

THE STRUCTURE OF THE WORD IN OLD KHMER

Since the publication in 1954 of M. Cœdès' sixth and last volume of edited and translated Cambodian inscriptions,¹ almost every legible text of Old Khmer which has so far come to light has been readily available in transcribed form to the student of the Khmer language. It is a humble and restricted literature which has come down to us, as the praises of gods and kings and most of the important edicts were written in Sanskrit while the native tongue was found useful chiefly for recording (and ensuring that everyone understood) such matters as the duties of temple slaves, the revenues due to the foundations, and the pious deeds of the local gentry. The language lacks uniformity because the field covered by the inscriptions is both geographically and temporally extensive, spreading over the whole area of Cambodia and spanning the seventh to the fourteenth centuries.

The orthography used for Old Khmer came from outside and did not fit at all perfectly. The Sanskrit consonant-system provided sufficient symbols for the Khmer language but the implication in the consonant-symbol of an 'inherent' vowel to be pronounced after it (except when another consonant or a vowel-symbol or a 'killer' sign was joined to it in writing) was unsuitable for the writing of Khmer consonant clusters. The vowel-system of Sanskrit was inadequate for Khmer and resulted in a certain amount of confusion. It is, for example, difficult to know whether a subscript *v* or *y* is to be regarded as representing the first sound in a diphthong or a second consonant.

Both the inscriptions and the language in which they were written are usually classed as either 'pre-Angkorian' or 'Angkorian'. This convenient chronological and historic division is supported linguistically by the marked difference in the vowel-system of the Angkorian dialect as compared with the language of all the pre-Angkorian inscriptions.

The aim of this study is to produce a system or systems based as far as possible on the internal evidence of the texts rather than on deductions drawn from a diachronic approach. The two dialects are discussed together because the same system of initial consonants and final consonants and the same type of syllable-structure operated in both. The different vowel-systems constitute the only fundamental distinction between them.

Simple initial consonants. Pre-Angkorian and Angkorian dialects

The initial consonant system may be presented as follows²:

¹ G. Cœdès, *Inscriptions du Cambodge*, VI, Paris, 1954.

² The distinction between phonological postulates and the transliterated characters (for which M. Cœdès' system is followed) is maintained throughout by using bold type for the former and italics for the latter.

	Phonology				Orthography			
Velar	k	g	ṅ		<i>k</i>	<i>g</i>	<i>ṅ</i>	
Palatal	c	j	ñ		<i>c</i>	<i>j</i>	<i>ñ</i>	
Dental	t	d	n		<i>t</i>	<i>d</i>	<i>n</i>	
Labial	p	b	m		<i>p</i>	<i>v, b</i>	<i>m</i>	
Semivowel	y	r	l	v	<i>y</i>	<i>r</i>	<i>l</i>	<i>v, hv</i>
Spirant	s	h			<i>s</i>	<i>h</i>		
Glottal stop	ʔ				Initial-vowel symbol			

Orthographically two groups of consonants occur in the spelling of native Khmer words which are not represented in the above tables. They are the aspirated occlusives and the retroflex consonants. They have been rejected for different reasons as follows:

(1) Aspirated occlusives. These do occur as initials in the writing of Khmer words but they are not regarded as representing simple initial consonants. They are used in two ways:

(a) to represent a 2-place initial sequence,¹ **kh**, **gh**, **ch**, etc. The separateness of the occlusive and the aspirate is seen in the infixed forms: e.g. *p-an-hem*, *d-an-hum*.²

(b) to represent in the Angkorian dialect aspiration occurring at the junction between certain 2-place initials, usually an unvoiced occlusive followed by a liquid or nasal: e.g. *khñum*, *thlvāñ*, *phlū*.

The pre-Angkorian *kñum*, *tlōñ*, and *plu* indicate that 2-place sequences, *kñ*, *tl*, *pl*, were here involved, not 3-place sequences (**k** + **h** + **ñ**, etc.)

(2) Retroflex consonants. There are a few occurrences, in words which have a Khmer-like form, of retroflex consonant symbols, e.g. *teñ*, *ñiñ*, *kraññ*.³ The instances are not sufficiently numerous, however, to justify the inclusion of retroflex consonants in the phonological table.

Few instances of the use of palatal or retroflex sibilants have been noted in the spelling of Khmer words.

Two items in the phonological table need further comment:

(1) **b** (*v*, *b*) and **v** (*v*, *hv*). Many words which in the modern language are spelt with initial *b*⁴ were written in Old Khmer with *v*:

¹ Borrowing the useful term of E. J. A. Henderson, 'The main features of Cambodian pronunciation', *BSOAS*, XIV, 1, 1952, 149-174.

² The hyphen is here used to help in demonstrating the infix. Elsewhere, in transliterated words it is used exactly as in Cœdès, *Inscriptions du Cambodge*; in the phonology it is used to indicate disyllables.

³ The word *rnoc* is in the Angkorian dialect often spelt *rñnoc*, but this is an alternative for the spelling with dental *n*.

⁴ The fact that the modern pronunciation of this character is unvoiced does not seem relevant here; the important point is that modern *b* and *v* represent distinct initial consonants.

e.g.	OK	<i>vave</i> <i>valvel</i>	MK	<i>babε</i> ¹ <i>babil</i>	goat taperholder
------	----	------------------------------	----	--	---------------------

There are on the other hand words which have *v* in both Old and Modern Khmer:

OK	<i>vā</i>	(male slave title)	MK	<i>vā</i>	he, she, it, they (derogatory)
	<i>svāy</i> <i>lvā</i>			<i>svāy</i> <i>lvā</i>	mango fig

The decision to assume that two initial consonants, **b** and **v**, operated in Old Khmer (and not just **b** or **v**, represented by *v*) rests in the first instance on some slight evidence in Old Khmer orthography. Thus:

(a) the character *b* occurs as an initial in native Khmer words very occasionally; e.g. in one of the pre-Angkorian inscriptions of Hīn K'ôn, K. 388, the word *ber* 'two' is written several times with initial *b*² (cf. modern *bīr* 'two').

