The Vietnamese etymological dictionary And "new" Language families

Sikkhamarn, Krisanaa

Abstract

The Austroasiatic language family which includes Vietnamese has been the subject of comparative investigation since the last century. However, it is only now , at the end of the twentieth century, that linguists are able to undertake truly detailed comparative study of the family. From this point of view, Austroasiatis is a typical 'new' language family, in contrast to the 'old' families with long and successful comparative traditions : Indo-European, Semitic and many others. The label 'new' does not indicate that the family is young , and formed only by closely related languages, but rather provides us with information about the level at which comparative studies in the family's linguistic prehistory have been conducted.

The following features are common to 'new' family :

- Lack of detailed synchronic descriptions for the majority of languages included in the family. There are usually good and reliable descriptions (phonological studies,grammars and dictionaries) for just a few of the major languages of the family , with the rest known only superficially. In the Austroasiatic family less than 15 languages are known well enough to be included in comparative study.
- No generally accepted genetis classification. In the Austroasiatis case, linguists are able to identify primary groups of closely related languages, like Vietmuong or Munda,but it remains unclear how these groups are related to each other.

- Reconstruction are available for some of the primary groups of the family. Only three reliable reconstruction of Austroasiatic groups are available to me : Vietmuong(Sokolovkaja MS), Monic (Diffloth 1984) and Katuic (Pejros 1996)

- There is no convincing reconstruction of the proto- languages of the family, and linguists usually have to deal with hypothetical structured which are not based on thorough comparative investigation.

- A lack of reconstruction means that it is difficult (and sometimes even impossible) to identify forms borrowed from one related languages into another.

One of the basic principle of modern comparative linguistics is the aspiration towards completeness which applies both to the data and its interpretation. The comparative method requires that a reconstruction should be based on a thorough study of all the languages include in the family, and a detailed comparison of each pair of the. 'Mass' comparison (on which a word from language A is compared with a word from language B, while another word of A compared with a word from language C without any attempt to find its counterpart in language B) is not a valid procedure. All comparison should be conducted between pair of language (A and B, B and C, A and C, etc.), and the overall investigation should include a though comparison for each of the possible pairings of the chosen languages. The reasons for this requirement are clear: the aim of a proper comparative study is to find regular patterns which connect the systems of all language. This necessitates a knowledge of relationships between the systems of all the languages studied. Only a comparison of each and every pair of these languages, one by one, is sure to discover all possible connections between their systems.

Completeness is thus central principle of modern comparative studies. Applying this principle to the study of 'new' families, the first stage of our investigation must involve:

(i) compilation of a provisional comparative dictionary, which includes morphemes of presumed common origin found in any two, three or more languages of the family;

(ii) establishment of a set of phonological correspondences which connect all the phonemes of all language of the family.

These two components lead linguists to:

(iii) identification of systematic phonological correspondences which are traces of phonological distinction of the proto-language

(iv) reconstruction of the system of photo phonemes; and

(v) reconstruction of lexical and grammatical morphemes of the proto-languages (with the aid of the phonological reconstruction and provisional comparative dictionary). A collection of these reconstructed proto-forms is usually called an 'etymological dictionary' of the family. An etymological dictionary differs considerably from the provisional comparative dictionary in its completion, as it comprises only genetically related morphemes of the languages, omitting resemblances which are due to borrowing, chance similarities, local influences or other non-genetic factors.

It is clear, however, that everything in comparative linguistics depends on success in the completion of a provisional comparative dictionary, and on our ability to interpret its evidence. This is basic upon which a linguist can conduct a morphological reconstruction, create a genetic classification of the family, and engage in other comparative research.

An etymological dictionary for a particular language –a culmination of comparative investigation into that language's history –is based on intensive lexical

studies and on a deep knowledge for the historical phonology of the language and its family. Does this imply that attempt to compile an etymological dictionary for a language of a 'new' family is a priori premature? I do not think so and for several years now I have been completing an etymological dictionary of Vietnamese. The theoretical foundations of this project are discussed below.

