

Linguistic Research on the Origins of the Vietnamese Language: An Overview

Author(s): Mark Alves

Source: *Journal of Vietnamese Studies*, Vol. 1, No. 1-2 (February/August 2006), pp. 104-130

Published by: University of California Press

Stable URL: <https://www.jstor.org/stable/10.1525/vs.2006.1.1-2.104>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

University of California Press is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Vietnamese Studies*

MARK ALVES

Linguistic Research on the Origins of the Vietnamese Language: An Overview

As can be seen by a glance through popular encyclopedias,¹ the linguistics community places the Vietnamese language in the Mon-Khmer sub-branch of the Austroasiatic language family, thereby linking Vietnamese in its origins with over 160 languages spoken throughout mainland Southeast Asia.² However, there have been opponents of this position, opponents who have proposed different affiliations and who group Vietnamese variously with the Austronesian, Tai-Kadai,³ or Chinese⁴ language groups. What has made the affiliation of Vietnamese with Mon-Khmer difficult to prove with absolute certainty is the rich lexical array of Vietnamese vocabulary that has roots in Austroasiatic, Chinese, Tai-Kadai, and, to a lesser extent, Austronesian. Furthermore, the linguistic typology of Vietnamese, a monosyllabic and tonal language, makes its Austroasiatic origins even less obvious, as Mon-Khmer languages tend to be bisyllabic and nontonal. Only through application of historical linguistic reconstructive tools can the Mon-Khmer origins of Vietnamese be more clearly demonstrated.

In fact, although a majority of linguists in Vietnam share the position that Vietnamese is an Austroasiatic language,⁵ official maps in Vietnam that show the geographic distribution of ethnic groups⁶ and the language groups to which they all belong leave uncertain the matter of the grouping of Vietnamese, indicating only that Vietnamese and Mường are related but implying that these two languages are somehow distinct in origin from all other

Journal of Vietnamese Studies, Vol. 1, Numbers 1-2, pps. 104-130. ISSN 1559-372x, electronic ISSN 1559-3738. © 2006 by the Regents of the University of California. All rights reserved. Please direct all requests for permission to photocopy or reproduce article content through the University of California Press's Rights and Permissions website, <http://www.ucpressjournals.com/reprintInfo.asp>.

language groups in Vietnam. This lingering uncertainty is significant since Vietnamese has the largest population of speakers (over 82 million) of any language in mainland Southeast Asia. It is, therefore, worth reviewing the available evidence for the various competing hypotheses.

While most of the ideas presented in this article can be found partially summarized in various publications, to date there have been no collective statements on the reasons for a claim of Mon-Khmer origins rather than other competing hypotheses. The purpose of this article is to provide a general, less technical description (with explanations for certain necessary terminology) of the result of historical linguistic research and debate on the linguistic affiliation of the Vietnamese language (i.e., with what languages does it share common origins), with reference to various hypotheses. Overall, based on conservative application of standard historical linguistic methodology and plausible scenarios of human migration and interethnic contact, the hypothesis of Mon-Khmer origins of Vietnamese remains the most viable; indeed, the evidence exemplifies the kind of data one would expect to find in identifying common linguistic origins. Moreover, this hypothesis provides an interesting example of the development of a tonal from a nontonal language and the reduction from a polysyllabic to a monosyllabic language.

Summary of the History of Research on the Origins of Vietnamese

The most complete summary of the early research on the historical linguistic origins of Vietnamese comes from an article by William Gage.⁷ The question of the genealogical linguistic origins of Vietnamese goes back to the early nineteenth century, when Bishop Jean-Louis Taberd first claimed that Vietnamese was a variety of Chinese.⁸ Then, in 1856, James Logan suggested that “Annamese” was part of the Mon-Khmer language family, what he termed the “Mon-Annam formation.” This position was certainly speculative considering the limited amount of data available at the time,⁹ and it was contested in the early twentieth century. In 1912, the renowned French historian and linguist Henri Maspéro wrote a monograph in which he posited that Vietnamese belonged to the Tai language group.¹⁰ It was not until the middle of the twentieth century that another French linguist,

André Haudricourt, published three articles¹¹ that provided the linguistic foundation under which Vietnamese could be convincingly shown to be Austroasiatic in origin according to both lexical and phonological data.¹² Since then, research on the topic has focused largely on the proposed Vietnamese/Mon-Khmer connections, with Michel Ferlus,¹³ Gage,¹⁴ Gérard Diffloth,¹⁵ Nguyễn Văn Lợi,¹⁶ and Nguyễn Tài Căn¹⁷ providing some of the most important publications on the issue and the works of many other scholars providing additional support. However, a number of complicating factors leave the issue unresolved in the eyes of several scholars.

Theoretical Tools in Historical Linguistics

As in archeology, approaches in the field of historical linguistics rely on weight of evidence and reasonable suppositions.¹⁸ Ultimately, there can be no absolute certainty in speculation about the distant past, but logical assumptions can be made and possibilities excluded based on existing linguistic, cultural, and historical data. Using such information, statements can be made about the linguistic structure of languages in the past (what are called “reconstructions”), contact between languages and language groups (e.g., borrowing of lexical and structural linguistic elements), and linguistic genealogical affiliation (i.e., connection with a certain language family). The two essential elements that allow researchers to place two languages in the same language family are (1) a common set of basic vocabulary and (2) sets of recurring sound correspondences between the two languages (e.g., a sound in one language is the same or similar to a sound in numerous cognates [i.e., words with common origins] in another language, thereby establishing a pattern of correspondence). Extensive time depth may leave some issues in doubt since numerous changes may take place over time, but even so, patterns of sound correspondences should be complete enough to allow for reconstructions of a protolanguage, that is, a language in the past from which related languages arose. Additionally, when possible, historical, archeological, and/or anthropological details should provide corroborating evidence and feasible scenarios to account for the separation of one language into more than one.

The term “basic vocabulary”¹⁹ is subjective and even impressionistic, but for well-established language families—such as the Indo-European, which connects languages as geographically far apart as Hindi and English, and the

Sino-Tibetan, a group of over two hundred languages²⁰ that includes, as the name suggests, Chinese and Tibetan—the term is easily exemplified. Basic vocabulary includes the kinds of words in a language that are least likely to be lost over periods of centuries and millennia and more likely to remain for longer periods of time than other types of non-core vocabulary. Words that can be considered truly basic include simple numerals of one to ten, body parts, common natural phenomena and animals, and basic actions/activities, among other such universal semantic classes and aspects of human existence. These are precisely the words that not only unite Mon-Khmer languages into a language family but also connect them to Vietnamese. In determining linguistic affiliation, words that are semantically more general are more useful²¹ than more specific but related words that are subcategories within larger classes. For example, the word *chim* [bird], a Mon-Khmer etymon, refers to an entire semantic field, while *bồ câu* [pigeon], probably of Tai origins, is one specific type of bird, which makes the Mon-Khmer etymon more useful. Basic vocabulary tends to be resistant to moderate language contact (i.e., interethnic contact that results in linguistic exchange), a situation in which some lexical borrowing occurs.²² This makes the category of basic vocabulary more useful for the purpose of identifying common linguistic origins of languages. Finally, lexicostatistics, which is the assessment of percentages of shared basic vocabulary among languages, is a supplementary tool that can be used to identify degrees of relatedness between languages and to provide some measure of evidence for linguistic affiliation. This straightforward statistical approach can certainly be the foundation for further inquiry but does not necessarily lead to certainty since languages do borrow words and sometimes large numbers of them.

