

Vietnamese as a Mon-Khmer language

Gérard Diffloth

Cornell University

During roughly the last millennium and a half, the Vietnamese language has had, it seems, a more dramatic history than other languages in Southeast Asia. Current scholarship shows that the language has changed from having the typical sesquisyllabic word-structure of an ordinary Mon-Khmer language with a great variety of rimes, no tones, and complex initials possibly containing affixes, to being strictly monosyllabic, having few final consonants, six (or five) tones and hardly any initial clusters.

Notions about the history of the language also have had their own, equally spectacular history. British scholars like Logan and Forbes were the first to upset the traditional view, echoed by Taberd (1838), that Vietnamese was simply an offshoot of Chinese. They proposed instead that VN was related to Mon, to Khmer and to several other languages of the Indochinese peninsula. Their idea of a Mon-Anam family (Logan 1856, Forbes 1878, 1881) was looked at with a good deal of suspicion, partly because Vietnamese had tones, and tone languages were considered to be radically different from others. It took another century, and Haudricourt's insistence, for the traditional view to be overturned and the original insight of Logan and Forbes' to gain credibility.

The history of the emergence of tone thus plays, and will surely continue to play, a prominent role in studies of VN historical phonology; I will offer some

views on the problem in another paper. But now looms the larger and more difficult question of finding a place for Vietnamese in the Austroasiatic family, or more precisely for the Vietic branch¹ away from, or among, the various known branches of the Mon-Khmer family.

Haudricourt was elusive on the matter. But Ferlus (1991) suggested "de solides affinités" between Katuic, Bahnaric and Việt-mường² without providing any evidence. He was presumably thinking about the lexicon, and I would share that impression, especially with regards to Vietic and Katuic.

Respect for strict methodology would lead us to reject the question as being undecidable for the moment. The current classification of Mon-Khmer into eleven branches (Thomas and Headly 1970, Diffloth 1974), though still valid, rests to some extent on now generally discredited lexicostatistical techniques, and no one has yet proposed a set of innovations distinguishing each branch of Mon-Khmer from the rest. The higher grouping into three major divisions: North MK, East MK, South MK, (Diffloth 1979) also remains little more than a guess. But fresh linguistic data³ collected on the lesser-known languages of the Vietic branch may allow us to break new ground.

The first phonological correspondence I will now document serves to identify all and only the languages which belong to the Vietic branch; it probably represents a Proto-Vietic innovation. The other correspondence sets apart all Vietic and all Katuic languages together, and no others in the family. Depending on reconstructions, this could indicate that Vietnamese, far from being a remote offshoot of the Austroasiatic phylum, as is sometimes suggested, belongs not simply to the Mon-Khmer family, but in fact to a subdivision within it. We could say that Vietnamese, and with it the Vietic branch, is a member of a Northeastern subdivision of Mon-Khmer consisting of two branches,

Vietic and Katuic; this in turn would be a part of the larger Eastern Mon-Khmer division which also includes Khmeric, Bahnaric and perhaps Pearic, the other two major divisions of the family being Northern and Southern Mon-Khmer.

This proposal entails, I realize, some drastic revisions of commonly held views on Vietnamese proto-history, cultural links, and early movements, but it is based on the most rigorous arguments available at present, which unfortunately is not to say very much. And other scenarios remain possible.

The Proto-Vietic *k-/ʔ- correspondence.

There are several Mon-Khmer, or even Austroasiatic, etyma which begin with a glottal stop in all branches of the family, but are found, surprisingly, with an initial *k- in every known Vietic language. Only six or seven examples have been found so far, but several of them are often recorded and well documented words. Cognates from all branches of Mon-Khmer are quoted here and give an idea of how Vietic fits, or does not fit, in the rest of the family.

"fire, firewood"

Vietic:

- VN: củi, ; Mường Khoi: kuy⁴, Mường Khên: kuy⁴⁵
- Thavưng⁴: kuy⁵
- Maleng: kuy⁵, Rục: kurh, Arem (Ferlus 1991): kuh
- Tum: ku:c31

The breathy register of Thavưng does not necessarily indicate a proto-voiced initial (contra Ferlus, 1974) but is conditioned in part by the quality of the proto-vowel. The Vietic forms above indicate a Proto-Vietic *ku:s, agreeing with Ferlus (1991).