(b) the digraph *hv* was used fairly frequently as an alternative spelling for *v* in some words which in the modern language have *v*.³

e.g. OK	<i>rval</i> or <i>rhval</i> <i>veñ</i> or <i>hveñ</i> cf. also <i>hvek</i>	MK	<i>raval</i> <i>veñ</i> <i>vek</i>	busy long ladle
---------	--	----	--	-----------------------

It would seem reasonable to suppose that the purpose of such a digraph was to distinguish two labial consonants. The modern language, in which *v* (OK *v* and *hv*) and *b* (OK *b* and *v*) represent two phonologically distinct initials, lends support to this assumption.

(2) Glottal stop. In many words the initial consonant character is followed by an initial-vowel symbol, written either below it in the conjunct consonant position or next along the line: e.g. 𑀓𑀺𑀓𑀺⁴ *taet* and 𑀓𑀺𑀓𑀺 *ta-et* (alternative spellings of a slave name most probably), *saap* 'to hate', *paem* 'sweet', and the slave names *kaol*, *laāc*, *saah*, *saot*, etc.

It seems that an initial consonant is involved here, part of an initial sequence. It is proposed to assume that this was a glottal stop⁵ and that the initial-vowel symbol was used to represent glottal stop + vowel. The above-mentioned words would therefore be transcribed *tʔet*,¹ *sʔap*, *pʔem* *kʔol*, etc. The occurrence of the glottal stop can be the only differentiating feature between two words, as in *tem* 'tree' and *tʔem* (slave name).

¹ The same system of transliteration is used for Modern Khmer as for Old Khmer with the additions of the phonetic symbols, *ε*, *v* and *ua* used to represent the vowels

𑀓𑀺, 𑀓𑀺, 𑀓𑀺, 𑀓𑀺, and 𑀓𑀺, 𑀓𑀺.

² The numeral occurs on many other pre-Angkorian inscriptions but is usually not written in word-form. In Angkorian Khmer the word-form is *vyar*.

³ The digraph *hv* is also used to represent a final consonant in Old Khmer which is, in Modern Khmer, *v*; e.g. OK *jāhv*, MK *jāv*, 'to barter'.

⁴ The modern script is used here to facilitate printing.

⁵ As in modern Khmer.

In a word in which the first character is an initial-vowel symbol, e.g. ʔ *oy*, the glottal stop is regarded as operating as a simple initial (phonological transcription, ʔ*oy*).

Vowels

Here there are two separate systems to be considered.

(1) Vowel system of the pre-Angkorian dialect. The following vowels of the Sanskrit syllabary occur in the writing of native Khmer words²:

<i>a</i>	<i>ā</i>
<i>i</i>	<i>ī</i>
<i>u</i>	<i>ū</i>
<i>e</i>	<i>ai</i>
<i>o</i>	<i>au</i>

Of these vowel-symbols the following are not held to represent units of the Khmer system (either pre-Angkorian or Angkorian) and are therefore left out of the final table given below.

ī: This vowel-symbol is used in the writing of the following words: *gī*, *cī*, *nī*, *mī*, *rī*, *sī*, *tnī*, *tgī*; *kamvīc*, *līñ*, *chīñ*; *damrīñ* (for *damrīn*).

These occurrences are few enough to make one suspect, even without further examination, that the symbol *ī* does not represent a unit of the vowel-system. One is not surprised to find that for most of the words with zero final alternative spellings with short *ī* occur with equal frequency (*gi*, *ci*, *ni*, *mi*, *si* have been noted). This would suggest that the vowel-length which was heard in these syllables and recorded in the writing was in fact a prosodic feature, occurring with zero final.

The final consonant may be held responsible for the use of the long vowel-symbol in the case of the three words *kamvīc*, *līñ*, and *chīñ*. The utterance of a palatal consonant after a close front vowel might well be heard as a prolongation of the vowel.

ū: Again, most instances occur in open syllables in words for which an alternative spelling, given in parentheses, has been noted: *kū* (*ku*), *gū*, *cū* (*cu*), *ñū*, *nū* (*nu*), *pū* (*pu*), *rū* (*ru*), *kdū*, *tamrū*, *srū* (*sru*), *laddū*; *cañkūy*, *tpūr* (for *tpur* ?); *kañjū* (for *kañje*, an isolated instance). *ū* in *cañkūy* remains unexplained.

ai and *au*: These vowels do not operate with final consonants. The exceptions to this statement are in words for which an alternative spelling either without the final or with a different vowel is more usual. Thus *daiy* and *kantaiy* are alternatives of *dai* and *kantai* and *dauñ*, *jaun*, and *jauhv* are written in place of the more common *doñ*, *jon*, and *jāhv*. *ai* and *au* are therefore held to represent the vowel *a* with final *y* and *v* respectively (*ay* and *av*).

The other Sanskrit vowel-signs are regarded as representing units of the Khmer system:

¹ For the phonological interpretation of both *taet* and *ta-et* as *tʰet*, see p.10.