Evidence for cognates in two languages involves the identification of phonological patterns of parallel contrasting sounds. Moreover, the more examples of such patternings that emerge, the more strength an argument for cognate status has. Thus, for example, finding several instances of words that show the dental sound /d/ in one language and the dental sound /n/ in another gives strength to the argument that these languages have some kind of connection, though whether genealogical or contact-based cannot be stated with certainty. To determine genealogical affiliation, sound correspondences among sets of basic vocabulary in two languages should include

enough lexical items to allow for reconstructions of complete sound systems and, ideally, some kind of sociocultural reconstruction (i.e., a reconstruction of a cultural system and lifestyle rather than the linguistic system alone) based on the basic vocabulary. This also requires the comparison of reconstructed early stages of groups of languages and not simply two languages in their modern forms. Having cognates with sound correspondences among many languages, including those geographically distant (thereby reducing the possibility of borrowed elements), can strengthen claims of genealogical affiliation. Without this sort of evidence, such claims are weakened.

While the identification of basic vocabulary and regular sound correspondences is still the basis for any judgment of linguistic affiliation, in the specific case of Vietnamese, additional criteria are needed to further reduce possible uncertainties. One of the dangers in a *Sprachbund*, or linguistic area, such as Southeast Asia is the fact that words in different languages can appear similar in both sound and meaning because of typological similarities in sound systems and semantic, grammatical, and pragmatic linguistic functions but still not have related etymologies. The phenomenon of “false cognates” or “look-alikes,” instances of chance phonological and semantic similarity, is a noticeably troublesome problem in Southeast Asian linguistics because many languages there have substantially large numbers of monosyllabic words, in addition to other typological similarities. There is a tendency for certain words or types of words to be so widespread among languages in a region, such as the Southeast Asian linguistic area, a place with a tremendous amount of interethnic exchange, that they actually decrease our ability to establish genealogical affiliation. This is the case around the world for the word for “mother,” which very often has an initial /m/ sound, and for “father,” which very often has a /b/ or similar sound. In East and Southeast Asia, the words that appear similar among various language families include those for “eye” and the demonstratives “this/that,” and it is not possible to say yet, based on existing data and methodology, whether these and other instances of similarity are the result of even more ancient affiliations among the language families or simply of language contact.

Words with more than one possible linguistic origin are best excluded or at least acknowledged as weaker support for a claim of affiliation between distinct language families, a position not often considered in Southeast

Asian historical linguistic research. It is absolutely crucial in a geographic setting such as Northern Vietnam to check for possible lexical connections in all neighboring languages—without preference for one group over another—since that area is a geographic link between China, the Southeast Asian mainland, and insular Southeast Asia. Finally, the phenomenon of onomatopoeia is one more issue that weakens claims of cognates; for example, there is the Vietnamese word *cắt* [cut] as well as similar phonetic forms for the same word in Mon-Khmer languages (e.g., Khmer *kat*), Tai (e.g., *tăt*), and Southern Chinese (e.g., Cantonese *chit* 切). A more conservative approach would amass potential cognates but then exclude or treat carefully as many as possible according to exclusionary criteria (see list below). The remaining data would thus be more likely to have value as evidence of linguistic affiliation. The criteria for determining genetic linguistic affiliation would thus include the following:

1. Quantity of basic vocabulary (inclusive)
2. Patterns of phonological correspondences of basic vocabulary (inclusive)
3. Number of languages in a language family having shared cognates (inclusive)
4. Multiple possibilities among different language families (exclusionary)
5. Onomatopoeia (exclusionary)

In addition to the general methodology outlined above, historical scenarios of human migration and contact must also be considered. If linguistic evidence points toward a shared historical background for two languages, what kind of interethnic relationship does such evidence suggest? How viable are hypotheses in the face of available historical data, and what is known about linguistic, social, and geographic circumstances? Though such research is certainly speculative, the viability of each of the linguistic-genetic hypotheses in human terms must be given some consideration.

Chinese

Vietnamese cultural contact with the Chinese goes back more than two thousand years—one thousand years of Chinese domination followed by another thousand years of tribute-state status and continued cultural contact

with the powerful neighbors to the north. The assumption of Vietnamese genealogical affiliation with Chinese, as noted above, goes back to the colonial era and seems to linger in popular assumptions. Such a view is for the most part based on the tremendous visible quantity of Chinese vocabulary in Vietnamese, significant typological linguistic similarities between the two languages, and, at least until the beginning of the twentieth century, the use by the Vietnamese of Chinese characters (along with Vietnamese *Nôm* script based on Chinese writing).

These aspects, however, are all problematic. Words can be borrowed, and it is the type of shared words, not just the quantity, that supports claims of common origins. Next, while both Chinese and Vietnamese do indeed have tones, so do the neighboring Tai-Kadai and Hmong-Mien languages. Finally, a written script (which is a transferable human artifact and not a genetic endowment) common to two languages is certainly not an indicator of genetic linguistic affiliation; only the spoken language that the writing represents is useable for this purpose. It is also important to keep in mind that similarities between Vietnamese and Chinese are not necessarily the result of Chinese influence. Some may be the result of general tendencies in the geographic region regardless of language family, or these similarities may also be the result of natural, language-internal changes. At the very least, to say that the current linguistic state of Vietnamese is determined solely by contact with Chinese grossly oversimplifies the situation.²³

While considering the noticeable similarities between Vietnamese and Chinese is important and interesting, it is equally important to consider what is not similar to or what in Vietnamese does not originate in Chinese. We must recognize in particular those elements that are required to establish linguistic affiliation between Vietnamese and Chinese. Except in a few literary and specialized instances, no pronouns or numerals in Vietnamese are of Chinese origin. Vietnamese words of truly basic, natural phenomena are also not Chinese, such as *chó* [dog], *chim* [bird], *củi* [firewood], *nước* [water], *lá* [leaf], *rễ* [root], and others, which are in fact of Mon-Khmer origin.²⁴ In general, the kinds of basic vocabulary that hold together varieties of Chinese into one language group are, for the greater part, missing in Vietnamese.²⁵ As for grammar, the noun phrase order in Vietnamese of noun before modifier

(a characteristic of both Mon-Khmer and Tai languages in the region) is the opposite of the order in Chinese, in which modifiers precede nouns. Indeed, the kind of vocabulary Vietnamese borrowed from Chinese is, for the most part, precisely the kind of vocabulary that is commonly borrowed by languages: vocabulary of culturally specific artifacts and cultural constructs (e.g., household utensils and objects or aspects of medicine, government administration, and the arts), and not basic vocabulary, which is relatively more resistant to borrowing.

Rather than words that indicate common origins with Chinese, there exist layers of Chinese loanwords. Researchers such as Wang Li,²⁶ Haudricourt,²⁷ Đào Duy Anh,²⁸ and others utilizing the Chinese historical phonological tradition have identified an early layer of Chinese vocabulary, going back to the Han Dynasty (206 BCE–220 CE), by identifying patterns of change among initial consonants and tones. Words such as those for “kind/type,” “paper,” and “well” show the pattern of initial palatalization and the *sắc* tone,²⁹ as shown in Table 1. Such words are often considered by Vietnamese to be part of the *Nôm* written tradition, which are characters representing spoken Vietnamese, not Chinese. This is in contrast with the later layer of Chinese loanwords from the Tang Dynasty (618–907), what can be called standard Sino-Vietnamese, in which each word is a Chinese character [*Hán tự* 漢字] read with Vietnamese pronunciation.

It is important to note here that the Chinese historical phonological tradition must be taken into account to establish the certainty of the origins of these early loanwords.³⁰ Without application of historical linguistic research on the Chinese “rhyming dictionaries”³¹ going back nearly fifteen hundred years and related studies on modern varieties of Chinese as well as Japanese and Korean readings of Chinese characters, no legitimate claim for genetic linguistic affiliation with Chinese can be made. In some cases,

TABLE ONE: Sino-Vietnamese and Old Sino-Vietnamese (SV) Loanwords

| Gloss | SV | Old SV | Character |
|-----------|-------|--------|-----------|
| kind/type | chủng | giống | 種 |
| paper | chỉ | giấy | 紙 |
| well | tỉnh | giếng | 井 |

through sporadic borrowing of words from another language, individual words may change their phonological forms in irregular ways, but a substantial assortment of recurring natural, typologically common patterns of sound changes must be identified to present a feasible case.

