Vietnamese in Mon-Khmer Perspective

William W. Gage

Center for Applied Linguistics

Establishing the Parentage of Vietnamese.

Understanding of the historical development of the Vietnamese language was profoundly affected by three articles by André-Georges Haudricourt published in 1953-1955:

La place du vietnamien dans les langues australasiatiques;
De l'origine des tons en vietnamien;
Comment reconstruire le chinois archaïque.

The last in particular illustrated the way in which vocabulary resemblances between Tai languages and Vietnamese, noted by Maspéro (1912), could be due to separate early borrowings from Chinese (antedating the stage of development of Chinese reflected in the standard Vietnamese readings of Chinese characters known as Sino-Vietnamese).

These investigations tended to defuse arguments against regarding Vietnamese as having originally been a Mon-Khmer language. Statistical evidence has subsequently been brought to bear to strengthen this position, as vocabularies from a large number of Mon-Khmer languages have become available. Word lists showing striking resemblances between Vietnamese and its near neighbors in the southern highlands of Vietnam appeared in David Thomas' article
on "Mon-Khmer Subgroups in Vietnam" (1966). Specific estimates of percentages of related words on such lists were given in "More on Mon-Khmer Subgroupings" (Thomas & Headley, 1970) and other counts by Huffman (1976). When one takes standardized lists that have the appearance of representing basic vocabulary, one can count perhaps twenty percent of items for which the Vietnamese word and the word in a given Mon-Khmer language appear to be related. No such degree of resemblance can be found in the vocabulary of a Tai language: (I have tried comparing a Lao and a Vietnamese list and could find only 7% of plausibly relatable entries.) In the light of such figures, it is hard not to conclude that Vietnamese bears a close relationship to the Mon-Khmer family.

Earlier Discussion.

The quest for an affiliation for the Vietnamese language has a long and variegated history. Early investigators, finding it written with Chinese characters and pervaded with Chinese vocabulary generally concluded that it was related to Chinese. So Bishop Taberd, introducing the Pigneaux-Taberd Dictionarium, declared (1838, i): "Dubium non est quin Anamitarum lingua à Sinensi ortum ducat." [There is no doubt but what the language of the Annamites derives its origin from the Chinese.]
The first to connect Vietnamese with the Mon-Khmer languages was James R. Logan, who propounded his views of "The Mon-Anam Formation" in 1852. In my opinion, we generally give Logan too much credit for his insights. His "formations" were not language families as we have been understanding that term for the last hundred years—he was a pretty thorough-going wave theorist avant la lettre. Many of the linkages he proposed were spurious. And as far as distinguishing between a Tai connection and a Khmer connection for Vietnamese, it is to be recalled that for Logan "Lau" was one division of the "Mon-Anam Formation". [Many of his spellings merit a sic after them by today's conventions.]

For Logan the Munda family of languages in India was part of "Dravirian", although similarities between Munda and Mon-Khmer languages were also noted. These soon came to be given a more prominent place in language comparison. In 1856 Schott suggested a specific relationship between Vietnamese and Mundari—on the basis of similarities in numerals and pronouns only.

It was in 1889 that Ernst Kuhn presented a description of Mon-Khmer linguistic relationships that is consistent with recent research. In fact, his schema of divisions within the family is almost isomorphic to that given by Headley (1976a, 434-5). Since Headley's is based on lexical innovations and Kuhn judged by lexical clusterings, there ought to be convergence, even though Kuhn had extremely meager data to work with.
Kuhn's outline differs from Headley's in only two significant particulars. One of these is easy to explain. Headley postulates a subfamily with four components: Bahnaric, Katuic, Monic, and Viet-Muong. Kuhn grouped the other three together, but separated what we now call Katuic as a main division of the "so-called Mon-Annam languages" (191). Clearly this arrangement was prompted by a reliance on numerals as highly diagnostic of language relationships—a prevalent view at that time, following, no doubt, from its success with Indo-European. The Katuic languages have in common a uniquely innovating set of numbers from six through nine. The other mismatch is Kuhn's joining of Khmer and Pearic. He even refers to Pearic languages as dialects of Khmer (194). This discrepancy is harder to understand. Perhaps it reflects the bias of his primary sources (e.g. Crawfurd, 1828). The lack in Khmer of any unitary words for the numbers six through nine that could be compared with other languages might also have contributed.

Kuhn also maintained the relationship of Mon-Khmer to Munda, although he considered it premature to conclude that this was a sure genetic connection—into what Schmidt (1906) christened the Austroasiatic family. This more remote relationship, linking Mon-Khmer including Vietnamese with Munda (and Nahali?) is beyond the scope of the present paper.
In general Kuhn's rather clear ideas have had less subsequent influence than one might have hoped. The second half of his 1889 article is devoted to demonstrating the Malayo-Polynesian affinity of Cham. He also specifically notes (193) that Sedang is close to Bahnar. Yet, as late as 1924, we find Przyluski postulating an Eastern subgroup of Mon-Khmer including Sedang and three Malayo-Polynesian languages: "Cham, Jarai, Radé, Sedang".

