

How to Tell Lai: An Exercise in Classification

Paul K. Benedict
Ormond Beach, Florida

Lai is almost too strange to be true. The big question is: what is it?

Liang (1984) has recorded the language in the southeastern China province of Guangxi, but the traditional homeland is to the west, in southwest Guizhou and southern Yunnan. The speakers reportedly number only several hundred, all in Guangxi, the language having been replaced by Yi (Lolo) in Yunnan. As Liang points out, the typology of the language is thoroughly Zhuang-Dong (Kam-Tai): monosyllabic morphemes; limitation of non-vocalic finals to glides (written *-i*, *-u*) and *-p*, *-t*, *-k*; *-m*, *-n*, *-ŋ*; six tones (only five before stop finals)¹; aspirated vs. unaspirated surd stop initials, with separate postvelar series; apical affricate series along with /b/ and /d/ (both prenasalized). The eight vowels (in native words) show a simple length distinction for /u/, /i/, /e/ and /ə/ and a highly idiosyncratic three-way length distinction for /a/, with both long /a:/ and half-long /a·/; the word order is VO and in noun compounds the 'possessor' normally follows (*suk¹ bu¹* 'hair head') but occasionally precedes (*tɔk¹ suk¹* 'mouth-hair' = 'beard'), the latter probably through Chinese influence. Liang also notes, however, that the lexicon reveals few similarities with Kam-Tai, Gelao (KD² language of China), Miao-Yao or Yi, and he attributes the Austroasiatic features in the numeral system to contacts with Vietnamese or other languages of that stock. He concludes his remarks with the observation that the affiliations of the language constitute a 'complex and richly interesting problem.'

This is indeed an intriguing problem, yet an answer appears to lie at hand, even on the basis of the limited available material in this sketch of the language. To begin with, one can quickly dispose of both Sino-Tibetan and Austro-Tai as possible congeners, on the basis of the lack of any substantial amount of shared basic vocabulary. As can be expected, Lai is well stocked with loans from Chinese, including those for the numerals above '1,000', but

¹ The Lai tones are numbered as follows: /1/ high-level; /2/ mid-level; /3/ low-level; /4/ high-falling; /5/ low-falling; /6/ low-rising (lacking in forms with stop final).

² Abbreviations: AA = Austroasiatic; AC = Archaic Chinese; AT = Austro-Tai; ATLC (see Benedict 1975); BG = Bodo-Garo; BL = Burmese-Lolo; CT = Central Thai; JH = Jch-Halang; KD = Kadai; KS = Kam-Sui; MC = Middle Chinese; MK = Mon-Khmer; MY = Miao-Yao; N. = Northern; NB = North Bahnaric; Nic. = Nicobarese; NT = Northern Tai; P. = Proto-; Pal. = Palaung; PK = Proto-Karen; PT = Proto-Tai; PW = Proto-Waic; S. = Southern; SB = South Bahnaric; ST = Sino-Tibetan; STC (see Benedict 1972); SWT = Southwest Tai; VM = Viet-Muong; Vn. = Vietnamese; WB = Written Burmese; WT = Written Tibetan.

none from basic vocabulary. Rather surprising, however, are the apparent early loans from Tibeto-Burman, including those for some 'non-culture' items:³

'1,000': *tɔ:ŋ*⁴; cf. PTB **s-toŋ*. This loan also appears in two Kadai languages of North Vietnam: Laqua *təŋ*, Latl *tuŋ*.

'100': *ʒɔʔ*²; cf. PBL **-rya* (WB /*ra*/ but /*rya*/ in Burmese inscriptions); with /*ʒ*/ < **y* or **ry*, /*ɔ*/ < **a*. This loan is also represented by PW **ryah*, from an **s-* prefixed (> final *-h*) prototype in TB.

'sky/rain': *qɔʔ*¹; cf. PTB **m-ka*: WT *mkha* 'sky', Garo *mikka* (= *mka*) 'rain'; with /*ɔ*/ < **a* and marked as relatively early by /*q*/ < **k*.

'ox/cattle': *la:i*⁶; cf. PTB **(g-)lwa:y* 'buffalo', itself an early loan from a KD source: PT **ɣl/rwaay*⁴ 'id.', with the medial maintained in Ahom *khrai*; with regular loss of medial **w*.

'fowl': *kye*⁴; cf. PBL **(k-)rak*, with 'animal prefix' **k(a)-* (see below), medial /*y*/ < **r* and regular *-e* < **-ak*, with the anticipated tone /4/ (see below); the initial *ky-* points to a relatively late loan (Lai has *k-* < **kr-*).

'head': *bu*¹; cf. PTB **d-bu*.

'needle': *ka:p*⁵; cf. PTB **s-kap*; marked as relatively late by /*k*/ < **k*. Khmu *skam* appears to be related to this root but the phonology remains obscure.

'run': *ga:t*⁵; cf. PTB **gat* (Jinghpaw *gat*, Mikir *kat*) ~ **kat* (BG **kat*); marked as relatively early by /*q*/ < **g* or **k*.

One is struck by the fact that these loans as a whole relate to TB generally rather than to the BL languages in the vicinity of the Lai, with forms often mirroring reconstructed PTB roots rather than the much modified forms of languages such as Loloish. It appears that some of these loans, at any rate, were made at an early period, when the ancestral Lai were located well to the west of the Lai of the present day. Still other loans are possible, even probable, but are listed below because of possible alternative etymologies. If the set of Lai forms relating to 'eat' is indeed of TB origin, which seems more than likely, an actual transfer of a morphological element (prefixed **a-*) to Lai must be inferred, indicating a very close and prolonged contact between some early group(s) of TB and the ancestral Lai.

Austro-Tai is represented by a rather larger number of Kadai loans, all from Tai and/or Kam-Sui⁴:

'ax': *khvən*⁵; cf. Thai (Siamese) *khwaan* < PT **xwaan*⁴; the vocalism is better matched by Sui *kwa:n* ~ *kuən* (tone **A*).

³ For the TB forms, see especially Benedict 1972, with modifications and additions as indicated by cited forms.

⁴ For the Tai and KS forms, see especially Li 1977 and Li 1965, respectively, with the writer's modifications (see ATLC: 181) and additions, as indicated by cited forms.

'firewood': *vi*¹; cf. PT **vɛɛl*¹ (final **-l* maintained in Saek); *-i* < **-ii* < **-il*, /i/ < **ɛ* in this loan and 'fish scales' (below), perhaps both from a prototypical **i(i)* rather than **ɛ(ɛ)*. This loan and that for 'ax' look like a 'loan package'.

'cotton': *va*:*i*⁶; cf. Diol (NT) *way* < **hwaay*^C; also PKS **hwaay*^C: Ten *xvaai*, Mak *waai*, contrasting with SWT/CT **faay*^B; Sul *fai* < **faay*^C. The tone /6/ of this Lai form apparently reflects an initial **hw-* in the prototype (see below).

'animal enclosure (pen)': *qo*:*k*⁵; cf. PT **yook* < **Gook*, with a doublet **Groom* as reflected in Lao *hok* < **room* 'coop' and the early loan represented by Khmu *groom* 'stable; prison'; marked as relatively early by /q/ < **G*, probably via **g*, but after the **o* > /u/ shift; apparently not from the doublet form: **groom*, which should have yielded initial /k/ in Lai.⁵

'horn': *kau*¹; cf. PT **xaw*^A < **ghaw*^A; marked as relatively late by /k/ < **k*, apparently from NT, which has initial /k/ in this root.

'fish scales': *ɛi*:*p*; cf. SWT **hlɛɛp* (Lao *lɛɛp*, Shan *lɛp*, both on 'high' tone) 'scale or peel off'; with /ɛ/ < **hl-*, /i/ < **ɛ* (cf. 'firewood', above); apparently from a NT (less likely: KS) language that had retained a nominal meaning in this root (from PAT **qulap*; see ATLC: 370), found elsewhere at the present time only in Li (Hainan): White Sand Li *luap* < [q]ulap (with 'vocalic transfer') 'fish scales'.

'father's y. brother': *ʔa*:*u*¹; cf. SWT/NT **ʔaaw*^A ~ CT **ʔaaw*^B.

'child': *ka*¹ *lye*²; cf. NT **lɛk* < **lyak*; also PKS **laak*; with prefixed *ka*¹- (see below); *-ye* < **-yak*, reflecting an earlier (reconstructed) NT root, or **-e* < **-a(a)k*, a regular shift with secondary medial /y/ after prefixed *ka*¹-.

'inside': *le*² *da*:*i*¹ (*le*² is locative); cf. Tho (CT) *dai*; also Wu-ming (NT) *ʔdaɛ* (Lai lacks *-aɛ*).

'dumb (mute)': *ɲam*⁶; cf. PT **ɲwam*^C; with regular loss of medial **w* (cf. 'ox/cattle', above).

'come': *mu*ɔ² ~ *ma-* (preceding the negative *ʔo*²; see below); cf. SWT/CT **maa*^A ~ NT **hmaa*^A; PKS **hmaa*^A; with /ɔ/ < **a*.

'go out': *ʔok*¹; cf. PT **ʔook*; marked as late by /ɔ/ < **o*.

'open (door)': *ha*:*i*⁵; cf. Mak (KS) *hai* (< tone **A*), as contrasted with Thai *khay* (< tone **A*).

