

THE MON-KHMER SUBSTRATE IN CHAMIC: CHAMIC, BAHNARIC & KATUIC CONTACT

Paul Sidwell

Centre for Research in Computational Linguistics (Bangkok)
& Australian National University
<paulsidwell@yahoo.com>

1 Introduction

The publication of Thurgood's (1999, henceforth "Thurgood") comparative-historical study of the Chamic languages puts the field of historical Chamic studies on a very solid foundation, yet also raises many new and important questions in relation to the history of language contact in Mainland Southeast Asia. Recent advances in the phonological and lexical reconstruction of Bahnaric and Katuic (Sidwell 2002, Sidwell & Jacq 2003, Sidwell 2005)¹ have clarified the internal genetic classifications of these groups, allowing inferences to be made about their historical locations and contacts with Proto-Chamic (henceforth "PC"). Comparison with the PC lexicon strongly suggests an ancient Sprachbund involving especially strong PC influence upon North and Central Bahnaric, but rather less direct contact with Katuic. The source of a substantial proportion of the apparently MK stratum in PC remains a mystery, and may well have been an extinct language (or languages) spoken in areas now home to Katuic or Bahnaric speakers.

2 Champa & Proto-Chamic

The Chamic languages are a Malayo-Polynesian sub-grouping, with speakers located today in Vietnam, Cambodia, Hainan Island (China) and Sumatra (Aceh Province of Indonesia). It appears that more than 2000 years ago a people speaking a language closely related to Proto-Malayic began to colonise the central Vietnam coast, and by the middle of the first Millennium they had established a series of allied coastal polities that we refer to conventionally as Champa. For around a thousand years the Cham settlements were important religious, military and trading centre of the region, rivaling Cambodia and Vietnam in importance. However, Champa entered an eight century long period of intermittent decline in the face of pressure from Vietnam, beginning with sacking of the northernmost centre of Indrapura (located near present day Da Nang) in 982, ultimately terminating in a rump Cham state persisting at Phan Rang into the 1830s. As the Vietnamese advanced southward Cham speakers variously resettled into the highlands, along the Mekong, and abroad by sea. Today we find the ironic situation that Chamic languages are virtually unknown in the zone of their ancient glory but are more or less alive and well on it's geographical periphery.

¹ These references were unpublished but circulating in manuscript form at the time of the SEALSXII meeting. Taking advantage of the unfortunate delay in the publication of these proceedings I took the liberty of updating the text of this paper to reflect the reality of these publications, as well as useful comments I received from colleagues, particularly Graham Thurgood and Anthony Grant.

It is reasonably assumed that during the first Millennium the ethnic Cham peoples spoke a more or less mutually intelligible language reconstructable as PC. A version of PC was offered by Lee (1966) and Burnham (1976) based upon Highland Chamic languages. Thurgood revised and expanded the reconstruction, taking into account a more representative range of Chamic languages, including Acehnese, which Thurgood controversially includes within the family. This proto-language has extraordinary characteristics: although self-evidently related to Malayic, it contains hundreds of Mainland SEAsian loans (many more than does Malay), and underwent partial phonological restructuring towards Mainland SEAsian type. This creates a very special opportunity for comparative linguistics, including the possibility that the inferred ad-strata may be subjected to internal and comparative investigation, revealing their identities and periodisation of contact influence.

Which were these contact language(s)? —the principle candidate according to Thurgood is Bahnaric, and secondarily Katuic. This is indicated in part by the fact that of the 267 MK-PC lexical parallels identified by Thurgood, the overwhelming majority have excellent Bahnaric matches. For the purposes of this paper I will assume that prolonged and intense Bahnaric-Chamic linguistic contact is accepted as a given, the outstanding issue being to identify the direction of Bahnaric-Chamic borrowings. Understandably Thurgood suggests that:

The early arriving pre-Chamic peoples most likely landed south of Danang and thus probably encountered Bahnarics. Given the restructuring of the arriving Austronesian language that took place, these pre-Chamic people must have become socially dominant, with this dominance leading many most probably Bahnaric-speaking people to shift to Cham, [p.251]

This conclusion is dependent on showing that the substantial common Bahnaric-Chamic lexicon is borrowed from Bahnaric. None-the-less it is clear that these two groups have (and continue) to experience significant linguistic contact. However, it is not so clear just how extensive Chamic-Katuic contact may have been. Thurgood's appendices list some 62 Katuic parallels, which appear to be the principle motivation for Thurgood's confidently assertion that:

...early contact between Chamic and Katuic speakers needs to be recognized anyway to account for the presence of Katuic borrowings in Chamic and Chamic borrowings in Katuic. (p.241)

and

Of particular interest are the Pacoh forms as they indicate intense contact between ancient Chams and the Katuic. (p.240)

Reference is also made to Reid's (1994) claim of an apparent morphological strata in Katuic, which Thurgood presumes is due to Chamic influence. I don't wish to offer a detailed critique of Reid's position here, but I read that paper as arguing for an Austric (and hence relation at much greater time depth) explanation for the MK-An morphological

parallels (which were earlier raised by Schmidt 1905, 1906). Notwithstanding the higher proportion of Chamic loans into Pacoh (clearly a local phenomenon), Katuic morphology is derived via its MK heritage, and not borrowed via later An contacts.