Still, in spite of agreements in basic vocabulary, there has been a marked reluctance to consider Vietnamese as a Mon-Khmer language.

Kuhn himself, at the end of his discussion, rather backs away from his placement of Vietnamese, remarking on the relatively scant representation of that language in his eighty-five vocabulary sets apart from the numerals (219). There are, in fact, nine more instances where a Vietnamese term could be added to the comparison; we can only speculate on whether it would have in any way allayed his doubts if he had had that information. He also alludes to the belief of some historians that the Vietnamese had entered Việt-nam from China at a relatively late date.

The main stumbling block for acceptance of a genetic tie between Vietnamese and Mon-Khmer has always been the great typological gulf that separates them. These differences were all adduced by Henri Maspéro (1912, 115-6), although for him the conclusive point was that
the tones in Vietnamese could be derived from proto-Tai, while he saw no possibility of tones arising in a Mon-Khmer language. This argument was not really dealt with until Haudricourt's 1954 article on origine des tones. As Haudricourt has pointed out (1974, 6), Maspéro was the first to take the position that the tones in Vietnamese words demanded a serious historical explanation. Most others merely disregarded them.

Maspéro's views exercised a considerable influence in the field, apparently leading Schmidt, who in 1905 had classified Vietnamese as a Mon-Khmer language, to align it with Tai languages in his 1926 work. Those views also seem to lie behind the declaration of Pinnow much more recently (1963, 141) that: "we may unhesitatingly regard Vietnamese and Müöng as Thai languages". He included Vietnamese forms in his comparative studies (1959; 1965), but considered Vietnamese a Tai language with extensive Mon-Khmer borrowings paralleling the case of Cham, a Malayo-Polynesian language with a great deal of Mon-Khmer vocabulary--most recently discussed by Headley (1966b).

Idiosyncrasies of Vietnamese.

If it were not a thorny question, the lineage of Vietnamese would not have been for so long a bone of contention. Following present-day theory that a preponderance of basic vocabulary must indicate the origin of a language, we assign Vietnamese to the
Mon-Khmer stock. As, in recent years, more lexical information about many Mon-Khmer languages has become available, this relationship seems increasingly secure.

It is, however, clear that Vietnamese has, as it were, undergone a mutation that strikingly differentiates it from its relatives.

These drastic changes were brought about primarily by the influence of Chinese. For nearly a millenium Vietnam was ruled from China, and in the subsequent millenium it has constantly been subject to Chinese cultural influences. It was indubitably in imitation of Chinese that Vietnamese words came to consist normally of a single syllable, whereas the typical Mon-Khmer pattern is a major syllable preceded by a weak one. The structure of syllables tended to be limited in the direction of what was possible in Chinese at any given time. The use of prefixation and infixation as grammatical processes disappeared. The cultural vocabulary became almost entirely Chinese, and the basic vocabulary, too, was heavily influenced.

The way in which Vietnamese has been moved away from its Mon-Khmer heritage is interestingly illustrated in studies on similarities in the basic parts of the lexicon. In Huffman's research reported in 1976, he counted apparent cognition between each pairing of eighteen languages on both a 500-word list and a 100-word list. With each of the other seventeen languages, Vietnamese
scored lowest in comparisons made on the 500-word list. Yet, with the more basic 100-word list, it was often not at the bottom of the cognate ranking. As many as five other languages placed below it—the figure for the Chaobon list (567). On the average three percentage points—that is, on a 100-word list, three more apparently cognate words—separated Vietnamese from the language showing fewest cognates with each given language. On the 500-word list it averaged eleven fewer apparent cognates than whatever language made the next worst showing. Thus is the general impression confirmed that when we consider very basic words any Mon-Khmer language shows a considerable similarity to Vietnamese, but, as we move away from these, further cognates become hard to find. Like the typological changes, this too must most probably be laid to the influence of Chinese, although no investigations to prove that contention in detail have yet been conducted.

The genetic connection of Vietnamese to Mon-Khmer gives us an unusual language family, in which the largest language is a maverick. With speakers now numbering nearly 50,000,000, Vietnamese accounts for around four-fifths of the total for the combined grouping. The two languages with long written traditions, Khmer and Mon, now have at most, respectively, 7,000,000 and 700,000 users. The fourth—though recent—literary language, Khasi, may be spoken by 500,000. All the other Mon-Khmer populations have been considered "tribal". The largest among the perhaps seventy
other languages is probably Kuy, which may number over 250,000 in Thailand and Cambodia. Figures at this point and beyond are, however, mostly in the realm of wild guesses.

The Lexical Core.

