PROTO-KATUIC PHONOLOGY AND THE
SUB-GROUPING OF MON-KHMER LANGUAGES[1]

Paul Sidwell
Australian National University
paul.sidwell@anu.edu.au

Summary

The Katuic languages are a branch of the Mon-Khmer (MK) family spoken by more than one million people living in Thailand, Cambodia, Laos and Vietnam. While the majority of Katuic speakers live in eastern Thailand and Cambodia, the greatest diversity of Katuic languages lies in the Salavan and Sekong provinces of Laos and adjacent border areas of Vietnam, part of a complex patchwork of small ethnic communities. From a comparative-historical point of view Katuic has particular importance, as between them the languages appear to have conserved some very ancient phonological and lexical features. At the same time some Katuic languages have been remarkably innovative and developed some of the richest vowel systems in the world. The recent advances in the reconstruction of Proto Katuic (Sidwell 2005) potentially allow us to investigate the sub-grouping of Katuic within Mon-Khmer on the basis of comparative phonology. However, the results are somewhat ambiguous, and do not support any special sub-grouping of Katuic within Mon-Khmer.

Classification of Katuic with the Mon-Khmer family

During the first major phase of comparative-historical work on the MK languages, which lasted into the 1960s (effectively beginning with the work of Schmidt (1901, 1904, 1905 etc.) until Pinnow (1959) and Shafer (1952, 1965)), there was no coherent account of the real extent and internal structure of the Mon-Khmer family.

Thomas and Headley (1970) established a new paradigm when they successfully applied lexicostatistics to the emerging body of new field data, distinguishing nine branches: *Pearic, Khmer, Bahnaric, Katuic, Khmuic, Monic, Palaungic, Khasi* and *Viet-Mương*. Adding *Aslian* and *Nicobarese* (not examined by Thomas and Headley although already long recognised as MK), Diffloth's (1974) expanded listing became the received classification[2]. The Munda languages of India are also generally recognised as related to MK, although opinion is divided over how close that relationship is. All together they are recognised as forming the Austroasiatic phylum, but in this paper I am only concerned with analysis up to the MK level.

---

[1] I would like to thanks Mark Alves for valuable advice and comment on this paper, in addition to his ongoing cooperation and encouragement in respect of my broader Mon-Khmer research activities. This paper was also made possible by assistance from the Department of Linguistics of the Max Planck Institute (Leipzig) in the form of stipend support during 2005.

[2] Since the 1980s some minor languages of China have come to light that may or may not constitute a new branch.
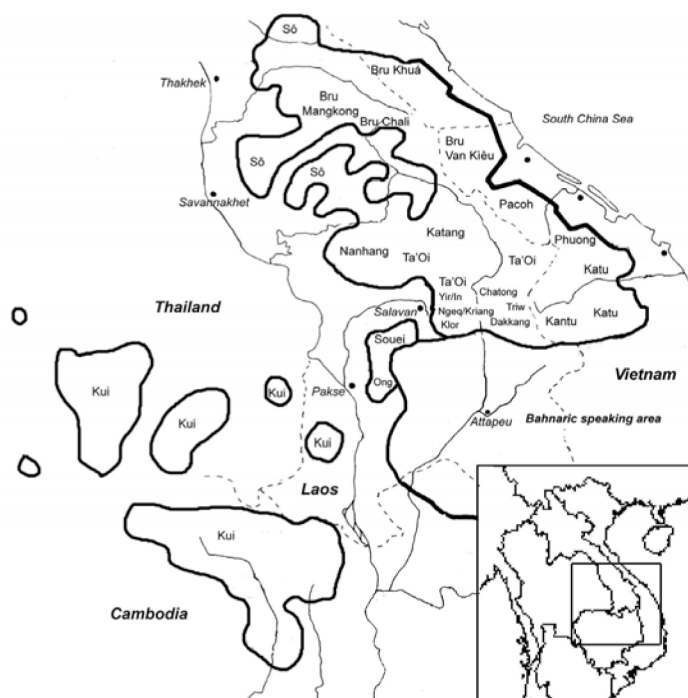
Figure 1: *Map showing approximate locations of Katuic languages*