Currently, given the logical application of historical linguistic methodology, any arguments for Vietnamese having Chinese origins and thus having connections to the larger Sino-Tibetan language family simply do not hold water.

Tai-Kadai

The strongest statement on the Tai origins of Vietnamese was made by Maspéro in a 120-page article in the *Bulletin de l'École Française d'Extrême-Orient*.³² In that work, Maspéro identified approximately one hundred words he considered shared cognates of Vietnamese and Tai-Kadai. Though he recognized Mon-Khmer cognates in Vietnamese as well, he posited that, based on the number of apparent Tai etymologies and shared phonological characteristics (such as the presence of tones and a predominantly monosyllabic syllable structure), Vietnamese belonged to the Tai language family. However, Haudricourt³³ pointed out that a significant number of the supposed Tai forms were actually Chinese in origin. Moreover, it can be seen that Maspéro's data shows many instances of weak lexical or phonological evidence, cases of onomatopoeia (e.g., *mèo* [cat]), or forms that are seen in languages throughout Southeast Asia, as noted in the section on theoretical tools in historical linguistics. Another complicating possibility is that some of the Tai languages in Vietnamese territory borrowed from the Vietnamese, a possibility generally not discussed by scholars in the literature on Southeast Asian historical linguistics. In such a situation, data from Tai languages that borrowed from Vietnamese might create a false impression of Tai loanwords in Vietnamese. Finally, the lack of shared basic vocabulary, such as pronouns, numbers, body parts, or other various basic nouns and verbs, weighs heavily against placing Vietnamese in the Tai-Kadai language family.

This is not to say that Vietnamese has not had contact with Tai languages. A more recent word list was amassed by Nguyễn Tài Căn,³⁴ who collected approximately two dozen possible Tai loanwords in Vietnamese. A comparison of those words with proto-Tai reconstructions³⁵ shows that indeed there

are some forms that are strong loanword candidates. The best approach would be to compare reconstructed forms of these Vietnamese words with the reconstructed proto-Tai forms. Otherwise, the patterns of sound change cannot be precisely estimated, and incorrect assumptions are harder to exclude. Such an endeavor awaits a complete reconstruction of proto-Vietic (i.e., the language group consisting of Vietnamese, Mường, and the two dozen Chứt languages).³⁶ Many of the Tai loanwords in Vietnamese are related to agriculture (e.g., *đồng ruộng* [field] and *mương* [ditch/canal]) and animal husbandry (e.g., *vịt* [duck] and *đực* [male animal]), a situation that strongly suggests a language contact but not one with significant impact on the Vietnamese language.

The typological similarities of Vietnamese and Tai, particularly tone, have been considered by scholars in this school of thought to be indication of some affiliation, or perhaps the result of such long-term contact that Vietnamese has developed Tai characteristics, but the geographic region of tonal and isolating/monosyllabic languages throughout China and mainland Southeast Asia is so large that there is no way to use merely the presence of tones (without other factors, specifically, phonological correspondences) as a way to prove linguistic affiliation. Indeed, even some Mon-Khmer languages and the Austronesian Chamic languages have also shown tendencies toward monosyllabicity and the development of tones.³⁷ The use of this areal typological feature for support of claims of linguistic affiliation is hazardous at best.

Based on the geographic proximity of Vietnamese and Tai groups and some of the lexical evidence mentioned above, it can be stated with relative certainty that the ancestors of the Vietnamese were in contact with Tai peoples in the region of modern-day southern China and northern Vietnam two millennia ago. Phạm Đức Dương discusses what appears to be a linguistic basis—lexical evidence as discussed above—for showing cultural borrowing of wet rice agriculture between Vietnamese and Tai peoples.³⁸ Biogenetic evidence also suggests possible close links between the Vietnamese and Tai groups, but other studies show links between Mon-Khmer and Tai-Kadai groups.³⁹ Still, overall, based on current data and research methods, the Vietnamese-Tai relationship can only be considered one of mild language contact at best.

Pacific Origins: Austronesian and Japanese

There exists another group of related, overlapping theories, all suggesting Pacific Ocean origins of Vietnamese, including claims of common origins with the Austronesian language family and even Japanese. Keith Taylor, at the beginning of his *Birth of Vietnam*, wrote the following based on his analysis of cultural traditions and folklore: “The mythical traditions surrounding Lac Long Quan and the origin of the Hung kings reveal a sea-oriented culture coming to terms with a continental environment. Civilization arrived with a culture hero from the sea who foiled a continental power by seizing his foe’s wife and making her the mother of his heirs. . . . The mythical origin of the Hung kings reflects a maritime culture based on political accretions from continental influences.”⁴⁰

Others have posited lexical evidence to support such a notion. Primary works to support the connection of Vietnamese with Austronesian are those of Nobuhiro Matsumoto, Nguyễn Lộc Bình, and Nguyễn Ngọc Bích.⁴¹ Matsumoto went furthest, suggesting a relationship between Vietnamese and both Japanese and Malayo-Polynesian. He provided over seventy possible cognates to link Vietnamese and Japanese, but only nine posited forms to support the Vietnamese-Austronesian link.⁴² The evidence consisted of some quasi-basic vocabulary but no semantically basic forms, and no numbers, pronouns, body parts, or other truly basic vocabulary. Another obvious problem confronting this list of words is the lack of phonological correspondences, notably with respect to tones, for which no feasible explanation has been provided.

Consider the lexical data in Table 2, taken from Nguyễn Lộc Bình.⁴³ “Chin” is a body part term, but not a semantically basic body part, like “face” or “arm,” and other body part terms are lacking. Vietnamese *bạn* [friend] is a standard Sino-Vietnamese loanword (cf. Mandarin *bàn* 伴 [companion]). Phonetically similar forms for the word “honey” are seen in various languages in the region, including both Mon-Khmer and Chinese, and cannot be used to support claims of shared historical origins or even borrowing. “Island” is clearly a Chamic loan since it is one of the few two-syllable words in Vietnamese (i.e., the two syllables do not have individual meaning nor is the word a reduplicant). “Leaf” is most likely Mon-Khmer; the Chamic form is provocative, but possibly misleading.⁴⁴ Excluding these examples makes “tree” and “mountain” seem like chance similarities.⁴⁵

TABLE TWO: Proposed Vietnamese-Japanese-Austronesian forms (Matsumoto)

| English | Vietnamese | Malayo-Polynesian | Japanese | Likely Source |
|----------|------------|----------------------------|-----------|-----------------|
| chin | cầm | dagu, anka', anko' | ago, agi | unknown |
| friend | bạn, bạn | têman | tomona-fu | Chinese |
| honey | mật | kemet, kemut | kimo | Southeast Asian |
| island | cù lao | (Cham) kalau, kulau, pulau | ikuri | Cham |
| leaf | lá | (Cham) hala | *pa > ha | Mon-Khmer |
| mountain | non | mênum, bênum, bénom | mine | unknown |
| tree | cây | (Malay) kayu | ko, ke | unknown |

Another problem is that the lexical items in Malay and Chamic are inconsistent, whereas considering the relative linguistic closeness of the two languages (both are closely related parts of the Malayic sub-branch of Austronesian), more instances should be present. Next, as is the case for “tree” and “island,” considering the significant time depth involved (more than two thousand years), the sound correspondences should actually be less obvious, although explainable and systematic patterns of phonological change should remain. Thus, the evidence is questionable at best, and what we have is most likely an assortment of false cognates. Finally, taken together, these certainly do not constitute a collection of words that are typical Austronesian etyma, which would be persuasive if not required to make claims of linguistic affiliation.

