Formation of Ethnonyms in Southeast Asia

Michel Ferlus
Independent researcher
(retired from Centre National de la Recherche Scientifique, France)

In the Southeast Asian Sinosphere, we can observe a circulation of ethnonyms between local languages and Chinese. A local Southeast Asian autonym is borrowed into Chinese, undergoes sound changes affecting the Chinese language, and then returns to the original populations, or to other populations. The result is a coexistence of ethnonyms of highly different phonetic outlook but originating in the same etymon.

We will examine four families of ethnonyms, mostly Austroasiatic and Tai-Kadai. Some of these ethnonyms are still in use today, others are known through Chinese texts where they are transcribed by phonograms. We will use Baxter [1992] and Pulleyblank [1991] to reconstruct Old Chinese (OC) and Middle Chinese (MC) forms. The phonetic restitution of the phonograms are typeset in boxes, for clarity.

1. The *b.rak (> bái 白) or *p.rak (> bǎi 百) family

From these etymons originate the following ethnonyms:

1.1 Sino-Vietnamese Lac is the most ancient term to name the Vietnamese people. Lac Việt (luòyuè 麗越) “population of ancient Vietnam” [Maspero 1918, p.7]; also spelled 駃 or 歸 [Madrolle 1937, p.323].

From these etymons originate the following ethnonyms:

1.2 *prɔːk “autonym of the Wa people” and rɔ:k, “a Khmu sub-group”. In the Waic dialects [Diffloth 1980], the etymon *p.ʁɔːk can be realized pəzaŋ, pərauŋ, pʰalok, etc. The shift aː > oː is attested in Bolyu [Edmondson 1995], an Austroasiatic language which has affinities with Vietic. This marked change might have existed in a larger area in South China.

<table>
<thead>
<tr>
<th></th>
<th>Waic</th>
<th>Khmu</th>
<th>Viet</th>
<th>Bolyu</th>
</tr>
</thead>
<tbody>
<tr>
<td>“fish”</td>
<td>kaʔ</td>
<td>kaʔ</td>
<td>cá</td>
<td>qo</td>
</tr>
<tr>
<td>“house”</td>
<td>naʔ</td>
<td>nhà</td>
<td>ɲa</td>
<td>ɲo</td>
</tr>
<tr>
<td>“leaf”</td>
<td>hlaʔ</td>
<td>hlaʔ</td>
<td>lâ</td>
<td>10</td>
</tr>
<tr>
<td>“road”</td>
<td>kraʔ</td>
<td>sá</td>
<td>kya</td>
<td></td>
</tr>
</tbody>
</table>

I cordially thank Frédéric Pain (Catholic University in Leuven, Belgium), a linguist specialist in Southeast Asia, who read the text over with the greatest attention.
The filiation of \texttt{prø:k} and \texttt{rc:ik} with the etymon *\texttt{b.rak}/*\texttt{p.rak} can be proven by the vowel shift \texttt{a} \rightarrow \texttt{o}.

1.3 Chinese texts from the 13th and 14th centuries inform us about populations named \textit{Bái yì 白衣 “White Clothes”, Bái yí 白夷 “White Barbarians” or Bái yí 白夷 “The Hundred Barbarians” [Luce 1958, pp.174-5, notes 11-13]. The existence of these variants suggests that the characters are used as phonograms to transcribe an ethnonym, i.e. *\texttt{p.rak} or *\texttt{b.rak}. Likewise for \textit{Bái mán 白蠻 [Luce 1961, p.24]. As far as the characters \textit{yì 衣 “clothes” and yí 寓 “barbarian” are concerned, they may simply transcribe a word meaning “People” (see §2 below).

1.4 In my talk held at the 41st ICSTLL in London [Ferlus 2008], I interpreted \textit{Báiyuè 百越 as “the Hundred Principalities”. I now think that it should be “the Principality of the \textit{p.rak}”.

1.5 The ethnonym Bai/Bai (\textit{Báizú 白族}) is also likely to originate in *\texttt{p.rak}/*\texttt{b.rak}. The Bai people was formerly named Minjia (\textit{Minjiă 民家}).

Note: Nowadays, some ethnic groups in Southeast Asia and South China are named with ethnonyms including a term interpreted as “white” : White Yi, White Thai, etc. We wonder whether it had originally acted like the ethnonym *\texttt{b.rak} which, over the course of the centuries, would have been reinvested by the first meaning of its phonogram \textit{bái 白}. Thereafter, ethnonymic compounds comprising the term “black” would have been introduced to satisfy a need of harmony.

2. The \texttt{*k.ri}: “(old name of) Tai-Kadai (peoples)” family

The alleged original form is well preserved, totally or partially, in some Mon-Khmer languages in Laos and Vietnam [Ferlus 2006].

\texttt{ri}: refers to the Lao in the Keneng language (Khmuic, Houa Phan province, Laos).

\texttt{raj} refers to the Tay/Tai (here Tây Mường) in the Ôdu/Őđu language (Khmuic, Nghế An province, Vietnam).

