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 austroasiatiques;

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.