

NOTE ON THE transliteration OF KHMER

Michel Antelme

Maître de conférences, Institut National des Langues et Civilisations Orientales, Paris

In order to understand why in some scholarly texts on Khmer studies one can find Khmer words written in Roman characters, in a system called “transliteration,” but which do not seem to relate to the pronunciation of these Khmer words, it will be necessary to have a quick historical and linguistic overview. First of all, words such as “transcription” and “transliteration” need to be defined. According to *The Oxford Companion to the English Language* (McArthur 1992), the word “transliteration” in English dates from the 1860s and is composed of “Latin *trans* – across, *lit(t)era* a letter, and *-ation*” and is “the action, process, or result of converting one set of signs to another, usually involving at least one set of alphabetic letters. Where two writing systems have a common base, such as for Polish and English, which use variants of the Roman alphabet, transliteration is unnecessary, despite differences in sound-symbol correspondence. Transliteration becomes necessary when the systems differ greatly.”

A nuance has to be made with the word “transcription.” Dictionaries do not really distinguish transcription from transliteration, but in linguistics usage, “transcription” is a system used to write down spoken words. So a “transcription” is a script which represents spoken words. Transcription is a way to render sounds, whereas “transliteration” is a more limited notion as it is an operation through which a script used for one language is transcribed into another script, i.e. each vowel or consonant symbol in a script has its counterpart in another script. Moreover, and this is not explained in dictionaries, Khmer language specialists make a second distinction between what is simply called “transcription” and “phonetic or phonological transcription,”¹ which is based on the International Phonetic Alphabet (IPA). Words in IPA given in this text are written in back slashes.

To give an example, the name of the Cambodian capital ភ្នំពេញ, is “Phnom Penh” in transcription, using letters of the Roman alphabet and imperfectly rendering the pronunciation of the name; this name is “bhnam beñ” in transliteration (the reason for this spelling is the very subject discussed in this paper); and /pʰnum pɛŋ/ in IPA using symbols based on the Roman alphabet and whose values, – often unknown to the layman, give to linguists an accurate idea of the pronunciation of words.

However for Khmer pronunciation there has not been any consensus yet as to the way to render all sounds, such that there are different systems of phonetic or phonological transcription such as Henderson’s (1952) (also used by Jacob, and later on by Lewitz/Pou with minor modifications), Huffman’s,

¹ Linguists also make a distinction between phonetics and phonology. I will not go into this issue which is not essential to the present paper.

etc.² These differences are due to dialectal variations, although linguists when rendering what they consider as “standard Khmer” tend to imagine their transcription to tend toward the most standard form. “Standard Khmer” being nearly an artificial form, each author is however not exempt of influences from his own dialect or his informer’s dialect, hence different systems.

Indianization of Cambodia

Cambodia underwent a process of indianization from the beginning of the Christian era. Institutions, arts, religious beliefs and rites as well as language received strong influence from the Indian subcontinent.

The first dated inscriptions in Khmer discovered so far, date from the sixth century A.D. The scripts used in Cambodia from the beginning of Cambodian history also come from India.

Indian scripts

All Indian scripts stem from one same script, the Brāhmī script. These scripts have been used to note down different languages. In the North of India, the Indo-Aryan linguistic group prevails, while in the South languages belong to the Dravidian group. Southern languages have borrowed substantially from Sanskrit, which belongs to Indo-Aryan.

North Indian languages seem to derive from one same language – or at least from one same source – which was further embodied into one fixed language called Sanskrit. Sanskrit has been used in both India and Cambodia but never supplanted local languages. It seems that there was a repartition between the fields these languages were used for (administrative writings, kings’ eulogies, etc.). In Cambodia the vernacular language also found in stone inscriptions is Khmer, and in India local languages used in everyday life, as well as in stone inscriptions, were called *prākṛt*-s.

Khmer scripts and their Indian roots

According to Vasundhara Filliozat (2002: 10), “all the scripts of all Indian languages are derived from the same Brāhmī script of Aśokan edicts.” These edicts of King Aśoka were about the ideals of Buddhism and were engraved on stones all over India in the third century B.C. What the script was before that time is beyond the scope of the present study.

So many scripts derive from the Aśokan Brāhmī inscriptions. At the beginning these scripts all shared the same patterns, as follows.

In Cambodia, the script used in stone inscriptions to note both Sanskrit and Khmer is derived from the Cālukya and Pallava scripts from the South of India. A second script, the Nāgarī script from the

² The phonological transcription used for modern Khmer is based on Huffman (1970a, 1970b). I have made only small modifications, as Huffman’s transcription seems to be based on a dialect which differs slightly from the one I know. For the supposed phonological transcription of old Khmer, see Jacob (1960), Jenner (1980, 1981) and Ferlus (1992).

North of India, has been used in a group of digraphic inscriptions, i.e. with one side in the Southern script and the other side in the Northern script. These inscriptions are in Sanskrit. In Cambodia, two scripts have derived from the Southern scripts. These are called *mūl* មូល (meaning “round”) and *jriēn* ជ្រៀង (meaning “bent”).³ These two scripts vary only slightly on the shape of some letters, the *mūl* is used for titles or to note Pāli, whereas *jriēn*, which is more cursive, is used for lay texts or usual writing.

In the Nāgarī script as well as the Cālukya–Pallava scripts, there is one written sign for each phoneme and each sign can have only one pronunciation, contrary to English where a letter like “g” can have different pronunciations, for instance in words like *general* or *gambler*.