As mentioned above, a 'new' family know only partially with detailed reliable descriptions being unavailable for most of its language. In dealing with such a family, then, we need a strategy which will meet the principle of completeness. One strategy is to include data from every languages or dialect mentioned in the literature, regardless of the possibility that this data is by no means complete. If in our example, a languages is know only by a list of 200 words, we could include it in our investigation, but we would not expect to be able to draw detailed conclusions about this language. If the number of such poorly recorded languages is significant, then a provisional comparative dictionary and phonological correspondences based this strategy will be complete, but not adequate for a reconstruction.

Another possibility is to concentrate form their comparison. In such a strategy the emphasis is on the complete interpretation of the data rather than on the completeness of the data itself. Both of these strategies are represented in the literature, but I think that the second one is more appropriate in the circumstances of 'new' language families such as Austroasiatic. My project is therefore base on a study of several languages chosen as primary sources, and an attempt is made to identify all possible comparisons between these languages and Vietnamese.

The staring point of the project is a list of common Vietnamese morphemes represented in the major modern Vietnamese dictionary. To it I have added some archaic morphemes which interesting etymologies. Recent borrowings. mostly from European languages, are not included. Altogether the list comprises about 5000 entries.

Vietnamese, together with many close related languages to the Vietnamese group of Austroasiatic. In principle, this group should be investigated before a comparison with other Austroasiatic languages begins. How ever, it is not possible to compile a full scale Vietmuong comparative dictionary because published data is not available for most Vietmuong languages (Barker 1993). With the exception of Vietnamese, only Muong dialect has been described in any detail (Materialy 1987), although there are short dictionaries of Ruc (Nguyen et al. 1988) and Thavung (Ferlus 1979). Nonetheless the history of the group has been the subject of intensive investigation, and numerous reconstruction have already been proposed (Barker 1963; Barker 1970; Ferlus 1975; Thompson 1976; etc.). In my opinion the best reconstruction is Sokolovskaja's (MS), which is base on interesting and important data including Nguyen Van Tai's dissertation, and unpublished field materials of joint Russian-Vietnamese linguistic expeditions. Sokolovskaja identifies more than 700 cognates showing good semantic and phonological correspondences between the selected languages. The Proto-Vietmuong reconstruction includes many consonantal clusters, and some disyllabic roots. There is no evidence of tonal oppositions, and the tones of modern languages are explained as having developed from two suffixes *-x and *-? These suffixes, which follow the obligatory final consonant of the root and are uncertain grammatical meaning, have been maintained in Arem. Their existence in Proto Vietmuong was first suggested by Haudricourt (1954).

According to Sokolovskaja the development of Vietnamese tones can be represented as follows:

Figurel

Vietnamese

PVM	PVM terminals			
	*-N	*-N+h	*-N+?	*-C
p, ?p	a	ã	á	
*b, *m	à	ã	a	

-C –finals *-p, *-t or *-k

-N –any other final consonant or 0.

Figure 2 represents the development of modern Vietnamese initials from the Vietmuong system suggested by Sokolovskaja:

Figure 2

Development of Vietnamese initials

(after Sokolovskaja MS)

Vietn.	PVM	Vietn.	PVM
b	**p*	*d*	**t*
ph	**ph*	*th*	**th*
v	**b, *w*	*d*	**b, *j*
m	**m, *?p*	*n*	**n, *?t*
		l	**l, * lh*
		l~nh	**ml*
ch	**c*		
x	**ch*	*nh ~ r*	**?c*
gi	**j*	*gi ~ tr*	**pl*