This is a sizeable number of loans, to be sure, especially in view of the limited amount of available material, but the Lai inhabit the very heart of the NT homeland, with KS speakers also in the region, hence this borrowing should hardly come as a surprise. Additional possible loans from Tai or

⁵ This etymon has also been reconstructed for Proto-Loloish as **krok* (cf. Lahu *khô* [Lh. *kh-* reflects a **kr-* cluster], Akha *kúʔ*, Mpi *khôʔ*), though it was probably borrowed into PL from Tai (see Matisoff 1972 #16 and 1988:385). [Ed.]

Kam-Sui are presented below because of possible alternative etymologies. The precise source of the loan is difficult to determine in some cases, as in 'ax' and 'child'. Some of the loans appear to be of considerable antiquity; note especially the semantics of 'fish scales' and the phonology of 'dumb', with Lai **ɲam⁶* < **ɲwam* reflecting an earlier prototype than any of the known Tai forms of the present day: Ahom (SWT) *bum* < **wum*; CT *waam* ~ *vam* ~ *vɔɔm*, NT *ɲɔm* ~ *gom*.

The other mainland family of the AT stock, Miao-Yao, appears to have exerted little if any influence upon Lai. The Lai combining form: *tau¹* '(male) person', as in */tau¹ ɲam⁶/* 'a (male) mute' (see above), can be compared with 'Highland' (Chiengrai) Yao *tau* (< **taw⁴*) 'clf. for persons, et al.' but, in the absence of any pattern of borrowing, the two forms are more likely simply 'look-alikes'. This apparent lack of MY loans in Lai supports the evidence supplied by early TB loans (above) to the effect that the Lai came from the west, outside any prime MY-speaking area.

With both ST and AT thus disposed of as possible congeners of Lai we are left with the third great language stock of SEA: Austroasiatic. MK affinities are apparent both for the numerals, as pointed out by Liang, and the personal pronouns (see the discussion of both below) as well as for a number of other lexical items. Huffman (1976) has compiled a list of 30 'highly persistent' roots as represented in 19 different MK languages and in 10 of these for which Lai forms are available a MK/Lai cognate set is evident in five; cf. the following list, with MK forms as presented by Huffman, who makes use of 'a kind of canonical form, or phonological average' (cit. supra: 549):

'water': *de⁴*; MK */daak/*.

'fish': *qɔ⁴*; MK */kaa/*.

'dog': *tsu⁴*; MK */cɔɔ/*.

'eye': *ma:t²*; MK */mat/*.

'hair': *suk¹* 'body hair; (cp. with 'mouth') beard; (cp. with 'head') head hair'; MK */sɔk/* 'hair'.

Of the remaining five entries on Huffman's list, Lai has a loanword in one ('child'; see above) and shows a semantic shift in another ('fly'):

'day': *han⁶*; MK */ɲay/*.

'bird': *san⁴*; MK */ceem/*.

'fly', n.: *mi³*; MK */ruy/*; here the comparison is with MK forms for 'mosquito': Thavung (VM group) *mɔɣh* (< **mɔs*), Vn. *muối*, Bru *muas*, Chabon (= Nyah Kur) *muus* (cf. 'laugh', below).

'horse': *lyiŋ⁶*; MK */seh/*.

'child': *ka¹ lye²* (loan from KD; see above); MK */kɔɔn/*.

In addition, the Lai material includes several other lexical items with ties to widespread MK roots:⁶

'year': *nam*⁶; cf. Khmer *cnam*, Mon *hnam*, PNB **hanām*, Khasi *snem*, PW **nym*, Vn. *nām*, et al.

'stone': *mau*³; cf. PNB **tamo*, PW **smo*?, Mon *mɔ*?, Kuy *tamaʋ* et al.

'eagle': *muɔ*⁵ *lyɑːŋ*⁵ (see below for *muɔ*⁵); cf. PMK **k(a)laaŋ* 'bird of prey (generic); hawk (Pal., Waic, Khmu, Bru); eagle (Waic); to fly without wing movements (Mon); vulture (Nic.); white-bellied sea-eagle (Nic.)', as reconstructed and glossed by Diffloth (1980b), who cites Northern MK **klaaŋ* and, under VM, Thavung *kalaŋ*¹ (glossed by Ferlus as 'pelican'); with secondary medial /y/ after the original 'presyllabic' **k(a)-* (cf. 'sun', below); for the Mon gloss, cf. Lai *lyɑŋ*² 'fly'.⁷

'husk (of grain)': *qam*⁴; cf. Khmer *aŋkām*, S. Khmu, Lamet *ŋka:m*, Pal., Riang *kham*, PW **kam*, Mon *kām*, et al.

'house': *ña*²; cf. PW **ña*?, Lamet *ña*, PJH **ña*, Vn. *nhà* (= *-nà*), on huyền tone, reflecting a non-glottalized final.

'thread': *ya:i*¹; cf. PNB **bray*, Mnong (S. Bahnaric) *brai*, Pear *phray*, Pacoh (Katulic) *parai*, et al. 'id.'; Pal. *pra:y* 'twist of thread'; with /y/ corresponding to MK /r/, along with tone /1/, indicative of initial unvoicing; cf. 'deep' (below).

'deep': *yau*²; cf. Khmer *ja(m)rau*, PNB **jru*?, Mon *rjūh* (*rj- < *jr-*), Khmu *jru*, Ngeq (Katulic) *c'ruu*?, Riang *tsəru*?, PW **?rɛ*?, (**?ru*?), 'Sakai' (Aslian) *jəru*, et al., with /y/ corresponding to MK /r/, as in 'thread' (above).

'mountain': *qhɔŋ*⁵; cf. PW **ŋg(ɔ)ŋ* ('vowel uncertain').⁸

'salt': *mya:n*⁶; cf. S. Khmu *hmar*; with secondary medial /y/ after an original 'presyllabic' (note S. Khmu *h-*).

'egg': *tham*⁶; cf. PW **ktəm*, Pal. *kətəm*, Riang *tam*, Lamet *ntam*.

'rice (early)': *bə*⁴; cf. Aslian: North (Cheq Wong), South (Mah Meri) *bə* 'rice (unhusked)', apparently quite distinct from Central Aslian (Semai, Temiar, Semram, Sabum, Lanoh, Jah Hut) *ba* 'id.', a cognate of Khasi *kba*, Danaw *ba*¹ 'id.'.

'road': *ko*⁴; cf. PW **kra* 'road, way'.

'long': *lɔŋ*²; cf. PW **?laŋ*; Aslian: Semaq Beri, Mah Meri *jəlɔŋ*, Temoq *jəlɔŋ*, et al.

⁶ For the MK forms, see especially Benjamin, G. 1976 (Aslian); Delcros 1966 (S. Khmu); Diffloth 1977a (P-Semaic) and 1980a (PW); Ferlus 1979 (Thavung); Huffman 1977 (VM and MK cfs); Luce 1965 (Mon, Palaungic and MK cfs); Smalley 1961 (S. Khmu); Smith 1972 (PNB); Svantesson 1983 (N. Khmu); Thomas and Smith 1967 (PJH); also the special sources on numerals and pronominals cited above.

⁷ This widespread root also appears in TB and MY. Cf. STC #333 and note 225 (p. 72). [Ed.]

⁸ Cf. PLB **kaŋ*¹ 'mountain' (WB *khaŋ*, Lahu *qhɔ*). [Ed.]

'go': žu²; cf. S. Khmu yoh; Khmu final -h after /o/ reflects earlier *-s as well as *-h (Lindell, Svantesson and Tayanin 1976) but Lai has -i < *-s (see below), disambiguating here in favor of an original *-h in this root.

'rob': hy³; cf. S. Khmu ra² 'take by force from someone, snatch'; PW *mra² 'steal'; Pal. bra², Riang, Danaw ra² 'id.'; with /y/ corresponding to MK /r/.

'listen/hear': moŋ⁶; cf. PW *hmɔŋ.

'laugh': mi⁶; cf. Thavung (VM) khamayh¹ (< *khamas) (cf. 'fly', above).

Other possible Lai/MK cognate sets are cited below in connection with phonology, while still others have competing non-MK etymologies. Four of these involve kinship terms, all designating older-than-Ego kin, hence of the type that tends to be areal in SEA, e.g. TB and P-Waic share no fewer than five such terms (Benedict 1979a).

One term of this kind ('Fa.'s o. Bro.') clearly has been borrowed from Tai (see above), hence we can anticipate that others may also have been borrowed from the same source while, as elsewhere, loans from TB must also be considered a possibility.

'grandfather': ta⁵ (Chinese glossed as 祖父 'pat. grFa.' but perhaps also for 'mat. grFa.', for which no term is cited); cf. Old Khmer atā 'old man', Khmer, Pear taa 'grFa.'; Riang ta², Danaw ta¹, Pal. ta² 'id.'; PW *ta² 'grFa., old man'; Thavung (VM) taa¹ 'Fa.'s o. Bro.' (cf. Maonan); cf. also PT *taa¹ 'mat. grFa.' (—> 'wife's Fa.' through teknonymy); PKS *taa¹ 'id.'; Maonan also /ta/ (< tone *C) 'Fa.'s o. Bro.'; cf. also PTB *ta 'grFa./Fa./ o. Bro.' (secondary term) and note AC/MC 𑜋 tso/tsuo: < *sto: < *s-ta^B 'grFa./ancestor'. The vocalism, with /a/ for *a, suggests a recent Tai loan, but the tone /5/ points to an underlying parallelism with the Lai doublet for 'father' (below): *tɔ¹ ~ ta⁵, consistent with a MK origin for the term, as indicated also by the semantics: 'grFa.' or 'pat. grFa.' rather than 'mat. grFa.'. A TB source seems highly unlikely here, despite the development in Chinese.

