

What is the original relationship between Mon-Khmer and Kam-Tai?

Bao-ya CHEN
Peking University

1. Introduction

Are the similarities between some Mon-Khmer and Kam-Thai languages the result of a genetic relationship or a contact relationship? Schmidt (1905) proposed that the Austronesian and Austroasiatic families are genetically related as part of an Austric superstock. Benedict (1942 and later) countered with a proposal that Thai, Kadai, and Indonesian are all part of an Austro-Thai stock which does not include Austroasiatic.

The unity of the Kam-Thai family is now generally accepted, as a result of the work especially of Li Fang-Kuei (1943 and later). And the unity of the Austroasiatic family, including Mon-Khmer, is also generally accepted. The question is what is the relationship between them?

The purpose of this paper is to test a new variant of lexicostatistics, applying it to some Southeast Asian languages and seeing what light it sheds on the relationship between Kam-Thai and the Va-De'ang (Palaung-Wa) branch of Mon-Khmer.¹

2. Matching words in Kam-Thai and Mon-Khmer

There are many consistent correspondences in basic words between Kam-Thai and Mon-Khmer languages in southwestern China. Table 1 illustrates two of these consistent correspondences between Dai Xishuangbanna, Dai Dehong, and Va.

¹ The languages cited in this paper are **KAM-THAI: THAI GROUP:** BY: Buyi of Guizhou, Guangxi; DD: Dai of Dehong, Yunnan; DJ: Dai along the Jinsha River, Yunnan; DL: Dai of Lincan, Yunnan; DX: Dai of Xishuangbanna, Yunnan, DY Dai along the Yun River, Yunnan; TT: Thai of Thailand; ZL: Zhuang of Longzhou; ZW: Zhuang of Wuming, Guangxi. **KAM GROUP:** 2DR Dong of Rongjiang, Guizhou; MLH: Maolan of Huangjiang, Guangxi, MLL: Mulao of Luocheng, Guangxi; SS: Sui of Sandu, Guizhou. **LI GROUP:** LB: Li of Baoding, Hainan, LT Li of Tongshi, Hainan. **MON-KHMER: VA-DE'ANG GROUP:** AV Alva of Menghai, Yunnan, PL Plang of Menghai, Yunnan; VA: Va of Cangyuan, Yunnan. **SINO-TIBETAN: CHINESE GROUP:** SWC Southwestern Chinese.

TABLE 1. TWO SOUND CORRESPONDENCES BETWEEN TT, DX, DD AND VA
(the raised numbers represent the tone classes)

Meaning	TT	DX	DD	VA
	-ɔŋg	-ɔŋg	-ɔŋg	-ɔŋg
echo	kɔŋg ³	kɔŋg ³	kɔŋg ³	rɔŋg
thing	khɔŋg ¹ ; hɔŋg ¹	xɔŋg ¹	xɔŋg ¹	khɔŋg; khɔŋg
umbrella	tsɔŋg ³ ; tsɔŋg ⁵	tsɔŋg ³	tsɔŋg ³	ɔŋg
shine		thɔŋg ⁵	thɔŋg ⁵	thɔŋg
	plɔŋg ³	pɔŋg ³	pɔŋg ³	plɔŋg
Lancang river		xɔŋg ¹	xɔŋg ¹	krɔŋg
basket	kr'jɔŋg ⁶ ; khɔŋg ³		xɔŋg ³	khɔŋg
back (of body)	khɔŋg ¹	lang ¹		krɔŋg
forehead	tr'phɔŋg ²			ɔŋg
	-ɔŋg	-ɔŋg	-ɔŋg	-ɔŋg
lake	nɔŋg ¹	nɔŋg ¹	nɔŋg ¹	nhɔŋg
stream	tsn'ɔŋg ²	hɔŋg ⁶	hɔŋg ⁶	khɔŋg
mat	ɔŋg ²	ɔŋg ¹ ; hɔŋg ²	ɔŋg ¹ ; hɔŋg ²	lhɔŋg
room	hɔŋg ³	hɔŋg ³	hɔŋg ³	hɔŋg
shine	sɔŋg ⁵	sɔŋg ⁵	sɔŋg ⁵	sɔŋg
gaze	tr'ngɔŋg ⁵	tɔŋg ²	tɔŋg ⁵	tɔŋg
	-a	-a	-a	-a
rice shoot	kla ³	kla ³	kla ³	kl̩a
fish	pla ¹	pa ¹	pa ⁶	kaʔ; pa
quilt	pha ³	pha ³	pha ³	ph̩a
split	pha ⁵	pha ⁵	pha ⁵	ph̩a
mix	pha ⁵	pha ⁵	pha ⁵	ph̩a
soak	ma ⁵	ma ⁵		s'ma
five	ha ³	ha ³	ha ³	h̩a
snow			ha ¹	rha
late	la ⁴	la ³	la ³	lh̩a