² *r* and *l* do occur occasionally in Pre-Angkorian and Angkorian Khmer, e.g. *tr* (MK *tri* 'fish') but the examples are too few to need discussion here.

a and *ā*: There are many instances of confusion between *a* and *ā*, not only in the writing of open syllables, as with *i* and *ī*, *u* and *ū* but also with closed syllables, as in the alternative spellings:

jav and *jāhv* 'to barter'
ay and *āy* 'to, at'
travañ and *travāñ* 'lagoon'

Nevertheless it seems that there is an opposition between short and long *a*, illustrated by these pairs of words:

<i>tañ</i>	'to put up'	<i>tāñ</i>	(title)
<i>man</i>	'which'	<i>māñ</i>	'to have'
<i>cañ</i>	'to provide'	<i>cāñ</i>	'to guard'

which justifies their inclusion as two units in the table of vowels.

i: With *ī* excluded, *i* is a straightforward case. This vowel-unit has a fair distribution, being used with most initials and finals.

e.g.	<i>dik</i>	'water'	<i>piñ</i>	'pool'
	<i>sin</i>	'another, more'	<i>si</i>	(male slave title)

u: Used with all initials and all finals except *v*.

e.g.	<i>ku</i>	(female slave title)	<i>duk</i>	'to put'
	<i>ruñ</i>	'wide'	<i>duñ</i>	'to buy'

e: Examples:

<i>ket</i>	'to wax'	<i>tel</i>	'where, which'
<i>tem</i>	'tree'	<i>jeñ</i>	'leg'
<i>ge</i>	'person'		

In addition to the vowel-symbols provided by the Sanskrit syllabary, certain digraphs were used, one composed of two vowel-signs used with the same consonant (*u* and *i*, written in their normal positions, the one below and the other above the initial consonant), the others composed of *y* or *v* with another vowel-symbol following, e.g. *ya*, *yā*, *ye*; *va*, *vā*, *vo*. These non-Sanskritic digraphs are discussed below.

ui: This has a high frequency rate in the pre-Angkor inscriptions as it is used in the spelling of a common word, *gui* 'that is, namely'. However, only one other word in which *ui* is used has been noted. This is a slave name of unknown meaning, *latui*. One cannot conclude that this digraph represents a separate vowel-unit.

ya, *yā*, *ye*: The total number of instances of all three of these digraphs is small. If a diphthong is here involved, the character *y* has a dual role to play, that of a consonant *y*¹ as well as that of diphthongizer and in many cases one cannot be sure which role it is playing.

However, in certain words which have complex (orthographic) initial sequences, with *y* in the last place, it can be shown by reference to the permissible consonant-patterns in initial sequences² that *y* could not be a

¹ As in the words *yi*, *yuñ*, *yok* 'to take', *yoñ*, *yon*, *yol*.

² Discussed on pp. 11-13 below.

consonant and must therefore be part of the vowel-unit. These words are:
hyam: **h** does not occur in an initial complex with any other consonant in second place.

tgyap: A 3-place initial sequence always has either a nasal or **r** or **l** in second place and the three consonants are not then written as one conjunct cluster.

*kyrel*¹: The consonants here written could operate as an initial sequence, though no other examples of such a sequence have been noted. It would not then be written as one conjunct cluster, however.

kantyar: If this word were analysed as having a 4-place initial sequence it would be the only instance in which **y** came in fourth place.

chdyās: The aspiration here recorded is held to be a prosodic feature of junction. If **y** is here consonantal, the sequence **cdy** would be in operation; but a 3-place sequence always has a nasal or **r** or **l** in second place.

One may conclude, therefore, that an attempt was being made to write a diphthong for which the Sanskrit syllabary did not provide a suitable character. Any such attempt was breaking new ground and was likely to cause difficulties and confusion at first; hence the three methods of spelling.

The words cited above happen to have occurred on inscriptions from both north and south Cambodia. It is therefore assumed that the diphthongization was not confined to the one local dialect and that it must be recognized in the table of pre-Angkorian vowel-units.

In the above-mentioned instances of this diphthong a fair variety of initial and final consonants occur.

The diphthong seems to have had as its starting point a close front vowel. The variety of spellings suggests that the finishing point was probably neither **a** nor **ā** nor **e** but a vowel of more neutral colour. In the phonological table of vowel-units it is represented as **īə**.²

va, *vā*, *vo*: The same difficulty arises here as with the previous diphthong, since **v** can be consonantal³ (as in *tve* 'to do', *tamve* 'doer', where the infix is written in the manner which is usual when it occurs between two consonants) and it can represent part of a vowel-unit. Here, too, the total number of instances of all three digraphs is not large. The instances in which **v** can be shown to represent part of a diphthong are fewer than for **īə**:

kandvat: The modern form of this word, *kanduət* 'Phyllanthus distichus', has the diphthong *uə*.⁴

kandvac, *kantvoh*: If these words were analysed as having 4-place initial sequences, they would be the only instances in which **b** or **v** came in fourth place.

¹ Modern Khmer *kryel* 'crane', in which *ye* represents a diphthong.

² No precise phonetic knowledge of the realization of this or any other vowel-unit is claimed; the use of the length-mark here is due to the need to distinguish phonologically between **īə** (pre-Angkorian and Angkorian) and **ie** (Angkorian only). **ūə** (pre-Angkorian and Angkorian) and **uo** (Angkorian only) are also distinguished in this way.

³ Representing both **b** and **v**.

⁴ Contrast OK consonantal **v** > MK *v*, e.g. pre-Angkorian *svā*, *lvā*, MK *svā* 'monkey', *lvā* 'fig.'

⁵ see pp. 11-13 below for discussion of permissible patterns in initial sequences.

Some slight additional support may be drawn from the words: *jvor* (Modern Khmer *juar* 'row, line'?), *juān*, *juak*. The 2-place initial *ju* in the one other word in which *j* and *v* occur together initially with *a*, *ā* or *o* following, is written with the digraph *hv*: *jahvol*.

Slender though this evidence is, it seems reasonable to conclude that a diphthong with a close back vowel as its starting point and a neutral vowel as its finishing point was heard and recorded. This is entered in the vowel table as *uə*.¹

It might be argued that in the cases of both *ia* and *uə* although a diphthong may have been heard and recorded, it may not necessarily have been a separate vowel-unit. Here I feel that evidence provided by the Angkorian dialect weights the scales in favour of the recognition of functional *ia* and *uə*. Although in the Angkor period the scene of activity shifted and a different dialect was recorded, in many ways a continuing orthographic tradition is, as might be expected, clearly discernible. The early efforts made by the pre-Angkorian Khmers to distinguish *ia* and *uə* from the other vowels were carried on and developed by their successors. A far greater proportion of words were then written with these diphthongs without the exclusion of any other vowel-units which operated in pre-Angkorian Khmer.

