The geographical distribution of autonyms for 'person/people' in the Mon-Khmer languages of South East Asia and its contribution to the understanding of the original location (homeland) of the Luòyuè residents

TRẦN, Trí Dõi

(Faculty of Linguistics, USSH Vietnam National University Hanoi)

Abstract: The Sino-Vietnamese phonetic cluster Lạc Việt (雒 越) first appeared in the Hanshu (漢書, Hàn shū), a first-century CE Chinese work. At that time, this cluster was used to refer to a group of non-Sinitic residents in the Bách Việt (百越, Băi yuè) area, but there were no specific indications about the exact location of this group or the owners of this name. From the perspective of geolinguistics, this paper investigates autonyms that refer to 'person/people' in the Mon-Khmer languages of Southeast Asia. The aim of this paper is to prove that Lạc Việt, used by the Chinese in the 1st century, was used to record the autonym of the Mon-Khmer residents of Southeast Asia.

Key words: homeland, Lạc Việt (Luòyuè), Mon-Khmer, Hànshū, autonyms for 'person/people,' geolinguistic.

1. Introduction

When discussing the etymology of the name Lạc Việt (雒 越, Luòyuè) recorded in the Hanshu (漢書, Hàn shū), a Chinese book from the first century CE, Ferlus (2011) supposed that at that time, the Chinese people used these characters to record the cluster [*p.rak *wat]. The syllable [*p.rak] is the phonetic form of the Mon-Khmer autonym which etymologically means 'person/people.' His etymological explanation of the components of the Sino-Vietnamese phonetic cluster Lạc Việt was based on an analysis of its relations to other languages in southern China and Southeast Asia. He also believed that Southeast Asia cannot be constrained within today's administrative limits, and the geo-cultural region of Lạc Việt, including the Vietnamese language and the culture of the Viet (Nam) people, as recorded in ancient Chinese history books, must encompass the south of China, i.e., 'south of the Yangtze River' and present Southeast

Asia. The author's works, published in 2017 and 2022, were not only in strong agreement with Ferlus's explanation but developed his idea by offering an additional discussion on the concept of Lạc Việt (雒越) in prehistory. As pointed out in these two studies, several Austroasiatic languages in Southeast Asia still retain the etymological meaning of that ancient name, deriving the ethnonym from when the Chinese used the character Lạc (Luò 雒) to phonetically record that non-Chinese name (Trần Trí Dõi, 2017, 2022). Recently, several anthropological studies (Đinh Hồng Hải, 2018; Dinh & Kelley, 2021) have also considered this etymological interpretation of the name Lạc Việt proposed by linguists.

Following previously published studies, the author presented another study (Trần Trí Dõi, 2023) on what names are currently used to refer to 'person/people' in some languages of Southeast Asia and the southern part of China. There are various different phonetic forms in the Mon-Khmer languages themselves and other forms in the languages of the Tai group, the Tai-Kadai family. However, in the case of the Tai languages, there is a distinction between the phonetic forms borrowed from the Chinese language (or the Sinotic group) when the Chinese language recorded the Mon-Khmer self-names, and on the other hand, the phonetic forms that the Tai languages themselves borrowed directly from the Mon-Khmer languages when the Tai people migrated southward to Southeast Asia.

From the perspective of geolinguistics, this paper examines the geographical distribution of those autonyms for 'person/people' in the Mon-Khmer and Tai languages. An interpretation of geographical maps will be to provide further discussion on the 'homeland' or 'cradle' of the people referred to by the phonetic cluster [*p.rak *wat] 雒越 (Luòyuè) by the Chinese in the 1st century CE. The findings are expected to reveal that the real owners of that cluster are probably inhabitants of the Mon-Khmer linguistic region.

2. Geographic distribution of autonyms for 'person/people' in the languages of Southeast Asia

2.1. In the Mon-Khmer languages

As stated in the author's latest research (2023), at present, the autonyms for 'person/people' in the Mon-Khmer languages can be divided into three different types based on the structure of the second syllable in the disyllabic or sesquisyllabic words.

