

Remarks on Sanskrit and Pali Loanwords in Khmer

Piotr Woźnica

INSTYTUT JĘZYKOZNAWSTWA, UNIWERSYTET IM. ADAMA MICKIEWICZA
AL. NIEPODLEGŁOŚCI 4, 61-874 POZNAŃ

pwoznica@amu.edu.pl

Abstract

The complicated orthography of Sanskrit and Pali loanwords in Khmer language poses a significant didactical obstacle, as the standard rules of Khmer orthography are irrelevant to those words.

After a brief historical introduction, the relations between Sanskrit and Sanskrit > Khmer loanwords' orthography are explained. Further on, a preliminary classification of Indic loanwords is provided, based on their morphological status and patterns of usage. The most important inference is that a relatively small number of immutable borrowings occurs in most frequently used compounds of Sanskrit and Pali origins.

This approach yields two potential solutions to the aforementioned problem. Firstly, a learner of Khmer should study the basics of Sanskrit and Pali orthography and phonology, as they have a tremendous impact on the way borrowings are spelled. Secondly, the learner should note that many Indic loanwords in Khmer are in fact compounds, and observe the distribution patterns of their individual constituents.

1 Notes on transcription and transliteration

Since compatibility of the phonemic representation used within this paper with previous literature is not material to the subject, all transliterated and transcribed¹ examples in this article are presented in IPA (2005 revision).

¹This includes phonemic and phonetic transcription, indicated by slash symbols '/' and square brackets '['] respectively

The transcription of Khmer language follows the Cambodian-English Dictionary by Robert Headley Jr. (Headley Jr. 1977). The transliteration of Sanskrit and Pali words is an adaptation of the IAST (International Alphabet of Sanskrit Transliteration) to the IPA standard. It should be born in mind, that the transliteration of Sanskrit and Pali examples does not accurately reflect the phonetics of those languages at their historical shape. It does provide a coarse definition of the phonetic structure of those languages and it may be thought of as transliteration and transcription in one, hence the use of slashes '/'.

Certain Khmer and nagari glyphs cannot be typeset on their own. They must instead be attached to a consonant symbol. Because of that, all examples of such glyphs in isolation will be presented attached to the first consonant letters of the two writing systems: 'ñ' for Khmer and 'क' for the nagari script.

2 Introduction

For a non-Khmer student, one of the first striking features of Khmer language is the abundance of Sanskrit and Pali loanwords in the vocabulary. The fact that the standard rules of Khmer orthography are seldom applicable to Indic² borrowings, poses a significant difficulty for beginning to pre-intermediate level students. A closer look at the process of introduction of Indic lexical items into Khmer shows that Sanskrit textbooks are among the sources to search for a solution of this problem.

The history of Indic influence over the Khmer kingdoms of Funan, Chenla and the Angkor Empire spanned more than ten centuries of cultural development³. The Khmer assimilated Brahmanism and Hinduism together with Sanskrit and the *Pallava* script (the Indic prototype of the future Khmer script, see figure 1 for a comparison of glyph shapes), which quickly became the language and writing system of the ruling class (Audric 1979). The sphere of *sacrum* was clearly separated from that of *profanum* in terms of the preferred language: the stone inscriptions that praised the god-king and his numerous virtues were written in Sanskrit, while documents concerning local economic issues were chiselled in Khmer.

With the arrival of new concepts in such fields as politics, religious beliefs, military theory, etc. came the need for new vocabulary in Khmer language. As all the required meanings already had lexical forms in Sanskrit, the Khmer adopted a tremendous amount of words in many specialist fields (with *linguistics* being one of them).

Pali rose to prominence in the Empire of Angkor together with Buddhism in the 12th and 13th centuries, during the reigns of Angkor's first buddhist

²The term 'Indic' refers to Sanskrit and Pali, unless explicitly stated otherwise.