'Fa.'s o. Bro.'s wife': muɔ⁵ žɔ⁶ (muɔ⁵ 'mother'); also /žɔ¹/ as combining form for 'female'; cf. PNB *ya²; Riang, Danaw ya², Pal. ya²; PW *ya² 'grMo.'; cf. also PT *yaa^C 'pat. grMo.', appearing in Shan (SWT) as 'appellation given to elderly females' and in Ddoi (NT) used for 'wife' (cf. Eng. slang 'old lady'); cf. also PTB *ya² 'Mo./grMo.' (secondary term).⁹ The phonology here (/žɔ/ < *ya) is of no help, but the apparent underlying usage as a term of respect points to a Tai rather than MK source, with TB a poor third possibility.

⁹ The PTB etymon for 'mother/grandmother/maternal aunt' is to be reconstructed rather as *(y)əy (Matisoff 1985, #100).

'father': $pu\supset^1$ 'father', applied also to males of persons, birds ~ pa^5 $\supset a:u^1$ 'my father'; cf. PNB $*\supset aba\supset$; Mon $a(m)ba$ Riang pa^- , Danaw pa^4 (both < $*ba$); PW $*ba\supset$; cf. also PTB $*pa$ but PK $*ba$ (~ $*pha$) and AC/MC $\times b'i\ w\supset/b'i\ u < *ba$. The Lai doublet appears to reflect retention of $*a$ before close juncture (see below), with a parallel doublet indicated for 'grFa.' (above). The tone /1/ points to an earlier $*p$ - rather than $*b$ -, but an original $*b$ - cannot be excluded in this case in view of the uncertainty regarding the development of voiced stops in Lai (see below). The paired term for 'mother' (below) has the appearance of an old loan from TB and on this basis one might consider a TB loan to be the likelier possibility for this term as well.

'mother': $mu\supset^5$ 'mother', also used for the females of animals and plants, with extension to a general combining form, as in 'eagle' (above); cf. Khmu, Riang $ma\supset$, Pal. ma^2 , PW $*ma\supset$ 'mother'; apparently also PJH $*ma$ 'aunt'; cf. also PTB $*ma$ 'mother'. The MK 'mother' terms generally have a front vowel rather than /a/, however, with the Khmuic and Palaungic /ma/ forms representing possible areal influence from TB, where /ma/ prevails. In view of this a borrowing from TB appears to be more probable than a MK provenience.

Apart from the above kinship terms, three other lexical items in Lai have competing MK vs. non-MK etymologies:

'this': ni^1 ; cf. PAA $*ni/ne$ (Pinnow 1965); cf. also PT $*ni i^B$ 'here' ~ $*hni i^B$ (< prefixed form) 'there'. The vocalism provides no clue here (Lai $*-i < *-i \sim *-e$), but the tone /1/ points to an earlier unvoiced initial: $*?n-$, fitting phonologically with PT $*hn-$ (< $*hn- \sim *?n-$) but semantically rather with PT $*n-$.

'carry on back (child)': pe^4 ; cf. PMK $*tabak$, as reconstructed on the basis of Khasi $tabah$ (= $taba\supset$) 'throw, hang, throw or hang across the shoulder', Vn. $v\acute{a}c$ ($v-$ < medial $*b$ and $s\acute{a}c$ tone < initial $*t-$) 'carry on shoulder', cited by Shorto (1973), who adds Sre $ba\supset$ 'to carry in shawl on back', et al.; cf. also the restricted TB root represented by Jinghpaw $ba\supset$ 'carry on back (as child)' and the adjacent Rawang $ba\supset$ 'carry on back/shoulder', both from $*bak$. Shorto (cit. supra) regards the above forms as probable loans from MK, but the semantic development is rather different and, as can be seen from the above glosses, fits well with Lai. The final $-e < *-ak$ is regular (see below) while the tone /4/ points to an earlier unvoiced initial, perhaps reflecting an original 'presyllabic' $*t-$, as in Vn. In view of the uncertainty regarding the development of voiced stops in Lai, however, the tone is insufficient to rule out the possibility of borrowing from TB.

'eat': $ts\supset^1$ 'eat', $\supset\supset^4$ $ts\supset^1$ 'grain/food'; also $ts\hbar\supset^6$ 'feed'; cf. Mon ca , PNB $*ca$, Nic. $\check{s}a$, Old Khmer cya 'eat'. This hardly has the look of a 'native' Lai series, however, with $\supset\supset^4$ on a 'wrong' tone (see below; not

glossed separately) and the 'causative' derivative form with aspirated initial lacking any apparent Lai (or MK) basis. If one posits a loan/loan series or 'loan package' from a TB/Karen source, however, all can be explained, with /ʔɔʔ/ for the widespread PTB nominalizing *a- (= /ʔa-/) prefix and initial variation suggestive of typical TB secondary aspirates in initial position vs. non-aspirates after prefixation or from earlier sonants; cf. PTB *dzɿa^B (generally > /dza^B/ forms) 'eat' > PBL *dza^B; PBL also *dza^A 'uncooked rice/food', fitting a known pattern of tonal variation in TB (Benedict 1980). TB also exhibits a basic morphological feature of initial voiced = 'intransitive/stative' vs. unvoiced = 'transitive/causative' but this pattern obtains only in a limited number of roots, with many roots showing a contrariant pattern, e.g. *səy^A 'die', *dzɿa^B 'eat'. This fact apparently has led to an irregular unvoicing of the initial in the Tamang group (Nepal): P-Tamang *tsa^B 'eat' and (with secondary aspiration as an initial) *tsha^B '(cause to eat =) graze (cattle) (tr.)'. Karen also has an unvoiced initial in this root: P-Karen *(a-)tsha^B 'food/fodder' but here it is ambiguous for an original *dz ~ *ts. The tone /1/ of Lai tɕɔ¹ points to an earlier *ts- rather than *dz-, yet one cannot preclude the possibility of an original *dz- in this case (see below), hence the TB prototype may well have been *dza^B rather than *tsa^B. Lai tshɔ⁶, however, points to a TB prototypical set of Tamang type, if indeed these forms are to be viewed as loans from that family. It is possible that we are dealing here with a 'mixed bag' of forms, e.g. with /tɕɔ¹/ and /ʔɔʔ/ tɕɔ¹ of loan origin but /tshɔ⁶/ of MK provenience, yet the semantic shift would be difficult to explain. The striking similarity of the TB/Karen (also Chinese) and MK roots is suggestive of an areal factor, especially in view of the reconstruction: PMK *cyaʔ suggested by Shorto (1972), perhaps better: *tsyaʔ (see below for *c- vs. *ts-).

An eighth Lai item with competing etymologies is prefixed ka¹-, which appears in a wide range of lexical categories: both animate (fauna, flora) and inanimate, e.g. ka¹ ɿya:ŋ, 'sun' (above); kinship terms, e.g. ka¹ lyɛ² 'child' (above); time divisions, e.g. ka¹ han⁶ 'daytime' (han⁶ 'day'); locatives, e.g. ka¹ dau¹ 'above'. The k- 'preformative' of certain MK faunal terms, e.g. PMK *k(a)laaŋ 'bird of prey' (above), can be interpreted as an 'animal prefix', with Sedang and other MK languages of Vietnam showing some extension of this feature (Smith 1975). The initial syllable of Pal. kətəm 'egg' vs. Riāng tam (cited above: Lai has the unprefixed /tham⁶/ here), as contrasted with that of Pal. kətɛ 'day' vs. Riāng kətɛʔ, can be viewed in a similar light, although the widespread monosyllabizing trend shown by the MK languages casts some doubt upon instances of this kind (cited along with a caveat by Shorto 1963: 57). The greatly expanded role of this prefix in Lai is paralleled close at hand in Blang (Samtau), a Waic (Palaungic group) language (Zhou and Yan 1983), which has /kàʔ/ (low-

falling tone) with body part words (thigh, tongue, gall bladder), faunal terms (ox/cattle), kinship terms (son's wife), culture words (small weeding hoe) as well as with numerals (one, nine) and even with adjectives (far, rich man). Further afield, Khasi has prefixed *k(a)-* ~ *kh(a)-* in numerous terms for 'living things' (Henderson 1976): animal and plant names, kinship terms, all as in Lai, as well as in body part words, as in Blang but apparently not in Lai (not specifically excluded, however, by the author of the Lai sketch).

The case for a MK origin for this Lai prefix thus would appear to be a strong one, yet the phonology presents a problem: why not the anticipated **q>-*? The vocalism can be explained in terms of retention of **a* before close juncture (see below) but the initial */k/* rather than **/q/* requires something more. Perhaps initial **k-* was retained here because of the weak or 'zero' stress of this prefix, paralleling a variation exhibited by AT (see below). It should also be noted that the tone */1/* here points to an earlier **k-* whereas the low-falling tone of the Blang prefix is indicative of an earlier **g-*, perhaps through secondary voicing in this weakly ('zero') stressed element.