In another approach, Vietnamese could be considered as linked to Austronesian in view of the Austric hypothesis,⁴⁶ which links the Austroasiatic and Austronesian language families. The best evidence linking Austronesian and Austroasiatic is primarily morphological,⁴⁷ and lexical evidence is extremely scarce.⁴⁸ Only indirectly does Vietnamese show any possible morphology and general word-formation patterns, as seen in the closely related minor Vietic languages, which show remnants of the causative /pa-/ prefix, a common prefix in many Austronesian and Mon-Khmer languages. The lexical evidence, too, is quite scarce, with the only possibilities being words for “dog,” “eye,” “fish,” and a few others. Still, even if this hypothesis is tenable, the most immediate relationship is still that between Vietnamese and Mon-Khmer, not Austronesian.

Considering the geographic position of Vietnamese, it would indeed be in the right position to have Austronesian origins. The northern Vietnamese

coast was accessible to the Austronesian travelers, who, in current theories,⁴⁹ passed from Southern China to Taiwan and then expanded southward, to the Philippines and Indonesian archipelago. The Chamic peoples are indeed the ancestors of this great wave that reached the southern coast of Vietnam about two thousand years ago, and the Vietnamese were their immediate neighbors north of the Champa Empire until its fall at the hands of the Vietnamese in the fifteenth century. There may have been language contact between the Vietnamese and Chamic peoples, but the Vietnamese typically had an unfriendly political relationship with their southern neighbors. Contact with them was, according to the linguistic data, minimal; very few genuine instances of Chamic or Austronesian loanwords can be shown conclusively.⁵⁰ The lack of a solid foundation of basic vocabulary and patterns of phonological correspondences makes the claim that Vietnamese is Austronesian in origin extremely weak.

Austroasiatic and Mon-Khmer

The 150-plus Austroasiatic languages are spread throughout mainland Southeast Asia, from Vietnam into Yunnan Province in China, down to mountainous regions of the Malaysian peninsula, and even into eastern India, the latter being the location of the Munda sub-branch of Austroasiatic. The lexical and phonological evidence solidly indicates a relationship between Vietnamese and Austroasiatic. Some of the prime collections of this aspect of the Vietnamese lexicon include those of Maspéro (though he argued for the Tai origins of Vietnamese), Gordon Luce, David Thomas and Robert Headley, and Franklin Huffman.⁵¹ Using lexicostatistical methodology, Thomas and Headley showed that the percentage of basic vocabulary shared by Vietnamese and other Mon-Khmer languages is 25 percent, which is similar to the percentages that link other language branches of Mon-Khmer (ranging in general from 25 percent to 35 percent). Works that provide evidence of supporting phonological correspondences include those by Haudricourt, Ferlus, Gage, Diffloth, and Nguyễn Tài Căn.⁵² Ferlus and, in an entire book devoted to the topic, Nguyễn Tài Căn show the path of phonological development Vietnamese has undergone over the past two plus millennia from its Mon-Khmer roots. From these various sources, three hundred Austroasiatic etyma can be identified in Vietnamese,

ranging as far back as proto-Austroasiatic. Vocabulary dating back to other lower levels/sub-branches (e.g., Mon-Khmer, Eastern Mon-Khmer, etc.) can also be identified, thereby presenting a more complete historical linguistic scenario of the Mon-Khmer peoples' gradual migration throughout Southeast Asia.⁵³

The lexical foundation that links Vietnamese and Mon-Khmer consists of semantically generic basic vocabulary and is, significantly, the same vocabulary that relates other languages and sub-branches of Mon-Khmer languages to each other. Huffman's 1977 one hundred-word wordlist showed several dozen⁵⁴ potential Vietnamese/Mon-Khmer cognates that have numerous sound correspondences with more than a dozen other Mon-Khmer languages. What this evidence shows is (1) core vocabulary, (2) sound correspondences, and (3) a spread of this vocabulary in languages both within and between sub-branches of Mon-Khmer. That last fact makes this data methodologically meaningful. Examples of these issues are shown in Table 3 (scripts from original publications have been kept), which contains data from Rục, a Vietic language spoken in Quảng Bình Province;⁵⁵ Pacoh, spoken in Thừa Thiên and Quảng Trị provinces;⁵⁶ Mon, spoken in Thailand and Burma;⁵⁷ and Riang, also spoken in Burma.⁵⁸ Words missing in the data sources are indicated by "x." It is important to consider the fact that the other Mon-Khmer languages come from four different sub-branches of Mon-Khmer and cover a wide geographic range, yet all have most of these commonly seen Mon-Khmer etyma, a fact that strengthens the inclusion of Vietnamese within Mon-Khmer and reduces the likelihood of an intense linguistic borrowing situation.

Investigation of this list of cognates reveals representative patterns of phonological correspondences between Vietnamese and other Mon-Khmer languages. Ferlus described evidence for the reduction of Mon-Khmer consonant clusters of two or three sounds into retroflex sounds in Vietnamese (e.g., Vietnamese "tr," "s," and "r").⁵⁹ Certain categories of tones in Vietnamese correspond to certain syllable final sounds in Mon-Khmer, as Haudricourt noted in his seminal article positing his hypothesis of the genesis of tones in Vietnamese.⁶⁰ Examples of these claims are shown in Table 4, which contains the categories of correspondences exemplifying such instances in Table 3.⁶¹

TABLE THREE: Vietnamese and Mon-Khmer Comparative Evidence

| English | Viet | Rục | Pacoh | Mon | Riang |
|--------------------|-------------------|----------------------|---------|--------|----------|
| 1. shoot | bắn | pính ³ | pɛ:ŋ | pan | pəiñl |
| 2. four | bốn | pôn ³ | pɔ̃ən | pan | kǎʋpənʔ |
| 3. three | ba | pa ¹ | pe: | pi | kǎʋwaiʔ |
| 4. fish | cá | aka ³ | x | ka | kaʔl |
| 5. leg | chân | ching ² | ji:ŋ | juiŋ | tjə:ŋʋ |
| 6. grandchild | cháu | chu ³ | ʔa.caw | caw | x |
| 7. bird | chim | ichim ¹ | ʔa.cɛ:ʔ | gacem | s'i:mʔ |
| 8. dog | chó | acho ³ | ʔa.cɔ: | x | s'oʔl |
| 9. ripe | chín | chin ³ | cj:n | cin | s'i:nʔ |
| 10. child | con | kon ¹ | ʔa.kɔ:m | kon | kuanʔ |
| 11. weave | đan | tanh ¹ | ta:ŋ | tān | tañ |
| 12. wind | gió | kajo ³ | x | kyā | kurʔ |
| 13. two | hai | hal ¹ | ba:r | ḃā | kǎʋʔarʔ |
| 14. leaf | lá | ula ³ | ʔu.la: | sla | laʔl |
| 15. one | một | môc ⁴ | mɔ:j | moa | x |
| 16. new | mới | bôj ³ | təm.me: | tami | tynʋméʔʋ |
| 17. you (singular) | mày | mi ¹ | maj | beh | miʔʋ |
| 18. nose | mũi | mujh ¹ | môh | muh | x |
| 19. year | năm | nəm ¹ | x | cnam | x |
| 20. yawn | ngáp | táp ³ | ʔa.ʔap | k'a-ap | hap |
| 21. sun/day | ngày | x | ʔi.ŋaj | tjai | s'ŋyɪʔʋ |
| 22. water | nước | dak ³ | da:ʔ | ḃāk | x |
| 23. root | rễ | lech ² | rjəʃ | ruih | riasʋ |
| 24. snake | rắn | pusinh ³ | ku.se:ŋ | x | həiñl |
| 25. thunder | sấm | kruim ⁴ | kri:m | x | tyrʋnamʋ |
| 26. deep | sâu | djoru ¹ | tru: | sjuih | tjyruʔ |
| 27. hand | tay | si ¹ | ʔa.ti: | tai | tiʔl |
| 28. hair | tóc | usuk ³ | sɔk | sok | huk |
| 29. betel | trầu | thu ² | x | jablu | pluʋ |
| 30. hundred | trăm | kləm ¹ | ku.lam | klañ | x |
| 31. fruit | trái | puli ³ | ku.laj | x | plêʔl |
| 32. thatch | tranh | mlənh ¹ | pla:ŋ | x | planʔ |
| 33. far | ngái ⁱ | chóngaj ³ | jo:ŋ | sɲai | s'ŋajʋ</ |

NOTES

"x" indicates words missing in the data sources.

i. The word *ngái* is spoken in Vietnamese dialects in the mountainous areas of Nghệ An province in North-Central Vietnam.

