THE CURRENT STATUS OF VIETNAMESE GENETIC
LINGUISTIC STUDIES

Mark J. Alves-Nan Jeon(1)

INTRODUCTION

Throughout the twentieth century, controversy surrounded the origins of the Vietnamese language. Currently, there are several primary schools of thought, in which Vietnamese is considered to be an Austroasiatic, Tai-Kadai, or Austronesian language. In another approach, Vietnamese is considered to be a 'mixed' language which either cannot be shown to belong to any particular group or can be shown to have a less obvious linguistic substratum. The dominant and most well supported hypothesis so far--based on solid lexical and phonological evidence--has been that Vietnamese is a language of the Mon-Khmer branch of Austroasiatic. However, even at the end of the twentieth century, there are those who still consider Mon-Khmer to be a kind of superstratum and that deeper exploration reveals a Tai or Austronesian base.

This paper explores these issues, summarizing available evidence for each viewpoint, using primarily phonological and lexical evidence, though other factors (typology and historical feasibility) are considered as well. The conclusion reached in this paper is that, though some provocative evidence exists showing possible ancient connections with Austronesian and Tai-Kadai, Vietnamese is nonetheless a Mon-Khmer language, particularly in light of recently collected data from Minor Vietic languages as well as other language contact issues. This paper first discusses the methods used to determine genetic linguistic affiliation. Then, for each of the three main hypotheses (that Vietnamese is Mon-Khmer, Tai-Kadai, or Austronesian), previous work is discussed, stages of linguistic development are hypothesized, and a general summary is given.

For this paper, wordlists by Diguet (1910) and Maspero (1912) for Vietnamese and Tai-Kadai, Matsumoto (1928) for Vietnamese and Austronesian/Japanese, and Luce (1965) and Huffman (1977) for Vietnamese and Mon-Khmer have served as the primary sources of lexical material. The works of Nguyễn T. G. (1978, 1984) on Vietnamese lexical studies, Nguyễn V. L. (1995) on Vietnamese and the Minor Vietic language Rục, and Nguyễn T. C. (1995) on Vietnamese historical phonology were also influential. Data from each source is reviewed, and the said criteria have been applied to restrict what should or should not be used.

(1) College of Technology and Commerce
METHODOLOGY FOR DETERMINING GENETIC LINGUISTIC AFFILIATION

A widely held view in modern historical linguistics is that claiming certain languages share common genetic linguistic origins requires the identification of systematic correspondences in the lexical, phonological, morphological, and syntactic systems of those languages (as stated in Thomason and Kaufman 1988:8). In Vietnamese, the most useful evidence is lexical and phonological, while morphological and syntactic evidence is less useable. Vietnamese lacks agglutination and inflection, using primarily lexical compounding and reduplication. Syntactic evidence is complicated by the overall similarity of syntax in Southeast Asia. The basic sentence order ‘subject-verb-object’, topic-comment structure, and head-modifier word order is seen in a majority of languages. Thus, the primary focus of this work (as has been the case with previous works) is lexical and phonological, though a few assertions can still be made from morphological and syntactic data.

In the application of lexical and phonological data, two aspects have been standard in historical linguistics: (1) the use of basic vocabulary and (2) the identification of regular sound correspondences. Basic vocabulary is the most stable part of languages—it is least likely to be lost and least likely to be borrowed—and thus is the primary viable source of genetic linguistic research. Basic vocabulary must refer to semantically general phenomena common to a group of languages in a selected geographic region. Having identified possible cognates within this portion of the vocabulary, patterns of sound correspondences must be found to validate these words as true cognates. The task is also complicated by the fact that additional sound changes may have occurred over time, causing the sound patterns to be less and less obvious. In such cases, additional evidence is needed, such as different historical reconstructions of languages.

Another criterion is based on observations of the use of lexical data in searching for Vietnamese origins. In various works, words are argued to be cognates with data chosen inconsistently from varying sources or from varying historical stages of languages. To provide convincing evidence of shared genetic linguistic origins, words must be selected consistently from one language or one period of a language and not multiple sources. Words taken randomly from differing layers or subgroups increases the possibility of coincidental look-alikes.

While the theorems of basic vocabulary and regular sound correspondences are still the basis for any judgment of genetic linguistic relationships, in the specific case of Vietnamese, other criteria are needed to further reduce possible uncertainties. In particular, categorizing the type of vocabulary considered results in the inclusion or exclusion of certain words classes. Consider the following criteria.

Criteria for inclusion or exclusion
1. Quantity (requirement)
2. Number of languages having shared cognates (requirement)
3. Multiple possibilities (exclusion)
4. Sound-meaning association (exclusion)

First of all, without a minimum requirement of a number of sound correspondences or a number of words that actually support the claims for sound correspondences, any chance similarities could be used. A sound correspondence is only valid if at least a few words
support it, and conversely, a word only supports claims of sound correspondences if at least a few other words show the same pattern. Next, when dealing with a group of related or potentially related languages, we should expect to find cognates not merely in one language of a language family, but in several languages from a few subbranches. As for exclusions, first, there are words that are associated with more than one language family (cf. /ma/ or some such for 'mother' and /ba/ or some such for 'father', or the words for 'eye' and 'this' which show similar phonological forms in many language families in and around Southeast Asia). Second, words that are unmarked in terms of their sound-meaning association (e.g., words indicating contact through hitting, such as 'tap', 'slap', 'whap', etc.). These criteria, though still subjective, should help to reduce ambiguity and uncertainty.