The most convenient place to see how comparable a good deal of Vietnamese basic vocabulary is to that of other Mon-Khmer languages is in the lists in Gordon H. Luce's discussion of Danaw (1965). While the Vietnamese items in the "Further Comparisons" to his 245-word list were, of course, not guaranteed as related but only offered as interesting similarities, still a large number of significant cognates do appear in these pages.¹ Huffman (1977) assembled 99 Vietnamese items with suggested cognates in 18 other Mon-Khmer languages. Even though I have grave doubts about twenty or so of these sets,² as a whole they also provide a coherent look at Vietnamese relationships to Mon-Khmer lexical items. (Thirty-five entries that are also glosses from Luce's list provide additional breadth of coverage.) There are more Vietnamese words in etymological comparisons in Shorto's Dictionary of the Mon Inscriptions (1971), but it is cumbersome to derive an overview of their significance by reading through a dictionary.
Vietnamese Etymologies.

As a demonstration of the fit of Vietnamese with the sound system of Mon-Khmer, a fairly detailed listing will be given of Vietnamese developments from the parent language. Only a single Mon-Khmer cognate is give for any but a very few. Most examples can be illustrated from other languages and could be strengthened by parallel examples, although a few are unique in available data. Since Written Khmer has an archaizing spelling that often helps to clarify historical developments, and since extensive lexicographic resources are available for it, that language has been preferred for citations. (Of the many transcriptions of Cambodian that are found in the literature, that used here approximates the practice of the School of Oriental and African Studies in London.) Other languages are utilized mostly for examples whose connections to Vietnamese are easy to perceive.

A. Initial Consonants.

Most Mon-Khmer correspondences can be explained as reflexes of a fairly simple consonant system:

\[
\begin{array}{ccccccccc}
\text{k} & \text{c} & \text{t} & \text{p} & \text{?} \\
\text{g} & \text{z} & \text{d} & \text{b} & \\
\text{s} & \text{h} & \\
\text{q} & \text{n} & \text{n} & \text{m} & \\
\text{y} & \text{r} & \text{l} & \text{w} & \\
\end{array}
\]

The Mon-Khmer sources of Vietnamese initial consonants were discussed in Maspéro's 1912 monograph. As he established, originally voiced onsets are reflected by the Vietnamese tones huyên, năng, and ngã.

<table>
<thead>
<tr>
<th>Việt</th>
<th>PMK</th>
</tr>
</thead>
<tbody>
<tr>
<td>/</td>
<td>*</td>
</tr>
<tr>
<td>Việt</td>
<td>Gloss</td>
</tr>
<tr>
<td>/ /</td>
<td>k</td>
</tr>
<tr>
<td></td>
<td>&lt; k</td>
</tr>
<tr>
<td></td>
<td>&lt; g</td>
</tr>
<tr>
<td></td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>&lt; g</td>
</tr>
<tr>
<td></td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>&lt; d</td>
</tr>
<tr>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>&lt; b</td>
</tr>
<tr>
<td></td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>h</td>
</tr>
<tr>
<td>[?]</td>
<td>&lt; ?</td>
</tr>
</tbody>
</table>

In the next group of examples we have instances of the developments referred to as "spirantisation" by Ferlus (1975; 1976) and as "softening" by Pulleyblank (1981). It seems clear that these sounds derive from consonants preceded by certain prefixes.
at an early stage of Vietnamese. This explanation was first suggested by Haudricourt (1965, 171), who noted a preceding r in a number of Mon-Khmer cognates of such words. It now seems unlikely that r can be posited in all instances. Neither, apparently, did all prefixes have this effect, since many Vietnamese words with unaffected initials show prefixes in their apparent cognates throughout Mon-Khmer. Until the general history of Mon-Khmer prefixing has been to some degree unraveled, the nature of the pre-Vietnamese forms must remain obscure. Here they will be noted only by putting hyphens around the proto-consonants.

\[
\begin{array}{llll}
\gamma & < -k- & g\grave{a}u & \text{a bear} & \text{Sedang} & \text{rok} \grave{u} \\
 & < -g- & g\ddot{o} & \text{knock} & \text{WrKhmer} & \text{go}h \\
y & (gi) & ky & g\ddot{i}o & \text{wind} & \text{Old Mon} & \text{ky} \ddot{a}l \\
 & < -c- & g\ddot{i}t & \text{kill} & \text{Old Mon} & \text{kucit} \\
? & < -\ddot{g}- & g\ddot{i}a & \text{old} & \text{Old Mon} & pju \text{ mature} \\
(d) & y & d\ddot{a}i & \text{long} & \text{Katu} & \text{yaal} \\
 & < -t- & d\ddot{a} & \text{skin} & \text{Chrau} & nt\ddot{o} \\
 & < -d- & d\ddot{\grave{c}} & \text{monkey} & \text{Bahnar} & \text{?d\ddot{\grave{c}}k} \\
v & (v) & w & v\ddot{e} & \text{return} & \text{WrKhmer} & v\ddot{i}l \\
 & < hw & v\ddot{\acute{t}} & \text{hinder} & \text{WrKhmer} & sv\ddot{\acute{t}} \text{ tough} \\
(b) & -p- & vu\ddot{\acute{t}} & \text{rub} & \text{Old Mon} & s\ddot{u}mpot \\
 & < -b- & v\ddot{\grave{e}} & \text{flap(wings)} & \text{Rõngao} & \text{båh}
\end{array}
\]
Modern  that is the result of "softening" was generally written
with a special letter  in Alexandre de Rhodes' materials of
around 1650. At a late stage in pre-Vietnamese it is assumed
that numerous voiceless onsets before resonants merged, probably
to give voiceless resonants, indicated by *hw above and *hr &c.
in the following groups. The proto-consonant most frequently
acting this way was probably *s.