The phonetic forms corresponding to the reconstructed form *p.rak can be found in the Palaung-Waic, Khmuic, Vietic, Katuic, Bahnaric, and maybe Khmeric groups.[1]

In the Palaung-Waic group, which is located in the northernmost part of the Mon-Khmer linguistic area (currently southern China), as Ferlus (2011) indicated, the existing phonetic forms of the autonyms are *pəʐaək*, *parauk*, and *pʰalok,* depending on the dialects (Ferlus, 2011, p. 7). He also suggested that the phonetic form in the Khmuic group is *rɔːk.* However, as indicated by Nguyễn Văn Lợi (2009, p. 393), the Mang language, a member of the Khmuic group, currently uses the form [*ha³*]. Thus, in these two language groups from the northern part of the Mon-Khmer linguistic area, the existing forms are either a disyllabic form with a closed main syllable (*parauk*) or a monosyllabic form, which is closed or open (*rɔːk, ha³*).

In terms of the Vietic group, the author's fieldwork data show that the form used in the Arem language is [cmrawˀ], a disyllabic word in which the main syllable is open (Trần Trí Dõi, 2013, p. 121). Mã Liềng, another Vietic language, uses either [ɓrʊˀ] or [kari] depending on the dialects (Ferlus, 1995, p. 55) and also [kri] (Chamberlain, 2018). Those are all sesquisyllabic words in which the main syllables are all open. In the Katuic group, the form used in the Bru-Van Kieu language is [bru].

Meanwhile, in the Bahnaric languages, the Mạ ethnic group has another name, *châu Mạ* (Đặng Nghiêm Vạn, 2003, p. 223), which is most likely a syllabification of a phonetic cluster as in the case of the Arem language [cmrawˀ]: chau < [*cmrawˀ]. Sidwell and Jacq (2003, p.65) studied the Western Bahnaric subgroup and reconstructed the Proto-West Bahnaric word for 'person' as [*raa].

In the Khmer language, the form [mənuh] for 'person/people' is likely to be a Sanskrit loanword, but the form [proh], which means 'man, male human' could be an Austroasiatic-originated word.[2] Similar forms are used as autonyms for 'person/people' in six groups in the Khmer branch of the Austroasiatic family, and the phonetic correspondence between them and the aforementioned form [*p.rak] can be considered as strong proof that they originated from the same source. Specifically, the first consonant of monosyllabic autonyms or the first consonant of the main syllable in disyllabic autonyms must always be either [r] or [h, l]. The distribution of autonyms

---

[1] I has not had the opportunity to process data in other groups of the Mon-Khmer branch. However, the data of those six of nine groups of this branch can be considered as strong proof to the Mon - Khmer origin of these autonyms.

[2] I must thank Dr Frédéric Pain at the Laboratoire Langues et Civilisations à Tradition Orale (LACITO, Paris) for providing me with this data when we carried out a Khmer language fieldwork study in Southern Vietnam.
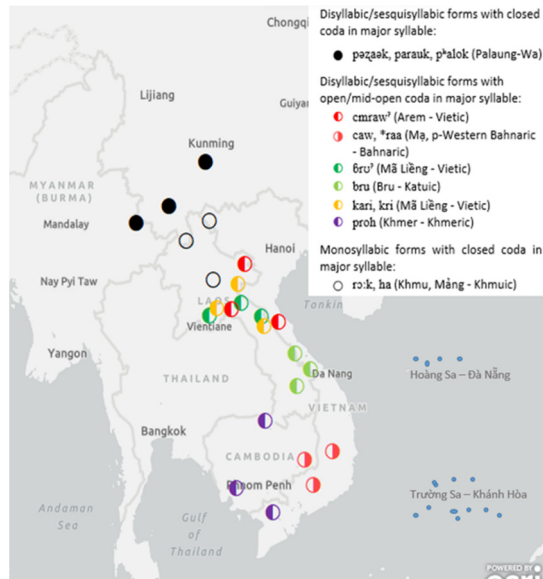
for 'person/people' in the six mentioned Mon-Khmer languages can be seen on the map below.
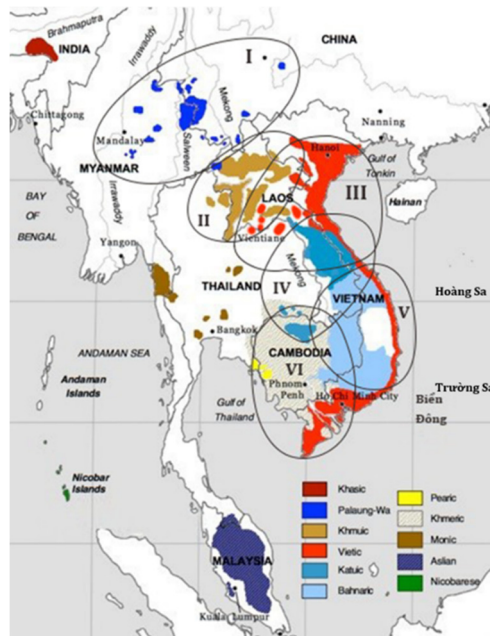
Map 1: Current location of six language groups in the Mon-Khmer branch
(Source: Gutman & Avanzati, 2013)