³There is some controversy among scholars regarding the exact time span and degree of this influence (Vickery 2003).



Figure 1: Khmer script and its ancestors (from left to right): brahmi script, Pallava script and three styles of modern Khmer script (Masica 1991; Huffman 1970a).

kings — Dharanindravarman II and Jayavarman VII (Audric 1979). A substantial amount of vocabulary was assimilated into Khmer. As in the case of Sanskrit, Pali loanwords in Khmer are usually associated with the sphere of religion, however this is not a universal rule.

An important point to make here is, that Sanskrit and Pali are closely related genetically. Pairs of Sanskrit–Pali cognates exist as borrowings in modern Khmer, in the form of alternative spellings of the same word (similar to the example of *shirt* vs *skirt* in English). This phenomenon is particularly interesting, because one can observe how some phonological and phonotactical relations between Sanskrit and Pali have crystallized in Khmer orthography (see table 1).

Table 1: Example of a Sanskrit–Pali cognate pair borrowed into Khmer

Source:	Sanskrit	Pali
Transcription:	/d ^h arma/	/d ^h am:a/
Khmer orthography:	ធម៌	ធម្ម
Khmer transcription and translation:	/thəə/, ‘law, standard, norm’	

3 The borrowing mechanism

The Khmer script and *devanagari*⁴ are not directly related, however both are descendants of the ancient brahmi script and accurately follow its structure. In terms of grapheme-to-grapheme relations, the Khmer script is a superset of devanagari. It contains the equivalents of all graphemes used in devanagari to write Sanskrit, plus additions created to compensate for the vast differences between Sanskrit and Khmer phonological systems⁵. Any word written in the devanagari script, can be transliterated into Khmer with no loss of orthographical information. Conversely, the exact orthography of a Sanskrit word can (with some exceptions) be reconstructed from the orthography of its Khmer counterpart (see figure 2). The only exception to this rule are the Sanskrit graphemes श and ष, corresponding to the phonemes /ç/ and /ʃ/. Their Khmer counterparts ឆ and ឝ have been superseded by ផ in the modern orthography (see table 2).

Both the Khmer and the nagari script are abugidas⁶, i.e. every consonant grapheme has the sound value of a consonant and an inherent vowel.

e.g. the Khmer consonant grapheme ‘ក’ stands for /ka:/ and the nagari ‘क’ for /ka/

Vowel graphemes are attached to consonant graphemes and their sound value replaces the inherent vowel.

e.g. Khmer: ក + ា = កា /ka:/, nagari: क + ो = का /ko/

In addition, nagari uses special graphemes to represent word-initial vowel sounds (इ ई उ ऊ ए ऐ ओ औ ऋ ॠ ॡ ॢ, see table 3 for examples). Their counterparts in the Khmer script, the independent vowels, function as combined consonant and vowel graphemes.

The mechanism of borrowing lexical items from Sanskrit to Khmer was based on the fact that in the Funan–Chenla–Angkor period, both languages were written with a common script (Chandler 1991). As a consequence of the aforementioned phonological disparity between Sanskrit and Khmer, the borrowings were subjected to extensive changes in pronunciation. Certain phonological contrasts specific to Sanskrit were not perceived by Khmer speakers. The most obvious case is the retroflex vs dental contrast. Of the five Indic graphemes linked with retroflex consonants, three were (and to

⁴After the introduction of printing technology in India, nagari (devanagari) became the standard script for writing Sanskrit (Coulson 1976). Devanagari will be used for all Sanskrit orthography examples in this paper, thus expression ‘Sanskrit orthography’ refers to the way it is written in devanagari.

⁵Sanskrit and Khmer belong to separate language families (Indo-European and Austroasiatic, respectively).