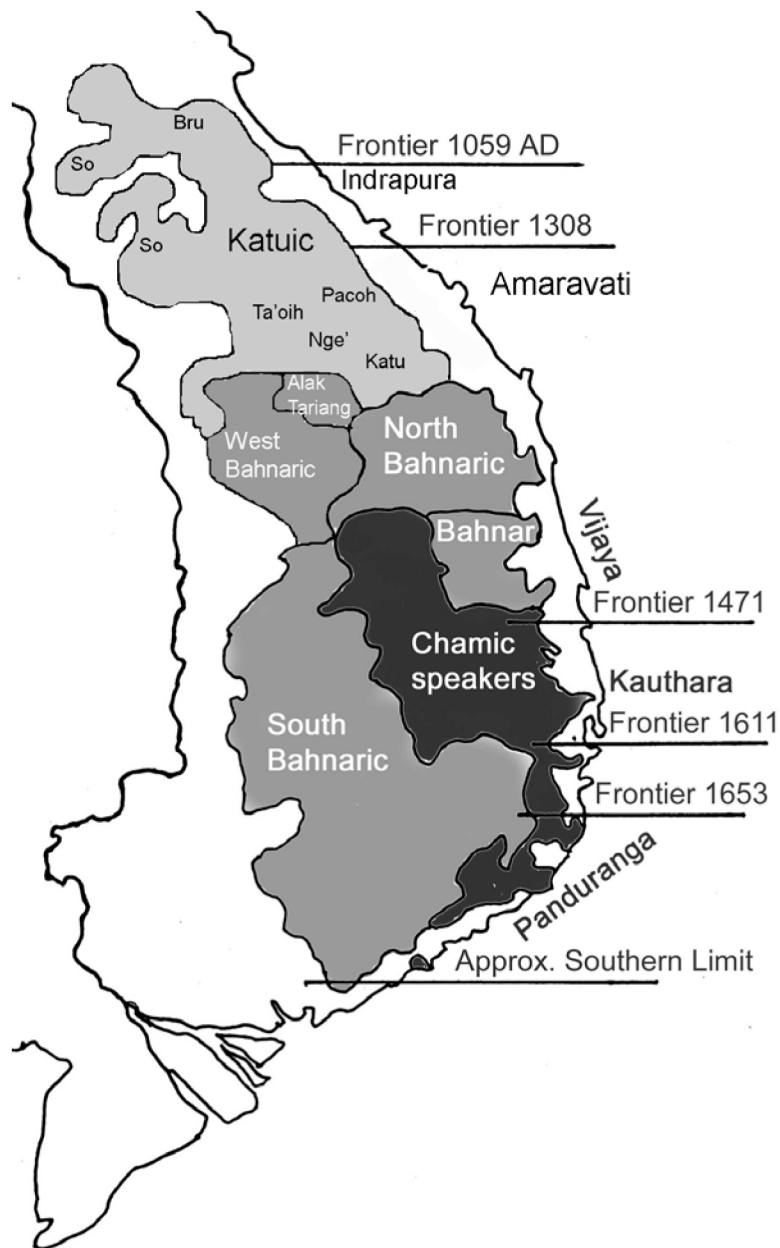


Figure 1: Map of approximate Bahnaric, Katuic and Chamic language distributions in Vietnam, and the historical Cham polities

I strongly suspect a geographical motivation strongly underlies ideas that Katuic speakers had an influence on ancient Cham, and I am sympathetic to the notion that we may gain some important insights by considering a broader synthesis of geographical and historical evidence in combination with linguistic data. It is tempting to place significance on the fact that the north-south distribution of Katuic and Bahnaric languages approximates

rather neatly the known extent of ancient Champa. The presently Katuic speaking areas lay adjacent to the old northern polities of Indrapura and Amaravati, while Bahnaric and remnant Chamic speaking areas are adjacent to Vijaya, Kauthara and Panduranga. Although the Katuic languages are certainly less Chamicised than the Bahnaric, this may simply be consistent with the lack of extensive Chamic resettlement into the Katuic zone, and the fact that the northern Cham polities were the first conquered by Vietnam.

Interestingly the main Highland Chamic intrusion (Ede and Jarai speaking) neatly corresponds in latitude to Kauthara, while the remnant coastal strip of Chamic speakers reflects the survival of an intact Panduranga into modern times.

We cannot assume that the modern distribution of Katuic and Bahnaric languages is unchanged from prehistoric times, but is rather the result of historical movements since the break-up of the respective proto-language speaking communities. Applying principles of dialect geography to the genetic classifications of these languages, we have a chance to suggest the homeland locations and migration paths of the daughter language speakers. These (necessarily tentative) results may then be correlated with what we know and presume concerning the geography of ancient Champa with what may be inferred from the distributions of loanwords. With this in mind I proceed to discuss the internal classifications of the Katuic and Bahnaric families.

3 The Katuic Family

There are approximately 15 Katuic languages spoken today, mostly in the area immediately north of the Bahnaric range and, by a huge number (perhaps a million?) Kui dialect speakers in a discontinuous spread from Laos, across into Isaan Thailand and northern Cambodia. The geographic centre of Katuic linguistic diversity lies in Laos in the hills east and north-east of Salavan, extending across into Vietnam.

It has been recognised consistently in comparative Katuic linguistics that Kui (also called Souei) sub-groups with Bru and So in a West-Katuic sub-branch (Thomas 1967 etc.). Phonologically the group is marked by a register split and vowel restructuring that is strongly parallel to that of Middle Khmer. Thus it is suggested that the Kui dialect chain, and the significant northern range of Bru and So must reflect historically late expansions of Katuic speakers under Khmer influence/dominance.

Although comparativists who have considered the issues using different methods (Thomas 1967, Efimov 1983, Diffloth 1982, Gainey 1985, Peiros, 1996, L-Tongkum 2000, Sidwell 2005) have offered various classifications, they have all agreed on the reality of the western sub-grouping and furthermore the identification of Katu (and dialects) as a distinct sub-group that branches from the highest node in the tree. On the basis of my recent historical phonological analysis (Sidwell 2005) I advocate the following classification into four more or less equidistant sub-groups: