

AMSTERDAM STUDIES IN THE THEORY AND
HISTORY OF LINGUISTIC SCIENCE

General Editor

E.F.K. KOERNER

(Zentrum für Allgemeine Sprachwissenschaft, Typologie
und Universalienforschung, Berlin)

Series IV – CURRENT ISSUES IN LINGUISTIC THEORY

Advisory Editorial Board

Lyle Campbell (Salt Lake City); Sheila Embleton (Toronto)
Brian D. Joseph (Columbus, Ohio); John E. Joseph (Edinburgh)
Manfred Krifka (Berlin); E. Wyn Roberts (Vancouver, B.C.)
Joseph C. Salmons (Madison, Wis.); Hans-Jürgen Sasse (Köln)

Volume 268

Thomas D. Cravens (ed.)
Variation and Reconstruction

VARIATION
AND
RECONSTRUCTION

Edited by

THOMAS D. CRAVENS
University of Wisconsin, Madison

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

VARIATION AS A REFLECTION OF CONTACT
NOTES FROM SOUTHEAST ASIA

GRAHAM THURGOOD
California State University, Chico

1. *Introduction*

Even historical linguists occasionally criticize reconstructed languages because much of the variation found synchronically in natural languages is not reflected in proto-languages. Although some variation shows up in proto-languages (see Ratliff, this volume), it is clear that significantly less variation is usually found in a proto-language than in any of the proto-languages' well-attested descendents. And, it is often true that historians could do a better job of incorporating the variation reconstructible back into the proto-language. However, to a large degree, this is as it should be: closer examination reveals that much, if not most, of the synchronic variation seen in Southeast Asian languages reflects the influences of language contact and the processes accompanying the incorporation and nativization of borrowings.

For Southeast Asian languages, as is the case with language change, it might be argued that the major source of variation is external contact, with internal variation playing but a minor role. It is impossible to find a well-attested Southeast Asian language in which contact has not played a major role, nor does it seem to be possible to find a language in which language contact has not resulted in variation. Indeed, contact-induced variation manifests itself in all the linguistic subsystems, and in various forms.

2. *The data*

Sometimes the variation reflects nothing more than an ongoing process of the nativization of the phonology of a borrowing. As a starting point, this paper examines parts of the variation in Phan Rang Cham (Blood 1967). Examination of the items in Blood's description in detail makes it clear that much, although not all of it, reflects the ongoing problems inherent in the

attempt to assimilate nonnative material, specifically Mon-Khmer loans, into contemporary Phan Rang Cham (cf. Thurgood 1999). For instance, final /-l/ and /-r/ and a retroflex /-ɾ/ vary, typically with /-n/. Most of this reflects the attempt of Cham speakers, without a final /-l/ or /-r/ in their native language, to deal with an /-r/ or /-l/ in borrowed Mon-Khmer forms. The final /-ɾ/ and /-l/ are made even more problematic by the coalescence of all three to /-n/ in the languages of the area. In short, these examples represent borrowed forms neither fully assimilated into Cham nor fully consistent with Cham phonotactics. Such variation, of course, does not show up in the reconstructions themselves (the material displaying regular sound-meaning correspondences), but it does, or at least should, show up in the discussions of the historical implications of the always present body of irregularly corresponding material (as it does here).

The Cham examples just discussed dealt with obvious, readily-recognizable borrowings. More difficult to see are patterns in which native forms have been influenced by bilingual contact. Again Cham, under the heavy influence of Vietnamese, in which the majority of Cham are bilingual, manifests this type of externally-induced variation in both borrowed and native vocabulary. The voiceless unaspirated stops are frequently unreleased utterance-finally; word-finally /-t/ and /-k/ may be replaced by a glottal stop. As is evident from spectrographic evidence, the glottal stop in final position varies with laryngealization, according to Blood, a characteristic of the languages of the area. This type of variation would not show up in a reconstruction, but might be manifested in the directionality of linguistic drift; that is, diachronically this sort of variation would most likely be seen in the movement of phonotactic systems in the direction of the systems of the contact languages.

