A Survey of Austroasiatic and Mon-Khmer Comparative Studies

DAVID D. THOMAS

O.   Introduction

The purpose of this paper is to present in general outline a short summary of the main trends and developments in comparative work on the Mon-Khmer languages, and on Austroasiatic languages in general with special reference to Mon-Khmer, with a few evaluative comments.

Only the more significant bibliographical information is given here. For full bibliographical data the reader is referred to Embree and Dotson, *Bibliography of the Peoples and Cultures of Mainland Southeast Asia* (Yale, 1950), and to Shorto, Jacob, & Simmonds, *Bibliographies of Mon-Khmer and Tai Linguistics* (Oxford, 1963).

1. General Austroasiatic Comparative Studies

Austroasiatic is a language family that has been postulated in Southeast Asia. The precise membership of this family is still a matter of some debate, with various linguists denying or affirming the membership of different groups of languages. But there is a fairly general agreement that such a family does exist, composed of Mon-Khmer, Palaungic, Munda, and some of the disputed groups.

The biggest handicap in this field is and always has been the lack of really adequate data.

The first suggestion regarding language relationships in Southeast Asia was made by Logan in 1856. This was taken up by Forbes (1881), Muller (1888) and Kuhn (1889). But the first one to attempt to put it on a sound footing was Wilhelm Schmidt, an anthropologist with some linguistic experience. He developed his theories mainly in a series of articles from about 1904 to 1907, of which the best known is 'Les peuples mon-khmer' which appeared in 1907. That article was an attempt to tie together all his linguistic, anthropological, and geographical information in a demonstration of Southeast Asian unity. He postulated an Austric superfamily composed of Austronesian (Malayopolynesian) and Austroasiatic. Schmidt divided Austroasiatic into the following groups :

I.   Mixed group : Cham, Radê, Jarai, Sedang.

II.   Mon-Khmer group : Mon, Khmer, Bahnar, Stieng, Samreh, Kha So, Kha Tampuen, Chong, Huei, Suc, Sue, Hin,

Nakhang, Mi, Khmu, Lamet, Bersisi, Jakun.

    III.   Senoi (Sakai) – Semang group.

    IV.   Palaung-Wa-Riang group.

    V.   Khasi group.

    VI.   Nicobarese group.

    VII.   Munda group. Divided into eastern and western.

And these he later regroups into :

    I.   a)   Semang

         b)   Senoi (Sakai, Tembe)

    II.   a)   Khasi

         b)   Nicobarese

         c)   Wa, Palaung, Riang

    III.   a)   Mon-Khmer

         b)   Munda

         c)   Cham, Radê, etc.

The appendices to the article give cognate sets for Nicobarese—Mon-Khmer, for Santali—Mon-Khmer, and for Austronesian-Austroasiatic. This article was so clear and specific and made such wide claims that it invited attack. And though it has been attacked from many quarters, much of it still stands.

Schmidt's intuitions may have been generally correct, and even his Austric super-family is still sometimes cautiously mentioned, but his proof was far from constituting real proof.

There are three major points at which his methodology must
be criticized :

1) He tacitly assumed that language and race are the
same thing, thus trying to use racial characteristics to prove
linguistic relationships, and vice-versa.

2) His cognate sets were randomly gathered from across the
whole Mon-Khmer area, making no attempt at determining
first whether the same sound in the various languages really
represented the same ancestral sound or whether they were
descended from different sounds. Thus a Khasi k would be
matched with a Mon k in one cognate set and with a Stieng
k in another set, with no word of justification. There seem to
be related symptoms in his *Grundzuge* (1905), which may
indicate that he was not persuaded of the generality of sound
laws, or at least that he didn't think that it would work for
Austroasiatic, and so he aimed at nothing higher than trying to
obtain a general impression of phonetic trends or general areas
of variation in Mon-Khmer.

3) He did not evaluate the reliability of his data. Thus
he was content to stake his reputation on wide hypotheses
built on partly unreliable data, and he would also mix reliable
and unreliable data haphazardly. This led him into several er-
rors of fact. His whole 'Mixed group' is fanciful, and is to
be completely abandoned, because the first three supposed
members are pure Malayo-Polynesian (cf. Pittman 1959), and
the fourth member, Sedang, is very closely related to Bahnar,
one of his prime examples of pure Mon-Khmer.

Schmidt, unfortunately, has plenty of company in other practitioners of haphazard comparative work. Relatively little of what has been done on Austroasiatic can lay much claim to being thorough, reliable, or scientific. Austroasiatic studies have consisted mainly of lists of supposed cognates to support one theory, and counter-lists of cognates to show that another theory is better, with no one making a reconstruction, and only Schmidt's *Grundzuge* making a careful comparison (before 1959).