Table 1 shows only two sets of consistent correspondences; many more such sets could have been adduced if time and space had permitted. How should we interpret this consistency?

3. Criteria for establishing genetic relationship

Historical linguists have generally relied on systematic sound correspondences to establish genetic relationships. Similarities or consistent changes in grammatical structure, usually taken to mean affix structure, is also taken as good evidence.

Having observed the living contact between Dai Dehong and Southwestern Chinese for several years, I found that loans from Southwestern Chinese into Dai Dehong which were borrowed in the same place at about the same time also show systematic sound correspondences between the two languages, as shown in Table 2, which shows some SWC *tɕh* : DD *ɕ* and SWC *tɕh* : DD *s* correspondences.

TABLE 2. SOUND CORRESPONDENCE BETWEEN BORROWED WORDS IN DAI DEHONG AND ORIGINAL WORDS IN SWC

Meaning	SWC	DD
district	/tɕh ⁵⁵ /	/ɕ ⁵⁵ /
flag	/tɕh ³¹ /	/ɕ ⁴² /
draw (money)	/tɕh ⁵³ /	/ɕ ⁵³ /
sledge	/tɕhiau ⁵⁵ /	/ɕiau ⁵⁵ /
to prize	/tɕhiau ²¹³ /	/ɕiau ²¹³ /
to warp	/tɕhiau ²¹³ /	/ɕiau ²¹³ /
buckwheat	/tɕhiau ³¹ /	/ɕiau ⁴² /
to invite	/tɕhin ⁵³ /	/ɕin ⁵³ /
skirt	/tɕhin ³¹ /	/ɕin ⁴² /
to advise	/tɕhian ²¹³ /	/ɕian ²¹³ /
fist	/tɕhian ³¹ /	/ɕian ⁴² /
melody	/tɕhio ³¹ /	/so ³¹ /
poor	/tɕhiong ³¹ /	/song ³¹ /

The /tɕh/ of Southwestern Chinese words corresponds to the /ɕ/ in Dehong Dai loans except for the last two items. The exception can be explained as a regular conditioned variant in that in Dehong there are no /io/ or /iong/ sequences after /ɕ/, so the initial consonant was reinterpreted as the permitted /s/.

A different problem appears when borrowing /uəi/ /iəu/ and /ən/ words from Southwestern Chinese. These appear in more than one form in Dai:

SWC	DD
/uəi/	/ui, oi/
/iəu/	/iu, eu/
/ən/	/um, ɣn/

The reason for these variants is that Dai, in some areas, is going through a merger of /oi/ and /ui/, /eu/ and /iu/, and /um/ and /ɣn/, these forms currently alternating freely. So the words being borrowed from Southwestern Chinese are showing this same alternation. As soon as this Dai sound change runs its course and settles down to one form, the loans from Chinese can be expected to do likewise, indistinguishable from native Dai words.

This situation, which is clearly a contact situation between SWC and DD, shows the fallibility of relying mainly on regular sound change to prove genetic relationship.

A different approach to discerning language relationships is that of comparing semantic shift. This approach assumes that borrowing between languages will tend to involve cultural vocabulary more than basic human vocabulary, so that basic vocabulary is more likely to reflect the genetic affiliation of a language. Swadesh (1952, 1955) drew up 100-word and 200-word lists of some presumably basic vocabulary. Attempts at refining the method for greater precision have not met with general acceptance, but the basic assumption seems to be true. The following discussion will build on this assumption.