A similar example, again from Phan Rang Cham, is the influence of the largely monosyllabic Vietnamese. Historically, as in the surrounding Mon-Khmer languages, Phan Rang Cham had words that were both disyllabic and iambic in keeping with the surrounding Mon-Khmer language. However, under the external influence of Vietnamese, this once predominantly disyllabic language is undergoing a rapid transition from disyllabic to monosyllabic. Doris Blood (1962:11) describes the phonetics of this transition graphically: “in normal speech a word that is sometimes heard as a two-syllable word is fused into one, as *pə̀rə̀w* > *p̄rə̀w* ‘new’ [...] The following variations have been observed: *pə̀rə̀w* ~ *p̄rə̀w* ~ *phirə̀w* ~ *phrə̀w* ~ *firə̀w* ~ *frə̀w*.” Blood notes both that the variation is particularly striking between the older and the younger generation and that it varies according to

social context, a reflection of the fact that the younger generation attending school has more contact with Vietnamese than does the older generation. None of this variation shows up directly in the proto-language, nor should it. Although clearly it needs to be accounted for in historical terms, the diachronic accounting is to be found in the discussion of contact phenomena, not in the reconstruction.

This is not to say that all such variation is excluded from the reconstruction. What does show up, however, is variation of the sort found in the reflexes of proto-Chamic **bahrɔw*; **barɔw* ‘new’ across the Chamic languages. Intriguingly, this same word also shows up in Malay, both with and without a reduced variant: *baharu* and *baru*. This particular form should be reconstructed as a doublet form in proto-Chamic as it not only occurs as a doublet in dialect after dialect, but it has a slightly different distribution in the shortened form: if the longer presumably more-stressed form is used, the meaning is ‘new’; if the meaning is not ‘new’ but ‘just’ (as in ‘just now’) the shortened, presumably less-stressed form is always used. For ‘new; just’ the reconstruction of a doublet is as it should be, but one suspects that other forms also displaying similar variation but with less clearly marked semantic domains should also be reconstructed as doublets: *mijah* (*jah*) ‘if, since’, *kayua* (*yua*) ‘because’, and *ɲan* ‘or’ (cf. Malay *dengan* ‘with’).

Another source of synchronic variation is the multiple borrowing of the same word either at different times or from different dialects. Tsat, an endangered Austronesian language of Hainan, provides some clear examples, having borrowed several words, once from a Mandarin dialect and once from a Min dialect, resulting in two phonologically distinct variants of what is synchronically the same word.

Zheng (1997:26) points out the existence of Tsat doublets, with the initial of one variant matching the variant typical of Mandarin dialects and the other typical of Min dialects. The initials of the Pinyin actually represent voiceless unaspirated initials, despite the orthography. The language labeled Official language is Zheng’s transcription of the Mandarin spoken by officials in Hainan, while the Hainanese is the local Min dialect. Note that other Min (sub)dialects such as Mai and Dan, also spoken on Hainan, could have been the actual donors, but in any case, it is clear that it was a Min dialect that the borrowing came from.

Pinyin	Official Lg. (Mandarin)	Tsat doublets	Hainanese (Min)	
bà	pa 23	pa 33 ?ba 33	?ba 214	'dam;
báo	pau 33	pa:u 33 ?ba:u 33	?bau 33	'wrap (surround)'
dǎng	taŋ 53	ta:ŋ 11 ?daŋ 11	?daŋ 31	'political party'

The initials suggest that the forms have been borrowed independently from the two dialects, with the tones of all three words simply reflecting the nearest Tsat equivalent. Although it would be nice to know if the two variants have developed different social values, for our purposes it is enough to say that forms such as these are not uncommon in Southeast Asia. When the doublet results not from a borrowing from another family, but through borrowing from a closely-related dialect, their identification is, of course, more difficult. In any case, it is clear that such variation is not reflected in a reconstruction of Tsat or of its ancestor proto-Chamic.

The examples thus far have come from phonology and phonotactics, mainly because the examples are typically more succinct. However, syntactic variation is also brought about through contact, with clear examples coming from both Tsat and Phan Rang Cham. The two comparative patterns of Tsat offer a relatively clear example of variation resulting from borrowing: one of the patterns is borrowed directly from Chinese. In contrast, Phan Rang Cham offers an example of variation, this time in the negation patterns, that seems to correlate with language contact; here, as with all such examples, the evidence is not as clear but it also seems convincing.

Tsat has two comparative patterns, one an inherited pattern, the other a pattern borrowed from Chinese. In the native pattern (1), Zheng (1997:75) notes that with comparatives, the word order is quality-marker-standard, that is, the quality being compared, followed by the preposition *la:u 32* (which serves as the comparative marker), followed by the standard of comparison, typically a pronoun. She gives the following examples, with Mandarin translations which follow the order marker-standard-quality.