The real problem with all of this, however, is that the velar prefix is an areal feature. TB has an 'animal prefix' **k(a)-*, well represented in BL and also encountered at times in other TB languages, notably Lepcha and Jinghpaw (see 'eagle', above, for an example of reanalysis in this 'bird of prey' loan from MK). It also appears in Lepcha and on occasion elsewhere in plant names (WB *krim* < **k-ri:m* 'rattan') and is even found in words for inanimate objects (WB *kyauk* < **k-lauk* < **-luŋ* 'stone') (see the discussion in STC: 113). It has been interpreted as a loan from MK, with reference especially to its appearance in Khasi, and it is perhaps significant that the languages in which it is best represented are either adjacent to MK-speaking areas (BL and Jinghpaw) or believed to have a MK substratum (Lepcha; see STC: fn. 24). It is possible, of course, that Lai *ka¹-* is a back-loan from some TB source(s); the expansion of function in Lai, however, is best matched within the MK family (Blang, Khasi).

As if all this were not complex enough, prefixed **qa-* with body part words is an absolute hallmark of the AT stock, represented even in Japanese (*a-* < **qa-*), with **ka-* as a variant (PAT **q>/k/* in weak or 'zero' stress position); cf. Lai, as discussed above. This **ka-* variant appears both in Kadai, e.g. Thai (Siamese) *kăduuk* 'bone', and the Miao languages, e.g. Phō *hlīe* 'collar', *kăhlīe* 'neck', while in Thai it is also found in faunal terms, e.g. *kătaay* 'hare', *kăcok* 'sparrow' (see the discussion in ATLC: 147).¹⁰ In view of this, one can hardly exclude the possibility that Lai *ka¹-* is a relatively late loan (hence the initial */k/*) from a KD or, less likely, a MY

¹⁰ The standard Thai forms are actually *krədūuk* 'bone', *krətāy* 'hare', and *krəcōk* 'sparrow'. [Ed.]

source. To sum up, this Lai prefix is of MK provenience, if it is not a loan from TB, provided of course that it is not a relatively late loan from either KD or MY. This is etymological ambiguity on a grand scale indeed, even for Southeast Asia!

Both the numerals (Zide 1976, Jenner 1976, Thomas 1976) and the pronominals (Pinnow 1965) of AA/MK have been the subjects of broad comparative studies, hence they offer promising leads in the analysis of the available Lai material, which is complete in these two lexical categories. The Lai numerals, clearly of MK provenience, are best analyzed in two sets: 1 - 5 and 6 - 10, as in the discussions by Jenner and Thomas. Lai *mə⁵* 'five' appears to be innovative, without apparent resemblance to any of the two or three groups of forms recognized by Thomas. The Lai numerals for 'one' through 'four', however, closely resemble the set of numerals included in the 'highly persistent' list of 30 MK roots cited by Huffman (above). The following table presents the MK forms cited by Huffman, along with the Proto-Khmer (Jenner), Old Mon, Bahnaric (PSB/PNB), Kuy (Katuic) (secondary nasalization omitted), Mal and Puoc (Khmuic) and Semaq Beri (Aslian), all as contrasted with forms from VM languages: Vn., Pong and Thavung:

	ONE	TWO	THREE	FOUR
Lai	<i>ma:i⁵/mə²</i>	<i>bi¹</i>	<i>pa:i¹</i>	<i>pu:n⁴</i>
MK (Huff.)	<i>/muəy/</i>	<i>/baar/</i>	<i>/pee/</i>	<i>/poon/</i>
Proto-Khmer	<i>*mooy</i>	<i>*βeer</i>	<i>*pii</i>	<i>*poon</i>
Old Mon	<i>/moy/</i>	<i>/ʔbar/</i>	<i>/piʔ/</i>	<i>/pɔn/</i>
PSB	<i>*mway</i>	<i>*bar</i>	<i>*pe</i>	<i>*puân</i>
PNB	<i>*ʔmoyʔ</i>	<i>*ʔbar</i>	<i>*pe</i>	<i>*pun</i>
Kuy	<i>muuy</i>	<i>ʔabia</i>	<i>ʔapay</i>	<i>ʔapoon</i>
Mal	<i>mooy</i>	<i>piar</i>	<i>pheʔ</i>	<i>phoon</i>
Puoc	<i>môt</i>	<i>biêl</i>	<i>pe</i>	<i>pôn</i>
Semaq Beri	<i>muy</i>	<i>hmar</i>	<i>mpɛʔ</i>	<i>mpon</i>
Vn.	<i>môt</i>	<i>hai</i>	<i>ba</i>	<i>bôn</i>
Pong	<i>motj</i>	<i>halʔ</i>	<i>pa</i>	<i>pôn</i>
Thavung	<i>muut¹</i>	<i>haal¹</i>	<i>pai¹</i>	<i>poon³</i>

Notes:

'ONE': The full Lai form (*mə²* is reduced form) shows regular loss of /u/, paralleling that for '1st pl. (excl.)' (below). The VM/Puoc form appears to reflect an earlier **muə* + unidentified particle.

'TWO': The preglottalized initial apparently reflects prefixed /ʔa-/, as found in Kuy for 'two' through 'four'; cf. also Semaq Beri *hmar* < **hmbar*. The vocalism is a problem here, as recognized by Shorto (1972), who devoted a full article to it. He cites Old Khmer *bər* - *vɤar* as the

problem forms, leading to the reconstruction: PMK **biʔaar > *ʔbia(a)r*, thereby explaining also the deviant Palaungic */ʔaar/* forms. Jenner's Proto-Khmer **βeer* reflects the fact that the earlier (pre-Angkorian) form is */ver/*, as contrasted with the Angkorian */vyar/ ~ /byar/*. Medial */ee/* will hardly do for the PMK level, however, with the Khmuic (Mal, Puoc) forms supporting the reconstruction with medial */iʔa/* or perhaps simply */ia/*; cf. also Aslian: Kintaq Bong *biyeh*, Che' Wong *ber*. Lai *bi¹* supplies additional support for this line of reconstruction, probably from **bir < *be(e)r < *biar*, with regular **e > /i/* shift; it further seems to preclude a possible interpretation of MK */ʔbiar/* types as register variants of a basic */ʔbaar/* in relation to the preglottalized initial, as suggested by the writer (ATLC: 482), although a register explanation may be applicable in isolated cases. As for the VM forms for this numeral, various attempts have been made to connect them to the general */ʔbaar/ ~ /ʔbiar/* root, but none has been very convincing, hence they must be viewed as innovative, along with the VM forms for 'one'.

The Lai forms for 'six' through 'ten' present a rather different picture, with ties specifically to VM rather than to MK generally (Khmer innovates for '10' and has compounds based on 'five' for 'six' through 'nine', while Mal and Puoc have loans from Tai; Semelai [Aslian] has forms only for 'six' and 'seven'):

	SIX	SEVEN	EIGHT	NINE	TEN
Lai	<i>piu⁴</i>	<i>pai¹</i>	<i>sa:m⁴</i>	<i>šen⁴</i>	<i>ma:n²</i>
Old Mon	<i>/tərcw/</i>	<i>/dəmpɔh/</i>	<i>/dəncam/</i>	<i>/dəncit/</i>	<i>/cɔs/</i>
PSB	<i>*tadrǎw</i>	<i>*poh</i>	<i>*pham</i>	<i>*sɪn</i>	<i>*məjɔt</i>
PNB	<i>*prǎw</i>	<i>*tapǎh</i>	<i>*tahɣam</i>	<i>*tac'in</i>	<i>*jǎt</i>
Kuy	<i>kapat</i>	<i>kapôôl</i>	<i>takual</i>	<i>takieh</i>	<i>fiɔt</i>
Semelai	<i>pruʔ</i>	<i>tmpəh</i>	--	--	--
Vn.	<i>sáu</i>	<i>báy</i>	<i>tám</i>	<i>chín</i>	<i>mười</i>
Pong	<i>prǎu</i>	<i>pal</i>	<i>sam</i>	<i>chin</i>	<i>mal</i>
Thavung	<i>phaluʔ¹</i>	<i>pih¹</i>	<i>saam³</i>	<i>ciin¹</i>	<i>sip¹</i>

Notes:

'SIX': Lai form also cited as */pi:u⁴/* (length not distinctive before */u/*). The primary root can be set up as PMK **paruʔ* (Thavung */l/ < *r* is regular), with secondary */-ǎw/* forms through 'vocalic transfer' (Benedict 1979b). In addition, the secondary **pr-* cluster yielded Vn. *s-* (= */š-/*, a regular shift), as well as Lai *pi-*, with */i/ < *-r-* as distinct from */y/ < *-r-* in the primary cluster (at the time of the loan) in 'fowl' (above). A dental (*d- ~ t-*) 'presyllabic' is evident in the Old Mon and PSB forms here, and appears elsewhere as well in other numerals in the 'six' through 'nine' group.