In addition to general methodology, human migration and contact must also be considered. If linguistic evidence points towards shared historical background between groups, what kind of relationship does it suggest? How viable are hypotheses in the face of available historical resources and social and geographic circumstances? Though certainly speculative, the viability of each of the linguistic-genetic hypotheses in human terms is given some consideration.

AUSTRONESIAN EVIDENCE

The primary works supporting the view that Vietnamese shares identifiable roots with Austronesian are those of Matsumoto (1928), Nguyễn B. L. (1971), and Nguyễn N. B. (1994). Matsumoto (Ibid.), suggest a relationship between Vietnamese and both Japanese and Malayo-Polynesian, a view Nguyễn N. B. supports. Matsumoto provided over seventy possibly cognates—consisting mainly of basic vocabulary (though closer inspection shows that these are not always semantically basic)—to link Vietnamese and Japanese, but only nine posited forms to support the Vietnamese-Austronesian link. The most obvious problem confronting this list of words is the lack of phonological correspondences. However, for these words to be useable, they must provide consistent phonological correspondences—either language to language or stage to stage of languages—neither of which is the case.

Consider the data (Nguyễn B. L. 1971:477) in Table 1. 'Chin' is not a very basic body part term, and other body parts are lacking. 'Friend' is a Sino-Vietnamese form (cf. Mandarin 伴 'companion'). 'Honey' is seen in various languages of Asia and cannot be used for supporting claims of shared historical origins or even direct contact. 'Island' is most likely a Cham loan. 'Leaf' is most likely Mon-Khmer, though the Cham form is interesting. Excluding most of these examples makes 'tree' and 'mountain' seem like a chance similarity. Notice the random selection from either Malay or Cham. There should be consistent cognate evidence in both Malay and Chamic. If there is a significant time depth, then the sound correspondences should not be so direct.

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Vietnamese</th>
<th>Malayo-Polynesian</th>
<th>Japanese</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chin</td>
<td>câm</td>
<td>dagu, anka’, anko'</td>
<td>ago, agi</td>
<td>?</td>
</tr>
<tr>
<td>friend</td>
<td>bốn, bàn</td>
<td>tman</td>
<td>tomona-fu</td>
<td>Chinese</td>
</tr>
<tr>
<td>honey</td>
<td>mật</td>
<td>kemet, kemut</td>
<td>kimo</td>
<td>SEA</td>
</tr>
</tbody>
</table>
island  cù lao  (Cham) kalau, kulau, pulau  ikuri  Cham
Leaf  lá  (Cham) hala  *pa > ha  Mon-Khmer
mountain  non  mnum, bnum, bnom  mine  ?
tree  cây  (Malay) kayu  ko, ke  ?

In another approach, Vietnamese could be considered linked to Austronesian in view of the Austric hypothesis (Schmidt 1906), which links the Austroasiatic and Austronesian language families. The best evidence linking Austronesian and Austroasiatic is primarily morphological (Reid 1994), while lexical evidence is extremely scarce (Diffloth 1994). Only indirectly does Vietnamese show any possible morphology, as seen in the closely related Minor Vietic languages, which show remnants of the causative /pa-/ prefix. The lexical evidence, too, is quite scarce, with the only the only possibilities being 'dog', 'eye', 'fish', and a few others. Still, even in this view, the most immediate relationship is still that Vietnamese is a Mon-Khmer language, not Austronesian.

Considering the issue of geography, Vietnamese would indeed be in the right position for Austronesian origins. The Northern Vietnamese coast would have been accessible to the Austronesian travelers, who, in current theories (e.g., Blust 1996), passed from southern China to Taiwan and then expanded southward. The problem then is to account for the huge amount of basic Mon-Khmer vocabulary and the possible development of a registral and tonal system,¹ even at early stages of contact with Chinese.

**TAI-KADAI EVIDENCE**

Though most of the previously considered evidence linking Vietnamese to Tai-Kadai has been primarily lexical, a more recent proposal using word families (Hồng 1998) has provided some food for thought regarding the Vietnamese-Tai relationship. Maspero (1912)² listed numerous potential cognates that Vietnamese shared with Tai-Kadai or Mon-Khmer. Haudricourt (1954) readily noted that some of the supposed Tai-Kadai forms are actually Chinese in origin. However, this does not account for the more than 100 instances of Tai cognates Maspero posited to be in Vietnamese. In Maspero's data, there are many instances of weak lexical or phonological evidence, cases of onomatopoeia, or forms that are seen throughout languages in Southeast Asia. Of the Tai languages Maspero used for comparison, some are within the sphere of Vietnamese influence and may have borrowed from Vietnamese.

Consider Table 2, which contains words Maspero posited as Vietnamese-Tai cognates. 'Chicken' in Tai languages is likely the Southern Chinese pronunciation of the Chinese word. Moreover, the supposed phonological link in Vietnamese appears weak, lacking the final off-glide. Supposed cognates for 'mouse' and 'partridge' occur in Tai languages not close to Vietnamese territory, while those close to Vietnamese territory do not have them, which makes them less likely Vietnamese-Tai candidates. Moreover,

---
¹ The likelihood that Vietic had phonologically significant phonation and possibly a precursor to a four-way registral/tonal system is suggested by the data of Minor Vietic languages (see Nguyễn V. L. 1988 for data on Rực and Areun).
² A less commonly cited work is that of Diguet 1910, which included what he considered cognates between Vietnamese and a few Tai languages.