

SOME CONSIDERATIONS ON THE CLASSIFICATION OF KHMER

Robert K. Headley, Jr.

The Catholic University of America

Until quite recently, in the schools of the Khmer Republic, students were taught that Khmer was related to Sanskrit and Pali. This view was evidently due to several reasons: the high proportion of words borrowed from Indo-European sources, especially Sanskrit and Pali; the use of an Indian-based script; and the tremendous cultural debt of Khmer civilization to India. All of these reasons, at least in the popular mind, obscured the fact that Khmer was really very different from and unrelated to the Indo-European languages of India. In this paper I shall discuss, first and briefly, the history of Mon-Khmer linguistic classifications, and secondly some new data on the position of Khmer within the Mon-Khmer Family.

In linguistic circles it had been supposed, at least as early as 1852,¹ that Khmer belonged to the same family of languages as Mon, Cham, several *montagnard* languages of Vietnam-Cambodia-Laos, and, possibly, Vietnamese. Wilhelm Schmidt in 1907² suggested the existence of an Austroasiatic Family of languages which included seven subgroups:

1. Mixed Group: Cham, Rade, Jarai, Seda
2. Mon-Khmer Group: Mon, Khmer, Bahnar, Stieng, Samre, Khmu, Lamet, etc.
3. Senoi (Sakai) - Semang Group
4. Palaung-Wa-Riang Group
5. Khasi Group

6. Nicobarese Group

7. Munda Group

In the same article³ he reclassified these seven groups into three larger groupings:

1. a) Semang
b) Senoi
2. a) Khasi
b) Nicobarese
c) Palaung-Wa-Riang
3. a) Mon-Khmer
b) Munda
c) Cham, Rade

Przyluski in 1924⁴ classified Austroasiatic into the main groupings: Munda, Mon-Khmer, and Annamite. Mon-Khmer was in turn divided into the following groups:

1. Central: Mon, Khmer, Bahnar, Stieng, Rengao, Moi, Kha, Kuy, Chong, Pear, Penong
2. Eastern: Cham, Jarai, Rade, Sedang
3. Malay Peninsula: Semang, Sakai, Jakun
4. Nicobarese
5. Middle Salween Basin: Palaung, Wa, Riang
6. Khasi

Schmidt in 1926⁵ made a further reclassification of Austroasiatic and proposed the following groupings:

1. Old Malacca (Semang, Senoi)
2. Central (Khasi, Nicobarese, Palaung)
3. Southeast and Northwest (including Mon-Khmer and Munda)
4. Northeast Mixed (Cham, Sedang)

Pinnow in his monumental work on Kharria⁶ divided Austroasiatic into an Eastern and Western division.

I. Western

- A. West-Northwest: Nahali
- B. Northwest: Munda

II. Eastern

C. Southeast: Mon-Khmer

a. Mon

b. Cambodian Group

1. Khmer

2. Pear

3. Chong

4. Anrak and Sauch

5. Stieng

6. Budeh and Dip

c. Chema Group

d. Mnong Group

e. Bahnar Group

f. Sedang Group

g. Brao Group

h. Jaru Group

i. Kuoy Group

j. Suoy Group

D. Northeast: Palaung-Wa

a. Western Group

1. Riang

2. Palaung

3. Angku

4. Wa

5. Danaw

6. Lawa

b. Eastern Group

1. Khmu

2. Lamet

E. Northern: Khasi

F. Southwest: Nicobar

G. Semang

H. Sakai

I. Jakun

Pinnow later⁷ modified his 1959 classification mainly by combining Mon-Khmer, Palaung-Wa, Khasi, Semang, Sakai, and Jakun into one sub-family which he called Palaung-Khmer. His 1963 provisional classification is given below:

Western Group (Nahali - Munda)

(A) West: Nahali (?)