In the rest of MK, the etymon has been identified for a long time (Schmidt 1905:22, Skeat and Blagden 1906:

F124). It is absent from Munda and the whole Northern-MK division⁵ except for Western Khmuic: Mal ʔo:ŋ, Prai: ʔo:t "fire"; otherwise, it is almost omnipresent in Eastern MK:

Khmer: ʔoh: "firewood"

Pearic: Chong: ʔu:t, Song, Samre: ʔu:h "firewood"

Katuic: Pacoh: ʔu:ŋ, Kuay: ʔu:h "fire, firewood"

Bahnaric has a cognate in only one language: Sre:

ʔo:s "fire", otherwise a similar sounding *ʔu:n has replaced this etymon.

It is omnipresent in the Southern division:

Monic: Nyah Kur: ʔuaŋ, Proto-Mon *ʔoh (Diffloth 1984:128) "firewood"

Aslian: Proto-Semai *ʔo:s, Temiar: ʔo:s, Jah Hut: ʔoh, Semelai: ʔus "fire"

Nicobarese: Central: ʔõh, Teressa-Bompaka: õsh "fire"

Note that all non-Vietic items have initial *ʔ-, and that the semantic extension from "fire" to "firewood", also found in Vietic, is older than the split between East and South Mon-Khmer.

"(part of the body) to **swell**"

Vietic:

- Thavưng: kɛh

- Maleng: kəɬ, Rục: kɛɬ, kəal

- Tum: kacɬ

The word is not attested in Việt or Mường, but the correspondence: Thavưng -h, Tum -c indicates a final *-s, and a probable Proto-Vietic form *kəs.

In the rest of MK, his etymon, absent from Khmeric and Monic, has received little attention so far, but goes back to Proto-MK. It is attested in the Eastern MK division:

Pearic: Chong: ʔət, Song: ʔəh

Bahnaric: Tarieng, Sre: ʔas "to swell", Brao: ʔəh "abcess"

Katuic: Pacoh ʔaŋ, Kuay: ʔa:h

in the Southern MK division:

Aslian: Proto-Semai *ʔas, Semelai ʔos "to swell"

Nicobarese: Central: ah-a

in the Northern MK division:

Khasian: Standard, Amwi: ʔat "to swell"

Palaungic: Ta-ang: ʔar, Ka-ang: ʔaɿ, Rumi: ʔo:ŋ;

Paraok: ʔuah, Mok: ʔek

Khmuic: Yuan: ès, Xieng-Khuang: ʔɛh

Most likely, the Proto-MK form was *ʔas.

"excrement"

Vietic:

- VN: cút

- Maleng: kic, kíc; Rục: kíc, (Ferlus, 1991: Arem
kíc, Sach: kíc1)

- Thavưng: kík

The item has been replaced in Mường: eg. M. Khoi: ʔɛ:4, but can be reconstructed as PVietic *kíc or *kíc, (Ferlus, 1991: *kíc)

In the rest of MK, this pan-Austroasiatic etymon has been identified for a long time (Schmidt, 1905:16-7, Skeat and Blagden, 1906: D114, B161), and is found in both branches of Munda, with forms like: f'j in the Northern branch, and like: ik in the Southern (Pinnow, 1959:85), where predictable initial glottal stops are not noted; it is absent from most of Northern MK, except Khasi: ʔec. It is omnipresent in the Eastern MK division, except for the Katuic branch where it is replaced, as in Kuay: ʔo:

Khmer: ʔac

Pearic: Chong, Song: ʔíc

Bahnaric: Tarieng, Brao, Cheng, Bahnar: ʔic, Hrê: ʔíc
(Chrau: ʔac < Khmer)

It is also omnipresent in the Southern division:

Proto-Monic *ʔic (Diffloth, 1984:89), Nyah Kur: ʔíc,

Mon: ʔoic

Aslian⁶: P-Semai *ʔəc, Temiar: ʔej, Che'Wong: ʔəc,

Semelai: ʔəc, MahMeri: ʔoʔoc

Nicobarese: Nancowry: ʔãc, Central: ʔac

The Proto-MK was *ʔVc, with V as a short, non-back vowel whose height is not yet determined.