Using Gutman and Avanzati's (2013) Austroasiatic language family map as a reference, the author has circled the six above-mentioned language groups and shown the expansion of the Vietic inhabitants to the west: (I) Palaung-Wa (Palaung-Waic); (II) Khmu (Khmuic); (III) Viet language group (Vietic); (IV) Katu (Katuic); (V) Bahnar (Bahnaric); (IV) Khmer (Khmeric) languages.

The next map shows the geographical distribution of the phonetic forms of autonyms for 'person/people' in each group. In this map as well as in the following maps, each form in each language group will be displayed in at least three representative locations.

Map 2: Geographical distribution of autonyms for 'person/people' among six language groups in the Mon-Khmer branch

2.2. Borrowed forms in the Tai and Sinitic groups

The distribution of loanwords for 'person/people' in the Tai group is shown in Map 3 below. Like Map 1, this map is by Gutman and Avanzati (2013); along with the Kam-Sui languages, they have also clearly mapped the boundary between the Southwestern Tai (SWT), the Central Tai, and the Northern Tai (NT) subgroups of the Tai group of the Tai-Kadai family. Among them, as Haudricourt (1953) indicated, the Southwestern Tai subgroup is part of the Tai population that migrated southward from the beginning of the first millennium CE until around the 10th century.

Map 3: Tai languages in Southeast Asia and Southern China
(Source: A. Gutman và B. Avanzati 2013)

As shown in the author's 2023 paper, there are now two different phonetic forms of the autonym 'person/people' within the current Tai languages. One is directly borrowed from the Chinese cluster 雒 越, which was used to record the Mon-Khmer cluster [*p.rak *wat]. Most of the Kam-Sui and Tai languages have borrowed forms in which the first consonant of the main syllable is [l]. This can be found in the case of [lak] in the Maonam or Lakkia languages (Ferlus, 2011, p. 7); [lak] in the Zhuang language; or [luk, lu] in the Tày, Nùng, White Tai, Black Tai, Lao, and Thai languages (Trần Trí Dõi 2022, pg.520; Trần Trí Dõi 2023, pg. 6–7). The other, [puok], was also borrowed from the Chinese language, according to Chamberlain[3] (Chamberlain, 2016), but it is only preserved in the Southwestern Tai subgroup. Thus, the two forms of the

---

[3] Chamberlain (2016) wrote that the Ksing Mul people were called *Puak,* but in Vietnam, they can be called *Puok*, so I chose this spelling.

autonym for 'person/people' (lak/lu/lu and puok) that the Tai languages borrowed from the Chinese language are now preserved in two different geographical regions.

In addition, three different forms, *xá*, *khạ,* and *kha,* related to the Austroasiatic autonym for 'person/people,' are currently used in the Southwestern Tai subgroup. As explained by Chamberlain (2016), this is the phonetic form that the Southwestern Tai languages borrowed directly from the Mon-Khmer languages when the Tai speakers migrated from southern China to Southeast Asia during the Proto-Southwestern Tai (PSWT) period. Chamberlain assumed that these two phonetic forms /*sa*/ and /*kha*/ that the inhabitants of the Southwestern Tai subgroup used to name the Mon-Khmer inhabitants are a consequence of the phonological change of the PSWT **khraaA1*: /*sa*/ or /*kha*/ < **khraaA1* PSWT. The form **khraaA1* in PSWT means 'Austroasiatic ethnic,' though it is used in other related languages as the general autonym for 'person.'

