

REFLECTIONS ON
THE HISTORY OF THE KHMER PHONEMIC SYSTEM*

Heinz-Jürgen PINNOW
Freie Universität, Berlin

The phonetic and phonemic systems of Khmer¹ are many-sided and complex to an uncommon degree; this is especially true of the vocalism. The situation is all the more complicated by the variety of transcriptions employed by different writers, who disagree in many cases in their interpretation of the articulation of vowels and consonants and of their place in the phonemic system. Various dialectal differences and orthographic variations also occur. This being the case, it ought to be useful to examine critically the systems favored by the principal authorities and weigh them one against the other. The following are considered in the presentation hereafter:

1. L. Finot, "Notre transcription du cambodgien," in *BEFEO*, II (1902).1: 1-15.
2. P. W. Schmidt, "Grundzüge einer Lautlehre der Mon-Khmer-Sprachen," in *Denkschr. d. kais. Akad. d. Wiss. Wien, Philos. hist. Kl.*, 51 (1905), Abhandlung III.
3. G. Maspero, *Grammaire de la langue khmère (cambodgien)* (Paris: Imprimerie Nationale, 1915).
4. J. Guesdon, *Dictionnaire cambodgien-français*, 2 volumes (Paris: Plon, 1930).
5. F. Martini, "Aperçu phonologique du cambodgien," in *BSLP*, XLII (1942-1945).1: 112-31.
6. G. Cambefort, *Introduction au cambodgien* (Paris: G.P.

*Translation of "Sprachgeschichtliche Erwägungen zum Phonemsystem des Khmer," in *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft*, 10 (1957).4: 378-91.

¹More accurately, Khmer, pronounced [khmaer] or [khmae] with a distinct [ae] diphthong, as in German Meyer.

Maisonneuve, 1950).

7. Eugénie J.A. Henderson, "The Main Features of Cambodian Pronunciation," in *BSOAS*, XIV (1952).1: 149-74.

A careful distinction must be made between *transliterations* on the one hand and, on the other, phonetic and phonemic *transcriptions*. Schmidt gives only a transliteration. Martini uses all three. Henderson uses the Cambodian character together with a phonological transcription, occasionally supplemented by a phonetic transcription. Maspero, Guesdon, Cambefort and Finot all give the Cambodian character along with a romanization which is supposed to be at once a transliteration and a phonetic and phonemic transcription, but which is at variance with the facts. The whole dispute between Schmidt and Finot over romanization is meaningless, since Schmidt uses only transliteration. When Finot asserts that "...la translittération *bhlön, *bhnam ne répond à aucune réalité ni moderne, ni ancienne,"² he is indeed right as far as pronunciation goes but not as far as written forms are concerned, and it was only the latter that were represented in Schmidt's system.

The Vocalism

In the writing system the following vowels are distinguished: ɔ, ɔ̃, a, ă, ĩ, i, ũ, u, (ř, ɾ, ĺ, ɿ), uo, ʋ, uʋ, ie, e, ε, ai, o, au, ũ̃, ɔ̃, ă, (ăŋ), ɔ̃h (ɔ̃ḥ, &c.).

In the older though never universally accepted transliteration these were represented a, a', ā, ā', i, ī, ʋ, ũ, u, ū, (ɾ, ř, ɿ, ĺ), uo, o', ʋo', ie, e, ê, ai, o, au, uṃ, aṃ, āṃ, (āṃṃ), aḥ, (a'ḥ, &c.). Various other systems have also been used.

The articulation of the vowels is now conditioned by the

²Finot, op.cit., 4.

nature of the consonants preceding them. These fall into two groups known as *series* or *registers*. To the First Register belong the original voiceless consonants k, kh, c, ch, t, th, t, th, p, ph, s and h, together with ŋ and ʎ and initial vowels. To the Second Register belong the originally voiced consonants g, gh, ʒ, ʒh, d, dh, d, dh, b and bh along with the nasals ŋ, ɲ, n and m as well as j [j], r, l and v. Of this Second Register the voiced stops, g and so on, have since lost their voicedness and hence become k and so on, while t and p are now realized as [d] and [b] respectively; see the discussion of the consonants below. To compensate for this shift, words of the Second Register are pronounced in a deeper voice.

The characteristics of the second register are a deep rather breathy or "sepulchral" voice, pronounced with lowering of the larynx, and frequently accompanied by a certain dilation of the nostrils. Pitch is usually lower than that of the first register in similar contexts.³

These relationships as represented by Schmidt, Martini and Henderson are shown in Table 1. The systems used by the others are generally similar but exhibit differences of detail; these are shown in Table 2.

The letters ř, ʀ, ʎ and ʌ are equivalent to combinations of r and l plus ũ and u. Maspero writes them as ʀũ, ʀu, ʎũ, ʎu, Guesdon as rŭ, rŭ, lŭ, lŭ.

With Henderson y = [j], and it is to be noted that : appears in her phonological transcription when a long vowel is followed by a final consonant. Her F stands for "final", that is, a following final consonant.

The realization of my /ĩ/, /ũ/, /ě/ and /õ/ is diphthongal: [iə], [uə], [ɛə], [ɔə]. /õ/ is realized variously as [o],

³Henderson, *op.cit.*, 151.

Tl.	phonol. R ¹ R ²		Transcriptions of						
			Schmidt	Martini				Henderson	
			Tl.	Tl.	phonet. R ¹ R ²	phonol. R ¹ R ²	phonet. R ¹ R ²	phonol. R ¹ R ²	
ɔ	a	ɔ	a	a	a: ɔ:	a ɔ	ɔ: ɔ:	ɔ ɔ	
ɔ̃	ã	ũ ɔ̃ ɛ̃	a	a'	a w _o o ə _a	a u ^o u ^o ə ^o	ɔ ɔ̃ uə	ɔF wəF	
a	a	iɣ	ā	ā	a: i: e	a ie	a: iə	a iə	
ã	ã	ũ ɔ̃ ɛ̃	ā	ā'	a ə _a ε _a	ã ă æ̃	a w ^o ə̃ é̃	aF ə̃əF é̃əF	
ĩ	ỹ	ũ ĩ	i	i	ə _e u ə _ε i	ě i ĩ	ɣ ɰ	ɣF ɰF	
i	ɣj	i	ī	ī	ε _i i:	ey i	ɣy i:	ɣy i	
ũ	(ỹ)	ũ	ü	i̇	- u	- i̇	(ɣ?) ɰ	(ɣ?) ɰF	
u	(ɣ)	u	ū	i̇	- u:	- i̇	(ɣ) ɰ:	(ɣ) ɰ	
ũ	õ	ũ	u	u	a ^u u	ɔ̃ ũ	ou ɰ oy w _o	oF ɰF	
u	o	u	ū	ū	o: u:	o u	o: ɰ:	o ɰ	
uo	uy	uy	uo	w	u: o u: o	uo uo	uə ɰə	uə ɰə	
ɣ	ay	ɣ	ö	œ	a ^ə ə:	ʌ ə	ay ɣ	ay ɣ	
uɣ	uɣ	uɣ	üö	yœ	u: ə u: ə	iə iə	uə ɰə	uə ɰə	
ie	ĩ	ĩ	ie	ye	ye ye	ĩe ĩe	iə iə	iə iə	
e	e ě	i ĩ	é	e	e: i:	e i	e: è: e i̇	e è eF i̇F	
ε	ae ě	ε	ē	ê	æ: ε:	æ ε	ae è: e	ae è eF	
ai	aj	uj	āi	ai	ai̇ ei̇	ăy ăy	ay ɰy	ay ɰy	
o	ao	u	ō	o	a ^o o ^u	ă u	ao ɔ:	ao ɔ	
au	av	uv	āu	au	a ^u ə ^u	av əv	av ɰv	av ɰv	
ũ	õm	ũm	ū	um̃	a ^u m um	ɔ̃m ũm	om ɰm	om ɰm	
õ	ãm	õm	ã	am̃	am ɔm	ãm ɔ̃m	ɔm ɔ̃ə̃m	ɔm ɔ̃ə̃m	
ã	ãm	ẽm	ã̃	ã̃m̃	am ə̃am	ãm ă ^o m	am è̃ə̃m	am è̃ə̃m	
ãŋ	ãŋ	ẽŋ	ã̃ñ	ã̃m̃ñ	aŋ εaŋ	ãŋ ẵŋ	aŋ è̃ə̃ŋ	aŋ è̃ə̃ŋ	
ɔh	ãh	ẽh	ah	ah	ah εah	ãh ẵh	ah è̃ə̃h	ah è̃ə̃h	