(B) East: Munda

(a) North

(b) South

Eastern Group (Khmer - Nicobar)

(A) West: Nicobarese

(B) East: Palaung-Khmer

(a) West: Khasi

- (b) North: Palaung-Wa
- (c) East: Mon-Khmer (Mon, Khmer, Bahnar, Sre, etc.)
- (d) South: Malacca (Sakai, Jakud, Semang)

Shorto and Jacob⁸ generally follow Pinnow's 1959 classification as far as the Cambodian Group is concerned, except that they omit the Budeh-Dip language. Thomas and Headley⁹ showed that Pear, Samre, Mng, and probably Anrak were distinct enough from Mon-Khmer to warrant setting up a separate branch of Mon-Khmer, for which the term Pearic was suggested. Budeh-Dip is probably a dialect of Stieng or Mng. Thomas and Headley used lexicostatistical methods to set up a classification of Mon-Khmer which recognized the following branches:

- I. Pearic
- II. Khmer
- III. Bahnaric
 - A. North Bahnaric
 - B. West Bahnaric
 - C. South Bahnaric
- IV. Katuic
- V. Khmuic
- VI. Monic
- VII. Palaungic
- VIII. Khasi
- IX. Viet-Muong

Headley, in an unpublished paper on lexical innovations, modified the Thomas-Headley classification as follows:

- 1. Pearic Subfamily
- 2. Khmer Isolate
- 3. Eastern Mon Khmer - Mon Subfamily
 - 3.1. Bahnaric
 - 3.1.1. North Bahnaric
 - 3.1.2. West Bahnaric
 - 3.1.3. South Bahnaric

- 3.2. Katuic
 - 3.2.1. North Katuic
 - 3.2.2. South Katuic
- 3.3. Monic
- 3.4. Viet-Muong
- 4. Northern Mon-Khmer Subfamily
 - 4.1. Palaungic
 - 4.2. Khmuic
- 5. Malacca Subfamily¹⁰
- 6. Khasi Isolate
- 7. Nicobarese Subfamily

I should now like to examine certain aspects of the Khmer language to see more precisely where it fits in the Mon-Khmer family. I hope to demonstrate that Khmer is a language isolate with some fairly close ties to certain northern and eastern Mon-Khmer languages.

There are several kinds of criteria that are useful in subclassifying languages. In general, a language may be classified on the basis of phonological, morphological, and lexical criteria. I am not convinced that phonological and morphological criteria are as useful or valid, at least for the Mon-Khmer languages, as are lexical ones. Nevertheless, I shall discuss a few phonological and morphological criteria.

I shall confine myself to three phonological phenomena: (1) the devoicing of original voiced stops in initial position, (2) the voicing of the original voiceless stops *t and *p in initial position, and (3) the loss of final *r. Khmer shares the devoicing of original initial voiced stops with Mon, many of the Palaungic languages, Sedang, Pearic, and, somewhat sporadically, with Vietnamese-Muong. Some examples are given below.

<i>Proto-Khmer</i>	<i>Khmer</i>	<i>Mon</i>	<i>other</i>	<i>gloss</i>
ap	toap	tòp	Pal. t̄ap	'army'
a:	tie	tèa	Lawa: tiē, Pear: ta:, tea	'duck'
lat	toat	tàt	Vn. tát	'slap'
lro:	tro:	krò	Pal. t̄ar̄ō	'violin'
lum	cum	---	Vn. chum	'around'
ji:k	ci:k	coik		'dig'
jri:	crey	sòa	Riang: c̄erl?	'fig'
je:ŋ	cə:ŋ	càŋ	Sed. c̄əaŋ, Lawa: chuəŋ	'foot'
ji:	ci:	cì	Sed. cí	'mint'
jar	coa	---	Pal. j̄ār	'resin'
guk	kuk	kək	Vn. ngục	'prison'
grah	kreah	krèh		'comb'
p_ñ	pñ	p̄oŋ	Sed. p̄eŋ	'full'
(m)bul	pul	---	Sed. p̄ou Pear: pu(:)l	'drunk'

Khmer shares the voicing of original *t and *p with Vietnamese, and, sporadically, with Lawa. However, the presence of prenasalization in Lawa suggests that the voicing might be a secondary development. Some examples are given below.