In 1926 Schmidt added the further refinement to his classification, labelling it :

I.   Old Malacca (Semang, Senoi)

II. Central (Khasi, Nicobarese, Palaungic)

III. Southeast & Northwest (Mon-Khmer, Munda)

IV. Northeast Mixed (Cham, Sedang, etc.)

In 1924 Przyluski, in *Les Langues du Monde,* divided Austroasiatic into the three main divisions Munda, Mon-Khmer and Annamite. He divided Mon-Khmer as follows :

1) Central: Mon, Khmer, Bahnar, Stieng, Rengao, Moi, Kha, Koui, Chong, Pear, Penong.

2) Eastern: Cham, Jarai, Radê, Sedang.

3) Malay Peninsula: Semang, Sakai, Jakun.

4) Nicobarese.

5) Middle Salween Basin : Palaung, Wa, Riang.

6) Khasi.

But he added that this classification was primarily geographical and not necessarily linguistic. He recognized that the linguistic facts were not thoroughly known. He gave as the criterion of Mon-Khmer membership the presence of certain prefixes and infixes, for which he gave somewhat inadequate definitions; also, affixes similar to these 'Mon-Khmer' affixes can be found in Malayo-Polynesian languages. So this is in effect just another intuitive revising of Schmidt's intuitions.

At the same time as Schmidt, Cabaton assembled quite a bit of information on Indochinese languages which he published in 1905. This was just a collection of miscellaneous cognate sets, comparing 28 languages and giving 416 cognate sets, some of them attested in as many as 15 languages. But Cabaton had this insight, however, that forms need to be seen in the light of the total phonemic structure of the language, so he started his article by presenting the sound systems of Cham, Khmer, Malay, Bahnar, Chrau, and Stieng ; these were the languages in which there were dictionaries or larger vocabularies extant. The sound systems he presented were not phonemic, but they give some basis for evaluation of his comparisons. Cabaton also made a provisional divison of the languages into Cham-borrowing, Cambodian-borrowing, Luang-Prabang area groups, but he did not expect it to stand as a permanent linguistic subgrouping. His grouping, however shows more accuracy in the light of present knowledge than the groupings of some of the other writers on the subject.

Georges Maspero was very skeptical of the work of Schmidt, and of Schmidt's immediate predecessors. His works on Thai

(1911), Vietnamese and Thai (1920) and Khmer (1915), made his position quite plain. He roundly condemned the 'Mon-Annam' family of Logan, Forbes, Muller, and Kuhn, which included Mon, Khmer, Vietnamese, Cham, etc. Maspero granted Mon-Khmer and Palaungic as being related, but nothing else certainly related. He felt that the work of Skeat and Blagden had removed the Senoi and perhaps also the Semang from Austroasiatic ; if Grierson's report was correct that Khasi had tones, then Khasi was automatically eliminated ; and he felt Munda's relationship to Austroasiatic still unproved despite the efforts of Schmidt and Konow. Maspero's skepticism was healthy, as the 'proofs' that had been adduced previously were far from being full proofs. His denial of Khasi, however, came from his *a priori* supposition (based on his assertion of a Thai – Vietnamese relationship) that Mon-Khmer languages cannot have tones. Recent research has shown Mon-Khmer languages with two-tone systems.

Hevesy (1930) attacked the problem of the Munda languages again, tore down Schmidt's Munda—Mon-Khmer cognate sets, and proposed other cognate sets to show that the Munda languages are related to Finno-Ugric, rejecting also Schmidt's proposed 'Austric' family. But Hevesy's work was not very convincing, either, so that Sebeok (1945) had no compunctions about calling it completely unreliable and worthless. Coedes, though, (1930) cautiously proposed that Munda might be Finno-Ugric on a Mon-Khmer substratum. And Briggs (1945) in an article right beside Sebeok's, was reluctant to completely abandon Hevesy's Finno-Ugric hypothesis.

Blagden (1929) proposed the linking of Achinese to Mon-

Khmer, a theory followed further by H.K.J. Cowan in 1947.

A book by Matsumoto also appeared (1928) proposing to link Japanese with Austric, but on rather dubious grounds.

Sebeok (1942) proposed a different subdividing of Austroasiatic :

1. Mon-Khmer : Cham, Semang, Nicobarese, Palaung, Khmer, Mon, Khasi, etc.

2. Munda

3. Mường-Annam : Vietnamese, Mường
His subdivisions within Mon-Khmer were apparently based on faulty data, as they widely missed the facts.

In 1945 Briggs summarized and restated Schmidt's theories and the discussions about them, but added nothing new.

After the war the question of Vietnamese relationships was reopened. The articles in the 1952 edition of *Les Langues du Monde* on Thai and Mon-Khmer had been written by Maspero just before the war, and he, as was mentioned above, considered Vietnamese to be related to Tai rather than to Mon-Khmer.