Table 1

Abbreviations: Tl. = transliteration; phonol. and phonet.
= phonological and phonetic transcription; R = register.

Tl.	Transcriptions of							
	Maspero		Finot		Guesdon		Cambefort	
	R ¹	R ²	R ¹	R ²	R ¹	R ²	R ¹	R ²
o	a	ò	a	o	â	o	â	o
õ	ã	õ	ã	õ	ã	õ	ã	õ
a	á	ea	á	ã	a	éa	e	ea
ã	ã	ã	ã	ã	ã	éa	ã	éa
ĩ	ẽ	ĩ	ẽ	ĩ	ẽ	ĩ	ẽ	ĩ
i	ei	i	ẽi	i	ẽi	i	ei	i
ũ	(ũ)	ũ	(ũ)	ũ	(ũ)	ũ	(ẽ)	ũ
u	(u)	u	(u)	u	(u)	u	(œ)	u
ũ	õ	ũ	õ	ũ	õ	ũ	õ	ũ
u	au	u	ó	u	âu	û	aû	u
uo	uo	uo	uo	ûo	uo	ûo	uo	ûo
y	eu	eu	ò	o	o	œu	o	o
uy	euo	euo	u ^o	ú ^o	eua	œua	u ^o	u ^o
ie	ie	ie	ie	îe	ie	îe	ie	îe
e	e	e	e	é	é	é	e	é
ε	è	è	è	ê	ê	ê	è	ê
ai	ai	ei	ai	ei	ai	ei	ãi	ẽi
o	ò	ou	ò	ô	o	ou	ò	ó
au	ao	òu	au	ou	au	ôu	ãu	õu
ũ	õᄀ	ũᄀ	õᄀ	ũᄀ	õᄀ	ũᄀ	õᄀ	ũᄀ
õ	aᄀ	òᄀ	aᄀ	oᄀ	ãᄀ	õᄀ	ãᄀ	õᄀ
ã	ãᄀ	eaᄀ	ãᄀ	ãᄀ	ãᄀ	eãᄀ	ãᄀ	éãᄀ
ãᄀ	ãᄀ	eaᄀ	ãᄀ	ãᄀ	ãᄀ	éãᄀ	ãᄀ	éãᄀ
ch	ã	ã	aᄀ	áᄀ	ãᄀ	eãᄀ	ãᄀ	éãᄀ

Table 2

Abbreviations: Tl. = transliteration; R = register.

[ou], [oy] and [ʊo]. e and ε become ě in the First Register before h, s; they become ĭ in the Second Register before h. ā becomes ě in the Second Register before k, ŋ, h; elsewhere it is ǎ.

Out of this considerable disparity between writing and pronunciation the following questions arise:

1. *What was the Old Khmer vowel system like around the time when writing was introduced?*
2. *What is the nature of the modern system?*
3. *How did the modern system develop out of the old?*

In seeking an answer to the third of these questions we may either start with the system as it exists today and trace its development backwards in time, or take the Old Khmer system as our point of departure. In either case the direction of our enquiry must be the former, though the latter is appropriate for a brief exposition.

Regarding the first question, it must be understood from the outset that, once adopted, the Indic writing system could never have been altogether adequate for the representation of the Khmer vowels.⁴ Then too it must be noted that the orthography of words has oftentimes varied not only in the course of historical development but also during given periods.

To begin with, ṛ̣, ṛ̣̣, ḷ̣ and ḷ̣̣ as well as ụ̃, ạ̌, ạ̄ and ɔ̣h can be dissociated from the rest of the writing system on the grounds that these represent combinations of consonants and vowels while nasalized vowels were and still are entirely wanting in Khmer; these symbols are purely Indo-Aryan. aɪ and au similarly represent combinations of a vowel plus j and v, and may hence be disregarded.

⁴On the Hindī vowel system and its representation in the *devanāgarī* vid. my paper "Über die Vokale im Hindī," in *ZfPh*, 7 (1953): 43 sq.

Of the remaining symbols, those corresponding to a, ā, i, ī, u, ū, e and o in the *devanāgarī* and other Indic writing systems represent Khmer ɔ, a, ĩ, i, ũ, u, e and o. The symbols ũ, w, v and ε are innovations, v being compounded of i and e, ε being a form of double e.⁵ By means of the diacritic ' (the *soṅkāt*) two additional vowels are distinguished: short ǔ and short ǎ. Finally, other symbols were invented for three special diphthongs: uo, ie and wv. Some of these symbols were devised in fairly recent times. This applies particularly to ũ, w and wv, believed to have been introduced by the Brōḥ Sūgōn (Pră sōkōn), who died in 1894.⁶ In the old inscriptions e was still being used for v.⁷ The relationship of e and ε is involved. While e is unquestionably the original symbol for e, i is frequently found in its place in the inscriptions.⁸ For this reason Schmidt⁹ decided in favor of ε as the proper symbol for Khmer e and saw e as having developed from i, which becomes e (Schmidt writes *é*) when it occurs with a voiced initial.¹⁰ The diphthongs ie and uo, as Schmidt has shown,¹¹ come from earlier ya [ja] and va [wa] and are hence not to be considered old.¹² The same holds true for wv.

The recent introduction of w and the other new symbols into the writing system does not of course constitute proof that the sounds they represent did not exist previously. Indeed, the latter were expressed formerly, though not accurately. From the

⁵Cf. the corresponding symbol in Thai.

