## 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 differentiate these languages among themselves, as well 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: Khamphakkut, Dongluang, Mukdahan, Thailand (1991)
- D. Bru: Khoksa-at, Phangkhon, 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, Kokohun, 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
Phangkhon	3	1,080
Phannanikhom	5	1,800
Kusuman	22	7,920
Tawngoy		
Khoksrisuphan		
Muang	2	720
TOTAL	34	12,240
1		

### 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] f. Kataang
- C. Central Katuic
- a. Ta'oih
- b. Ngeq [*N, 0*]
- 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. Suei (Sui)
- 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).

A kui: surin						(PERCENTAGES MATRIX)						
57 B so: phiangkaw												
56 79 C bru: khamphakkut												
ස	81	85	D b	ru:	khok	sa-a	t	_		ر S و	•	
57	78	81	87	E s	o: k	usum	an m	uang	kaw			
58	77	84	87	90	F s	o: n	akha	m		\		
56	76	81	82	84	87	G s	o: h	uayp	hra		\	
55	76	79	80	80	<b>8</b> 6	84	H s	o: p	hath	ay	Α,	~
54	72	75	80	73	74	72	72	Ib	ru:	nong	yang	
57	74	77	81	74	<b>7</b> 6	73	73	84	]J b	ru:	tiw	Bru
58	73	76	82	73	<b>7</b> 5	72	71	79	86	K b	ru:	dongluang \
61	69	70	78	71	71	71	68	74	75	73	L b	ru: woenbuek
58	æ	71	79	70	71	68	68	70	72	72	76	M bru: quang tri (VN)
52	54	57	ස	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 0 ngeq: pakse (L)
48	50	54	58	52	53	51	50	51	52	50	55	55 69 76 P nkriang (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 0) to a high of + 6.1 (between wordlists B and 0).