

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 The speakers reportedly number only several hundred, all in Guangxi, the language having been replaced by Yi (Lolo) in Yunnan.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-tɔŋ*. This loan also appears in two Kadai languages of North Vietnam: Laqua *təŋ*, Latl *tɔŋ*.

'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 **ɣam⁶* < **ɣwam* reflecting an earlier prototype than any of the known Tai forms of the present day: Ahom (SWT) *bum* < **wum*; CT *waam* ~ *vam* ~ *vɔɔm*, NT *ɣɔm* ~ *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 ɣam⁶/* '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/*.