\[ r < *r \quad rə \quad \text{root} \quad \text{WrKhmer} \quad rs \n\]
\[ < *hr \quad rən̩ \quad \text{tooth} \quad \text{Palaung} \quad hrąŋ^2 \n\]
\[ < *-s- \quad rən \quad \text{snake} \quad \text{WrKhmer} \quad msəŋ \n\]

Among the nasal-consonant etymologies, instances such as
*ŋŋ exemplify what James A. Matisoff has often referred to for
Tibeto-Burman languages as "nasal preemption".

\[ η < *ŋ \quad nəy \quad \text{day} \quad \text{WrKhmer} \quad thŋai \n\]
\[ < *ŋŋ \quad ngai \quad \text{far} \quad \text{WrKhmer} \quad chŋąy \n\]
\[ < *ŋŋ \quad ngi \quad \text{sit} \quad \text{WrKhmer} \quad ?oŋguy \n\]
\[ n < *ŋ \quad nəhə \quad \text{house} \quad \text{Lawa} \quad ŋIeq \n\]
\[ < *ŋŋ \quad nho \quad \text{remember} \quad \text{Bru} \quad saŋdə \n\]
\[ < *ŋŋ \quad nho \quad \text{small} \quad \text{Rôngao} \quad ąyoh \quad \text{young} \n\]
\[ < *ŋc \quad nhaŋ \quad \text{ring} \quad \text{WrKhmer} \quad ?qncian \n\]
\[ < *ŋŋ \quad nhe \quad \text{light in weight} \quad \text{Khmu? (T'èng)} \quad əjal \n\]
\[ n < *n \quad nəp \quad \text{bend over} \quad \text{Rôngao} \quad ąap \n\]
\[ < *hn \quad nəm \quad \text{year} \quad \text{WrKhmer} \quad chnəm \n\]
\[ < *ŋd \quad nə̀c \quad \text{water} \quad \text{Old Mon} \quad ək \n\]
\[ < *nt \quad nəŋ \quad \text{heavy} \quad \text{Souei} \quad nʊəŋ \n\]
Examples of nasal preemption can be adduced for *ŋɡ, *nt, *mp, but not for *ŋk, *nd, *mb. This fact may be a result merely of insufficient examination of the data.

Vietnamese s /s/ is the normal reflex of any initial consonant followed by *r when the cluster as a unit was ancestral to the Vietnamese form. (Sometimes widely attested initials disappeared, joining the correspondences for *hr or *r:

réo shout WrKhmer krev
rú wooded Bru brōu
rúng jungle Khasi kyreng forest.)

s < *kr sau after WrKhmer krau
< *gr sām reverbrating crash WrKhmer gram
< *cr sa fall(rain) Middle Mon croyh
< *gr sau deep WrKhmer jrau
< *dr sūŋ horn Old Mon draŋ tusk
< *pr sōc squirrel WrKhmer kəmpruk
< *br sōi thread Stieng bray
< *sr sāp arrange WrKhmer srap
In the Vietnamese of Alexandre de Rhodes around 1650 there were two apparent clusters, bl and tl, the former deriving from clusters of a labial with *l, the latter from most other consonants when followed by *l. In modern Southern Vietnamese both yielded tr /t/. In Northern Vietnamese the successor to bl is usually gi /z/, while that to tl is tr /č/.

\[
\begin{align*}
\text{t} (\underline{bl}) & \text{< *pl trä́i fruit WrKhmer phlè} \\
& \text{< *bl tr̀mì sky,heaven Khasi Blei God} \\
& \text{< *ml tr̀âu betel WrKhmer mlūv} \\
\text{t} (\underline{tl}) & \text{< *kl tram 100 Old Mon klam} \\
& \text{? *gl tre bamboo Chrau gle} \\
& \text{< *cl tr̀âu carabao WrKhmer chlūv year of the ox} \\
& \text{< *ʒl tràn invade WrKhmer jhlān aggressive} \\
& \text{< *tl trăn python WrKhmer thlan} \\
& \text{< *sl tré́o crossed WrKhmer slēv}
\end{align*}
\]

In de Rhodes' material there is also a cluster ml. In modern Vietnamese reflexes of his ml words appear with initial l or nh or as doublets having both spellings. Not much comparative evidence has been reported that bears on the earlier state of these words. In a 1966 article Henderson offered some suggestions of Khasi relationships.