Table of pre-Angkorian vowel-units

Phonology		Orthography	
a	ā	<i>a</i>	<i>ā</i>
i	e	<i>i, ī</i>	<i>e</i>
u	o	<i>u, ū</i>	<i>o</i>
	ia	<i>ya, yā, ye</i>	
	uə	<i>va, vā, vo</i>	

Note: *ai* and *au* should be mentioned as orthographic representations of *a + y*, *a + v* respectively.

(2) Vowel-system of the Angkorian dialect. All the pre-Angkorian vowel units occur as vowel-units in the Angkorian dialect. Less confusion arises over *a* and *ā*, and *uə*. *i* and *u* are again not regarded as representing units of the system as very few occurrences have been noted except in open syllables.

Many Angkorian words in which the vowel-units common to both dialects occur may be directly compared with the pre-Angkorian form:

	Pre-Angkorian	Angkorian	Meaning
a	<i>jmaḥ</i>	<i>jmaḥ</i>	name
	<i>tnal</i>	<i>tnal</i>	made road
	<i>psam</i>	<i>phsam</i>	to unite
ā	<i>mān</i>	<i>mān</i>	to have
	<i>mās</i>	<i>mās</i>	gold
	<i>slā</i>	<i>slā</i>	areca
i	<i>dik</i>	<i>dik</i>	water
	<i>kāmpit</i>	<i>kāmpit</i>	(slave name)

¹ See p. 6, n. 2.

	<i>pi</i>	<i>pi</i>	in order to
	<i>aṃvil</i>	<i>aṃvil</i>	tamarind
e	<i>teṃ</i>	<i>teṃ</i>	tree
	<i>jeṇ</i>	<i>jeṇ</i>	leg
	<i>sre</i>	<i>sre</i>	riceland
u	<i>sruk</i>	<i>sruk</i>	country
	<i>kñūṃ</i>	<i>khñūṃ</i>	slave
o	<i>roc</i>	<i>roc</i>	to wane
	<i>poñ</i>	<i>poñ</i>	(title)
	<i>sot</i>	<i>sot</i>	moreover, too
	<i>so</i>	<i>so</i>	white
īe	<i>syam, syām</i>	<i>syam, syām</i>	Siamese
ūe	<i>kandvat</i>	<i>kandvāt</i>	Phyllanthus distichus

The only fundamental difference between the Angkorian dialect and the pre-Angkorian lies in the occurrence of two extra diphthongs in the vowel system. These are:

(i) a vowel written *ya* or *yya*, but never *yā* (being thus distinguished from *īe* which is written *ya*, *yya*, and *yā* in the Angkor period).

(ii) a vowel written *va*, but never *vā* or *vo* (being thus distinguished from *ūe* which is written *va*, *vā*, and *vo* in the Angkor period, as in the pre-Angkor). Examples with pre-Angkorian equivalents:

	Pre-Angkorian	Angkorian	Meaning
(i)	<i>dnem</i>	<i>dneyam</i>	pair, team
	<i>ber</i>	<i>vyar</i>	two
	<i>vñe</i>	<i>vñya</i>	flower
	<i>valvel</i>	<i>valvyal</i>	taperholder
	<i>tamre</i>	<i>tamrya, tamryya</i>	elephant
(ii)	<i>kon</i>	<i>kvan</i>	child
	<i>jon</i>	<i>jvan</i>	to offer
	<i>paon</i>	<i>pha^avan</i>	younger sibling

These vowels are represented as *īe* and *ūe* in the table of vowel-units.

Table of Angkorian vowel-units			
Phonology		Orthography	
a	ā	<i>a</i>	<i>ā</i>
i	e	<i>i, ī</i>	<i>e</i>
u	o	<i>u, ū</i>	<i>o</i>
īe	īe	<i>ya, yya</i>	<i>ya, yya, yā</i>
ūe	ūe	<i>va</i>	<i>va, vā, vo</i>

Further examples of words in which the Angkorian vowels occur:

a	<i>ta</i> 'to'	ā	<i>slāp</i> 'to die'
	<i>tai</i> (female slave title)		<i>jā</i> 'as, being'
	<i>rat</i> 'to run'		<i>tahvāy</i> 'offering'
	<i>vnam</i> 'hill'		<i>prāk</i> 'silver'

i	<i>gi</i> 'that is, namely' <i>tiñ</i> 'to know' <i>liḥ</i> (small measure of weight) <i>damriñ</i> 'trellis'	e	<i>khnet</i> 'period of waxing moon' <i>ge</i> 'person' <i>neh</i> 'this' <i>pre</i> 'to use'
u	<i>kaṃluñ</i> 'middle' <i>nu</i> 'at' <i>yul</i> 'to understand' <i>ruñ</i> 'wide'	o	<i>lño</i> 'sesamum' <i>noḥ</i> 'that' <i>phon</i> 'also' <i>rañko</i> 'husked rice'
ia	<i>kanmyañ</i> 'young' <i>jyañ</i> (measure of weight) <i>paryyañ</i> 'oil' <i>ampyal</i> 'salt'	ia	<i>camryañ</i> , <i>camryyañ</i> , <i>camryāñ</i> 'singer' <i>dyāñ</i> 'candle' <i>ryām</i> 'brother' <i>canlyak</i> , <i>canlyāk</i> 'clothing' <i>cancyāñ</i> 'ring'
ua	<i>pandval</i> 'to inform' <i>svaṃ</i> 'to ask, to beg' <i>vvaṃ</i> 'not' <i>tvañ</i> 'old woman, grandmother'	ua	<i>kanmvāy</i> 'nephew' <i>dvot</i> 'great grand(parent), great grand(child)' <i>paṃmvās</i> 'clergy' <i>mvay</i> , <i>mvāy</i> , 'one'

Final consonants. Pre-Angkorian and Agkorian dialects

Phonology				Orthography			
k		ñ		k		ñ	
c		ñ		c		ñ	
t		n		t		n	
p		m		p		m, m̐	
y	t	l	v	y ¹	r	l	v, hv ¹
s	h	Zero		s	h, ḥ		

Notes on the orthography:

(i) Double final consonants occur but this is only a device to mark the end of a word—an alternative to the use of the 'killer'. Thus *nakk* is clearly the same word as *nak*, *nak* 'person'; e.g. K. 137, 1, 6, *nakk pamre* 'persons serving'. *rohḥ*, *nehḥ* 'all these', which occurs frequently, is the same as *roh neh*.