4. More basic vs. less basic vocabulary

It follows from the above that the elements in a vocabulary list which are more basic should show a higher rate of resemblance (whether inherited or borrowed) if reflecting a genetic relation between the two languages, and the less basic elements should show a higher rate if reflecting a contact relation. †Swadesh's 100 and 200 lists give a good starting point for testing this.

4.1 *In genetic relationships*

I took Swadesh's 100 list as most basic, and call it here the 1st 100; the remainder of his 200 list I call the 2nd 100. I then took five Dai dialects, comparing them on the 1st 100 list and then on the 2nd 100 list. The results are shown in Table 3, where the first figure in each pair is the number of resemblances in the 1st 100, and the second figure is the number of resemblances in the 2nd 100. In every case the figure for the 1st 100 is considerably higher than that for the 2nd 100.

TABLE 3. DAI DIALECT RESEMBLANCE PERCENTAGES IN THE FIRST 100 / SECOND 100 LISTS

DX				
88 / 71	DD			
91 / 68	92 / 72	DY		
85 / 71	94 / 85	91 / 68	DL	
84 / 66	91 / 69	88 / 68	88 / 68	DL

Then I looked at Swadesh's figures from his 100 list and his 200 list for some historically attested European languages. These are given in Table 4.

TABLE 4. SWADESH'S COGNATE PERCENTAGES

Languages compared	100 list	200 list
Old English/Modern English	86%	77%
Old German/Modern German	89	84
Old Swedish/Modern Swedish	94	85
Latin/Modern Romanian	71	56
Latin/Modern French	74	62
Old Greek/Modern Greek	71	69

Taking Swadesh's figures, and applying the formula $x=2y-z$,² I deduced the number of resemblances in the 2nd 100 words, as shown in Table 5.

TABLE 5. SWADESH'S COMPARISONS, SHOWING 1ST 100 WORDS VS. 2ND 100 WORDS

Languages compared	1st 100 words	2nd 100 words
Old English/Modern English	86%	68%
Old German/Modern German	89	79
Old Swedish/Modern Swedish	94	76
Latin/Modern Romanian	71	41
Latin/Modern French	74	50
Old Greek/Modern Greek	71	67

In each case Table 5 shows a higher percentage or resemblance for the 1st 100 words than for the 2nd 100 words, which is to be expected by my theory since these are in each case clearly genetically related languages.

Narrowing the scope down to languages within the clearly proven Germanic group, comparing each of them with Modern English,³ we see again from Table 6 that the 1st 100 in every case has a higher percentage than the 2nd 100, as we would expect from genetically related languages.

TABLE 6. GERMANIC COGNATE PERCENTAGES WITH MODERN ENGLISH

Languages compared	1st 100 words	2nd 100 list
Old English/Modern English	87%	71%
Modern German/Modern English	64	46
Old Frisian/Modern English	72	51
Modern Dutch/Modern English	71	54
Old Icelandic/Modern English	76	57
Gothic/Modern English	58	39

Turning our attention now to Asia, we look at the resemblance (presumed cognates) figures between some dialects of Chinese as shown in Table 7. Here, again, in these genetic relationships in every case the 1st 100 has a higher percentage than the 2nd 100.

² $x=2nd\ 100\ figure, y=Swadesh\ 200\ list\ figure, z=Swadesh\ 100\ list\ figure\ (=1st\ 100)$.

³ Barnhart 1988 was taken as the standard for cognacy decisions.

TABLE 7. COGNATE PERCENTAGES BETWEEN CHINESE DIALECTS

Mandarin						
75 / 56	Hakka					
61 / 56	69 / 51	South Min				
82 / 71	72 / 52	65 / 51	Hsiang			
82 / 71	71 / 59	62 / 52	87 / 64	Wu		
76 / 73	66 / 56	62 / 49	81 / 67	78 / 72	Kan	
83 / 71	70 / 55	60 / 49	78 / 68	76 / 72	71 / 69	Yue

4.2 *In contact relationships*

We now look at the differential between the 1st 100 and the 2nd. 100 in the contact relationships between Southwestern Chinese and Dai Dehong, Dai Jinsha, and Dai Yun. Table 8 summarizes the borrowings in the 1st 100 and 2nd 100 words for each of these Dai languages (see Appendix A for the data). In this contact relationship the percentages are the reverse of those in the genetic relationships shown in Sec.4.1; the 2nd 100 figure is in each case considerably higher than the 1st 100 figure, as predicted by my theory.

