SOUTHEAST ASIA FROM THE LINGUISTIC POINT OF VIEW

Richard S. Pittman
Summer Institute of Linguistics
University of North Dakota

The area which includes present-day Indonesia, Malaya, Burma, Thailand, Laos, Cambodia, Vietnam, Singapore, Hong Kong and the Philippines has often been considered to be a region of unusual linguistic diversity. Without disputing the general sense of this statement, we may, nevertheless, state that the linguistic homogeneity of the area is greater than most people realize. Only four language families are commonly distinguished in this part of the world: Tibeto-Burman, Malayo-Polynesian, Thai, and Mon-Khmer. The status of two others -- Karen and Miao-Yao -- is still uncertain. Much of the apparent diversity of the area is due simply to heavy borrowings between these families, and between each of them and Chinese.

Of the many proposals which have been made for uniting various of the language families of Southeast Asia, we will mention only three:

1. Schmidt (1905) proposed that Mon-Khmer, Malayo-Polynesian, and Munda be united into what he called the Austric family.

2. Przyluski (1923) considered that Mon-Khmer, Cham, Vietnamese and Munda should be united into what he called the Austro-Asiatic family.

3. Benedict (1942) suggested that Thai, Kaddai, and Malayo-Polynesian should be united into a single family.

Vietnamese is most commonly classified as a member of the Thai family (Maspero 1912, 1952) or the Mon-Khmer family (Przyluski 1923; Haudricourt 1954).

It does not seem to me that the evidence is sufficiently convincing, at this stage, to make any one of these proposals truly solid. It does seem, however, that there is sufficient evidence in each case to demonstrate a greater degree of unity among the languages of Southeast Asia than is sometimes recognized. In the case of Vietnamese, I am inclined to agree with Haudricourt's (1954) reasoning that the Thai and Cantonese tone patterns in Vietnamese are not sufficient evidence against a Mon-Khmer relationship. The considerable Mon-Khmer substratum, or basic vocabulary, in Vietnamese, gives much weight to the claim that Vietnamese and Mon-Khmer are members of the same family, even though the other languages of the family do not have tones.

It is not the purpose of this paper, however, to present evidence for or against any specific family grouping. We propose instead to illustrate, with examples from the languages of Vietnam and the Philippines, certain aspects of descriptive
linguistics with which the Western world should be better acquainted and certain phenomena of comparative linguistics with which the Vietnamese people should be better acquainted. Since the features indicated are common to many languages of Southeast Asia, the applications of the paper should extend far beyond the limits of Vietnam and the Philippines.

In this connection I should like to pay tribute to the scholars who formulated the Quốc Ngữ orthography of Vietnamese. The service which they have performed, both to the Vietnamese people and to linguists in general is very great, not only because they have provided an ideal medium for literacy and literature, but also because they have given an ideal medium for linguistic analysis and comparison.

The Mountain Partners

It is not uncommon to hear scholars lament the absence of ancient documents for the study of the history of their favorite languages. We cannot sympathize with this lament, however, when we remember that we are surrounded with countless living documents of ancient language history -- many of which we never look at. The reason we have this rich history is because each dialect, however despised it may be, preserves certain ancient forms which all others may have lost. And no two preserve exactly the same forms. When all living descendants of an ancient language are studied, it is possible to reconstruct, with remarkable fidelity, the ancient and modern history of the languages involved -- even without old documents.

A familiar illustration of the contribution of contemporary partners to the history of a language is that of the Tagalog word for 'roof' atip. We may say that part of the history of this word is lost in Tagalog, because the original vowel of the second syllable is not preserved in Tagalog. Nor is it preserved in the Cebuano cognate atop. But the Cebuano word does give an important clue. Original i and original o should be the same in modern Tagalog and Cebuano, as in the word hipon 'shrimp' which is identical in both languages. When a Tagalog i, therefore, is paralleled by Cebuano o instead of i, a comparative linguist knows that he has either a conditioned variation or evidence of a different vowel. In this case Tagalog atip beside Cebuano atop is evidence of a different historical vowel in the second syllable, and the evidence, especially when repeated in many other words in both languages, is as good as an ancient manuscript.

