Documentation and phonological description of Malieng, Vietnam, with focus on tone.

Upgrade Chapter

In submission for PhD candidateship on 15/05/2019

Department of Linguistics, School of Languages, Cultures and Linguistics

SOAS University of London

MPhil Candidate: Albert Badosa Roldós

First Supervisor: Dr Nathan Hill

Second Supervisor: Dr Chris Lucas

Table of contents

1. Aims

The aim of this PhD is to document and describe the Maleng language (ISO 639-3: *pkt* or *bgl*) also called Pakatan or Bo) which is spoken in Vietnam and Laos by only around 500 speakers. The project will focus on Malieng, the variety of the Pakatan language spoken in Hoanh Son, Vietnam. This variety has 200 remaining speakers living in half a dozen villages. The other varieties are spoken in Laos. The project envisions a large-scale linguistic and cultural documentation of the language community, including elicitation, narratives and traditional songs, among other genres. The description of the language will focus on phonology and more particularly on the tonal system, with notes on historical linguistics and tonogenesis (how the Vietic languages developed tonal features). The new gathered data will contribute to the widely developed theory about the origin of tones in Vietnamese and Asian one in general–tonogenesis–as deriving from ancient initial consonants. The contribution to this theory will be the main research question. There is almost no data on the Pakatan language, only a few elicited word-lists. The other Vietic languages (except for Vietnamese) are only barely documented, mainly by a Vietnamese and French scholars, with whom I am in contact.

All Vietic languages–except for Vietnamese–, together with other minority languages in the area, are endangered and facing extinction. The wide documentation of a Vietic language, Maleng, will contribute to Austroasiatic and Vietnamese linguistics (phonology, tones and historical linguistics in particular) and culture studies, minimising the scarce research carried out on Vietic languages. At the same time, the project also covers the need of addressing language loss in general and particularly in South-East Asia, as it may be of good use for future language support and revitalisation.

2. Vietic languages: classification and state of the art

2.1 Minorities and ethnic groups in Vietnam

2.1.1 The concepts of minority and ethnic group

There has been a long and intense controversy on the term "minority" and its meaning in (mainly Western) literature. Although there seems to be no valid or universally accepted general definition of the term *minority*, a commonly accepted point is that the main characteristic of a minority is the differentiation between dominance and subordination, and not primarily a numerical question of population size (Engelbert, 2016: 10). Engelbert identifies quantity and power–supremacy versus inferiority–as the two main characteristics defining majorities and minorities.

There are two other definitions that I consider relevant. Firstly, Markefka remarks that the social recognition of differences within a social system depends on ideologies and political constellations (Markefka, 1990: 12-15, after Engelbert, 2016: 10). Secondly, Ottersbach defines minorities with 3 key points: (1) there are 4 different types of minorities: quantitative or numerical, political or electoral, cultural, economic and social or discriminated as well as protected minorities; (2) there are 2 theoretical approaches to the definition of the minority: the analytical approach–minorities are constructed social orders–and the realistic definition–minorities create themselves by their own will; (3) he finally defines minorities as results of social processes, through the preservation or protection as well as the discrimination of differences (Ottersbach, 1997: 235, after Engelbert, 2016: 10).

With regard to the issue of the number of speakers/members of a minority or ethnic group, a new term has recently been coined applied to European minorities: the concept of minorisation instead of minority. Whereas 'minority' puts emphasis on the number of speakers of the language, which is normally not directly related to its degree of endangerment, the term 'minoritised' emphasises the power imbalance in relation to the dominant languages (Spolsky and Hult, 2010: 266). Then, there are languages that are spoken by large majorities but are minoritised, like Kurdish, for instance. Henceforth, I will use the term minoritised–coined by Aracil (1983)–along with minority.

Pascal Bordeaux remarks put an emphasis on the minorisation–discrimination by national or supranational powers–condition when defining minorities:

> "the concept of minority is normally used to qualify very diverse social and political phenomena–for a social group to emancipate themselves, to protest against something, to make statements based on the claim to a legitimate right to be different, to the right to individual freedom and even to citizens' equality." (Bordeaux, 2016: 9).

In the Western framework, the old Western division which considers 'indigenous or autochthonous minorities'–those groups found in a particular territory before the foundation of the nation-state–and 'diaspora or immigration minorities'–those established after the foundation of the nation-state is nowadays not acceptable (Windischer, 2009: 40-88, after Engelbert, 2016: 25). This distinction was already qualified as old by scholars such as Kymlicka, who proposes the distinction 'multinational' countries–long-established nations sharing a state due to colonisation, conquest or confederation–and 'poliethnic' countries–those with individual and familial immigration–or a combination of both (Kymlicka, 1995: 15). Moreover, Kymlicka qualifies the Old World-New World division as a dangerous oversimplification (ibid.: 20).

Keating, on the other hand, criticises the distinction between immigrant ethnic groups and national ethnic groups that authors like Windischer and Kymlicka make. He claims these differences are extremely relative depending on what the cut-off date is and who is doing the defining (Keating, 2001: 42). The biggest bias of the previous authors, however, is their West-centralised vision. Kymlicka himself only talks about "liberal democracies", as if multiculturalism and its management was only a problem of the West, but even in this narrowed-down context he avoids talking about, for instance, Asian liberal democracies such as Japan.

Nonetheless, Keating (2001: 39) criticises authors with the point of view that, in words of Kymlicka, "immigration is, in most cases, voluntary" (Kymlicka, 1995). Engelbert agrees with Keating: "[…] the basic idea of the term *minority* already implies the notion of an unequal relationship between social groups, and of dependence, subordination and potential discrimination." (Engelbert, 2016: 16). Thus, it is difficult to posit the existence of "free will" in a context of discrimination and repression, or, in other words, in a context of coercion (Keating, 2001: 39). Free will is also a relevant concept then discussing the–different–identity of a particular group, and at the same time 'identity' itself is also a controversial concept in the field (Engelbert, 2016: 11).

Outside of the West, a big question arises: is a minority the same as an indigenous group? Should they be treated the same way? The UN has a clear position on this issue, stating that indigenous populations should be treated separately from minorities (Windischer, 2009: 53, in Engelbert, 2016: 15). Again, however, we come back to the same point: who defines what an indigenous group or a minority is?

Engelbert claims to have found an answer to the problem:

> "if the existence of a minority is not regarded as a static process, but dependent primarily on its own will to form a social group distinct from the majority, then the manifested will to be treated as a distinct group would make a distinction between indigenous peoples, minorities and immigrants irrelevant." (Engelbert, 2016: 15-16).

Nonetheless, Vietnam seems to have found the answer in the state: Vietnamese researchers tend to affirm that qualifying a community as a minority corresponds to the state:

> "the purely numerical relationship of a group towards a majority alone does not constitute a minority. It can only be described as such if the state determines it as a minority, hence changing it from a numerical to distinctive social group" (Trần Thị Liên, 2016: 9).

The German author challenges the previous assumption:

> "Is the state responsible for 'creating' minorities or do these groups have a choice of their own or a word to say in this process? […] In societies like Vietnam, for instance that have not experienced such claims, […] this concept of a minority refers first of all to the difference between majority and minorities within one people." (Engelbert, 2016: 9).

With the idea that the terms 'ethnic minority' and 'indigenous community' have been and will be controversial for a long time, scholars seem to leave the discussion when addressing Vietnamese minorities, accepting the official will. However, as we will see, there are authors (mainly from the anthropological field) who write about the consequences of the official policies towards minorities.

2.1.2 Ethnic minorities in Vietnam

Vietnam concentrates great cultural diversity. There are 54 officially-declared 'ethnic minorities'–*dân tộc thiểu số*–, making up about 14% of the country's population. As a country, Vietnam is proud of this diversity in a folkloric way, showcasing it in museums, festivals or promoting it in the tourism sector. It is considered a positive value that contributes to the country, making Vietnam a multi-ethnic state (Taylor, 2008: 3).

The official discourse of the state stresses the contribution of the ethnic minorities to the Vietnamese national project, as a part of the Vietnamese nation. Their economic contribution is the main concern of the Vietnamese authorities, which implies heavy transformations of their traditional life. Internal colonisation is seen as "Viet/Kinh helping the ethnic minorities survive in the harsh conditions of the highlands". Every concession to ethnic minorities is celebrated as a sign of kindness towards all the Vietnamese, regardless of their origin. This official view is well-reflected in the book "Ethnic Minorities in Vietnam", related to the Vietnam Museum of Ethnology (Đặng Nghiêm Vạn et al., 2000: 1-11).

Behind the scenes, however, the government is promoting policies of assimilation, capitalist economic development and homogenisation that have a great impact on the ethnic minorities. The concept of 'national unity' is highly valued, and this concept does not allow for opposition to these policies. Demonstrations, protests and ethnic/religious movements are considered to foster disunity by the government and prosecuted. Along with being considered a national treasure, diversity is sometimes also conceived as an obstacle to development, as an obsolete relic of the past or a security problem for the nation's interests (*ibid*.: 1-11).

Philip Taylor therefore explains that the state has a paternalistic view of the minorities, who are denied a voice and agency:

> "Minorities have so far too much been described in their relations with the majority, or rather, as a dependent part of the Vietnamese national project. This has satisfied stereotypes, as if minorities have always only been subjugated, disciplined or circumscribed. They have too often been described as governmentalised subjects of social engineering. The situation of the highland peoples, in particular, has been portrayed in apocalyptic images. The settlement of lowlanders in the highlands has been narrated in strong language, using terms like 'internal colonialism' or 'Vietnamisation'." (Taylor, 2008: 7-13).

This view incarcerates the minorities within the trap of rigid framework of the nation-state, ignoring their transnational, multi-ethnic and multilinguistic socioeconomic and cultural practices which were built up long before the creation of the nation-states in South-East Asia (ibid.: 18-20). In addition, we must not forget that most Vietic groups are transnational, with a number of them living in the borders between Laos and Vietnam, and some in the Thai border. These policies of discrimination and assimilation have existed for many years and are not in any way a consequence of contact with the West, although they have been enhanced by the Western colonial forces (Engelbert, 2016: 17-18).

The policies of assimilation regarding minorities and the negligence of indigenous and minority rights are shared by the communist and pro-communist authorities and the anti-communist groups, in the same way that they were shared by the Soviet Union and the United States, for instance. The Vietnamese assimilatory practices and strategies resemble those in China: for instance, resettling majority Viet/Kinh population in Vietnam and Han population in China in highland/distant areas so they become a majority (ibid.: 21).

Since independence from France (1953/54), former Indochina–Vietnam, Cambodia and Laos–has attempted to bring the Central Indochinese highlands under the firm control of the lowland governing elites–Viet/Kinh, Lao and Khmer–. The strategies at assimilation and seizure of control include road building, economic development, resettlement with people from the lowland majorities and division of the highlands into different provinces (Engelbert, 2016: 76; Schliesinger, 1997: 35-38).

All Vietic groups except for the Vietnamese (Viet or Kinh) and the Mường fit into the category of highlanders. In fact, the majority of the officially recognised 54 ethnic groups in Vietnam are highlanders, the others being lowlanders–only the Viet/Kinh, Khmer, Cham, Hoa and Ngai (Schliesinger, 1997: 23). The Vietic highland groups are located in the Central-North Vietnamese regions and also the Central Indochinese highland area and are therefore affected by these policies–which were already laid out by the French colonial authorities. It is of utmost importance to consider this aggressive context of internal colonisation, which is accompanied by a denigration of the highlanders for the colonisation to be justified: they are said to have savage, uncivilised or outright reactionary customs and traditions (Engelbert, 2016: 21; 76).

In fact, some Vietic groups preserve traditional ways of life today considered ‘old’ and ‘primitive’. Chamberlain (1998: 109) has established a cultural typology of the Vietic groups

according to their way of life–which Chamberlain himself advises not to be construed as evolutionary in nature:

Vietic group	**Lifestyle**
Atel, Thémarou, Mlengbrou and possibly Cheut	Small-group foraging nomads
Arao, Maleng, Malang, Makang, Tơe, Ahoe, Phóng	Originally collectors and traders who have become emergent swidden sedentists
Kri	Swidden cultivators who move every 2–3 years among pre-existing village sites
Ahao, Ahlao, Liha, Phong (Cham), Toum	Combined swidden and paddy sedentists

Table 1: Cultural typology of Vietic-speaking ethnic groups

This characterisation of the ethnic minorities is linked to the universal phenomenon of colonisation, but also to the historical Vietnamese self-perception that continues nowadays of their culture and state organisation as superior in comparison to those of their neighbours (Engelbert, 2016: 18-19). In fact, this seems to be a common phenomenon in South-East Asia, as these views are also shared among minority groups towards other minority groups. Maleng, the language I am going to document, is spoken in Laos and Vietnam. Two of its varieties, Pakatan and Maleng, spoken in Laos, coexist in the same mountain range. The Pakatan variety refers to the Pakatan speakers living in the village of Pakatan, with a more sedentary life. Pakatan villagers consider their Maleng neighbours, who live in the mountains, as ‘less developed’. In fact, Maleng stands for ‘to be human’, just like several other denominations of indigenous groups in South-East Asia–Khamou or Kesing– (Ferlus, 2016: 8-13).

Research on the Indochinese Central Highlands has therefore its own particularities in this (internal) colonisation context. As the Vietic group that I will work with is established in the Vietnamese highlands, I will address the aforementioned research particularities in the following sections.

2.1.3 The research context in Vietnam

In the past decades, studies show that the single largest cause of ethnic conflict in the world today is the struggle of indigenous peoples for the protection of their land rights (Gurr, 1993; Nietschmann, 1987). These conflicts arise in all parts of the world and Vietnam is not an exception.

In Vietnam there is a great lack of research on ethnic minorities: "the minorities of Vietnam are among the least studied peoples on earth" (Taylor, 2008: 27). In order to evaluate the research and fieldwork contexts in Indochina, Vietnam and the Central highlands, anthropology is perhaps the most suitable field to look at, since it has produced wider and more in-depth work than linguistics. Both fields seem to share the same problems and face the same contextual specificities.

One of the factors that explains the paucity of research on these peoples is that research conditions in Vietnam are not very favourable to both Vietnamese and international researchers, although they have improved compared to past decades. Taylor, for instance, reports continuing problems of access to field sites as well as political constraints on reporting the results of research. He also warns that research is often carried out according to the rationales of state policy, instead of drawing attention to the ethnic minorities themselves, despite the minorities being the object of the research. This gives the idea of "minorities as governmentalised subjects of social engineering, applied anthropology, and official classification." There is also a lack of research on the impact of state policies on the ethnic minorities. These focuses magnify the power and voice of the state and ignores the interests of the ethnic minorities, who are the object of study at the same time (ibid.: 5-6; 25-27).

Taylor questions the nation-state as the desirable framework to do research on ethnic minorities, not only because they tend to have transnational and multilingual practices and life-styles, but also because it denies the voice of the minorities themselves (negatively affecting the research), and because it is a modern and recent framework that prevents the research from having the historical perspective of the time where nation-states did not exist (ibid.: 6; 20-22). As the author pictures it:

> "When minority groups sing and dance, they dramatize the geopolitical imaginary of the multi-ethnic nation. Here, the nation-state looms even larger as the very condition for existence of minority ethnicities, providing the language, categories, and contexts in which their identity can be imagined." (Taylor, 2008: 16-17)

It is true, on the other side, that the nation-state is radically transforming and affecting the life-styles, traditions and socioeconomic practices of the ethnic minorities at a fast rhythm. This includes the exposure of these groups to the wider world and its rapidly changing dynamics, which imply the contact with newcomers and the imposition of external standards. The Vietnamese policies of assimilation and economic development are the agents of this new, broader international context, which endangers the traditional ways of life, the languages and the cultures of the ethnic minorities (ibid.: 7; 16). Many scholars, thus, consider that the consolidation of the nation-state conflicts with the maintenance of such a great diversity; minorities are likely to perish in the nation-state (ibid.: 15-16).