Two of her examples seem compelling:

\[
\begin{align*}
\text{ml̀n} & \text{big,great Khasi mlan mlen robust,well built} \\
\text{mlàt} & \text{insipid Khasi blad}
\end{align*}
\]
On the basis of Khasi alone we cannot distinguish between the sources of ml and bl. The proto-language must in fact have had some onsets considerably more complicated than we can now account for. Even with better evidence it will still be risky to reconstruct these, since different languages often show different prefixes.

The only sequences that seem to have continued as clusters in modern Vietnamese are occasional examples of *kw and *gw.

\[
\begin{align*}
\text{kw} & \leftrightarrow *kw \quad \text{quay} \quad \text{roast on a spit} \quad \text{WrKhmer} \quad \text{khvai} \\
& \leftrightarrow *gw \quad \text{quàng} \quad \text{surround} \quad \text{WrKhmer} \quad \text{ghvăng deviate}
\end{align*}
\]

For Vietnamese kh, th, ph, and x, examples are scarce, and it is difficult to pin down the Mon-Khmer sources. At the time when a scholarly pronunciation for Chinese (called Sino-Vietnamese) was codified in newly independent Vietnam after AD 939, Vietnamese apparently had a series of aspirated stops: *k' (\(>\) kh /x/), *c' (\(>\) x /s/), *t' (\(>\) th /t'/), *p' (\(>\) ph /f/). Certain Mon-Khmer languages have added aspiration to their historical voiceless stops and devoiced their historical voiced stops (cf. Haudricourt, 1965), but otherwise the possible etymological source for such a sound-type appears to have generally been a cluster of a stop followed by *h. Such sequences seem not to have been very common. (The aspiration shown in many of the Written Khmer examples in this study for the prior members of clusters is a relatively late phenomenon.)
x < *kh khá feel better Old Mon khis in good shape
< *hk kháp join Bahnar hòkòp
s < *ch xéo slanting WrKhmer chev crooked
< *c? xương bone WrKhmer ch?mŋ
°t' < *th thời blow Bru thòr
° *s? thời rotten WrKhmer s?uy
f < *hp phui dust off Rôngao hòpuih sweep

The patterns of Sino-Vietnamese lead us to believe that Vietnamese once had /ʃ/ which became /t'/. If the rather scant examples of putative *s? > th are genuine, it seems likely that they were /ʃ/ rather than /t'/ before the merger.

A major study by Perlus (1978) dwells largely on the relationships of Vietnamese x. He posits as its main source */TS/. The sound sequence that in a living language might be interpreted as /tš/ might equally well be /c'/ or /ch/ in another, so his presentation differs from that implied in the table above primarily in how it views the sound-pattern of the proto-language. At the level of Proto-Mon-Khmer he believes there were two entities **/TS/ and **/TS/, with **/TS/ giving /h/ in Khmer, Bahnaric, and Katuic (summary on his p.26).
He further tentatively proposes three more Proto-Mon-Khmer phonemes: */š/, */ʃ/, and */q/. The etymologies offered for these last three suggested proto-phonemes and those that support the distinction between **/TS/ and **/TS/ involve a good many problems—not the least of which is the inclusion of some items that are probably Chinese borrowings. The present discussion seems no place to go into the technicalities of
the evidence, but I am unconvinced. (As noted above, I do believe that Vietnamese at one time had /ʃ/, the equivalent of Perlus' /ʃ/, but I am reluctant to see this as a unique preservation from Proto-Austroasiatic.) Certain other proposals for the existence of additional Proto-Mon-Khmer initials have been put in doubt in a study by Diffloth (1977).

B. Final Consonants.

For the most part, the word-final correspondences are more straightforward than those for initial consonants, except as some recent changes in Vietnamese may have clouded the picture.

\[
\begin{align*}
  k & < *k \text{ tóc} \quad \text{hair} \quad \text{WrKhmer} \quad \text{sok} \\
  \eta & < *\eta \text{ cong} \quad \text{bent} \quad \text{WrKhmer} \quad \text{kọŋ} \\
  t & < *c \text{ thit} \quad \text{meat} \quad \text{WrKhmer} \quad \text{sac} \\
  & < *t \text{ chát} \quad \text{acrid} \quad \text{WrKhmer} \quad \text{cot} \quad \text{sour} \\
  n & < *\eta \text{ 掸} \quad \text{weave} \quad \text{WrKhmer} \quad \text{tपङ} \\
  & < *n \text{ trán} \quad \text{python} \quad \text{WrKhmer} \quad \text{thlan} \\
  p & < *p \text{ SSIP} \quad \text{dam} \quad \text{WrKhmer} \quad \text{dɔp} \quad \text{stop up} \\
  m & < *m \text{ nǎm} \quad \text{year} \quad \text{WrKhmer} \quad \text{chnąm} \\
  i & < *y \text{ rùi} \quad \text{housefly} \quad \text{WrKhmer} \quad \text{ruy} \\
  & < *r \text{ või (CaCO}_3\text{)} \quad \text{lime} \quad \text{WrKhmer} \quad \text{kɔmpor} \\
  & < *l \text{ mùi} \quad \text{salt} \quad \text{WrKhmer} \quad \?ɔmpil \\
  u & < *w \text{ tréo} \quad \text{crossed} \quad \text{WrKhmer} \quad \text{slɛv}
\end{align*}
\]

After front vowels, *r and *l regularly disappear.
In some words where cognates in other languages show these consonants after non-front vowels, they are absent in Vietnamese in circumstances that are at present inexplicable.