'SEVEN': The primary form appears to be of /*ta(m)pah*/ type but both the VM and Lai forms point rather to /*pal*/, with Lai *-i* < final **-l* paralleling the regular shift in Vn.. The status of Thavung *pih¹* is unclear, however, and the Vn. *h¹oi* tone is indicative of an earlier final **-s* or **-h* rather than **-l*. A possible solution here is provided by setting up an underlying root such as (*t-*)(*m*)*pah¹*, with general loss of the final syllable in maintaining disyllabism after the 'presyllabic', along with simple loss of **-h-* in Lai and a parallel late loss (but reflected in the tone) in Vn. in a basic form without the 'presyllabic'.

'EIGHT': Both Lai and VM reflect a root of /*sa(a)m*/ type (Vn. /*t/* < **s* is regular), from a basic MK root of /*ta(m)sam*/ type, with Mon /*c/* < **s* under the influence of the regular /*c/* in the following numeral. The strange SB/NB pair can be tied in by positing a dissimilative **s* > /*h/* shift, as indicated by Mang (unclassified/isolate) *tâmham*, with /*mham/* > /*hgam/* ~ /*pham/*.

'NINE': Here the MK root clearly is of /*tacin*/ type, with Mon /*t/* < **n* under the influence of the regular **t* in the following numeral, later replaced by /*cɔs/*; Niakuol (= Nyah Kur), a closely related Monic language with conservative features, has /*njit/* 'nine', /*jat/* '10'.

'TEN': Two roots are in evidence here: the Monic/Bahnaric/Katuic, of /(*ñ*)*jat*/ type, and the VM/Lai, of /*ma:l*/ type, with regular final **-l* > *-i* shift in Vn. and **-l* > *-n* shift in Lai after the long vowel /*a:/*. Thavung *sip¹* is a loan from Tai.

As can be seen from all the above material on numerals, the critical question can be formulated as follows: are the VM/Lai agreements shown by the second ('six' through '10') group of numerals the result of parallelism, or do they represent shared innovations? If the above analysis is correct, the agreements in 'six', 'eight' and 'nine' are clearly of the former kind, probably along with that in 'seven', while one can also argue that in the case of '10' it is Monic/Bahnaric/Katuic rather than VM/Lai that has innovated. This line of argument is greatly strengthened by the fact that Lai has no trace of the striking VM innovations for both 'one' and 'two'.

The Lai personal pronouns are even more 'independent' than the numerals, as indicated by the following:

'1st sg.': ?*a:u¹*; cf. Sedang Rengao *aw*, P-Semaic **ăw*, as opposed to the widespread MK/Munda root of /*iŋ/* type (Pinnow, cit. supra).

'2nd sg.': *mi²*; cf. the widespread MK/Munda root of /*me/* type (Pinnow); with regular /*i/* < **e* shift.

'3rd sg.': ?*i¹*; cf. the widespread MK/Munda root of /*i/* ~ /*e/* type (Pinnow), with the Lai vocalism ambiguous for this pair of proto-forms (cf. '2nd sg.', above).

'1st pl. (excl.)': *pa·i¹*; cf. Old Mon /*poy/* 'we'; with regular loss of the initial labial element in a final /*-uay/* ~ /*-uəy/* type, yielding Mon

/-oŋ/; cf. the precisely parallel Lai and Mon forms for 'two' (above). Pinnow admits to having had "major difficulties" with the MK/Munda forms for this pronoun as well as for the following, surely indicative of widespread replacement of both roots. It is possible that the Lai form represents the basic MK root here but there is an alternative possibility, that Lai has innovated by modifying ('half-long' vowel) the numeral for 'three' *pa:i¹*, on the same tone; this would constitute a parallel (possible areal factor) to a basic MY feature, e.g. Yao (Chiengrai) has /*pua*/ (< **pua⁴*) 'three', /*bua*/ (< **mpua⁴*) 'plural marker' —> 'we'; White Miao has /*pe*/ (< tone **A*) 'three; we'.

'1st pl. (incl.):' *ʔa:i¹*; cf. the Palaungic/Aslian/Nic. root for '1st dual' of /*ai*/ type (Pinnow). The Lai cognate emphasizes the strikingly peripheral distribution of this root, which Pinnow speculates (without the Lai form!) may have originally been a plural form. This does indeed seem quite likely now, permitting the conclusion that Lai has maintained the original MK pair of '1st pl.' pronouns.

'2nd pl.': *ma:i²*; apparently distinct from the widespread MK/Munda root of /*pə*/ type (Pinnow), perhaps an innovation based on /*mi²*/ (< **mə*) '2nd sg.' through analogy with /*ʔa:i¹*/ 'we (incl.)'.

'3rd pl.': *cɔŋ¹*; apparently an innovation, distinct from the widespread MK/Munda root of /*ki*/ ~ /*ku*/ type (Pinnow).

The Lai material, although limited in amount, suffices for establishing the main outlines of the comparative phonology of the language vis-à-vis MK. The distribution of initials with respect to the six tones is highly revealing and furnishes an excellent starting point for our analysis. Seven different categories of initials can be set up on the basis of this distribution, as shown in the following table, which gives the numbers of forms on each tone (f- only in loans from Chinese; ɣ- only in /*ɣa:i¹*/ 'thread', above, and three forms on tone /2/; hɣ- only in *hɣɔ⁵* 'rob', above):

TONES:	/1/	/2/	/3/	/4/	/5/	/6/
1. Unasp. stops and affricates	45	28	5	28	22	--
2a. Nasals	3	16	8	2	12	10
/ <i>l</i> /	3	14	5	2	4	6
2b. / <i>v</i> /	3	2	2	--	3	6
2c. / <i>ʒ</i> /	--	3	5	3	7	6
3. Glottal stop	6	--	--	--	--	--
4. / <i>b, d</i> /	14	--	--	14	--	--
/ <i>ɬ</i> /	4	--	--	5	--	--

/s/	6	--	--	7	--	--
5. /š/	6	--	--	3	--	3
6. /h/	3	--	1	--	--	3
7. Asp. stops and affricates	--	--	9	--	--	20

Several of the Chinese loans in Lai are 'exceptional' in tone and these forms have been omitted in the above count in order to clarify underlying patterns. Similarly, two particles with glottal onset on tone /2/ have been excluded: /ʔo2/ (negative) and /ʔe2/ (interrogative).

Three pairs of tones are indicated by the above distribution:

/1/ and /4/: see Category 4, with /b/ and /d/ from earlier *ʔb- and *ʔd, as shown by two key roots (two, water); for Categories 2a/2b/2c one can posit earlier preglottalized nasals, lateral, and glides (*ʔw- > v-, *ʔy- > ž-), or surd stop 'presyllabics', thus establishing unvoiced/unaspirated as the initial feature for this pair of tones. The three initial h- forms on tone /1/ can be explained by setting up an earlier *ʔg- (> h-), paralleling the *ʔb- and *ʔd-.

As for the basic /1/ vs. /4/ tone assignment, four sets of data clearly point to final glottalization for the latter as opposed to lack of same for the former:

(a) five forms on tone /4/ correspond to MK forms with final -ʔ, as generally reconstructed (although Huffman, above, has /cɔɔ/ and /kaa/ as his 'canonical forms' for 'dog' and 'fish', respectively) (dog, fish, road, rob, six); similar forms on tone /1/ are all, apart from /ʔi1/ '3rd sg.', either definite (sky/rain, head) or possible/probable loans (father, eat, this) based on vowel-final prototypes.

(b) all three Lai forms showing (original) unvoiced initial and final *-ak > -e shift, presumably via an intermediate *-eʔ (water, fowl, carry on back), are on tone /4/.

(c) four of the numerals (four, six, eight, nine) have tone (4) corresponding to Vn. sǎc tone, long known to be associated with earlier glottalized finals, as contrasted with two numerals (three, seven) that have tone /1/ corresponding to other tones in Vn. The glottalization in the nasal-final forms for 'four', 'eight' and 'nine' is probably to be explained as a "transfer" from an earlier prefixed *ʔa-, as exemplified by the Kuy numerals for 'two' through 'four' (see table above); cf. also /qa:m4/ 'husk', which corresponds to MK forms of /ʔa(ŋ)ka:m/ type.

(d) Category 3, with representation on tone /1/ but not on tone /4/, is readily explained on the basis of dissimilation: *ʔV(C)ʔ > /ʔV(C)/.

/3/ and /6/: see Category 7; also 6 (with note above); the three *š- forms on tone /6/ perhaps reflect *šh- forms (this initial lacking in Lai); for Categories 2a/2b/2c one can posit earlier preaspirated nasals, lateral and glides (*hɰ- > /v/, *hy- > /ʒ/, or aspirated (> *h-) 'presyllabics' from whatever source, with the anticipated gap of Category 1 forms on tone /6/: *h-p- > p- but not *h-p- > *ph-, etc. This all indicates ASPIRATED as the initial feature for this pair of tones, with some support for forms on tone /6/ (year, salt, listen/hear, laugh) and one form on tone /3/ (stone); note also the loan for 'cotton' on tone /6/, apparently reflecting an initial *hɰ in the Tai/KS prototype. No correspondences have been uncovered for initial l- forms on tones /3/ or /6/, which must have had a somewhat different origin from initial ɬ- forms, the latter perhaps all from an earlier *śl- (see 'die', below, for one possible example).

As for the basic /3/ vs. /6/ tone assignment, one would like again to set up an earlier final glottalized vs. non-glottalized basis, but the available Lai material hardly lends itself to this view. It may be significant, however, that both Lai forms that correspond to MK forms with final *-s (fly, laugh) are on one or the other tones of this pair.