First, there are signs representing isolated vowels, i.e. not attached to a consonant in pronunciation:⁴

a	ā	i	ī	u	ū	e	ai	o	au	r	ṛ	l	ī
អ	អា	ឥ	ឺ	ឧ	ឺ ⁵	ឯ	ឺ	ឺ	ឺ	រូ	រូ	លូ	ឺ

Then there are the consonants:⁶

ka	kha	ga	gha	ṅa	ក	ខ	ក	ឃ	ង
ca	cha	ja	jha	ṅa	ច	ឆ	ជ	ឈ	ញ
ṭa	ṭha	ḍa	ḍha	ṇa	ដ	ឋ	ឌ	ឍ	ណ
ta	tha	da	dha	na	ត	ថ	ទ	ធ	ន
pa	pha	ba	bha	ma	ប	ផ	ព	ភ	ម
ya	ra	la	va		យ	រ ⁷	ល	វ	
śa	ṣa	sa	ha		គ	ឋ	ស	ហ	

In Sanskrit, these signs were to be read and pronounced with a short “a” which is often pronounced as a schwa /ə/ (unstressed vowel) in modern Indian languages. This inherent vowel in the graphic sign necessitated the use of particular techniques to represent consonant clusters. In Devanāgarī script the second consonant was attached to the lower part of the first, such that the two merged as a new sign; whereas in the script derived from Cālukya and Pallava scripts, which then became the Khmer script or កម្ពុជាក្បួរ Kamvujākṣara, the second consonant was written under the first one, sometimes undergoing some changes in shape.

³ The *Vacanānukram Khmaer* វចនានុក្រមខ្មែរ or *Dictionnaire cambodgien* (Khmer Dictionary) distinguishes a third script very close to the *mūl* one, called *kham* ខម (in Siamese, this name refers to Ancient Khmers, and is also the name given by the Siamese to the Khmer script in usage in their country).

⁴ I give the graphic representation of these signs in Modern Khmer script. Even though the form of these signs has evolved for well over a millennium, they remain representative of the vowels and consonants of Sanskrit, and there is complete concordance with the signs representing the same values in Devanāgarī for instance.

⁵ The graphic Khmer sign is more correctly ឺ but as the pronunciation of this sign has acquired a new phonetic value in Khmer, some scholars later proposed another form: ឺ, which is used for Indo-Aryan words and is distinguished from the former one used for Khmer words.

⁶ The consonants śa and ṣa are no longer used in modern Khmer writing. The consonant ឡ which does not appear in the table belongs to Pāli, not to Sanskrit. Moreover, in modern Khmer, the sign for the isolated vowel “a” is considered as a consonantal sign, which is in accordance with phonetics as it represents the glottal stop in Khmer.

⁷ Ra has a different shape when it is followed directly by another consonant in pronunciation without any intermediate vowel. In this case, it is written as a superscript, i.e. above the consonant which follows in speech: រៈ.

These signs in modern Khmer are written as follows:

ka	kha	ga	gha	ṇa
ca	cha	ja	jha	ṇa
ṭa	ṭha	ḍa	ḍha	ṇa
ta	tha	da	dha	na
pa	pha	ba	bha	ma
ya	ra	la	va	
śa	ṣa	sa	ha	



When the vowel in a syllable starting with a single consonant or a consonant cluster is other than the inherent vowel, a graphic sign is attached to the consonant sign(s) to represent it. In modern Khmer script, the sign can be placed before, above, below, after or surrounding the consonant sign. For instance, below are the signs representing the vowels from Sanskrit:¹¹

ā	ī	ī	u	ū	e	ai	o	au	ṛ
័	័	័	័	័	័	័	័	័	័

Scholars believe that there were many more vowels in Khmer than in Sanskrit since pre-Angkorian times. Certain consonant subscripts, such as ya ័ and va ័, are thought to have been used to represent some vowels or diphthongs.

Indian scripts used in mainland Southeast Asia can be used to transcribe Sanskrit and Pāli words and texts. Some of these were progressively adapted to the phonetics of the vernacular languages they were used for, and slowly with time, some letters from Sanskrit or Pāli such as the series of retroflex consonants (ṭa, ṭha, ḍa, ḍha, ṇa), were forsaken. The current Lao alphabet, which has undergone repeated reforms, is an extreme of such simplifications. However a religious script, adapted to both Pāli and Lao, has been maintained.

Thus there is a complete correspondance in vowels and consonants used for the notation of Pāli in the Khmer *mūl* script and the Lao religious script (called Tham from Pāli ធម្ម *dhamma*). Many signs have however been added to the particular scripts to render the vernacular languages. For example, in Tham script signs have also been designed to render consonants which do not exist in Sanskrit or Pāli such as /f/.

⁸ The shape of subscript ṇa was ័ in stone inscriptions and has been retained now as such only when the first consonant in consonant clusters is also a ṇa: ṇṇ ័.

⁹ Nowadays, the subscript shape of ṭa has been forsaken and cannot be distinguished from the one of ta, but in the *Inscriptions Modernes d'Angkor*, in manuscripts from the beginning of the twentieth century and in the first issues of កម្ពុជសុរិយា *Kambujasuriyā*, its shape resembles the lower part of the ṭa in *mūl* script.

¹⁰ The shapes of the subscript for śa and ṣa do not appear in the វិចិត្រនិក្ខេប *Vacanānukram Khmaer* but can be found in Maspero (1915).

¹¹ As it seems that the vowels after ័ do not appear attached to a consonant in Khmer inscriptions and texts, they are not represented in a subscript shape in this paper.

The transliteration of Sanskrit in Roman characters

When Westerners in the nineteenth century felt the need to study Sanskrit texts, especially German scholars at the beginning and then English scholars, they coined a system to represent each vowel and consonant used in Sanskrit. Indian scripts clearly represent Sanskrit pronunciation phonetically. Contrary to English or French spellings which are etymological and frequently no longer match pronunciation, the reading of a Sanskrit text does not allow any inaccuracy in pronunciation. Sanskrit is a highly formalized language which became artificially rigidified over time, losing the fluidity of a living language. Although there are fluent speakers of Sanskrit, virtually no one speaks Sanskrit as a mother tongue. And it is a well known fact that pronunciation in a language changes with time. How then, one might ask, can we be sure of the pronunciation of Sanskrit words used a few milleniums ago, when speakers of Sanskrit, usually Brāhmins have different Indian languages as mother tongues? We can be sure of it thanks to texts which explain how consonants and vowels should be exactly pronounced. Description of Sanskrit, and in particular its pronunciation, is provided in great detail in Pāṇini's classic grammar of the language written in the sixth century B.C. (or earlier, or around 400 B.C. according to some). And also thanks to very strict methods of oral transmission.¹² Moreover, although Indian languages have undergone some phonetic change with time, many of them seem much more "conservative" in pronunciation than – for example – modern Latin languages such as French or Spanish whose pronunciation has undergone great change since the period that Latin was still a living language.

