

LEXICOSTATISTIC ANALYSIS OF SOME KATUIC LANGUAGES

Brian Louis Migliazza
Lecturer
Thammasat University &
Summer Institute of Linguistics
Bangkok, Thailand

1. INTRODUCTION

A lexicostatistic comparison is made of some Thailand-based Katuic languages (specifically those that are variously labeled So and Bru) with those in Laos and Vietnam. This helps to differentiate these languages among themselves, as well as properly placing them within the larger context of Katuic languages in Southeast Asia (thus considerably enlarging the scope of the results of a previous paper given at the 24th ICSTLL). The data corpus consists of 18 wordlists (using the 281-item Southeast Asia Wordlist — revised) which were analyzed using the *Wordsurv* program. Nine of the wordlists are derived from the author's research this past year in three northeastern provinces of Thailand (Sakon Nakhon, Nakhon Phanom, and Mukdahan). Specifically, these nine wordlists represent nine villages in the main So/Bru districts of Tha-uthen, Phonsawaan, Kusuman, and Dongluang. The remaining nine wordlists are from other researchers. Below is a catalog of the various wordlists. First is the language name as reported by the researcher. Next is the village, district, province, and country of where the speaker came from. And finally, in parenthesis, is the date the wordlist was collected.

- A. Kui: Samrongthap, Surin, Thailand (1980)
- B. So: Phiangkaw, Kusuman, Sakon Nakhon, Thailand (1991)
- C. Bru: Khamphakut, Dongluang, Mukdahan, Thailand (1991)
- D. Bru: Khoksa-at, Phangkhan, Sakon Nakhon, Thailand (1988)
- E. So: Kusuman MuangKaw, Kusuman, Sakon Nakhon, Thailand (1991)
- F. So: Nakham, Phonsawaan, Nakhon Phanom, Thailand (1991)
- G. So: Huayphra, Tha-uthen, Nakhon Phanom, Thailand (1991)
- H. So: Phathay, Tha-uthen, Nakhon Phanom, Thailand (1991)
- I. Bru: Nongyang, Dongluang, Mukdahan, Thailand (1991)
- J. Bru: Tiw, Dongluang, Mukdahan, Thailand (1991)
- K. Bru: Muang, Dongluang, Mukdahan, Thailand (1991)
- L. Bru: Woen Buek, Khong Chiam, Ubon, Thailand (1981)
- M. Bru: Khe Sanh, Quang Tri, Vietnam (1968)
- N. Bru: Lawang, Kokphun, Thateng, Saravan, Laos (1978)
- O. Ngeq: Laksipha, Pakse, Laos (1978)
- P. Nkriang: Tray River, Laos (1969)

Q. Pacoh: Thua Thien, Vietnam (1972)

R. Katu (low): An Diem, Quang Nam, Vietnam (1968)

2. DEMOGRAPHICS

The following tables present the estimated population of the So and Bru peoples in Nakhon Phanom, Sakon Nakhon, and Mukdahan Provinces. The population figures are calculated by figuring an average village to consist of roughly 60 houses and each house to contain about 6 persons (van der Haak and Woykos 1990). Thus a typical village would contain approximately 360 people. Adding the figures for the three provinces together we arrive at a total of 98 'So' villages and a total population of about 35,000 people. The four districts with the highest concentrations are Dongluang (10,080), Kusuman (7,920), Phonsawaan (6,480), and Tha-uthen (2,880). This forms a broad U-shaped area starting from Tha-uthen through Phonsawaan and Kusuman, and then on around to Dongluang. A line in the village or people column indicates that there were no So or Bru in that district. An asterisk (*) by the number of people indicates that there were only a handful of 'So' reported for that area, so the entire village was not counted — thus the village number is placed in parenthesis (as in Nakae District of Nakhon Phanom, and Dontan District of Mukdahan).