"to have"

Vietic:

- VN: có, Mường Khoi: kɔ3

- Maleng: kɔʔ, Rục: kó

This item is much less widespread, and presents semantic and phonological problems; it may have to be later rejected:

Aslian: Jah-Hut: ʔoʔ, Che'Wong, Bateg: wəʔ, Kensiw,

Jahai: weʔ "to have"

Monic: Old-Mon: woʔ "this"⁷

All these words uniformly have an initial ʔ- in every language of the Austroasiatic phylum except in the Vietic branch, and I see no way of invoking presyllabic or prefixed material to explain the initial *k- found there. It goes without saying that there are also many MK etyma where Vietic *k- corresponds to *k- everywhere else, as in "child": *kɔ:n, "arrow": *kam, "to harvest, cut": *kac. Cases of Vietic initial *ʔ- corresponding to ʔ- elsewhere also exist, but often present problems of indirect cognacy (intra-MK borrowing or back loans) or iconicity. So we propose, tentatively, that the initial *k- of Vietic in the words above is the regular outcome, perhaps only in certain phonological contexts which remain to be identified⁸, of an earlier *ʔ- attested everywhere else in MK and even Austroasiatic.

There are other etyma where Vietic *k- corresponds to ʔ- elsewhere in Mon-Khmer, but these are words reconstructed with initial clusters or

presyllables at an early period, which later became monosyllabic; this question will require fuller treatment elsewhere; they include: "to winnow" (Mường Khoi: kum³, Tum: kom[?], cf. Khmer ʔom), "molar tooth" (VN cãm, Aho: kãm, cf. Katu: taʔa:m, Hrê: daʔem), and "mortar" (VN cối, Thavung ko:l[?], cf. Proto-Monic *knʔɿ), Diffloth 1984: N212 p.127)

The Proto-Katuic-Vietic *-s-/-h-/-ʔ- correspondence.

There is another set of etyma where a non-initial PMK *-ʔ- turns up in Vietic languages, again surprisingly, as an *-s-. The Katuic reflex in cognate words is an *-h-, which draws us to conclude that the evolution was probably *-ʔ- > *-h- > *-s-.

Well documented examples are few for the moment but, I think, etymologically sound:

"centipede"

Vietic:

- Thavung: kasɿ:p, Aho: kasɿ:p
- Maleng: kasɿ:p
- Tum: li:p-sɿ:p

The Rục word for "centipede": kasɿ:t, would be a possible cognate except for the final consonant, and VN rết "centipede" is directly related to it by spirantisation of medial *-s- to VN r-. Whether or not the finals of these two words can be explained is not really crucial, since the Thavung, Maleng and Tum words are evidently cognate to the Mon-Khmer word for "centipede":

Katuic: Pacoh: kahe:p, Kuay: kahɛ:p, Katu: kahip, Bru: kalheip, Sô: rahɿ:p

Bahnaric: Sre: kalʔe:p, Bahnar: kəʔɛ:p, Cheng: kaye:p,

Tarieng: siji:p, Hrê: kajɿp, Sedang: kəcep

Khmer: kʔæp

Northern MK Division:

Khmu: kʔi:p

Palaungic: Lamet: siʔi:p, Lawa: saʔaip

Southern MK Division:

Nicobarese: Nancowry: kaʔiáp

Aslian: P.-Semai: kʔɛ:p, Temiar: kʔɛb, Semelai: kaʔip,

Jah-Hut: kʔɛp

Since there are Mon-Khmer etyma with *kʔ- initials (eg. "crow" *kʔa:k) which do not yield an -h- in Katuic nor an *-s- in Vietic, the conditioning factor for the Katuic-Vietic innovation must be more specific. I would suggest for "centipede" a reconstruction like PMK *kaʔɛ:p supported by the Bru, the Sre, and indirectly the Sô evidence. Then, Proto-MK *-ʔ- would shift to *-h- in Katuic and eventually to *-s- in Vietic, only after a non-initial (or vocalic) liquid.