⁶The term ‘abugida’ has been introduced by Peter T. Daniels (Daniels 1990)

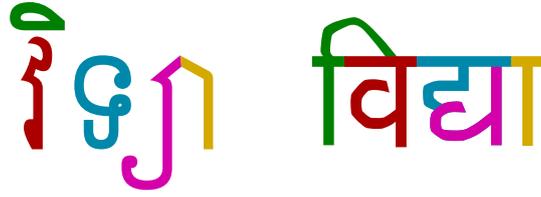


Figure 2: Grapheme-to-grapheme identity between Khmer (left — /vittjā~vicciə/) and Sanskrit (right — /vidjɑ:/) orthography — colours indicate the corresponding graphemes;

this day are) virtually redundant in native Khmer (see table 2 for a full reference). Other typical changes in pronunciation include the loss of a final syllable:

e.g. Sanskrit /a:-gne-ja/ > Khmer /ʔa:-kne:/ — ‘southeast’

Various vowel modifications are also possible (Coulson 1976; Huffman 1970a; Bühler 1977; Headley Jr. 1977).

The orthography-based assimilation of Indic vocabulary into Khmer, in combination with the aforementioned changes in pronunciation of loanwords, turned the Khmer script into a complicated system in terms of the relation between spelling and pronunciation. Rules that govern native orthography are hardly applicable to Indic loanwords and vice-versa. This is a substantial impediment for a Khmer language learner, especially since no optimization of the ‘memorize spelling and pronunciation, word by word’ method is present in traditional Khmer language teaching. Even grammar textbooks published recently in Cambodia scarcely address this problem (Cchun 2007). The following section is an attempt to provide an example approach to such optimization.

4 Identifying Indic borrowings in Khmer

Most loanwords in Khmer are immediately obvious as such for an advanced learner. Such identification is based on certain cues, which, if properly defined, should be usable even at the earliest stages of Khmer language learning.

The native Khmer lexicon is strictly mono- and disyllabic⁷. The word layout is /C(CC)V(C)/ for a monosyllabic, and /C(C)VC(CC)V(C)/ for a disyllabic word (elements in parentheses are optional), with the restriction that a short vowel may not occur in word-final position. The first syllable in a

⁷certain researchers are on the opinion, that Proto-Austroasiatic was essentially monosyllabic, whereas disyllabic lexical items are a relatively new, derivational development in languages such as Khmer (Donegan and Stampe 1983).

Table 2: Comparison between consonant graphemes in devanagari and modern Khmer script – grey colour indicates original retroflex graphemes, which became redundant with dentals in Khmer. Maroon indicates graphemes no longer used in Khmer. Note that this table does not contain graphemes specific to Khmer script.

Script →	devanagari					Khmer				
articulation ↓	Stops and nasals:									
Velar:	क	ख	ग	घ	ङ	ក	ខ	ក	យ	ង
Palatal:	च	छ	ज	झ	ञ	ច	ឆ	ជ	ឈ	ញ
Retroflex:	ट	ठ	ड	ढ	ण	ដ	ថ	ឌ	ណ	ណ
Dental:	त	थ	द	ध	न	ត	ថ	ទ	ណ	ន
Labial:	प	फ	ब	भ	म	ប	ផ	ព	ក	ម
	Other consonants:									
	य	र	ल	व	ह	យ	រ	ល	វ	ហ
	श	ष	स			ឆ	ថ	ស		
	Simplified phonemic transcription table (IPA):									
Velar:	/k/	/k ^h /	/g/	/g ^h /	/ŋ/	/k/	/kh/	/k/	/kh/	/ŋ/
Palatal:	/c/	/c ^h /	/ɟ/	/ɟ ^h /	/ɲ/	/c/	/ch/	/c/	/ch/	/ɲ/
Retroflex:	/ʈ/	/ʈ ^h /	/ɖ/	/ɖ ^h /	/ɳ/	/d/	/th/	/d/	/th/	/n/
Dental:	/t/	/t ^h /	/d/	/d ^h /	/n/	/t/	/th/	/t/	/th/	/n/
Labial:	/p/	/p ^h /	/b/	/b ^h /	/m/	/b/	/ph/	/p/	/ph/	/m/
	/y/	/r/	/l/	/v/	/h/	/y/	/r/	/l/	/v/	/h/
	/ç/	/ʃ/	/s/			/-/	/-/	/s/		

disyllabic word is unstressed and thus shortened and centralized (Huffman 1970a). Various types of deviations from the above rules occur in most foreign > Khmer loanwords:

- Polisyllabic words, eg. ករុណា /kaʔruʔna:/ ‘compassion, pity’ — ety. Sanskrit/Pali /karuṇa:/
- Long vowel or diphthong in the first syllable of a disyllabic word, e.g. ភាសា /phiəsa:/ ‘language’ — ety. Sanskrit /b^ha:ṣa:/

It must be stressed, that some loanwords are in perfect agreement with the aforementioned rules, e.g. ប្រទេស /prate:h/ ‘country, state’ — ety. Sanskrit /pradeṣa/

The spelling of loanwords in Khmer contains even more information on their etymology. The vast majority of consonant graphemes never occur in word-final position in native Khmer lexical items (see table 5 for examples). The Khmer script uses certain symbols exclusively in Indic borrowings (Huffman 1970a). The symbol ក្រិ indicates the grapheme used for /r/ in Sanskrit words, when it occurs initially in a consonant cluster. It is genetically related to the nagari symbol कर्, as in:

ទុក្ខិតិ /turəkuət/ ‘destitute’ — ety. Sanskrit दुर्गत /durgata/

Other spelling patterns that usually indicate Indic origin of a Khmer lexical item include:

- Clustered consonant graphemes at the end of a word (only the first grapheme is pronounced), e.g. មិត្តិ /mit/ ‘friend’ — ety. Sanskrit /mitta/
- Graphemes marked with the ក្រិ diacritic symbol, used to indicate unpronounced final graphemes in a word of foreign origin, e.g. ប្រយោជន៍ /prajaoc/ ‘usefulness’ — ety. Sanskrit /prajojana/
- Presence of independent vowel graphemes, e.g. ឯកិ /ʔaek/ ‘single’ — ety. Sanskrit /eka/

The case of independent vowels is somewhat problematic. The usage of independent vowels in Khmer spelling is not as strictly based on Sanskrit orthography as in the case of consonant and dependent vowel graphemes. The symbols related to Sanskrit long and short initial vowels of similar quality (e.g. ឿ vs ឺ or ឺ vs ឯ) are used in alternative spellings of certain Khmer lexical items, as in:

ឿតិហស vs ឺតិហស /ʔeʔteʔha:h/ — ety. Sanskrit /itiha:sa/ ‘chronicle’
ឺរាវណ vs ឯរាវណ /ʔe:ra:van/ — ety. Sanskrit /aira:va:ta/ ‘the elephant of Siva’

Table 3: Sanskrit vowel symbols and their Khmer counterparts, together with cognate pairs; Note that the relation is not fully regular

Khmer	Sanskrit		example		
independent vowels (word-initial vowels in Sanskrit)					
ឥ	इ	ឥន្ទ	/ʔintuʔ/	इन्दु	/indu/
ឦ	ई	ឦសាន	/ʔəjsa:n/	ईशान्य	/i:çɑ:nja/
ឧ	उ	ឧទាហរណ៍	/ʔuʔtiəhɑ:/	उदाहरण	/uda:haraŋa/
ឧ	ऊ	ឧន	/ʔu:n/	ऊन	/u:na/
ឧ/ឧ	ओ	ឧស្ម	/ʔaoh/	ओष्ठ	/oʃtʰa/
ឧ	औ	—	—	—	—
ឧ	ए	ឧក	/ʔaek/	एक	/eka/
ឧ	ऐ	ឧរាវណ	/ʔe:ra:van/	ऐरावात	/aira:va:ta/
dependent vowels					
កា	का	ការ	/ka:/	कार	/ka:ra/
កិ	कि	វិរាម	/viʔriəm/	विराम	/vira:ma/
កី	की	វេទី	/ve:ti:/	वेदी	/vedi:/
កុ	कु	គុណ	/kun/	डुण	/guna/
កូ	कू	ភូមិ	/phu:m/	भूमि	/bʰu:mi/
កោ	को	យោគីន	/jo:kin/	योगिन्	/jogin/
កេ	के	ទេស	/te:h/	देश	/deçɑ/