- | | | |
|-----|----------------------------|--------------------------|
| (1) | Tsat (inherited): | Mandarin: |
| | nau 33 ma 43 la:u 32 ha 33 | tā bǐ' nǐ' pàŋ |
| | he fat exceed you | he compare you fat |
| | 'He is fatter than you.' | 'He is fatter than you.' |

- (2) Tsat (inherited):
 ?a 11 ko 33 phai 43 kha:n 21 na:i 32 la:u 32 ?a 11 thai 11
 older.brother reads books good exceed younger.brother
 'Elder brother studies more than younger brother.'

Mandarin:
 gēge xuéxí bǐ dìdì hǎo
 older.brother study compare younger.brother good
 'Elder brother studies more than younger brother.'

However, Zheng (1997:75) goes on to note that due to the influence of Mandarin, comparatives often follow the Mandarin order, using *pi 11* 'compare' (borrowed from Mandarin) to mark the comparison. She provides the examples in (3), with a Mandarin translation.

- (3) Tsat (Mandarin influenced):
 kau 33 pi 11 ha 33 tsat 24-tso 33 kiə 33 sun 33.
 I compare you short three inch
 'I am three inches shorter than you.'

Mandarin:
 wǒ bǐ nǐ āi sān cùn
 I compare you short three inch
 'I am three inches shorter than you.'

The word order and the comparative marker of the Tsat example are Mandarin. Needless to say, this synchronic variation does not appear in the proto-language, nor should it.

Phan Rang Cham has two dominant negation patterns. The more colloquial pattern is the historically-innovated use of the sentence-final *o* 'NEG' (Doris Blood 1977:40).

- (4) mi kaw poy? o. (1977:40)
 father I scold NEG
 'My father won't scold.'
- (5) min o_h dom təpià?. (6.1.120)
 but NEG speak out
 'but didn't say anything'

As Doris Blood notes, the more formal pattern in (5) involves preverbal negation, an inherited pattern.

While the synchronic distribution correlates nicely with formality distinctions, the diachronic origins of the innovated verb final negation pattern are not as obvious. Most likely, however, final negation originated in dynamics of second language acquisition: the most colloquial matches the first stage in the acquisition of negation—whole utterance negation with the negation marker occurring either initially or finally. The more formal pattern is pre-verbal negation, the second stage in second language acquisition; this more formal register reflects the older, less contact-influenced register. Both patterns of negation reconstruct to proto-Chamic (but not beyond), a reconstruction that correlates with other indications that proto-Chamic evolved during a period of intense language contact between Austronesian-speaking Chamic speakers and speakers of various Mon-Khmer languages (cf. Lee 1996).

3. *Systematic variation in proto: probably contact*

At least for the languages of Southeast Asia, non-systematic variation tentatively ascribed to proto-languages usually turns out to be contact. Certainly, despite the methodological requirement that all sound change be regular, in practice historians seem quite willing entertain the notion of variation in the proto-language: it just has not worked out in practice.

Two Tai-Kadai examples come to mind in which the variation is an immediate and ultimately accurate indicator of earlier contact. The first is Gedney's puzzle (1989[1972]), a detailed examination of what Strecker (1986:107) describes as "certain anomalous vowel correspondences" in the Tai branch of the Kadai family. In discussing Gedney's Tai puzzle, Strecker (1986) draws attention to similar patterns of vowel variation in Kam-Sui, a distinct branch of Tai-Kadai. Two things are remarkable about the vowel anomalies: first, although the anomalies are similar typologically, with few exceptions the Kam-Sui examples and the Tai examples involve different words; second, an overwhelming number of these forms are borrowings. Strecker has seven tables of forms which he specifically labels as Han (Chinese) loans while making it clear that additional forms are suspect; still others have been identified as Han loans in Thurgood (1994). In another table, Strecker (1986:108) lists four more forms along with their Austronesian counterparts; Thurgood (1994) has argued that these are Austronesian loans into Tai-Kadai. In short, the irregular correspondences that bothered Gedney are the result not of variation but of contact, more specifically, of borrowing.

A second set of data, this time with the variation in the initials, was noted in passing by Li (1977), and then written about in considerable detail by Gedney (1989 [1979]). In his paper entitled "Evidence for another series of voiced initials in proto-Tai", Gedney tentatively tried to account for roughly fifty-one forms with a pattern of otherwise unique initial and tonal correspondences. Although the anomalous nature of the data was also apparent to Li, in his *Handbook* (1977) he simply reconstructs voiceless aspirated stops for these forms in Southern Tai (that is, Southwestern (SWT) and Central Tai (CT)), but with voiced stop initials in Northern Tai (NT). As Gedney pointed out, although accurate as far as it goes, at the proto-Tai level, Li's analysis is problematic: usually a Northern Tai voiced initial obstruent corresponds to a Southern Thai voiced initial, but in some fifty or so cases it corresponds instead to a voiceless unaspirated obstruent.