The reflexes of *-s and *-h are tonal. When the pre-Vietnamese initial was voiceless we have the [hoi] tone and when it was voiced, [ngā].

There is some support, particularly from Khmu?, for a distinction in the proto-language between presence and absence of a final glottal stop. The situation is not entirely clear, as I have endeavored to show in a paper I hope to get published elsewhere. We assume that original final *? is reflected as Vietnamese sá̄c tone after originally voiceless initials and năng after voiced ones, while originally open syllables receive ngang or huyên tone respectively, as Haudricourt was the first to suggest (1954).
C. Vowels.

In a study published in 1976, Harry L. Shorto showed that a vowel system rather close to that of written Khmer could account for the developments of vowels in Mon and in Khmer—and, he believed, in the rest of the family as well. His system contains the following units:

\[
\begin{align*}
\text{uu} & \quad \text{u} \quad \text{i} \quad \text{ii} \\
\text{ua} & \quad \text{ea} \quad \text{i}e \\
\text{oo} & \quad \text{e} \quad \text{ee} \\
\text{oo} & \quad \text{o} \quad \text{a} \quad \text{ai} \\
\text{aa} &
\end{align*}
\]

The principal mergers giving the Written Khmer system are:

\[
\begin{align*}
\text{oo} & \quad \text{e} \quad \text{ee} \quad \text{after voiceless onsets} \\
\text{oo} & \quad \text{e} \quad \text{ee} \quad \text{after voiced onsets} \\
\text{oo} & \quad \text{i} \quad \text{voiceless onsets}
\end{align*}
\]

Old Mon pronunciation was marked by disappearance of the short vs. long contrasts, as each short vowel apparently hastened to merge with some phonetically adjacent long vowel. Also *ue and *ie became monophthongs. The vowel systems of both the modern spoken languages have undergone violent rearrangements.
No details have as yet been worked out showing the appropriateness of this framework for any other Mon-Khmer language. Shorto assumes that numerous roots in the proto-language exhibited vowel alternations. The recent publication of A lexicon of Khmer Morphology by Jenner and Pou makes clear the extent of such variation in this one language. The entry for each root contains a list of all other entries that are reasonably similar in sound and meaning; in many instances these show only vowel differences with near synonymy. The degree of such variation and the uncertainty as to which alternatives are candidates for comparison in historical reconstruction make the extension of the Shorto system to other languages somewhat problematic. It will be discouraging if ever more and more Mon-Khmer roots seem to exhibit a wide range of proto-vowels interchanging in ways we have no means of accounting for.

A preliminary attempt to fit Vietnamese into the Shorto framework is set forth below. For some of the presumed proto-phonemes, I have as yet no compelling Vietnamese reflex. Furthermore, it is clear that many Vietnamese developments require something other than what can be established on the basis of Khmer and Mon alone. Consider for example the following set for which those two languages agree completely:

<table>
<thead>
<tr>
<th>Old Mon</th>
<th>Spoken Mon</th>
<th>Written Khmer</th>
<th>Spoken Khmer</th>
<th>Viet</th>
</tr>
</thead>
<tbody>
<tr>
<td>new</td>
<td>*tmi?</td>
<td>tami</td>
<td>kəmoe?</td>
<td>thmi</td>
</tr>
<tr>
<td>3</td>
<td>*pi?</td>
<td>pi?</td>
<td>poe?</td>
<td>pų</td>
</tr>
<tr>
<td>earth</td>
<td>*ti?</td>
<td>ti?</td>
<td>toe?</td>
<td>tų</td>
</tr>
</tbody>
</table>
(In the last example, the Vietnamese word is very likely not directly comparable. Since there is other evidence that the proto-language had variation in the final consonant between *ti? and *tik, it may well have had *tit as well. Thavung, a language in the same (Mường) group as Vietnamese has /atak/; a final *k is implied also by Souei /ktɛɛ'/ and Bru /kuteig/. Both /thɛɛ/ from final *ʔ and /ndik/ are found in Tin.)

At any rate, a great deal more work is required before any clear picture of the overall development of the Vietnamese vowel system can emerge. Present data allow only an outline of likely principal developments and a very few of the conditioned shifts, which were probably quite numerous.