The nature of the lost vowel becomes plain, in this case, when the Ilocano equivalent, atep, is inspected. The original word atep became atip in Tagalog and atop in Cebuano, but remained unchanged in Ilocano. Certain parts of the story are added by Cebuano, others by Ilocano, and still others by each of the languages of the Philippines.
The same is true of languages all over the world. Greek and Latin could not testify to the history of the consonant $\mathbf{s}$ in those two languages. But Sanskrit preserved the $\mathbf{s}$. Sanskrit, on the other hand, could not tell the history of its own vowels. But Greek and Latin supplied much of the missing information.

Vietnamese is no different from any other language in this respect. Alone, it cannot give more than a very fragmentary testimony to its own history. When the witness of its contemporaries is added, however, the history of many of its features is revealed.

A very nice illustration of the importance of data from the 'mountain partners' for the understanding of the history of the Vietnamese language is the word $\mathbf{nu\text{\textcircled{c}}\,}'$ 'water'. On a map of Vietnam one sees many place names beginning with $\mathbf{Dak}$, which also means 'water'. At first glance one would say that these are two completely different words. The witness of other mountain partners, however, shows them to be variations of one word. Much of the history of the shift between these two apparently different words is almost certainly revealed in the following comparison: Bahnar $\mathbf{dak}$, Mnong Gar $\mathbf{daak}$, Cua $\mathbf{ndak}$, Hre $\mathbf{diak}$, OVN $\mathbf{nu\text{\textcircled{c}}\,}$, MVN $\mathbf{nu\text{\textcircled{c}}\,}$.

Another nice example is the word for 'blow' or 'fan a fire'. Between the Vietnamese word $\mathbf{th\text{\textcircled{o}}\,}$ and the Jeh word $\mathbf{hul}$ there is no apparent connection. But when the witness of Hre $\mathbf{thui}$, Bahnar $\mathbf{thuol}$, and Bru $\mathbf{thor}$ are added, the historical relation between them becomes readily apparent. The shift between a final $\mathbf{-r}$ or $\mathbf{-l}$ and the final $\mathbf{-i}$ of Vietnamese is very well attested in numerous comparisons of Vietnamese and Muong.

One of the most fascinating sound-shifts in the history of the Vietnam languages is the well-known shift from $\mathbf{bl-}$ to $\mathbf{tr-}$. It may be seen in the following words: 'fruit' Old Vietnamese $\mathbf{blai}$ Modern Vietnamese $\mathbf{trai}$; 'sky' OVN $\mathbf{bloi}$ MVN $\mathbf{troi}$; 'betel' Proto-Mon-Khmer $\mathbf{b\text{-}lu}$ Vietnamese $\mathbf{tr\text{\textcircled{a}}\,}$; 'boy' $\mathbf{blai}$ $\mathbf{trai}$; 'moon' in which Old Vietnamese $\mathbf{bl\text{\textcircled{a}}\,}$, as recorded by Alexandre de Rhodes in his dictionary, has become modern Vietnamese $\mathbf{tr\text{\textcircled{a}}\,}$. This last word has a special interest because of strong resemblance to Proto-Malayo-Polynesian $\mathbf{b\text{\textcircled{a}}\,}$ 'moon'.

Vietnamese function words

A feature which is common to many languages in this part of the world is a tendency to use monosyllabic morphemes. This tendency gives rise to a common statement that many of these languages have no morphology.

Vietnamese is an excellent example of a language which some would say has no morphology and all syntax. Another way in which it might be described would be to say that it has only morphemes and arrangements of morphemes (Harris, 1948, p.87). One recent view (Pike 1954) would describe it as all morphemes and tagmemes, which are classes of morphemes in relevant distributional arrangements.
Errata: The last part of the first sentence in the second paragraph below should read: in italics are what Fries (1952) would call 'function words':

None of these views, in my opinion, does justice to Vietnamese. All of them, I think, fail to discern the fact that in any given Vietnamese text certain words serve a primarily lexical function while others serve a primarily grammatical function. That is to say, the distribution of lexical elements in a text is not accomplished solely by arrangement, but, to a large degree, by 'grammatical' words which 'distribute' or 'structure' the 'lexical' words into sentences and into discourses.