2.1.4 Theesearch context in the Vietnamese Central Highlands

We now move to the Central Highlands region, which until the XIXth century was the almost exclusive habitat of various tribal peoples, linguistically of Austroasiatic and Austronesian origin. They–and by extension all highlanders in South-East Asia–were generally called ‘savages’ by the lowland populations, which translates as ‘Mọi’ in Vietnamese, ‘Kha’ in Lao or ‘Phnong’ in Khmer. These terms have gradually been replaced by others such as ‘mountain people, ‘highlander’, ‘lowlander’, etc. in the Khmer, Lao and Vietnamese languages, in an effort to include these people in the national projects and not to use derogatory terms (Engelbert, 2016: 77; Schliesinger, 1997: 23).

The contact with French missionaries and explorers was one of the great changes the highlanders experienced, integrating them irreversibly into colonial, imperial and nation-state dynamics. The intention of the missionaries was that of Vietnamisation by way of Vietnamese Christians. They had goals to proselytise, explore and subjugate the tribal peoples (Engelbert, 2016: 84-102).

The context in the Indochinese Central Highlands is similar in all South-East Asia –namely Thailand, Laos and Vietnam– if we look at the conceptualisation of the “highlands”. In a comparison between Thailand and Vietnam, Tan and Walker note that the highlands are seen as a different world to the lowlands. Lowlanders think that societies living in the hills are fundamentally timeless and only change if externally induced, which the authors demonstrate to be ungrounded. Lowlanders also tend to think of highlanders as either easily assimilable or resistant –following the dichotomy established by the French missionaries on the nature of the ‘hill tribes’, either ‘martial’ or ‘peaceful’ (Engelbert, 2016: 82)–. This conception

disregards the highlands people as isolated and with a culture that can only be adopted from below, from the cultural lowland centres where it diffuses from (Tan and Walker, 2011).

There are, however, two main differences regarding the nation-state. First of all, the assimilation and economic development practices that affect the ethnic minorities may vary. For example, Vietnam applies "development policies" which intend to foster the migration of lowlanders to the highlands, not to speak of the imposition of the neoliberal model and pharaonic cultural and infrastructural engineering projects. The other difference is that the Central Highlands in Vietnam and to a minor extent in Laos, were devastated by the Vietnam War and its consequences for its inhabitants and the environment, namely destruction and pollution (Taylor, 2008: 7-11).

In fact, we cannot talk of 'Vietnamese highlands' (and by extension, 'Lao' or 'Thai' highlands) until very recently, when the Viet Cong power managed to consolidate control over the region in the 1970s and brought the first nation-state policies. Before that, relationships with *montagnards* were mainly left to the French colonial interests or the Cham merchants (Parkin, 1991: 90).

The impact of the nation-state policies over minorities is severe, with the official narrative putting efforts to minimise it and fit it into the greater national project. In particular, migration from the lowlands provokes discrimination and inequality between two different groups: the highlanders, seen by the lowlanders as 'simpleminded and ingnorant', and the lowlanders, seen by the old-residents as 'stingy and deceptive'. Meanwhile, the government narrative is to describe the internal migration as "fraternal solidarity" (McElwee, 2004).

Although most Vietic groups live in the Central Highlands, an area quite thoroughly researched by anthropologists, I have not found any remarkable anthropological study on them. The only exception is the number of studies on the Mường–who live on the plains south of Hanoi–, especially on their identity (see for instance Trần Thị, 2004–studies on Mường's identity and agency).

2.2 Historical linguistic classification of the Austroasiatic phylum

2.2.1 Problems with the classification of Austroasiatic languages

Austroasiatic languages are an established phylum in South and Southeast Asia, with more than 150 languages over a dozen branches. Their language domain is divided and overlain by speakers of other groups (Indic, Dravidian, Sino-Tibetan, Tai, Hmong-Mien, Austronesian), creating a vast region of ethnolinguistic diversity (Sidwell, 2009: 1).

Southeast Asia as an ethnolinguistic diverse area, with language and culture contacts taking place, which have led to confusions when trying to classify this rich diversity. The lack of data has been another issue that has complicated the classification of Asian languages and mislead some researchers (ibid.: 15-16). Moreover, there is also the historical tendency of some scholars to treat data as a personal property and not to share it with the wider community, which also hinders the progress in Asian comparative linguistics (ibid.: 43).

Paul Sidwell is aware of the several problems that Austroasiatic classification is facing. He is concerned that after more than a century of comparative Austroasiatic studies, there is still no explicitly justified and comprehensive internal genetic classification of the phylum. There is also no scholarly consensus on (1) the relations between Austroasiatic branches, (2) the absolute age or diversity of Austroasiatic, or (3) an appropriate program for addressing these issues. Sidwell adds that "consequently, the field is yet to benefit significantly from extensive multidisciplinary research" (ibid.: 2). The field of Austroasiatic studies has always lacked adequate survey works, with much of the vital literature to be found in hard-to-find books, journals and dissertations. Little of the data used for competing classifications has ever been published, and therefore cannot be evaluated by peer review.

> "To a great extent, we remain reliant on classifications that owe their direct source to typological and lexicostatistical studies of the 1960s and 1970s, and almost nothing in terms of robust cladistic studies of phonology or lexicon that may be readily reviewed or assessed." (Sidwell, 2009: 1-2).

Sidwell's frustrations are joined by those of Blench (2008: 117-118):

> "Austroasiatic languages are the most poorly researched of all those under discussion. Many are not documented at all and some recently discovered in China are effectively not classified. The genetics of Austroasiatic speakers are almost unresearched. Austroasiatic is conventionally divided into two families, Mon-Khmer (in SE Asia) and Munda (in India).

> Diffloth (2005, 79) now considers Austroasiatic to have three primary branches but no evidence for these realignments has been published. Indeed Austroasiatic classification has been dodged by a failure to publish data, making any evaluation of competing hypotheses by outsiders a merely speculative exercise."

Before going into the historical debates over the classification of Austroasiatic, it is appropriate to list the different language families that compose this language phylum. There is little discussion on the belonging of these groups to Austroasiatic, the discussion being more focused on its internal classification and subgrouping (Sidwell, 2009: 3):

Branch	**Main regions where spoken**
Aslian	Malay Peninsula
Bahnaric	Central Indo-China
Katuic	Central Indo-China
Khasic	Maghalaya state of India
Khmer	Cambodia and neighbouring areas
Khmuic	Northern Laos
Monic	Southern Myanmar and central Thailand
Munda	Eastern and Central India
Nicobaric	Nicobar Islands of India
Palaungic	Shan State of Myanmar
Pearic	Cambodia and Thailand
Vietic	Vietnam and Central Laos

Table 3: Branches of Austroasiatic according to Sidwell (2009: 3)

Following Sidwell (2009: 3-4), the further steps the research community has to take are (1) how the branches of Austroasiatic relate to each other–the suggestions made so far are poorly substantiated, (2) advance in a detailed reconstruction of Austroasiatic, lacking at this time, (3) see how and why internal branches vary considerably in internal diversity, and adequately document and reconstruct each of them.

The following sections provide a chronological history of the Austroasiatic classification. Although there is a separate section on Vietic, recurrent referrals to this subgroup will be made as it has been a controversial issue in the history of Austroasiatic classification. Therefore, the issues regarding the classification of Vietic languages within Austroasiatic will

be outlined in the following paragraphs, whereas the issues regarding the internal Vietic classification will be addressed in section 2.4.

2.2.2 Early classification of Austroasiatic

One of the pioneers in finding correspondences between Austroasiatic languages was James Richardson Logan, who in the 1850s wrote about a 'Mon-Annam Formation', which comprised Munda, Mon, Khmer, Vietnamese, Khasi, Nicobarese, Pearic and Aslian, anticipating the description of the Austroasiatic Phylum. In the same period, Mason demonstrated the link between Munda and the languages of Indo-China, and developed the Kol-Annam or Mon-Annam theory (Sidwell, 2009: 5-6).

Some scholars who supported the Mon-Annam theory had doubts about Vietnamese belonging to this proposed group. Kuhn, for instance, considered Vietnamese a sort of creole which was developed with the contact of migrant Vietnamese groups from the north-east with Mon-Annam speakers. His work was very influential and was followed until mid-XXth century. This influence made the classification of Vietnamese very controversial, as for a long time it was considered for many scholars that it did not belong to the Mon-Khmer group despite Logan and Mason's early studies. In fact, Kuhn, like other contemporary orientalists, classified Vietnamese as 'Thai-Chinese' (ibid.: 13-14).

The publication of the Linguistic Survey of India–edited by Grierson and published between 1868 and 1928 in eleven volumes–was an important improvement of this situation, with data on the westernmost Austroasiatic languages. However, this survey mistakenly classified Munda with Dravidian and Mon-Khmer as a subgroup of an Indo-Chinese family, together with Tibeto-Burman and Thai-Chinese (ibid.: 16).

Despite the discussions, disagreements and controversies, by 1900 there were enough demonstrations of what was later called the Austroasiatic phylum and there had already been interesting discussions on its internal classification and typology, as well as representative–although small–data from most of its branches (ibid.: 19).

In this context, Wilhelm Schmidt started to make historical comparative analyses for each of the evident Austroasiatic grouping. His large work–which also intended to link Austroasiatic with Austronesian–caused a schism between neogrammarians and diffusionist tendencies.

Schmidt's huge comparative work contributed to a reconstruction of proto-consonantism and an analysis of some morphological correspondences between many–although not all– Austroasiatic groups, among others. He also suggested a genetic classification of Austroasiatic. In his classification, he includes all the branches accepted nowadays with exception of Vietic (a controversial discussion was being had on Vietnamese which left it as 'unclassified', with the other Vietic languages being still unknown at the time) and includes the Chamic group, now considered Austronesian. Moreover, his subgroupings have radically changed since his original proposal. Schmidt also coined the term 'Austroasiatic' to refer to the phylum. His work had unfortunately been put aside by diffusionists such as the highly reputed Blagden (ibid.: 20-22). Finck followed Schmidt classification but correctly maintained Vietnamese within Austroasiatic (ibid.: 33).

Blagden, contemporary to Schmidt and very well-renowned, also made huge contributions to Austroasiatic studies. He was, however, mislead by the time's obsession with race and misinterpretations of some phenomena caused by language contact. For instance, he identified Sakai and Semang, two Aslian Austroasiatic languages, as mixed with a huge influence of Mon-Annam–and classified as Austroasiatic by Schmidt. In spite of his mistakes, and perhaps as a consistent diffusionist, he, together with other diffusionists, correctly established that diffusion and mixing had had an important role in establishing the linguistic distribution of Southeast Asia (ibid.: 25-26).

2.2.3 Classification of Vietnamese

A parenthesis here needs to be made, as in the first half of the XX century there was a major debate over the classification of Vietnamese that conditioned the discussions over Austroasiatic. While authors like Finck claimed it to belong to the Austroasiatic group, most scholars classified it into Thai or Sino-Tibetan or a combination of both–Thai-Chinese. Some others like von Hevesy even questioned Austroasiatic and the link between Munda and Mon-Khmer (Sidwell, 2009: 35).

The essential question was if Vietnamese was an Austroasiatic language with Tai and Chinese influence or a Tai language–or even Chinese–with Austroasiatic influence. One of the problems was its typology– Vietnamese being an analytic language with no initial clusters and short syllables–, in a time when it was considered to be an argument for genetic

classification. The tones were also an issue, as authors like Maspero believed that a language cannot acquire tones if it previously lacks them (Maspero, 1912, after Parkin, 1991: 90).

Maspero, one of the most prominent researchers on Vietnamese and representative of the French academy views, defended the idea that Vietnamese is Tai with arguments related to tone. Maspero was the first scholar to work on Vietnamese tonogenesis, a solid work that later proved to be in the right path and granted him a good reputation. His mistake about the classification of Vietnamese prevented him from completing his work on the origin of the tones before Haudricourt did in the 1950s (Sidwell, 2009: 33-36). Blagden followed Maspero in arguing for a Thai origin for Vietnamese, although both of them accepted Austroasiatic substrate (Parkin, 1991: 89-90).

Przyluski's work in the 1920s was one of the first to include Mường, a Vietic language and the only known close relative to Vietnamese at the time–and which is phonologically more conservative than Vietnamese. These new data were relevant in building up more arguments on the affiliation of Vietnamese with Austroasiatic. The Viet-Muong subgroup then became a member of the Austroasiatic phylum, although the position of Vietnamese was contested until the second half of the past century. Przyluski also contributed to the reconstruction of proto Viet-Muong with data from the early missionaries. He then demonstrated that the monosyllabic structure of Vietnamese could not be explained by Vietnamese belonging to the Tai or Chinese groups (Sidwell, 2009: 33-34).

Przyluski's work also contributed to expand the Mon-Khmer group and consider it a family with many subgroups, which were considered direct groups of Austroasiatic before. Austroasiatic was reduced to two main families: Munda and Mon-Khmer, with discussion over the Vietic one, which was considered a subgroup of Mon-Khmer, a family within Austroasiatic or out of the phylum (ibid.: 35).

Sebeok's 1942 work had a strong relevance and was cited and endorsed by many scholars. He questioned the existence of the Austroasiatic family, manifesting that there were no strong evidences to link Munda with Mon-Khmer, followed Blagden on his views about Aslian as Austronesian instead of Austroasiatic and, based on typological arguments, classified Viet-Muong into Tai. Haudricourt, in his work in the 1950s, offered robust data that linked Viet-Muong to Mon-Khmer and rescuing the Austroasiatic hypothesis, which recovered prestige and relevance (ibid.: 37).

In the end Haudricourt's work (1954, 1961) could convincingly explain and demonstrate the tonogenesis process, and also convincingly argue that tone was an areal feature that developed in a similar way in Thai, Vietnamese and Chinese. After his publications, all scholars finally accepted and defended the belonging of Vietnamese and Viet-Muong to Austroasiatic. Data from Muong and the newly described Vietic languages, which have more conservative features, which make it resemble more the other Mon-Khmer languages, also helped improving the tonogenesis demonstrations and therefore convince the academic community (Sidwell, 2009: 33-36; Parkin, 1991: 89-91).

2.2.4 Modern classification of Austroasiatic

After Haudricourt's contribution, Pinnow's work in the late 1950s and 1960s represented the next big advancement in Austroasiatic studies. He provided a reconstruction of proto-Munda and proto-Austroasiatic etyma and also presented an Austroasiatic etymological dictionary. He designed a new Austroasiatic classification and sub-grouping that went beyond Schmidt's and adopted some of Przyluski's findings to make important advances, such as consolidating Przyluski's clear division of Munda and Mon-Khmer as separate families within Austroasiatic. Vietic/Viet-Muong, however, was not included, following Maspero and Sebeok. A big gap that was still to be covered and was noticed by Pinnow was the lack of data from eastern Austroasiatic languages, which was essential for their proper classification (Sidwell, 2009: 38-40).

Pinnow's claims were heard and in the 1960s young fieldworkers together with The Summer Institute of Linguistics, bringing along with them new methods like lexicostatistics, started to collect data in Southeast Asia. Thomas' work is a good example, showing how lexicostatistics, easily applied to analytic languages, allowed for a subgrouping of Mon-Khmer in a Northern–Katuic–and a Southern group–Bahnaric–, a distinction that could not have been possible years before. Thomas' paper transformed the classification premises by excluding any geographical and typological factors and relying on purely linguistic data, *signifié et signifiant* (ibid.: 43).

Further studies were carried out on other Austroasiatic groups, using a modified Southeast-Asian Swadesh list. Thomas and Headley 1970's study identifies 4 families composing Austroasiatic: Munda, Mon-Khmer, Malacca (Aslian) and Nicobarese, with Viet-Muong

being a subgroup within Mon-Khmer. This paper consolidated the Austroasiatic phylum and derived the discussion to its sub-groupings (ibid.: 44-45).

Thomas and Headley's paper influenced two of the most prominent researchers on Austroasiatic: Ferlus–who worked on Vietic languages, among others–and Diffloth. Both of them published separate classifications in 1974, which have been of reference for many years and will be discussed in the following section. There were also scholars, like Huffman, who were sceptical about the lexicostatistic method (ibid.: 46-47).