⁶Maspero, *op.cit.*, 63.

⁷*Ibid.*, 106.

⁸*Ibid.*, 95.

⁹Schmidt, *op.cit.*, §§165 sq, pp. 159 sq.

¹⁰*Ibid.*, 160.

¹¹*Ibid.*, §§199 sq, pp.180 sq; §§225, pp.198 sq.

¹²*Vid.* also Maspero, *op.cit.*, 105.

absence of a perfect parallel between Mon, Khmer, Bahnar and Stieng γ , Schmidt infers that γ is comparatively recent in these languages.¹³ The same holds true for w , which Mon does not have. Schmidt advances no opinion on w . It is nevertheless likely that both sounds existed in Khmer long before they were represented in writing, possibly even before the adoption of the writing system.

As to \circ , \check{o} and a , \check{a} (represented by a , a' and \bar{a} , \bar{a}' in Martini's transliteration), the separation of these two pairs forces itself upon us, as it were. The first vowel of the *devanāgarī*, namely a , represents a short sound in Sanskrit and kindred languages; in Khmer however it is a long vowel which can be shortened by means of the *soṅkāt* ($'$). The second vowel of the writing system, \bar{a} , can also be shortened by $'$. It follows that a^{14} cannot stand in the same relationship to \bar{a} as it does in Sanskrit; it must have a different articulation. The probable articulation of Old Khmer a^{14} is not hard to define: it must have been an $[\circ]$ type vowel. Neither Sanskrit nor the Middle Indic dialects had an \circ , and in the New Indic period a was first pronounced in this manner in Bengali. Hence there were only two possibilities for the representation of Old Khmer $*\circ$: either by means of the symbol o , or by means of the symbol a . One reflection of this is the frequent vacillation in the orthography between a and o , as in $pha\eta^{14} \sim pho\eta$ 'together', $ap \sim op$ 'to embrace', and $\tau a\eta \sim \tau o\eta$ 'handle'.¹⁵ Vacillation is also noted between a in one language and o in another; for example, Khmer an 'to weaken' : Mon on , Khmer $\tau a\eta$ 'handle' : Stieng $to\eta$.¹⁶ The same sort of vacillation is seen within Mon,

¹³Schmidt, op.cit., §§253 sq, pp.218 sq.

¹⁴Still using Martini's transliteration.

¹⁵Maspero, op.cit., 81 sq.

¹⁶Id.

of which the name itself is written *mon* or *man*.¹⁷ On the basis of such comparisons Schmidt¹⁸ concluded that these languages originally had no independent *o* at all. In reality, these comparisons seem to show that *a* as well as *o* and *ɔ* were in existence but that words with *ɔ* had to be written with either *a* or *o*. This situation persisted in Mon, but in Khmer the disadvantages it entailed were overcome by the *soṅkāt*, thanks to which *a* could be reserved for *ɔ* while *ā* could be reserved for *a*. Short syllables then in existence could be marked with *´*. That the latter dispensed with sundry anomalies, aberrations and mistakes is clear. The existence of an old *ɔ* sound side by side with an *o* sound is supported by the coexistence of *ɛ* and *e*. In this connection we may well wonder why a new symbol was not also found for *ɔ* when one had been found for *ɛ*. That one was not found for *ɔ* may have been owing to the form (*e + ā*) of the symbol *o*.

It will be clear from the foregoing that transliterating the sounds in question by *a*, *a´* and *ā*, *ā´* is altogether inappropriate and confusing. It is in fact downright absurd, since *ā´* = *ǎ* or *ǎ:* (!). To be preferred is the transliteration of these sounds by *ɔ*, *ɔ̃* and *a*, *ǎ*, which fulfill the conditions stated above and have been used in the foregoing exposition. The forms cited above should hence not be transliterated *phaṅ*, *ap*, and the like, but *phoṅ*, *ɔp* and so on. Since the "normal" vowel in Khmer is long, there is an advantage in leaving the long vowels unmarked (as by *:*) and in marking the short vowels with *˘*.

So the old vowel system as it existed prior to the devoicing of the originally voiced stops was substantially as follows. First, the normal (long) vowels:

¹⁷Cf. Maspero, *op.cit.*, 96.

¹⁸Schmidt, *op.cit.*, §230, p.195.

i	[i:],	written i
e	[e:],	written e
ε	[ε:],	written ε
u	[u:],	written i (u)
ɤ	[ɤ:],	written e (ɤ)
a	[a:],	written a
u	[u:],	written u
o	[o:],	written o
ɔ	[ɔ:],	written ɔ

In addition to the above, the diphthongs ie, wy, uo developed from jɔ (ja) and wɔ (wa), written yɔ (ya), vɔ (va). The orthographic diphthongs ai and au are combinations of vowel plus j or w.

Once the devoicing of the originally voiced stops had set in, this change affected the vowels. These split up into the two so-called registers, differentiated by pitch and openness. The First Register, here marked ´, had greater openness and higher pitch, while the Second Register, marked ` , had less openness and lower pitch. Thus the total number of vowels was doubled, resulting in a system which, with the diphthongs included, was structured as follows:

ì	ù	ù
í ¹⁹	ú	ú
ìe	ùɤ	ùo
íe	úɤ	úo
è	ý	ò
é	ý	ó
è	à	ò
é	á	ó

¹⁹That i after voiceless consonants (First Register) is rare

In the course of time these sounds underwent further changes of various kinds, culminating in the following inventory in which diphthongization plays an important part. The changes are best shown in tabular form:

<i>old</i> → <i>modern</i>		<i>old</i> → <i>modern</i>		<i>old</i> → <i>modern</i>	
	M ²⁰ H	M H		M H	
ì → i	ì	ù → i̇	ù	ù → u	ù
í → εj	ɤj	(ú)		ú	
ìe → ìe	ìə	ùɤ → i̇ə	ùɤ	ùo → uo	ùə
íe → ìe	iə	úɤ → i̇ə	ɤɤ	úo → uo	uə
è → ɪ	è	ỳ → ə	ỳ	ò → u	ò
é → e	e	ý	(ɤ)	ó	o
è → ε	è	à		ò → ɔ	ò
é → æ	ae	á	Λ	ó → ǎ	ao
	a			ó → a	ɔ

According to Martini the old à became ie, which hence differs from the ìe from old ìe.

The reader will recognize the clear parallel between the central and the back vowels. í develops into a diphthong or, more accurately, into a vowel plus j, thus deviating from the

does not indicate that Mon-Khmer originally had no i in this environment, as Schmidt supposed (op.cit., §178, p.168). What it does suggest is that even before the adoption of writing Khmer and Mon i after voiceless consonants had become diphthongized to ai [aj] (cf. Schmidt, op.cit., §242, p.211). Cf. Khmer ʈai, Mon tai (toa) 'hand' with Bahnar ti, Khaɽia ti?, Santali ti; Khmer cai, Mon cai 'louse' with Bahnar fi, Santali se, Mundari si-ku.