In the following Vietnamese sentence, for example, the words in italics are 'function' words, or what Fries would call (1952) 'parts of speech': Trái cam này ngọt làm. (Fruit orange this sweet, very) 'This orange is very sweet.' The words trái, này and làm all serve as 'determiners' (Nida p.111, Ly p.132) helping to structure the words cam and ngọt into a sentence.

The criteria for distinguishing function words from content words have been given in the author's paper 'On Defining Morphology and Syntax' (Pittman 1959). They consist essentially of testing the substitution possibilities of each word in context. If the substitution possibilities are many, it is a content word in that context. If they are few, it is a function word in that context. Thus, in the exclamation Nhiều cua làm! 'Such riches!', cua is a content word, but in the sentence cái mũ cua tôi 'my hat', it is a function word.

It is not sufficient, however, to distinguish only between function words and content words. A further subdivision must be made dividing function words into those which mark classes (Ly's 'witness words') and those which mark sequences. These two kinds of function words may be called 'class markers' and 'construction markers'.

Dr. Ly has already made an excellent description of the class markers in his book Le Parler Vietnamien, and Dr. Hoa has extended the description further in his valuable article 'Classifiers in Vietnamese'. Examples of Vietnamese construction markers are thì, là, cua, nhưng, với, với, and other conjunctions and prepositions.

When class and construction markers have been distinguished on the basis of whether they identify classes or constructions, it is not difficult to distinguish, in a language like Vietnamese, the morphology from the syntax. A content word plus any class markers is morphology. A sequence of two content words plus any construction marking signal is syntax.

In the phrase süt mới cua tôi 'my invitation', we have a nice example of both morphology and syntax. süt is a morphological signal because it relates mới to all the other nouns in the language. cua is a syntactic signal because it relates mới and tôi together in a sequence.

In doing this, the words süt and cua are functioning in exactly the same manner as affixes do in some other languages. süt,
for example, is analogous to the suffix -ness, and cùa is analogous to the suffix -'s in the English phrase the boy's helpfulness.

APPLICATIONS

It is appropriate to ask if there are any useful consequences of these views. Two come immediately to mind:

First, dictionary or grammar entries may be so written as to distinguish content words from function words. Thus, one may write cùa... to distinguish the possessive use of this word from cùa meaning 'riches'. The various functions of không could be indicated: không... 'not', không 'n'est-ce pas?', V không 'free' (as in ăn không 'eat without paying'), N không 'plain' (as in com không 'rice without accompanying dishes').

Second, a shift of emphasis in language teaching from content words to function words would accelerate language learning. A person may learn hundreds of words, for example, like duong, nha, cach, com, xe, che, nuoc, banh without being able to say a single sentence of Vietnamese. But if he learned the functions of even a few score words like...di, cai,...dau,...gi,...va..., co...,...khong,...khac, he would be far into a useful grasp of the language. I estimate that there are less than 400 function words in Vietnamese. If top priority were given to these words in a Vietnamese course, I believe the student would learn the language with maximum speed.

There is still a third application which I believe may be made of this view. The comparative and historical study of the languages of Southeast Asia is much complicated by the fact that content words are borrowed from one language into another, with classifiers or other function forms attached. This may be readily illustrated by the Tagalog words ang lamesa 'the table' and ang silyas 'the chair' which have been borrowed from Spanish. The first was borrowed with the Spanish article la and the second with a plural suffix -s, neither of which now functions in Tagalog grammatical structure. Each of them has become an integral part of its word base.

The identification of la and -s is easy in these words because they are well-known. I believe that identification of function word components in the content words of Southeast Asia can also be done very readily when the function words of the 'mountain partners' are studied for their witness to history.

Bibliography

Benedict, Paul, 1942: Thai, Kadai and Indonesian, American Anthropologist 44.576-601.


----., 1943: Meaning, Monatshefte fur Deutschen Unterricht 35.101-06.


-----, 1954b: De l'origine des tons en vietnamien, *Journal Asiatique* CCXLII.69-82.

Hjelmslev, Louis, 1953: *Prolegomena to a Theory of Language*. IJAL Memoir No.7 (Supplement to IJAL 19. No.1)


Thomas, David D., 1958: Preliminary Proto-Kha Morpheme List, Ms. in Summer Institute of Linguistics workpapers, Grand Forks, North Dakota. 2-10.