TABLE FOUR: Vietnamese-Mon-Khmer Phonological Correspondences

| Vietnamese | Mon-Khmer | Instances in Table 3 |
|--|--|----------------------|
| Retroflex sounds ("tr," "s," and "r") | Consonant clusters and presyllables | #2–26, #30–33 |
| Voiced stopped sounds ("b" and "d") | Voiceless stopped sounds (/p/ and /t/) | #1–4, #12 |
| Tones sắc and nặng | Syllable-final stopped sounds (/p/, /t/, /c/, /k/, and /ʔ/) | #21, #23, #29 |
| Tones hỏi and ngã | Syllable-final fricative sounds (/h/ and /s/) | #19, #24 |

The quality of the Mon-Khmer basic vocabulary in Vietnamese is also high. For example, in contrast with the possible Tai loanwords *gà* [chicken] or *công* [peacock], from Mon-Khmer, we find the semantically least marked, more universal word *chim* [bird]. Among Mon-Khmer cognates in Vietnamese, we also find the truly basic animal terms *chó* [dog], *cá* [fish], and *rắn* [snake], as opposed to more specific varieties of these animals. Of body parts, we find *chân* [leg/foot], *tay* [arm/hand], *tai* [ear], and *mũi* [nose], which are connected throughout Mon-Khmer and even as far away as the Munda languages of East India. Though fewer examples exist, there are also basic Mon-Khmer verbs, such as *ngồi* [to sit], *mắng* [to listen],⁶² and *chết* [to die]. David Thomas has demonstrated that the numerals from one through ten in Vietnamese all belong to the Mon-Khmer language group or a Mon-Khmer sub-branch.⁶³ Though the pronoun system in Vietnamese has largely been replaced by a system of socially conditioned terms derived mostly from terms of family relations, the intimate/vulgar terms *mày* [you] and *nó* [he/she/it] are Austroasiatic,⁶⁴ certainly withstanding the effect of a complete restructuring of the terms of address and reference in Vietnamese.⁶⁵

The phonological correspondences of the above discussed basic vocabulary are plentiful. Gage listed dozens of individual instances of sound correspondences between words in Vietnamese and words in Mon-Khmer languages.⁶⁶ These instances fall into several broad categories, which is exactly the kind of support needed for evidence of relationships and not just chance similarities. These phonological correspondences provide not only evidence to link Vietnamese and Mon-Khmer but also a means of

reconstructing Vietnamese to stages several centuries ago and earlier.⁶⁷ Some early Mon-Khmer prefixes may also be recoverable in Vietnamese.⁶⁸

Linguistic data from the closely related but extremely conservative linguistic relatives of Vietnamese, the highland Chứt languages⁶⁹ (such as the language Rục in Table 3) in the North-Central Highlands of Vietnam and bordering areas of Laos,⁷⁰ provide evidence that changes earlier hypotheses about the stages and timing of certain phonological developments in Vietnamese. That Vietnamese had consonant clusters is noncontroversial since some dialects have had them into the twentieth century, and they were clearly present in written records from the seventeenth century (as in de Rhodes' 1651 dictionary).⁷¹ When presyllables ceased to exist in Vietnamese is more difficult to posit, though their presence in the closely related Chứt languages strongly suggests that this was several centuries ago or earlier. The possibility of bisyllabic structure in Vietnamese is only hinted at in a Ming Dynasty (1368–1644) Vietnamese-Chinese glossary, in which two characters, and thus two syllables, were used to represent certain words in Vietnamese at that time, and this was more likely a way to represent consonant clusters.⁷² In considering these earlier stages of Vietnamese phonology, the apparent modern similarity between Vietnamese and either Tai or Chinese is thus given a new perspective. The modern stage of Vietnamese phonological structure is simply the recent endpoint of an extended progress of monosyllabification that was likely due to internal reduction following typological tendencies as well as the gradual incorporation of hundreds upon hundreds of monosyllabic Chinese words over two thousand years.

There are in this category two additional pieces of information that highlight the Mon-Khmer origins of Vietnamese. First, Vietnamese has evidence of both word-initial clusters and prefixes or infixes, some of which correspond with Mon-Khmer forms.⁷³ Such evidence was mostly lost through the process of monosyllabification. Second, the patterns of alternating reduplication (known in Vietnamese as *láy*), in which a single syllable is copied partially but typically with one changed sound, are an aspect that suggests typological similarity with Mon-Khmer languages. Admittedly, this is more the case with Mon-Khmer languages spoken in and close to Vietnamese territory than those more geographically distant, suggesting a type of areal influence and not necessarily genetic linguistic affiliation.⁷⁴ Perhaps more

persuasive is the phenomenon of phonologically separable multiword reduplicants, a characteristic seen in Vietnamese and the Mon-Khmer language Pacoh, a noticeably peculiar linguistic feature.⁷⁵

Summary and Comparison of Hypotheses

At this point, we are in a position to compare the different hypotheses in terms of the strength of the linguistic arguments and the feasibility of the anthropological-historical scenarios. Table 5 provides this author's evaluation of the quality of linguistic evidence for the various possible source language groups. Only Mon-Khmer has been shown to share a solid core of basic vocabulary with Vietnamese, with a sizeable quantity of regular phonological correspondences, whereas such is not the case for any other language group. To dismiss the basic Mon-Khmer vocabulary seen in Vietnamese would require the exclusion of basic comparative methodology, which, while far from perfect, is nevertheless important for the systematic collection of supporting evidence. Regarding the phonological and general typological issues, overlap exists among all the groups, and this is perhaps the greatest source of confusion in this inquiry. It is perhaps better to see these linguistic features in the Southeast Asian linguistic area as the result of long-term language contact rather than the result of shared linguistic origins.

TABLE FIVE: Quality of Evidence for Linguistic Affiliation between Vietnamese and Other Language Families

| | Mon-Khmer | Chinese | Tai-Kadai | Austronesian |
|--------------|--|---|--|--|
| Lexical | Solid: Substantial amount of shared basic vocabulary | Weak: Little core vocabulary, highly visible borrowings | Weak: Little shared basic vocabulary | Very weak: No shared core basic vocabulary |
| Phonological | Solid: Numerous correspondences, but some lingering uncertainties | Some: Noticeable overlap in phonological systems | Weak: Similar typology but no patterns of sound correspondences among basic vocabulary | Weak: No identified phonological correspondences |
| Typological | Some: Morphological evidence, affixes and reduplication patterns, post-nominal modifiers | Weak: Various typological differences | Some: General Southeast Asian typology | Weak: General Southeast Asian typology |