For instance, ā ॐ corresponds to /aa/, i.e. a long vowel; i ॐ to /i/, short vowel; while ī ॐ represents long /ii/, etc.

As for consonants, ka ॐ represents an unvoiced unaspirated guttural stop /k/; kha ॐ is an unvoiced aspirated guttural stop /kh/; ga ॐ is a voiced unaspirated guttural stop /g/; gha ॐ is a voiced aspirated guttural stop /gh/; ṇa ॐ is a guttural nasal /ŋ/, etc.

We shall see below that the Khmer vocalic system is richer than the Sanskrit one and very gradually the Khmers have added new signs for these vowels, by modifying existing signs and sometimes borrowing them from the Siamese who appeared to be particularly creative in this field. This is why scholars had to add new transliteration signs to the original transliteration system of Sanskrit in order to render Khmer texts. I shall not give an exhaustive list of scholars who have worked on this. Most significantly, we should note: Aymonier (1900, 1901, 1904), Martini (1942-45), Au Chhieng (1953) and Lewitz (1969). The transliteration proposed by Lewitz is the one used in this article. Other authors have added even more signs to try to get a transliteration as accurate as possible. For example Jenner and the Library of Congress make the distinction between ॐ ī, and ॐ which is composed of the isolated vowel "a" ॐ and the vowel sign for ī, i.e. the vowel sign used in combination with a consonant sign. To distinguish the latter, an apostrophe is used by Jenner: 'ī, and the letter q is used by the Library of Congress: qī.

¹² For a discussion of Sanskrit, its pronunciation and Pāṇini's work, see Allen (1953) and Staal (1988).

Pronunciation of Sanskrit words in Khmer

Any speaker of Khmer can only be surprised after these explanations about Sanskrit pronunciation, as a Sanskrit word such as *ganḡā* “water; the Ganges” written កង្កែវ in Khmer script is pronounced /kuəŋ-kia/ in standard Khmer, never /gəŋ-gaa/. To understand such a discrepancy, one must turn to historical phonology.

Pronunciation of pre-Angkorian and Angkorian dialects

Pronunciation of consonants

When the Khmers borrowed their script from India to write Sanskrit as well as Khmer, it is sure that ka ក was used to represent an unvoiced unaspirated guttural stop /k/, kha ខ an unvoiced aspirated guttural stop /kh/,¹³ ga ហ a voiced unaspirated guttural stop /g/, gha ហា a voiced aspirated guttural stop /gh/, ña ញ a guttural nasal /ɲ/, etc. Of course, it is possible that some consonants in Indian languages did not exist in Khmer and Khmer used those signs to represent phonetically close consonants. For instance va វ which is read as /w/ in some modern Indian languages, may have been /β/ in pre-Angkorian Khmer. However it seems that the Indian script was more or less sufficient to write down Khmer consonants and consonant clusters. As for final consonants, Khmer can only have one final consonant in pronunciation. Some consonants were doubled (or followed by their aspirated counterpart) perhaps used to shorten the preceding vowel, or to indicate the end of the word.

Pronunciation of vowels

It seems that since ancient times, Sanskrit vowel symbols could not represent all the vowels of Khmer. However the ancient Khmers made few innovations to more accurately represent the Khmer vocalic system, apart from the following combinations:

- ui ឺ, found only in one very common word: *gui* គឺ meaning “that is to say, i.e.” and which is *gī* គី in modern Khmer with the phonemic value: /kii/.¹⁴

- ya យ, yā យា, ye យេ in pre-Angkorian for /iiə/; ya យ, yya យ្យ for /iə/, and ya យ, yya យ្យ, yā យា for /iiə/ in Angkorian according to Jacob (1960).

- va វ, vā វា, vo វោ in pre-Angkorian for /uuə/; va វ for /uə/, and va វ, vā វា, vo វោ for /uuə/ in Angkorian according to Jacob (1960).

On the value these vowel symbols could have in old Khmer, we also have the works of Jenner (1980, 1981) but the latest research is that of Ferlus (1992) who generated the following after comparing different modern dialects and the epigraphical evidence:¹⁵

¹³ Linguists actually think that it represented a consonant cluster in Khmer, i.e. unvoiced unaspirated guttural stop /k/ plus spirant /h/. See Ferlus (1992: 82).

¹⁴ Jacob (1960) signals another word: *latui* លត្ថី, a personal name of unknown meaning.

¹⁵ The table below is differently organised from that of Ferlus (1992: 87). I put the graphic symbols first, followed by their phonetic values, whereas he gives phonetic values first to show how they were represented in the writing system.

Transliteration	Khmer script	Phonemic value in pre-Angkorian	Phonemic value in Angkorian
a (inherent vowel)	-	/a/ or /ɔ/	/a/ or /aa/ or /ɔ/ or /ɔɔ/
a (inherent vowel) + double consonant in writing as word -ending	-	/a/	
ā	័	/a/ or /aa/	/a/ or /aa/
i	័	/i/ or /ii/ or /i/ or /i/ or /i/ or /i/	/i/ or /ii/ or /i/ or /i/
iy	័		/ii/ or /i/
iy	័		/ii/ or /i/
ī	័	/ii/ or /i/	/ii/ or /i/
īy	័		/ii/
u	័	/u/ or /uu/ or /uo/	/u/ or /uu/
ui	័	/i/	
ū	័	/uu/	/uu/
e	័	/ə/ or /ie/ or /iə/ or /ce/ or /εε/ or /ia/	/ə/ or /ce/ or /εε/
[ai] ¹⁶	័	/aj/ or /aaj/	/aj/ or /aaj/
o	័	/ɔ/ or /ɔɔ/ or /oo/ or /uo/ or /ua/	/ɔ/ or /ɔɔ/ or /oo/
[au] ¹⁷	័	/aw/ or /aaw/	/aw/ or /aaw/
ya	័	/ia/	/ie/ or /ia/
yā	័		/ia/
ye	័	/ia/	
yya ¹⁸	័		/ie/ or /ia/
yyā ¹⁹	័		/ia/
va	័	/uo/ or /ua/	/uo/ or /ua/
vā	័	/ua/	/ua/
vo	័	/uo/	

THE RECONSTRUCTION OF VOWELS IN OLD KHMER
Based on Ferlus (1992).