\texttt{yì}: \texttt{(yì:<*ri)} designates the Tay/Tai (here Black Tay) in the Kesenig Mul language (Khmuic, Sơn La province, Vietnam).

\texttt{kòri}: in the expression Maleng Kari \texttt{mòleŋ kòri}, autonym for the Kha Phong (Vietic, Quang Bình province, Vietnam), in which Maleng means “Human being, person, people”.

The three words, \texttt{raj}, \texttt{ri}, and \texttt{yì}, clearly means “Tai/Thai speaking population”. This meaning is probably the same in the expression Maleng Kari \texttt{mòleŋ kòri}, that can be interpreted “Maleng (who lives near) Lao”, the same pattern found in Maleng Brô \texttt{mòleŋ bu}: “Maleng (who lives near of) Bru”, the autonym of a neighbouring Katuic population (Khammuon province, Laos). In a wider sense, the specifying term \texttt{Kari} is also used in its Vietnamese form \texttt{Trì}, to name the \texttt{Tia Rì, or Chà Lý}, a local branch of the Bru (Katuic, Quảng Bình and Quảng Trị provinces).

To explain the phonetic shifts from the original forms to the modern terms we just have to use some simple sound shifts attested in Chinese and in the Sinospheric at large.

- Vowel shift -i- \rightarrow -aj
- Dialectal random shift r- \rightarrow l-
- Regular shift of l (medial or initial) from OC to MC [Baxter 1992]:
  1. Medial in tense sesquisyllable : OC \texttt{-l} \rightarrow MC \texttt{d-}
  2. Initial in lax monosyllable : OC \texttt{l-} \rightarrow MC \texttt{j-}
- Regular shift r- > l- from OC to MC
- Loss of the first syllable in sesquisyllables

From the *k.ri: etymon originate the following ethnonyms:

2.1 Hlai tfoot{tai}, or Lai, is the autonym of a Kadai population (Tai-Kadai family), located in Hainan (Guangdong). According to Matisoff [1988: 289, note 3], who reconstructed Proto-Hlai initials and tones, t- would be a voiceless lateral spirant which determines high series tones. Thus, this initial must originate in a cluster of *plosive+I. Etymologically, the form tfoot{tai} perfectly results from k.ri: by a succession of typologically well-known changes: k.ri: > kʰli: > tfoot{i}i: > tfoot{taj}. These shifts needn’t to be explained by a remodelling through the Chinese language.

2.2 Li (lixir, li 黎) is the name given to the Hlai by Chinese. This ethnic naming appears in the Han dynastic records [HCTE, VI §2] and also during the Song period.

| Llixir < MC liX < OC *rjι [rj?] |
| Lixi < MC leJ < OC *C-rij [C.ri] |

2.3 Tai/Thai (or Tay/Thay), from Proto-Thai tfoot{daj^A} (< *daj), is the autonym of populations of the Thai branch (Li Fangkuei’s Southwestern Tai) and of the Tai branch (LFK’s Central Tai). It is absent from the Yay branch (LFK’s Northern Tai). Its formation can be explained by changes specific to Chinese on the basis of a dialectal alternation kɔliː > kɔdiː/kɔdaj (-l- -> -d- in tense sesquisyllable) > diː/daj (loss of first syllable) > tfoot{daj^A} (Proto-Thai) which evolved to tfoot{taj^A2} in Thai/Siamese and in Lao and to tfoot{taj^A2} (in other Tai languages).

Note: a deeply rooted belief in Thailand has it that the term ‘Thai’ derives from the last syllables -daya in Sukhodaya/Sukhothai (สุโขทัย), the name of the first Thai Kingdom. The spelling emphasizes this prestigious etymology by writing 俸n (transliterated ai-d-y) to designate the Thai/Siamese people, while the form 俸n (transliterated ai-d) is occasionally used to refer to Hlai/Thai speaking ethnic groups. Lao writes 俸n (transliterated ai-d) in both cases.

2.4 Yi, spelled yì 彝 or 夷, an ethnic naming of some Tibeto-Burman groups, may originate in the etymon *k.ri: by the shift l- > j- (lax monosyllable) specific to Chinese.

| yì 夷 < MC yij [jì] < OC *ljj [li] |

The interpretation of the ethnonym ‘Yi’ as a member of the *k.ri family remains speculative, however it is not as firmly established as for Hlai and Tai/Thai.

3. The Kraw *k.raːw family

From the etymon *k.raːw originate the following ethnonyms:

3.1 Gelao: autonym of a Kadai population scattered from Guizhōu (China) to North Vietnam.

3.2 Lao: ethnonym and autonym of the Lao people (ລາວ); nationality of the inhabitants of Laos. Formed by the monosyllabization of the etymon *k. raw. The peoples named Lao (lào 獵), supposed to be the ancestors of Lao and some other Tai-Kadai populations, settled in the upper Tonkin and in parts of Yunnan and Guizhōu during the Tang times [Pelliot 1904, p.136].
lao 狼 <MC lawX <OC *C-raw? [C.raw2]

This reconstruction of the pronunciation for the phonogram 狼 confirms that ‘Lao’ originates in the etymon *kraw.