TABLE 8. RESEMBLANCE PERCENTAGES BETWEEN SOUTHWESTERN CHINESE AND SOME DAI LANGUAGES.

	Dehong	Jinsha	Yun
SW. Chinese	2 / 10	0 / 9	3 / 8

5. Kam-Thai and Mon-Khmer

In the data presented so far we have seen that in each case of a genetic relationship the resemblance or cognate percentage is higher in the 1st 100 words, and in each case of a contact relationship the percentage is higher in the 2nd 100 words. Can this be taken as a universal or near-universal theory?

5.1 *Within Kam-Thai languages*

Most linguists today accept the existence of a Kam-Thai family including a Thai (Zhuang-Dai) branch, a Kam-Sui (Dong-Sui) branch, and a Li branch. I compared the 1st 100 and 2nd 100 basic vocabulary in eleven languages from the Kam-Thai family, with the results shown in Table 9.

TABLE 9. COGNATE PERCENTGES BETWEEN KAM-THAI LANGUAGES

ZW												
86/69	ZL											
90/81	78/61	BY										
78/53	80/55	72/53	DX									
76/54	72/58	72/51	88/71	DD								
61/46	54/38	56/46	52/36	48/34	DR							
56/48	52/40	52/47	51/34	48/33	74/56	MLL						
57/54	56/41	56/52	53/44	50/40	80/59	76/50	SS					
56/55	46/44	54/52	48/37	47/37	79/57	73/59	79/62	MLH				
49/27	48/25	52/25	48/26	51/31	40/18	38/18	37/21	37/19	LB			
46/32	46/24	50/27	49/29	48/30	39/21	37/18	38/24	38/22	90/90	LT		

Between these genetically related languages, as expected, the percentages for the 1st 100 are significantly higher than for the 2nd 100, except one case in which they are equal (LB:LT).

5.2 Kam-Thai and Austronesian

Applying this method to a more debated area I compared two Austronesian languages, Indonesian and Malay, with nine Kam-Thai languages, with the results as shown in Table 10.

TABLE 10. AUSTRONESIAN AND KAM-THAI RESEMBLANCES

	ZW	ZL	BY	DX	DD	DR	MLL	SS	MLB
Indonesian	14/6	14/6	13/6	15/5	15/5	14/6	14/4	11/6	13/6
Malay	12/6	11/6	10/5	12/5	12/5	11/6	12/4	10/6	11/5

Here the figures from the 1st 100 words are significantly higher (double or more) than the 2nd 100, which would indicate a genetic relationship between Kam-Thai and Austronesian. The figures are very low but are consistent.

6. Kam-Thai and Mon-Khmer

We come back now to our opening question, that of the resemblances between Kam-Thai and the Mon-Khmer languages in Yunnan. Are the resemblances a result of contact or of genetic relationship? We applied our test, using the Mon-Khmer languages Va, Plang, and Alva, comparing them with nine of the previously mentioned Kam-Thai languages, with results as shown in Table 11. The data from which these figures were obtained is given in Appendixes B and C.

TABLE 11. MON-KHMER AND KAM-THAI RESEMBLANCES

	ZW	ZL	BY	DX	DD	DR	MLL	SS	MLH	LB	LT
VA	8/15	9/11	7/13	10/16	9/17	2/11	2/7	2/11	3/9	6/8	7/7
PL	7/14	8/12	5/12	8/16	8/15	3/11	3/9	3/10	3/10	4/10	6/8
AL	7/12	8/9	6/10	9/14	8/16	2/11	2/8	2/9	3/7	2/8	4/6

Table 11 shows the 2nd 100 words significantly higher than the 1st 100 words, except one case in which they are equal. This is the opposite of the Austronesian figures and points to a contact rather than a genetic relationship between Kam-Thai and Mon-Khmer. The figures are very low, as with Austronesian but consistent.