/2/ and /5/: see Category 2a, with nasal-initial and l-initial forms best represented on tones 5 and (especially) /2/, pointing to VOICED as the initial feature for this pair of tones. This also accounts for the Category 2b and 2c forms on these tones, the latter mainly on /5/, as well as for the three initial ɣ- forms on tone /2/ (see above). The fact that Category 1 initials are well represented on these tones would suggest that original voiced stop and affricate initials were simply unvoiced in Lai, with transphonologization of the voicing to tones /2/ and /5/, but it appears that complex conditioning factors were at work here (see below).

Although numerous problems in Lai phonology remain unsolved, the following points are worthy of note:

1. Lai lacks final palatals and cognate sets for establishing the reflexes for these characteristic MK elements appear to be lacking.

2. Lai has transphonologized final *-ʔ at least in part (see above under tones /1/ and /4/), but appears to have retained all other non-palatal final stops and nasals, apart from the *-ak > -e shift (water, fowl, carry on back, child).

3. Final *-h > Ø in the one available example (go).

4. Final *-s > -i, with raising of the preceding vowel (fly, laugh).

5. Final *-r and *-l merged in Lai, yielding -n after long /a:/ (salt, ten) but -i (*-ir ~ *-il > *-ii > -i) elsewhere (seven, two, firewood).

6. *ʔb- > b- (two), *ʔd- > d- (water), perhaps also *ʔg- > h- (see above under tones /1/ and /4/). On the basis of Lai b- or d- in a

Lai/MK cognate set one can reconstruct the preglottalized initial at the PMK level, as in 'rice (early)'.

7. *p > p- in several good cognate sets (three, four, six, seven, 1st sg. (excl.)).

8. *b > p- in two possible Lai/MK cognate sets (father, carry on back) for which a loan (< TB) origin cannot be excluded, especially since they involve the /1/ ~ /4/ pair of tones, associated with (original) surd initials. It is possible, however, that in both these forms the initial was unvoiced by a 'presyllabic': *ʔaba > puɔ¹ 'father'; *tabak > be⁴ 'carry on back'. In a third likely cognate set, a 'presyllabic' r- may have produced aspiration: cf. Lai phyə⁶ 'clothes', PW *r(m)beʔ 'id.' (provisional reconstruction), with secondary palatalization.

9. *t- > th- in the one certain MK/Lai cognate set (egg), contrasting with *p- > p- (above) as well as with *t- > t- in one certain loan (thousand) and one possible loan (grandfather); cf. the initial tsh- form under 'eat' (below). No MK/Lai cognate sets have been uncovered for the anticipated *d > t- or th- shift.

10. Lai has two reflexes for PMK *c, as customarily reconstructed: /ts/ in 'dog' and perhaps /ts/ and/or /tsh/ in the set of forms under 'eat' (possible loans from TB) but /ʃ/ in 'nine'. The latter cannot be viewed simply as a substitute for /ts/ before /ə/ in view of Lai tsən¹ 'get' and ʒɔ³ tsəm⁴ 'female shaman' but one might argue that it has been conditioned by an original following *i, as indicated for this MK root. It is also possible, however, that Lai has maintained distinct reflexes for a PMK *ts vs. *c contrast, as provisionally set up by the writer (ATLC: 472 ff.).¹¹ PMK 'presyllabic' *c- is perhaps represented by Lai s-; cf. Lai sa:i⁴ 'know/know how'; S. Khmu chay (< *c-hay) 'know', with *sha:y as a likely intermediate form (Lai lacks *sh-). In yet another root PMK *ts- or *c- may be reflected in secondary ɬ- (< *sɬ- or *ʃɬ-); cf. Lai ɬet¹ 'die', perhaps an -l- infix derivative of the MK root represented by Mon kcit, Bahnar kəcit, Sre cət, Thavung cəət, Vn. chăt, et al.

11. *k- > q- in two MK roots (fish, husk) and one early loan (sky/rain), but *k- > k- in two relatively late loans (needle, horn) as well as in prefixed ka¹-, apparently as the result of reduced stress. Lai also has *kr- > k- (road), the pattern of velar shifts here precisely paralleling that

¹¹ The MK/Munda comparisons appear to require the setting up of *ts vs. *c at the PAA level, at any rate, while certain "irregularities" in MK reflexes apparently point in the same direction. Diffloth (1977b) has attempted, without much success, to explain away these variations, in part making use of ad hoc arguments. Ferlus (1978), on the other hand, has reconstructed both *ts and *tʃ at the PMK level, along with *c, but in a p.c. (6/3/82) does a volte-face, reconstructing simply *c (or *tʃ) "for the moment". It is to be hoped that in time Lai will supply a definitive answer to this problem, presumably in favor of PMK *ts > Lai /ts/ vs. *c > Lai /ʃ/.

found in Lahu, a BL language spoken in China (Yunnan) as well as in Burma, Thailand and Laos (possible areal factor).¹² The **kr-* > *k-* shift was early, antedating the final **-ak* > *-e* shift, as shown by the loan for 'fowl': *kʏeʔ*, with *-y-* for **-r-*, from a TB prototypical **krak* < **k-rak*. In view of the evidence for these shifts one can rule out the comparison: Lai *kɛ:tʔ* 'pry/prize'; Khmu *kɛʔt* 'id.', inasmuch as Khmu has /*kr*/ as well as /*k*/.

12. **g-* > *qh-* in the only available MK/Lai cognate set (mountain), supplying further evidence for reflexes of this kind in Lai (cf. 'clothes', above). Lai has *q-* for a prototypical **g* or **k* in loans for 'run' and 'animal enclosure', with **gr-* a less likely alternative for the latter. It would appear, in any event, that Lai /*k*/ does not serve as the reflex for MK **g*, thus weakening the case for a possible loan from Kadal: Lai *kuiʔ* 'tiger/leopard'; P-Li *[g]uiʔ* (< **guiʔ*) 'leopard': *gui* ~ *hui* ~ *ɣui*. The tonal pattern (above) suggests an early **?g-* > *h-* shift in Lai but no comparative evidence in support of this has been uncovered.

13. Lai has /*ʒ*/ for a prototypical initial **ry-* in the loan for '100', with the medial **-y-* probably playing a role here; cf. the regular /*ʒ*/ for **y-* in 'go'. In another loan from TB (fowl), medial **-r-* is represented by Lai *-y-*. In a secondary cluster of MK provenience, **pr-*, Lai has the vocalized equivalent /*i*/, with maintenance of the 'presyllabic' **p-* (**paruʔ* > *piuʔ* 'six'). In two similar roots with voiced 'presyllabic', however, Lai has /*y*/ after **b-* in 'thread', with loss of the 'presyllabic', and /*hy*/ after **m-* in 'rob', apparently with replacement of the 'presyllabic'. In view of the evidence for PMK **r* vs. **y*, at least in final position (ATLC: 470), however, one can hardly without additional data preclude the possibility that Lai has maintained distinct reflexes for this pair of phonemes.

14. Lai has a curious secondary *-y-*, always before /*a:*/, in words of MK provenience (after initial *l-* in 'eagle', *ɬ-* in 'sun' and *m-* in 'salt'); also perhaps after *l-* in the loan for 'child', where there is an alternative view of the *-y-* as prototypical (*ka¹ lye²* < **lyak*). Prefixed *ka¹-* also appears in 'sun', apparently after loss of an earlier 'presyllabic' of the same type (**khl-* > *ɬ-*), while the MK root for 'eagle' has 'presyllabic' **k(a)-*, replaced in Lai by /*muɔʔ*/ (see above under 'mother'). It would seem that this element has triggered the secondary *-y-* in these forms, with perhaps another 'presyllabic' element conditioning the same feature in 'salt' (S. Khmu has /*hma:r*/).

15. The most prominent vowel shift in Lai is that from **-a* to *-ɔ* ~ *-uɔ* (after labials) in final position in loans (100, sky/rain, come) and before final **-ʔ* (with loss of **-ʔ*) in forms of MK provenience (fish, house, road, rob), with representation also in a group of forms of uncertain etymology

¹² See Matisoff 1969, section 6.2. [Ed.]

(father, mother, Fa.'s o. Bro.'s wife, eat). This shift appears to be relatively late, with **a* maintained before close juncture (> crasis), as shown by /*mau*⁵/ 'not come', from /*mu*ɔ²/ 'come' + ʔo³ (negative), with shift to tone /5/. The same phenomenon apparently underlies the doublet forms: puɔ¹ ~ pa⁵ 'father', the latter in /pa⁵ ʔa:u¹/ 'my father', with the same shift to tone /5/. In addition, the puzzling form ta⁵ 'grandfather', on the same tone /5/, looks like the doublet of a missing (or unreported) *tɔ¹. It is perhaps significant that at least one (come), possibly all three, of this trio of forms/doublets is of loan origin, suggesting an earlier **a* > /ɔ/ shift for forms of MK provenience.

16. Another prominent vowel shift involves the raising of **o* (four, go) and **ɔ* (dog) to /u/, operative before final **-k* (hair) but not before final **-ŋ* (mountain). It is possible that **o* has been maintained in /ʔo⁵/ 'not go' < /ʔu/ 'go', paralleling the form for 'not come' (above), but in this case the shift can be viewed as simply assimilative, with further data required to settle the point.