The distribution of the forms /puok/, /lak, luk, lu/, and /sa, kha/ can be seen in Map 4 below.
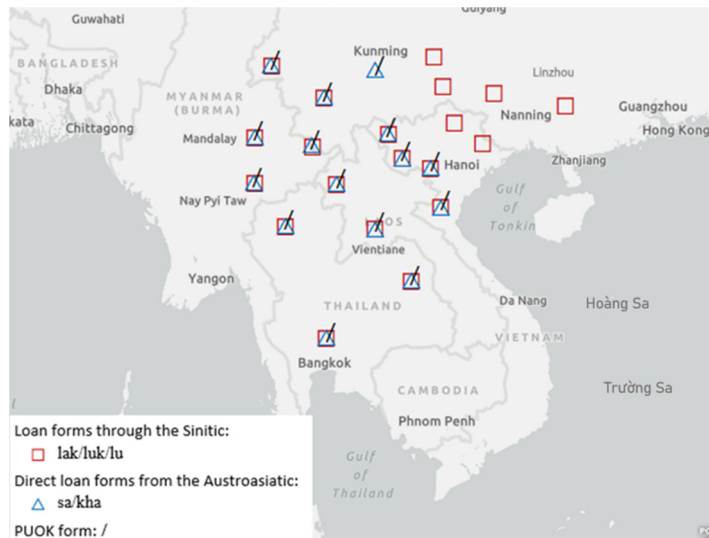
Map 4. Geographical distribution of loan words in the Tai languages

Regarding the Sinitic language, as aforementioned, from the perspective of historical linguistics, the Sino-Vietnamese cluster 雒 越 (Lạc Việt) was used by the Chinese people in the Old Chinese period (OC) and was recorded in the Hanshu (1st century CE) (Trần Trí Dõi 2017; 2022). The character 雒, vocalised in OC as [*p.rak], was an autonym referring to 'persons/people' who spoke the Mon-Khmer language. The form [puok], as Chamberlain (2016) explained, is also very likely the autonym for 'person/people' that the Chinese used to refer to the Austroasiatic inhabitants of the

Dian kingdom, which was pacified by the Chinese later than the 1st century CE (this area is now Yunnan province, China).

Map 5 shows the general geography of the Sino-Tibetan language family, and the geographical distribution of the Sinitic forms borrowed from Austroasiatic languages is shown in Map 6 below.

Map 5: General geography of the Sino-Tibetan language family

Map 6. Distribution of autonyms for 'person/people' in the Sinitic languages

3. Discussion

The distribution of autonyms for 'person/people' in the Mon-Khmer, Tai-Kadai, and Sino-Tibetan languages in Southeast Asia and southern China, as shown in Maps 2, 4, and 6, could contribute greatly to further discussion regarding the homeland and ethnic group of the people who used these autonyms during the Old Chinese period. First, consider the six language groups belonging to the Mon-Khmer branch examined in this paper. Taking into account the simultaneous existence of both monosyllabic and disyllabic or sequisyllabic autonyms in all six linguistic groups examined, it can be argued that [*p.rak], the reconstructed Old Chinese phonetic rendering of 雒, is related to the autonyms for 'person/people' in the Mon-Khmer languages. Additionally, the current distribution of those autonyms in those Mon-Khmer language groups which are distant from southern China (the Palaung-Waic, Khmuic, Vietic, Katuic, Bahnaric, and Khmeric groups) suggests that the autonyms must have originated in the Mon-Khmer languages. Furthermore, this argument seems to be strengthened by considering the geographical distribution of names for 'people' in the Tai-Kadai languages. While the forms [lak, luk, lu] now exist in the Tai-Kadai language family, including both the Kam-Sui group in the north and the Tai group in the south, the form /sa, kha/ is found only to a limited extent in the Southwestern Tai subgroup of the Tai group. In other words, the forms currently used in the Mon-Khmer language groups are essentially similar, while there are distinctions between the Southwestern Tai subgroups and the rest of the Tai-Kadai family. A possible explanation for this might be that those two forms which did not appear in Kam-Sui and two other subgroups of Tai were directly borrowed from Mon-Khmer languages while the form [lak, luk, lu] was a borrowed form of [*p.rak] from the Old Chinese language. These findings are further supported by the geographical distribution of the two forms [*p.rak] in OC and [puok].

These results taken together suggest that the historical linguistic proof does not support the argument of some Chinese researchers, including the authors Liang and Li (2017), that Zhuang-speaking members of the Tai group were the people referred to by the autonym Lạc Việt (雒 越). There are no Mon-Khmer or Tai historical documents identifying which group was called 'lạc việt/luòyuè' during the last centuries BCE and first centuries CE; thus, in this case, linguistic data can be considered as the most reliable basis for making the identification. Historical linguistics, as well as geographical distribution, provides evidence to support the notion that Lạc Việt is an autonym for Mon-Khmer speakers of the Austroasiatic language family, not members of the Tai language group.