"bone"

Vietic:

- VN: xương

- Mường Khoi: sɤ̃əŋ2

- Maleng: sə:ŋ, Rục: sə:ŋ 11, Arem (Ferlus 1991):

siaŋ

- Tum: siaŋ

Ferlus (1991) proposes *jʔa:ŋ for Proto-Vietic, but I think this word displays another example of Proto-Vietic *-s- corresponding to Katuic *-h-:

Katuic: Kuay, Bru, Pacoh: ŋhə:ŋ, Katu: ŋhə:ŋ

This *-h- in turn corresponds to a non-initial *-ʔ- in the rest of Mon-Khmer, except Bahnaric⁹:

Khmeric: cʔəŋ

Northern MK division:

Khmuic: Khmu: cʔa:ŋ, Mal: siʔiaŋ

Palaungic: Ta-ang: kə(n)ʔaŋ, Rumi: kəʔa:ŋ, Paraok:

siʔəŋ, Lamet: cəŋʔə:ŋ

Southern MK Division:

Aslian: P.-Semai: *jʔa:gʷ, Temiar: jʔa:k, Che'Wong:

jʔegʷ, Semelai: jʔæŋ, MahMeri: jʔak

Nicobarese: Nancowry: ʔuŋʔiŋ

Here again, there are reasons for positing a complex initial containing more than two segments, and reconstructing PMK *jŋʔa:/iəŋ see Diffloth (1990). The conditioning factor seems to be a preceding non-initial (or vocalic) nasal. This is confirmed by another etymon, for which, unfortunately I have not found a Vietic cognate:

"to cough"

Katuic: Kuay: ŋhɔʔ, Bru: ŋhɔk, Yeu: kahɔʔ, Pacoh:

kahɔ:ʔ

Khmer: kʔo:ʔ

Bahnaric: Brao: kʔɔk, Hrê: kʔɔk

Pearic: Chong: kəmp'k, Song: khamo'k (with -m- infix and loss of *-ʔ-)

Here, one would reconstruct PMK *k(-N)ʔɔk, with an aspectual nasal infix in Katuic and Pearic. The evolution *-ʔ- > *-h- > *-s- does seem phonetically odd, but it is restricted to a specific context, i.e. after a vocalic continuant which is itself preceded by another segment. It will have to be placed in the context of the re-syllabification processes these languages have been undergoing.

The second stage of this evolution, which affects only Vietic: P-Katuic-Vietic *-h- > P-Vietic *-s-, is attested in other items which do not descend from earlier *-ʔ- but from clusters containing a non-initial *-h-, as in:

"blood"

Vietic:

- Maleng: ʔasa:mʔ, Rục: asa:mʔ 4

- Rho: pasa:mʔ

Ferlus (1991) suggests VN tuəm "to ooze" as a cognate although the vowel and the meaning are out of line, and a sǎc tone would have been expected.

Katuic: Kuay, Bru: ŋhə:m, Pacoh: ʔahə:m, Katu: ʔahə:m

Bahnaric: Brao, Bahnar: pha:m, South Bahnaric

*məha:m

Khmer: cheam

Pearic: Chong: məhə:m, Sog: hə:m

Southern MK Division:

Monic: *chim (Diffloth, 1984:103)

Aslian: P.Semai: *bhi:b^m, Jah-Hut: nihim, Ch'Wong:

bhi:b^m, Semeali: mhæm

Northern MK Division:

Khmuic: Khmu: mə:m, Prai: miəm (with regular Khmuic loss of *-h-)

Other Northern MK languages have forms going back to *sna:m with different affixes. With the data at hand one could reconstruct "blood" as PMK *j(-m-)ha:/iəm.

"to fart"

Vietic:

- VN: rǎm

- Tum: som?

Except for the vowel of VN, which is unexplained (ludic distortion?), one could reconstruct a Proto-Vietic *psom?.

Khmer: phaom

Pearic: Chong, Song, Samre: phy:m

Bahnaric: Bahnar, Brao, Stieng, Biat: pho:m, Sedang:

phúam, Halang: phu:m

Northern MK Division:

Palaungic: Ta-ang: pham, Ka-ang: pho:m, Na-ang: pha:m,

Rumai: pham, Paraok: phum, Lawa: phəum

Khmuic: Khmu, Mal: pu:m (again with Khmuic loss of *-

h-)

Southern MK Division:

Aslian: P.Semai *hɛ:b^m, Lanoh: phe:m, Jah-Hut: (p)hũm,
Che'Wong: pəŋhəb^m

The proto MK word was evidently of the form
*ph0:m, with 0: as a long back rounded vowel of yet
undetermined height.