nh	**I*
t	**s*
r	**r, *rh*
c,qu,k	**k*
kh	**kh*
g	**g*
ng	**I*
O	**?*
h	**h*

nh ~ l	**ml*
s	**pr, *kr*
r ~ nh	**?c*
tr	**kl*
tr ~ gi	**pl*

Using these two figures and a chard of development of Vietnamese finals. One can predict a proto form for any Vietnamese morpheme. These forms can then be checked (where possible) against Sokolovskaja's Proto-Vietmuong reconstructions. The Vietnamese root 'to bark', for example, can be traced back to hypothetical **[p,k]-rhua-x* which can be compared with Wa *ruah* 'to bark and Written Khmer *bruh* 'to bark'. In this case the Vietmuong form is know, and is reconstructed as **k=rhua-x*. the proto-form of Vietnamese trum 'to cover', 'to wrap' is not know, but we can postulate it as **k-lum* and thus to compare it with Pakohk lum 'to cover completely', and Nyakur forms like khl 'im 'cover the ridge of a roof. Most Vietnamese word are not represented in Sokolovskaja's manuscript or any other Vietmuong study, and in these cases I am forced to accept the strategy of historical extrapolation of forms.
Dictionaries of four Vietmuong languages are included in my list of primary sources and are to be investigated in detail:

(1) Vietnamese

(2) Ruc (Nguyen et al. 1988)

(3) Thavung (Ferlus 1979)

(4) Muong (Materialy 1987)

Aadditionally seven of the main Mon-Khmer languages are identified as primary source, and lexicons will also be included in my database:

(5) Khmer, represented in traditional orthography;

Several Khmer dictionaries are available to me, but for purely practical reasons Gorgoniev' s Khmer-Russian dictionary (1984) forms the main source of information, and its data data already been thoroughly investigated.

(6) Mon. with the addition of Diffloth's Monic reconstruction;

Two Mon dictionaries by Shorto (1962; 1971) are the main source of lexical information. Since Nyakur vocabularies have been investigated by Diffloth I do not plan to conduct additional investigation.

(7) Proto-Katuic represented by my reconstruction (Pejros 1996); The reconstruction is based on the direct comparison of four Katuic languages (Bru,Kui,Pakoh), each prescnted in a substantial dictionary. Reconstructed proto-form as well as forms of these spoken languages are included in the database.

(8) Khmu;

Only the dictionary of Northern Khmu (Suwilai Premsrirat) is currently available to me. I plan to use dictionary of Southern Khmu published in Laos if I can obtain access to it.

(9) Parauk-Wa;

As it is represented in the Chinese dictionary (Yan et al 1981). To this I will add my comparative Proto-Palaungic data, which includes etymologies gicen in the literature (Luce 1965,Diffloth 1980,Svantesson 1988; Paulsen1989, etc.) and those obtained through direct comparison of major Palaungic sources (Milne 1931; Zhou & Yan 1986 ; Li et al 1986;Chen et al.1986). Additional investigation is needed to obtaain a thorough Proto-Palaungic reconstruction.

(10) Khai;

Two Khasi dictionaries are available to me (Singh 1906;1920).

(11) Comparative Bahnaric lexicon;

The project conducted by Paul Sidwell (Meldourne) will ultimately include a comparative phonology and lexicon of the main Bahnaric languages and a phonological reconstruction. An intermediate version (mid 1995) is at my disposal.

Some other Mon-Khmer or Austroasiatic languages, such as Aslian, Pearic and Nicobaric , will not be included in the database , as no reliable and sufficient lexical information about them is available. I have dictionaries of two Munda languages- Santali (Bodding 1929-36) and Sora (Ramamurti 1938)-which could be used in the project, but I have serious concerns about the quality of the transcriptions. This concern, together with the absents of a reliable Proto Munda reconstruction, makes it difficult to use the Munda data in the project.

All of the lexical sources listed will be included in a single comparative database which when complete will include about 30 000 entries. The database was designed for me by Dr. Durie (Mrlbourne) and is based on the'4D Database' program for the Macintosh.

The principle of completeness will apply to this whole database: each pair of languages is to be thoroughly compared and all possible lexical similarities are to be identified.At this stage of investigation,it is not important that main goal of the project is the history of Vietnamese: the data for all the languages included are equally important. The identification of comparisons is perhaps the following pairs languages:

Vietnamese – Khmer	Vietnameses - Mon	Vietnamese - Katuic
	Khmer – Mon	Khmer – Katuic
		Mon – Katuic

Many hundreds of forms found in any two, three or in all four languages have been identified, with the majority of these not known from the published source. The next language, probably Wa, will be compared with each of the four languages. At the end of this stage (approximately by the middle of 1997) I should have a representative collection of lexical comparisons, with most of them of Mon-Khmer origin. They will be sued to work out a provisional phonological reconstruction of the family. Using this reconstruction I will be able to separate Vietnamese word of common origin from those borrowed from related Mon-Khmer languages.