¹⁶ This graphic symbol has not been studied by Ferlus, but when I compare old and modern words, it seems that the written distinction with *āy* ័ /-aa/, i.e. its supposedly long counterpart, was not really made. For instance, the female title *tai* ័ is probably to be found in Modern Khmer *māy* ័ /mā-daa/ "mother," and old Khmer *ai* ័ has probably given both *ai* ័ in *ai* ័ ័ /ʔaj dɛɛ/ "in, at (obsolete)" and *āy* ័ /ʔaaj/ "over here." It is also highly probable that this symbol represented the combination vowel + final consonant /j/ in speech, and it was sometimes followed by the graphic symbol for consonant y. This use is attested from the pre-Angkorian period to become nearly systematic in the middle period.

¹⁷ This graphic symbol has not been studied by Ferlus either, but when I compare old and modern words, it seems that the written distinction with *āy* ័ /-aaw/, i.e. its supposedly long counterpart, was not really made. The interpretation is the more difficult when for instance the word for "out, outside" is *krau* ័ /kraw/ in Central Khmer, but **krāv* ័ /kraaw/ in the Surin dialect, the word for "shirt" is *āv* ័ /ʔaaw/ in Central Khmer but **au* ័ /ʔaw/ in the Surin dialect.

¹⁸ This graphic combination is possible only when the first letter of the syllable is the *repha* subscript sign: ័.

¹⁹ *Ibid.*

There would appear to have been no organized effort in ancient times to consistently develop the writing system in accordance with the phonetic system. Nearly all the vowels in old Khmer can be noted with different symbols and one symbol can represent different sounds. It has taken centuries for the Khmers to adapt the alphabet to their vowel system. Actually, it was only in the twentieth century that the alphabet came to note in a consistent and rather satisfactory way the vocalic system, in part thanks to borrowings of diacritics from the Siamese neighbors who adapted the Khmer alphabet to their own language and who, from the very beginning of the formation of a Siamese political entity, started to invent new symbols and diacritics to match their language.

For the linguist, this lack of accuracy of the old Khmer writing system produces obstacles in attempts to reconstruct the phonology and phonemics of a word, or to find its meaning.

Middle Khmer: Unvoicing of Consonants, Bipartition of Voice Registers and Vocalic System

Consonants became progressively unvoiced in several Mon-Khmer languages.²⁰ In Khmer it occurred during the middle period, i.e. the Post-Angkorian period.

The voiced consonants *g* 𑜀 /g/, *j* 𑜁 /j/, *d* 𑜂 /d/, *b* 𑜃 /b/, and their aspirated counterparts, became unvoiced, and were merged in pronunciation with their unvoiced counterparts.

When voiced consonants disappeared from the language to become unvoiced:

ga 𑜀 /g/ became similar to *ka* 𑜁 /k/ in pronunciation.

gha 𑜂 /gh/ became similar to *kha* 𑜃 /kh/ in pronunciation.

ja 𑜁 /j/ became similar to *ca* 𑜂 /c/ in pronunciation.

jha 𑜃 /jh/ became similar to *cha* 𑜄 /ch/ in pronunciation.

da 𑜂 /d/ became similar to *ta* 𑜃 /t/ in pronunciation.

dha 𑜄 /dh/ became similar to *tha* 𑜅 /th/ in pronunciation.

ba 𑜃 /b/ became similar to *pa* 𑜄 /p/ in pronunciation (but *pa* 𑜄 also took the value /b/ (see below), so for /p/ in the first series, the following spelling is used: *pa* 𑜅).

bha 𑜄 /bh/ became similar to *pha* 𑜅 /ph/ in pronunciation.

In modern Khmer there are still two voiced consonants: 𑜆 and 𑜇,²¹ represented by the consonantal signs *pa* 𑜆 and *ta* 𑜇. This would mean that devoicing was not complete. One possibility is that in proto or

²⁰ And also in several Thai/Tai languages. Some other Mon-Khmer languages have, however, yet to be so affected.

²¹ In my transcription I actually write them as b and d for easy reading as this paper is not intended for phoneticians but for a broader audience.

old Khmer they were pre-glottalised consonants: ‘b and’ d.²² These two voiced consonants are never found in first position in an initial cluster. So when *pa* ប is written as the first consonant in a consonant cluster, it is pronounced /p/.²³

At this stage, the Khmer language phonologised a distinction between two voice registers conditioned by the consonant. The vowels following the series of former unvoiced consonants were pronounced with a “clear” voice (*samlēn srāl* សម្លេងស្រាល or *samlēn tūc* សម្លេងតូច in Khmer) whereas those following the former voiced consonants were pronounced with a “breathy” voice²⁴ (*samlēn dbñan* សម្លេងធ្ងន់ or *samlēn dham* សម្លេងធំ). This register distinction has now disappeared from most Khmer dialects, but is still present in, for example, the Khmer dialects of the Cardamom range.

The Khmer vocalic system was itself split into two, i.e. the vowels took different values when coming after a former unvoiced or voiced consonant. Here is one hypothetical sequence of events:

At the beginning, with the unvoicing of consonants, *kā* កា /kaa/ remained /kaa/, pronounced with a clear voice, whereas *gā* កា /gaa/ became /kàa/ with a breathy voice. Later, *gā* កា /kàa/ became /kiə/ retaining the breathy voice. This register was then lost in most dialects to become /kiə/. There was no change in spelling; the graphic signs simply had new phonetic values.

So the vocalic system nearly doubled. At the same time, some vowels also underwent other changes in their pronunciation due to a variety of factors such as the influence of the final consonant. Another example is that some words with /e/ changed to /ɛɛ/, then in Standard Khmer the /ɛɛ/ became /ae/ when associated with a former unvoiced consonant and /ɛɛ/ when associated with a former voiced consonant.

