NOTE ON THE TRANSLITERATION OF KHMER

Michel Antelme
Maitre de conferences, 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 – attion)” 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 សម្រាប់ ក្រុង, is “Phnom Penh” in transcription, using letters of the Roman alphabet and imperfectly rendering the pronunciation of the name; this name is “bhnan bern” in transliteration (the reason for this spelling is the very subject discussed in this paper); and /phnum pen/ 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. 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 prākrit-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 Feledus (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 mül ញរ (meaning “round”) and jríen ញត (meaning “bent”). These two scripts vary only slightly on the shape of some letters, the mül is used for titles or to note Pāli, whereas jríen, 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:

\[
\begin{array}{cccccccccccc}
\text{a} & \text{ā} & \text{i} & \text{i} & \text{u} & \text{ū} & \text{e} & \text{ai} & \text{o} & \text{au} & \text{r} & \text{ṛ} & \text{l} & \text{l} \\
\end{array}
\]

Then there are the consonants:

\[
\begin{array}{cccccccccccc}
\text{ka} & \text{kha} & \text{ga} & \text{gha} & \text{ṇa} & \text{ṛ} & \text{ṝ} & \text{l} & \text{l} \\
\text{ca} & \text{cha} & \text{ja} & \text{jha} & \text{nā} & \text{ṛ} & \text{ṝ} & \text{l} & \text{l} \\
\text{ṭa} & \text{ṭha} & \text{ḍa} & \text{ḍha} & \text{nā} & \text{ṛ} & \text{ṝ} & \text{l} & \text{l} \\
\text{ṭa} & \text{ṭha} & \text{ḍa} & \text{ḍha} & \text{nā} & \text{ṛ} & \text{ṝ} & \text{l} & \text{l} \\
\text{ṭa} & \text{ṭha} & \text{ḍa} & \text{ḍha} & \text{nā} & \text{ṛ} & \text{ṝ} & \text{l} & \text{l} \\
\text{ṭa} & \text{ṭha} & \text{ḍa} & \text{ḍha} & \text{nā} & \text{ṛ} & \text{ṝ} & \text{l} & \text{l} \\
\end{array}
\]

In Sanskrit, these signs were to be read and pronounced with a short “a” which is often pronounced as a schwa /a/ (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 ០� Kamvujāksara, the second consonant was written under the first one, sometimes undergoing some changes in shape.

---

3 The Vamānākrama Khmer ឃ្លេច្រោះព្រឹត្តិការ៍ or Dictionnaire cambodgien (Khmer Dictionary) distinguishes a third script very close to the mül one, called khm ឈុ (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 ៣ but as the pronunciation of this sign has acquired a new phonetic value in Khmer, some scholars later proposed another form ១, which is used for Indo-Aryan words and is distinguished from the former one used for Khmer words.

6 The consonants ០ and ១ are no longer used in modern Khmer writing. The consonant ២ 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.
These signs in modern Khmer are written as follows:

\[
\begin{array}{cccccc}
ka & kha & ga & gha & \acute{n}a & \\
ca & cha & ja & jha & \acute{n}a & \\
ta & tha & da & dha & \grave{n}a & \\
t\acute{a} & th\acute{a} & da & dha & ma & \\
pa &pha & ba & bha & ma & \\
y\acute{a} & ra & la & va & \\
\acute{s}a & \acute{sa} & \acute{sa} & \acute{ha} & \\
\end{array}
\]

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:\(^8\)

\[
\begin{array}{ccccccc}
\ddot{a} & i & \ddot{i} & u & \ddot{u} & e & \ddot{ai} & o & \ddot{au} & \ddot{r} \\
\end{array}
\]

Scholars believe that there were many more vowels in Khmer than in Sanskrit since pre-Angkorian times. Certain consonant subscripts, such as \(\ddot{a}\) and \(\ddot{a} \ddot{i}\), 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 (\(\ddot{a}, \ddot{ha}, \ddot{a}, \ddot{ha}, \ddot{n}\)), 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 correspondence in vowels and consonants used for the notation of Pāli in the Khmer mūl script and the Lao religious script (called Tham from Pāli \(\dddot{a} \dddot{g} \ ddh\ddot{a} \dddot{m} \dddot{m}\) and \(\dddot{a} \dddot{g} \ ddh\dddot{a} \dddot{m} \dddot{m}\)). 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 \(/\dddot{f}\/\). \(^9\)

\(^8\) The shape of subscript \(\ddot{n}\) was \(\dddot{a}\) in stone inscriptions and has been retained now as such only when the first consonant in consonant clusters is also a the \(\dddot{a}\). \(^9\) Nowadays, the subscript shape of \(\ddot{a}\) has been forsaken and cannot be distinguished from the one of \(\ddot{a}\), but in Inscriptions Modernes d'Angkor, in manuscripts from the beginning of the twentieth century and in the first issues of \(\dddot{a} \dddot{g} \ ddh\dddot{a} \dddot{m} \dddot{m} \ Kambujasiri\ddot{w}, its shape resembles the lower part of the \(\dddot{a}\) in mūl script.