(ii) No opposition occurs between *m* and *m̐*, *h* and *ḥ*; usually final *m* and *h* are written *m̐* and *ḥ*.

(iii) The digraph *hv*, already discussed in the section on initial consonants, is often written instead of *v*, e.g. *kāhv*, *jāhv*, *kandehv*.

Notes on the phonology:

No complete statement on the frequency and distribution of the finals can be made since the nature of a preceding vowel is in many words not precisely known (e.g. sometimes one cannot tell whether *ya*, in an Angkorian word for which no other spelling occurs and of which the modern equivalent is not known, represents *ia* or *īa* or consonant *y* and *a*).

¹ *ai* and *au* are held to represent *a* + final *y* and a final *v* respectively and should therefore be mentioned here.

However, the following is general indication of the comparative distribution of the final consonants:

(i) **k, ŋ, ŋ, m**. Velar finals have the greatest frequency. They and **n, m**, whose frequency is rather less, occur following all vowel-units, i.e. both pure vowels and diphthongs.

(ii) **ñ, v, h**, Zero. These finals have the most restricted use, and follow mainly the pure vowels. Instances of *hv* following *o* are suspected of being mis-spellings of **av** or **āv** (e.g. *johv* for *jāhv*, *yohv* for *yau*).

(iii) In between these two extremes comes **c, t, p, y, r, l, s**, which occur following all pure vowels and some diphthongs.

Word-structure. Pre-Angkorian and Angkorian dialects

A. Monosyllables

(1) CV(C), simple initial. The monosyllable with simple initial is also the simplest form of word. All the initial consonants may occur as simple initials.

e.g.	<i>vā</i> , vā	(male slave title)	<i>cāp</i> , cāp	to seize'
	<i>hau</i> , hav	'to call'	<i>oy</i> , ʔoy	'to give'

(2) CCV(C), 2-place initial. Orthographically, three forms of word occur here:

(i) Words with a 2-place initial sequence may be orthographically quite straightforward:

<i>cdin</i> , cdin	'stream'	<i>tmāt</i> , tmāt	(vulture)
<i>jhe</i> , jhe	'tree'	<i>lvā</i> , lvā	'fig'

(ii) A junctional prosodic aspiration, occurring especially between an unvoiced occlusive and a following nasal or liquid, is recorded, almost exclusively in the Angkorian dialect, by the use of an aspirated occlusive character. The aspiration is not noted in the phonological transcription.

<i>thpvañ</i> , tpuən 'south'	<i>chmām</i> , cmām 'guardian'
<i>phlū</i> , plu 'road'	<i>khloñ</i> , kloñ (title)

(iii) Many words are orthographically disyllabic, with the 'inherent' vowel in the first syllable:

၂၂၂ *vave* 'goat', *karom* 'below', *raman* 'Mon', *tadai* 'different', *cake* 'dog'.

There are numerous cases where alternative spellings of such words as monosyllables are current:

ကံ krom, ကြ rmañ, တို t dai, ငှံ cke.

This suggests that here we have an attempt to record another prosodic feature of junction, probably a short neutral vowel occurring at the junction of the two consonants. This feature is not noted in the phonological transcriptions.

The inherent vowel is written in this way between all kinds of initial consonants, without apparent restriction. Thus, for example, it is found between two occlusives (*kadam*, **kdam** 'hut', and *tadai*, **tdai** and *cake*, **cke**, mentioned above); it occurs where a sonant occlusive is followed by a liquid

or nasal (*jamaḥ*, **jmaḥ** 'name', *daluh*, **dluh** 'through')¹; and it is especially common where *r* is involved (*karol*, **krol**, 'park, enclosure', *rapam*, **rpam** 'dance', *marām*, **mrām** 'finger').

The formula suggested for all three types of word described above, whether the junction is marked by aspiration, by the occurrence of an anaptyctic vowel, or by nothing, is CCV(C).

Note: The consonants which occur in first and second positions are very varied, the only limitations being that **h**, **v**, **y** and **ñ** occur only in second place.² This will be referred to again.

The behaviour of the glottal stop (represented by the initial-vowel symbol), when it occurs in first place, needs further comment. In several common words there seems to be a doubt about whether it is to be taken as present:

cf. 𑀓𑀺𑀓 ^a*tā*, alternative form of 𑀓𑀺𑀓 *tā* 'old man'

𑀓𑀺𑀓 *anak* and 𑀓𑀺𑀓 ^a*nak*, alternative forms of 𑀓𑀺𑀓 *nak* 'person'

𑀓𑀺𑀓 *anau*, alternative form of 𑀓𑀺𑀓 *nau* 'to stay, to live'

𑀓𑀺𑀓 *aji* and 𑀓𑀺𑀓 *aji*, alternative forms of 𑀓𑀺𑀓 'grandparent'

𑀓𑀺𑀓 *ame* and 𑀓𑀺𑀓 ^a*me*, alternative forms of 𑀓𑀺𑀓 *me* 'mother';

(It will be observed that alternative spellings of the initials with and without the 'inherent' vowel are current for several of these words).

Other words of less certain meaning seem to have the same alternative initial, glottal stop + consonant or single consonant:

akān, alternative form of *kān* 'young lady'?

alān " " " *lān* 'to wash'?

Compare also the slave names of unknown meaning, *adah* and *dah*, *adu* and *du*, *adat* and *dat*, *asā* and *sā*.