To give an idea of the doubling of the vowel system, below is a table with the traditional alphabetical order of vowels associated with consonants *ka* and *ga*.²⁵ Each syllable is given first in transliteration,

²² This suggestion comes from Pinnow (1957). I am not going to detail this point even though it deserves more reflection, as I cannot myself propose any hypothesis. One can notice that many Sanskrit and Pāli words borrowed into Khmer have their consonants t and p turned into /θ/ and /b/, such as *tārā* “star” which has become *tārā* តារា /daa-raa/ and is heard as such in careful speech and in songs (even though it becomes /tə-raa/ in fast speech as the first syllable is unstressed and even though the Khmer dictionary gives only the pronunciation /taa-raa/); and *pālī* “Pāli” which has become *pālī* បាលី /baa-lej/ in Modern Khmer; these words are *tārā* តារា /daa-raa/ and *pālī* បាលី /baa-li/ in Siamese. However in other words, these consonants have “retained” their “original” pronunciation, such as Pāli *pīti* បីតិ “happiness” which has become *peti* បេតិ /pac-tej/ in modern Khmer. This is probably due to different periods of borrowings and/or influence from Buddhist monks concerned with the proper pronunciation of a sacred language (on this subject, see Bizot (1993). Even though he focuses on Thai Buddhism, these religious reforms had a very strong impact on neighboring countries such as Cambodia). This can be seen even now. When Dhammayut monks (a reformist sect coming from Siam at the end of the nineteenth century) recite Pāli prayers their pronunciation of vowels is based on the Indian or more probably Singhalese pronunciation, and they try to do the same for the consonants, whereas the Mahānikāy monks pronounce these Pāli prayers according to the Khmer rules of pronunciation, with the exception of the Pāli p pronounced as such, whereas this has become /b/ in the common Khmer language. This has led to confusion, and some neologisms still are not fixed in pronunciation, such as Pāli *pātukamm* បាតុកម្ម “demonstration,” which is pronounced as /paa-tok-kam/ by some and /baa-tok-kam/ by others. As for “Khmer” words with the consonants t and p, it is probable that some are borrowings from non Indic languages, but others seem to be Khmer. So the written consonants t and p from Sanskrit were probably used for at least two different consonants in pronunciation.

²³ This is a very brief explanation of the phenomenon. This bipartition between the two registers can also lead to two different tones. For more details, see Henderson (1952) and Ferlus (1992: 62).

²⁴ We have good examples with borrowings from Sanskrit. These Sanskrit words were borrowed by Siamese via Khmer but retained their original vowel pronunciation. For instance, Sanskrit *betu* “fact, reason” is now pronounced /haet/ in Khmer and still /hēt/ in Siamese (the loss of the brief u at the end of the word is a natural phenomenon in the two languages).

²⁵ I make no mention of the two registers in the transcription, contrary to Henderson (1952) and Jacob (1968), as in the standard variety of Khmer I use, these are absent. The traditional table of vowels gives some combinations which actually correspond to vowel + consonant, or vowel + semi-vowel. /ə/ appears in *gəh* ក្រែង /kəh/ but also before final /r/ as in *gāk* ក្រែង /kəh/ and final /ŋ/ as in *gām* ក្រែង /kəh/; /o/ appears in the recitation in *gām* ក្រែង /koəm/, but also appears before all the final phonetic consonants other than /r/ and /ŋ/. Absent from the recitation table is /ə/ as in *gār* ក្រែង /kəh/, which has in writing a final r no longer pronounced in most dialects. But the final r is still present in some northern dialects where it reveals that it was the brief diphthong /oə/ which lengthened with the loss of the final consonant in pronunciation (in some northern dialects, it is pronounced /əh/, as the standard /oə/ corresponds to /əh/ in this area).

then in Khmer script, followed by phonetic transcription.

ka	ក	/kɑɑ/,	ga	គ	/kɔɔ/,
kā	កា	/kaa/,	gā	គា	/kiə/, ²⁶
ki	កិ	/keɿ/, ²⁷	gi	គិ	/kiɿ/,
kī	គិ	/kej/,	gī	គិ	/kiɪ/,
kj	គ	/kɛɿ/,	gi	គ	/kiɿ/,
kj	គ	/kɛɛ/,	gī	គ	/kiɪ/,
ku	កុ	/koɿ/,	gu	គុ	/kuɿ/,
kū	គុ	/koo/,	gū	គុ	/kuu/,
kuo	គូ	/kuə/,	guo	គូ	/kuə/,
koe	គើ	/kaə/,	goe	គើ	/kɔə/,
kj̄oe	គើ	/kiə/,	gīoe	គើ	/kiə/,
kīe	គើ	/kiə/,	gīe	គើ	/kiə/,
ke	គេ	/kee/,	ge	គេ	/keɛ/,
kae	គៃ	/kae/,	gae	គៃ	/keɛ/,
kai	គី	/kaj/,	gai	គី	/kej/,
ko	គោ	/kao/,	go	គោ	/kɔə/,
kau	គៅ	/kaw/,	gau	គៅ	/kɔw/,
kum̄	គុំ	/kom/,	gum̄	គុំ	/kum/,
kaṃ	កំ	/kɑm/,	gaṃ	គំ	/kum/, ²⁸
kām̄	កាំ	/kam/,	gām̄	គាំ	/kɔəm/,
kaḥ	កះ	/kah/,	gaḥ	គះ	/kɛəh/.

To illustrate this, below is an excerpt from a *cpāp* studied by Jenner (1976: 697) given first in modern Khmer script and spelling, followed by transliteration, followed then by a phonetic transcription of a hypothetical middle Khmer pronunciation before devoicing, and finally, by a phonetic transcription of standard modern Khmer pronunciation:

²⁶ Pronounced /keə/ in some dialects – for instance the Kompong Chhnang dialect – thus making the difference with *gīe* គើ /kiə/ and *kīe* គើ /kiə/.