\(^10\) The shapes of the subscript for \(\ddot{a}\) and \(\ddot{a} \ddot{a}\) do not appear in the \(\dddot{a} \dddot{g} \ ddh\dddot{a} \dddot{m} \dddot{m}\) Vatangānundu Khmer but can be found in Maspéro (1915).

\(^11\) As it seems that the vowels after \(\ddot{a}\) 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 millennia ago, when speakers of Sanskrit, usually Brāhmīns 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. 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 → corresponds to /aa/, i.e. a long vowel; ī → /i/, short vowel; while ē → represents long /ii/, etc.

As for consonants, ka ṅ represents an unvoiced unaspirated guttural stop /k/; kha ṇ is an unvoiced aspirated guttural stop /kh/; ga ṅ is a voiced unaspirated guttural stop /g/; gha ṇ is a voiced aspirated guttural stop /gh/; na ṇ is a guttural nasal /n/, 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 ṇī and ṇ, which is composed of the isolated vowel “a” ṇ and the vowel sign for i, 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.

\[\text{\textsuperscript{12} For a discussion of Sanskrit, its pronunciation and Pāṇini’s work, see Allen (1953) and Staël (1988).}\]
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 gaṅgā “water; the Ganges” written ế in Khmer script is pronounced /kʰoŋ-kia/ in standard Khmer, never /gaŋ-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 ế was used to represent an unvoiced unaspirated guttural stop /k/, kha ế an unvoiced aspirated guttural stop /kh/,13 ga a voiced unaspirated guttural stop /g/, gha a voiced aspirated guttural stop /gh/, ṇa a guttural nasal /ŋ/, 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 ṭ which is read as /w/ in some modern Indian languages, may have been /ʋ/ 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 shortten 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:
- ʊ, found only in one very common word: gui ế meaning “that is to say, i.e.” and which is ế in modern Khmer with the phonemic value: /ki/.14
- ya -j, yā -j, ye t in pre-Angkorian for /io/; ya -j, yya w for /io/, and ya -j, yya w, yā -j for /iio/ 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: lāuti ế, a personal name of unknown meaning.
15 The table above 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.
<table>
<thead>
<tr>
<th>Transliteration</th>
<th>Khmer script</th>
<th>Phonemic value in pre-Angkorian</th>
<th>Phonemic value in Angkorian</th>
</tr>
</thead>
<tbody>
<tr>
<td>a (inherent vowel)</td>
<td>-</td>
<td>/a/ or /ɔ/</td>
<td>/a/ or /aa/ or /ɔ/ or /ɔɔ/</td>
</tr>
<tr>
<td>a (inherent vowel) + double consonant in writing as word-ending</td>
<td>-</td>
<td>/a/</td>
<td>/a/ or /aa/</td>
</tr>
<tr>
<td>ā</td>
<td>-</td>
<td>/a/ or /aa/</td>
<td>/a/ or /aa/</td>
</tr>
<tr>
<td>i</td>
<td>-</td>
<td>/i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/</td>
<td>/i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/</td>
</tr>
<tr>
<td>iy</td>
<td>-</td>
<td>/i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/</td>
<td>/i/ or /ii/</td>
</tr>
<tr>
<td>iyy</td>
<td>-</td>
<td>/i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/</td>
<td>/i/ or /ii/</td>
</tr>
<tr>
<td>ɨ</td>
<td>-</td>
<td>/i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/</td>
<td>/i/ or /ii/</td>
</tr>
<tr>
<td>i̯</td>
<td>-</td>
<td>/i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/ or /i/ or /ii/</td>
<td>/i/ or /ii/</td>
</tr>
<tr>
<td>u</td>
<td>-</td>
<td>/u/ or /uu/ or /oo/ or /uo/</td>
<td>/u/ or /uu/</td>
</tr>
<tr>
<td>u̯</td>
<td>-</td>
<td>/u/ or /uu/</td>
<td>/u/ or /uu/</td>
</tr>
<tr>
<td>ɨ̯</td>
<td>-</td>
<td>/u/ or /uu/ or /oo/ or /uo/</td>
<td>/u/ or /uu/</td>
</tr>
<tr>
<td>e</td>
<td>-</td>
<td>/ɛ/ or /ɛɛ/ or /ɛɛ/ or /ɛɛ/ or /ɛɛ/</td>
<td>/ɛ/ or /ɛɛ/ or /ɛɛ/ or /ɛɛ/</td>
</tr>
<tr>
<td>ai⁴⁶</td>
<td>-</td>
<td>/ai/ or /aai/</td>
<td>/ai/ or /aai/</td>
</tr>
<tr>
<td>o</td>
<td>-</td>
<td>/ɔ/ or /ɔɔ/ or /ɔɔ/ or /ɔɔ/ or /ɔɔ/</td>
<td>/ɔ/ or /ɔɔ/ or /ɔɔ/ or /ɔɔ/</td>
</tr>
<tr>
<td>[au¹⁷]</td>
<td>-</td>
<td>/aw/ or /aaw/</td>
<td>/aw/ or /aaw/</td>
</tr>
<tr>
<td>ya</td>
<td>-</td>
<td>/ia/</td>
<td>/ia/</td>
</tr>
<tr>
<td>yā</td>
<td>-</td>
<td>/ia/</td>
<td>/ia/</td>
</tr>
<tr>
<td>ye</td>
<td>-</td>
<td>/ia/</td>
<td>/ia/</td>
</tr>
<tr>
<td>yyy⁴⁸</td>
<td>-</td>
<td>/ia/</td>
<td>/ia/</td>
</tr>
<tr>
<td>yyy⁴⁹</td>
<td>-</td>
<td>/ia/</td>
<td>/ia/</td>
</tr>
<tr>
<td>va</td>
<td>-</td>
<td>/u/ or /ua/ or /uu/ or /uo/</td>
<td>/u/ or /ua/</td>
</tr>
<tr>
<td>vā</td>
<td>-</td>
<td>/u/ or /ua/ or /uu/ or /uo/</td>
<td>/u/ or /ua/</td>
</tr>
<tr>
<td>vo</td>
<td>-</td>
<td>/u/ or /ua/</td>
<td>/u/ or /ua/</td>
</tr>
</tbody>
</table>