Headley demonstrated that different methodologies–phonological features, cognate percentages and lexical innovations–showed very different *stammbäume*, one of the big issues concerning sub-grouping. Most classifications using lexicostatistic methodology were based on lexical innovations, such as Diffloth's. Headley updated his 1970 classification with Thomas in 1976, arguing that Khmer is an isolate inside Austroasiatic (ibid.: 49).

In the 1970s and 1980s, a SOAS-based scholar, Harry Shorto, was working on a vast work, the Mon-Khmer Descriptive Dictionary, which was posthumously published in 2006. He also made some classification attempts at classification using lexicostatistics, mostly unpublished, which shed some light into Mon-Khmer branching issues (ibid.: 50-54).

2.2.5 Recent classifications

Austroasiatic classification is still nowadays a source of controversies, mainly regarding subgrouping proposals. Here I present some classifications provided by the most relevant authors in the field. Some of them have revised these classifications over time. I present their most recent classifications and provide a brief discussion on the characteristics of their proposals.

Gérard Diffloth is one of the experts on Austroasiatic languages. In his first classifications were published in the 1970s and 80s, when he argued for a Northern Mon-Khmer branch including Palaungic, Khmuic, Khasic and Viet-Muong. Later, he grouped Viet-Muong with the Eastern Mon-Khmer branch, opposed to the Northern and Southern branches, which he maintained until his last classification of 2005. His grouping of Katuic of Vietic is interesting for the h : s correspondences he found. Other internal grouping have changed in the author's

classifications. Unfortunately, the evolution of Diffloth's classifications is not much discussed (Diffloth, 2005; Sidwell, 2009: 54-55).

His last classification, presented below, rescues his early classification between Northern and Southern Mon-Khmer/Austroasiatic languages, collapsing again the Eastern and Southern branches into one. He also argues for the elimination of any distinction between the concepts of Austroasiatic and Mon-Khmer (Diffloth, 2005).

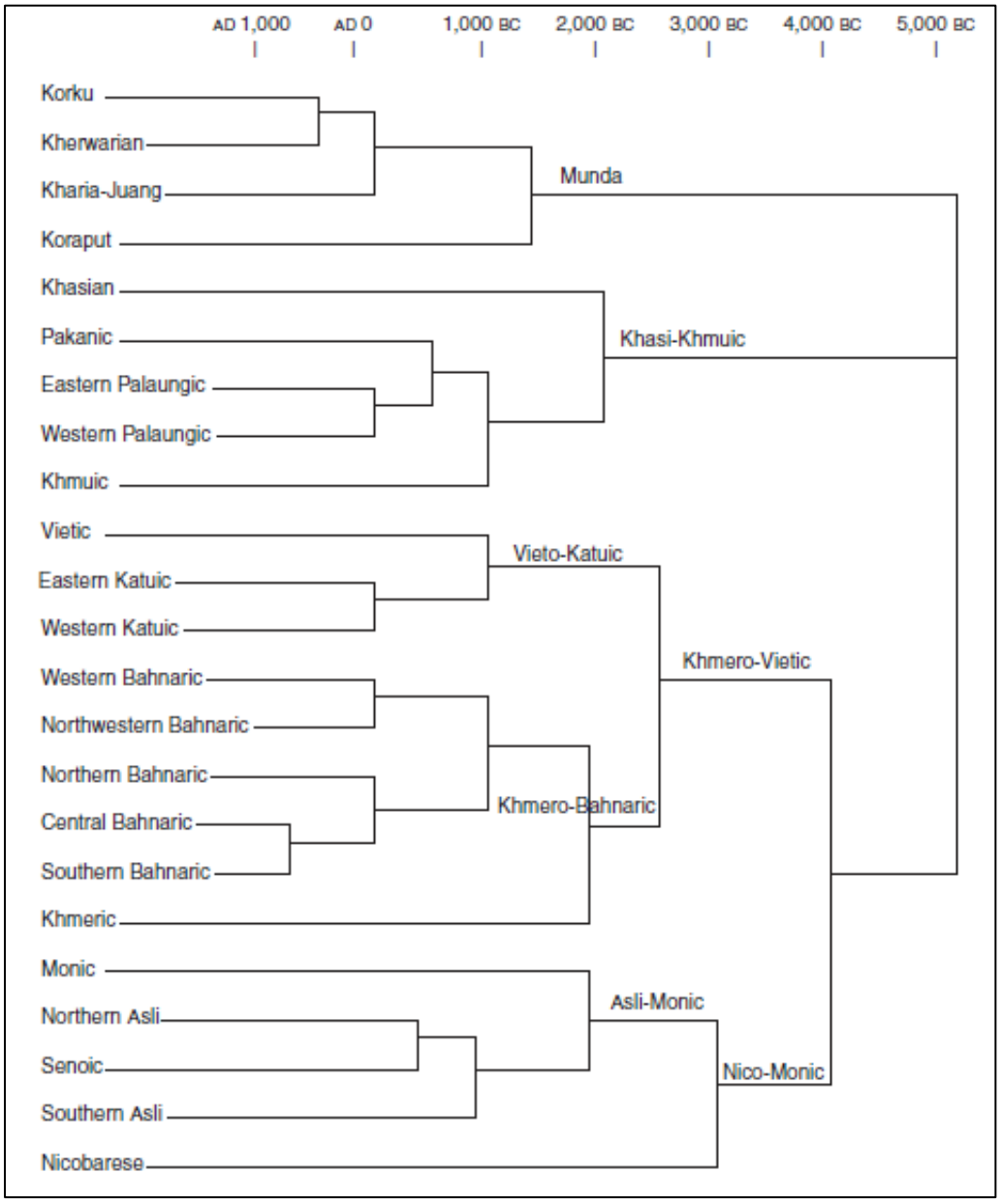


Figure 2: Austroasiatic classification according to Diffloth, 2005.

Ilia Peiros is another linguist who has worked on the classification of Austroasiatic and the whole of Southeast Asia using lexicostatistical methods. One of his goals is to propose a

macro-southeast-Asian phylum called Miao-Austroasiatic. He attempted two classifications: one in 1998 and another one in 2004. The latter includes the Nicobaric branch, absent in the first study. His 2004 diachronic classification using lexicostatistical methods is the following (Sidwell, 2009, adapted from Peiros, 2004–in Russian):

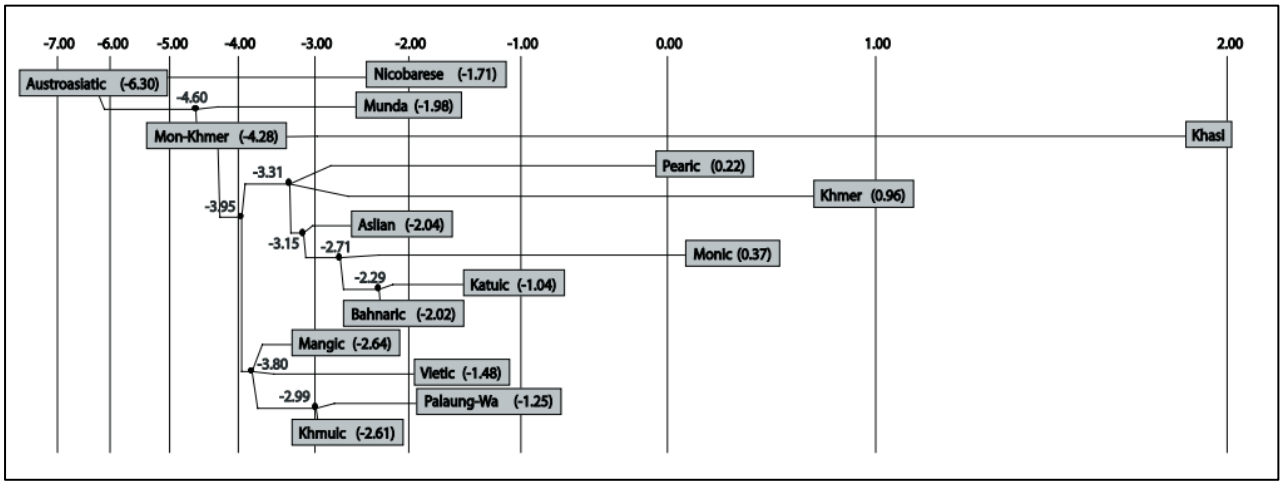


Figure 3: Austroasiatic diachronic classification by Peiros, 2004 (retrieved in English from Sidwell, 2009: 43).

Michel Ferlus and Paul Sidwell have proposed similar classifications. Ferlus' classifications were proposed in the 1970s and have not been revised ever since. In a first classification with Matras, they propose 12 Mon-Khmer subgroups (adapted from Sidwell, 2009: 46, adapted from Matras and Ferlus, 1971):

A	Mon
B	Khmer
C	Pear
D	Mnong-Mà
E	Bahnar-Sedang
F	Laven-Brao
G	Katu
H	Khamou
I	Palaung
J	Mang
K	Viet-Muong
L	Khasi

Table 4: Austroasiatic groups according to Matras and Ferlus, 1971.

In 1974, Ferlus revised the internal subgrouping. His most notable contribution is the proposal of a Northern Austroasiatic sub-group, which includes Palaungic, Khmuic and Mang. This proposal ressembles Peiros' one. In fact, Ferlus used some lexicostatistical methods, together with looking at lexical innovations (Ferlus, 1974; Sidwell, 2009: 46).

Sidwell has been revising the Austroasiatic classifications for the past decade. His last classification dates from 2015:

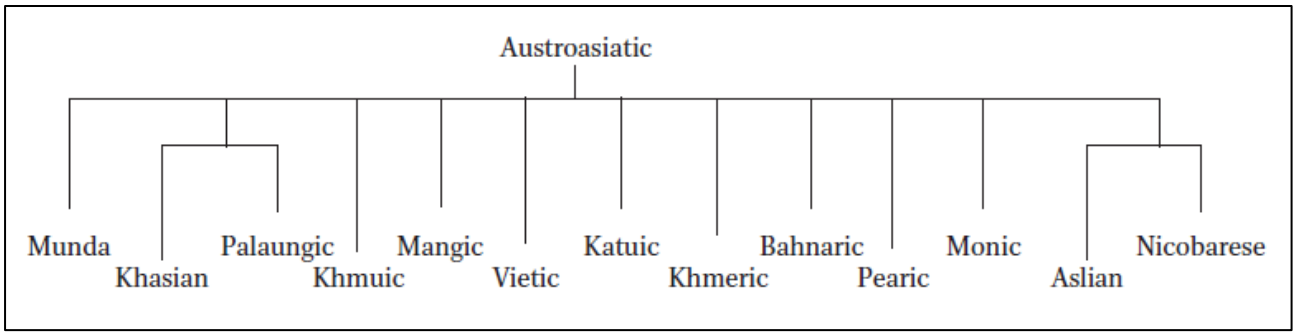


Figure 4: Sidwell's classification of Austroasiatic (Sidwell, 2015: 179).

His non-lexicostatistical proposal can be considered synthetic, taking into consideration what other authors have contributed to the question, together with his work, and proposing a consensual subgrouping (Sidwell, 2015: 179). He is currently working on internal subgroup classifications.

2.3 Vietnamese minority classifications

In this section I will outline the classifications of Vietnamese ethnic minorities. I will first discuss the problems that minority classification in Vietnam faces. I will move then to a chronological history of the classification, beginning with the first missionaries and finishing with the Vietnamese government official classification.

2.3.1 Problems with the classification of Vietnamese ethnic minorities

Like in other parts of the world, subgrouping and dividing linguistic continua are the most challenging tasks when classifying the groups belonging to a particular language family. The Vietnamese area is not an exception, and the distinction between languages and dialects has

always been controversial, especially with, as Engelbert mentions, the Austronesian groups, whose languages are mutually intelligible (2016: 85).

The classification of ethnic groups in Vietnam is not only an academic issue, but also a political one. As explained earlier, in Vietnam an ethnic minority only exists if it is officially recognised. Moreover, Taylor explains that:

> It has been assumed that the minorities are inescapably part of the national project. The delineation of national borders, the census, and the ethnological survey have constituted "national minorities," who are assigned a name, place, and rank within the modern nation, written into national history, and reformed in line with national mores. […] They are subject to a systematic official project of ethnic classification and counting. Their age-old ways of life are depicted in development reports funded by international agencies as "poor," "backward," and "deficient" in relation to the ethnic majority. (Taylor, 2008: 4).

Classification of ethnic minorities in Vietnam is an official pursuit for the government, and its political use affects the research. Nevertheless, it is also an official pursuit for a number of researchers, implying again another set of biases with regard to research on ethnic minorities (Taylor, 2008: 28-29).

2.3.2 Minority classification in Southeast Asia and Vietnam: first attempts

We owe the first attempts to study and classify the Vietnamese minorities to (mainly French) missionaries–and also some explorers–, who did some work–mainly ethnographies and word lists–on this issue. Catholic missionaries were also the first to start building permanent bases in remote areas–the highlands, for instance–, paving the way for the later French and Vietnamese domination and homogenisation of minority groups. We must not forget that the missionaries and explorers were conscious advocates of colonial expansion (Engelbert, 2016: 76-88).

Alexandre de Rhodes is one of the earliest and most important missionaries who worked in Vietnam during the XVII century. He developed the currently in use Vietnamese Latin alphabet–quốc ngữ–as well as a Vietnamese-Latin-Portuguese dictionary. He also wrote on Vietnamese history and developed Catholic missions in Vietnam. Other missionaries worked with minority groups, mainly converts to Catholicism, and wrote dictionaries and

ethnographies, as well as religious studies and catechisms for conversion purposes (Engelbert, 2016: 99-100).

The missionaries and explorers started to group and classify the ethnic minorities under –mixed–criteria of language, natural habitat, economy and culture. One of the recurrent classifications, as seen, was to label the tribes as "peaceful races" (*races paisibles*) or "martial races" (*races guerrières*), depending on their position as captors or prisoners in tribal conflicts. This binary classification was useful for the colonisers to know which groups were easy to negotiate with. Some of the first groupings made back then–with their biases and colonial intentions–continue nowadays (Engelbert, 2016: 81-82). For instance, the Chứt group is an officially recognised minority group in Vietnam, which comprises in fact several Vietic languages–Rục, Sách, Salang, Mày, Mã Liềng, etc.–. The word 'Chứt' means 'mountian dweller', and it is recognised as one group with one language despite the group's internal diversity (Ferlus, 1996: 14-15). Scholars such as Hayes, Sidwell or Ferlus follow the Vietnamese official classification, grouping all the ethnic groups under the label 'Chứt', whereas others such as Peiros or Chamberlain treat the varieties as separate languages (see section 2.4.2: Vietic historical classification).

In the case of the highlands, the first groups to start negotiations with the French were those defined as "peaceful races", easier to control for the French colonial authorities–and later also the Vietnamese and Lao authorities–. They were also the first to begin being converted to Catholicism. Thus, they were also the first to be studied and classified. The Bahnar group is good examples of these first contacts (Engelbert, 2016: 86).

2.3.3 Official classifications

There have been 3 classifications of ethnic groups so far, in 1959, 1973 and 1979. The Vietnamese national censuses follow these classifications.

The official classification of 1959 included 63 ethnic groups, which were grouped by the criteria of language, culture and geography. The groups make up a total of 3,298,546 people belonging to these ethnic minorities. This classification was the first comprehensive attempt to list all the minority groups of Vietnam. These groups, adapted from Schliesinger (1997: 4-13), are:

a) Han-Tang languages, which includes the groups (1) Tang-Mien, (2) Chinese-Tay (with a Tay and a Chinese branch), (3) Meo[1]-Giao and (4) Lac-Viet.
b) Mon-Khmer languages, which includes the groups (1) Xo-Dang – Ma-Puoc and (2) Khmer.
c) Malay-Polynesian languages, which includes only one subgroup, the E-de – Gio-rai – Ra-glai group.