<i>Proto-Khmer</i>	<i>Khmer</i>	<i>other</i>	<i>gloss</i>
t_h	dah		'breast'
p_n	buən	Vn. bốn	'four'
pɔ:k	ba:k	Vn. bóc	'to peel'
-tam	sdam	Lawa: ndəm	'right'
pañ	bañ	Vn. bắn ¹¹	'to shoot'
pak	bak 'broken'	Lawa: mbuək (?)	'to stab'
tap	dap 'trim'		'to stab'
pu: (?)	bav	Vn. bú; Pal. b̄ū	'to suck'
p_	bəy	Vn. ba	'three'

<i>Proto-Mon-Khmer</i>	<i>Khmer</i>	<i>other</i>	<i>gloss</i>
*tə:m	daem		'tree trunk'
*kp_n	kbən		'loincloth'

Loss of final *-r is shared by Khmer, Mon, Vietnamese, and Sedang.

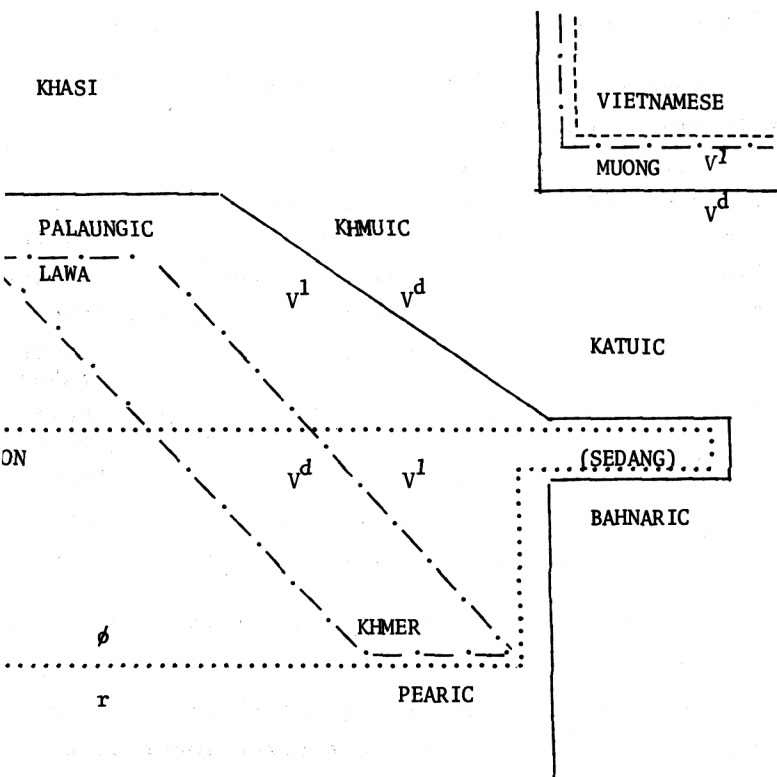
<i>Proto-Mon-Khmer</i>	<i>Khmer</i>	<i>Mon</i>	<i>other</i>	<i>gloss</i>
*jar	coa	---		'resin'
*kor	kao	---	Vn. cạo; Sed. kúa	'to shave'
*ba:r	pi:	ba	Vn. hai; Sed. p'ea	'two'
*w_r	vie	---	Vn. v'ê	'abstain taboo'
*kəmor	kmao	---	Vn. m'oi	'beetle, termite'
*kəta:r	kda:	hətp		'board'
*səg_r	sko:	---	Sed. roka	'drum'
*k_por	kambao	həpp	Vn. v'oi	'lime'

In many cases Vietnamese has the reflex -j of *-r and Muong has the reflex -l. See Map 1.