17. A parallel raising of **e* to /i/ in Lai is found in 'you' and perhaps 'this' and '3rd sg.' as well. In addition, Lai bi¹ 'two' shows a similar shift from an earlier **ʔbiar* or the like (see above)¹³, probably from **bir* (with regular -i(i) for final **-r*) via an intermediate **bœr* (cf. early Khmer *ver*). An original **e* was retained in Lai, however, before final **-ʔ*, as shown by 'rice (early)' and perhaps also 'clothes' (see above). The distinct Lai reflexes here for 'you' and 'rice (early)' provide excellent evidence for setting up a final **-V* vs. **-Vʔ* distinction at the PMK level.

18. Before final **-s*, both **a* and **ɔ* were first fronted (> **e*) before a secondary **-i* (< **-s*), then raised to /i/, as above (fly, laugh).

19. Lai has /au/ for /au/ ~ /o/ or /u/ elsewhere in two forms (stone, deep) and a parallel /a:i/ for /ai/ ~ /e/ elsewhere in a third (three), suggesting PMK reconstructions of /ə/ plus glide type, perhaps with a length (:) factor.

20. Lai has dropped the **u* of an original **-uay* ~ **-uəy* after initial labial in 'one' and '1st pl. (excl.)', paralleling the loss of a prototypical **-w-* after **l-* in the loan for 'ox/cattle' and after **ŋ-* in the loan for 'dumb'.

¹³ Shorto (1976) has reconstructed both **iə* and **uə* at the PMK level while Diffloth (1977) sets up the former for P-Semai, but MK comparativists have shied away from reconstructing either **ia* or **ua* at an early level. Thus Shorto reconstructs PMK **biʔaar* > ʔbiə(a)r 'two' rather than the simpler **ʔbiar* or **ʔbiar*. Actually, a third falling diphthong: **ia* (~ **iə*) seems to be required to handle the MK forms for 'water', where Shorto favors **diʔaak*, to explain Khmer *dək* in the face of /daak/ forms generally (see text, but note Vn. *núoc*, with *n-* < **ʔd-*); here PMK **ʔdɛak* is more natural (Lai *də** can be derived simply from **ʔdaak*; see text). A PMK system of falling diphthongs of this type: **ia* ~ *iə*, **ua* ~ *uə* and **ia* ~ **iə* can be viewed simply as an expansion (MK "specializes" in vowels) of an areal feature represented also in Tai: PT **ia* ~ **ua* ~ **ia* or (Li 1977) **ia* ~ **ue* ~ **ie*. Again it is to be hoped that in time Lai will throw some light upon this murky problem.

21. Lai has /i/ for prototypical *ɿ in two loans from Tai (firewood, fish scales) but no cognate sets involving Lai or MK /ɿ/ have been uncovered.

22. Finally, Lai shows secondary vowel length, of uncertain origin, in a number of forms, e.g. /ma:t²/ 'eye'. It should be noted that in Mon-Khmer, in sharp contrast to the adjoining Sino-Tibetan stock, it is long rather than short vowels that are 'regular' (much more common), as reflected in P-Semai (Diffloth 1977a) and elsewhere, hence it may be that Lai is simply pursuing a basic MK trend in its secondary vocalic lengthening.

We do now, after this review of all the available material, have an answer to the question: what is Lai? It is at heart a MK language but has been "Kadaicized": reduced largely to monosyllabic forms, thus stripping it of virtually all of its morphology, and provided with a complex tonal system. In all of this it closely resembles Vietnamese, with a specific parallelism in the tonal development inasmuch as initial voicing and final glottalization are prominent conditioning factors in both cases. Each language, in addition, has a large corpus of loans from Tai, but Lai has added a further complication: a body of loans from TB. Maspero, a pioneer in the Southeast Asian field, pronounced Vietnamese to be a congener of Tai; one can only wonder what he would have done with Lai.

Still another question remains to be answered: where does Lai fit within the MK family? As in the field of TB studies, there is a consensus as to what the writer has called "nuclear" groups within MK, but much disagreement as to any higher order arrangement. Ferlus (1979) simply places all these "nuclear" groups on a single line, thereby evading the issue; others have tended to separate off one or both of the two geographically distant groups (Khasi, Aslian), to classify Khmer with Pearic, Palaungic with Khmuic ("Northern MK") and Monic with Bahnaric and Katuic, while either keeping VM separate or grouping it with Monic, et al. Headley (1976) shows VM as separated on the basis of shared cognate percentages but classified with Monic, et al. on the basis of lexical innovations. Additionally, there appears to be a consensus that the "center of gravity" of the family lies in Monic/Bahnaric/Katuic.

One can hardly carry out a study of shared cognate percentages with the limited available material on Lai. One can, however, take into consideration the striking fact that Lai has cognates for only half (5) of the "highly persistent" lexical items, excluding the numerals, for which Lai forms are available (total of 10), in a list compiled by Huffman on 19 MK languages (see above), with representatives from all the above groups other than Khasi and Aslian. Huffman states that these core items are "highly persistent", with perhaps one or two exceptions each, hence the low rate of retention exhibited by Lai surely is to be regarded as exceptional, placing the language well outside this "inner grouping" of the MK family.

The same Huffman list also includes the numerals from 'one' through 'four' (see above), well maintained in Lai and most MK groups, including even Aslian (Semaq Beri), but exhibiting marked innovation in Palaungic and VM as well as in Khasi. Lai has innovated for 'five', but for 'six' through 'nine' appears to have retained the original set of MK numerals, along with Monic, Bahnaric and VM, while for '10' it shares a root with VM as opposed to a Monic/Bahnaric/Katuic root (see details above). Apart from 'five' and perhaps '10' Lai thus has been unusually retentive in its treatment of numerals. If the VM/Lai root for '10' is the innovative one of this pair, it would indeed constitute a strong VM/Lai link, but one can hardly exclude the possibility that Lai has also been retentive here, as elsewhere in the numeral system and in the pronominals (below). As Zide (1976) has argued, one must posit the existence of a decimal system at the PMK stage despite the competing forms for '10', the only question being whether either of the above pairs of roots represents the PMK root for that numeral. A parallel situation obtains in the TB/ST field, where the 'winner' among a number of competing TB roots or forms for '10' is PTB **(s-)gip*, with markedly restricted distribution, on the basis of the Chinese cognate: AC/MC **sɣiəp/zɣiəp*. Can it be that Lai/VM **ma:l* is the **(s-)gip* of MK?

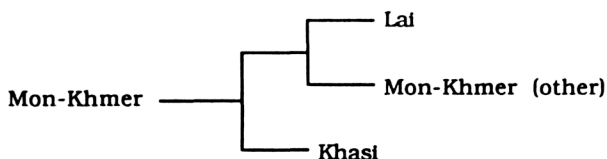
The Lai personal pronouns pose even more difficult problems. Lai has retained both the 2nd and 3rd person sg. forms while innovating for the corresponding plural forms, but specific links shown by the 1st person pronouns are altogether remarkable: root of */ʔaw/* type for '1st sg.', with links to Bahnaric and Aslian, contrasting with the widespread MK/Munda root of */iŋ/* type; root of */puay/* type for '1st pl. (excl.)', with link only to Monic ('we'); root of */ʔay/* type for '1st pl. (incl.)', with link to '1st dual' forms in Palaung, Aslian and even Nicobarese, speculatively derived by Pinnow from earlier '1st pl.' forms.

What to make of all this? The */ʔaw/* root must be regarded as innovative at an early (possibly PMK) level, probably related to status, paralleling later distinctions of this kind made by Mon, Bahnar, Mnong and Khasi (see Headley 1976: table on p. 441). As for the '1st pl. (excl.)' and '1st pl. (incl.)' forms, Pinnow was struck by the great variety of forms for these two pronouns throughout MK/Munda, and it is entirely possible that Lai has maintained both pronouns, as noted above. This makes good sense of it all, at any rate, requiring only a loss of the excl. vs. incl. distinction in the Mon cognate ('we') and a semantic shift to '1st dual' in Palaung, et al., as per Pinnow's speculation.

Apart from the numerals and pronominals, the specific ties shown by other lexical elements in Lai deserve special mention. These include four with Palaungic (egg, listen/hear, mountain, road), one with Palaungic/Khmuic (rob), two with Khmuic (go, salt), one with Palaungic/Aslian (long), one with Aslian (rice) and two with VM (laugh, sun).

This survey is preliminary in nature and it is not unlikely that cognates for one or more of these roots will in time turn up in Bahnaric or elsewhere in MK. The peripheral, especially northern, distribution of these special lexical links does appear to be of some import, however, e.g. the PNB/PJH (Bahnaric) sources are fully as rich, taken together, as the PW (Palaungic) source, yet they have failed to provide even one specific tie of this sort. The overall impression here, as with the numerals and pronominals, is of a MK language relating to the family as a whole, apparently at a considerable time-depth, exhibiting a high degree of independence.

It appears, then, that one must at the very least view Lai as a distinct branch of MK, on an equal footing with any of the others, rather than to attempt to group it with Palaungic, Khmuic, VM or some combination of these. It is even possible that additional material on the language in the future will provide evidence for a classification scheme of the following type:



The position to be assigned Lai is a matter of some importance in view of the loan evidence (above) that points to a western origin for the language. This fact, together with the location of Khasi in Assam, suggests that the MK homeland may have been the Salween basin, where the Waic languages are located, shortening the distance eventually traveled by the Khasi and those more distant MK congeners, the Munda peoples of India.