The Han-Tang language group collapses the Tay/Thai groups with the Sino-Tibetan ones, together with Hmong-Mien and Vietic. At that time, Vietnamese was still considered a "Chinese-Tay"[2] language by many scholars (see next section). Therefore, the Vietic minorities–called Lac-Viet in this classification–belong to the Han-Tang group and not the Mon-Khmer group. This classification contemplates 2 subgroups of the Lac-Viet, the Mường and the Dan-Lai. The Mường group includes the Ao-ta, Nguôn, Sách and Thổ. The Dan-lai corresponds to the Liha, Li-ha or Ha-do Vietic group. Other Vietic groups such as Ruc, "discovered" in 1957 (Schliesinger, 1997: 28), were still unknown or unclassified.

The 1973 classification was built on new data from scientists, and again using linguistic, cultural and geographic criteria. This time 59 ethnic groups were listed, including the Viet/Kinh–to provide the picture of the whole ethnic composition, not only the minorities–, and were grouped into 3 different language families (Schliesinger, 1997: 14-17):

a) The South-Asian language family, which includes the language groups (1) Viet-Muong, (2) Mon-Khmer, (3) Tay-Thai, (4) Meo-Zao and (5) Kadai.
b) The Sino-Tibetan language family, which includes (1) Tibeto-Burman and (2) Han (Chinese).
c) The Malay-Polynesian language family, with no subgroups.

In this classification, the Tay-Chinese family is separated, leaving a Sino-Tibetan family, which was already accepted at that time, but proposing a macro-South-Asian family which includes the Tai-Kadai/Kra-Dai groups, the Hmong-Mien and the Mon-Khmer/Austroasiatic. This macro-grouping follows the Austric macro-phylum proposal (although without the Austronesian group), proposed in 1906 by Schmidt and still followed by some scholars, such as Benedict or Starostin. Viet-Muong, however, is considered a subgroup of South-

[1] *Meo* is a term used for the Hmong ethnic group
[2] Blagden's terminology

Asian/Austric and not a subgroup of Mon-Khmer. The Viet-Muong minorities included in this list are the Mường, the Thổ and the Chứt.

The 1979 classification is the one in force ever since it was established. It defines 54 ethnic groups grouped into three main language families, which are divided into eight language groups (Schliesinger, 1997: 18-22):

a) The Austro-Asian language family, including the (1) Viet-Muong, (2) Mon-Khmer, (3) Tay-Thai, (4) Meo-Zao and (5) Kadai language groups.
b) The Austronesian language family, including the Malay-Polynesian language group.
c) The Sino-Tibetan language family, which includes the (1) Chinese language group and the (2) Tibeto-Burman.

This is a very similar classification to the previous one of 1973. The language families do not change, and there are small changes only in the group names and merging of two ethnic groups into one, reducing them from 59 to 54. The classification takes the data from the 2009 national Vietnamese census (SPH, 2010). The table outlining the full picture Vietnam's ethnic composition according to the 1979 classification can be consulted in the appendices (section 7.1).

The Vietic groups in the 2009 census comprise 74,943,870 people, which represents 87.3% of Vietnam's population. The Viet/Kinh majority group alone represents 85.72%. The Vietic minority groups–excluding the Kinh–represent therefore only a 2.1% of Vietnam's total population. By groups, the Mường is the largest, making up a 1.48%; the Thổ make up 0.09% and the Chứt 0.007%. The group Chứt defined by Vietnamese linguists is a broad group created because of cultural and geographical factors. It includes the Rục, Mày and Sách languages, among others. The word 'chứt' is a Cham word for 'mountain dweller' (Sidwell, 2009: 140).

2.4 A historical approach of Vietic languages

2.4.1 Vietic languages: an overview

Vietic languages are a subgroup of the Austroasiatic family. They are spoken in South-East Asia, mainly in Vietnam and Laos, but also in Thailand and Cambodia. The Vietic group is also called Viet-Muong (the historical denomination) or Kri-Mol (coined by Chamberlain,

2018: 9, arguing that the term 'Vietic' can cause confusion with Vietnamese and Viet-Muong). 'Vietic' is nowadays the most common denomination and has been since Hayes (1982) coined it (Sidwell, 2009: 140).

The number and nomenclature of the Vietic languages remains controversial. The big number of ethnonyms, exonyms, names of different varieties or simply ways of writing them often lead to much confusion. For practical reasons, I will follow Ferlus' 1996 classification of Vietic languages and their varieties. Also, because Ferlus is the scholar who has carried out fieldwork the most on Vietic languages. According to him, there are 8 Viet-Muong languages (in parentheses other variants):

Maleng (Pakatan, Bo, Malieng, Kha Phong...)	Hung (Pong, Phong, Tay Pong, Lihà)
Arem (Chomrau, Chombrau, Umo)	Thổ (Cuối, Cuối Chăm, Mọn)
Chứt (Sách, Rục, Mày, Salang)	Mường (Mol, Mual, Mon, Nguồn)
Aheu (Thavưng, Sô)	Vietnamese

Table 2: List of Vietic languages by Ferlus, 1996: 12.

The Vietic language family was only defined in the 70's after the recognition/discovery of the last members of this group. The first complete list of the Vietic language varieties was offered by Ferlus in his 1974 paper. Vietic languages can be divided into two groups: (1) Vietnamese and Mường and (2) the rest of the language varieties. The first group is the oldest of which the research community has notice. Both Vietnamese and Mường have a large number of speakers (around 80 million and 1 million respectively) and have been historically heavily influenced by Tai and Sino-Tibetan languages, especially by Chinese. The second group started to be spotted in the early 20th century by French scholars, and its last members were recognised by Vietnamese scholars in the 60's and 70's. All these languages are spoken in high areas in the Central parts of Laos and Vietnam, and all of them have a very reduced number of speakers and are severely endangered. Also, they have been isolated from the influences of Tai and Sino-Tibetan studies, and they present a set of features closer to the other Austroasiatic languages (Sidwell, 2009: 140).

Vietic

Viet-Muong

Vietnamese (vie)
Mương (mtq)
Bo (bgl)
Nguon (nuo)

Pong-Toum

Pong (hnu)
Cuối (tou)

Chut

East Chut

Arem (aem)
Maleng (pkt)
Chứt (scb)

West Chut

Thavung (thm)

The following map, adapted from Ferlus, (1996: 22), shows the geographic distribution of the Vietic languages–excluding Vietnamese:

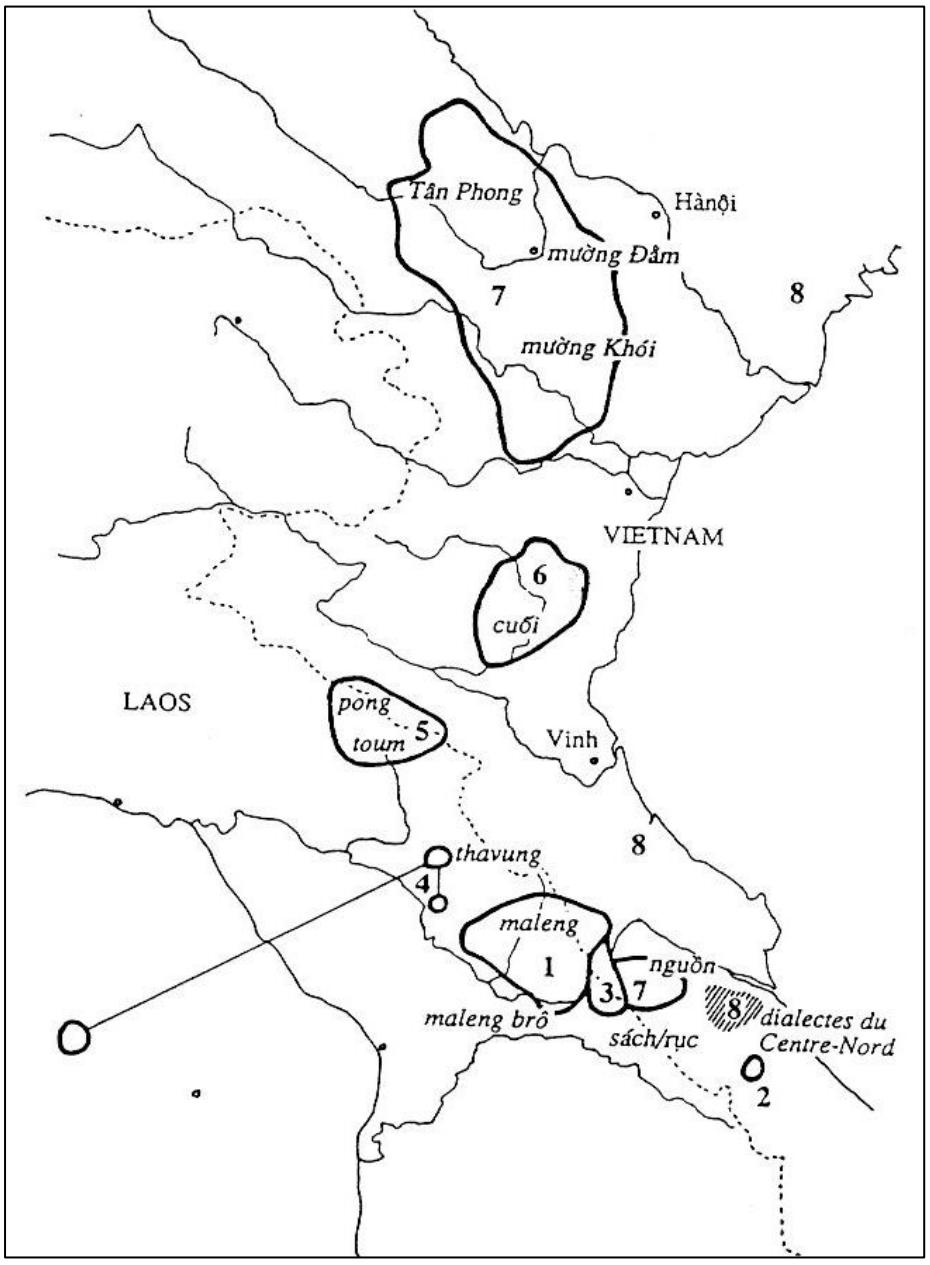


Figure 1: Map with the geographic location of Vietic languages (Ferlus, 1996: 22).

2.4.2 History of the classification of the Vietic languages

Vietic languages are an attested subgroup of the Austroasiatic/Mon-Khmer phylum spoken around the border between Laos and Vietnam–and Vietnamese as the national and only official language of Vietnam. For many centuries, Vietnamese was the only known language of this subgroup, until Mường was documented in the early 20th century. Vietnamese appears to be the least representative of the Vietic languages (Sidwell, 2009: 140). In fact, some authors classify it as a Sino-Austroasiatic creole developed on the coasts of North Vietnam (Chamberlain, 2003). As we have seen with the problems in the classification of Vietnamese during the XXth century, the 'discovery' of the other Vietic languages helped defining their role within Austroasiatic. This late discovery, however, also means that the classification of these languages is currently an ongoing task which still largely has to be developed.

Between 1900 and 1970, several Vietic languages were attested and partly documented. The first was the Mường language(s), attested after 1905. In the early XXth century, the term Viet-Muong was coined as a synonym of the Vietic languages–since at that time only those two languages were known. Nowadays, Viet-Muong stands for a subgroup of the Vietic languages that includes Vietnamese and Muong; this has been the case since Hayes proposed the renaming, although not all scholars use it in this way (Hayes, 1992: 212-213).

Some of the languages 'discovered' over the following decades were Thavưng, Chứt. Sách, Rục, Mường, Harème or Thổ. The first comprehensive list of languages was offered by Ferlus in 1974. Michel Ferlus was the first scholar, together with other Vietnamese scholars such as Trần Trí Dõi, ever to carry out basic fieldwork on all known Vietic groups at the time. The basic gathered data was mainly used for historical linguistic purposes, and in 1979 Ferlus presented the first classification of the Vietic languages which includes the vast majority of them–although some "new" languages spoken on the Nakai Plateau have recently come to light, such as Atel, Atop, Makang, Arao and Thémarou (Chamberlain, 2018: 9).

Ferlus' 1979 classification–highlighting the position of Thavung–(Ferlus, 1979: 81, adapted by Sidwell, 2015: 204).

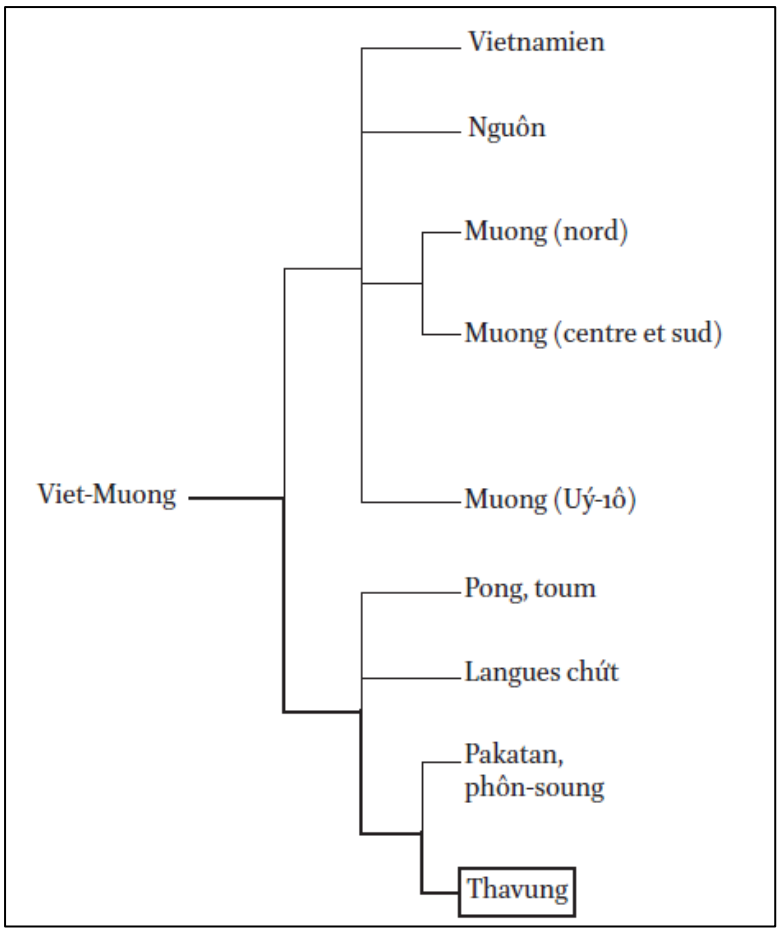


Figure 5: Ferlus' 1979 classification–highlighting the position of Thavung–(Ferlus, 1979: 81, adapted by Sidwell, 2015: 204).

Ferlus argues that there are two main Vietic groups, the Northern Vietic languages, with Vietnamese and Mường and its related languages and the Southern Vietic languages, with the languages spoken in the Centre-North of Vietnam, generally the languages with the lowest number of speakers located in the highlands. The Northern languages were influenced by Sinitic languages and the Southern languages are more conservative, having escaped the Sinitic influences and maintaining more Austroasiatic features (Ferlus, 1996: 10-12; Sidwell, 2009: 141).

This view can be observed in Ferlus' typology of his proposed 5 subgroups, published in 1989/90. In this classification, 1.1 and 1.2 represent the Southern Vietic languages out of the Sinitic influence, whereas groups 2, 3 and 4 form the Northern Vietic groups with heavy influence from Sinitic languages:

1.1 Eastern archaic: Mãliềng, Arem, Chứt
1.2 Western archaic: Thavung, Pakatan
2. Pong-Toum group
3. Mương
4. Vietnamese

Figure 6: Vietic subgroups according to Ferlus, 1989/1990 (adapted from Sidwell, 2015: 204).