Turning to the morphology of Khmer, there are two criteria of interest: the arrangement of the classified noun phrase and the type of pronoun system. By "classified noun phrase" I mean a phrase which consists of a noun, a number word and a classifier. There appear to be two main arrangements used in Southeast Asia for the classified noun phrase. These may be represented by the formulae:

- (1) NUM + CLF + NOUN
- (2) NOUN + NUM + CLF

The chart below groups several Mon-Khmer languages according to which arrangement they use.



————— Voiced::Voiceless (voiced > voiceless)

-.-.-. Voiced::Voiceless (voiceless > voiced)

----- R :: Ø (-r > Ø)

Map.1. Distribution of Three Phonological Features.

NUM + CLF + NOUN

Vietnamese

Bru

Katu

Cua

Hre (?)

Bahnar

Mnong (?)

Chrau, Koho

Khasi¹²

Sakai

NOUN + NUM + CLF

Mon

Palaung, Wa, Danaw

Khmu

Boloven

Kuy¹³

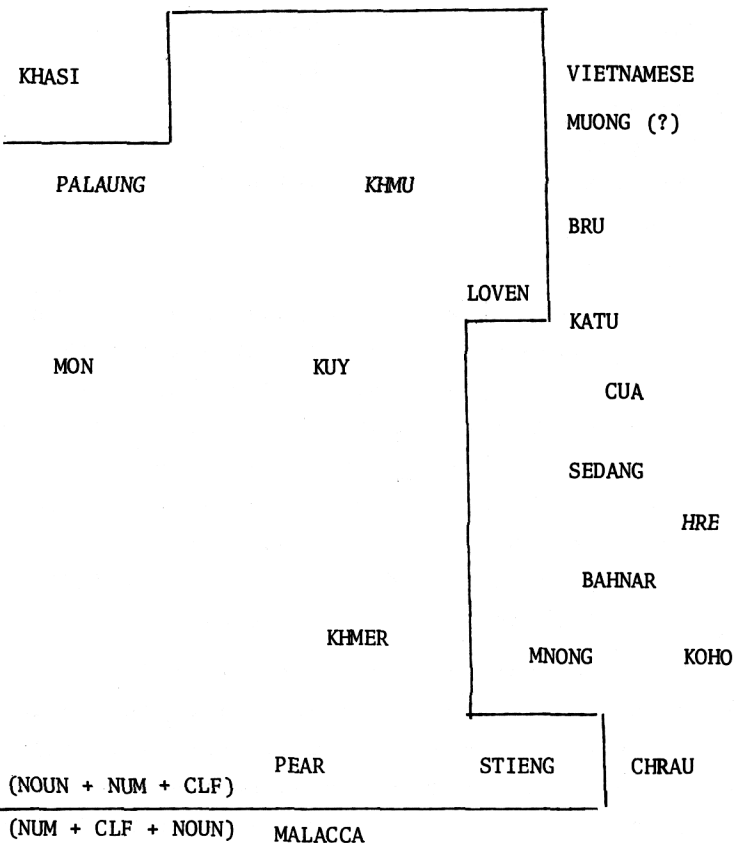
Khmer

Pear

Stieng¹³

The non-Mon-Khmer languages Lao and Malay fall into the second category--NOUN + NUM + CLF--while the Austronesian languages Jarai and Rade fall into category one--NUM + CLF + NOUN. It may be significant that Jarai and Rade are practically surrounded by non-related Mon-Khmer languages that have the same arrangement for the classified noun phrase. This feature may perhaps be easily borrowed or modified. Map 2 shows the geographical ranges of both categories.

There are four features that are of interest in classifying the pronoun systems: number (singular-plural versus singular-dual-plural systems), inclusiveness (exclusive and inclusive categories versus simple, non-marked forms in the first person dual and plural), sex (masculine and feminine), and status. Studies on the pronominal systems of Mon-Khmer languages are few, but some information can be abstracted from other linguistic studies. The following languages are those for which adequate data are available.