²⁷ All syllables with a short vowel in stressed position must have a final consonant. In reading the alphabet, when no final consonant is written, the glottal stop is pronounced.

²⁸ In some peripheral dialects, the pronunciation is /kɔm/, thus different from *gum̄* គុំ /kum/.

ពាក្យនេះជាច្បាប់	ទូន្មានឱ្យគាត់	កូនចៅឯក្រោយ
អាពុកឥតទ្រព្យ	វិភាគទុកឱ្យ	តែពាក្យនេះដោយ
	ព្រះបាលីថា ។	
លោកថាភ្លើងភ្លឺ	មែនពិតមោះប៉ុ	ពុំស្មើសុរិយា [...]

Transliteration:

Bāky neh jā cpāp'	dūnmān oy gāp'	kūn cau e kroy
ābuk it draby	ṽi gāp' duk oy	tae bāky neh tōy
	brah pālī thā.	
Lok thā bhločn bhli	maen bit moh r̄	bum smoe suriyā [...]

Transcription in hypothesised middle Khmer:

baak neh jaa cpap	duunmaan ʔoj gap	kuun caw ²⁹ ʔee krooj
ʔaabuk ʔit drap ³⁰	vii gap duk ʔooj	tee baak neh dooj
	brah paalii thaa.	
look thaa blɛŋ blɛ	meen bit moh rɛ	bum smɛ surijaa [...]

Transcription in modern Standard Khmer:

piəʔ nih ciə c ^h bap	tuunmiən ʔaoj koəp	koon caw ʔae kraoj
ʔaapuk ʔət troəp	vəj koəp tuk ʔaoj	tae piəʔ nih daoj
	preəh baalej thaa.	
lœwʔ thaa p ^h lɔʔŋ p ^h lɛ	meen pit ³¹ muəh rɛ	pum smaə soʔreʔjaə [...]

Conclusion

Transliteration of Khmer texts in Roman characters proves to be useful on four points:

1. As Khmer has been noted down at least since the sixth century of the Christian era with Indian scripts and has borrowed heavily from Sanskrit and Pāli, and to a lesser extent from Indian Prākṛt-s, the transliteration of Khmer texts for historical and linguistic purposes helps to link all these languages. Moreover as neighboring languages such as Siamese, Lao, Burmese, Mon, etc., also use Indian scripts, sometimes borrowed from the same region in India or even via the Khmers, transliteration allows us to see how all these languages have adapted Indic words. As many different scripts have been adapted from the Brāhmī script and have evolved in different ways, people might know one or two or three different scripts but it is uncommon to find people who can read all Indian scripts.³² Transliteration is a solution to unite all these scripts. On a more phonological and phonetic ground, transliteration used alongside phonetic

²⁹ Jenner gives the pronunciation /caw/.

³⁰ Jenner gives the pronunciation /droəp/, but then it does not rhyme with /gap/.

³¹ In some Northern Khmer dialects, the pronunciation is clearly /pit/. In the Central dialects, the /i/ systematically turns into /i/ when followed by a final consonant with the exception of /ʔ/, /c/, /p/, /h/.

³² For a table of different Indian scripts, see for example <http://www.geocities.com/Athens/Academy/9594/ brahmi.html>. For a discussion of the origins of the Brāhmī script and phonological considerations, see <http://www.scrindian.com/sa-research/sa0aa19.htm>.

transcription helps to shed light on the ways all these languages have phonetically evolved. On the other hand, it is important to recall that those who write research papers in Khmer can note Sanskrit or Pāli words perfectly in Khmer characters, as the Khmer alphabet was in fact created, in part at least, to render Sanskrit. Pāli has also been written in Cambodia in Khmer script for centuries. Thus it is important that the new generation of Cambodians does not break this link with the past.

2. Since the twentieth century, Khmer writing has had relatively consistent spelling, which is at once phonetic and etymological. Khmer language itself has different dialects and different accents, although intercomprehension is not a problem. There is a standard norm of pronunciation which is the one heard on radio, television and taught at school. Although it is not entirely clear which regional accent the standard one is based upon, it is a well-known fact that the colloquial language of the capital Phnom Penh is not standard at all. It seems that the standard accent is from the center of Cambodia, or even maybe the south of Battambang province, though each region maintains its own accent, and in each region several minor accents also exist. There are more or less accepted norms, such as trilling the “r” which is not done in the Phnom Penh colloquial language. As for vowels, the issue is much more complex. Even when trying to speak with a “standard” – we should even say an “articulate” – pronunciation, Phnom Penh residents do not make the distinction between several vowels that are distinguished in writing and in some other spoken dialects. On the contrary, in Phnom Penh colloquial dialect, some vocalic distinction corresponds to writing, while in Siem Reap that same distinction will not be made, and so on. When in dictionaries, for instance in bilingual dictionaries, a phonetic or phonological transcription of Khmer words is given, it can be considered as nearly artificial, making many vocalic distinctions not covered by a single dialect. On the other hand, some distinctions are frequently not made in bilingual dictionaries which give phonetic transcription – even though spelling and dialectology prove that they exist. This is often due to the influence of one peculiar accent on the dictionary compilers.

To give examples, in formal speech the Phnom Penh dialect does not make the vocalic distinction between *rān* ពន់ /riən/ “merchandise stand” and *rīen* រៀន /riən/ “to learn,” whereas in the Surin dialect of Thailand, the first word is pronounced /riən/ and the second one /run/. In turn, some Surin dialects do not make the difference between the words *bhliēn* ភ្លៀង “rain; to rain” and *bhlei* ភ្លេង /phluŋ/ “music; song,” whereas the Phnom Penh dialect does make it as the latter is pronounced /phlēŋ/ as opposed to *bhliēn* ភ្លៀង /pʰliəŋ/. Spelling overcomes quite satisfactorily all these phonetic differences.