The next problem I need to address is the identification of loans which came to Vietnamese from languages other than Mon-Khmer. I plan to investigate it following the principle as discussed above: maximum completeness of data interpretation, at the expense of the rang of languages studied.

The Chinese influence is very strong in Vietnamese, which makes Chinese data essential for the project. The list of Vietnamese word included in the project has already been analysed from this point of view, and for each Chinese loan I have the following information:

- Modern, Middle and Old Chinese readings of the corresponding character;
- information about the period from which the word is recorded in Chinese;
- information about the possible period of borrowing into Vietnamese. This part of the project is the work of Starostin(Moscow).

Theoretically, there could be loans in Vietnamese from any other Southeast Asian languages, even if we have no have no good evidence of then yet. For this reason, I have included in my list of primary source the following languages:

(12) This (Siamese) with my Proto-Kadai reconstruction and etymologies(Pejos to appear);

(13) Yao (Lombard 1968) with my Proto Miao-Yao reconstruction and Miao and Miao-Yao etymologies(Pejos to appear);

(14) the main Austronesian reconstructions.

These will be included in the main database and compared with Vietnamese and other Mon-Khmer languages along the lines discussed above.

The result of this project will be a comprehensive set of historical data, including:

(i) all Mon-Khmer comparisons from the database, with or without Vietnamese;

(ii) various loans from one Southeast Asian languages into another, regardless of the direction or period of borrowing.

From these result I plan to extract the cases where the Vietnamese form are know, and to publish then with necessary as a Vietnamese Etymological Dictionary.

BIBLIGRAPHKY:

Barker, Milton E.

1963 Proto-Vietnamuong Muong tone correspondences. *Van-hoa Nguyet-san* 12:491-500.

1966 Vietnamese and Muong tone correspondences. In Zide, ed 1966-25.

Barker, Milton E and Muriel A. Barker

1970 Proto-Vietnamuong (Annamuong) final consonants and vowels. *Lingua 24/3:268-285.*

Barker, Miriam A.

1993 Bibliography of Muong and other Vietic language groups, with note. *Mon-Khmer* Studies 23:197-243.

Bodding,Paul O.

1929-36 *A Santal dictionary*. 5 vols. (1993 edition: Delhi: Gian Publishing House).

Chen Xiangmu et al.

1986 *Sketch grammar of Dean* (in Chinese). Beijing: Nationalities Publishing House.

Diffloth, Gerard

1980 The Wa languages. *Linguistics of* the *Tiberto-Burman Area* 5/2:1-182.

1984 *The Dvaravati Old Mon language and Nyah Kur*. Bankok: Chulalongkorm University.

Ferlus, Michel

1975 Vietnamien et proto-viet-muong. *Asia du sud-Est et Monde Insulindien* 6/4:21-55

1979 Lexique francais-thavung. *Cahiers de Linguistique Asie Orientale* 5:71-94.

Haudricourt, Andre-Georges

1954 De I'origine des tone vietnamien. *Journal Asiatique n.s.*243:69-82.

Jenner, Philip H. et al.,eds

1976 *Ausroasiatic studies* 2 vols. (Oceanic Linguistics Special Publications) 13). Honolulu: The University Press of Hawaii.

Li Daoyung et al.

1986 *Sketch grammar of Blan* (in Chinese). Beijing: Nationalities Publishing House.

Lombard, Sylvia J.

1968 Yao-English dictionary (Data paper 69, Linguistics series 2). Ithaca: Cornell University Southeast Asia Program.

Luce, Gordon H.