"breath, to breathe"

Vietic:

- Thavung: pəsɿ:m?
- Maleng: pəsɿ:m?

Katuic: Kuay: pəhɔ:m, Bru, Sô: pahɿ:m, Pakoh: palhɔ:m
Bahnar: Tarieng: cəhɿ:m, Hrê: hihəm
Khmer: ជំងឺរាង

Southern MK Division:

Aslian: P.Semai: *ləhãm, Jah-Hut: grhɨm
Nicobarese: Car: ʔuhɔ:m

Northern MK Division:

Palaungic: Ta-ang: phɨm, Ka-ang: phɛ:m, Rumi: phɛ:m,
Parak: mphɔm, Lamet: phɛ:m, U: ʔarsem53, Mok:
ge:m

Khasi: mɔ̃se:m (cf. Khmuic: Khmu: hrɨɨam "heart"
with a different affix)

These forms suggest a Proto-MK *pɾhə:m. The -s-
found in Khasi and Angkuic (U, Mok) is also the result of
an independent innovation *-h- > -s-, regular in Angkuic,
but poorly documented in Khasi. This begs, of course, the
question of alternate reconstructions leading to
different classifications, and to different answers to
our initial question concerning the place of Vietnamese
in Mon-Khmer.

If we reconstructed a PMK *-s- in "blood", "fart"
and "breathe", then, Angkuic, Khasi and Vietic would be
conservative and all other branches and languages would
have undergone the shift of *-s- to -h- separately, and
on distinct occasions. This is not difficult, since the

innovation is phonetically a very natural weakening. But for "centipede" and "bone", and also for "cough" the evolution *-h- to -ʔ- seems difficult phonetically, and would have to occur separately and repeatedly in Khmeric-Bahnaric, and in the Northern and the Southern Division. Whichever reconstruction is adopted, some phonetically unusual changes would have to take place. And the fact of Vietic being conservative does not help us in placing it anywhere within or outside Mon-Khmer, it simply leaves the question unanswered.

My first proposal, namely, that Vietic and Katuic form a subdivision within Eastern Mon-Khmer, is more specific and easier to disprove. I would therefore choose it for the time being.

It also agrees with the general impression one gathers from inspecting the lexicon of Vietic languages like Thavưng, Maleng and Tum, namely, that there are many unique and remarkable agreements between them and the Katuic languages. These agreements do not seem to favor one sub-branch of Katuic or another, but are distributed throughout the branch. There are also cases of unique lexical agreements between Mường and the Katuic languages, and occasionally even of Vietnamese agreeing alone with the Katuic branch. Arguments about lexical replacement require ample space and time to be presented convincingly, which I hope to be able to do on another occasion.

Notes

¹ The term is originally due to Hayes (1984), it currently includes: - a Việt-Mường sub-branch consisting of all forms of Vietnamese, Nguôn and all Mường dialects, - a Chứt sub-branch which includes Arem, Rục, Sách, Mày, Mã Liêng and Maleng, - a Thavưng sub-branch which includes Thavưng, Ahlau and Ahor, and - a Tum sub-

branch which includes Tum, Cuối, Đan Lai and Poọng. The Maleng language, which is what Ferlus calls "pakatan", is spoken in a dozen villages of Bolikhamsay Province in Laos, including *bân pàak ʔatan*, near the mouth of the Atan river, which I visited in 1990; the language does have a name, *malɛŋ*, and I will use it rather than "pakatan"; Ferlus' "phone soung" is the name of a village, *pho:n sũ:ŋ*, which no longer exists but used to be speaking Ahoʔ (*ʔahə:*) according to Ahoʔ speakers from other villages, a statement confirmed by the linguistic data I collected in the area.

² In that paper, Ferlus uses the term "viet-muong" for what we call Vietic, and introduced the term "muong-viêt", which he credits to "les linguistes vietnamiens", for what is usually, and here, called Việt-mường.