To give one last example of how important dialectal variations can be, a sentence like “the hen sits on eggs” *mān’ krāp bān* មាន់ក្រាបពង, which from a lexical viewpoint has no dialectal variations, will be pronounced:

/moən kraap pəwŋ/ in the so called “standard accent,”

/məən kraap puəŋ/ in Surin or in some parts of Siem Reap and Battambang provinces,

/moən kśəp pəwŋ/ in Phnom Penh,

/məən kraap pəwŋ/ in Moat Chrouk province (now in Vietnam and called Chau Doc in Vietnamese), etc.

So it appears that transliteration in a French or English text about Khmer studies is a good solution if one cannot include Khmer characters or if the text is addressed to readers who are not supposed to know Khmer script. A phonetic transcription based on one particular dialect can be added, but transliteration will help overcome dialectal phonetic differences.

3. Transliteration also subsequently helps to better understand linguistic relationships between ethnic groups speaking languages from different linguistic groups. For instance, Khmer belongs to the Mon-Khmer family whereas Siamese belongs to the Tai-Kadai family. These two languages have heavily borrowed from one another in terms of vocabulary and their productive relationship went up to the point of back-borrowing from Siamese to Khmer. Transliteration helps to see immediately words similar to both languages via borrowings even though current pronunciations diverge. Sometimes, some words common to both languages do not come from one or the other, or from Indic languages, but instead from other languages such as Chinese. Transliteration helps to see their common roots.

For instance, “to think” is គិត /kít/ in Khmer and คิด /khít/ in Siamese. In transliteration, the word in both languages is *git*. Due to the fact that both languages use Indic-based scripts transliteration shows that the word was originally roughly pronounced /git/ when one borrowed it from the other.

The word for “craftsman” is ជាតិ /ciəŋ/ in Khmer and ช่าง /châaŋ/ in Siamese. In transliteration, the word is *jān* in Khmer and *jā¹n* in Siamese.³³ This word, maybe of Chinese origin, was probably pronounced /ʔaaŋ/ when it was borrowed by both languages (regardless of the tone in Siamese).

4. Because the Khmers borrowed their script before the unvoicing of consonants, transliteration helps make the link with other Mon-Khmer languages which are more “conservative” phonetically. Some of these languages recently acquired their own script based on the Roman one, thanks to missionaries and linguists.

If one looks at lists of words in Mon-Khmer languages at random, one can find in Mnong Ro’lo’m (spoken in northeastern Cambodia and south central Vietnam) the word *jūt* “to wipe” (Blood 1976: 11) which can be compared to Khmer /cuut/ (same meaning) which is written ជូត and transliterated *jūt*. In Jeh, a North Bahnaric language of the Mon-Khmer family spoken in Kontum province in Vietnam, one can find the word *gap* (pronounced with a long vowel) “to hold (with slit stick)” (Gradin 1976: 27) which can be compared to Khmer /kiəp/ “to press, compress, hold by means of pliers or tweezers,” which is written កាប and transliterated *gāp*. In Mo’dra and Didrá, two To’drah dialects, a North Bahnaric language also spoken in Kontum province, the word for “three” is *pi* (long vowel) (Gregerson and Smith 1973: 175) to be compared with modern Khmer /bəj/ (same meaning), written បី and transliterated *pī*.

³³ The number in the Thai transliterated word is for the tonal diacritic.

References³⁴

- Allen, W. Sidney, 1953, *Phonetics in Ancient India*, Oxford University Press, London Oriental Series, vol. 1, London.
- Au Chhieng, 1953, *Catalogue du fonds khmer*, Bibliothèque nationale, Département des manuscrits, Paris.
- Aymonier, Étienne, 1900, *Le Cambodge, I, Le royaume actuel*, Ernest Leroux, Paris.
 ———— 1901, *Le Cambodge, II, Les provinces siamoises*, Ernest Leroux, Paris.
 ———— 1904, *Le Cambodge, III, Le groupe d'Angkor et l'histoire*, Ernest Leroux, Paris.
- Bizot, François, 1993, *Le Bouddhisme des Thaïs. Brève histoire de ses mouvements et de ses idées des origines à nos jours*, Éditions des Cahiers de France, Bangkok.
- Blood, Henry F., 1976, "The Phonemes of Uon Njuñ Mnong Ro'lo'm," *Mon-Khmer Studies*, V, edited by Kenneth Gregerson, David Thomas, The Summer Institute of Linguistics, Manila, pp. 5-23.
- Buddhist Institute ព្រះសាសនបណ្ឌិត្យ, 1967-68, *Dictionnaire cambodgien*, volumes I and II, 5th edition, Institut bouddhique, Phnom Penh.
 វិចិត្រក្រមខ្មែរ ភាគ ១ និង ភាគ ២ ព្រះពុទ្ធភ្នាលី ៥ ព.ស. ២៥១១-២៥១២ គ.ស. ១៩៦៧-១៩៦៨ ការផ្សព្វផ្សាយរបស់ព្រះសាសនបណ្ឌិត្យ ភ្នំពេញ ។
- Dhanan Chantrupanth and Chartchai Phromjakgarin, 1978, *Khmer (Surin)-Thai-English Dictionary*, Indigenous Languages of Thailand Research Project, Chulalongkorn University Language Institute, Bangkok.
- ทะนิน จันทรูปันธุ์ และ ชาตชาย พรหมจักรินทร์ พ.ศ. ๒๕๒๑ พจนานุกรมเขมร(สุรินทร์)-ไทย-อังกฤษ โครงการวิจัยภาษาไทยและภาษาพื้นเมืองถิ่นต่าง ๆ สถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย ก.ท.ม.
- Ferlus, Michel, 1992, "Essai de phonétique historique du khmer (du milieu du premier millénaire de notre ère à l'époque actuelle)," *Mon-Khmer Studies*, XXI, Institute of Language and Culture for Rural Development, Mahidol University at Salaya (Thailand), pp. 57-89.
- Filliozat, Vasundhara, 2002, "Indian Epigraphy and its influence on Cambodia," *សិក្សាបុរាណ Siksācākṛ*, No. 5, Center for Khmer Studies, Siem Reap, pp. 10-13.
- Gradin, Dwight, 1976, "Word Affixation in Jeh," *Mon-Khmer Studies*, V, edited by Kenneth Gregerson, David Thomas, The Summer Institute of Linguistics, Manila, pp. 25-42.
- Gregerson, Kenneth J. and Kenneth D. Smith, 1973, "The Development of To'drah Register," *Mon-Khmer Studies*, IV, Language Series No. 2, edited by David D. Thomas, Nguyê n-Dinh-Hoà, Center for Vietnamese Studies, Southern Illinois University, Carbondale, Illinois, pp. 143-184.