²⁰M = Martini; H = Henderson.

pattern. For all practical purposes $\grave{\text{i}}\text{e}$ and $\acute{\text{i}}\text{e}$ and the other pairs are identical, the sole difference between them being their register. Like $\grave{\text{a}}$, old $\acute{\text{e}}$, $\acute{\text{y}}$, $\acute{\text{o}}$ are diphthongized. For the ə in Henderson's iə and uə it is preferable to write v , since no new ə phoneme is involved. In the same way, it is better to use w and v for Martini's i and ə .

With Martini the modern vowel system (long vowels) consists of the following:

i	u	u
ie	wv	uo
ɪ		ʊ
e	v	o
ɛ		ɔ
æ	ʌ	ɑ
a		ɑ

For his ĭe (iě) see the short vowels.

But if with Henderson we take register rather than openness as the relevant characteristic, we obtain a vowel system which is quite different from Martini's, namely:

<i>monophthongs:</i>	i	u	u
	e	v	o
	ɛ	a	ɔ
<i>diphthongs:</i>	iv	wv	uv
	ae	av	ao

To the above the symbol ` must be added to mark the Second Register. As the First Register is not abnormal in any way, it need not be marked.

Thus the curious result is that the modern system matches the old one tolerably well, but that register — which, however, does not belong directly to the vowel system — now be-

comes operative in place of the earlier voicing. There was a return, in a sense, to the old system after the vowel changes consequent upon the old devoicing had been carried out.

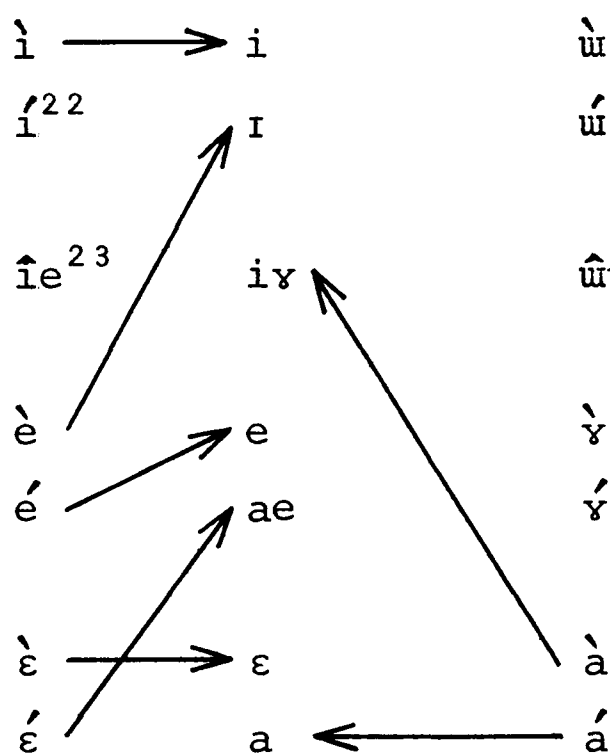
Yet the system proposed by Henderson suffers in one respect. Out of a total of nine vowels (monophthongs) register is relevant in only three cases: $e : \grave{e}$, $o : \grave{o}$, $\text{ɔ} : \grave{\text{ɔ}}$. The six other vowels always take either the Second Register (\grave{i} , \grave{u} , \grave{y} , $\grave{\text{e}}$) or the First Register (a). Among the diphthongs, ae , av , ao are invariably of the First Register. The three remaining diphthongs occur with both registers, but the phonetic distinction between them ($i\text{v} : \grave{i}\text{v}$, $uv : \grave{u}\text{v}$, $uy : \grave{u}\text{y}$) is virtually nonexistent and carries no weight in favor of including register within the phonemic system.²¹ Since the relative openness of the sounds in question — $e : \grave{e}$, $o : \grave{o}$, $\text{ɔ} : \grave{\text{ɔ}}$ — is what is perceptible, we shall give preference to Martini's interpretation and with him set up ɪ for \grave{e} , ʊ for \grave{o} , and ɑ for ɔ , while ɔ will be used for Henderson's $\grave{\text{ɔ}}$. This ɪ and ʊ do not represent somewhat centralized sounds, but degrees of openness midway between i and e on the one hand and between u and o on the other.

Henderson's three diphthongs ae , av , ao are preferable to Martini's interpretation which posits $/\text{æ}/$, $/\text{ʌ}/$, $/\text{ɑ}^\circ/$ as phonemic monophthongs realized phonetically as diphthongs, inasmuch as no useful purpose is served in treating real diphthongs as monophthongs. What is more, ae , av , ao occupy a place in the system which is analogous to that of $i\text{v}$, uv , uy .

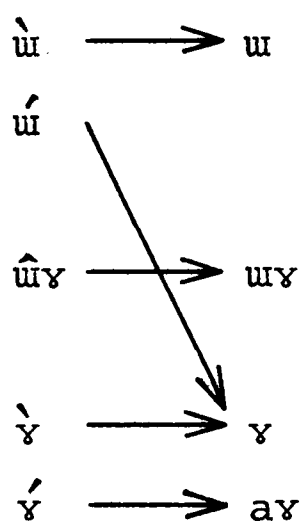
In this way the following scheme of changes results:

²¹But cf. Henderson, *op.cit.*, 154 sq.

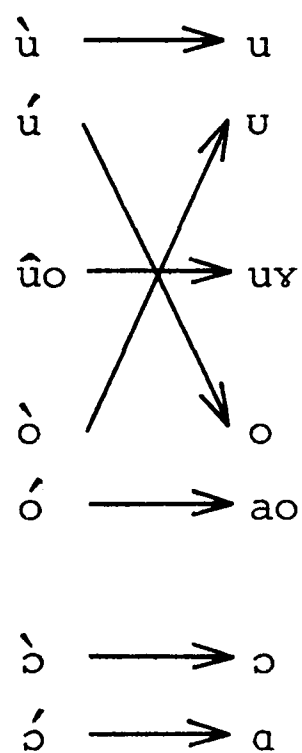
old → modern



old → modern



old → modern



For the long vowels, therefore, the modern phonological system is as follows:

<i>monophthongs:</i>	i	u	u
	I		U
	e	y	o
	ε		o
	a		a
<i>diphthongs:</i>	iy	uy	uy
	ae	ay	ao

This system is valid only for the long vowels, the only ones that can be deemed normal. Short vowels occur only before a final consonant and are hence marked with $\tilde{}$. They must be considered next.

²²*i* becomes $\tilde{y}j$.

²³The symbol $\hat{}$ represents both registers, i.e. $\acute{}$ and $\grave{}$. \hat{ie} becomes \tilde{i} [iə], which is shorter than iy. Vid. the discussion of short vowels.