References

Chamberlain, James R. (2016) Kra-Dai and proto-history of South China and Vietnam. *Journal of the Siam Society* 104: 27–76.

Chamberlain, James R. (2018) A Kri-Mol (Vietic) bestiary: Prolegomena to the study of ethnozoology in the Northern Annamites. *Kyoto Working Papers on Area Studies*, 133: 1–171. URI: http://hdl.handle.net/2433/231260

Đặng, Nghiêm Vạn (2003) *Cộng đồng quốc gia dân tộc Việt Nam* [*The Vietnamese national community*]. Ho Chi Minh City: Publishing House of Vietnam National University.

Đinh, Hồng Hải (2018) Lạc Việt: Từ tộc danh đến căn tính tộc người [Lạc Việt: From the ethnic name to ethnic identity]. *Journal of Social Sciences and Humanities* 6: 710–726.

Dinh, Hong Hai and Kelley, Liam C. (2021) Competing Imagined Ancestries: The Lạc Việt, the Vietnamese, and the Zhuang. In J. Gillen, L. C. Kelley, & L. H. Phan (Eds.), *Vietnam at the Vanguard: New Perspectives Across Time, Space, and Community* (pp. 89–107). Singapore: Springer.

Ferlus, Michel (1997) Le maleng brô et le vietnamien. *Mon-Khmer Studies* 27: 55–66.

Ferlus, Michel (2011, June 30–July 1) *Les Băiyuè, ou les "pays des (horticulteurs/mangeurs de) tubercules"* [*The Băiyuè, or the "country of (horticulturists/eaters of) tubers"*] [Conference presentation]. 24th Day of Linguistics of East Asia, East Asian Linguistic Research Center (EHESS CNRS), Paris, France.

Ferlus, Michel (2014) Arem, a Vietic language. *Mon-Khmer Studies* 43.1: 1–15 (ICAAL5 special issue).

Gutman, Alejandro, and Avanzati, Beatriz (2013) *Austroasiatic languages*. The Language Gulper. URI: http://www.languagesgulper.com/eng /Austroasiatic.html

Gutman, Alejandro, and Avanzati, Beatriz (2013) *Tai-Kadai languages*. The Language Gulper. URI: http://www.languagesgulper.com/eng/Taikadai.html

Gutman, Alejandro, and Avanzati, Beatriz (2013) *Tibeto-Burman languages*. The Language Gulper. URI: http://www.languagesgulper.com/eng/Tibeto.html

Haudricourt, André-Georges (1953) La place du vietnamien dans les langues austroasiatiques [The place of Vietnamese in the Austroasiatic languages]. *Bulletin de la Société de Linguistique de Paris* 1: 122–128.

Liang, Tingwang and Li, Sheng (Eds.) [梁庭望、厉声（主编）] (2017) 骆越方国研究（上下册）[A study of Luoyue Fangguo, Vol 1 & 2]. Beijing: The Ethnic Publishing House].

Nguyễn, Văn Lợi, Nguyễn, Hữu Hoành, Tạ, Văn Thông (2008) *Tiếng Mảng* [*The Mang language*]. Hanoi: Social Science Publishing House.

Sidwell, Paul, and Jacq, Pascale (2003) *A handbook of comparative Bahnaric*, *Vol.1: West Bahnaric*. Canberra: Pacific Linguistics.

Trần, Trí Dõi (2017) Trao đổi thêm về từ nguyên của yếu tố "lạc (luò 雒/駱)" trong tổ hợp "lạc việt (luòyuè 雒越)" [A further discussion on the etymology of "lạc (luò 雒/駱)" in the cluster "lạc việt (luòyuè 雒越)"]. *Journal of Museum and Anthropology* 18:41–53.

Trần, Trí Dõi (2022) *Lịch sử ngôn ngữ người Việt góp phần tìm hiểu văn hóa Việt Nam* [*The history of Vietnamese language in contributing to understanding Vietnamese culture*]. Hanoi: Vietnam National University Publishing House].

Trần, Trí Dõi (2023) Những dạng thức lưu giữ tên tự gọi (autonym) chỉ "người (person/people)" của những ngôn ngữ Môn-Khmer ở Đông Nam Á [Retainers of autonyms referring to people in the Mon-Khmer languages of Southeast Asia]. *Journal of Language and Life* 335: 5–13.