The glottal stop consonant (អ) combined with an appropriate dependent vowel is also a common spelling alternative for an independent vowel. Since the sound values of independent vowel graphemes vary from word to word, a single independent vowel grapheme may have up to three such alternative spellings (Huffman 1970b).

e.g. the independent vowel ឺ /ʔə/ or /ʔi/ or /ʔəj/ may be substituted with ឺ, ឺ or ឺ, depending on the word it is used in.

5 Preliminary classification of Indic borrowings in Khmer

Indic borrowings (including compounds) often have synonyms in the Khmer native lexicon. The usual difference is that loanwords belong to a literary or formal register, while native words are more colloquial (Headley Jr. 1977). Some of the borrowings' meanings are so fundamental, that they are virtually unavoidable in day-to-day communication.

e.g. ពេល /pe:l/ 'time' — ety. Sanskrit/Pali /vela:/

The loanwords can be divided into three groups, based on their occurrence in compounds and function in modern Khmer⁸. Note that this is not a categorization in the strict sense — the exact borders between the following types of borrowings cannot be clearly defined. The placement of certain borrowings in this simplified model is ambiguous.

1. **lexical borrowings:** items used in their full lexical meaning and not occurring in compounds, e.g. អវកាស — /ʔaʔveəʔka:h/ — 'outer space, space void';
2. **fragment borrowings:** items used both in compounds with other Sanskrit/Pali derived words, and in their full lexical meaning, e.g. សាស្ត្រ — /sa:h/ — 'science, forms the names of science disciplines'
3. **modifier borrowings:** items predominantly used as derivational affixes (they usually retain their full lexical meaning as well), also in conjunction with native words, with a varied degree of productivity, e.g. កា — /ka:/ — 'nominalizing prefix deriving nominal phrases from verbs or verb phrases'

The first case is the least interesting for the point of view adopted here. Borrowings belonging to this type cannot be dissected into meaningful parts, and at the same time do not form compounds. They include widely used vocabulary, especially in the written language.

⁸Listing all (or nearly all) the elements of the three groups described here is beyond the scope of this paper.

e.g. សប្តាហ៍ — /sap:da:/ — ‘week’

Indic loanwords belonging to the second category generally occur as constituents of compounds. The fact that they usually form compounds with other Sanskrit/Pali derived words, and the presently non-productive⁹ character of this process, suggest, that the compounds were native to the source language, rather than a Khmer development. The most striking feature of the fragment borrowings is that a relatively small number of them occurs repeatedly in numerous compounds (Headley Jr. 1977). Eight of the most common fragment borrowings (see table 4)¹⁰ are constituents of over a thousand compounds. The most frequently recurring fragments can be usually associated with certain types of meanings of the compounds they form, however they do not necessarily *always* adhere to such patterns. Fragment borrowings tend to have a fixed position within compounds (e.g. ក្រី occurs exclusively in final position). The orthographic and phonological form of a fragment borrowing is not influenced by the other part of the compound, however, some consonant clusters created at the boundary are divided by a weak epenthetic *schwa*, as in the following example:

/cət/ + /sa:h/ > /cətəsa:h/ — ‘heart, mind’ + ‘science’ > ‘psychology’

There are also cases of an otherwise unpronounced final grapheme, or graphemes, to become fully pronounced when occurring medially within a compound (Headley Jr. 1977).