In terms of scenarios of interethnic contact, for the two hypotheses of Tai-Kadai or Austronesian origins to be feasible, there would have to have been a period in which the ancestors of modern Vietnamese speakers were in a very socially subordinate position. That is, it would have to have been the case that a group of Chinese, Tai, or Austronesian speakers became so influenced through social contact with a Mon-Khmer language or group of languages that these speakers borrowed most of their basic vocabulary before becoming influenced by contact with Chinese. Of course, it cannot be disproved conclusively that, three millennia ago, Vietnamese was a Tai, Austronesian, or Chinese language that later developed Mon-Khmer language characteristics due to intense language contact, but that could be the case for any language in the region. Without any other evidence, such suppositions are highly speculative. While other hypotheses can never fully be disproved, the strongest positive evidence is still that Vietnamese is of Austroasiatic origins. In fact, research on DNA haplotypes in Southeast Asia has shown substantial commonality between the Vietnamese and the Malaysian Aborigines (the Orang Asli), whose languages are Austroasiatic—a fact that corresponds well with the hypothesis posited here.⁷⁶

Through consideration of the categories of vocabulary and their language group origin, a simple and clear historical picture emerges. A group of speakers of a kind of Mon-Khmer language who most likely practiced swidden agriculture, as many Mon-Khmer groups do, and who were developing a more sophisticated civilization came into contact with Tai speakers in the region of modern Northern Vietnam. The Vietic group gained from the Tai concepts and associated words for wet-rice agriculture and animal husbandry practices, the practices that may have allowed further development of the early Vietnamese civilization. However, there does not appear to be any solid evidence for intense contact with Tai languages specifically that changed Vietnamese linguistic structure, but rather a kind of general process of linguistic homogenization in which languages of many origins in the region became similar typologically. Such linguistic similarities between all of the languages appear to be the result of millennia of contact among various language groups. Next came contact between the Vietnamese and the Chinese. In the past several centuries in particular, the Vietnamese lexicon has been saturated with Sino-Vietnamese vocabulary and

did end up with phonological and morphological/word-formation patterns that resemble modern-day Chinese.⁷⁶ Essentially, Vietnamese is a Mon-Khmer language with a smattering of Tai loanwords and a heavy layer of Chinese lexical and some structural influence—a simple but reasonable statement in light of existing data and analytical tools. ■

MARK ALVES is Associate Professor, Department of Reading, ESL, Foreign Languages, and Philosophy, Montgomery College.

ABSTRACT

While the majority of linguistic researchers both inside and outside Vietnam consider the Vietnamese language a Mon-Khmer, Austroasiatic language, some counterarguments and general confusion continue to exist in the public domain. This article looks at four hypotheses regarding the linguistic origins of Vietnamese, hypotheses that place Vietnamese variously within the Austroasiatic, Austronesian, Chinese, or Tai-Kadai language families. Based on linguistic methodology and plausible scenarios of interethnic contact, the commonly held position—Vietnamese is an Austroasiatic language—remains the most tenable.

KEY WORDS: *Vietnamese, historical linguistics, linguistic affiliation, Chinese, Mon-Khmer, Austroasiatic, Austronesian, Tai-Kadai*

Notes

1. Some examples include the *Encyclopedia Britannica*, *MSN Encarta*, and the online encyclopedia found at www.Wikipedia.org.
2. For more on the number of Mon-Khmer languages, numbers of speakers of those languages, and their geographic locations, see the website of the Summer Institute of Linguistics' "Ethnologue" database (www.ethnologue.com), an enormous if still somewhat incomplete collection of the world's languages. For more on linguistic research of the Mon-Khmer language family and the sub-branches, visit Paul Sidwell's website (Australia National University), <http://www.anu.edu.au/%7Eu9907217/languages/languages.html> (accessed May 24, 2006).
3. The term "Tai-Kadai" refers to three branches of a complete language family, including Kam-Tai, Kadai, and Hlai. In this paper, the commonly used term "Tai," a subgroup within the Kam-Tai subgroup, is the preferred term.

- “Tai” should not be confused with “Thai” (with the “h”), the latter referring to the national language of Thailand and only one of fifty Tai languages.
4. The term “Chinese” in this article refers to a group of related languages—several major groups (e.g., Mandarin, Yue, Min, etc.), each with dozens of dialectal/regional varieties—rather than a single variety of Chinese.
 5. This is an admittedly speculative statement, but it is what seems to be the case among the linguists at the Institute of Linguistics in Hà Nội and among most linguistic researchers in other universities in Vietnam. For a sampling of opinions on this subject, see the website of the Institute of Linguistics in Hà Nội (www.ngonngu.net), which does show acceptance of the Austroasiatic origins of Vietnamese.
 6. Dang Nghiem Van, Chu Thai Son, and Luu Hung, *The Ethnic Minorities in Vietnam* (Hà Nội: Foreign Languages Publishing House, 1984).
 7. William W. Gage, “Vietnamese in Mon-Khmer Perspective,” *Southeast Asian Linguistics Presented to André-G. Haudricourt*, eds. S. Ratankul, D. Thomas, and S. Premisarat (Bangkok: Mahidol University, 1985), 493–524.
 8. Ibid.
 9. Ibid.
 10. Henri Maspéro, “Études sur la Phonétique Historique de la Langue Annamite: Les Initiales [Studies on the Historical Phonology of Annamese: The Initials],” *Bulletin de l’École Française d’Extrême-Orient* 12 (1912): 1–127.
 11. André G. Haudricourt, “La Place du Vietnamien dans les Langues Austroasiatiques [The Place of Vietnamese in the Austroasiatic Languages],” *Bulletin de la Société de Paris* 49, no. 1 (1953): 122–128; André G. Haudricourt, “De l’Origine des Tons en Vietnamien [The Origin of Vietnamese Tones],” *Journal Asiatique* 242 (1954): 69–82; André G. Haudricourt, “Comment Reconstruire le Chinois Archaïque [A Comment on the Reconstruction of Archaic Chinese],” *Word* 10, no. 2–3 (1955): 351–364.
 12. In his 1954 article “De l’Origine des Tons en Vietnamien,” Haudricourt also proposed a hypothesis about the origins of Vietnamese tones, a theory that has had significant impact on historical linguistic research on East and Southeast Asian languages.
 13. Michel Ferlus, “Vietnamien et Proto-Viet-Muong [Vietnamese and Proto-Viet-Muong],” *Asia Sud Est un Monde Insulindien* 6, no. 4 (1975): 21–55; Michel Ferlus, “Sự biến hóa của các âm tắc giữa (obstruents mediales) trong tiếng Việt [Changes of Medial Obstruents in Vietnamese],” *Ngôn Ngữ Học* 2 (1981): 1–21.
 14. Gage, “Vietnamese in Mon-Khmer Perspective.”
 15. Gérard Diffloth, “Proto-Austroasiatic Creaky Voice,” *Mon-Khmer Studies* 15 (1989): 139–154; Gérard Diffloth, “Vietnamese as a Mon-Khmer Language,”