But Haudricourt has taken sharp issue with Maspero's position, publishing a series of articles on the subject (1953, 1954, 1956). In these articles he takes Maspero's examples of Thai – Vietnamese cognates and shows most of them to be general Southeast Asian vocabulary ; and then he takes Maspero's tonal argument and shows correspondences between Vietnamese

tones and Mon-Khmer final consonants. He also shows that the tones in Thai and Chinese arose by somewhat similar processes, thus making tonality an areal trend. And indeed there seems to be a trend going on right now toward tonality in some of the non-tonal languages of South Vietnam. Haudricourt pictures the development of Vietnamese in the following stages :

1)   An initial stage with no tones.

2)   Final consonants *-s*, *-s̃*, *-h* replaced by *hỏi-ngã* tone, final glottal stop replaced by *sắc-nạng* tone, and all others having a mid tone *bằng-huyền*.

3)   These three tones split, conditioned by the voicing of the initial consonant, with *hỏi*, *sắc*, and *bằng* appearing after voiceless consonants, and *ngã*, *nặng*, and *huyền* after voiced consonants.

4)   Many voiceless initial consonants (with *hỏi*, *sắc*, or *bằng*) became voiced, making the 6 tones fully phonemic.

The discussion of the 3rd and 4th stages was presented by Haudricourt and Martinet (1946) as an areal trend.

These articles by Haudricourt are a big step in the right direction in that they are fairly careful analyses of small points rather than sweeping generalizations. But Haudricourt gives too few examples to be really able to carry the weight of the extent he would have them cover, though the argumentation sounds reasonable. Also his handling was not strictly phonemic, as phonemically his steps 3 and 4 in the outline of Vietnamese tonal development would seem to be phonemically one step, with

step 3 being an allophonic distinction which became phonemic when the conditioning environment disappeared in step 4.

Shafer (1953) undertook a preliminary reconstruction of Palaungic (Palaung, Wa, Riang, etc.) and also presented a number of presumed Khasi—Mon-Khmer cognates. The cognate list is just a random citing of languages. But the reconstruction work on Palaungic is valuable, while covering only a small amount of data. Shafer (1960) now says that he has completed a comparative study of the Austroasiatic languages, which has not yet been published.

H. J. Pinnow's *Versuch einer Lautlehre der Kharia-Sprache* (1959) is a milestone in Austroasiatic comparative studies. He gives a detailed study of Kharia, one of the Munda languages, then carefully reconstructs Proto-Munda, pointing out similarities with other Austroasiatic languages in passing, and gives a few cautious suggestions toward possible Proto-Austroasiatic reconstructions. The detailed thorough scholarship of this work is the first, I trust, of other such scholarly comparative works to come in the field of Austroasiatic. Pinnow also attempts an outline listing of the whole Austroasiatic family.

A volume of Austroasiatic comparative studies (Norman Zide, editor) is now in press, which we hope will significantly advance scholarly Austroasiatic study.

2. Mon-Khmer Comparative Studies

The term 'Mon-Khmer' has been used in different ways.

Some writers use it in a very limited sense, and others make it almost synonymous with Austroasiatic. Since the whole problem of Austroasiatic relationships has mostly been a matter of intuitive conjecture up to the present, I will use 'Mon-Khmer' in the sense that Schmidt used it in his *Grundzuge* and came close to establishing scientifically — i. e. the four languages Mon (Talaing, Peguan), Khmer (Cambodian), Bahnar, and Stieng, and any other languages which would fall within the degree of relationship bounded by this grouping. There is some question whether the Palaungic languages in Burma and the languages like Khmu in northern Laos fall within this group, but in the absence of any reconstruction of Mon-Khmer it is difficult to make any sure statements about the degree of relationship of these languages to Mon-Khmer; so for the purposes of this paper they will not be considered Mon-Khmer.

There has been relatively little comparative work done within Mon-Khmer. The oldest and still most complete work is Schmidt's *Grundzuge einer Lautlehre der Mon-Khmer Sprachen* (1905), in which he assembled and compared cognate sets from Mon, Khmer, Bahnar, and Stieng. The work is divided into three parts: 1) An introduction in which he traces the history of previous Mon-Khmer studies, gives as his aim the putting of Mon-Khmer studies on a solid basis, and explains and justifies his transcriptions of Mon and Khmer. 2) A listing of the sets of presumed cognates from Mon, Khmer, Bahnar, and Stieng. He gives perhaps 1000 cognate sets from two or more of the four languages, but at least one in each set had to be Mon or Khmer, as his Bahnar and Stieng data was less

trustworthy. 3) A discussion of the correspondences in terms of initial consonants, final consonants, and vowels. (This distinction between initial and final consonants is important in Mon-Khmer languages.)