One must not forget, however, the possibility of opposition here, between the 2-place initial (glottal stop + consonant) and simple consonant. This surely occurs between the words 𑀓𑀺𑀓 ^a*muḥ*, **ṛmuḥ**, and





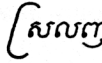
𑀓𑀺𑀓 *muḥ*, **muḥ** ? **ṛmuḥ**, M. Cœdès shows,³ should be an infixed form of **ṛuḥ* 'to heat' (**m** infix). **muḥ** 'nose' ? occurs in a place-name (K. 426, 1, 2) *cuḥ muḥ* 'end of nose'.

¹ There is an interesting distinction here between the behaviour of surd and sonant, a distinction which the writing suggests but which no longer operates in Modern Khmer. Thus while the junction of a surd and a liquid or nasal is often marked by aspiration, that between a sonant and a liquid or nasal is often marked by the anaptyctic vowel.

² Many female slave names are preceded by *ya*, a Mon female slave title (I am indebted to Professor G. H. Luce for this suggestion); e.g. *ku ya māy*. Orthographically one cannot tell whether two words, *ya māy*, or one word *yamāy*, is intended, but phonologically **ya māy** is suggested.

³ op. cit, III, p. 75, n.3.

(3) CCCV(C). Many words resemble the last group, (2(iii)), in that they are orthographically disyllables and have the 'inherent' vowel in the first 'syllable'. They are different in having three consonants before the main vowel:

	<i>danhum</i>	'perfumed'		<i>damnuñ</i>	'purchase'
	<i>dalmāk</i>	'mahout'		<i>samlo</i>	'soup'
	<i>sralañ</i>	'to love'			

The analysis of the 3-place initial sequences is based on the restricted nature of the second consonant. This is either

(i) a nasal (preceded by the anaptyctic vowel)
or (ii) *r* or *l* (usually *r* is followed and *l* preceded by the anaptyctic vowel but this is not always so).

(i) Nasal consonant in 2nd place, CNCV(C).¹

There is a strong tendency for the nasal to be written as the homorganic of a following occlusive:

<i>laŋgau</i>	<i>lŋgav</i>	'copper'	<i>kañje</i>	<i>kñje</i>	'basket'
<i>pandval</i>	<i>pndual</i>	'to inform'	<i>campok</i>	<i>cmpok</i>	' <i>Buchania fastigiata</i> '

When the consonant in third place is not an occlusive, the orthography of the nasal is usually as follows:

Velar nasal, <i>ŋ</i>	before <i>v</i>
Dental nasal, <i>n</i>	" <i>m</i>
Labial nasal, <i>m</i>	" <i>ñ</i> or <i>n</i>
An alternance of <i>n</i> or <i>m</i>	" <i>y, r, l, s, h</i> and initial-vowel symbol.

There is clearly not a 4-term system here, but on the other hand one cannot assume that mere nasality, with the exact articulation dictated by the context, is in operation either, as there is evidence of grammatical opposition between *n* and *m* (see below).

(ii) *r* or *l* in 2nd place, Cr/lCV(C).

The instances of the occurrence of *l* are few but are consistently spelt with *l* and not *r*:

<i>kalmon</i>	<i>klmon</i>	'wax'	<i>dalmāk</i>	<i>dīmāk</i>	'mahout'
<i>kalpau</i>	<i>klpav</i>	'Hydrocarpus anthelmintica'	<i>klamum</i>	<i>klmum</i>	'maiden'
<i>valvel</i>	<i>blbel</i> ²	'taperholder'			

There are many instances of *r* in second place:

<i>krapas</i>	<i>krpas</i>	'cotton'	<i>krapi</i>	<i>krpi</i>	'buffalo'
<i>prahveñ</i>	<i>prveñ</i>	'length'	<i>karcol</i>	<i>krcol</i>	(Meaning not known. Cited here because of position of anaptyctic vowel.)

¹ *N* is used with the implication 'any one of the four nasal consonants, *ñ, ŋ, n, m*'.

² Modern Khmer *babil* supports the phonological transcription *b* for *v*.

(4) CCCV(C). Four-place initial consonants do occur, but comparatively very rarely. Of the occurrences which have been noted, a large proportion have **h** or **r** in 4th place:

kandhar, **kndhar**
kamphac, **khmpac**
kamphik, **kmphik**
kambhan, **kmbhan**
kambhat, **kmbhat**
kanthok, **knthok**
kandha, **kndha**
kandhan, **kndhan**
kandhas, **kndhas**

kandhik, **kndhik**
kandhu, **kndhu**
candhan, **cndhan**
rambha, **rmbha**
kantran, **kntran**
kaṃprvat, **kmpruat** or **kmprūt**
kaṃvrau, **kmbrav**
kaṃvrah, **kmbrah**

The words *kaṃsten*, *kantakoy*, *kaṃphsik*, and *laṅklaḥ* could be analysed as monosyllables with 4-place initials although the occurrence of *t*, *k*, *s*, and *l* in 4th place seems to be peculiar to the examples here cited.

It will be observed that all the examples have a nasal in second place. Their precise formula is therefore CNCCV(C). Perhaps in the words *craṃṇep*, *kraṇṇāṇ*, *kaṃrateṇ*, one may be permitted to see three examples of 4-place initial sequences in which both *r/l* and *N* occur: **crṃṇep**, **krṇṇāṇ**, **Cr/INCV(C)**; **kmrteṇ**, **CNr/ICV(C)**.

The rare forms *kathor*, *kabhap*, *rabhass*, seem to belong to neither section 3 nor 4. Although they have only three initial consonants, the second is not *N* or *r/l*. The modern form of *kathor*, it is interesting to note, could be analysed as a monosyllable with a 4-place initial (MK *kanthor*), CNCCVC, 'spittoon'.