Map 2. *Distribution of Classified Noun Phrase Types.*

	Number		Sex m/f	Excl/ Incl	Status
	sg/pl	sg/dl/pl			
Palaung		X		X	
Riang-Lang		X			
Praok		X			
Khmu		X	X		
Pacoh		X			
Katu		X		X	
Khasi	X		X		(16)
Mon	X			X	X
Vietnamese	X			X	
Boloven	X				
Khmer	X		(15)		
Mnong	(14)		X	X	X
Chrau	(14)		X	X	
Sedang		X		X	
Bahnar		X		X	X

The geographical limits of the singular/dual/plural and exclusive/inclusive systems are shown on Map 3.

Lexically, cognate percentages, shared innovations and numeral systems are of interest. Thomas and Headley (1970) found that cognate percentages between their nine branches of Mon-Khmer averaged out at about 25%. This would appear to indicate that all of the branches separated from each other at approximately the same time. Khmer patterns here as a distinct branch on the same level as Pearic, Katuic, and the others. Cognate percentages were slightly higher between Khmer and Pearic and slightly lower between Khmer and Viet-Muong. The higher percentage between Khmer and Pearic can probably be explained by extensive borrowing by Pearic from the more prestigious Khmer.

KHASI

VIETNAMESE
(excl/incl)

PALAUNG
(excl/incl)

KHMU

RIANG

(sg/dual/pl)

(sg/pl)

LOVEN

KATU
(excl/incl)

PACOH

MON

(excl/incl)

SEDANG

BAHNAR

MNONG

CHRAU

KHMER

Map 3. Distribution of Two Pronoun Features.

In an unpublished article on lexical innovation Headley determined that Khmer had one of the highest numbers of unique innovations for any Mon-Khmer grouping. A test list of 45 items, most of which are on the Swadesh 100-word list, was used in this study. The results, given in Chart 1,* show that Pear had 15 unique forms, Khmer and Vietnamese had 13 each, and the rest had much lower figures.

The unique forms in Khmer are:

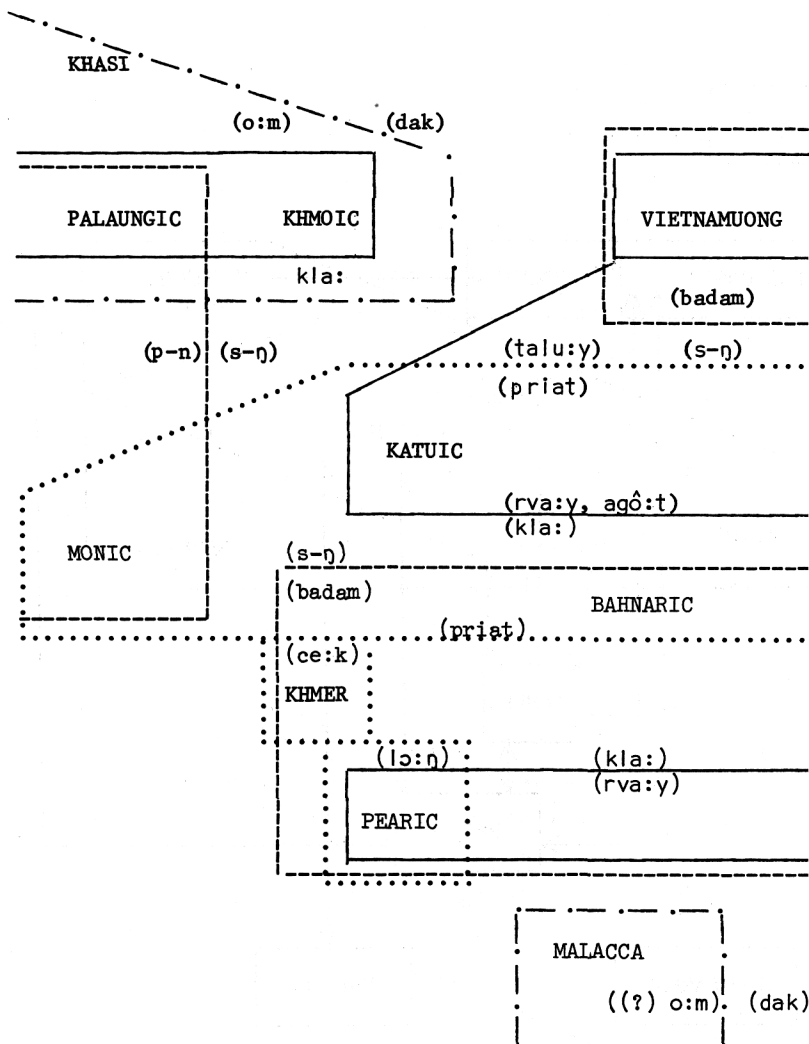
pruəñ	'arrow'	ce:k	'banana'
slap	'die'	pɔ:ŋ	'egg'
dəmɾəy	'elephant'	co:l	'enter'
pne:k	'eye'	plə:ŋ	'fire'
trəy	'fish'	pteah	'house'
pləv	'road'	pkə:y	'star'
sa:	'white'		

