

UNITED NATIONS

Third Session of the Ad Hoc Group of Experts
on Geographical Names
2 to 12 February 1971

Information Paper No. 12

Transliteration of Khmer Writing

Transliteration of Khmer Writing

Carl Page

Geographic Names Division
Department of Technical Services
U.S. Army Topographic Command

December 1970
(corr. January 1971)

Transliteration of Khmer† Writing

1. The following symbology is employed in this paper:

- x = any Khmer consonant character or consonant character plus 'foot'.
- [] = graphic(ally), when enclosing Roman letters.
- [C] = any Roman consonant letter.
- [V] = any Roman vowel letter.
- = is (are) transliterated.
- * = a hypothetical construction. Each Khmer word cited following a colon is a Khmer lexical entry, unless marked * .

2. The Modified SGK 1959 System, published as the BGN/PCGN 1962 Transliteration System for Cambodian in Romanization Guide, BGN, 1967, is fairly serviceable with regard to representation of the phonemic elements of the Khmer language, and might, with a few additions, be adapted to represent without ambiguity the (more or less) conventional Khmer orthography.

3. Whereas idiosyncratic spelling in the established languages of Europe and the Americas is largely confined to place names, proprietary names, and to commercial advertising, spelling in Khmer is somewhat more flexible: ឆ្មើវ --[chhiěv] is also written ឆ្មើវ --[chiěv]. Both ស្វាយ and ស្វាយ --[svay]. Spellings such as ត្បែង (literally *[tngâm]), and ត្បែង (literally *[tngãm]) are graphic variants of the word ត្បែង --[teăng], and approach ideographic writing, as do such words as [tough trough through thorough] in English. Such items as the variant spellings of [teăng] can hardly be accommodated within a romanization system, and must be treated as special exceptions wherever identified. Fortunately they appear to be infrequent in Khmer writing.

4. Pages 2-6 present the Modified SGK System, as set forth in Romanization Guide, 1967.

† The term [Khmer] is here used in reference to things linguistically, ethnically, or otherwise culturally of the Khmer people. The term [Cambodian] is used only in reference to the country, Cambodia. The various Vietnamese, Thai, Lao, Upland Mon-Khmer, and Chinese speaking peoples who live in the country may be Cambodian in the latter sense, without being Khmer in any sense.

2. The symbol x represents any consonant. In the case of the column illustrating the vowels of the â-series, x represents only consonants of the â-series, whereas in the case of the column illustrating the vowels of the ô-series, x represents only consonants of the ô-series.
3. To supply the missing independent forms of the vowels of the â-series, the dependent vowel-symbols are written in conjunction with the 'vowel bearer', 𐌺𐌺. The 'vowel bearer' itself is not shown in transcription. Thus: 𐌺𐌺 â; 𐌺𐌺 a; 𐌺𐌺 ei. To supply the independent forms of the vowels of the ô-series, the dependent vowel-symbols are written in conjunction with the 'vowel bearer', modified by means of the symbol ˘ placed above it. When this would conflict with another superscribed symbol, the symbol 𐌺 is substituted, placed below the 'vowel bearer.' Thus: 𐌺𐌺 ˘ ô; 𐌺𐌺 ˘ éa; 𐌺𐌺 ˘ i.
4. The symbol ˘ written over a consonant of the â-series changes it to the ô-series. Thus: 𐌺𐌺 ˘ hō. The symbol ˘ written over a consonant of the ô-series changes it to the â-series. Thus: 𐌺𐌺 ˘ nhâ. When these symbols would conflict with another superscribed symbol, the symbol 𐌺 is substituted, placed below the consonant. Thus: 𐌺𐌺 ˘ hi; 𐌺𐌺 ˘ yei.
5. The symbol ˘ is also written above the symbol 𐌺 to produce the consonant p in the â-series. Thus: 𐌺 ˘ pâ. The symbol 𐌺 is also here substituted in cases where a conflict would result. Thus: 𐌺 ˘ pē. In the case of the cluster pr in the â-series, the diacritic mark is not used. Thus: 𐌺 ˘ prâ, not brâ.