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 ɨ̯ /-ɨ̯/ < -aai/ of its supposedly long counterpart, was not really made. For instance, the female title ai ɨ̯ in is probably to be found in Modern Khmer miːt [mii] / miːt / maːdaːj/ “mother,” and old Khmer ai ɨ̯ has probably given both ai ɨ̯ in ai ɨ̯ /iː/ daːj / daːj / daːj “in, at (obsolete)” and ɨ̯ ɨ̯ /-aai/ “over here.” It is also highly probable that this symbol represented the combination vowel + final consonant /u/ in speech, and it was sometimes followed by the graphic symbol for consonant ɨ̯. 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 ɨ̯ /-ɨ̯/ < -aaw/ of its supposedly long counterpart, was not really made. The interpretation is the more difficult when for instance the word for “out, outside” is krau [krɔ] / krau/ in Central Khmer, but *krau [krɔ] /krau/ in the Surin dialect, the word for “shirt” is ɨ̯ ɨ̯ /-aaw/ in Central Khmer but *eau *eaw / eaw/ in the Surin dialect.

18 This graphic combination is possible only when the first letter of the syllable is the rapha 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. In Khmer it occurred during the middle period, i.e. the Post-Angkorian period.

The voiced consonants /g/, /j/, /d/, /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:

\[ ga \rightarrow kha, gh \rightarrow kh, ja \rightarrow ca, jh \rightarrow ch, da \rightarrow ta, dh \rightarrow tha, ba \rightarrow pa, bh \rightarrow ph \]

In modern Khmer there are still two voiced consonants: /b/ and /d/, represented by the consonantal signs /p/ and /t/. 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.' These two voiced consonants are never found in first position in an initial cluster. So when pa ឬ is written as the first consonant in a consonant cluster, it is pronounced /p/.