Khmer shares the words for 'breast,' 'buffalo,' 'dig,' 'five,' 'tiger,' and 'water' with eastern Mon-Khmer languages and the word for 'good' with northern Mon-Khmer languages. Map 4 shows the isoglosses for four key words and Map 5 shows the sub-groupings and inter-group relationships based on lexical innovation.

**Key to Chart 1:* Figures to the right of the double line indicate the number of words shared between individual languages. Thus PRC in line 1 shares 15 words only with itself; 11 words with KHM; 5 with NBC, and so on. Figures to the left of the double line indicate the average number of words shared between language groups/isolates. The following abbreviations are used: PRC=Pearic, KHM=Khmer, NBC=North Bahnaric, JH=Jeh-Halang, WBC=West Bahnaric, SBC=South Bahnaric, KAT=Katu, KMU=Khmu, PAL=Palaung, LAW=Lawa, VN=Vietnamese, MLC=Malacca, KHA=Khasi, NIC=Nicobarese, BNC=Bahnaric, KTC=Katuic, KMC=Khmuic, MNC=Monic, PLC=Palaungic.

PRC	15	11	5	7	8	9	8	11	6	6	9	2	4	9	4	3	4
KHM	11	13	7	7	9	11	6	6	1	4	8	2	4	7	7	3	5
(NBC)			2	31	21	27	11	9	4	4	14	4	5	10	7	3	5
(JH)																	
BNC	7.2	8.5		1	24	29	16	12	5	5	17	6	7	13	7	3	5
(WBC)					4	25	16	14	4	4	15	3	5	12	7	5	5
(SEC)						2	16	13	4	4	19	5	8	13	8	3	6
(KAT)							6	29	5	6	16	8	8	8	9	3	5
KTC	9.5	6			13.3												
(BRU)								5	7	9	14	8	10	6	9	3	5
(KMU)																	
KMC	6	2.5			4.2		6.7		8	13	7	14	10	8	5	4	4
(MAL)										6	7	10	8	8	6	3	5
MNC	9	8			16.2		15			7	9	9	9	11	9	3	8
(PAL)																	
PLC	3	3			5.3		8.5		10.5	9	4	4	24	8	5	4	3
(LAW)													9	5	5	4	4
VM	9	7			12		7		8	11	6.5	13	2	2	2	4	4
MLC	4	7			7.2		9		5.5	9	5	2	5	4	5	5	5
KHA	3	3			3.5		3		3.5	3	4	2	4	1	1	1	1
NIC	4	5			5.2		5		4.5	8	3.5	4	5	1	4	4	4

Chart I



- _____ 'tiger' (Katu has unique form agô:t)
 - - - - - 'five'
 - . - . - 'water'
 'banana' (Lawa has unique form ləmoih)

Map 4. Isoglosses for Four Words.

The Khmer numeral system from six to ten inclusive, is unique. The method of forming the numbers six, seven, eight, and nine is by combining the forms for one, two, three and four with the form for five. The word for ten is unique, dap.¹⁷

The implications of these criteria for sub-grouping are discussed below.

