherwise, they are romanized *oă*: រោក *rôpeăk*, ម៉ាត *moăt*, វើង *veăngk*, ភ័ញ *phoăpv*.
7. The combination ័ is romanized *r* before a consonant: ធ័ម *thôrm*. The combination ័័ is romanized as *r* before a consonant preceded by a shortened vocalic nucleus: កាវ័ *koărr*.
8. The symbol ័ in syllable-initial position is ignored in romanization: ស័ *sâ*, ស័សស *sâsâs*. In syllable-final position ័ indicates that the consonant is vocalised, 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: បុណ្យ *bôny*, ពោធិ៍ *pouthi*, ភូមិ *phumi*.
10. The independent character ឿ is romanized either *ô* 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 ង / ង *ngô*, 6 ច / ច *châ*, 8 ជ / ជ *chô*, 11 ដ / ដ *dâ*, 18 ទ / ទ *dô*, 20 ណ / ណ *nô*, 29 វ / វ *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 | IV.4 | ័ | œ | oe–ue |
| III.3 | ឌ | õ, ù | o | IV.5 | ័ | œ | eu–ueu |
| III.- | ឌ | (not given) | ou | IV.7 | ឡ | o–u | ou–u |
| III.5 | ឡ | rœ | rue | IV.10 | ឡៀ | œă | oea |
| III.6 | ឡ | rœ | rueu | IV.13 | ័ | ê | ae–eae |
| III.7 | ឡ | lœ | lue | IV.16 | ័ | au–ôu | au–ov |
| III.8 | ឡ | lœ | lueu | V.2 | ា៉ | ă–oă,eă | (not given) |
| III.9 | ង | ê | ae | V.(2) | ាក | (ăk–eăk) | ak–eak |
| III.12 | ឌ | au | ov | V.3 | ៊• | ă–oă,eă | a–oa |

V.(3) ឺយ (not given) ai–ey

V.- ឺ័័័័ (not given) eh–is

V.10 ឺ័័័័័័ aôh–ôăh aoh–uoh

V.- ័័័ (not given) ak–eak

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.