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 (samlei srāl សំឡើងស្រាប់ or samlei tīc សំឡើងតូច in Khmer) whereas those following the former voiced consonants were pronounced with a "breathy" voice (samlei dhnīn សំឡើងដូង or samlei dhan សំឡើងត្រចូល). 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 ការ /kaa/ remained /kaa/, pronounced with a clear voice, whereas gā ការ /gaa/ became /kā/ with a breathy voice. Later, gā ការ /kāa/ became /kā/ retaining the breathy voice. This register was then lost in most dialects to become /kə/. 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/c changed to /ee/, then in Standard Khmer the /ee/ became /ac/ when associated with a former unvoiced consonant and /e/c 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. Each syllable is given first in transliteration,

---

---
then in Khmer script, followed by phonetic transcription.

<table>
<thead>
<tr>
<th>ka</th>
<th>/kəa/</th>
<th>ga</th>
<th>/kəa/</th>
</tr>
</thead>
<tbody>
<tr>
<td>kā</td>
<td>/k̂a/</td>
<td>gā</td>
<td>/k̂a/</td>
</tr>
<tr>
<td>ki</td>
<td>/k̂i/</td>
<td>gī</td>
<td>/k̂i/</td>
</tr>
<tr>
<td>kī</td>
<td>/k̂i/</td>
<td>gī</td>
<td>/k̂i/</td>
</tr>
<tr>
<td>kō</td>
<td>/k̂o/</td>
<td>gu</td>
<td>/k̂u/</td>
</tr>
<tr>
<td>kū</td>
<td>/k̂u/</td>
<td>gi</td>
<td>/k̂u/</td>
</tr>
<tr>
<td>kuo</td>
<td>/k̂u/</td>
<td>guo</td>
<td>/k̂u/</td>
</tr>
<tr>
<td>koe</td>
<td>/k̂o/</td>
<td>goe</td>
<td>/k̂o/</td>
</tr>
<tr>
<td>kīoe</td>
<td>/k̂o/</td>
<td>giōe</td>
<td>/k̂o/</td>
</tr>
<tr>
<td>kē</td>
<td>/k̂i/</td>
<td>gē</td>
<td>/k̂i/</td>
</tr>
<tr>
<td>kac</td>
<td>/k̂a/</td>
<td>gac</td>
<td>/k̂a/</td>
</tr>
<tr>
<td>kāi</td>
<td>/k̂a/</td>
<td>gai</td>
<td>/k̂i/</td>
</tr>
<tr>
<td>ko</td>
<td>/k̂a/</td>
<td>go</td>
<td>/k̂o/</td>
</tr>
<tr>
<td>kau</td>
<td>/k̂a/</td>
<td>gau</td>
<td>/k̂a/</td>
</tr>
<tr>
<td>kum</td>
<td>/k̂a/</td>
<td>gum</td>
<td>/k̂a/</td>
</tr>
<tr>
<td>kām</td>
<td>/k̂a/</td>
<td>gām</td>
<td>/k̂a/</td>
</tr>
<tr>
<td>kah</td>
<td>/k̂a/</td>
<td>gah</td>
<td>/k̂a/</td>
</tr>
</tbody>
</table>

To illustrate this, below is an excerpt from a cūpū' 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 gīə tri /kia/ and kīə tri /k̂o/.

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 gūm /kum/.
Transliteration:

<table>
<thead>
<tr>
<th>Khmer</th>
<th>Roman</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>បែក នៃ ឆ្វេង</td>
<td>dūmnā m oy gōp</td>
<td>kūn cau e kroy</td>
</tr>
<tr>
<td>អំប៉ុល ឈិត ដរយ</td>
<td>vi gōp duk oy</td>
<td>tae bāky នៃ តឹយ</td>
</tr>
<tr>
<td>លក្ខ ដាល ធ្វើលូង</td>
<td>maen bit mo xī</td>
<td>būm smoe suriyā</td>
</tr>
</tbody>
</table>

Transcription in hypothesised middle Khmer:

<table>
<thead>
<tr>
<th>Khmer</th>
<th>Roman</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>baak neh jaa cpap</td>
<td>duunmaan jōj gōp</td>
<td>kuun cau z9 tēe krooj</td>
</tr>
<tr>
<td>ណាញីកត្រានៃ ដរា</td>
<td>vi gōp duk jōj</td>
<td>tēe baak neh dooj</td>
</tr>
<tr>
<td>brah pālī thā.</td>
<td>brah pālī thā.</td>
<td></td>
</tr>
<tr>
<td>look thaa blesq blī</td>
<td>meen bit mo xī rī</td>
<td>būm smoe surijaa</td>
</tr>
</tbody>
</table>

Transcription in modern Standard Khmer:

<table>
<thead>
<tr>
<th>Khmer</th>
<th>Roman</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>pīo2 nih cīo c'hap</td>
<td>tuunmaan jōj kōa p</td>
<td>koon cau tēe kroaj</td>
</tr>
<tr>
<td>ṛaa būk ṛat kroaj</td>
<td>vi kōa tuk jōj</td>
<td>tae pīo2 nih daoj</td>
</tr>
<tr>
<td>pri kīa bāluaj thāa.</td>
<td>pri kīa bāluaj thāa.</td>
<td></td>
</tr>
<tr>
<td>loe2 thaa plām plī</td>
<td>meen pit3 tō me khā rī</td>
<td>pūm smoa sôrējaa</td>
</tr>
</tbody>
</table>

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 Prākrit-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. 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 /cau/.
30 Jenner gives the pronunciation /droap/, but then it does not rhyme with /gap/.
31 In some Northern Khmer dialects, the pronunciation is clearly /pi/. In the Central dialects, the /i/ systematically turns into /i/ when followed by a final consonant with the exception of /p/, /c/, /g/, /k/.
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.scribd.com/sa-research/sa0819.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 interpenetration 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 râm ṭhâ /rîam/ “merchandise stand” and rînu ʔj̬ːm /rîan/ “to learn,” whereas in the Surin dialect of Thailand, the first word is pronounced /riam/ and the second one /rûm/. In turn, some Surin dialects do not make the difference between the words bblieâ t̪j̬ː “rain; to rain” and bblieâ t̪j̬ː /pʰlûŋ/ “music; song,” whereas the Phnom Penh dialect does make it as the latter is pronounced /pʰłeŋ/ as opposed to bblieâ t̪j̬ː /pʰlûŋ/.

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” mûnʼ krap bâi រែូស្ពែក ៈ/, which from a lexical viewpoint has no dialectal variations, will be pronounced:

/mûn krap pœn/ in the so called “standard accent,”
/mûn krap pœn/ in Surin or in some parts of Siem Reap and Battambang provinces,
/mûn kâap pœn/ in Phnom Penh,
/mûn kâap pœn/ 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 ឃុំ /kit/ in Khmer and อำเภอ /khít/ in Siamese. In transliteration, the word in both languages is giit. Due to the fact that both languages use Indic-based scripts transliteration shows that the word was originally roughly pronounced /giit/ when one borrowed it from the other.

The word for “craftsmen” is នាយ /cin&/ in Khmer and ชาว /ch&an/ in Siamese. In transliteration, the word is j&i in Khmer and j&i in Siamese. This word, maybe of Chinese origin, was probably pronounced /jaan/ when it was borrowed by both languages (regardless of the tone in Siamese).

4. Because the Khmers borrowed their script before the invoicing 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 Mnon Ro’lo’m (spoken in northeastern Cambodia and south central Vietnam) the word jil “to wipe” (Blood 1976: 11) which can be compared to Khmer /cuut/ (same meaning) which is written ឃុំ and transliterated jïl. In Jeh, a North Bahnaric language of the Mon-Khmer family spoken in Kontum province in Vietnam, one can find the word g& (pronounced with a long vowel) “to hold (with plst stick)” (Gradin 1976: 27) which can be compared to Khmer /k&/ “to press, compress, hold by means of pliers or tweezers,” which is written ឃុំ and transliterated g&. In Mo’dra and Didra, two To’drah dialects, a North Bahnaric language also spoken in Kontum province, the word for “three” is pit (long vowel) (Gregerson and Smith 1973: 175) to be compared with modern Khmer /bej/ (same meaning), written ຕ and transliterated pi.

---

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


---

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.


——— 1970b, វិជ្ជាចារ្យាល វិជ្ជាចារ្យាល (Vījā khemarabhāsa), Cambodian System of Writing and Beginning Reader, Yale University Press, London/New Haven.


——— 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.


Long Scam ស្រែម ប្រែក, [2000], Dictionnaire du khmer ancien (D’après les inscriptions du Cambodge du vie-vie siècles), Académie royale du Cambodge, Institut de langue nationale, Université royale des Beaux-Arts, Faculté d’archéologie, Phnom Penh.


National Library, 2000, Book on Different Forms of Ancient Characters, Section of Ancient Languages, National
Library, Department of Fine Arts, Bangkok.


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


[There are no visible references to pages or paragraphs in the visible text.]

[Additional text not transcribed due to its illegibility or partial visibility.]

[Additional text not transcribed due to its illegibility or partial visibility.]