In what follows comparison is generally to written Khmer. Possible disambiguating Mon forms are supplied, if they could be found, where a merger has obscured the source of the Khmer vowel. Since the correspondence have been worked out only for these two languages, other documentation is not likely to be helpful here.

a. Examples before Proto-Mon-Khmer oral consonants:

<table>
<thead>
<tr>
<th>Việt</th>
<th>PMK</th>
<th>Gloss</th>
<th>Written Khmer</th>
<th>Spoken Mon</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ /</td>
<td>* Việt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u &gt; *uu run worm brûn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; *u run shake grun</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Since there are two nearly certain examples, it seems there was a special development of *u before *y.

\[ uə < *u(y) \quad ruəi \quad \text{housefly} \quad \text{ruy} \]
\[ aə < *u \quad \text{tail} \quad \text{kənduy} \]
\[ uə < *uə \quad vuət \quad \text{rub} \quad \text{buət} \]
\[ ð < *oo \quad \text{to multiply} \quad \text{bor} \quad \text{overflowing} \quad \text{pō} \quad \text{plentiful} \]
\[ o < *o \quad \text{cong} \quad \text{bend} \quad \text{kəŋ} \quad \text{kaŋ} \]
\[ \text{here[ʔuŋ]} \quad \text{rōŋ} \quad \text{pure} \quad \text{rōŋ} \]

No convincing example of *u.

\[ a < *aa \quad \text{gən} \quad \text{weave} \quad \text{tpréŋ} \]
\[ uə < *aa(k) \quad \text{nùc} \quad \text{water} \quad \text{Old Mon} \quad \text{dāk} \]
\[ ã < *a \quad \text{bán} \quad \text{shoot} \quad \text{pəŋ} \]
\[ i < *ii \quad \text{chín} \quad \text{ripe} \quad \text{ch'ın} \quad \text{cin} \]
\[ i < *i \quad \text{lim} \quad \text{disappear} \quad \text{lüm} \quad \text{dim} \]
\[ õ [w] < *i(ŋ) \quad \text{cùng} \quad \text{hard} \quad \text{kəŋ} \]
\[ iə < *iə \quad \text{mìt} \quad \text{crush} \quad \text{sbiət} \]
\[ e < *əə \quad \text{đến} \quad \text{come} \quad \text{tvû} \quad \text{get up} \]
\[ ?? *ee \quad \text{them} \quad \text{in addition} \quad \text{them} \]
\[ a [Λ] < *ə \quad \text{đáp} \quad \text{dam} \quad \text{dɔp} \quad \text{stop up} \quad \text{hədəŋp} \quad \text{close} \]
\[ e [ɛ] < *e \quad \text{trễo} \quad \text{crossed} \quad \text{sləv} \quad \text{hèle} \quad \text{oblique} \]

No example located suggests *ai.

Khmer has no historical Mon-Khmer sources for ɯə, so that when Vietnamese ɯə [ɯə] and Khmer ɯə match up, it seems always to be the case that Khmer has borrowed the Vietnamese word. (Cf. Sakamoto, 1977.)
b. Examples before *h.

(In the Shorto schema the possibilities occurring before laryngeal consonants are more limited than elsewhere.)

\[
\begin{align*}
    u &< *uuh & u & to warm & s\textsuperscript{2}uh & stuffy \\
    \theta &< *uh & l\theta & hole & lu\textsuperscript{1}h & perforate & kal\theta & perforate \\
    &< *oh & l\theta & lose in business & lo\textsuperscript{1}h & ransom & l\textsuperscript{2}h & come unwound \\
    o &< *ooh & bo & toss & po\textsuperscript{1}h & throw & pu\textsuperscript{1}h & shoot with a pellet bow \\
    &< *oh & g\theta & knock & go\textsuperscript{1}h & Chrau & g\textsuperscript{4}h \\
    a &< *ah & va & slap & \ddot{p}ah \\
    e &< *eh & r\ddot{e} & split & re\textsuperscript{1}h & carve out
\end{align*}
\]

No example located suggests *ih.

c. Examples before *? or without a final consonant:

\[
\begin{align*}
    u\textsuperscript{2} &< *u? & l\textsuperscript{2}a & rice(crop) & sr\textsuperscript{2}uv & s\textsuperscript{2}o? \\
    o &< *o & mo & spathe & kha\textsuperscript{3}5 \\
    a &< *a? & ha & open one's mouth & ha \\
    i &< *i? & ri & spirit & var\textsuperscript{2} & custodian & kari & guide \\
    &< *e & b\textsuperscript{4} & calf & h\dot{a}be? & goat & bob\textsuperscript{4} & goat suggests *ee
\end{align*}
\]

Only \( \acute{\theta} \) < *a? is common.

One correspondence suggests the existence of *u\textsuperscript{2}?, not provided for by Shorto.

\[
\begin{align*}
    u\textsuperscript{2} &< *u\textsuperscript{2} & u\textsuperscript{2} & faded & ?u\textsuperscript{2} & stale
\end{align*}
\]
Shorto's reconstruction of *uu?, *ii?, rather than *aw, *ay, (1046, 1047, 1062) may still be open to question. In several branches of Mon-Khmer some languages show diphthongs and others monophthongs. The same is true even between dialects of Vietnamese. It is also interesting to note that only *uu? and *ii? develop without a final glottal stop in Mon. These correspondences are fairly common.