- Papers from the First Annual Meeting of the Southeast Asian Linguistics Society*, eds. Martha Ratliff and Eric Schiller (Tempe: Arizona State University, 1990), 125–139.
16. Nguyễn Văn Lợi, *Tiếng Rục* [The Ruc Language] (Hà Nội: Khoa Học Xã Hội, 1993).
 17. Nguyễn Tài Cẩn, *Giáo trình lịch sử ngữ âm tiếng Việt* [Textbook of Vietnamese Historical Phonology] (Hà Nội: Giáo Dục, 1995).
 18. One crucial difference between physical archeology and historical linguistics is that in archeology, age of objects can be assessed through carbon dating. Historical linguistics is ultimately a field with no such diagnostic tool. The closest objective dating tool is written records, but only when the dates of such materials can be verified.
 19. See Hans Heinrich Hock, *Principles of Historical Linguistics* (New York: Mouton de Gruyter, 1991); Terry Crowley, *An Introduction to Historical Linguistics* (New York: Oxford University Press, 1992), or other historical linguistic textbooks for more discussion on basic vocabulary, lexicostatistics, and other related principles and practices in historical linguistic methodology.
 20. The Sino-Tibetan Etymological Dictionary and Thesaurus, “Introduction,” <http://stedt.berkeley.edu/> (accessed May 24, 2006).
 21. The phrase “more useful” here should be understood in a relative sense. Since what is at issue is weight of evidence, it is necessarily possible that somewhat less basic vocabulary could be cognates. Extreme statements on any side of the issues are not reasonable, but we can hope that accumulation of details is enough to support a position.
 22. For more relevant discussion, see Sarah Grey Thomason and Terrence Kaufman, *Language Contact, Creolization, and Genetic Linguistics* (Berkeley and Los Angeles: University of California Press, 1988).
 23. Mark J. Alves, “What’s So Chinese about Vietnamese?” *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society*, ed. Graham W. Thurgood (Tempe: Arizona State University, 2001), 221–242.
 24. Such evidence comes from comparison with numerous Mon-Khmer languages. In some cases, words such as those for “fish” and “leaf” may have even broader connections with other language groups in the region (“fish” in Austronesian and “leaf” in both Sino-Tibetan and Austronesian), but only systematic studies with rigorous critical application could determine this more definitely. Furthermore, small numbers of possible cognates are not enough to determine whether the issue is language contact or common linguistic origins. Laurent Sagart has even posited an ancient relationship between Sino-Tibetan, Austronesian, and Tai-Kadai. “The Higher Phylogeny of Austronesian and the Position of Tai-Kadai,” *Oceanic Linguistics* 43 (2004), no. 2: 411–444.

25. A good list of typical Sino-Tibetan etyma is provided by Jerry Norman; of those he lists, virtually none have any apparent corresponding cognates in native Vietnamese vocabulary. Jerry Norman, *Chinese* (Cambridge: Cambridge University Press, 1988), 13.
26. Wang Li, “Hanyueyu yanjiu [Research on Sino-Vietnamese],” *Lingnan Xuebao* 9, no. 1 (1948): 1–96.
27. Haudricourt, “Comment reconstruire le Chinois Archaïque.”
28. Đào Duy Anh, *Chữ Nôm: Nguồn gốc, cấu tạo, diễn biến* [Nôm Script: Origins, Formation, and Transformations] (Hà Nội: Khoa Học Xã Hội, 1979).
29. Mei Tsu-Lin, “Tones and Prosody in Middle Chinese and the Origin of the Rising Tone,” *Harvard Journal of Asiatic Studies* 30 (1970): 86–110, specifically 95–96.
30. The most complete historical linguistic work on Sino-Vietnamese, with a complete discussion of the Chinese historical phonological tradition, is that of Nguyễn Tài Cẩn, *Nguồn gốc và quá trình hình thành cách đọc tiếng Hán Việt* [The Origins and Process of Development of Sino-Vietnamese Readings] (Hà Nội: Khoa Học Xã Hội, 1979).
31. The rhyming dictionaries indicated categories of pronunciation for Chinese characters in the Tang and Song dynasties. The exact realizations of the sounds represented can be inferred only by comparing those categories with modern varieties of Chinese, a subject of historical linguistic inquiry that goes back to the late 1800s. See S. Robert Ramsey, *The Languages of China* (Princeton, NJ: Princeton University Press, 1987), 116–142, ch. 7.
32. Maspéro, “Études sur la Phonétique Historique de la Langue Annamite: Les Initiales.” A less commonly cited work is that of Edouard Jacques Joseph Diguët, *Étude de la Langue Thô* [A Study on the Tho Language] (Paris: Librairie Maritime et Coloniale, 1910), which included what the author considered cognates between Vietnamese and a few Tai languages.
33. Haudricourt, “De l’Origine des Tons en Vietnamien.”
34. Nguyễn Tài Cẩn, *Giáo trình lịch sử ngữ âm tiếng Việt*, 322.
35. The Tai loanwords in Vietnamese posited by Nguyễn Tài Cẩn were compared with the online “Proto-Tai-O-Matic,” a database of reconstructed proto-Tai words maintained by the Center for Research in Computational Linguistics Inc., available at <http://www.seasrc.th.net> (accessed August 1, 2006).
36. As discussed by Nguyễn Tài Cẩn, *Giáo trình lịch sử ngữ âm tiếng Việt*.
37. Theraphan L. Thongkum, “An Instrumental Study of Chong Registers,” *Austroasiatic Languages: Essays in Honour of H.L. Shorto*, ed. Jeremy H.C.S. Davidson (London: University of London, School of Oriental and African Studies, 1991), 141–160. For similar discussion, see Đoàn Văn Phúc, “Đón tiết

- hóa và vấn đề ngôn điệu trong các ngôn ngữ Cham [Monosyllabification and Questions of Tone in the Chamic Languages],” *Những vấn đề ngôn ngữ các dân tộc Việt Nam và khu vực Đông Nam Á* [Linguistic Inquiries of the Peoples of Vietnam and the Area of Southeast Asia] (Hà Nội: Khoa Học Xã Hội, 1988), 86–108; and Graham Thurgood, “Language Contact and the Directionality of Internal Drift: The Development of Tones and Registers in Chamic,” *Language* 71, no. 1 (1996): 1–31.
38. Phạm Đức Dương, “Cội Nguồn Mô hình Văn Hoá—Xã hội Lúa nước của người Việt qua cứ liệu ngôn ngữ [The Formative Cultural Origins: The Wet-Rice Society of the Vietnamese through Linguistic Data],” *Nghiên Cứu Lịch Sử* 5 (1982): 43–52.
39. Luigi L. Cavalli-Sforza, *The History and Geography of Human Genes* (Princeton, NJ: Princeton University Press, 1994), 225, 234.
40. Keith W. Taylor, *Birth of Vietnam* (Berkeley and Los Angeles: University of California Press, 1983), 1.
41. Nobuhiro Matsumoto, *Le Japonais et les langues Austroasiatiques: Etude de vocabulaire compare* [Japanese and the Austroasiatic Languages: A Comparative Lexical Study] (Paris: P. Geuthner, 1928); Nguyễn Lộc Bình, *Nguồn gốc Mã Lai của dân tộc Việt Nam* [The Malay Origins of the Vietnamese] (Sài Gòn: Bách Bộc, 1971; repr., Los Alamitos, CA: Xuân Thu, 1987); Nguyễn Ngọc Bích, “Tiếng Việt, Tiếng Nhật, và họ Mã-Lai Đa-Đảo [Vietnamese, Japanese, and Malayo-Polynesian],” *Tuyển-Tập Ngôn-Ngữ và Văn-Học Việt-Nam* 2 (San Jose: Đông Viện, 1994), 437–480.
42. This large number of posited Japanese-Vietnamese cognates by a Japanese speaker is perhaps more an indication of the creative capacity for linguists to see patterns in languages they know best rather than of meaningful insight. The ability to find similar sounding words is certainly strengthened when the number of possible phonemes are smaller, as is the case in Japanese, which has only five vowels, in contrast with Vietnamese’s eighteen vowels, depending on regional variety.
43. Nguyễn Ngọc Bích, “Tiếng Việt, Tiếng Nhật, và họ Mã-Lai Đa-Đảo,” 477.
44. Hypothetically, the Chamic word could be a Mon-Khmer loanword, as the Chamic peoples came into intense contact with Mon-Khmer peoples in Central Vietnam over the past several centuries. See the work of Graham Thurgood, “From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change,” *Oceanic Linguistics Special Publication No. 28* (Honolulu: University of Hawai’i Press, 1999).
45. Another aspect to consider is which syllable of a bisyllabic word is lost. Among Mon-Khmer and Chamic languages, for instance, bisyllabic words receive stress on the second syllable, leading to loss of the initial not the final syllable.

Assuming that Vietnamese has followed this typological pattern, as historical linguistic evidence strongly suggests, a word such as cây [tree] should not have preserved the first syllable.

