



**HAL**  
open science

## Formation of ethnonyms in Southeast Asia

Michel Ferlus

► **To cite this version:**

Michel Ferlus. Formation of ethnonyms in Southeast Asia. 42nd International Conference on Sino-Tibetan Languages and Linguistics, Payap University, Nov 2009, Chiang Mai, Thailand. halshs-01182596

**HAL Id: halshs-01182596**

**<https://shs.hal.science/halshs-01182596>**

Submitted on 1 Aug 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

## 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

<i>bái</i> 白 < MC <i>bæk</i> < OC <i>*brak</i> [ <b>b.rak</b> ]
---

<i>bǎi</i> 百 < MC <i>pæk</i> < OC <i>*prak</i> [ <b>p.rak</b> ]
---

From these etymons originate the following ethnonyms:

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

<i>luò</i> 雒 < MC <i>lak</i> < OC <i>*C-rak</i> [ <b>C.rak</b> ]
--

Sino-Vietnamese *Lạc* **\*rak** results from the monosyllabization of **\*b.rak** or **\*p.rak** by loss of the first element in the iambic cluster.

1.2 **\*prɔ:k** “autonym of the Wa people” and **rɔ:k**, “a Khmu sub-group”. In the Waic dialects [Diffloth 1980], the etymon **\*p.rɔ:k** can be realized **pəzəək**, **parauk**, **p<sup>h</sup>alok**, etc. The shift **a: > ɔ:** 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.

	*Waic	Khmu	Viet	Bolyu
“fish”	<b>ka?</b>	<b>ka?</b>	<i>cá</i>	<b>qɔ</b>
“house”	<b>pa?</b>		<i>nhà</i>	<b>ɲɔ</b>
“leaf”	<b>hla?</b>	<b><sup>h</sup>la?</b>	<i>lá</i>	<b>lɔ</b>
“road”	<b>kra?</b>		<i>sá</i>	<b>kɲɔ</b>

The filiation of **prɔ:k** and **rɔ:k** with the etymon **\*b.rak/ \*p.rak** can be proven by the vowel shift **a: > ɔ:**.

1.3 Chinese texts from the 13th and 14th centuries inform us about populations named *Bái yī* 白衣 “White Clothes”, *Bái yí* 白夷 “White Barbarians” or *Bǎi yí* 百夷 “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. **\*p.rak** or **\*b.rak**. Likewise for *Bái mán* 白蠻 [Luce 1961, p.24]. As far as the characters *yī* 衣 “clothes” and *yí* 夷 “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 *Bǎiyuè* 百越 as “the Hundred Principalities”. I now think that it should be “the Principality of the **\*p.rak**”.

1.5 The ethnonym Bai/Bai (*Báizú* 白族) is also likely to originate in **\*p.rak/ \*b.rak**. The Bai people was formerly named Minjia (*Mínjiā* 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 **\*b.rak** which, over the course of the centuries, would have been reinvested by the first meaning of its phonogram *bái* 白. Thereafter, ethnonymic compounds comprising the term “black” would have been introduced to satisfy a need of harmony.

## 2. The **\*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].

**ri:** refers to the Lao in the Kenieng language (Khmuic, Houa Phan province, Laos).

**raj** refers to the Tay/Tai (here *Tây Mường*) in the Ödu/*Ođú* language (Khmuic, Nghệ An province, Vietnam).

**yí:** (**yí:** < **\*ri:**) designates the Tay/Tai (here Black Tay) in the Kesiing Mul language (Khmuic, Sơn La province, Vietnam).

**kǎri:** in the expression Maleng Kari **mǎle:ŋ kǎri:**, autonym for the Kha Phong (Vietic, Quảng Bình province, Vietnam), in which Maleng means “Human being, person, people”.

The three words, **raj**, **ri:** and **yí:**, clearly means “Tai/Thai speaking population”. This meaning is probably the same in the expression Maleng Kari **mǎle:ŋ kǎri:**, that can be interpreted “Maleng (who lives near) Lao”, the same pattern found in Maleng Brô **mǎlè:ŋ brɔ:** “Maleng (who lives near of) Bru”, the autonym of a neighbouring Katuic population (Khammuon province, Laos). In a wider sense, the specifying term *Kari* is also used in its Vietnamese form *Trì*, to name the *Tia Ri*, or *Chà Ly*, 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:** > **-aj**

- Dialectal random shift **r-** > **l-**

- Regular shift of **l** (medial or initial) from OC to MC [Baxter 1992]:

