MON-KHMER SUBGROUPINGS IN VIETNAM

by

DAVID D. THOMAS

It has been generally assumed, and rightly so, that the majority of the indigenous languages of southern Vietnam belong to the Mon-Khmer family; but apart from an unpublished study by Phillips[1] the internal classification of them has been largely a matter of conjecture until now. Data on these languages has recently become available in the form of survey word lists taken largely by the writer, by Harvey Taylor, and by Richard Phillips, sampling most of the dialect areas.[2] On the basis of a study of these, a tentative outline of the language relationships will be attempted.

Language relationships can only be established with certainty by a study of phoneme shifts and mergers, as their imprint is indelible, while lexical and syntactic features are more easily erased. These word lists, however, are not phonemic, so cannot be used for accurate phonological study. So this study is perforce a lexico-statistical study, hence only tentative, yet it is submitted with the confidence that the main outlines of it will stand when phonological comparisons can be made.

A glance at the cognate percentages shows a clearcut clustering of the percentages, with one large group clustering around 28-34%, a smaller group clustering around 43-50%, and a still smaller group around 60%. The percentages from 22% to 37% reveal a clean split between a northern group of languages (Katu, Brôu (Bru), Pacoh, etc.) and a southern group of languages (Chrau, Bahnar, Hrê, etc.), with all of the comparisons be-

---

[1] In the report by Richard L. Phillips on a survey undertaken for the Christian and Missionary Alliance. Phillips also presented a brief oral summary of it at a 1959 meeting of the Saigon linguistic club, a meeting which the writer was unfortunately not able to attend. The conclusions presented in this present paper were arrived at independently, but they agree to a large extent with Phillips' conclusions. Because of the shortness of the lists used (av. 130 words compared), the cognate percentages are higher than if Swadesh's full 200-word list had been used.

[2] These lists are available for consultation at the University of Saigon and the University of North Dakota.

tween these two groups and only the comparisons between these two groups falling within this range, as in the following chart:

	Katu	High Katu	Phương	Pacoh	Brôu (1)	Brôu (2)
Chrau	30	31	29	23	26	27
Koho (1)	31	30	29	25	29	29
Koho (2)	28	32	28	22	26	27
Cua (1)	37	35	32	24	31	32
Cua (2)	32	30	29	26	29	30
Hrê	35	32	30	27	28	29
Sedang	31	33	29	27	31	31
Bahnar	32	34	31	27	34	34
Monom	36	34	33	29	34	34
Jeh	33	33	32	26	30	32

Considering the limitations of the method and the probable inaccuracies in the word lists, this surprisingly close clustering of percentage figures must be taken as significant evidence for a sharp break between the two language groups. The line between the two groups falls geographically between the Jeh and the Katu, i.e., an east-west line about halfway between Quảng Ngãi and Tourane. The existence of this break was suggested previously by Phillips.

This split is most strikingly demonstrated in the numerals, where the numbers 1, 2, 3, 4, 10 are apparently cognate almost unanimously. Then one set of words for 5-9 appears unanimously in the northern group, and a different set appears nearly unanimously in the southern group. This accords with my previous observation[3] that numerals tend to be among the most persistent parts of the vocabulary of a language; this might perhaps indicate that early Mon-Khmer counting was based on a system of 4, and that decimal systems were adopted subsequent to the splitting up of proto-Mon-Khmer unity. The southern set of numerals is apparently cognate with both Mon and Vietnamese (except for Mon 'five', which is the northern word); and Khmer has the southern 'five', but uses compound numerals ('five plus one', etc.) for 6-9.