Note: Certain alternative spellings occur—some with great frequency—in which a nasal or other consonant in the initial sequence is reduplicated (e.g. *rddeḥ*, *rṃmam*, *taṃmrya*, *paṃmre*, *kaṃmrateṇ*, alternative spellings of *rdeḥ*, *rmam*, *taṃrya*, *paṃre*, *kaṃrateṇ*). The analysis of such words is based on the spellings without the reduplicated consonant (**rdeḥ**, **rmam**, **tmrya**, **pmre**, **kmrteṇ**); cf. MK *radeḥ*, *taṃrī*, *paṃmr̥*.

The distribution of initial consonants described above reveals a striking inequality between the various units of the initial system. Thus while the plosives and *s* occur in both first and third positions of a 3-place sequence and have either *N* or *r/l* in second place, the restricted consonants **h**, **v**, **y**, and **ṇ** which, it was mentioned, occur only in the second position of a 2-place sequence, are found only in the third position of a 3-place sequence.

The consonants **r** and **l** do not occur with *r/l* in second place (as might be expected), nor do the nasals **m**, **n** and **ṇ** occur with *N* in second place. The glottal stop again behaves slightly irregularly in that it, like *r*, *l*, and the nasals, does not play a full role. It occurs in both first and third positions but when it is in first place, only *N*, and not *r/l*, can be in second place. Examples:

(i) Plosives and *s*:

(a) Occurring in 1st place with *N* in 2nd place:

kañjan, kñjan *camnat, cmnat* *jamhañ, jmhañ* *taṃrvac, tṃrṭəc*
damnuñ, dmnuñ *paṃnos pṃnos* *vamñā, bṃñā*¹ *sam-ap, smʔap*

(b) Occurring in 3rd place with N in 2nd place:

camkā, cmkā *paṅgam, pṅgam* *kañcas, kñcas* *kañjes, kñjes*
kantur, ktur *kandat, kndat* *sampok, smpok* *kansom, knsom*

(c) Occurring in 1st place with r/l in 3rd place:

krapi, krpi *gragar, grgar* *crakan, crkan* *jranyañ, jrniñ*
tralā, trlā *dradap, drdap* *pradul, prdul* *vraneñ, brneñ*¹
sralañ, srlañ

(d) Occurring in 3rd place with r/l in 2nd place:

krakap, krkap *srageḥ, srgeh* *kracok, krcok* *krapi, krpi*
*travañ, trbañ*² *krasop, krsop*

(ii) The restricted initials **h**, **v**, **y**, **ñ** occurring in last place only:

kraham, krham *krahvañ, krvañ* *kraññañ, krññañ*

(ii) **r** and **l**. Occurring with N but not with r/l:

(a) Occurring in last place with N in 2nd place:

raṃnoc, rṃnoc *laṃñāc, lṃñāc*

(b) Occurring in 3rd place with N in 2nd place:

pamre, pmre *panlas, pnlas*

(iv) Nasals. Occurring with r/l but not with N:

(a) Occurring in 1st place with r/l in 2nd place:

mratañ, mrtañ *mrasiṛ, mrsiṛ* *ñrañor, ñrñor*

(b) Occurring in 3rd place with r/l in 2nd place:

prañā, prñā *krañañ, krñañ* *klamum, klmmum*

(v) Glottal stop:

(a) Occurring in 1st place with N in 2nd place:

amnoy, ʔmnoy

(b) Occurring in 3rd place with N in 2nd place:

sam-ap, smʔap *kan-in, knʔin*

(c) Occurring in 3rd place with r/l in 2nd place:

kalaoh, klʔoh *sra-ip, srʔip*

Affixation

The composition of initial sequences is directly concerned with grammar. There is evidence to be drawn from certain pairs of words of infixed **m** and **n** and of prefixed **pr** and **pN**³ having an effect on the meaning at the grammatical level; e.g.:

m infix (agent)

2-place initial

Simple initial		with infixed consonant
<i>soñ, soñ</i>	to pay back	<i>smon, smon</i> he who reimburses
<i>cer, cer</i>	to transgress	<i>cmer, cmer</i> a transgressor

¹ By definition, so to speak, **v** must be construed as **b** here (since **v** is an initial of restricted use). Modern Khmer *bhñā* corroborates the **b** in *bṃñā*.

² Modern Khmer *drabañ* establishes the **b** here.

³ There seems to have been a prefix **kN** too; cf. *je, je* (capacity measurement) and *kañje, knjë*, basket used for measuring. A very large proportion of words have a 3-place initial of which the first two consonants are **kN**.

The structure of the word in Old Khmer

<i>som</i> , som	to ask	<i>smom</i> , smom	a beggar
<i>jon</i> , jon	to offer	<i>jmon</i> , jmon	a person offering

3-place initial

2-place initial		with infixed consonant	
<i>pre</i> , pre	to use	<i>pamre</i> , pmre	servant
<i>tve</i> , tve	to do	<i>tamve</i> , tmve	a do-er

n infix (utensil, means)

Simple initial		2-place initial with infixed consonant	
<i>vās</i> , vās	to measure	<i>vnās</i> , vnās	(measurement)
<i>jāhv</i> , jāv	to barter	<i>jnāhv</i> , jnāv	price
<i>roc</i> , roc	to wane	<i>rnoc</i> , rnoc	period of waning moon

2-place initial		3-place initial with infixed consonant	
<i>sre</i> , sre	riceland	<i>sanre</i> , snre	ricefield
<i>tloñ</i> , tloñ	20 (capacity) measures	<i>tanloñ</i> , tnloñ	(classifier) ¹

mn infix (nominalizer)

Simple initial only		3-place initial with 2 infixed consonants	
<i>jon</i> , jon	to offer	<i>jamnon</i> , jmnnon	offering
<i>pos</i> , pos	to enter the religious life	<i>pamnos</i> , pmnos	monk ²
<i>duñ</i> , duñ	to buy	<i>damnuñ</i> , dmnuñ	purchase
<i>oy</i> , oy	to give	<i>amnoy</i> , ʔmnoy	gift

Note: No anaptyctic vowel occurs between the nasals.

pr prefix (nominalizer ?)