*uu?    cháu    grandchild    cau
*uu    sau    after    krau
sâu    deep    jrau
*ii?    chấy    head louse    cai
*ii    ngày    day    thọai

The differences between au and ău, ay and ăy, as the Vietnamese reflexes, seem not to be systematic. One should recall that the Vietnamese word for 'this' is Northern ndef, Southern nday.

As we learn more about Mon-Khmer historical phonology, some long-standing etymologies invoked to show the relation of Vietnamese will require reinterpretation.

The Vietnamese word for 'foot', chân, is not directly connected to the common Mon-Khmer etymon as seen in Written Khmer蹶, Old Mon jun, but only to the form found in the Northern Aslian languages of the Malay peninsula (formerly know as Semang) such as Jehai /can/ (Benjamin 1976, 107). Vietnamese đá 'rock, stone' has a possible exact cognate in Written Khmer tài 'flat expanse of rock', but the usual Mon-Khmer word represented by Written Khmer thm3, Old Mon
tmo' can be related only by assuming not only an -m- infix--common in Mon Khmer--but also a different original vowel. The non-agreement of the final t in ɗat 'earth' was mentioned earlier.
Notes

Since the material provided by Luce goes so far in providing a general view of Mon-Khmer commonalities in basic vocabulary, it seemed appropriate to list here additional Vietnamese items that are relevant to the comparisons, and also some (given in parentheses) that do not really share a Mon-Khmer history with the forms they accompany, or at least should be viewed with caution. Numbers are those of his list.

1. 2  vãi a couple  135. thread sợi; lua silk
2. 3  ba
3. 7  bảy
4. ye  bay
5. father bố
6. mother mã
7. ear tai
8. breasts vú
9. elephant(tùtông is Sino-Vietnamese)
10. tiger khái [obsolete]
11. voi 'elephant' related to some forms
12. squirrel sóc
13. butterfly butóm
14. crab(tôm with Bru asuom 'shrimp')
15. peafowl công
16. ant mot termite
17. fruit trái
18. thatch grass tranh
19. bamboo(trúc is Sino-Vietnamese)
20. hill field rẫy
21. paddy lúa
22. taro [khoai] sọ
23. cotton vãi
24. 136. fig tree si sp. banyan
25. 139. sky trénn upon
26. 147. wind gió
27. 152. hole lỗ
28. 157. silver(tién is Sino-Vietnamese)
29. 162. village(tình SV though irregular)
30. 168. pot lọ
31. 171. sweep phuỗi dust off
32. 186. far ngái
33. 200. this nay
34. 205. borrow vay
35. 206. bury đáp cover
36. 211. dream mờ
37. (mồng is Sino-Vietnamese)
38. 213. to fly bay
39. 219. lick lưỡi tongue [?]
40. 223. in pain chịu endure
41. 224. weave chăn
42. 227. go home về
43. 229. putrid thởi likely the reflex
44. (rather than hôi cf. Northern Khmu' huur)
Some of the Vietnamese forms given accord with only one among the entries cited by Luce for a given number.

2 Particularly I question items 18, 37, 75, which are Sino-Vietnamese, 96, an earlier borrowing from Chinese, and 56, 67, and 86, where the other examples show hardly any resemblance to Vietnamese. His 46 'to tear' is possible if nhạc-xo 'ragged' be substituted for xé. Several others posit correspondences not otherwise supported.

3 The phonetics for Vietnamese given here is essentially an abstraction based on the orthography. As a reality it can be heard in the studied pronunciation of some educated Southern speakers who partially imitate Northerners when trying to make clear the identity of the words they are mentioning.

4 The problem of the most perspicacious phonemic notation for Vietnamese has little relevance to the present discussion and will simply be ignored. It seems clear that at one time ong words must have had a pronunciation [ɔŋ]. It may be that o < *e is a regular Vietnamese development in some circumstances, paralleling the Khmer merger of *ɔ and *e.

5 Not only is this correspondence not confirmed by finding Old Mon ð (or i), Spoken Mon /ɛ/ (or /i/), but, in the absence of the Khmer form, them would certainly be considered a loan from Chinese and a byform of Sino-Vietnamese thém 'increase' from the late Middle Chinese /t'iam/ that yields Modern tían. Possibly the Khmer and the Vietnamese are independent loans from Chinese 添 or the Khmer (or even both) may be from Thai /thēm/ 'add'.
References


Kuhn, Ernst. 1889. "Beiträge zur Sprachenkunde Hinterindiens." Sitzungsberichten de philosophische-psilologischen und historischen Klasse der königlichen bayerischen Akademie der Wissenschaften, 2.1, 189-236.


Serampore: Marshman.


Bulletin de l'Ecole Française d'Extrême-Orient, 7(1907), 213-63; 8(1908), 1-35.]

——. 1926. Die Sprachfamilien und Sprachenkreise der Erde. Heidelberg:

Carl Winter.

Schott, W. 1856. "Ueber die sogenannten indo-chinesischen Sprachen insonderheit das Siamesische." Philosophische Abhandlungen der königlichen Akademie der Wissenschaften, Berlin, 161-