3.3 Sino-Vietnamese Giao, in Kiao-tche/ Giao Chi (jiāo zhī 交趾) “ancient name of Vietnam”, has been attested since 207 BC [Lê Thành Khôi 1981, p.66], subsequently to Lạc (see 1.1).

jiāo 交 <MC kæw <OC *kraw [k.raw]

The expression Giao Chi is represented by Catti- in Ptolemy’s Cattigara, and by Cochi- in Cochinchina, one of the names for the Đại Việt transmitted by the Portuguese.

3.4 Keo/ Kew ke:wʌ1 (kæw) : an ethnonym formerly given to the Vietnamese by the Lao, currently slightly derogatory. In Pupeo (Kadai branch), on the other hand kew is used to name the Tay of North-Vietnam. This ethnonym comes from the MC form of *kraw.

4. The *-raŋ ‘Human being, person’ family

It can be identified in the following expressions:

4.1 Sino-Vietnamese Ván-Lang (wén láng 文郎), name of a province and its population south of the territory of the Giao Chi, also North Vietnam (T'ang dynasty) [Maspero 1918]. This province might have been located in the centre of Laos.

láng 郎 <MC lang <OC *C-rang [C.raŋ]

4.2 T'ang-ming/ Táng-míng 堂明: according to the Records of Three Kingdoms (sānguózhì 三國誌), the T’ang-ming was a state situated in the north of Cambodia. A Chinese author from the 13th century identified this T’ang-ming with the Tao-míng/ Đào mìng 道明 of the Táng times, also located in the north of Cambodia [Hoshino 1986, p.32].

míng 明 <MC mjæ <OC *mrjang [m.raŋ]

4.3 Maleng: autonym of an endangered Vietic sub-group scattered on the Nakai plateau, in the centre of Laos. This ethnic naming can be identified to *m.raŋ through the iambic late form *m.leŋ.

4.4 Maleň/ Malyān: these terms, attested in the Old Khmer epigraphy, point to a country (and its inhabitants) situated south to the Great Lake. The Pre-Angkorian Maleň, reconstructed as *mšljan, is attested in the 7th century; the Angkorian malyān *mšljen had been attested from the 9th to the 11th century [Jenner 2009a, 2009b]. The Maleň/ Malyān kingdom was subjected by Jayavarman II in the 8th century. We don’t know anything about the language spoken by this population. There is an obvious phonetic relationship between the epigraphic Maleň/ Malyān and the Maleng in Laos, but we cannot draw any conclusion from this comparison.

4.5 Mrâîmâ: the Burmese epigraphy attest two variants for the ancient name of Burma, mraîmâ (since 1190) which have remained until today, and mrâîmâ (1342) [Luce 1959, p.53]. The base of the latter, mraî- (in which mî has the same value as m), is transcribed with the phonograms mî-lâng 木浪 [Luce 1958, p.106].

lâng 浪 <MC lanj <OC *C-rang [*C.raŋ]
We might think that the validity of the late form mran(mā) is uncertain, as it is rarer than the normal form mran(mā), but its transcription in Chinese is in perfect conformity with its spelling in Old Burmese.

4.6 Orang/Urang *ʔuranj “Human being, person” [Dyen & McFarland 1979; Wurm & Wilson 1975]. Although attributed to Proto Austronesian by some authors, the forms resulting from this etymon remain limited to the Malayic area (Melayu, Minangkabau, Madurese). It is difficult to place *ʔuranj in one of the linguistic families in Southeast Asia.

5. Conclusions

The first observation that emerges from this survey is that there is considerable phonetic variety within each family of ethnonyms.

Among the four fundamental etymons, only *k.ri: has been fully preserved until today, coexisting with its daughter forms Hlai/Li and Tai/Thai which name Tai-Kadai populations (leaving aside the dubious case of ‘Yi’).

The use of the daughter forms of *k.raw is common to Tai-Kadai (Gelao, Lao) and to Austroasiatic ancestors of the Vietnamese.

The etymon *p.rak/b.rak is mainly associated with Austroasiatic populations, and secondarily with Tibeto-Burman ethnic groups. It can be hypothesized that this situation reflects a spread of Tibeto-Burman languages over an Austroasiatic substratum.

The etymon *-raŋ is shared between the Asian mainland and the Western Indonesian world. It is attested today in three linguistic families: Tibeto-Burman, Austroasiatic and Austronesian.

References


Ferlus, Michel. 2008. Etymology of *wat/yuè 越 "people, principality" (as in Báiyuè 百越). 41st ICSTLL, September 17-21. SOAS, University of London.

Ferlus, Michel. 2009. What were the Four Divisions of Middle Chinese? Diachronica 26(2): 184-213.


*Some corrections : 2011, 2015*