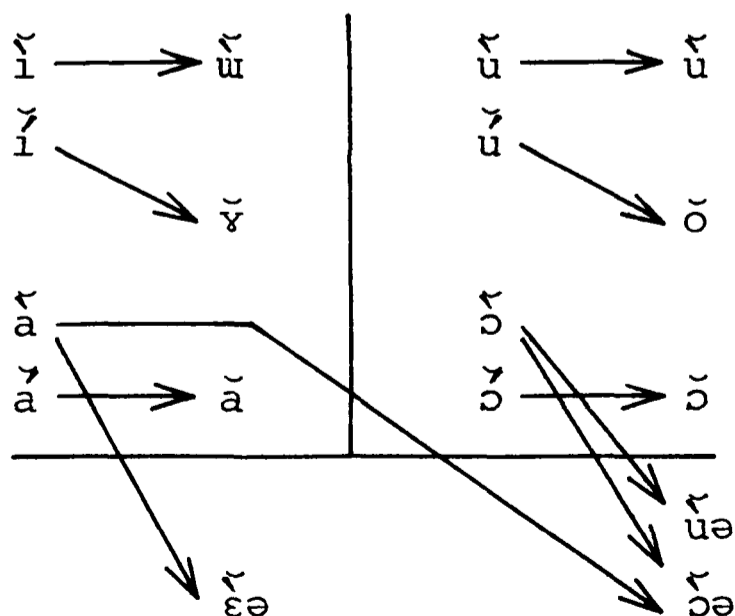
The old system regularly distinguished ĭ, ŭ, ǎ and ǔ. The long phonemes u, y and o as well as e and ε appear to have had no short counterparts in early times; hence graphemes for *ŷ, *ō, *ě and *ē are not found. The modern writing system has a symbol for ŭ, while the sorĭkāt (´) is also used at times to distinguish short ě, as in ʦĕc 'to break off'. The old system of short vowels was therefore:

ĭ [i], written ĭ
 ǎ [a], written ǎ
 ŭ [u], written ŭ
 ǔ [ɔ], written ǔ

With the advent of register, consequent upon the loss of voicing, these developed into:

ǐ	ǔ
Ǒ	ǖ
ǎ	ǔ
Ǒ	ǔ

With the lapse of time these sounds, like their long counterparts, then underwent various modifications, culminating in the following inventory:



The front vowels ǐ and Ǒ were thus altered to nonlabial back

vowels (\check{u} , \check{y}); \check{i} and \check{u} were lowered to ($*\check{e}$ >) \check{y} and \check{o} ; while \check{a} and $\check{\text{ö}}$ were diphthongized, the former to $\check{e}\text{ə}$ or $\check{\text{ö}}\text{ə}$, the latter usually to $\check{e}\text{ə}$ before h, to $\check{\text{ö}}\text{ə}$ before m, and to $\check{u}\text{ə}$ ²⁴ elsewhere. \check{o} is phonetically [o], [ou], [oy] or [^uo]. The resulting system was therefore:

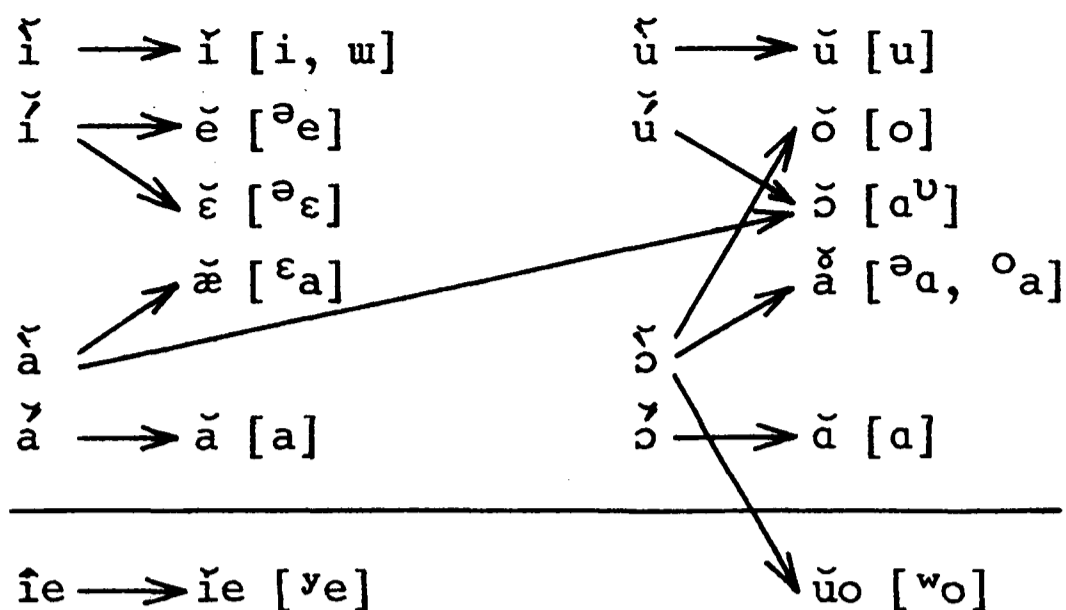
<i>monophthongs:</i>	\check{u}	\check{u}
	\check{y}	\check{o}
	\check{a}	$\check{\text{ö}}$
		$\check{u}\check{v}$
<i>diphthongs:</i>	$\check{e}\check{v}$	
		$\check{\text{ö}}\check{v}$

\check{y} , \check{o} , \check{a} , $\check{\text{ö}}$ belong to the First Register, the others to the Second Register.

There was an urge to adjust this new system of short vowels to that of the long vowels, inasmuch as front $*\check{i}$, $*\check{e}$, $*\check{\text{e}}$ were lacking. Old \check{i} did not invariably become \check{u} , but in some cases remained as \check{i} . Also, the old long è before h was shortened and raised to \check{i} , so that a definite contrast arose between \check{i} : \check{u} : \check{u} . The old é and é before h and s were similarly shortened, resulting in the contrast \check{e} : \check{y} : \check{o} as a secondary development. The old shortened é , however, developed into \check{e} , with the result that there was no $*\check{\text{e}}$ to complete the \check{a} : $\check{\text{ö}}$ contrast.

The system presented here accords with Henderson's interpretation in all essentials. Martini's analysis of the short vowels is quite different. According to him, the following developmental scheme obtains:

²⁴These three diphthongs are properly $\check{e}\check{v}$, $\check{\text{ö}}\check{v}$, $\check{u}\check{v}$.



According to Martini, the old \check{i} became \check{e} before k, η, h but \check{e} elsewhere. The old \check{a} had two replacements while \check{o} had three. The realization of \check{i} is [i] or, more often, [w] (i.e., \check{i}). Martini maintains that the diphthong ie is short and contrasts with $i:e$ from old \check{a} . His system hence looks as follows:

\check{i}	\check{u}
\check{ie}	\check{uo}
\check{e}	\check{o}
$\check{\epsilon}$	$\check{\circ}$
$\check{æ}$	\check{a}
\check{a}	\check{a}

This corresponds to Henderson's system as below:

Old	→	Modern	H ²⁵
\check{i}		\check{i}	\check{i}
\check{u}		\check{u}	\check{u}
\check{y}		\check{e}	\check{y}
\check{i}		$\check{\epsilon}$	
\check{e}		-	\check{i}
\check{e}		-	\check{e}

²⁵Henderson's F is replaced by \check{y} .