1. Medial in tense sesquyllable : OC **-l-** > MC **d-**

2. Initial in lax monosyllable : OC **l-** > MC **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 **ɬai**, or Lai, is the autonym of a Kadai population (Tai-Kadai family), located in Hǎinán (Guǎngdōng). According to Matisoff [1988: 289, note 3], who reconstructed Proto-Hlai initials and tones, **ɬ-** would be a voiceless lateral spirant which determines high series tones. Thus, this initial must originate in a cluster of *plosive+1*. Etymologically, the form **ɬai** perfectly results from **kri:** by a succession of typologically well-known changes: **kri:** > **k<sup>h</sup>li:** > **ɬi:** > **ɬaj**. These shifts needn't to be explained by a remodelling through the Chinese language.

2.2 Li (*lǐ* 隸, *lí* 黎) is the name given to the Hlai by Chinese. This ethnic naming appears in the Hàn dynastic records [HCTE, VI §2] and also during the Sòng period.

<i>lǐ</i> 隸 <MC liX <OC *C-rji? [ri?]
<i>lí</i> 黎 <MC lej <OC *C-rj [C.ri]

2.3 Tai/Thai (or Tay/Thay), from Proto-Thai **ɗaj<sup>A</sup>** (<**\*ɗaj**), 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 **kǎli:** > **kǎdi:/kǎɗaj** (-l- > -d- in tense sesquisyllable) > **di:/ɗaj** (loss of first syllable) > **ɗaj<sup>A</sup>** (Proto-Thai) which evolved to **t<sup>h</sup>aj<sup>A2</sup>** in Thai/ Siamese and in Lao and to **taj<sup>A2</sup>** (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*/ Sukhothay (สุโขทัย), the name of the first Thai Kingdom. The spelling emphasizes this prestigious etymology by writing ไทย (transliterated *ai-d-y*) to designate the Thai/ Siamese people, while the form ไท (transliterated *ai-d*) is occasionally used to refer to Tai/ Thai speaking ethnic groups. Lao writes ໄທ (transliterated *ai-d*) in both cases.

2.4 Yi, spelled *yí* 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.

<i>yí</i> 夷 <MC yij [ji] <OC *ljij [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.ra:w** family

From the etymon **\*k.ra:w** originate the following ethnonyms:

3.1 Gelao : autonym of a Kadai population scattered from Guìzhōu (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 (*lǎo* 獠), supposed to be the ancestors of Lao and some other Tai-Kadai populations, settled in the upper Tonkin and in parts of Yúnnán and Guìzhōu during the Táng times [Pelliot 1904, p.136].

lǎo 獠 < MC lawX < OC \*C-raw? [C.raw<sup>2</sup>]

This reconstruction of the pronunciation for the phonogram 獠 confirms that ‘Lao’ originates in the etymon \***k.ra:w**.

3.3 Sino-Vietnamese *Giao*, in Kiao-tche/ *Giao Chi* (*jiāo zhǐ* 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/ Kæw **kɛ:w**<sup>A1</sup> (ᨗᩢ᩠ᨦ) : 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 *Văn-Lang* (*wén láng* 文郎), name of a province and its population south of the territory of the *Giao Chi*, also North Vietnam (Táng 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-ming/ *Dào míng* 道明 of the Táng times, also located in the north of Cambodia [Hoshino 1986, p.32].

*míng* 明 < MC mjæŋ < 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 *Maleñ/ Malyāñ*: these terms, attested in the Old Khmer epigraphy, point to a country (and its inhabitants) situated south to the Great Lake. The Pre-Angkorian *Maleñ*, reconstructed as \***mǎliəŋ**, is attested in the 7th century ; the Angkorian *malyāñ* \***mǎliəŋ** had been attested from the 9th to the 11th century [Jenner 2009a, 2009b]. The *Maleñ/ Malyāñ* 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 *Maleñ/ Malyāñ* and the Maleng in Laos, but we cannot draw any conclusion from this comparison.

4.5 *Mranmā*: the Burmese epigraphy attest two variants for the ancient name of Burma, *mranmā* (since 1190) which have remained until today, and *mranmā* (1342) [Luce 1959, p.53]. The base of the latter, *mran-* (in which *m̄* has the same value as *n̄*), is transcribed with the phonograms *mù-làng* 木浪 [Luce 1958, p.106].