NAKHON PHANOM

DISTRICT	VILLAGES	PEOPLE
Banphaeng	0.5	180
Srisongkhram	0.2	72
Nawa	—	—
Tha-uthen	8	2,880
Phonsawaan	18	6,480
Muang	4	1,440
Plapa	0.5	180
Nakae	(1)	15*
Thatphanom	—	—
Renunakhon	—	—
TOTAL	31.2	11,247

SAKON NAKHON

DISTRICT	VILLAGES	PEOPLE
Songdaw	2	720
Phangkhn	3	1,080
Phannanikhom	5	1,800
Kusuman	22	7,920
Tawngoy	---	---
Khoksrisuphan	---	---
Muang	2	720
TOTAL	34	12,240

MUKDAHAN

DISTRICT	VILLAGES	PEOPLE
Dongluang	28	10,080
Muang	2	720
Dontan	(1)	6*
Nikhomkhamsoy	3	1,080
Khamcha-i	---	---
Wanyay	---	---
Nongsung	---	---
TOTAL	33	11,886

3. LANGUAGE GROUPS

According to Smith's 1981 classification, the Katuic languages can be roughly divided into five main groups. The capital letters after the language name corresponds to the wordlists listed above in the introduction. Thus the wordlists in this sample represent each of these five groups.

A. Katu

- a. Katu (Low) [R]
- b. Kantu, High Katu

B. Pacoh

- a. Pacoh (Bo River Van Kieu)
- b. Phuang (Huu River Van Kieu) [Q]

C. Central Katuic

- a. Ta'oih
- b. Ngeq [N, O]
- c. Nkriang [P]
- d. Ong (?Tong)
- e. Inh (?Ir)
- f. Kasseng, Talieng

D. North Katuic

- a. Bru [I, J, K, L, M]
- b. Makong
- c. Bruu
- d. So [B, C, D, E, F, G, H]
- e. Sui (Sui)
- f. Kataang
- g. Siliq
- h. Lor, Klor
- i. Leun

E. West Katuic

- a. Kuy [A]
- b. Nyeu
- c. Kuay

4. LEXICOSTATISTICS

Using the *Wordsurv* program to count the apparent cognates results in the following percentages matrix. All villages are in Thailand unless indicated by VN (Vietnam) or L (Laos).

(PERCENTAGES MATRIX)

57	B so: phiangkaw																	
56	79	C bru: khamphakut																
63	81	85	D bru: khoksa-at													So		
57	78	81	87	E so: kusuman muangkaw														
58	77	84	87	90	F so: nakham													
56	76	81	82	84	87	G so: huayphra												
55	76	79	80	80	86	84	H so: phathay											
54	72	75	80	73	74	72	72	I bru: nongyang								Bru		
57	74	77	81	74	76	73	73	84	J bru: tiw									
58	73	76	82	73	75	72	71	79	86	K bru: dongluang								
61	69	70	78	71	71	71	68	74	75	73	L bru: woenbuek							
58	69	71	79	70	71	68	68	70	72	72	76	M bru: quang tri (VN)						
52	54	57	63	56	58	55	54	54	58	57	61	59	N bru: thateng (L)					
49	51	52	56	52	53	51	51	51	53	52	57	54	71	O ngeq: pakse (L)				
48	50	54	58	52	53	51	50	51	52	50	55	55	69	76	P nkr:iang (L)			
49	54	58	60	58	59	56	54	53	57	58	59	60	64	60	60	Q pacoh (VN)		
37	41	42	43	41	42	39	38	39	40	40	40	42	38	40	43	47	R katu (VN)	

These cognate percentages are imperfect estimates of the true cognate percentage. Using standard statistical procedures, *Wordsurv* computes a range of error for these measured percentages which is shown in the following variance matrix. The range of error is based on a) the cognate percentage, b) the sample size (total number of words compared), and c) the reliability code. The reliability code is a number from 1 to 4, with 1 being the highest and four being the lowest. The analyst assigns each wordlist a number based upon his judgement of the accuracy of the wordlist. Thus since most of these wordlists were collected in a quick survey-type visit to the language group, they were given the lowest reliability number of four. The only wordlists that received a higher reliability code were lists A, M, P, Q, and R. These five were given reliability ratings of one. Thus a comparison of wordlist A (Kui) with wordlist B (So of Phiangkaw) shows that the 57% cognate figure (from the percentages matrix above) could actually vary from about 53% to 61%. The range of error varies from a low of ± 2.7 (between wordlists P and Q, and M and Q) to a high of ± 6.1 (between wordlists B and O).