Schmidt's work was only partially successful, for several reasons. His Mon and Khmer data used a straight transliteration from the non-phonemic Mon and Khmer scripts. This led to a few anomalies in the consonants (e. g. a final -b in Khmer); and in the vowels it definitely led him astray, though the degree of inaccuracy is debated — Maspero felt it to be quite inaccurate, Pinnow (1957) seems to feel it valid in the main, and Jacob (1961) would seem to indicate that it is partly valid and partly invalid. In a number of words Khmer orthography varies freely between o and a, which could very possibly indicate an $ɔ$ phoneme which the borrowed Devanagari script could not distinguish. So this makes all his statements about o and a suspect. This is particularly unsettling when a is by far the most frequent vowel, being found in at least 1/3 of the words he cites.

Another difficulty was that Schmidt's Bahnar and Stieng sources were not completely reliable, especially the latter, though he used the best that was available at the time.

Still another difficulty, one that besets all Mon-Khmer comparativists, is the complexity of the vowel shifting that has taken place in Mon-Khmer, making it very difficult to establish regular patterns. Schmidt, after a careful survey of the situation, had to content himself with just statements about possible general trends,

establishing no sound-laws. Other comparativists have stated flatly that regular sound-laws simply do not exist in Mon-Khmer vowels, and, indeed, no one has yet succeeded (in print, anyway) in establishing a regular pattern in Mon-Khmer vowel comparisons. I feel that the solution, however, lies in starting at the very lowest level of comparison, working on adjacent languages to establish proto-forms at that level, then using these reconstructions as the basis for comparison on the next level. Only in this way, I feel, will the Mon-Khmer vowels be able to be solved.

So in general Schmidt's work on initial consonants can be considered sound, the final consonants can be considered valid at most points, and the vowels are as he left them — chaotic.

After Schmidt's *Grundzuge* for a long time no further work was done in the field of Mon-Khmer comparison. But in the last few years interest has revived again.

Reynaud's *Etude sur les phonemes Vietnamiens* (1962) devotes quite a little discussion to Bahnar, Jarai, and Sedang, attempting to form sound-laws within them and between them and Vietnamese. Jarai, however, is Malayo-Polynesian, not Mon-Khmer, so his arguments lose a lot of their force.

Mlle. Piat's comparison of Bru and Khmer (1962) found many regularities between Bru (Brôu) and Khmer consonants, but, like Schmidt, she found the vowels not amenable to predictable rules. Brôu is linguistically quite some distance from Khmer, so this result is not surprising. This also was done without the benefit of a phonemic analysis of Brôu.

In 1961 Richard Phillips made a survey of the Mon-Khmer languages of Vietnam (unpublished), which I have followed up with a short lexicostatistical study to be published soon. Further studies by other linguists are in progress, which should help to settle some of the question regarding Mon-Khmer relationships and reconstructions.

A phonemic base is necessary for adequate comparative work. The lack of careful phonemic (or prosodic) descriptions has been the main hindrance to progress in Mon-Khmer and Austroasiatic comparisons. But fortunately this lack is now being rapidly filled from several quarters, so the next few years should see rapid progress and hopefully the solution especially of the vowel problem.

3. References

Selected important references, and recent references not in the pubished bibliographies.

Haudricourt, A. G., 'La place du vietnamien dans les langues austro-asiatiques', *B. S. L. P.* 49 : 122-128 (1953).

Haudricourt, A. G., ' De l'origine des tons en vietnamien'. *J. Asiatique* 242 : 69-82 (1954).

Piat, Martine, 'Quelques correspondances entre le Khmer et le Bru, langue montagnarde du Centre-Vietnam', *B. Soc. Etudes Indoch.* 37 : 311-323 (1962).

Pinnow, H. J., *Versuch einer Lautlehre der Kharia-Sprache.* Wiesbaden, 1959.

Pittman, Richard S., 'Jarai as a Member of the Malayo-Poly-nesian Family of Languages', *Asian Culture* 1: 59-67 (1959).

Reynaud, le Docteur, 'Etude des phonemes vietnamiens, par confrontation entre le vietnamien et quelques dialectes des Hauts-Plateaux du Sud-Vietnam', *B. Soc. Etudes Indoch.* 37: 117-253 (1962).

Schmidt, Willhelm, 'Grundzuge einer Lautlehre der Mon-Khmer Sprachen', *Denksch. kaiserl. Akad. Wiss. Wien* (Phil-hist. Kl.) Vol. 51, No. 3 (1905).

Schmidt, Wilhelm, 'Les peuples Mon-khmers', *B. E. F. E. O.* 7: 213-263 ; 8 : 1-35 (1907-8).

Sebeok, Thomas, 'An examination of the Austro-Asiatic language family', *Language* 18: 206-217 (1942).

Shafer, Robert. 'Etudes sur l'austroasien', *B. S. L. P.* 48: 111-158 (1952).

Zide, Norman, editor, *Studies in Comparative Austroasia-tic Linguistics.* To appear in 1964 (Mouton).