There is no yet general agreement among specialists concerning how the MK branches further sub-group with each other. Peiros (1998) also applying lexicostistics, identified five coordinate MK groups (plus Munda):

1. Central (Bahnaric, Katuic, Aslian, Monic)
2. Vietic
3. Northern (Palaung-Wa, Khmuic)
4. Khmer
5. Khasi
6. Munda

Figure 2: *Austroasiatic classification of Peiros (1998)*

In contrast Diffloth has developed a model, based largely on identifying lexical innovations (lexicostistics counts only retentions, at least as practiced by Peiros), that distinguishes only three coordinate MK groups (plus Munda), which in their most recent incarnation (2005) are designated "Khasi-Khamuic", "Khmero-Vietic" and "Nico-Monic" (see Figure 2).

One notes that the placement of Katuic is radically different in each the above schema: Peiros grouping it with Bahnaric, Aslian & Monic, while Diffloth sub-groups it with Vietic (*Viet-Mương*). Alves (this volume) also cites lexical evidence in favour of Diffloth's hypothesis. However, it is striking that two different lexical approaches, one based upon identifying retentions (Peiros) and the other innovations (Diffloth) reach such contrary results. In these circumstances a fresh approach is called for. I have recently completed a comprehensive phonological reconstruction of Proto-Katuic

194

(Sidwell 2005), and a preliminary phonological reconstruction of Proto-Mon-Khmer is also available (Shorto f.c.), so an attempt at a classification based upon historical phonology is possible. Diffloth (1991) did invoke some phonological data in respect of the claim of Vieto-Katuic subgrouping, so this will be discussed first before moving on to presenting my historical-phonological analysis.
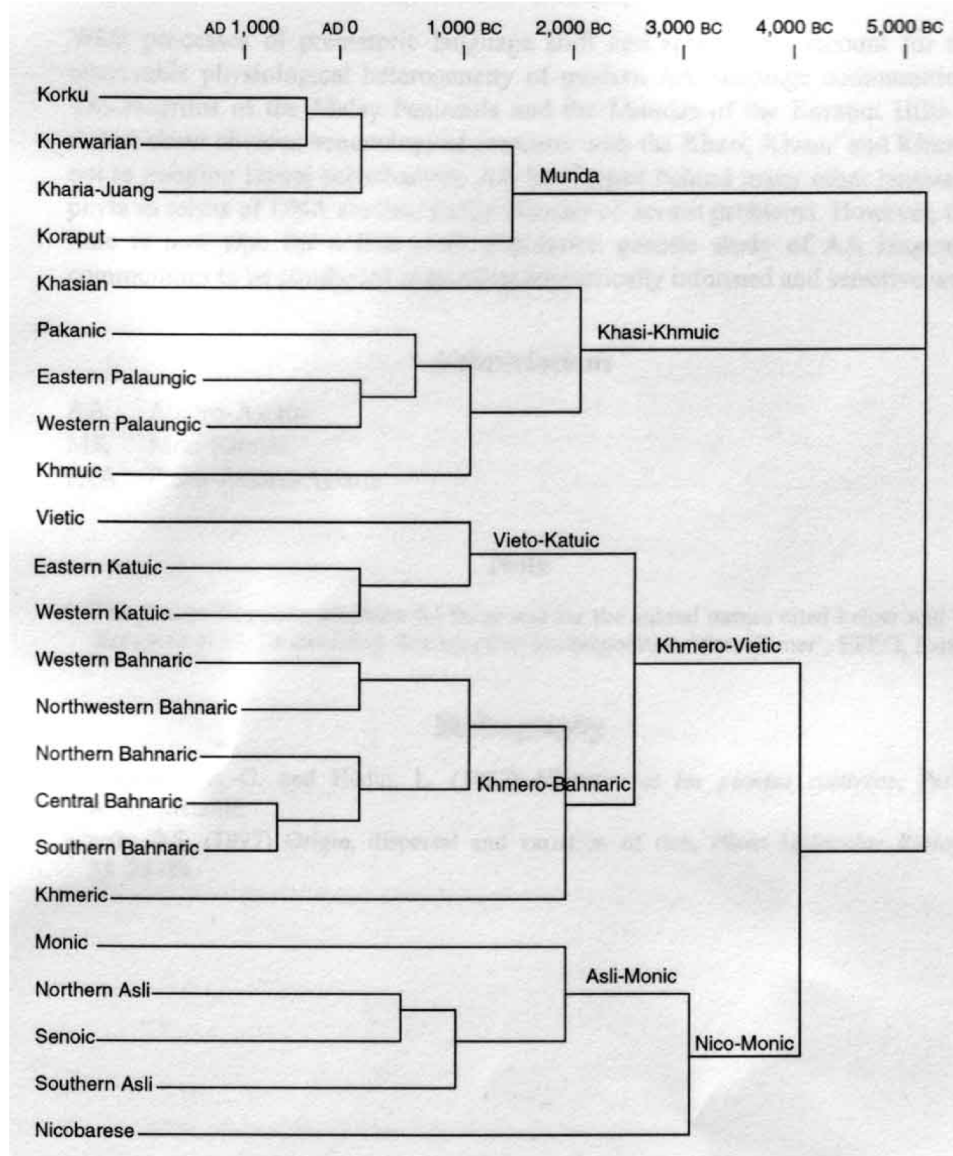