Two general points remain to be made here. The first involves the usefulness of numeral systems in an 'exercise in classification' such as this. It would appear that SEA linguists have become unduly skeptical in this area of the lexicon, largely perhaps as a result of having seen their colleagues led astray in the past both by the Tai/Kam-Sui and the Miao-Yao numerals, which in each case point to ST while the core lexicons point elsewhere, specifically in the direction of Austro-Tai. In general, however, numeral systems have been reasonably well maintained in SEA, e.g. Vn. has retained the basic MK numerals, with some innovation for the first two (see above), despite the massive "Kadaicization" at an early period and the later "Sinicization". One can even lay down a rule here, viz. numeral systems (especially substantially complete ones) + (other) core vocabulary items = genetic relationship. In the case of Tai/KS, the first two numerals "escaped" the replacement process, thereby arousing an element of suspicion in the knowledgeable comparativist; in the case of MY, at least four of the first five numerals cannot be fitted into any borrowing scheme, which

involves the numerals from 'six' through '10' - a red flag of warning for comparativists. Lai has innovated only for 'five', for which other MK languages show much variation, and perhaps (along with VM) for '10', although this is a moot point. It also has a sizeable number of forms with evident MK cognates of "core vocabulary" type, quite different as an assemblage from the groups of TB and KD loans to be found in the language. The phonology is all "good KD" but can be tied in with that of MK without too much difficulty, as shown above, even with the limited available Lai material. Of MK morphology there appears to be hardly a trace; even the prefixed *ka¹-* is part of a widespread SEA areal feature. This is all to be anticipated, however, in view of the leveling effect of monosyllabization. Actually, if one were to follow the rule (above) in looking over the Lai material, the basic MK provenience of the language would be evident from the very outset.

The second and final point concerns the usefulness of a limited corpus of material such as we now have for Lai. The writer has faced this problem from the very beginning of his studies in the SEA field, especially in the case of three key northern Kadai languages (Gelao, Laqua and Lati), each of which was known only from very limited word lists and odds and ends of other material. He could have postponed any analysis of these languages until such time as more material might become available but, fortunately, he chose not to; only very recently has there been any significant change in this situation and only in the case of Gelao, with both Laqua and Lati still largely unreported. In a sense, then, one is indulging in a form of "linguistic prediction" (Benedict 1979c) when he works with limited material in this fashion, e.g. in the case of Lai he is "betting" that the choice of illustrative examples by Liang was not by chance "skewed" in some manner, thus having the effect of making the language look "more MK" than it really is or, perhaps, minimizing specific links with this or that group within MK. One ordinarily does have a complete set of numerals, however, at times also the pronominals, as in the case of Lai, hence all is not entirely by chance. For Lai, perhaps also now for Laqua and Lati, we have some prospect of obtaining much more material in the not-too-distant future; one can only hope to see his predictions confirmed.

REFERENCES

Abbreviations:

AS	Austroasiatic Studies (see Jenner, et al. 1976)
MKS	Mon-Khmer Studies
MZYW	Minzu Yuwen

- Benedict, P. K. 1972. *Sino-Tibetan: A Conspectus* ("STC"). Contributing Editor, James A. Matisoff. Cambridge University Press.
- _____. 1975. *Austro-Thai: Language and Culture* ("ATLC"). New Haven: HRAF Press.
- _____. 1976. "Austro-Thai and Austroasiatic." AS 1:1-36.
- _____. 1979a. "Austroasiatic loanwords in Sino-Tibetan." Helsingør Symposium on Austroasiatic Linguistics and Literature. To be published by the Scandinavian Institute of Asian Studies, København (J.-O. Svantesson, Ed.).
- _____. 1979b. "Vocalic transfer: a Southeast Asia areal feature." *Acta Orientalia* 40:229-52.
- _____. 1979c. "Linguistic prediction: the case of Saek." Sino-Tibetan Conference IX. Paris.
- _____. 1980. "The PST tone/accent system: additional TB data." Sino-Tibetan Conference XIII. Charlottesville.
- Benjamin, G. 1976. "Austroasiatic subgroupings and prehistory in the Malay Peninsula." AS 1:37-128.
- Delcros, H. 1966. *Petit dictionnaire du langage de Kmhmu' de la région de Xieng-Khouang*. Vientiane (mimeo.).
- Diffloth, G. 1977a. "Towards a history of Mon-Khmer: Proto-Semai vowels." *Tōnan Ajia Kenkyū* (Southeast Asian Studies) 14:463-95.

- _____. 1977b. "Mon-Khmer initial palatals and 'substratumized' Austro-Thai." *MKS* 6:39-57.
- _____. 1980a. "The Wa languages." *Linguistics of the Tibeto-Burman Area* 5.2:1-182.
- _____. 1980b. "Austroasiatic, the indigenous Southeast Asian language family." Association for Asian Studies, 32nd Annual Meeting, Washington, D.C.
- Ferlus, M. 1978. "Reconstruction de /t s/ et /t š/ en mon-khmer." *MKS* 7:1-38.
- _____. 1979. "Lexique thavung-français." *Cahiers de Linguistique Asie Orientale* 5:71-94.
- Headley, R. K., Jr. 1976. "Some considerations on the classification of Khmer." *AS* 1:431-51.
- Henderson, E. J. A. 1976. "Vestiges of morphology in modern standard Khasi." *AS* 1:477-522.
- Huffman, F. E. 1976. "The relevance of lexicostatistics to mon-khmer languages." *AS*, 539-74.
- _____. 1977. "An examination of lexical correspondences between Vietnamese and some other Austroasiatic languages." *Lingua* 43:171-98.
- Jenner, P. N. 1976. "Les noms de nombre en khmer." In G. Diffloth and N. H. Zide (Eds.) *Austroasiatic Number Systems*. *Linguistics* 174.
- Jenner, P., S. Starosta, and L. Thompson (Eds.). 1976. *Austroasiatic Studies*. 2 vols. Honolulu: University Press of Hawaii.
- Li, F-K. 1965. "The Tai and Kam-Sui languages." *Lingua* 14:148-79.
- _____. 1977. *A Handbook of Comparative Tai*. Honolulu: University Press of Hawaii.
- Liang, M. 1984. "A brief description of the Lai language." *MZYW* 1984.4 (in Chinese).

- Lindell, K., J-O. Svantesson, and D. Tayanin. 1976. "Phonology of Kammu dialects." Sino-Tibetan Conference IX. København.
- Luce, G. H. 1965. "Danaw, a dying Austroasiatic language." *Lingua* 14:98-129.
- Matisoff, J. A. 1969. "Lahu and Proto-Lolo-Burmese." *OPWSTBL* Vol. I, 117-221. Ann Arbor.
- . 1972. *The Loloish Tonal Split Revisited*. Research Monograph #7, Center for South and Southeast Asian Studies. Berkeley: University of California.
- . 1985. "God and the Sino-Tibetan copula, with some good news concerning selected Tibeto-Burman rhymes." *Journal of Asian and African Studies* (Tokyo), 29.1-81.
- . 1988. *The Dictionary of Lahu*. Berkeley, Los Angeles, London: University of California Press.
- Pinnow, H. J. 1965. "Personal pronouns in the Austroasiatic languages." In G. B. Milner and E. J. A. Henderson (Eds.) *Indo-Pacific Linguistics Studies*. Amsterdam: North-Holland Publishing Co.
- Shorto, H. 1963. "The structural patterns of Northern Mon-Khmer languages." In H. Shorto (Ed.) *Linguistic Comparison in South East Asia and the Pacific*. London: School of Oriental and African Studies.
- . 1972. "The word for 'two' in Austroasiatic." In L. Burnet and J. M. C. Thomas (Eds.) *Langues et Techniques, Nature et Société*. Paris: Klincksieck.
- . 1973. "Mon-Khmer contact words in Sino-Tibetan: Commentary on Paul K. Benedict, *Sino-Tibetan*." (Mimeo.)
- . 1976. "The vocalism of Proto-Mon-Khmer." *AS* 2:1041-67.
- Smalley, W. 1961. *Outline of Khmu? Structure*. New Haven: American Oriental Society.
- Smith, K. D. 1972. *A Phonological Reconstruction of Proto-North-Bahnaric*. Santa Ana: Summer Institute of Linguistics.

- . "The velar animal prefix relic in Vietnam languages." *Linguistics of the Tibeto-Burman Area* 2:1-18.
- Svantesson, J-O. 1983. *Kammu Phonology and Morphology*. Lund: Travaux de l'Institut de Linguistique, XVIII.
- Thomas, D. 1976. "South Bahnaric and other Mon-Khmer numeral systems." In G. Diffloth and N. H. Zide (Eds.) *Austroasiatic Number Systems*. Linguistics 174.
- Thomas, D. and M. Smith. 1967. "Proto-Jeh-Halang." *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 20:157-75.
- Zhou, Z-Z. and Q-X. Yan. 1983. "A brief description of the Blang language." *MZYW* 1983.2 (in Chinese).
- Zide, H. H. 1976. "Introduction." In G. Diffloth and N. H. Zide (Eds.) *Austroasiatic Number Systems*. Linguistics 174.