Phonology appears to be of very limited usefulness. Most of the Mon-Khmer languages share the same stock of phonemes. Indeed, even unrelated languages such as Rade, Jarai, and Thai do not differ very greatly from Mon-Khmer in their phoneme inventories. This is especially true for the consonants. Some of the phonological features, e.g., the loss of final *-r, are probably independent features not indicative of any close relationship but rather suggestive of some broader areal trend.¹⁸ If the devoicing of initial voiced stops and the voicing of original *t and *p are shared innovations and not merely general Mon-Khmer tendencies, then Khmer appears somewhat closer to the western and northern Mon-Khmer languages--especially Mon and Lawa. A tree diagram based on the above-mentioned phonological features might look like Figure 1 below.

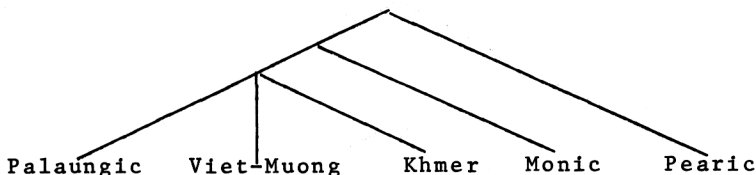


Figure 1.

Morphology does not seem to be any more useful for sub-grouping Mon-Khmer languages than phonology. In the arrangement of the classified noun phrase

er again appears to fall in with the western and northern
 -Khmer languages. The Khmer pronominal system has apparently
 ferred considerable modification from the basic Mon-Khmer
 tem, but among the traits discussed above there are enough to
 ggest a closer relationship of Khmer to Mon, Vietnamese and
 sibly Khasi than to Palaungic, Katuic, Bahnaric or Khmuic.
 the very least, it is clear that Khmer has developed a very
 ferent kind of complex pronominal system.

A genetic tree based on cognate percentages is given below
 Figure 2.

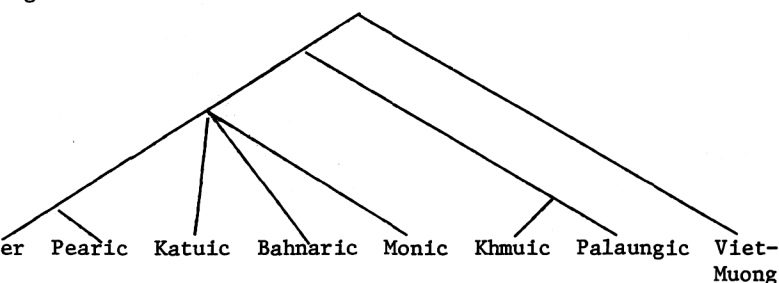


Figure 2.

this method, Khmer appears somewhat closer to Pearic and to
 clustering Katuic-Bahnaric-Monic. The genetic tree in
 ure 3 below is based on lexical innovations and agrees in
 t respects with the tree in figure 2.

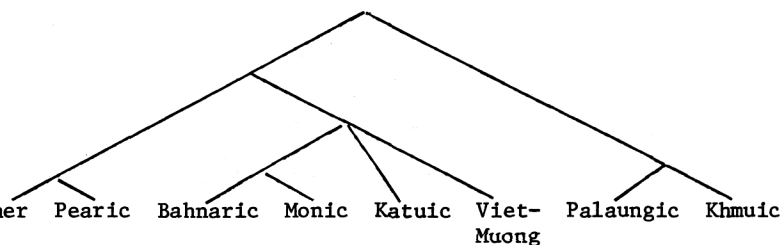


Figure 3.

The major difference in the two trees is the interchange of Viet-Muong and Palaungic.