e.g. ភូមិ + សាស្ត្រ > ភូមិសាស្ត្រ — /phu:m/ + /sa:h/ > /phu:mi?sa:h/ — ‘village, region’ + ‘science’ > ‘geography’

Modifier borrowings are special in that, they are used as derivational affixes in a productive manner¹¹, which is very untypical in the predominantly analytical Khmer morpho-syntax. This means that, unlike fragment borrowings, usage of modifier borrowings is not restricted by the etymology of the lexical item they are affixed to. The two most prominent loanwords of this type are កើត — /ka:/ — ‘happening, act, job, activity’ and កិរិយា — /phiəp/ — ‘state, condition, aspect, form, manner’ (Headley Jr. 1977; Huffman 1970b). The former is used to derive nouns (names of activities) from verbs and verbal phrases, while the latter derives abstract nouns from adjectives describing states. Theoretically, both can be used in their full lexical meaning, however, it is so broad, that it is difficult to create an example sensible from the point of view of day-to-day communication.

This quite unsophisticated model of how Indic loanwords function in modern Khmer was not thought up just for the sake of classification. The main purpose behind it, is to point out possible shortcuts in this problematic aspect of Khmer orthography, and optimize the didactic process.

⁹It is generally accepted among scholars that compounds of this kind are crystallized and unmodifiable in modern Khmer (Huffman 1970b).

¹⁰This list is by no means exhaustive.

¹¹The possible limits of this productivity require further detailed research.

Table 4: Sample list of the most common fragment borrowings, their full lexical meaning, and etymology (generalized meaning in compounds is given below an entry, if applicable)

Khmer		etymology	
កែវ	/ka:/	hand, maker	/kara/
forms names of occupations			
ធម្ម/ធម៌	/thɔə/	law	Sk. /d ^h arma/ ~ Pa. /d ^h am:a/
forms abstract nouns related to standards or habits			
ទេស	/teh/	country, land	Sk. /deṣa/
កម្ម	/kam/	action, deed	Pa. /kam:a/
forms abstract nouns related to actions			
ផល	/phal/	fruit, product	Sk. /p ^h ala/
forms nouns related to products, results or inferences			
ដ្ឋាន/ហ៊ាន	/tha:n/	location	Pa. /t ^h a:na/
forms names of locations and official buildings			
វិទ្យា/វិជ្ជា	/vitjiə/	knowledge	Sk. /vidja:/
forms names of science disciplines			
សាស្ត្រ	/sa:h/	science	Sk. /ṣa:stra/
forms names of science disciplines and school subjects			

6 Indic loanwords in Khmer glottodidactics

The Khmer script is not a complicated system in terms of grapheme to phoneme relations. Given a phonotactically valid sequence of phonemes, there is little to no choice as to its correct graphic representation¹². Conversely, the pronunciation of a Khmer word is in most cases immediately obvious from the spelling (Huffman 1970a).

The above is true only for the native lexicon of Khmer. But, the Khmer word ជ្រុំជ្រួត (*'jealousy'*, from Sanskrit /*pracaṇḍa*/) should be pronounced /*praca:nda*/ instead of /*pracan*/ according to the standard rules (Headley Jr. 1977). In native vocabulary, Khmer uses a single grapheme to represent each word-final stop. Therefore the word /*prak*/ (*'silver, money'*) can only be spelled ប្រាក់. However, in the case of Indic loanwords, the final stop /-k/ can be spelled with any of the four graphemes: ក, ខ, ក, ឃ (see table 5). Such examples are so numerous and widespread in the language, that treating them simply as exceptions from the general orthography rules is not even remotely feasible from a didactical point of view.

One way to address this problem is to formulate rules for Indic loanword spelling and pronunciation based on the borrowing-related phenomena outlined in section 3. The most simple example of such a rule would be that the inherent vowels¹³ /*a*/ and /*ɔ*/ are pronounced as /*a*/ and /*oə~eə*/ respectively, in words of Sanskrit or Pali origin.