A kui: surin

(VARIANCE MATRIX)

4.2	B so: phiangkaw
4.2	4.0 C bru: khampakkut
3.8	3.8 3.4 D bru: khoksa-at
4.1	4.1 3.8 3.1 E so: kusuman
4.1	4.2 3.5 3.2 2.8 F so: nakham
4.2	4.3 3.8 3.7 3.5 3.1 G so: huayphra
4.2	4.3 4.0 3.9 3.8 3.2 3.4 H so: phathay
4.3	4.6 4.4 3.9 4.5 4.5 4.6 4.6 I bru: nongyang
4.2	4.5 4.2 3.8 4.5 4.3 4.6 4.6 3.5 J bru: tiw
4.1	4.6 4.2 3.7 4.5 4.4 4.6 4.6 4.0 3.3 K bru: dongluang
3.9	4.9 4.7 4.0 4.7 4.7 4.7 4.9 4.5 4.4 4.5 L bru: woenbuek
2.9	3.5 3.3 2.8 3.4 3.4 3.5 3.5 3.4 3.3 3.3 3.0 M bru (VN)
4.3	6.0 5.8 5.3 5.8 5.7 5.9 5.9 5.9 5.8 5.8 5.4 4.0 N bru (L)
4.4	6.1 5.9 5.6 5.9 5.9 6.0 6.0 6.0 6.0 6.0 5.6 4.1 4.6 O ngeq (L)
3.2	4.4 4.2 3.9 4.2 4.2 4.3 4.3 4.3 4.3 4.1 2.9 3.4 2.9 P nkriang (L)
3.1	4.2 4.0 3.8 4.0 4.0 4.1 4.2 4.2 4.1 4.0 3.9 2.7 3.7 3.8 2.7 Q pacoh (VN)
3.5	4.8 4.7 4.6 4.8 4.8 4.9 4.9 4.9 4.9 4.8 4.8 3.3 4.9 4.6 3.3 3.2 R katu (VN)

5. SUMMARY

This analysis shows that these nine *So* and *Bru* villages of Northeastern Thailand fall squarely in the North Katuic grouping and most neatly divide into either *So* or *Bru*. *So* of Phiangkaw (wordlist B) is not as closely related to *So* as the others, and what is called *Bru* of Khampakkut (wordlist C) and *Bru* of Koksa-at (wordlist D) seem to be closer to the *So* of Kusuman (wordlist E). One of the other wordlists, labeled *Bru* of Thateng in Laos (wordlist N) seems to fit more in the Central Katuic branch, closer to Ngeq and Nkriang.

Many groups of minority peoples living in Northeast Thailand have been classified as *So* speakers. Frequently these Katuic peoples are labeled *So* in the area of Kusuman district (Gainey 1985, Malai 1980) and *Bru* in the area of Dongluang district (Chollada 1986, Darunee 1986). This study confirms this division in that two large language classifications emerge: 1) *SO* -- as spoken in Tha-uthen, Phonsawaan, and Kusuman Districts; and 2) *BRU* -- as spoken in most of Dongluang District.

6. REFERENCES

- Chollada Jungprasert (1986) **Clauses in So (Bru) of Dong Luang, Mukdahan Province**. Mahidol University: M.A. thesis.
- Darunee Kulachol (1986) **The phonology of So at Amphoe Dongluang, Mukdahan Province**. Mahidol University: M.A. thesis.
- Gainey, Jerry (1985) **A comparative study of Kui, Bruu and So phonology from a genetic point of view**. Chulalongkorn University: M.A. thesis.
- van der Haak, Feikje and Brigitte Woykos (1990) 'Kui dialect survey in Surin and Sisaket'. *Mon-Khmer Studies* 16-17:109-142.
- Malai Lerthirunwong (1980) **A syntactic description of So: an Austroasiatic language in Thailand**. Mahidol University: M.A. thesis.
- Pattiya Jimreivat (1981) **Clauses and phrases in Bruu**. Mahidol University: M.A. thesis.
- Smith, Kenneth (1981) 'A lexico-statistical study of 45 Mon-Khmer languages'. In Andrew Gonzalez and David Thomas (eds), **Linguistics across continents: studies in honor of Richard S. Pittman**. Manila, Philippines: Summer Institute of Linguistics and Linguistic Society of the Philippines, 180-205.
- Wimbish, John (1989) **Wordsurv: a program for analyzing language survey word lists**. Occasional Publications in Academic Computing, Number 13. Dallas, TX: Summer Institute of Linguistics.