Within the southern group there appears to be a further division between a northerly group (Bahnar, Sedang, Hrê, Cua, etc.) and a southerly group (Chrau, Koho, Mnong, and Stieng). Geographically these two groups are separated by the large area of Malayo-Polynesian

[3] "Basic Vocabulary in some Mon-Khmer Languages", *Anthropological Linguistics*, 2, no. 3 (1960), pp. 7-11. Also noted by Kroeber for Yokuts and Athabascan ("Semantic Contribution of Lexicostatistics", *IJAL*, 27, 1-8 (1961).

languages (Rade, Jarai, etc.), but statistically the break between the groups does not appear as sharp as that between the major groups discussed above. The percentages between these two groups cluster in the 40 's, as in the following chart:

	Cua (1)	Cua (2)	Hrê	Sedang	Bahnar	Mơnơm	Jeh
Chrau	44	43	50	44	49	47	45
Kơho (1)	43	38	48	39	46	45	47
Kơho (2)	42	39	47	39	43	43	43
Mnong Rơlơm	39	39	48	46	46	46	44
Stieng	44	44	50	44	50	45	47

Cognate percentages between languages within the southern group range from 57% to 68%. The Kơho percentages are consistently low (57-60), but it is possible that the higher Mnong, Chrau, and Stieng figures are a result of the fact that these three lists, alone among the lists used in this study, were compiled by linguists personally acquainted with the languages concerned. This may have skewed the results, though in the direction of greater reliability.

Within the northerly group Cua appears to be slightly divergent; although it shows 56-61% cognates with Hrê, it is only 50-53% cognate with the other languages in the group. Cua also shows more phonological deviation from the other members of the group, which possibly led to non-recognition of cognates. The others of the group, including Jeh, Rengao, and Halang, are 55-60% cognate with each other.[4]

In the northern major group there is similarly a binary split (also noted by Phillips), as shown by the following figures:

	Katu	High Katu	Phương (1)	Phương (2)
Pacoh (1)	40	45	50	55
Pacoh (2)	41	42	48	55
Brôu (1)	46	48	49	48
Brôu (2)	43	47	48	47

The two Phương lists gave only 75% cognate between themselves, but they are assumed to be the same language. The percentages between languages of the Katu group range from 62 to 72%. The percentages between Pacoh and Brôu range in the low 60's.

This then would suggest the following classification of the languages concerned:

[4]   As published in my "Các ngữtọc trong tỉnh Kontum", *Văn-hoa A-châu*, 1959.

I. Katuic

A. ~~Brôuan~~

1. Brôu (Bru, Kalơ, Leu, Galler, Muong Kong, Quang Tri Van Kieu)
2. Pacoh (Pko, Bô River Van Kieu; subgroups Pacoh Pahi, Pacoh Ndyong, Pacoh Ta-oih)[5]
3. Ta-oih (in Laos)

B. ~~Katuan~~

1. Katu (Teu)
2. High Katu
3. Phương (Hưu River Van Kieu, Phuang)

II. Bahnaric

A. ~~Bahnaran~~ North ~~South~~ Bahnaric

1. Bahnar (subgroups Gơlar, Alakong, Tolo, etc.)
2. Rengao
3. Sedang
4. Halang (Kyon)
5. Jeh (Dié, Strieng?)
6. Mơnơm (Bơnam)
7. Kayong (Same as Duân?)
8. Hrê (Davak)
9. Cua (Traw, Kor)

B. ~~Stiengan~~ South Bahnaric

1. Stieng (Budip, Budeh, Bulach, Bulơ)
2. Central Mnong (Pnong, Preh, Nong, Bunơr, Rơhong)
3. Biat (Mnong Biat)
4. Mnong Rơlơm (Rlâm)
5. Mnong Gar
6. Mnong Khwanh
7. Kơho (Sre, Maa, Chau-Ma, Tring, Chil, Sop, Nop, Lat, Pru, Rion, Laya)
8. Chrau (Bajieng, Ro, Mrư, Jro, Butwa', Buham, Bu-prơng, Bla, etc.)

If Pinnow's subgroupings[7] are right, then the Jelung, Halong, Hagu, Dedrah, Kemrang, and Hejung should be added to the Bahnaran subgroup; and Kasseng, Alak, Kontu, Lavé, and So should be added to the

---

[5] Findings by Phillips and by Richard Watson indicate the unity of Phương. Pacoh and Ta-oih seem to melt into each other.
[6] I am indebted to Richard Phillips and Henry Blood for the information on the Mnong languages.
[7] Heinz-Jurgen Pinnow, *Versuch einer historischen Lautlehre der Kharia-sprache* (Wiesbaden, Otto Harrassowitz, 1959). See especially the outline of the Austroasiatic family on pp. 1-6.