5. Khmer consonants are (for historical reasons) written in two series, here called the \hat{a} -series, and the \hat{o} -series. A vocalic nucleus is indicated in part by a vowel marker, and in part by the consonant character in conjunction with which the vowel marker is written (except in the cases of the nine independent vowel characters, which appear alone). Approximately two-thirds of the 25 vowel markers represent different pronunciations in each of the two consonant series. In addition, the series of a Khmer consonant character may be reversed by the employment of a diacritical mark: $\tilde{x} \tilde{x} \tilde{x} \tilde{x}$. Unlike the languages of some neighboring peoples, Vietnamese, Thai, and Laotian, Khmer does not have morpheme-bound phonemic tone, thus eliminating an element that considerably complicates fully meaningful writing of these other languages in Roman script.

6. While the Modified SGK System attempts (not entirely successfully) to indicate the pronunciations of the various Khmer vocalic nuclei by the use of vowel digraphs and of Roman alphabet diacritical marks, it simplifies the writing of the consonants (without loss of phonemic representation) at the cost of accurate graphic representation. None of this, of course, takes into account sporadic inconsistencies in Khmer vowel representation, which cannot be accommodated in any transliteration system.

7. Inasmuch as neither the series of Khmer consonants nor the presence of Khmer diacritical marks is indicated in the Modified SGK System (although these may be inferred in perhaps 6 out of 10 instances on the basis of an adjacent vocalic nucleus), such ambiguities as the following inevitably occur in reverse transliteration:

[kâk] -- កក, កគ, កីក, កឹក.

[tuôl] -- តួល, តួឡ, តួល, តួឡ, តួល, តួឡ, តួល, តួឡ.

Clear indication of both consonant series and Khmer diacritical marks is thus necessary to unambiguous romanization, and is essential to accurate work where reference must be made to items lexicalized in Khmer, whether names (geographical or personal), words

9. Items a b c d f g i j m n improve graphic representation, making reversibility possible. Items e h i j k l n improve phonemic representation, making pronunciation of romanized Khmer easier for the user. Items e k l m n tend to bring names in the new GND gazetteer for Cambodia into closer harmony with those found on TOPOCOM 7011 and 7016 series 1/50,000 scale maps.

10. It may be objected that the suggested system further complicates the already complicated Modified SGK System. In spite of the fact that the suggested system is nowhere more complicated than the Khmer orthography from which it is derived, this is a legitimate objection with regard to applications in which indication of diacritical marks is deemed too demanding.

11. It is a purpose of this paper to suggest a system which may be satisfactory in the most demanding applications, such as bibliography, and which might be simplified for other applications in order to satisfy particular requirements. Inasmuch as the Modified SGK System appears to be the only romanization system in common use for Khmer, it may at this time be possible to establish a single transliteration system which will suit the needs of the various U.S. Government agencies, and which may obviate the proliferation of systems which has resulted in other instances, and which might eventuate here also.

Supplementary Notes on Changes in the Modified SGK System

1. In romanization of those Khmer consonants which are paired in the â-series and the ô-series in the Transliteration Table, underlining of the ô-series consonant has been suggested. In general the â-series consonant appears to be statistically the more common. The reverse is true only in the cases of 𑜀 --[n̄] and 𑜁 --[l̄], which appear much more frequently than 𑜂 --[n] and 𑜃 --[l].

It might be suggested that underlining of the 13 consonants in every appearance is unnecessary, since the vocalic nucleus following a consonant letter enables one to infer the series of the consonant in about 2/3 of the instances (in syllable initial position only). Omission of these underlines would not cause ambiguity (so long as Khmer diacritical marks are still indicated), but would introduce an element of inconsistency into an already complicated process.