Figure 2: *Sub-grouping of Mon-Khmer (and Austroasiatic) languages according to Diffloth (2005)*

Diffloth's Vieto-Katuic hypothesis

Diffloth (1991) proposes that in some cases, conditioned by unknown factors, prevocalic *ʔ became *h in Proto-Vieto-Katuic, and subsequently *s in Vietic. The changes are attested in a small number of well distributed MK etyma, e.g.:

Khmer	Katuic	Vietic
chʔɤŋ 'bone'	Katu *ŋhaaŋ*	Ruc *saaŋ*, Viet. *xương*
kʔaep 'centipede'	Katu *kahip*	Maleng *kasiip*

It is argued that phonetically it is more likely for Vietic **s** to have come from a Proto-Katuic-Vietic **h**, rather than independently from PEMK **ʔ**. However, the potential counter-examples to this rule outnumber the examples, e.g. there was no such change in etyma such as PK **kaʔaaʔ** 'crow', **ʔŋʔuur** 'wasp', **sʔaap** 'yawn',**ʔuus** 'fire, firewood' and others which have solid MK etymologies. In the absence of a motivated phonetic change it is very difficult to explain this **h:s** corespondence. My own hypothesis is that it resulted from some cluster reduction, but in what language and at what stage? The very limited lexical attestations of the change may rather indicate that the affected words were borrowed, perhaps a case of dialect mixing or a language shift. Borrowings from one or more related language can create small odd clusters of correspondences that stand out as in this case. On balance I suggest that this **h:s** correspondence is simply too ambiguous in its significance to base a classification upon it, and instead I suggest that the strongest weight should be placed upon clearly motivated systemic changes. This requires a well developed model of the historical phonology.

The Proto-Katuic consonants and PMK

The comparative phonological reconstruction of Proto Katuic (Sidwell 2005) reveals a mostly conservative system. The proto-consonant inventory is equivalent to what Shorto (f.c.) posits for PMK, with the addition that a palatal implosive is indicated, which I provisionally reconstruct to PMK—it appears that Shorto did not have extensive Katuic sources and missed the distinction.

*p	*t	*c	*k	*ʔ
*b	*d	*ɟ	*g	
*ɓ	*ɗ	*ʄ		
*m	*n	*ɲ	*ŋ	
*w	*r *l	*j		
	*s			*h

Table 2: *Proto Katuic consonant inventory*

The major changes in the consonant system that occurred as PMK split into separate branches of two types: 1) reorganisation of the system of oral stops, and 2) lenitions/mergers of palatals and fricatives. The sonorants appear to have been much more stable, and are therefore less useful for discussion of sub-grouping.

It is significant that Katuic retains the PMK three stop series intact—in fact it survives unaltered in various Katu dialects, but is reduced to two series by the merger of plain voiced and voiceless stops in the rest of Katuic. Outside of Katuic the old three series distinction is maintained in Bahnar, and in the orthography of Old Mon (see Shorto 1971 *Introduction* for discussion). In most MK languages it has reduced to two