The differential between the 1st and 2nd 100 may occasionally be as low as 0 (i.e. equality), but in no case in our data is the 1st 100 lower in a genetic relationship, and in no case in our data is the 2nd 100 lower in a contact relationship. Even though a rare case may show up contrary to our thesis, yet the general pattern seems unmistakable.

REFERENCES

- Barnhart, R.K. 1988. *The Barnhart Dictionary of Etymology*. New York: Scott Foresman.
- Benedict, Paul K. 1942. "Thai, Kadai, and Indonesian: a new alignment in Southeastern Asia." *AA* 44:546-601.
- Benedict, Paul K. 1972. *Sino-Tibetan: a conspectus*. New York: Cambridge University Press.
- Chen, Bao-ya. 1994. *On language contact and language alliance*. Beijing University. Ph.D dissertation.
- Haudricourt, André-G. 1954. "De l'origine des tons en Vietnamien." *JA* 242:69-82.
- Li, Fang-Kuei. 1943. "The hypothesis of a pre-glottalized series of consonants in primitive Tai." *BIHP* 11:177-88.
- Li, Fang-Kuei. 1977. "Sino-Tibetan." Paper presented at the 1st Japan-US Joint Seminar on East and Southeast Asian Linguistics. Tokyo.
- Schmidt, Wilhelm. 1905. "Grundzuge einer Lautlehre der Mon-Khmer-Sprachen." *Denkschr.* kaiserl. Akad. Wiss. Wien, Phil.-hist. Kl. 51:1-233.
- Swadesh, Morris. 1952. "Lexico-statistic dating of prehistoric contacts, with special reference to North American Indians and Eskimos." *Proceedings of the American Philosophical Society* 96:452-463.
- Swadesh, Morris. 1955. "Towards greater accuracy in lexicostatistic dating." *IJAL* 21:121-137.
- Thomason, L. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Weinreich, Uriel. 1953. *Languages in Contact*. The Hague: Mouton.

Received 26 June 1995

Chinese Department
Peking University, 100871
Beijing, China

Appendix A. Basic words in Dai languages borrowed from Southwestern Chinese

Dai Dehong: 1st 100 words

Meaning	SWC originals	loans in Dai	distribute condition
hair	毛 /mau ³¹ /	/mau ⁴² / (knitting wool)	/mau ⁴² ɕian ²¹³ /
liver	肝 /kan ⁵⁵ /	/kan ⁵⁵ / (hepatitis)	/kan ⁵⁵ ian ⁵⁵ /

Dai Dehong: 2nd 100 words:

Meaning	SWC originals	loans in DD	distribute condition
father	爹 /tie ⁵⁵ /	/te ⁵⁵ /	
float	漂 /phiau ⁵⁵ /	/phiau ⁵⁵ /	
dull	慙 /xan ⁵⁵ /	/xan ⁵⁵ /	
turn	轉 /tsuan ²¹³ /	/tsuan ²¹³ /	
ice	冰 /pin ⁵⁵ /	/pin ⁵⁵ /	
sea	海 /xai ⁵³ /	/xai ²¹³ /	
tie	捆 /khun ⁵³ /	/xun ⁵³ /	
squeeze	壓 /jia ³¹ /	/jia ³¹ /	
if	如果 /zu ³¹ ko ⁵³ /	/zu ³¹ ko ⁵³ /	
because	因為 /jin ⁵⁵ uei ²¹³ /	/jin ⁵⁵ vui ²¹³ /	

Dai Jinsha: 1st 100 words

Meaning	SWC originals	loans in DJ
		none

Dai Jinsha: 2nd 100 words

Meaning	SWC original	loans in DJ
rub	擦 /tsha ³¹ /	/tsha ¹¹ /
rotten	爛 /lan ²¹³ /	/lan ³³ /
think	想 /ɕia ⁵³ /	/ɕia ⁵⁵ /
sea	海 /xai ⁵³ /	/xai ⁵⁵ /
split	扯 /tshə ⁵³ /	/tshə ⁵⁵ /
few	少 /sau ⁵³ /	/sau ⁵⁵ /
swell	腫 /phau ⁵⁵ /	/phau ⁵⁵ /
turn	轉 /tshuan ²¹³ /	/tshuan ³³ /
cut	砍 /kan ⁵³ /	/kan ⁵⁵ /