³ During a research project on an etymological lexicon of Khmer financed partly by the National Endowment for the Humanities, and partly by the National Science Foundation, I was fortunate to work with speakers of Thavưng and Maleng for several months in Thailand (1990). At the end of that year, I went, accompanied by Dr. Thongphet Kingxada of the Lao Comittee of Social Sciences, to the area of Bolikhamsay Province (Laos) where Thavưng, Maleng and Tum are spoken, and was able to make recordings of these languages.

⁴ Ferlus (1991) gives *kujh1*; even if one used his tone notation for Thavưng which is not a tone but a register language, the word should have been noted *kujh2*. I have personally recorded and checked this and other Thavưng forms on many occasions, both with refugees in Thailand, and in their original village in Laos.

⁵ Shorto (1971:25) suggests that Palaung (i.e. *Ta-ang*): *kəʔwɔ̃*, and Praok: *siao* could be cognates; however, both words mean "to be warm", and cognates within Palaungic: Parack: *siʔy*, Bulang: *səʔél*, Phang: *səʔ+h*, Phalok: *ləʔuy*

show that the proto-final is an *-r, not an *-s, thereby excluding cognacy.

⁶ In most of Aslian and in Nicobarese, the forms also mean "stomach", a semantic development which is one of several innovations common to these two sub-branches of the Southern MK division.

⁷ Christian Bauer (pers. comm. 1989) suggests that this deictic may also have the syntactic functions of an existential verb.

⁸ There is also a possibility that this correspondence may represent the reflex of a yet unidentified proto-MK element, let us say a uvular *q-; in this case, the Vietic merger of this *q- with Proto-Vietic *k- could have occurred at any proto-period. With either reconstruction, the evidence given here is only useful for identifying Vietic, not for clarifying its position within MK.

⁹ The Bahnaric cognates: South Bahnaric *ntiŋ, West Bahnaric *kətiŋ, North Bahnaric *kasiŋ, present a special and yet unexplained evolution, as mentioned in Diffloth, 1990.

References

- Diffloth, G. 1974. "Austroasiatic languages." *Encyclopedia Britannica* III. Chicago.
- 1979. "Aslian languages and Southeast Asian prehistory." *Federated Museums Journal* n.s. 24: 3-16.
- 1984. *The Dvaravati Old-Mon language and Nyah-Kur*. Bangkok: Chulalongkorn University Printing House. 402 pp.
- 1990. "What happened to Austric?" *Mon-Khmer Studies* 16: 1-11.
- Ferlus, M. 1974. "Problèmes de mutations consonantiques en Thavung." *Bulletin de la Société de Linguistique de Paris* 69.1: 311-23.

- 1991. "Vocalisme du Proto-Viet-Muong," paper circulated at the 24th International Conference on Sino-Tibetan Languages and Linguistics, Bangkok.
- Forbes, C. J. F. S. 1878. "On the connection of the Mons of Pegu with the Koles of Central India." *Journal of the Royal Asiatic Society*, n.s. 10.
- 1881 *Comparative grammar of the languages of Further India: a fragment*. London: W. H. Allen. 192 pp.
- Hayes, L. V. H. 1984. "The register system of Thavung." *Mon-Khmer Studies* 12: 91-122.
- Logan, J. R. 1856. *Ethnology of the Indo-Pacific Islands* Part III, Chapter VI, Appendix B: "Comparative vocabulary of miscellaneous words of the Mon-Anam formation" *Journal of the Indian Archipelago*, n.s. 1.
- Pinnow, H.-J. 1959. *Versuch einer Lautlehre der Kharia Sprache*. Wiesbade: O. Harrassowitz. 514 pp.
- Schmidt, P. W. "Grundzuge einer Lautlehre der Mon-Khmer Sprachen." *Denkschr. Acad. Wiss. Wien Phil-hist kl.* 51: 1-233.
- Shorto, H. L. 1971. *A Dictionary of the Mon Inscriptions*. London: Oxford University Press. 406 pp.
- Skeat, W. W. and Blagden, C. O. 1906. *Pagan Races of the Malay Peninsula* Vol. 2. London: Macmillan.
- Taberd, J. L. 1838. Introduction to: Pigneaux, P. J. and Taberd, J. L. *Dictionarium Anamitico-Latinum*. Serampore: Marshman.
- Thomas, D. D. and Headley, R. K. 1970. "More on Mon-Khmer subgroupings." *Lingua* 25.4: 398-418.