Table 5: Examples of various spellings of word-final /-k/ in Indic loanwords

transcription	orthography	translation	etymology
/ʔaɛk/	ឯក	one, one person	/eka/
/muk/	មុខ	face, front	/muk ^h a/
/niək/	នាគ	naga	/na:ga/
/me:k/	មេឃ	sky, cloud	/meg ^h a/

A second (but in no way conflicting) approach stems from the divisibility

¹²As opposed to English, which uses far fewer graphemes, but has more complex orthographic rules.

¹³An inherent vowel is the default vowel sound pronounced if no vowel grapheme is present in a syllable.

of most Indic borrowings into smaller, recurring and meaningful elements. This approach is best illustrated with an example of a procedure of introducing new vocabulary in a didactic context:

1. presentation of a new lexical item (spelling, pronunciation, meanings): រ៉ូតឃ្លន់, /rəʔtʂəʊn/, ‘*automobile*’
2. presentation of the word’s etymology (original transliteration): Sanskrit /rat^ha/ ‘*cart, chariot*’ + Pali /janta/ ‘*device, machine*’
3. further elaboration on the occurrence of the individual elements in other compounds: រ៉ូត is the first element of compounds describing various ground-based vehicles. ឃ្លន់ is the last element of compounds related to machinery, especially engine-powered.

The original transcription of a loanword is important because it can serve as a speech-based reference for the Khmer spelling of the word¹⁴. In the above example, the spelling cannot be inferred from the pronunciation — in a native Khmer word the pronunciation /rəʔtʂəʊn/ should yield the spelling រ៉តឃ្លន់ instead of រ៉ូតឃ្លន់. The transliteration of the Indic etymological source of a loanword gives full information about its Khmer spelling, thus providing a systematic link between the graphical and speech-based representations of a meaning.

For the language learner this means that, instead of a very large lexicon full of arbitrary exceptions in spelling, he/she will be memorizing a much smaller lexicon governed by consistent orthography rules. For the teacher this means, that he/she will not need to shrug off questions like “why does the word /prəhəʊh/ (*Thursday, the planet Jupiter*) contain three mute letters?”.

Bibliography

- Audric, J. (1979). *Angkor i Imperium Khmerskie*. Państwowy Instytut Wydawniczy.
- Bühler, G. (1977). *Podręcznik Sanskrytu*. Cambridge University Press.
- Cchun, L. (2007). *វេយ្យាករណ៍ខ្មែរ (Khmer Grammar)*. បណ្ណាគារបន្ទាយស្រី (Bantey Srey Publishing House).
- Chandler, D. (1991). *ប្រវត្តិសាស្ត្រខ្មែរ (A History of Cambodia)*. Westview Press.
- Coulson, M. (1976). *Sanskrit — An Introduction to the Classical Language*. Oxford University Press.

¹⁴Provided that the learner has some basic knowledge about Sanskrit and Pali phonology and orthography.

- Daniels, P. T. (1990). Fundamentals of grammar. *Journal of the American Oriental Society CX*, 727–731.
- Donegan, P. J. and D. Stampe (1983). Rhythm and the holistic organization of language structure. *Parasession on the Interplay of Phonology, Morphology and Syntax*, 337–353.
- Headley Jr., R. (1977). *Cambodian-English Dictionary*. Catholic University of America Press.
- Huffman, F. E. (1970a). *Cambodian System of Writing*. Yale University Press.
- Huffman, F. E. (1970b). *Modern Spoken Cambodian*. Yale University Press.
- Masica, C. P. (1991). *The Indo-Aryan Languages*. Twayne Publishers.
- Vickery, M. (2003). Funan reviewed: Deconstructing the ancients. *Bulletin de l'Ecole française d'Extrême-Orient XC*, 101–143.