Dai Yun 1st 100 words:

Meaning		SWC originals	loans in DY	Distribution Condition
oil, grease	油	/jəu ³¹ /	/jəu ³³ / (pork fat)	/pən ⁵³ jəu ³¹ /
kill	殺	/sə ³¹ /	/sə ³¹ /	
water	水	/sw ⁵³ /	/swəi ⁵³ / (water in ditch)	/pən ⁵³ swəi ⁵³ /

Dai Yun 2nd 100 words

Meaning		SWC originals	loan in DY	Distribution Condition
grass	草	/tshau ⁵³ /	/tsha:u ⁵³ / (straw sandals)	/tsha:u ⁵³ xə:i ³¹ /
think	想	/ɕiang ⁵³ /	/ɕa:ŋ ⁵³ /	
sea	海	/xai ⁵³ /	/xai ⁵³ /	/xə:i ⁵³ tsl ⁵³ / (lake)
and	和	/xo ³¹ /	/xo ³¹ /	
old	老	/lau ⁵³ /	/la:u ⁵³ / (old man) /la:u ⁵³ ja ⁵³ / (old woman)	/la:u ⁵³ pu ²⁴ thau ¹¹ /
squeeze	壓	/ja ³¹ /	/ja ³¹ /	
father	爹	/tie ⁵⁵ /	/ʔa ³³ tje ³³ /	
dull	笨	/pən ²¹³ /	/pən ²⁴ /	

Appendix B. Mon-Khmer and Kham-Thai similar words in 1st 100 words

	VA	PL	AL	ZW	ZL	BL	DX	DD	DR	MLL	SS	KHL	LB	LT
water	rom	um ¹	rom ¹	ram ⁴	nam ⁴	zam ⁴	nam ⁴	lam ⁴	nam ⁴	nam ⁴	nam ⁴	nam ³	nam ³	nom ³
you	mai	mi ⁷²	me ⁷¹	mumg ²	maur ²	mumg ²	mumg ²	maur ²				meu ¹	meu ¹	meu ¹
green		khui ¹		heu ¹	kheu ¹		xeu ¹	xeu ¹	su ¹	hou ¹	gu ¹	ju ¹	khiau ¹	khiau ¹
one	ti	ka ⁷⁴ l ⁷⁴	te ⁷²	deu ¹		diau ¹								tsheu ³ , tsuu ²
moutain	gong	nkong ³	gong ²		domg ²		kong ²					kong ⁵ - pja		
sand		sai ³	sai ¹		tai ²		sai ²	sai ²						
cold	kuat	kuat ¹	si ² kiat ²				kat ⁷	kat ⁷						
ash	nau ⁷	ka ⁷³ u ²	no ⁷²	tau ⁶	pjau ⁶	tau ⁶	tau ⁶	tau ⁶						tsu ² - tau ³
yellow	lhvng	lvng ¹	ngel ¹		lo ^{ng} ¹		lo ^{ng} ¹	lo ^{ng} ¹					te ^{ng} ⁴	ze ^{ng} ¹
red	raung			ding ¹	deng ¹	ding ¹	deng ¹	deng ⁶					ge ^{ng} ⁴ , de ^{ng} ³	de ^{ng} ³
many	ne			lan ¹	lai ¹	lai ¹	lai ¹	lai ¹					tai ¹	toi ¹
bite	kiat	ket ²	kiat ²						kit ¹⁰	ca ⁸	ti ⁸	oi ⁸		
flesh	ne ⁷	ni ⁷		no ⁶	nu ⁴	no ⁶	no ⁴	lo ⁴						
that	ʔan	on ²	an ¹	han ⁴			nan ⁴	lan ⁶						
root	riah	xel ^{h2}	rial	Yaak ⁸	lak	za ⁷⁸	haak ⁸	kaak ⁸						