Simple initial		3-place initial with 2 prefixed consonants	
<i>veñ</i> , veñ	long	<i>prahveñ</i> , prveñ	length ?
<i>dul</i> , dul	supporting	<i>pradul</i> , prdul	a support ?
		pN prefix	
<i>jāhv</i> , jāv	to barter	<i>pañjāhv</i> , pñjāv	things acquired by exchange

B. Disyllables

A disyllable is here defined as two monosyllables operating together as one word or free form; each of the two parts can be analysed in the terms applied above to monosyllables. The formula is therefore

(C)(C)(C)CV(C)–(C)(C)(C)CV(C)

¹ Numerals follow *tloñ* but precede *tanloñ*, e.g. *tloñ* 2 *tanloñ*.

² Angkorian dialect, *pvas*, **pūas**, *pamnvās*, **pmmūas**

The implication of V in the formula is that one of the vowel-units shown in the tables of vowel-units, pre-Angkorian and Angkorian, occurs in each syllable.

It is not always easy to know from the orthography whether a disyllable of this kind is intended or not, as the delimitation of words is not clear. No spaces occur between words. The end of the word may be recognized if there is a final consonant other than **m**. Thus *h*, the final **h** symbol, marks the end of a word; and the diacritic ⁻, written above a consonant-symbol other than *h* or *m*, implies that the consonant shall be uttered alone, not followed by the 'inherent' vowel, **a**, e.g. កេត *ket*, not *keta*; this marks the end of many Khmer words. The anusvara, *m̐*, is used for final **m** (e.g. សំ *sam̐*, **sam**) but also occurs when **m** is in the second place of a 3-place sequence (e.g. សំរុក *samruk̐*, **smruk**). One cannot therefore rely on it. The writing of a double consonant, as in *nakk*, **nak**, has been mentioned below as an alternative to the diacritic ⁻, for marking the end of a word.

If the final consonant is zero, there is no orthographic means of showing it. Thus តោតិល *totil*, could be *to-til*, or *to til*. The repeated occurrence of the two syllables together is helpful. M. Cœdès transcribes *vodi* as one word, since the two syllables always occur together and seem to have a simple semantic meaning, 'vase'. This would also seem desirable in the case of ឧលិក *ulik*, **u-lik** 'melon' and មីម៉យ *mimoy*, **mi-moy** 'each'.

In other cases the disyllabic interpretation seems not so certain. In the words

<i>kurāk</i>	<i>kuñjāhv</i>	<i>gulut</i>
<i>kurek</i>	<i>kunti</i>	<i>duvau</i>
<i>kuruñ</i>	<i>tuñnot</i>	<i>vuyuk</i>

it is possible that the short *u* is used, as it is used in Old Mon, to indicate the occurrence of the anaptyctic vowel and that the phonological interpretation should be monosyllabic:

krāk ?	kñjāv ?	glut ?
krek ?	knti ?	dvav ?
kruñ ? (Mod. Khm. <i>kruñ</i> 'city, protection')	tñnot ? (Mod. Khm. <i>tnot</i> 'sugar-palm')	byuk ?

On the other hand some other disyllables in M. Cœdès' transcription might, according to the present analysis, have to be regarded as two words: *cidok*: *dok* 'boat' occurs separately.

ci frequently occurs in titles, e.g. K. 137, 16, *ku ci mratān*; K. 22, 15, *va ci poñ... va ci dat*. In K. 30 where *cidok* occurs, it is in a phrase where one expects a title, *sre amnoc* (for *amnoy*) *cidok*, **sre** **mnoy ci dok**, 'riceland presented by Ci Dok'.

yakāy, *yañau*, *yanoy*, *yanoc*, *yajuñ*, *yadat*, *yaroñ*, *yaleñ*, *yapoñ*:

All these words follow the female slave title, *ku*, and, as *ya* is a Mon female slave title and *y* does not elsewhere occur as the first consonant of an initial sequence, it would seem safer to separate them into two words:

ya kăy, ya, ñav, etc.

Orthography is helpful when two medial consonants (the final of the first syllable and the initial of the second syllable) can be written as a conjunct consonant sequence,

e.g. រត្ថន្ធ *ustai*, រុស្ស *us-tay*

យក្កេ *yakke*, តក្កេ *tak-ke*

វេក្ក *vekta*, វេក្កា *vek-ta*

មក្កា *makkañ*, មក្កា *mak-kañ*

All these words can be safely classed as disyllables.

The total number of disyllables thus analysed is small, but it includes three variations of the structural formula, as follows:

(1) (C)CV(C)–CV(C):

ulik, រុល្ល *u-lik* melon
vodi, វូដិ *vo-di* (vase)

mimoy, មីមូយ *mi-moy*
khalju, ក្កាល្យ *khal-ju*

each
acid-bitter?
(a slave name)

(2) CV(C)–CV(C) with reduplicated consonants, C₁VC₂–C₁VC₂:

kapkep, ក្កាបក្កេប *kap-kep*, *cakcek*, ចាកចេក *cak-cek*, *cancuñ*, ចាន់ចួន *can-cuñ*, *taktek*, តាកតេក *tak-tek* (and Angkorian *taktyak*; តាកត្មិក *tak-tiək*), *tañteñ*, តាន់តេន *tañ-teñ*, *tañton*, តាន់តូន *tañ-ton*, *taltol*, តាល់តូល *tal-tol*, *dapdip*, ដាបដិប *dap-dip*, *damdam*, ដាមដាម *dam-dam*, *lolì*, ល៉ូលី *lo-li*.

All these words are slave names of uncertain meanings. Some of them could fit into the group CNCV(C), or, in one case, Cr/lCV(C), but in every instance the nasal consonant would be the 'wrong' (i.e. non-homorganic) one and the repetition of the consonants would be an unusual coincidence.

(3) CCCV(C)–CCCV(C):

One example of this form has been noted, *gralingraloñ*, ក្ក្រិលក្ក្រិល *griiñ-grloñ* 'Gracupia nigricollis'.