1965 Danaw, adying Austroasiatic language. *Lingua* 14: 148-176.

Materialy

1987 *Materialy sovertsko-v'etnamskoj lingvisticheskoj ekspedicit. Jazyk myong*. Moscow: Nauka.

Milne, Leslie

1931 *A dictionary of Palaung-English and English-Palaung*. Rangoon: Government Printing.

Nguyen Phu Phonh et al.

1988 *Lexique Vietnamien-Ruc-Francais.* Paris: University de Paris VII.

Paulsen, Debbie L.

1989 A phonological reconstruction of Proto-Plang.MA thesis. The University of Texes at Arlington.

BIBLIGRAPHKY:

Barker, Milton E.

1963 Proto-Vietnamuong Muong tone correspondences. *Van-hoa Nguyet-san* 12:491-500.

1966 Vietnamese and Muong tone correspondences. In Zide, ed 1966-25.

Barker, Milton E and Muriel A. Barker

1970 Proto-Vietnamuong (Annamuong) final consonants and vowels. *Lingua 24/3:268-285.*

Barker, Miriam A.

1993 Bibliography of Muong and other Vietic language groups, with note. *Mon-Khmer* Studies 23:197-243.

Bodding,Paul O.

1929-36 *A Santal dictionary*. 5 vols. (1993 edition: Delhi: Gian Publishing House).

Chen Xiangmu et al.

1986 *Sketch grammar of Dean* (in Chinese). Beijing: Nationalities Publishing House.

Diffloth, Gerard

1980 The Wa languages. *Linguistics of* the *Tiberto-Burman Area* 5/2:1-182.

1984 *The Dvaravati Old Mon language and Nyah Kur*. Bankok: Chulalongkorm University.

Ferlus, Michel

1975 Vietnamien et proto-viet-muong. *Asia du sud-Est et Monde Insulindien* 6/4:21-55

1979 Lexique francais-thavung. *Cahiers de Linguistique Asie Orientale* 5:71-94.

Haudricourt, Andre-Georges

1954 De I'origine des tone vietnamien. *Journal Asiatique n.s.*243:69-82.

Jenner, Philip H. et al.,eds

1976 *Ausroasiatic studies* 2 vols. (Oceanic Linguistics Special Publications) 13). Honolulu: The University Press of Hawaii.

Li Daoyung et al.

1986 *Sketch grammar of Blan* (in Chinese). Beijing: Nationalities Publishing House.

Lombard, Sylvia J.

1968 Yao-English dictionary (Data paper 69, Linguistics series 2). Ithaca: Cornell University Southeast Asia Program.

Luce, Gordon H.

1965 Danaw, adying Austroasiatic language. *Lingua* 14: 148-176.

Materialy

1987 *Materialy sovertsko-v'etnamskoj lingvisticheskoj ekspedicit. Jazyk myong*. Moscow: Nauka.

Milne, Leslie

1931 *A dictionary of Palaung-English and English-Palaung*. Rangoon: Government Printing.

Nguyen Phu Phonh et al.

1988 *Lexique Vietnamien-Ruc-Francais*. Paris: University de Paris VII.

Paulsen, Debbie L.

1989 A phonological reconstruction of Proto-Plang.MA thesis. The University of Texes at Arlington.

Pejros, Ilia
1996 *Katuic etymological dictionary*. Pacific Linguistics C-
to appear Linguistic prehistory of Southeast Asia. A monograph to be published with Pacific Lingustics.
Ramamurti, G. V.
1938 Sora-English dictionary.(1986 edition: Delhi: Mittal Publication).
Shorto, Harry L
1962 *A dictionary of Modern spoken Mon*. London: Oxford University Press.
1971 *A dictionary of Mon inscriptions, from the sixth to the sixteenth centuries, incorporating materials collection by the late C. O. Blogden.* London Oriental Series 24 London: Oxford University Press.
Singh Nissor U.
1906 Khasi-Enhlish dictionary. (1986 edition : Delhi: Gain Publishing House)
1920 English-Khasi dictionary. (1994 edition: Delhi: Mittal Publications).
Sokolovskaja, Natalia K.