³⁴ Some of the works cited in the bibliography have not been cited in this paper. I have included them for readers interested in pursuing the topic.

Henderson, Eugenie J.A., 1952, "The Main Features of Cambodian Pronunciation," *Bulletin of the School of Oriental and African Studies*, vol. xiv (University of London), London, pp. 149-174.

Huffman, Franklin E., 1970a, *Modern Spoken Cambodian*, Yale University Press, London/New Haven.
——— 1970b, វិជ្ជាខេមរភាសា [*Vijjā khemarabhāsā*]. *Cambodian System of Writing and Beginning Reader*, Yale University Press, London/New Haven.

Jacob, Judith M., 1960, "The Structure of the Word in Old Khmer," *Bulletin of the School of Oriental and African Studies*, XXIII, 2, London.

——— 1968, *Introduction to Cambodian*, Oxford University Press, London.

Jenner, Philip N., 1976, "The Relative Dating of Some Khmer *Cpā'pa*," pp. 693-710 in P.N. Jenner, L.C. Thompson, S. Starosta, eds: *Austroasiatic Studies, Part II*, University Press of Hawai'i, Honolulu.

——— 1980, *A Chrestomathy of pre-Angkorian Khmer. I, Dated Inscriptions from the Seventh and Eighth Centuries (A.D. 611-781)*, Southeast Asia Paper No. 20, part I, University of Hawai'i, Asian Studies Program, Southeast Asian Studies, Center for Southeast Asian Studies, Honolulu.

——— 1981, *A Chrestomathy of pre-Angkorian Khmer. II, Lexicon of the Dated Inscriptions*, Southeast Asia Paper No. 20, part II, University of Hawai'i, Center for Asian and Pacific Studies, Southeast Asian Studies, Honolulu.

Lewitz, Saveros, 1969, "Note sur la translittération du cambodgien," *Bulletin de l'École française d'Extrême-Orient*, LV, Paris, pp. 163-169.

Long Seam ឡុង សៀម, [2000], *Dictionnaire du khmer ancien (D'après les inscriptions du Cambodge du vie-viie siècles)*, Académie royale du Cambodge, Institut de langue nationale, Université royale des Beaux-Arts, Faculté d'archéologie, Phnom Penh.

វិចិត្រក្រមខ្មែរបុរាណ (តាមសិលាចារឹកបុរាណសតវត្សទី ៦-៨) រាជបណ្ឌិតសភាកម្ពុជា វិទ្យាស្ថានភាសាជាតិ សាកលវិទ្យាល័យភូមិន្ទវិចិត្រសិល្បៈ មហាវិទ្យាល័យបុរាណវិទ្យា ភ្នំពេញ ។

Martini, François, 1942-45, "Aperçu phonologique du cambodgien," *Bulletin de la Société de Linguistique de Paris*, 42, pp. 112-131.

McArthur, Tom (Editor), 1992, *The Oxford companion to the English language*, Oxford University Press, Oxford, New York.

Maspero, Georges, 1915, *Grammaire de la langue khmère (cambodgien)*, Imprimerie nationale, Paris.

Monier-Williams, Monier, 1995, *A Sanskrit-English Dictionary, Etymologically and Philologically Arranged with Special Reference to Cognate Indo-European Languages*, Motilal Banarsidass Publishers Private Limited, Delhi, (1st ed.: 1899, Oxford University Press).

หอสมุดแห่งชาติ พ.ศ. ๒๕๔๓ หนังสือแบบอักษรโบราณ ส่วนภาษาโบราณ หอสมุดแห่งชาติ กรมศิลปากร ก.ท.ม. [National Library, 2000, *Book on Different Forms of Ancient Characters*, Section of Ancient Languages, National

Library, Department of Fine Arts, Bangkok.]

Pinnow H.J., 1957, "Sprachgeschichtliche Erwagungen zum Phonemsystem des Khmer," *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft*, 10 (4), pp. 378-391.

Pou, Saveros, 1992, *Dictionnaire vieux khmer – français – anglais, An Old Khmer – French – English Dictionary*, វចនានុក្រមខ្មែរចាស់- បារាំង- អង់គ្លេស, Centre de documentation et de recherche sur la civilisation khmère (Coll. "série B : Travaux et recherches"), Paris.

สวัสดี วิเศษวงษ์ พ.ศ. ๒๕๒๔ ตำราเรียนอักษรโบราณ ขอม พม่า มอญ ลานนา อักษรธรรม ลาว สำนักพิมพ์พงศ์สว่าง ก.ท.ม. [Sawat Wisetwong, 1981, *Treatise on learning Ancient Characters: Khom, Burmese, Mon, Tham characters, and Lao*, Pong Sawang Publishing House, Bangkok.]

Staal, Frits, 1988, *Universals: studies in Indian logic and linguistics*, University of Chicago Press, Chicago.

Tonkin, Derek, 1962, *Modern Cambodian Writing*, Phnom Penh.

Thongluang Boonprom, [1994], *Thai-Northern Khmer-Cambodian-English Dictionary*, Rewat Karn Phim, Buriram (Thailand).

ทองเหลือง บุญพร้อม [พ.ศ. ๒๕๓๗] พจนานุกรมภาษาไทย-เขมรถิ่นไทย-เขมรกัมพูชา-อังกฤษ พิมพ์ที่ เรวัตการพิมพ์ อ. เมือง จ. บุรีรัมย์

ចង្ហែង ប៊ុនប្រម [១៩៩៤] វចនានុក្រមភាសាថៃ-ខ្មែរសុទ្ធ-ខ្មែរកម្ពុជា-អង់គ្លេស ពេទ្យភូមិវេត្យកាសិ ប៊ុនរ៉ាំ ។