46. Pater Wilhelm Schmidt, *Die Mon-Khmer-Völker* (Braunschweig: Druck und Verlag von Friedrich Vieweg und Sohn, 1906).
47. Lawrence A. Reid, "Morphological Evidence for Austric," *Oceanic Linguistics* 33, no. 2 (1994): 323–344.
48. Gérard Diffloth, "The Lexical Evidence for Austric, So Far," *Oceanic Linguistics* 33, no. 2 (1994): 309–321.
49. Robert Blust, "Beyond the Austronesian Homeland: The Austric Hypothesis and Its Implications for Archaeology," *Prehistoric Settlement of the Pacific*, ed. W. Goodenough (Philadelphia: Transactions of the American Philosophical Society 86, no. 5 (1996): 117–140).
50. Kenneth Gregerson, in his "On Austronesian Lexicon in Vietnamese," discussed three dozen Austronesian words in Vietnamese, but for the most part, these words must be excluded either because many could come from more than one language group or because most would have phonological problems if they were to be reconstructed. Kenneth Gregerson, "On Austronesian Lexicon in Vietnamese," *Austroasiatic Languages: Essays in Honour of H.L. Shorto*, ed. Jeremy A.C.S. Davidson (London: School of Oriental and African Studies, 1992), 81–94.
51. Maspéro, "Études sur la phonétique historique de la langue Annamite: Les initiales"; Gordon H. Luce, "Danaw, a Dying Austroasiatic Language," *Lingua* 14 (1965): 98–129; David D. Thomas and Robert K. Headley, Jr., "More on Mon-Khmer Subgroupings," *Lingua* 25 (1970): 398–418; Franklin E. Huffman, "The Relevance of Lexicostatistics to Mon-Khmer Languages," *Lingua* 43 (1976): 171–198; Franklin Huffman, "An Examination of Lexical Correspondences between Vietnamese and Some Other Austroasiatic Languages," *Lingua* 43 (1977): 171–198.
52. Haudricourt, "De l'Origine des Tons en Vietnamien"; Ferlus, "Sự biến hóa của các âm tắc giữa (obstruentes mediales) trong tiếng Việt"; Gage, "Vietnamese in Mon-Khmer Perspective"; Diffloth, "Vietnamese as a Mon-Khmer Language"; Nguyễn Tài Cần, *Giáo trình lịch sử ngữ âm tiếng Việt*.
53. Mark Alves, "The Vieto-Katuic Hypothesis: Lexical Evidence," *SEALS XV: Papers from the 15th Meeting of the Southeast Asian Linguistics Society*, ed. Paul Sidwell, The Australian National University, Research School of Pacific and Asian Studies (Canberra: Pacific Linguistics Publishers, 2005), 169–176.
54. Some two dozen cognates are considered less likely due to other complications, the kinds of excluding factors presented in the "Theoretical Tools in Historical Linguistics" section above.

55. Nguyễn Văn Lợi, *Tiếng Rục*.
56. Nguyễn Văn Lợi, Đoàn Văn Phúc, and Phan Xuân Thành, *Sách học tiếng Pakoh-Taôih* [Text for Studying Pakoh-Taôih] (Vietnam: Ủy Ban Nhân Dân, Tỉnh Bình Trị Thiên [The People's Committee, Binh Tri Thien Province], 1984).
57. Luce, "Danaw, a Dying Austroasiatic Language."
58. Ibid.
59. Ferlus, "Sự biến hóa của các âm tắc giữa (obstruents mediales) trong tiếng Việt."
60. Haudricourt, "De l'Origine des Tons en Vietnamien."
61. Some questions regarding the data remain. Why are there words with the *sắc* tone in Vietnamese that do not have final stop consonants (such as *bốn* [four] and *lá* [leaf]), as posited by Haudricourt in "De l'origine des tons en Vietnamien"? Why are some Mon-Khmer word-initial stops also stops in Vietnamese, but others are nasals (such as *nước* [water] and *nó* [it])? What led to the reduction of clusters to retroflex consonants (i.e., Vietnamese "r," "s," and "tr"), as opposed to other possibilities, particularly since retroflex sounds are rare in neighboring Mon-Khmer, Tai, and Chamic languages? Diffloth, in "Proto-Austroasiatic Creaky Voice," provided data that may or may not account for the first question. The second issue will require further research. The third issue has not been fully explained, though the massive borrowing and subsequent contact with Chinese, which has been reconstructed with retroflex sounds (only northern Chinese has maintained these retroflex sounds), may have been a factor.
62. This is an older form, more rarely used than the standard *nghe*.
63. David Thomas, "South Bahnaric and Other Mon-Khmer Numeral Systems," *Linguistics* 174 (1976): 65–80.
64. H.J. Pinnow, "Personal Pronouns in the Austroasiatic Languages: A Historical Study," *Lingua* 14 (1965): 3–42.
65. This restructuring was in large part due to contact with Chinese, from which many familial terms were borrowed, though they have developed functions not seen in Chinese.
66. Gage, "Vietnamese in Mon-Khmer Perspective."
67. For more discussion, see Ferlus, "Sự biến hóa của các âm tắc giữa (obstruents mediales) trong tiếng Việt"; and Nguyễn Tài Căn, *Giáo trình lịch sử ngữ âm tiếng Việt*.
68. Michel Ferlus, "L'infixe instrumental -m- en Khamou et sa trace en Vietnamien [The Instrumental Infix -m- in Khmu and Its Trace in Vietnamese]," *Cahiers de Linguistique, Asie Orientale* 2 (1977): 51–55.
69. The term "Chứt" is the ethnonym used in Nguyễn Tài Căn's *Giáo trình lịch sử ngữ âm tiếng Việt*, and the entire Vietic language group thus consists of the

- Việt-Mường and Chứt language sub-branches. I used the term “minor Vietic” to refer to that group of languages. Mark J. Alves, “Ruc and Other Minor Vietic Languages: Linguistic Strands between Vietnamese and the Rest of the Mon-Khmer Language Family,” *Papers from the Seventh Annual Meeting of the Southeast Asian Linguistics Society*, eds. Karen L. Adams, Thomas John Hudak, and F.K. Lehman (Tempe: Arizona State University, 2003), 3–30.
70. See, for example, data on the language Rục in Nguyễn Văn Lợi, *Tiếng Rục*. Also, for discussion of how such data does provide links, see Alves, “The Vieto-Katuic Hypothesis: Lexical Evidence.”
71. As noted by Maspéro, “Études sur la phonétique historique de la langue Annamite: Les initiales,” 3.
72. Jeremy H.C.S. Davidson, “A New Version of the Chinese-Vietnamese Vocabulary of the Ming Dynasty-1,” *Bulletin of the School of Oriental and African Studies* 38, no. 2 (1975): 296–315 and no. 3 (1975): 586–608.
73. Ferlus, “Sự biến hóa của các âm tắc giữa (obstruents mediales) trong tiếng Việt.”
74. See Laurence E. Thompson, *A Vietnamese Reference Grammar* (Seattle: University of Washington, 1965) for discussion of these patterns.
75. Sonny Vu, “Doing Splits at PF: Compound Separability in Vietnamese,” paper presented at the Ninth Annual Conference of the Southeast Asian Linguistics Society, University of California at Berkeley, May 21–23, 1999.
76. Theodore G. Schurr and Douglass C. Wallace. “Mitochondrial DNA Diversity in Southeast Asian Populations,” *Human Biology* 74, no. 3 (June 2002): 431–452.
77. However, in “What’s So Chinese about Vietnamese?” I claim that (a) much of the modern-day linguistic structure of Vietnamese can be attributed to natural linguistic processes rather than contact with Chinese, (b) Chinese played a relatively minor role in the nonlexical aspects of Vietnamese, and (c) many of the changes that make Vietnamese resemble Chinese structurally were only realized in the past few centuries rather than during the era of the Han to Tang dynasties.