After considering all of these criteria for the sub-grouping of the Mon-Khmer languages we are faced with several contradictions. On phonological and morphological grounds a case could be made for grouping Khmer closer to Monic, Palaungic, Viet-Muong, and Pearic. Yet, on lexicostatistical grounds Viet-Muong and Palaungic are shown to be quite distant from Khmer. I again suggest here that phonological and morphological criteria lose much of their traditional validity when one is dealing with Mon-Khmer languages. Sedang serves to point this out. It shares two of the phonological criteria, yet in all other respects it is quite different from Khmer. Also, at least two of the phonological criteria--loss of final *-r and devoicing of original initial voiced stops--took place very late in Khmer, much later, in fact, than the splits between the various Mon-Khmer languages. The present study points up the fact that lexicostatistics offers the most useful and valid methods for sub-grouping.

Primarily on lexicostatistical grounds, therefore, I suggest that Khmer stands alone as a language isolate. It has its closest ties with the Eastern Mon Khmer-Mon Subfamily. Khmer exerted a strong influence on its neighbors, especially Pearic. This accounts for the high lexicostatistical figures between Khmer and Pearic. For the future, certainly we have reason to hope that a complete reconstruction of Proto-Mon-Khmer will enable us to refine our phonological criteria; and, as more lexical information becomes available, we shall be able to determine the direction of loans. The future will bring the

le hazy picture of Mon-Khmer interrelationships
o sharper focus.

¹ Logan, J.R., "Ethnology of the Indo-Pacific
lands," in *The Journal of the Indian Archipelago
Eastern Asia*, 1852-1855, VI: 658ff., VII: 199ff.,
I: 28ff., 200ff., IX: 1ff., 162ff., 359ff., and
Journal of the Indian Archipelago, N.S., I (1856):

² Schmidt, Wilhelm, "Les peuples mon-khmêr, trait
union entre les peuples de l'Asie centrale et de
Australonésie," in *BEFEO*, VII (1907): 213-63, VIII
(1908): 1-35.

³ Schmidt, *op.cit.*, 227-8.

⁴ Przulski, J., "Les langues austroasiatiques,"
A. Meillet et M. Cohen, *Les langues du monde*
Paris: Champion, 1924), 385-403.

⁵ Schmidt, Wilhelm, *Die Sprachfamilien und
Sprachkreise der Erde* (Heidelberg: Carl Winter,
1906).

⁶ Pinnow, Heinz-Jürgen, *Versuch einer
historischen Lautlehre der Kharia-Sprache* (Wiesbaden:
Franz Steiner Verlag, 1959).

⁷ Pinnow, Heinz-Jürgen, "The Position of the
Khasi Languages within the Austro-Asiatic Language
Family," in *Linguistic Comparison in South East Asia
and the Pacific*, edited by H.L. Shorto (London:
Oxford University Press, 1963), 140-52.

⁸ *Bibliographies of Mon-Khmer and Tai Linguistics*.
Compiled by H.L. Shorto, Judith M. Jacob, and E.H.S.
Shorto (London: Oxford, 1963).

⁹ Thomas, David D. and Robert K. Headley, Jr.,
"Evidence on Mon-Khmer Subgroupings," in *Lingua*, 25.4
(1960): 398-418.

¹⁰ "After hearing the papers by Dr. Diffloth and
Dr. Benjamin at the First International Conference on
Austroasiatic Linguistics, I would now change this to
Khasi Subfamily."

¹¹ However, note Muong p'anh.

¹² Khasi actually has the somewhat divergent
formula NUM + CLF + ARTICLE + NOUN.

¹³ I suspect that Kuy and Stieng both originally
had a system like that of other Katuic and Bahnaric

languages, NUM + CLF + NOUN, but that they have been influenced very strongly by Khmer.

¹⁴Pluralizing particle used.

¹⁵With titles used as pronouns and with kinship terms.

¹⁶In second person singular.

¹⁷For further discussion of Mon-Khmer numeral systems see David Thomas, "South Bahnaric and Other Mon-Khmer Numeral Systems," paper presented at the First International Conference on Austroasiatic Linguistics, January 1973.

¹⁸See Thomas and Headley, *op.cit.*, 407.