*làng* 浪 < MC laŋ < OC \*C-rang [\*C.raŋ]

We might think that the validity of the late form *mram(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 \*ʔuraŋ “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 \*ʔuraŋ 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

- BEFEO*: Bulletin de l'École Française d'Extrême-Orient.
- Baxter, William H. 1992. *A Handbook of Old Chinese Phonology*. Berlin & New York: Mouton de Gruyter.
- Cœdès, George. 1948/1989. *Les états hindouisés d'Indochine et d'Indonésie*. Paris: de Boccard.
- Diffloth, Gérard. 1980. The Wa Languages. *Linguistics of the Tibeto-Burman Area* 5(2).
- Dyen, Isidore & Curtis McFarland. 1979. Proto-Austronesian etyma constituting an Austronesian cognate Finder list. Yale University. Mimeographed.
- Edmondson, Jerold A. 1995. English-Bolyu Glossary. *Mon-Khmer Studies* 24: 133-159.
- Ferlus, Michel. 2006. Sur l'origine de quelques ethnonyms: tai/thai, li/hlai, yi, gelao, lao, ... (Chine du sud, Asie du Sud-Est). *Vingtièmes Journées de Linguistique Asie Orientale*, 22-24 juin 2006. Paris: CRLAO (EHESS-CNRS).
- Ferlus, Michel. 2008. Etymology of \***wat/ yuè** 越 "people, principality" (as in *Bǎiyuè* 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.
- Haudricourt, André G. 1984. La tonologie des langues de Hainan. *Bulletin de la Société de Linguistique de Paris* 79(1): 385-394.
- Hoshino, Tatsuo. 1986. *Pour une histoire médiévale du Moyen Mékong*. Bangkok: Éditions Duang Kamol.

- [HCTE] *Hlai (Li) Chinese Thai English*. 2003. A Collaborative Research Project between : Institute of Language and Culture for Rural Development (Mahidol University at Salaya, Thailand) and Kam-Tai Institute (Central University for Nationalities, People's Republic of China).
- Jenner, Phillip N. 2009a. *A dictionary of pre-Angkorian Khmer*. Doug Cooper ed., Pacific Linguistics 597. Australian National University.
- Jenner, Phillip N. 2009b. *A dictionary of Angkorian Khmer*. Doug Cooper ed., Pacific Linguistics 598. Australian National University.
- Langages and Cultures of the Kam-Tai (Zhuang-Dong) Group : A Word List*. 1996. A Collaborative Research Project between : Institute of Language and Culture for Rural Development (Mahidol University at Salaya, Thailand) and Kam-Tai Institute (Central University for Nationalities, People's Republic of China).
- Lê Thành Khôi. 1981. *Histoire du Viêt Nam*. Paris: Sudestasië.
- Luce, Gordon H. 1958. The Early *syām* in Burma's History, *Journal of the Siam Society* 46(2): 123-214.
- Luce, Gordon H. 1959. Note on the Peoples of Burma in the 12th-13th Century AD. *JBR* 42(1):52-74.
- Luce, Gordon H (translated by). 1961. *Man Shu (Book of the Southern Barbarian)*. Data paper n° 44, Southeast Asia Program. Cornell University.
- Luce, Gordon H. 1976. Sources of Early Burma History. *Southeast Asian History and Historiography, Essays presented to D.G.E. Hall*. C.D. Cowan & O.W. Wolters ed. Ithaca & London: Cornell University Press.
- Madrolle, Cl. 1937. Le Tonkin ancien. *BEFEO* 37(2): 263-332.
- Maspero, Henri. 1918. Etudes d'histoire d'Annam, IV Le Royaume de Văn-Lang. *BEFEO* 18(3-1).
- Matisoff, James A. 1988. Proto-Hlai Initials and Tones: A first approximation. *Comparative Kadai: Linguistic Studies beyond Thai*. Edmondson & Solnit ed., pp. 289-322.
- Pain, Frédéric. 2008. An Introduction to Thai Ethnonymy: Examples from Shan and Northern Thai. *Journal of the American Oriental Society* 128(4): 641-661.
- Pelliot, Paul. 1904. Deux itinéraires de la Chine en Inde à la fin du VIIIe siècle. *BEFEO* 4: 131-413.
- Pulleyblank, Edwin G. 1991. *Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: UBC Press.
- Taylor, Keith Weller. 1976. *The Birth of Vietnam*. Berkeley, Los Angeles & Oxford: University of California Press.
- Wurm, S.A. & B. Wilson. 1975. *English Finderlist of Reconstructions in Austronesian Languages*. Pacific linguistics series C, n° 33. The Australian National University.