<i>Old</i>	→	<i>Modern</i>	
		M	H
ǎ		ǎ̄	ǎ̄ə
ǎ		ǎ̄	ǎ̄ə
ǎ		ǎ̄	ǎ̄
ǔ		ǔ̄	ǔ̄
ǔ		ǔ̄	ǔ̄
ǔ		ǔ̄	ǔ̄ə
ǔ		ǔ̄	ǔ̄ə
ǔ		ǔ̄	ǔ̄ə
ǔ		ǔ̄	ǔ̄ə

Old ǐ, ǔ and ǎ are in correspondence, while the interpretations of ǔ̄, ǎ̄, ǎ̄, ǎ̄ and ǎ̄ differ in ways which are relatively unimportant, as also does ǎ̄. Markedly different in interpretation are the correspondences for ǐ̄, ǔ̄ and ǔ̄: here the realizations which Martini advances say more in favor of Henderson.

Henderson's system is accepted, then, but it is preferable to follow Martini to the extent of taking ǎ̄ instead of Henderson's ǔ̄.²⁶ Thus we obtain the following system of short vowels:

<i>monophthongs:</i>	ǐ̄	ǔ̄	ǔ̄
	ǎ̄	ǎ̄	ǔ̄
	ǎ̄	-	ǎ̄
<i>diphthongs:</i>			ǎ̄ǎ̄
	ǎ̄ǎ̄		
			ǔ̄ǎ̄

Instead of ǎ̄ǎ̄ it might be better to posit ǎ̄ǎ̄, since the first

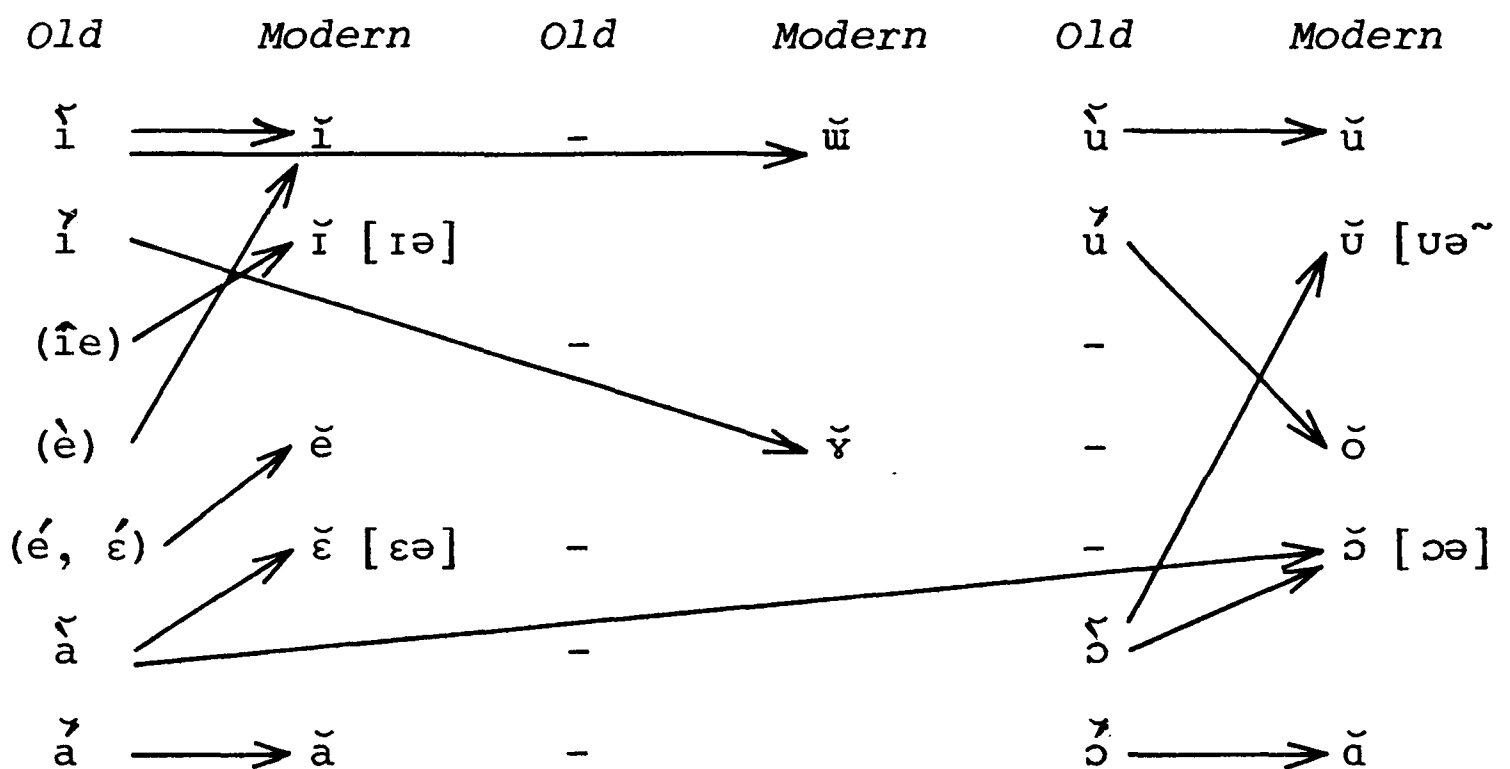
²⁶Cf. also the long vowels!

member of this short diphthong is somewhat open.

Diphthongization may be said to be a basic drift with the short vowels, as is evidenced by the realization of \tilde{o} and by Martini's setting up such phonetic values as $[\text{əe}]$, $[\text{ə}\epsilon]$, $[\text{əa}]$ and $[\text{ə}\alpha]$, as above. For this reason it would not be amiss to set up the phonemes $/\tilde{\epsilon}/$ $[\text{ε}\text{v}]$, $/\tilde{o}/$ $[\text{o}\text{v}]$ and $/\tilde{u}/$ $[\text{u}\text{v}]$ rather than Henderson's short diphthongs. To be sure, these are realized as diphthongs, but their relevant features are openness and advancement. In addition, Martini's statement that his \tilde{ie} is a short diphthong — hence $\tilde{i}\text{v}$ or, more accurately, $\tilde{i}\text{v}$ — can be accepted, and there is nothing to prevent our setting it up as phoneme $/\tilde{i}/$. It is true, however, that this $/\tilde{i}/$ differs from the other short vowels in that it can stand in final position.

Thus the realizations of \tilde{i} , \tilde{u} , $\tilde{\epsilon}$, \tilde{o} , all four of which exhibit phonetic lowering, are always diphthongal, with a short $[\text{ə}]$ or $[\text{v}]$ as their second member.