Underlining of [m̄ n̄ ŋ̄] is suggested as a means of differentiating such romanizations as the following: 𑜀 --[k̄om̄]; 𑜁 --[k̄om̄]; 𑜂 --[r̄ōāŋ̄]; 𑜃 --[r̄ēāŋ̄].

2. The characters 𑜄 --[th̄], 𑜅 --[h̄], 𑜆 --[t̄h̄]; 𑜇 --[l̄h̄] are very rare.

3. 𑜈 in syllable final position does not appear with consistency in Khmer writing. Inasmuch as 𑜉 in syllable final position is not pronounced, Khmer speakers are occasionally uncertain whether to write it or not. This phenomenon is similar to the uncertainty as to whether [tranquil(l)ity] has one or two [l].

4. In every appearance the consonant 𑜊 represents a lottal stop. It has been customary in romanization of Khmer to ignore 𑜊 in word initial position, and often elsewhere as well. Representation of 𑜊 by [ʔ] in the â-series, and of 𑜋 by [ʔ̄] in the ô-series maintains a consistent symbol-to-symbol relationship throughout the system, as well as marking the presence of this consonant phoneme. This usage also differentiates graphically between word initial 𑜊 bearing a vowel marker, and an independent vowel character: 𑜌 --[ēs]; 𑜍 --[ēsei].

5. In all respects the suggested system varies as little as possible from the Modified SGK System. However, it will be noted that rather more diacritical marks are employed in representation of vocalic nuclei than are necessary to fully reversible transliteration. A number of these might be omitted in the interest of graphic economy, or, indeed, a different and more accurate system of graphic notation might be developed. Since most romanization of Khmer up to the present time has employed graphic conventions very similar to those of the Modified SGK System, it seems most practical to introduce only those changes necessary to unambiguous Roman transliteration.

6. If the possibility of introducing considerable change in the representation of Khmer vocalic nuclei is considered, the changes cited on the following page in the columns for vocalic nuclei and independent characters might be considered, with the following results:
 - a. Reduction of diacritical marks from 43 to 22 (at the cost of increase in the number of graphically different vocalic nuclei from 47 to 55, with retention of underline to indicate anusvara and visarga).
 - b. Much improved representation of vowel phonemes.
 - c. Full reversibility without underlining of the letters underlined in the consonant column, except when in syllable final position.

The following alterations would affect the Transliteration Guide, pp. 10-11:

- a. Minor changes in the spellings of some of the Khmer examples cited.
- b. The following paragraph added to note 1: "The letters underlined in the consonant column are underlined only in syllable final position."

The following alterations would affect the Pronunciation Guide, p. 12:

- a. Deletion of the words [crude] in item A, and [ordinarily] in line 1, item D.
- b. Line 3, item D: "[œ eu] [ɛ:u:] similar to Vietnamese [œ u] in [sœ su]."
- c. Line 5, item D: "[é é ê e] [e:e:ɛ:ɛ] similar to French [é é ê e] in [été été brêlle bref]."
- d. Deletion of line 8, item D: [6] [uə], etc.
- e. Line 9, item D: "In combinations of [i a o u œ eu] plus another vowel letter, this last vowel letter represents a vocalic glide or schwa (IPA [ə]). The diacritical mark [˘] appearing over such a vowel letter marks graphic distinctions only, and is of no significance to vowel pronunciation." 13

Bibliography

1. Tandar, S., Dictionnaire cambodgien-français, Imprimerie Albert Portail, Phnom Penh, 1935. Two volumes.
2. Guesdon, J., Dictionnaire cambodgien-français, Les Petits-Fils de Plon et Nourrit, Paris, 1930. Two volumes.
3. Institut Bouddhique, Dictionnaire cambodgien, second edition, Institut Roudhique, Phnom Penh, 1951. Two volumes.
4. Gorgoniyev, Y.A., The Khmer Language, Nauka Publishing House, Central Department of Literature, Moskow, 1966.
5. Jacobs, J.M., Introduction to Cambodian, Oxford University Press, London, 1968.