Gedney attempted to solve the problem through positing a fifth series of obstruents for proto-Tai, a solution that he was clearly unhappy with, and he himself suggested that the possibility of borrowing might be looked into. And, quibbling about details aside, it seems obvious that the overwhelming majority if not all of these forms were borrowed into various Tai subgroups and languages after the breakup of proto-Tai. Of the fifty-one forms tentatively assigned to this set, forty-eight have counterparts in other language families, or have irregular correspondence patterns within one or more Tai-Kadai branches, or both.

The two cases just discussed invite a solution that involves positing internal variation; in both cases that solution would be wrong.

4. *Conclusions*

There is nothing unusual or atypical about these examples of contact-related variation in Southeast Asia. In fact, most Southeast Asian variation is the result of borrowing, language shift, or the linguistic convergence associated with bilingualism. Even in cases of apparently internally-motivated variation, the impetus for the variation is often the pressure to adjust to external language norms. In short, the majority of Southeast Asian variation is contact motivated, not internally motivated. From this, it follows that there should be less variation in the reconstructed proto-languages than in extant synchronic systems. There is a corollary, however: historical work should lose some of its excessive preoccupation with the reconstructed proto-language and pay far more attention to the irregularities in the data, as this is where we will find the bulk of the evidence for earlier contact patterns and thus the bulk of the evidence for non-linguistic history.

REFERENCES

- Blood, David L. 1967. "Phonological Units in Cham." *Anthropological Linguistics* 9.8:15-32.
- Blood, David L. 1977. "A Three-dimensional Analysis of Cham Sentences." *Papers in South East Asian Linguistics No. 4: Chamic Studies*, ed. by David Thomas, Ernest W. Lee & Nguyen Dang Liem. Pacific Linguistics Series A, No. 48:53-76.
- Blood, Doris Walker. 1962. "Reflexes of Proto-Malayo-Polynesian in Cham." *Anthropological Linguistics* 4.9:11-20.
- Blood, Doris Walker. 1977. "Clause and Sentence Final Particles in Cham." *Papers in South East Asian Linguistics No. 4: Chamic Studies*, ed. by David Thomas, Ernest W. Lee & and Nguyen Dang Liem. Pacific Linguistics Series A, No. 48, 39-51.
- Gedney, William J. 1989 [1979]. "Evidence for Another Series of Voiced Initials in Proto-Tai." *Selected Papers on Comparative Tai Studies*, ed. by Robert J. Bickner, John Hartmann, Thomas John Hudak, and Patcharin Peyasantiwong. Michigan Papers on South and Southeast Asian Studies, The University of Michigan. Number 29, 229-270.
- Gedney, William J. 1989 [1972]. "A puzzle in Comparative Tai Phonology." *Selected Papers on Comparative Tai Studies*, ed. by Robert J. Bickner, John Hartmann, Thomas John Hudak & Patcharin Peyasantiwong. Michigan Papers on South and Southeast Asian Studies, The University of Michigan. Number 29, 271-280.
- Lee, Ernest W. 1996. "Bipartite negation in Chamic." *Mon-Khmer Studies* 26:291-317.
- Li, Fangkuei. 1977. *A Handbook of Comparative Tai*. Honolulu: The University Press of Hawaii.
- Strecker, David. 1986. "Gedney's Puzzle in Kam-Sui." *Comparative Kadai: Linguistic studies beyond Tai*, ed. by Jerold A. Edmondson & David B. Solnit. Summer Institute of Linguistics and The University of Texas at Arlington Publications in Linguistics No. 86, 107-128.
- Thurgood, Graham. 1988. "Notes on the Reconstruction of Kam-Sui." *Comparative Kadai: Linguistic studies beyond Tai*, ed. by Jerold A. Edmondson & David B. Solnit. Summer Institute of Linguistics and The University of Texas at Arlington Publications in Linguistics No. 86, 179-218.
- Thurgood, Graham. 1994. "Tai-Kadai and Austronesian: The nature of the historical relationship." *Oceanic Linguistics* 33.2:345-368.
- Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two thousand years of language contact and change*. Oceanic Linguistics Special Publication No. 28. Honolulu: University of Hawai'i Press.
- Zheng Yiqing. 1997. *A study of Huihui [Tsat]*. Shanghai: Far East Publishers.