This yields the following developmental scheme:



Thus the modern short-vowel system matches that of the long vowels perfectly except for the fact that phonemic diphthongs are lacking here.

The end result of all these complex changes is an essen-

tially simple system:

normal (long)			short			diphthongs		
i	u	u	ĩ	ũ	ũ	iy	uy	uy
ı		u	ĩ		ũ			
e	ɣ	o	ě	ỹ	õ	ae	aɣ	ao
ε		o	ě		õ			
a		a	ã		ã			

Thus the vowel system, comprising thirty phonemes, can be expressed without ambiguity with thirteen symbols: a, a, e, ε, i, ı, o, o, u, u, ɣ, u and ˘.

The Consonantism

In the writing system the following consonants are distinguished:

p	b	t	d	ṭ	(ḍ)	c	ḥ	k	g	ʔ
ph	bh	th	dh	(ṭh)	(ḍh)	ch	ḥh	kh	gh	
	m		n		ṇ		ṅ		ṅ	
			l		ḷ					
			r							
		s		(ṣ)		(ś)				ḥ h
	v							y		

To these are to be added the consonant symbols taking the diacritic " (the sōḷāp):

ṁ (ñ) ḡ ḥ (ĩ) ṛ ṽ ḡ ś ḥ ḡ (b̄)

Also to be noted are the diacritic ˘ and the symbols ṛ, ṛ, ḡ and ḡ (vid. the vowels). ˘ is equivalent to -m. The symbols for the sonorant r and l of Sanskrit are graphic only, being

realized as /r, l/ plus /ũ/ or /u/. For all practical purposes ḍ, ṭh, ḍh, ṣ and ṣ are obsolete; ñ, ï and ð are listed in the *Abhidhānaśabda* but are equivalent to ṇ, ḷ and þ — though ñï is usually written for ṇḷ. ʔ is the vowel support, also used in ligatures.

As regards articulation, original b, d, ṣ, g, bh, dh, ṣh, gh have become voiceless, merging with p, t, c, k, ph, th, ch, kh. The retroflex symbols are purely graphic and have no counterpart in pronunciation, where they are realized as the corresponding dentals: ṭ [ṭ], ḍ [ḍ], ṭh [ṭh], ḍh [ḍh], ṇ [ṇ], ḷ [ḷ], ṣ [s]. Palatal ṣ = [s].

According to Martini c is [c] or [t͡ʃ], being close to Russian ТЬ, and is hence often romanized *ti* or *tch*. Again according to Martini, ṣ is [ç] while ch and ṣh are both [t͡ʃ]. Henderson, however, indicates that c and ṣ are both [c] and that ch and ṣh are [ch]. Dialect differences may be involved here. According to Martini again, ṭ (as well as ḍ) and p are [ḍ] and [b] respectively; according to Henderson they are voiced implosive [ḍ] and [b], represented hereafter as preglottalized 'd and 'b. Graphic ḥ is [h], while ~ is [m]; v is either bilabial [w] or labiodental [v] generally without friction: [v]. The ligature hv is [f]. þ is [p]; ñ, ï, ð and so on are realized as the corresponding symbols without the sũḷāp.

In final position only the following phonemes appear:

p	t	c	k
m	n	ṇ	ṇ
	l		
	(r)		
	(s)		h
w		j	

Final *r* and *s* occur in only a few dialects. Elsewhere final orthographic *r* is silent and final *s* is replaced by /h/. Final /w, j/, phonologically consonants, form diphthongs.

The consonants fall into two groups, the registers, according to their effect on the vowels following them.²⁷ The *sõļāp* (‘’) serves to change the register while preserving the consonant's value. Thus *m̄, j̄, ŋ̄, r̄, v̄* and *ȳ* belong with the First Register, *š* and *h̄* with the Second. The sole exception is *p̄*, which remains with the First Register; in this case the *sõļāp* indicates that it is to be pronounced as voiceless [p] rather than as implosive [ɓ] ('b). This ‘’ was introduced in recent times,²⁸ and most of the forms with which it occurs are loans or derivatives in which an earlier consonant cluster is embedded; should the cluster be lost through the dropping of a prefix or addition of an infix, the second member of the cluster presupposes the same vowel as before. To mark this, ‘’ is used, as in *thvay* /twaj/ 'to offer' (First Register) → *ṭəŋvay* /'dəŋwaj/ 'offering' (First Register).²⁹ Otherwise *v* (Second Register) + *a* must be read /iɤ/. Compare also *prv* 'to order' (First Register) → *põr̄v* 'servant' (First Register),³⁰ *m̄əp* = *hm̄əp* 'folded over, doubled'.

It is possible that in *m̄, r̄, v̄* and the like we have the vestiges of former voiceless consonants. Yet, since all these take the First Register, it is not impossible that *m̄, j̄, ŋ̄* and *ŋ* as well as *r̄, l̄, v̄* and *ȳ* were intended to represent former preglottalized implosives ('m, 'ɲ, 'ŋ, 'n, 'r, 'l, 'w, 'j) still

²⁷Vid. the vowels.

²⁸Maspero, *op.cit.*, 130.

²⁹*Ibid.*, 158.

³⁰*Ibid.*, 178.

actually found in other Mon-Khmer languages such as Bahnar.³¹ In Mon and in Khmer alike it is the rule that the implosives (except Mon 'b), whatever their origin, belong to the First Register, which is otherwise the domain of the voiceless consonants. Khmer 'b and 'd, which developed from former voiceless consonants, are now voiced but still belong to the First Register. Mon 'b and 'd, which developed from former voiced consonants, are treated in the same way. Compare:

- Khmer ʈi, Mon tɪ 'land', Mundari ɔte 'field';
Khmer ʈ /'d/ < *t.
- Khmer puon, Mon pɔn, Mundari upun 'four';
Khmer p /'b/ < *p.
- Mon ɖaʔ, Khmer dũk, Mundari daʔ 'water';
Mon ɖ /'d/ < *d.
- Mon 'ba, Khmer bir, Mundari bar 'two';
Mon 'b /'b/ < *b.

All of these Mon and Khmer forms belong with the First Register. So the principle is not violated, as Schmidt mistakenly supposed,³² but is as follows:

implosives:	<i>First Register</i>
nonimplosives:	
formerly voiceless:	<i>First Register</i>
formerly voiced:	<i>Second Register</i>

Thus voice plays no part in determining the register of the implosives.

The number of forms with " is not inconsiderable. Examples:³³

³¹Cf. André-G. Haudricourt, "Les consonnes préglottalisées en Indochine," in *BSLP*, 46 (1950): 172-182.

³²Schmidt, *op.cit.*, §143, p.140.

³³My examples are taken from Guesdon, whose romanizations are given in parentheses. Cf. also Maspero, 69, 72, 130 sq.

- ñ ñak (*ngak*) 'onomatopoeia for dull sounds or cries';
 ñai (*ngai*) 'day', for the more usual thñai;
 phlīt ñoy (*phlēt ngōy*) 'monk's fan'.
- j̄ j̄oŋ (*nhanh*) 'elder sister';
 thw̄y j̄oŋj̄ōk (*thvœ nhonhāk*) 'to toady';
 j̄oŋeŋ (*nhonhēng*) 'to look sour, sullen';
 j̄aŋ (*nhanh*) 'grimace'.
- ŋ = ñ ŋoŋ (*nây*) 'coat';
 ŋa (*na*) 'where, who';
 ŋoh̄ (*nōh*) 'that';
 ŋoŋ (*nōng*) 'firm, solid'.
- m̄ m̄o (*mâ*) 'medicine';
 m̄uv (*mâuv*) 'furrow';
 m̄au (*mau*) 'to be angry'.
- ȳ ȳoŋ (*yâng*) 'species of plant';
 ȳav (*yav*) 'slow';
 paŋyap (*bayap*) 'idol'.
- r̄ r̄oŋ (*râp*) 'to prepare';
 r̄os (*râs*) 'alive';
 r̄aŋ (*rang*) 'hole, pit';
 r̄uv (*râuv*) 'right, good'.
- ï p̄oŋl̄a (*bânla*) 'thorn'.
- ʌ = ï ʌok (*lâk*) 'to scratch, wound';
 ʌay (*lay*) 'all';
 ʌek (*lêk*) 'numeral', Sanskrit *lekha*; the ʌ preserves
 the /e/.
- v̄ v̄aŋ (*vang*) 'to prowl around';
 v̄īk (*věk*) 'to break in, train';
 loŋva (*lova*) 'species of banana'.

It goes without saying that the question of former implosives does not arise in the case of š and ħ.

š šok (sok) 'to insert';

ši (si) 'to eat'.

ḥ ḥaŋ (héang) = haŋ (hang) 'shop, store';

ḥan (héan) 'to dare; bold';

ḥam (héam) 'to flow, gush'.

ḥ is voiceless /p/.

ḥ ḥoŋ (pāng) 'to wish, desire';

ḥoḍḥuc (pāthōch) = proḍḥuc (prāthōch) 'to risk, be in danger';

ḥothom (pāthām) = prothō (prāthām) 'first', Sanskrit *prathama*.

Consonant clusters are important in Khmer. To them belong the aspirated stops, which are combinations to be interpreted not as unit phonemes but as stops + /h/, inasmuch as they may be dissociated, as with dhō 'big' → dōhō 'size'. Clusters comprise only two consonants, though between these a phonetic [h] may intervene. As prior member r, y, v, ŋ, ɲ, n, 'b may not occur (in terms of pronunciation, not of orthography), while the latter member may not be one of the aspirated series. The transition between the two members may be simple — without a phonetic bridge — or it may be effected by junctural [h] or [ə]. Henderson represents this purely phonetic interpolation by underscored h and ə, both devoid of phonological relevance. Possible combinations, in phonemic form rather than transliteration, are the following:

1. *Simple*: kh, ch, th, ph; kr, cr, tr, pr; sk, st, s'd, sp, s'b, sŋ, sɲ, sn, sm, sj, sw, sr, sl.

2. *With junctural h*: ck, tk, pk; kc, kj, tj, pc, pj; kt, ks, pt, ps; kp, kw, cp, cw, tp, tw; kl, cl, tl, pl. These are realized [k^hc], [k^hj], &c., and written chk, chg, thk, thg, phk, phg, khc, khj, phc, phj, &c.

3. With junctural ə: k'd, k'b, c'd, c'b, t'b, p'd, 'd'b. These are realized as [kə'd], [kə'b], &c., and written kt, kp, ct, cp, tp, pt, t̥p.

Clusters with nasals call for separate treatment. With these [h] or [ə] is again interpolated, but usage fluctuates considerably.

1. With h: cŋ, tŋ, pŋ; kŋ, pŋ; kn, cn, tn, pn; km, cm, tm.

2. With ə: kŋ, lŋ; kn, tn, pn; km, tm, lm.

Thus bnau [pənəu] or bhnau [pʰnəu] 'Aegle marmelos, Malabar orange' and gman [kʰmi:ən, kəmi:ən] 'to have not'.³⁴

3. Clusters with m as the prior member show junctural ə: mr, mc, m'd, mn, realized [mər], [məc], &c. Combinations of m + aspirated stops (e.g., mkh) also occur.

Clusters with l as the prior member include lv and lk, for example lvɛŋ /lwɛŋ/ [lɔwɛ:ŋ] 'far, distant'. This is according to Martini. Henderson³⁵ reports the optional insertion of junctural ə or simple clustering, as in lva [lɛv̥iə, lv̥iə] 'fig-tree'.

Clusters with ʔ as the second member are articulated with an intercalated vowel according to Martini but without any vowel according to Henderson. Thus Martini writes kʔ, cʔ, tʔ, pʔ, sʔ, rʔ, lʔ as [kaʔ], [tʃaʔ], [taʔ], [paʔ], [saʔ], [rɔʔ], [lɔʔ]. For example, lʔɔ /lʔa/ 'good, fine' is [lɔʔa:] for Martini, [lʔɔ] for Henderson.

Henderson's analysis disagrees with Martini's in a few other details. For example, she treats r as a possible cluster

³⁴The pronunciation is as given by Martini, op.cit., 128.

³⁵Henderson, op.cit., 167.

initial, as in rəʈuv [rədʋv] 'season', and so forth.³⁶

In the light of the above, the consonant system of Khmer may be said to comprise the following:

p	t	c	k	ʔ
m	n	ɲ	ŋ	
b	d			
	l			
	r			
(f)	s			h
w		y		

Plain b and d may be used for 'b and 'd inasmuch as no confusion can arise. As in the case of the vowels, then, the true consonantism is different from what the writing system would suggest.

The older consonant system may be set up as follows:

p	b	t	d	c	ɟ	k	g	ʔ
	m		n		ɲ		ŋ	
	'b		'd					
	'm		'n		'ɲ		'ŋ	
			l					
			'l					
			r					
			'r					
		s						h
	v					y		
	'v					'y		

³⁶For details see Henderson, *op.cit.*

Admittedly, the positing of the implosives other than 'b and 'd is problematic. ? could be taken as earlier *'g.

The changes which culminated in the modern language may be recapitulated in the following terms. The old voiced stops became voiceless after they, together with the explosive nasals and l, r, w and y, had brought the Second Register into being. Meanwhile, 'm, 'n and the other implosives except 'b and 'd coalesced with m, n and the other explosives but kept the First Register.

Objection to this theory on the grounds that the introduction of the s̄l̄ap is of recent date scarcely affects its validity, inasmuch as we are concerned not with the symbol but with register, which can be older than the diacritic. It is likely that register was marked only in later times and for the sake of precision and the avoidance of ambiguities. In this light, the assumption that words which now take the s̄l̄ap may formerly have had the other register would be mistaken. For the theory to be demonstrated, however, extensive etymological work would have to be done to throw more light on these complex questions. Also to be carried out is more reliable research in the phonology of related languages.