In 1982, Hayes published a classification of the Vietic languages using lexicostatistical methods and coined the term 'Vietic' (Hayes, 1982). Ten years later, Hayes revised it and proposed 3 sub-branches of Vietic (Hayes, 1992: 220-221):

1. West Vietic.
1.1. Thavưng (Kha Tha Vưng).
1.2. Kha To(o)ng Luông, Phon Soung (Phôn Xúng).
1.3. Kha Bô, Kha Mường Ben (Bên), Kha Nặm Ớm, Pakatan.
1.4. Harème, Kha Phọng.

2. Central Vietic.
2.1. Đan Lai, Katiam Pong Houk, Ly Hà, Tà P(o)ọng.
2.2. Hung, Không Khêng
2.3. Toum (Tày Túm, Ktum).
2.4. Cọi, Cuổi, Tày Chăm, Tày Pụm.

3. East Vietic.
3.1. Chứt.
3.1.1. Arem, Mã Liềng, Mày, Rục, Sách.
3.1.2. Kha Mụ Già
3.2. Viet–Muong.
3.2.1. Mường–Nguồn.
3.2.1.1. Mường dialects.
3.2.1.2. Nguồn.
3.2.2. Vietnamese.
3.2.2.1. Centrolineal dialects (Hanoi, Huê, Saigon, etc.)
3.2.2.2. Archaic dialects (Haut Annam).

Figure 7: Hayes' classification of Vietic languages (Hayes, 1992: 220-221).

At Hayes' time, there were adequate descriptions of only 4 languages (Vietnamese, Mường, Rục and Thavưng) out of the 30 languages named by Ferlus in his 1979 classification (Hayes, 1992: 212). Parkin discussed in 1991 the internal relationships of 20 languages identified as Vietic, but did not propose any classification (Parkin, 1991).

In the same decade, Diffloth (1989–non-published) proposed a classification dividing the Vietic languages in five groups, reproduced in Chazée (1999) and adapted below. Diffloth had already proposed an Austroasiatic classification in 1980. The main difference is in the Vietic group–called Viet-Muong in 1980 and Vietic in 1989–, which was classified as an

independent Mon-Khmer subgroup, out of the 'Northern MK' and the 'Southeast MK'–and its subgroups 'Eastern MK' and 'Southern MK'–. In 1989, Vietic forms a subgroup of 'Eastern MK' together with Katuic:

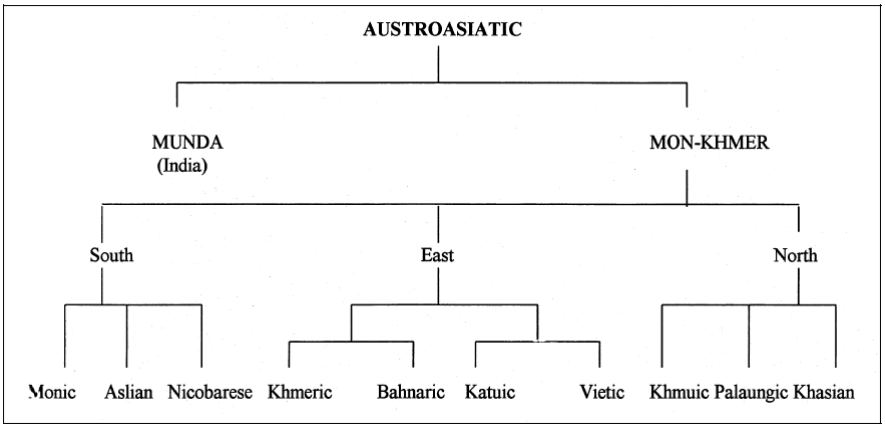


Figure 8: Austroasiatic classification of Diffloth (1989), obtained from Chazée (1999).

In the same book in which Peiros proposed a classification of the Austroasiatic phylum based on lexicostatistical methods, the author also proposed a classification of each of the sub-families, including the Vietic (Peiros, 2004: 37, adapted from Sidwell, 2009: 142):

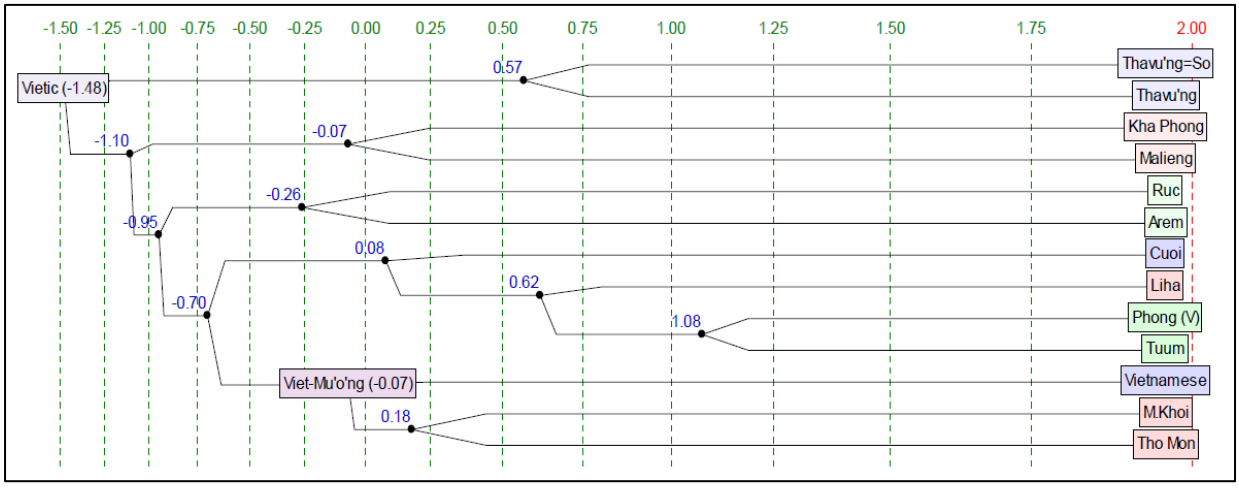


Figure 9: Vietic classification from Peiros, 2004: 37, after Sidwell, 2009: 142.

Sidwell has also proposed a Vietic classification synthesizing the present state of knowledge. His classification resembles that of Ferlus, arguing for a Viet-Muong subbranch, a Pong-Toum group and a Chứt group, divided into East and West. Both Sidwell and Ferlus follow the official Vietnamese classifications in grouping the Southern Vietic languages under the Chứt label, although contrarily to Ferlus leaves Arem out:

Vietic
- Viet-Muong: Vietnamese (various dialects), Mương Muốt, Mương Nàbái, Mương Chỏi etc.
- Pong-Toum: Phong, Đan Lai, Hung, Toum and others
- Chut:
  - East: Mãliềng, Maleng, Arem, Kri, Chứt (Mày, Rục, Sách, Mụ Già)
  - West: Thavung, Pakatan

Figure 10: Vietic classification according to Sidwell (2015: 205).

Chamberlain's classifications are the most recent and differ from the ones of Sidwell and Ferlus. His 2003 classification presented below (Chamberlain, 2003: 422) classifies the Vietic languages in 6 sub-groups, with no mention about their relationship between each other. Maleng is classified in the Southwest branch–not to be confused with the Chứt-related language Malieng, here belonging to the Southeast branch. Kri was then a newly described language and was added as the sixth group, one more than the 5 proposed earlier by Diffloth (1989).

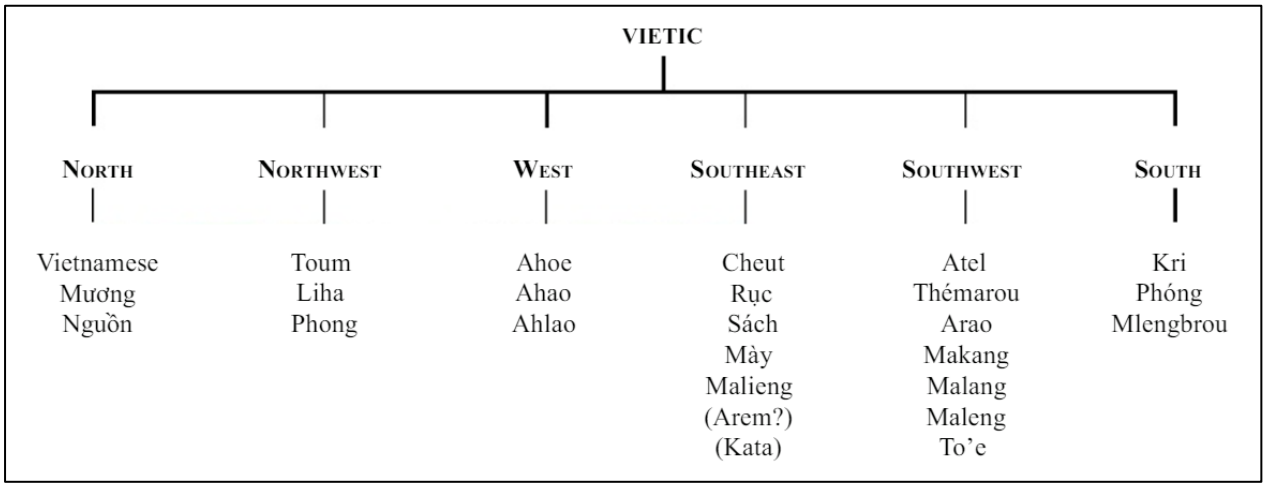


Figure 11: Chamberlain's 2003 Vietic classification.

Chamberlain's 2018 classification revises his 2003 one. He added the newly described languages mentioned above, proposed an internal subgrouping and renamed some of the terms used up until now–the most important change being to rename the Vietic group as Kri-Mol. He also proposes a wider classification of the Vietic languages with Katuic, named Vieto-Katuic.

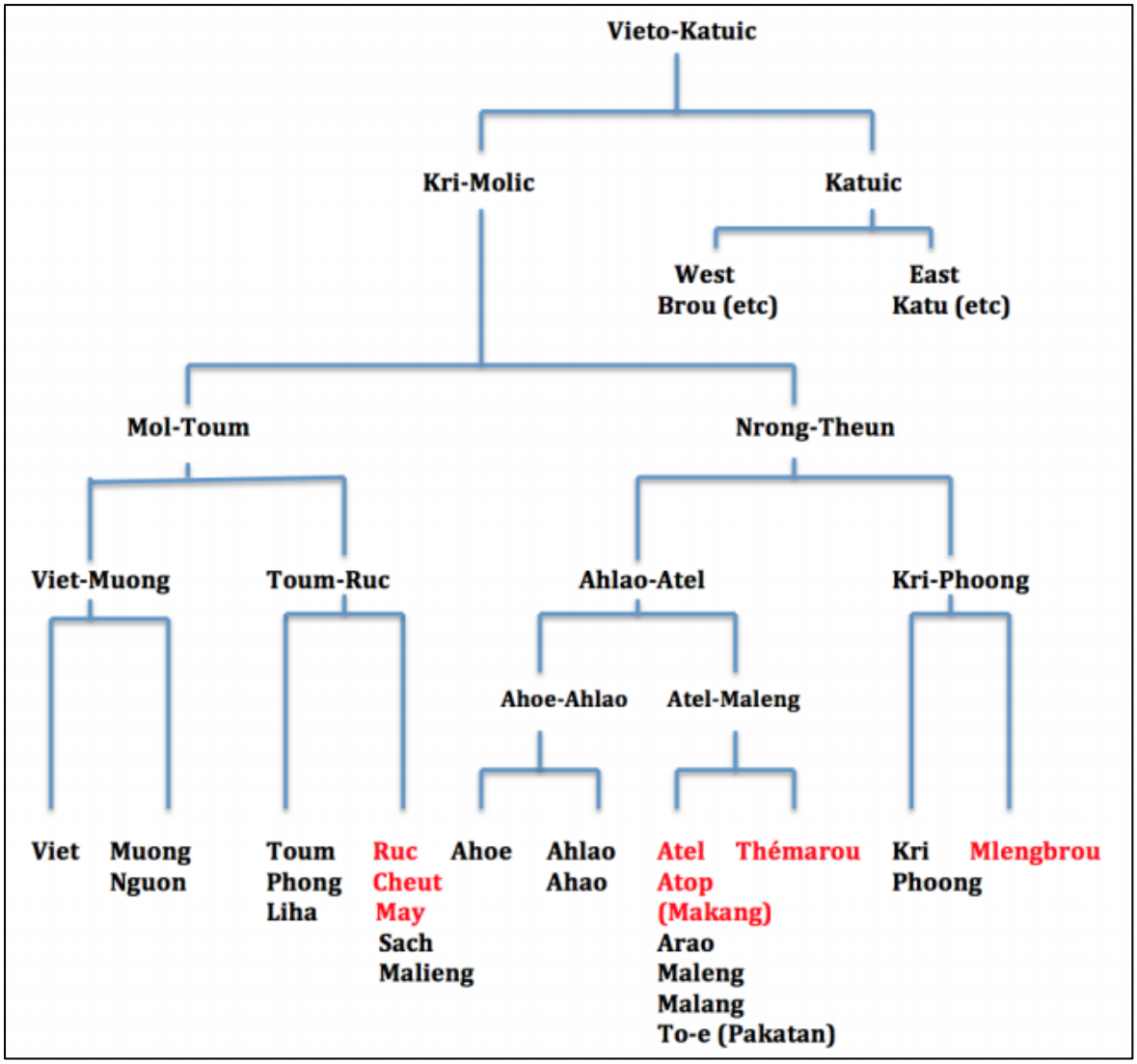


Figure 12: Austroasiatic classification according to Chamberlain 2018: 12.

Regarding the relationship with other Austroasiatic groups, Ferlus argues that Vietic is closer to Khmuic, whereas Alves, Chamberlain and Diffloth group it with Katuic (Chamberlain, 2018: 9).

As one can observe, there are many controversies among scholars on internal Vietic classification. In any case, it is clear that new data is urgently needed, especially because of the degree of endangerment of these languages.

3. Methodology and ethics of the research

Documentation: The project contemplates a large-scale linguistic and cultural documentation of the Malieng community, including: elicitation, traditional stories and narratives, traditional songs, linguistic interaction (dialogues, conversations), traditional practices and activity descriptions, cultural and traditional knowledge (such as medicine or mountain farming) and linguistic elicitation. All the data is intended to be given to the community and also deposited into the ELAR archive, based at SOAS, with the consent of the community.

Data analysis: The data analysis will be focused on phonology, and most of the elicitation sessions will be phonology-oriented. More particularly, I am interested in the tonal system and the role of Maleng in the tonogenesis theories. I will therefore provide a detailed phonological description with focus on tone and also on historical linguistics, contributing to the Vietnamese tonogenesis (the origin of the tonal system) theories (Haudricourt, 1954; Ferlus, 1998; 2004; Alves, 1995) for Vietnamese but also for the other Vietic languages and for historical linguistics (proto-Vietic, Old and Middle Vietnamese). It will also contribute to general as well as South-East Asian linguistics, historical linguistics, phonology and the study of Asian tone. The documentation of the language will at the same time provide relevant data for future research on other aspects of language and culture and serve as a basis for future research on Vietic communities, in linguistics or other disciplines.

Theoretical framework: The theoretical basis needed to carry out this project has two main general topics: theoretical phonology and tonal studies and Austroasiatic Linguistics, including Southeast Asian Linguistics for a more general overview. The theoretical research on phonology and phonetics will provide me the tools to choose a proper framework to work with. For this upgrade chapter, I have evaluated research on Austroasiatic Linguistics and Southeast Asian Linguistics to see what frameworks and analysis are most commonly used. I have investigated about tones and tonal phonology, especially Vietic and Vietnamese phonology. The research done on Vietic languages and Vietnamese is very focused on historical linguistics and the comparative method. I have included these studies into my thesis, but I have also looked for more modern analysis on phonetics and phonology (for instance Nguyễn, 2015). I have read all of Michel Ferlus' and Trần Trí Dõi's work on Vietic languages, since they are the only experts Vietic languages who have also carried out

extensive fieldwork in Southeast Asia, some of it on Malieng and on the Quang Binh[3] languages. I have also evaluated the work of reputed linguists with expertise in Southeast Asia and Austroasiatic, such as Paul Sidwell or Mark Alves, among others.

Overseas fieldwork: I will carry out fieldwork in Vietnam during the second and the third years of the PhD. During the second year, I will be approximately 9 months in the Quang Binh province, plus 3 months in a second field trip during the third academic year. During fieldwork, I will gather linguistic and cultural data from the Malieng community. The archived data will be of use for a range of different disciplines (linguistics, anthropology and cultural studies, religion studies, botany, etc.), and my field methods approach will be designed to be as interdisciplinary as possible, though a special focus will be given to linguistics, particularly phonology. I will familiarise when preparing fieldwork with other approaches from other disciplines, especially anthropology and cultural studies, in addition to my training on linguistic fieldwork.

These data will be archived at SOAS, at the ELAR archive for endangered languages, with whom I am in contact. I am also in contact with researchers in Paris (Michaud and Ferlus) and Hanoi (Trần Trí Dõi), whom I am eager to collaborate with. The data analysis will be carried out using Praat, Toolbox, ELAN and Say More, among other relevant tools. The training and experience in field methods and language description I have received at SOAS and thanks to the program 'Engaged Humanities' will be clue and determining.

Ethics: Regarding ethical aspects, I will explain the project to the community and ask them for consent to participate in it and also for the data archiving (a copy of all the data will be given to the community). I will compensate the participants for their hours of work with me with a salary and gifts for them and the community. The local culture, customs and religion will be absolutely respected and will not be interfered. I will also be aware of the effects of the presence and role of the researcher, as well as the minorisation and endangerment situation of the Malieng culture and language. The project meets the SOAS Ethical regulations, which have been approved by the Research Ethics Panel.

[3] Province where Malieng is spoken.

4. Contribution, significance and limitations of the research

It is important to take into consideration that all Vietic languages except for Vietnamese are severely endangered and facing extinction. This situation makes necessary a documentation project to be carried out, not only to record linguistic data but also to preserve the language and the culture. The project is relevant first and foremost for the community, as it may increase the prestige and value of the language to the speakers and non-speakers from the community and raise awareness of language endangerment and language loss. Moreover, the project will create recorded materials for future language support and revitalisation projectsas well as establish ties with the community that will allow for future collaborations and projects.

From the academic point of view, my PhD proposal will provide new linguistic data from an under-documented area. The work and the contacts made there could serve as a basis on further language support and revitalisation programmes to be carried out. My research will cover the gap of widely documenting an under-documented Vietic language in Vietnam with sensitivity towards language endangerment. As Sidwell points out, now is a crucial time for comparative Austroasiatic linguistics (Sidwell, 2009: 3).

I am working on phonology because it is the language dimension I have always enjoyed the most and I have been specialising on. The other reason is because I want to focus and contribute more on tone studies, since tone is a less-studied phenomenon in linguistics. This is also the reason why I started studying Vietnamese three years ago, and why my language skills will be of good use, especially during fieldwork. The project will contribute to linguistic studies on tone, especially tonogenesis, and historical phonology of Austroasiatic languages.

For different reasons, I am also committed to work with endangered languages. Language death is a serious wide-spread phenomenon, with 50-90% of the languages predicted not to survive the XXIst century. As a Catalan speaker, my concern is perhaps more intense, as it affects my own language community. This is why I have been specialising in endangered languages and interested in language documentation and description as well as language support and revitalisation. The project will contribute to understanding language endangerment and loss in indigenous communities, and also on language documentation strategies in minoritised communities.

This research project faces the limitation of the lack of previous data and research on the Malieng community, and Vietic ethnic minorities in general. This means that the previous knowledge prior to fieldwork and data analyses is scarce, and extra effort will be needed to ensure the familiarisation with cultural and linguistic contextual particularities.

This paucity of resources also means that all fieldwork planning is subject to modifications and adaptations, which may affect the later data analysis.

The project also confronts the difficulty of (1) working with an endangered and minoritised community with a reduced number of speakers, (2) the difficulties of carrying out fieldwork in Vietnam, factors that may also condition fieldwork and consequently the data analysis.

5. Research schedule plan

Schedule: The first year of the PhD has been dedicated to the upgrade, mainly theoretical framework and literature review of the topic, as well as research project designing and fieldwork scholarship applications. I have mainly read and written about Vietic languages an communities, Vietnamese and Vietic historical linguistics and phonology and tonogenesis theories. I have also taken the research seminar and the last available course in Vietnamese at SOAS (Vietnamese language and texts), which has helped me improve my Vietnamese language skills. After passing the upgrade I will start planning with detail my research trip to Vietnam during the second year and obtain the required permissions from the Vietnamese authorities.

The second year will be primarily dedicated to data collection. I plan to stay around nine months in Vietnam for the first field trip. For this first trip I will need to prepare for the first term of the second year, as I will go to Vietnam at the start of the dry season, namely spring 2020 (fieldwork is much more difficult during the rainy season). During the first term, I will also consolidate my intermediate-advanced Vietnamese level. Once in Vietnam, I will visit Hanoi before and after the field trip to make contacts with Vietnamese scholars. I will move to the community in Quang Binh and start fieldwork, where I will detail my project and obtain the community consent in order to start the data collection.

I will finish the first field trip during the first term of year 3. I will spend the rest of year 3 archiving and analysing the data and writing up the thesis. I will also go on a second short

field trip (3 months) by the end of year 3 in order to fill in the gaps and check the data. During year 4 I will finalise analysing the data and writing up the thesis and I will submit it by the end of the academic year. The next table summarises the plan for the next 3 years of the PhD.

	Term 1	Term 2	Term 3 (summer)
Year 1	Working on the upgrade Learning Vietnamese	Working on the upgrade Learning Vietnamese	Preparing Fieldwork Vietnamese intensive language course
Year 2	Preparing fieldwork Consolidating Vietnamese	Fieldwork	Fieldwork
Year 3	Data analysis and archiving	Data analysis and archiving	Fieldwork
Year 4	Completing the data archiving Data analysis and writing up	Completing the data analysis Writing up	Completing the writing up and submitting

Table 5: PhD schedule plan.

6. References

Alves, M. 1995. 'Tonal Features and the Development of the Vietnamese Tone.' *Working Papers in Linguistics*: *Department of University of Hawaii at Manoa,* Vol. 27: 1-13.

Alves, M. 2003. 'Ruc and other Vietic languages: Linguistic strands between Vietnamese and the rest of the Mon-Khmer language family', Papers from the Seventh Annual Meeting of the Southeast Asian Linguistics Society, ed. by Karen L. Adams et al. Tempe, Arizona, 3-19. Arizona State University, Program for Southeast Asian Studies.

Alves, M. 2006. 'Linguistic Research on the Origins of the Vietnamese Language: An Overview.' *Journal of Vietnamese Studies*, vol.1, issue.1-2: 104-130.

Aracil, Ll. V. 1983. "Sobre la situació minoritaria", Dir la realitat. Barcelona: Edicions dels Països Catalans: 171-206.

Barker, M.E. 1966, 'Vietnamese-Muong tone correspondences,' in *Studies in Comparative Austroasiatic Linguistics*, ed. N.H. Zide, 9-25. Mouton & Co.

Barker, M. E. 1968. 'Vietnamese and Muong tone correspondences.' *Studies in comparative Austroasiatic linguistics*, N. Zide (ed.), 9-27, The Hague: Mouton.

Barker, M. E. and M. A. Barker 1970. 'Proto-Vietnamuong (Annamuong) final consonants and vowels.' *Lingua* 24.3: 268-285.

Blench, Roger. 2008. 'Stratification in the peopling of China: how far does the linguistic evidence match genetics and archaeology?' In *Human migrations in continental East Asia and Taiwan. Matching archaeology, linguistics and genetics*. Sanchez-Mazas, A., Blench, R.M., Ross, M.D., I. Peiros and M. Lin (eds). 105-132. London: Routledge.

Bordeaux, P. 2016. 'The French-speaking Reformed Church, a Neglected Source of Protestantism in Vietnam: Observations on the Origins of a Minority Religion.' Vietnam's Ethnic and Religious Minorities. A Historical Perspective. Engelbert, T. (ed.) 2016. Frankfurt am Main: Peter Lang.

Chamberlain, J. R. 1998, 'The origin of Sek: implications for Tai and Vietnamese history', *The International Conference on Tai Studies*, ed. S. Burusphat, Bangkok, Thailand,

pp. 97-128. Institute of Language and Culture for Rural Development, Mahidol University.

Chamberlain, J. R. 2003. 'Eco-Spatial History: a nomad myth from the Annamites and its relevance for biodiversity conservation,' in *Landscapes of Diversity: Proceedings of the III MMSEA Conference,* X. Jianchu and S. Mikesell, eds. , 25–28 August 2002. Lijiand, P. R. China: Center for Biodiversity and Indigenous Knowledge. pp. 421–436.

Chamberlain, J. R. 2018. 'A Kri-Mol (Vietic) Bestiary: Prolegomena to the Study of Ethnozoology in the Northern Annamites.' *Kyoto Working Papers on Area Studies* No. 133. Kyoto: Kyoto University.

Chazée, L. 1999. *The peoples of Laos: Rural and Ethnic Diversities*. Bangkok: White Lotus.

Đặng Nghiêm Vạn, Chu Thái Sơn and Lưu Hùng. 2000. *Ethnic minorities in Vietnam*. Hanoi: Thế Giới Publishers.

Diffloth, G. 1982. 'Registres, dévoisement, timbres vocaliques: leur histoire en Katouique.' Mon-Khmer Studies XI: 47-82.

Diffloth, G. 1989. 'Proto-Austroasiatic creaky voice.' *Mon-Khmer Studies* 15: 139-54.

Diffloth, G. 1990. 'Vietnamese tonogenesis and new data on the registers of Thavung.' *Sino-Tibetan Conference*. University of Texas at Arlington.

Diffloth, G. 2005. The contribution of linguistic palaeontology to the homeland of Austroasiatic. In *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics,* edited by L. Sagart, R. Blench and A. Sanchez-Mazas: Routledge/Curzon.

Enfield, Nick J. & Gérard Diffloth. 2009. 'Phonology and sketch grammar of Kri, a Vietic language of Laos.' *Cahiers de linguistique - Asie Orientale* 38: 3-69.

Engelbert, T. (ed.) 2016. *Vietnam's Ethnic and Religious Minorities. A Historical Perspective*. Frankfurt am Main: Peter Lang.

Engelbert, T. 2016. ‘Catholic Missionaries in the Indochinese Highlands: The First Period (1850-1893).’ Vietnam’s Ethnic and Religious Minorities. A Historical Perspective. Engelbert, T. (ed.) 2016. Frankfurt am Main: Peter Lang.

Ferlus, M. 1974. ‘Problèmes de mutations consonantiques en thavung.’ BSLP; LXIX.1: 311-23.

Ferlus, M. 1979. ‘Lexique thavung-français.’ *Cahiers de Linguistique, Asie Orientale* 5: 71-94.

Ferlus, M. 1982. ‘Spirantisation des obstruantes médiales et formation du systèm consonantique du vietnamien.’ *Cahiers de Linguistique, Asie Orientale* 11.1: 83-106.

Ferlus, M. 1989-90. ‘Sur l'origine géographique des langues Viet-muong.’ *Mon-Khmer Studies* 18-19: 52-59.

Ferlus, M. 1992. ‘Histoire abrégée de l'évolution des consonnes initiales du vietnamien et du sino-vietnamien.’ *Mon-Khmer Studies* 20: 111-125.

Ferlus, M. 1994. ‘Quelques particularités du cuôi cham, une langue viet-muong du Nghê-an (Vietnam).’ *Neuvièmes Journées de Linguistique de l’Asie Orientale,* May 1994, Paris.

Ferlus, M. 1996. ‘Langues et peuples viet-muong.’ *Mon-Khmer Studies* 26: 7-28.

Ferlus, M. 1997. ‘Le maleng brô et le vietnamien,’ *Mon-Khmer Studies* 27: 55-66.

Ferlus, M. 1998. ‘Les systèmes de tons dans les langues viet-muong.’ *Diachronica* 15.1: 1-27.

Ferlus, M. 2001. ‘The Origin of Tones in Viet-Muong.’ *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistic Society*, 2001: 297-313.

Ferlus, M. 2004. ‘The Origin of Tones in Viet-Muong.’ In *Papers from the Eleventh Annual Conference of the Southeast Asian Linguistics Society 2001,* Somsonge Burusphat (ed.), 297–313. Tempe, Arizona: Arizona State University Programme for Southeast Asian Studies Monograph Series Press.

Ferlus, M. 2014. ‘Arem, a Vietic Language.’ Mon-Khmer Studies 43.1:1-15 (ICAAL5 special issue).

Gurr, T. 1993. Minorities at Risk: A Global View of Ethnopolitical Conflict. Washington DC: Institute of Peace Press.

Haudricourt, A. 1954. ‘De l'origine des tons en viêtnamien.’ *Journal Asiatique* 242: 69-82.

Haudricourt, A. 1961. ‘Bipartition et Tripartition des Systèmes de Tons dans quelques Langues d'Extrême Orient.’ *Bulletin de la Société de Linguistique de Paris* 56: 163-80.

Hayes, LV. 1982. ‘The mutation of *r in pre-Thavung.’ *Mon Khmer Studies 1* (1): 83-100.

Hayes, LV. 1992. ‘Vietic and Viet-Muong: a new subgrouping in Mon-Khmer.’ *The Mon-Khmer Studies Journal*, 21: 211-228.

Honda, K. 2005. ‘Tone correspondences and tonogenesis in Vietic,’ In *SEALSXV: papers from the 15th meeting of the Southeast Asian Linguistics Society*, Paul Sidwell (ed.), 177-191, Canberra, Pacific Linguistic, 2005.

Kasuga, A. 1994. *A Study on Arem Phonology*. Master’s thesis submitted to Tokyo University.

Keating, M. 2001. ‘So many nations, so few states: Territory and nationalism in the global era’. Multinational Democracies, A. Gagnon & J. Tully (eds.): 39-64. Cambridge: Cambridge University Press.

Kirby, J. 2011. ‘Vietnamese (Hanoi Vietnamese). Illustration of the IPA’. *Journal of the International Phonetic Association* (2011) 41/3: 381–392.

Kymlicka, Will. 1995. Multicultural Citizenship: A Liberal Theory of Minority Rights. Oxford: Oxford University Press, 1995.

Liên, T. T. 2016. ‘The Vietnamese Catholics, Nation State and National Identity: Some Reflections on the State of a Religious Minority.’ Vietnam’s Ethnic and Religious Minorities. A Historical Perspective. Engelbert, T. (ed.) 2016. Frankfurt am Main: Peter Lang.

Markefka, M. 1990. Vorurteile, Minderheiten, Diskriminierung. Ein Beiträge zun Verständnis sozialer Gegensätze. Neuwied: Luchterhand (6. Auflage), 1990.

Maspero, Henri. 1912. ‘Étude sur la phonétique de la langue annamite : Les initiales.’ *Bulletin de l'Ecole Française de l'Extrême-Orient* 12(1): 1-127.

Matisoff, J. 1970. ‘Glottal dissimilation and the Lahu high-rising tone: A tonogenetic case-study.’ *Journal of the American Oriental Society*. 90 (1): 13–44.

Matisoff, J. 1973. ‘Tonogenesis in Southeast Asia,’ in *Consonant Types and Tone*, Larry M. Hyman, (ed.), 71-95. Southern California Occasional Papers in Linguistics, No. 1. Los Angeles: UCLA.

Matras, J., & Ferlus, M. 1971. ‘Les langues austroasiatiques.’ *Asie du Sud-Est et Monde Insulindien*, 2.4: 53-93.

McElwee, P. 2004. ‘Becoming Socialist or Becoming Kinh? Government Policies for Ethnic Minorities in the Socialist Republic of Vietnam,’ in *Civilizing the Margins: Southeast Asian Government Policies for the Development of Minorities*, ed. Christopher R. Duncan, 182-213, Ithaca, NY: Cornell University Press.

Nguyễn M-C. 2016. ‘The tone system of Kim Thượng Mường: an experimental study of fundamental frequency, duration, and phonation types.’ *Linguistics*. HAL, archives-ouvertes.

Nguyễn V. T. 1982. *Ngữ âm tiếng Mường qua các phương ngôn* [Phonetics of the Mường language through its dialects]. PhD thesis. Hanoi: University of Hanoi.

Nietschmann, B. 1987. ‘The Third World War.’ *Cultural Survival Quarterly*, number 11(3): 1-16.

Ottersbach, M. 1997. *Gesellschaftliche Konstruktion von Minderheiten*. Opladen: Leske/Budrich.

Parkin, R. 1991. *A Guide to Austroasiatic Speakers and Their Languages*. Oceanic Linguistics Special Publication, n.23. Honolulu: University of Hawaii Press.

Peiros, I. J. 2004. *Geneticeskaja klassifikacija avstroaziatskix jazykov*: Moskva: Rossijskij gosudarstvennyj gumanitarnyj universitet (doktorskaja dissertacija).

Schliesinger, J. 1997. *Hill Tribes of Vietnam. Introduction and Overview*, Bangkok: White Lotus, Volume 1.

Sidwell, P. 2009. *Classifying the Austroasiatic languages*: history and state of the art. Munich, Lincom Europa.

Sidwell, P. 2015. 'Austroasiatic classification,' in *The Handbook of Austroasiatic Languages*, Jenny, M and Sidwell, P. (eds.), 144-220, Leiden: Brill.

Solntsev, V. 1996. 'Some remarks on the Ruc language.' Mon-Khmer Studies, No. 26: 29-32.

Spolsky, B and Hult, F. M. 2010. *The Handbook of Educational Linguistics*. John Wiley & Sons.

Statistical Publishing House (SPH). 2010. The 2009 Vietnam Population and Housing Census: Complete Results. Hanoi: Central Population and Housing Census Steering Committee.

Suwilai P. 1996. So (Thavung) preliminary Dictionary. Bangkok: Institute of Language and Culture for Rural Development, Mahidol University, Thailand; Melbourne: Department of Linguistics and Applied Linguistics, University of Melbourne, Australia.

Tan, S. and Walker, A. 2008. 'Beyond hills and plains: rethinking ethnic relations in Vietnam and Thailand,' in *Minorities at Large: New Approaches to Minority Ethnicity in Vietnam*, Philip Taylor (ed.), 117-157, Singapore: ISEAS Publishing.

Taylor, P. 2008, 'Minorities at Large: New Approaches to Minority Ethnicity in Vietnam', in *Minorities at Large: New Approaches to Minority Ethnicity in Vietnam*, Philip Taylor (ed.), 3-43, Singapore: ISEAS Publishing.

Thompson, L. C. 1976. 'Proto-Viet-Muong Phonology.' *Austroasiatic Studies* 1976b: 1113-1204.

Thurgood, G. 2002. 'Vietnamese and tonogenesis: Revising the model and the analysis.' *Diachronica*, Volume 19, Issue 2: 333-363.

Trần T. D. 1996. 'Les initiales */s, z/ et /h/ du Proto Việt-Mường (PVM) et leurs changements dans le Vietnamien.' *Mon-Khmer Studies* 25: 263-268.

Trần Thi T. T. 2004. ‘Local Politics and Democracy in a Muong Ethnic Community,’ in *Beyond Hanoi: Local Government in Vietnam*, eds. Benedict J. Tria Kerkvliet and David Marr, 137-166, Singapore: ISEAS.

Wilson, R.S. 1966, ‘A comparison of Muong with some Mon-Khmer languages,’ in *Studies in Comparative Austroasiatic Linguistics*, ed. N.H. Zide, 203-213. Mouton & Co.

Windischer, R. 2009. Old and New Minorities: Reconciling Diversity and Cohesion. Baden-Baden: Nomos.

7. Appendices

7.1 Ethnic minorities 1979 classification table

Group	People	Population (2009 Census)	Distribution (2009 Census)
	Total	**85,846,997**	
A. **Vietic**	1.Kinh/Viet	73,594,427	Throughout Vietnam
	2.Chứt	6,022	Quảng Bình (5,095 persons, comprising 84.6% of all Chứt in Vietnam)
	3.Mường	1,268,963	Hòa Bình (479,197 persons, comprising 63.3% of the province's population), Thanh Hóa (328,744 persons, comprising 9.5% of the province's population), Phú Thọ (165,748 persons, comprising 13.1% of the province's population), Sơn La (71,906 persons, comprising 8.2% of the province's population), Ninh Bình (46,539 persons)
	4.Thổ	74,458	Nghệ An (59,579 persons, comprising 80.0% of all Thổ in Vietnam), Thanh Hóa (9,652 persons, comprising 13.0% of all Thổ in Vietnam)
B. **Tai–Kadai**	5.Bố Y	2,273	Lào Cai (1,398 persons, comprising 61.5% of all Bố Y in Vietnam), Hà Giang (808 persons, comprising 35.5% of all Bố Y in Vietnam)
	6.Giáy	58,617	Lào Cai (28,606 persons, comprising 48.8% of all Giáy in Vietnam), Hà Giang (15,157 persons, comprising 25.9% of all Giáy in Vietnam), Lai Châu (11,334 persons), Yên Bái (2,329 persons)

Group	People	Population (2009 Census)	Distribution (2009 Census)
	7.Lao	14,928	Lai Châu (5,760 persons, comprising 38.6% of all Lào in Vietnam), Điện Biên (4,564 persons, comprising 30.6% of all Lào in Vietnam), Sơn La (3,380 persons, comprising 22.6% of all Lào in Vietnam)
	8.Lự	5,601	Lai Châu (5,487 persons, comprising 98.0% of all Lự in Vietnam)
	9.Nùng	968,800	Lạng Sơn (314,295 persons, comprising 42.9% of the province's population and 32.4% of all Nùng in Vietnam), Cao Bằng (157.607 persons, comprising 31.1% of the province's population and 16.3% of all Nùng in Vietnam), Bắc Giang (76,354 persons)
	10.Sán Chay	169,410	Tuyên Quang (61,343 persons, comprising 36.2% of all Sán Chay in Vietnam), Thái Nguyên (32,483 persons, comprising 19.2% of all Sán Chay in Vietnam), Bắc Giang (25,821 persons),
	11.Tày	1,626,392	northern Vietnam
	12.Thái	1,550,423	northern Vietnam
C. **Kadai** (Kra)	13.Cờ Lao	2,636	Hà Giang (2,301 persons, comprising 87.3% of all Cờ Lao in Vietnam)
	14.La Chí	13,158	Hà Giang (12,072 persons, comprising 91.7% of all La Chí in Vietnam), Lào Cai (619 persons), Tuyên Quang (100 persons)
	15.La Ha	8,177	Sơn La (8,107 persons, comprising 99.1% of

Group	People	Population (2009 Census)	Distribution (2009 Census)
			all La Ha in Vietnam)
	16.Pu Péo	687	Hà Giang (580 persons, comprising 84.4% of all Pu Péo in Vietnam), Tuyên Quang (48 persons)
D. **Austroasiatic**	17.Ba Na	227,716	Gia Lai (150,416 persons, comprising 11.8% of the province's population and 66.1% of all Ba Na in Vietnam), Kon Tum (53,997 persons, comprising 12.5% of the province's population and 23.7% of all Ba Na in Vietnam), Phú Yên (4,145 persons, comprising 12.5% of the province's population and 23.7% of all Ba Na in Vietnam)
	18.Brâu	397	Kon Tum (379 persons, comprising 95.5% of all Brâu in Vietnam)
	19.Bru	74,506	Quảng Trị (55,079 persons, comprising 73.9% of all Bru-Vân Kiều in Vietnam), Quảng Bình (14,631 persons, comprising 19.6% of all Bru-Vân Kiều in Vietnam), Đăk Lăk (3,348 persons)
	20.Chơ Ro	26,855	Đồng Nai (15,174 persons, comprising 56.5% of all Chơ Ro in Vietnam), Bà Rịa-Vũng Tàu (7,632 persons), Bình Thuận (3,375 persons)
	21.Co	33,817	Quảng Ngãi (28,110 persons, comprising 83.1% of all Co), Quảng Nam (5,361 persons)
	22.Cờ Ho	166,112	Lâm Đồng (145,665 persons, comprising 12.3% of the province's population and

Group	People	Population (2009 Census)	Distribution (2009 Census)
			87.7% of all Cơ Ho in Vietnam)
	23.Cơ Tu	61,588	Quảng Nam (45,715 persons, comprising 74.2% of all Cơ Tu in Vietnam), Thừa Thiên-Huế (14,629 persons, comprising 23.8% of all Cơ Tu in Vietnam)
	24.Giẻ Triêng	50,962	Kon Tum (32,644 persons, comprising 62.1% of all Giẻ Triêng in Vietnam), Quảng Nam (19,007 persons, comprising 37.3% of all Giẻ Triêng in Vietnam)
	25.Hrê	127,420	Quảng Ngãi (115,268 persons, comprising 90.5% of all Hrê in Vietnam)
	26.Kháng	13,840	Sơn La (8,582 persons, comprising 62.0% of all Kháng in Vietnam), Điện Biên (4,220 persons, comprising 30.5% of all Kháng in Vietnam)
	27.Khmer Krom	1,260,640	Sóc Trăng (397,014 persons, comprising 30.7% of the province's population and 31.5% of all Khmer in Vietnam), Trà Vinh (317,203 persons, comprising 31.6% of the province's population and 25.2% of all Khmer in Vietnam), Kiên Giang (210,899 persons, comprising 12.5% of the province's population and 16.7% of all Khmer in Vietnam), An Giang (90,271 persons), Bạc Liêu (70,667 persons), Cà Mau (29,845 persons) each comprising less than 10% of all Khmer in Vietnam)
	28.Khơ Mú	72,929	Nghệ An (35,670 persons, comprising 48.9% of all Khơ Mú in Vietnam), Điện Biên (16,200 persons, comprising 22.2% of all Khơ Mú in Vietnam), Sơn La (12,576

Group	People	Population (2009 Census)	Distribution (2009 Census)
			persons), Lai Châu (6,102 persons)
	29.Mạ	41,405	Lâm Đồng (31,869 persons, comprising 77.0% of all Mạ in Vietnam), Đắk Nông (6,456 persons), Đồng Nai (2,436 persons)
	30.Mảng	3,700	Lai Châu (3,631 persons, comprising 98.1% of all Mảng in Vietnam)
	31.Mnông	102,741	Đăk Lăk (40,344 persons, comprising 39.3% of all M'Nông in Vietnam), Đăk Nông (39,964 persons, comprising 38.9% of all M'Nông in Vietnam)
	32.Ơ Đu	376	Nghệ An (340 persons, comprising 90.4% of all Ơ Đu in Vietnam)
	33.Rơ Măm	436	Kon Tum (419 persons, comprising 96.1% of all Rơ Măm in Vietnam)
	34.Tà Ôi	43,886	Thừa Thiên-Huế (29,558 persons, comprising 67.4% of all Tà Ôi in Vietnam), Quảng Trị (13,961 persons, comprising 31.8% of all Tà Ôi in Vietnam)
	35.Xinh Mun	23,278	Sơn La (21,288 persons, comprising 91.5% of all Xinh Mun in Vietnam), Điện Biên (1,926 persons)
	36.Xơ Đăng	169,501	Kon Tum (104,759 persons, comprising 24.4% of the province's population and 61,8% of all Xơ Đăng in Vietnam), Quảng Nam (37,900 persons, comprising 22.4% of all Xơ Đăng in Vietnam), Quảng Ngãi (17,713 persons)

Group	People	Population (2009 Census)	Distribution (2009 Census)
	37.X'Tiêng	85,436	Bình Phước (81,708 persons, comprising 95.6% of all Xtiêng in Vietnam)
5. **Hmong–Mien**	38.Dao	751,067	northern Vietnam
	39.Hmong	1,068,189	northern Vietnam
	40.Pà Thẻn	6,811	Hà Giang (5,771 persons, comprising 84.7% of all Pà Thẻn in Vietnam), Tuyên Quang (877 persons)
6. **Malayo-Polynesian**	41.Chăm	161,729	Ninh Thuận (67,274 persons, comprising 41.6% of all Chăm in Vietnam), Bình Thuận (34,690 persons, comprising 21.4% of all Chăm in Vietnam), Phú Yên (19,945 persons), An Giang (14,209 persons)
	42.Chu Ru	19,314	Lâm Đồng (18,631 persons, comprising 96.5% of all Chu Ru in Vietnam)
	43.Ê Đê	331,194	Đăk Lăk (298,534 persons, comprising 17.2% of the province's population and 90.1% of all Ê Đê in Vietnam), Phú Yên (20,905 persons)
	44.Gia Rai	411,275	Gia Lai (372,302 persons, comprising 29.2% of the province's population and 90.5% of all Jrai in Vietnam), ngoài ra còn có ở Kon Tum (20,606 persons), Đắk Lắk (16,129 persons)
	45.Ra Glai	122,245	Ninh Thuận (58,911 persons, comprising 48.2% of all Raglai in Vietnam), Khánh Hòa (45,915 persons, comprising 37.6% of all Raglai in Vietnam), Bình Thuận (15,440

Group	People	Population (2009 Census)	Distribution (2009 Census)
			persons)
7. **Chinese**	46.Hoa	823,071	Hồ Chí Minh City (414,045 persons, comprising 50.3% of all Hoa in Vietnam), Đồng Nai (95,162 persons), Sóc Trăng (64,910 persons), Kiên Giang (29,850 persons), Bạc Liêu (20,082 persons), Bình Dương (18,783 persons), Bắc Giang (18,539 persons)
	47.Ngái	1,035	Thái Nguyên (495 persons, comprising 47.8% of all Ngái in Vietnam), Bình Thuận (157 persons, comprising 15.2% of all Ngái in Vietnam)
	48.Sán Dìu	146,821	Thái Nguyên (44,131 persons, comprising 30.1% of all Sán Dìu in Vietnam), Vĩnh Phúc (36,821 persons, comprising 25.1% of all Sán Dìu in Vietnam), Bắc Giang (27,283 persons), Quảng Ninh (17,946 persons), Tuyên Quang (12,565 persons)
8. **Tibeto-Burman**	49Phunoi	2,029	Lai Châu (1,134 persons, comprising 55.9% of all Cống in Vietnam), Điện Biên (871 persons, comprising 42.9% of all Cống in Vietnam)
	50.Hà Nhì	21,725	Lai Châu (13,752 persons, comprising 63.3% of all Hà Nhì in Vietnam), Lào Cai (4,026 persons), Điện Biên (3,786 persons)
	51.La Hủ	9,651	Lai Châu (9,600 persons, comprising 99.5% of all La Hủ in Vietnam)
	52.Lô Lô	4,541	Cao Bằng (2,373 persons, comprising 52.3% of all Lô Lô in Vietnam), Hà Giang (1,426

Group	People	Population (2009 Census)	Distribution (2009 Census)
			persons), Lai Châu (617 persons)
	53.Phù Lá	10,944	Lào Cai (8,926 persons, comprising 81.6% of all Phù Lá in Vietnam), Yên Bái (942 persons), Hà Giang (785 persons), Điện Biên (206 persons)
	54.Si La	709	Lai Châu (530 persons, comprising 74.8% of all Si La in Vietnam), Điện Biên (148 persons, comprising 20.9% of all Si La in Vietnam)

Table 1: Vietnam's 1979 minority classification with crossed data from the national census (SPH, 2010).

7.2 Vietic historical phonology

7.2.1 Vietic phonological systems

In this section I will outline the main phonological characteristics of each of the currently described Vietic languages, with some basic information their speakers and other aspects of the language. I am treating each of the described languages separately and not grouping them into sub-groups. For instance, the Chứt are a group of languages classified by most scholars under the word 'Chứt' (mountain dweller) (see section 2.4.2: Vietic historical classification). The languages of this group include Rục, Sách, Arem, Mã Liềng, Chứt or Mày. Each of the languages/varieties has its own section, as they have been described as separate entities.

Vietnamese:

Vietnamese was first known to the West via missionaries and trade in the XVIth century under the name Annamite, denomination which persisted until the end of the French colonisation in the mid XXth century. Annam–'Pacified South' in Sino-Vietnamese–was the name of a Chinese province and then name of one of the French protectorate. Due to the fact

that the North of Vietnam was more than 1000 years under the control of the Chinese dynasties, the Vietnamese language has major influences from the Sinitic family. In fact, some authors defend Vietnamese was a coastal creole originated in the South of China–North of Vietnam region (Chamberlain, 2003; Sidwell, 2009). These influences, together with other Southeast Asian areal linguistic features, made the classification of Vietnamese a controversial issue (see section 2.2.3: Classification of Vietnamese).

The first phonological study of Vietnamese was that of Alexandre de Rhodes, who in 1651 published a Annamite-Portuguese-Latin dictionary with comments and developed the quốc ngữ (national writing system), which is now in use in Vietnam with minor changes from his proposal (Engelbert, 2016: 99-100). We have to wait until the XXth century to see the next important linguistic descriptions of Vietnamese, until Maspero published his major work *Étude sur la phonétique de la langue annamite: Les initials* in 1912.

Nowadays, Vietnamese is the only Vietic official language in the world and it is spoken by more than 70 million people in Vietnam, where it acts as a *lingua franca*. Its linguistic description has been developed the past decades, especially regarding the tonal system (Kirby, 2011: 381).

I have taken the phonological data on Vietnamese from the Journal of the International Phonetic Association. The author, James Kirby, summarises a description of the Standard Hanoi Vietnamese phonology from a number of studies. Hanoi Vietnamese preserves all six tones but has lost consonant distinctions. The other varieties have less tones and preserve consonant distinctions lost in Hanoi Vietnamese, although at the same time have lost others. The next table offers a picture of the Hanoi Vietnamese initials:

	Labial	Labio-dental	Dental	Alveolar	Palatal	Velar	Glottal
Plosive	ɓ		t t^h	ɗ	tɕ	k	ʔ
Nasal	m		n		ɲ	ŋ	
Fricative		f v		s z		x ɣ	h
Approximant	w						
Lateral approximant			l				

Figure 1: Hanoi Vietnamese initials (Kirby, 2011: 382).

Hanoi Vietnamese has 8 finals: three unreleased voiceless obstruents /p t k/ ([p^ t^ k^]), three nasals /m n N/, and two approximants /j w/ (Kirby, 2011: 383). In the next F1 and F2

schemes, single vowels are represented on the left and diphthongs on the right. The vowel /a/ can be short or long:

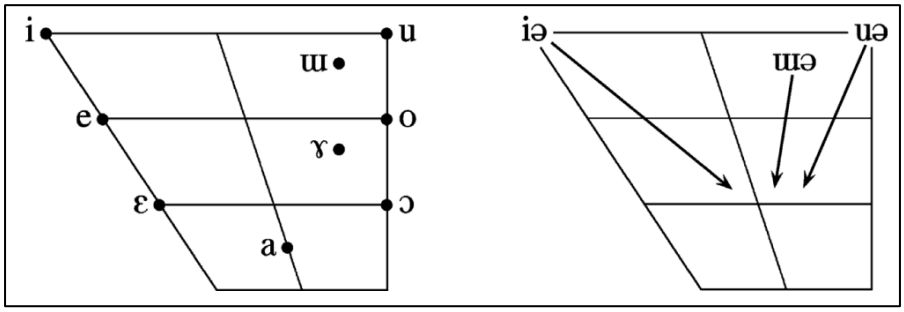


Figure 2: Hanoi Vietnamese vocalic system (Kirby, 2011: 384).

Vietnamese has undergone the well-described tonogenesis process (see section 7.2.3: Vietic Tonogenesis), resulting in a 6-way tonal system divided up into 2 series: high and low. The three tones from the high series and the three tones from the low series can appear in open syllables or closed in nasals (columns A, B and C), but the syllables closed in obstruents only present two possible tones (column D). The tones are represented in the next table adapted from Ferlus (1982: 103) with both the Vietnamese orthography and the Vietnamese names for each tone. In the following diagram, tone frequencies for each of the tones are given:

	A	**B**	**C**	**D**
High	a *ngang*	á *sắc*	ả *hỏi*	á *sắc*
Low	à *huyền*	ạ *nặng*	ã *ngã*	ạ *nặng*

Table 2: Northern Vietnamese tonal system

Below

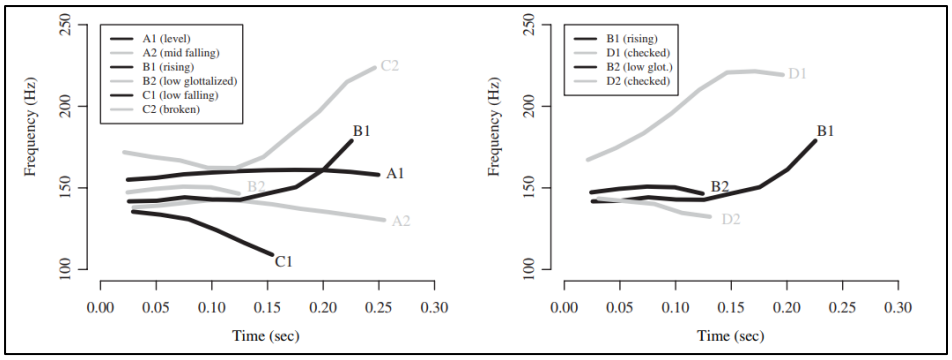


Figure 3: Hanoi Vietnamese tone frequencies (Kirby, 2011: 384).

Mường:

Mường is an official long-established ethnic minority from the North of Vietnam. In fact, Mường represents a group of language varieties very close to Northern Vietnamese. Along with the Việt/Kinh, they are the only Vietic groups with a significant number of speakers (more than 1 million for Mường, the second largest ethnic minority after the Tày) and more than 70 million for Vietnamese (SPH, 2010). The term 'Mường' has a Tai origin, meaning 'principality', and it is used for Vietic as well as for Tai groups in both Laos and Vietnam (Nguyễn, 2015: 10).

The following table, adapted from Ferlus, 1992, shows the Mường tonal system. The same picture is presented by Barker, 1966; Wilson, 1966 and Nguyễn, 2015. The tonal system has undergone the same tonogenesis process as Modern Vietnamese, although it has collapsed the correspondent *ngã* and *nặng* Vietnamese tones. The table is organised in 4 categories: A, B and C for open syllables or closed in nasal or sonorant, and D for closed syllables in a voiceless stop:

	A	**B**	**C**	**D**
High	a	a˝	aˀ	at˝
Low	a`	a.		at.

Table 3: Mường tonal system, adapted from Ferlus, 1982: 103.

According to Wilson (1966: 210-211), the initials in Mường are the following (presented in *quốc ngữ* or Vietnamese ortography): *b, ch, d, h, k, l, m, n, ng, p, r, s, t, th, tl*. On the other hand, the finals are: *k, l, m, n, ng, p, t*.

The following tables, adapted from Nguyễn (1982), present the phonemic system of Mường. Nguyễn's study takes into consideration different varieties of the language. As for the vowels, note that /ɤ/ and /a/ can be long or short. The diphthongs described are /iə, ɯə, uə/.

m /m/	*n* /n/	*nh* /ɲ/	*ng* /ŋ/	
p /p/	*t* /t/	*ch* /c/	*c* /k/	
ph /pʰ/	*th* /tʰ/		*kh* /kʰ/	

b /ɓ/	*đ* /ɗ/		*g* /ɡ ~ ɣ/	
	x /s/			*h* /h/
v/*w*/*o*/*u* /β/	*d*/*gi*/*i*/*y* /z ~ j/			
	l, *tl* /l, tl ~ kl/			

Table 4: Mường consonantic system, adapted from Nguyễn, 1982: 9.

i /i/	*ư* /ɯ/	*u* /u/
ê /e/	*ơ*, *â* /ɤ, ɤ̆/	*ô* /o/
e /ɛ/	*a*, *ă* /a, ă/	*o* /ɔ/

Table 5: Mường vocalic system, adapted from Nguyễn, 1982: 10.

The phonological evolution of the different Mường varieties has been to a larger extent in parallel to Vietnamese but has at the same time resulted in a more conservative closer in some respects to proto-Northern Vietic (Nguyễn, 2015: 13).

Arem:

One of the most well-documented Vietic languages is Arem or Cmbrau, which according to Ferlus is undoubtedly a Viet-Muong language. It has around a hundred speakers and these speakers are multilingual: they know both dialectal and standard Vietnamese and some also know Khùa and/or Lao. It is considered endangered (Ferlus, 2014: 1-2). Its phonology has been studied by Ferlus (2014) and Kasuga (1994). There is a lexicon which has been created by the same authors.

According to Ferlus (2014: 3), Arem vowels can be classified into two series, 1 and 2, the first including those vowels with high-clear register phonation and the second including those with a low-breathy register phonation. These series can be further divided up in two groups, (a) syllables ending with voiced finals and (b) syllables ending with voiceless finals or (–h), which check the syllable. Group (a) shows a contrast between a (aa) modal tone and (ab) a glottalised tone. Finally, the combination of the contrast /clear~breathy/ and /unchecked~glottal/ forms a four-tone system, as shown in the next table (adapted from ibid. 3):

tone a	high, clear, unmarked (corresponds to MV tone ngang)
tone aˀ	high, glottalized, slightly raising (corresponds to MV tone sắc)
tone à	low, breathy (corresponds to MV tone huyền)
tone àˀ	low breathy glottalized (corresponds to MV tone nặng)

Table 6: Arem tonal system, adapted from Ferlus, 2014: 3.

The following tables (taken from ibid. 3) compare Arem's tonal system with the one of Vietnamese:

Arem	*voiced finals* (# w j l m n ɲ ŋ)		*unvoiced finals* (h) / (p t c k)	
series 1 (hight-clear)	a	aˀ	ah	aC
series 2 (low-breathy)	à	àˀ	àh	àC

Vietnamese	*voiced finals*			*unvoiced finals*
series 1 (*tones a á ả*)	*ngang*	*sắc*	*hỏi*	*sắc*
series 2 (*tones à ạ ã*)	*huyền*	*nặng*	*ngã*	*nặng*

Figure 4: Arem tonal system, adapted from Ferlus, 2014: 3.

The following tables (taken from ibid. 4-5) list Arem's phonemic inventory:

Initial consonants

Simple initials

p^h	t^h			k^h	
p	t	$tʃ$	c	k	$ʔ$
b	d		$ɟ$		
m	n		$ɲ$	$ŋ$	
	s	$ʃ$			h
v	(z)	$(ʒ/ʑ)$	j		
	l	r			

Clusters with r / l

pr	tr	cr	kr	
br				
pl	tl		kl	
	t^hl			

Pre-nasalized initials

	^nth				^ŋh
^mp	^nt	$^ntʃ$	^ɲc	^ŋk	$^ŋʔ$
^mb	nd		$^ɲɟ$		
^nv	^nl	$^nr/^ndʒ$	^ɲj		

Figure 5: Arem monosyllables: initial consonants (Ferlus, 2014: 4).

Finals

					h
p	t		c	k	$\#(ʔ)$
$m(ʔ)$	$n(ʔ)$		$ɲ(ʔ)$	$ŋ(ʔ)$	
$w(ʔ)$		$l(ʔ)$	$j(ʔ)$		

Figure 6: Arem monosyllables: finals consonants (Ferlus, 2014: 4).

Pre-syllables

Note: infrequent pre-syllables are placed in the brackets:

a- *i-* *u-*

pa- (*pi-*) (*ba-*) (*ma-* / *mu-*)

ta- (*tam*) N- (first segment of pre-nasalized)

ci- (*cm*)

ka- *ku-* (*kə-*)

la- (le-) (*ra-* / *rə-*)

Pre-syllables in decreasing order, with the number of occurrences in the lexicon:

a- (240), N- (141), ka- (83), u- (79), ci- (54), pa- (44), ta- (36), i- (30), la- (30), ku- (14).

Figure 7: Arem pre-syllables (Ferlus, 2014: 5).

Vowel system

breathy vowels						*clear vowels*					
ì	*ì:*	*i̤*	*i̤:*	*ù*	*ù:*	*i*	*(i:)*	*ɨ*	*ɨ:*	*u*	*u:*
						ɪ	*ɪ:*	*ʉ*	*ʉ:*	*ʊ*	*ʊ:*
	ìe		*i̤ə*		*ùo*		*ie*		*ɨə*		*uo*
è	*(è:)*	*(ə̀)*	*ə̀:*	*ò*	*ò:*	*(e)*	*e:*	*ə*	*ə:*	*o*	*o:*
		v̤	*v̤:*			*ɛ*	*ɛ:*	*(ɐ)*	*ɐ:*	*ɔ*	*ɔ:*
						æ	*æ:*	*ʌ*	*ʌ:*	*ɑ*	*ɑ:*
								a	*a:*		

Figure 8: Arem vocalic system (Ferlus, 2014: 5).

Ferlus' 2014 paper also studies Arem in contrast with Proto-Viet-Muong (PVM henceforth) and Modern Vietnamese. He describes several correspondences in consonants, vowels and tones. Arem appears to be more conservative than Vietnamese when comparing both with PVM. The phonological description of Arem contributes to justifying the reconstructed PVM and to explaining the tonogenesis of both Arem and Vietnamese tonal systems. Ferlus' interest in Arem seems to be in relation to its contribution to the reconstruction of PVM. The author explains the *heterogeneity* of the language (i.e. Arem has borrowings from modern standard Vietnamese, a local dialect yet to be described and also Katuic languages, Lao and Cham) as the main problematic facing the reconstruction of PVM. The author also encourages further documentation of this language (ibid. 14).

Aheu:

Another Vietic language that has been studied is Aheu, So or Thavung (Aheu henceforth), studied by Ferlus (1979), Hayes (1982) and Suwilai (1996). Aheu is spoken in Laos (some one hundred speakers, Ferlus, 1979: 71) and Thailand, (around 1000 speakers; Suwilai, 1996: 163). Its speakers are shifting to Laos and Thai in Thailand, where they form a multilingual community in an area with high language contact, especially Tai languages. Therefore, the So people also know Laos and Nyoh, and the youngest also central Thai (ibid. 163-164). I will use 'So' to refer to the variety spoken in Thailand and 'Thavung' for the one in Laos following Ferlus' and Suwilai's criteria, although none of the authors specify where they did the documentation, so the location of the different varieties remains unclear.

Suwilai identifies three main So varieties: on one hand there is the continuum between "big So", spoken by elderly people; and "small So", spoken by younger people and used by elderly and middle-age people when talking to the young members of the community, and is heavily influenced by Laos and Thai. On the other hand, there is also a "mixed So-Nyoh" variety (ibid. 1996: 165).

Suwilai has focused on "big So" to describe So's phonology. Its syllable structure can be more complex than other Viet-Muong languages, having words with up to three syllables: (pre-syllable)+(pre-syllable)+monosyllable. The syllabic structure can be also expressed this way: [CV(C)]+[CV]+[CV(C)] for trisyllabic words and as [CV(C)]+[CV(C)] for disyllabic ones (ibid. 166). The underlined segments represent the suprasegments 'stress' and 'tone'. Thavung is clearly more monosyllabic than So (ibid. 168).

The following tables taken from Suwilai, 1996 (168-172) show So's phonemic inventory:

		bilabial	alveolar	palatal	velar	glottal
stops	vl unaspirated	p	t	c[tɕ]	k	ʔ
	vl aspirated	ph	th		kh	
	vd unaspirated	b	d			
	fricatives	(f)	s [s, ʃ]			h
	voiced fricative	v				
	nasals	m	n	ɲ	ŋ	
	lateral		l			
	semivowels	(w)		j		

Figure 9: So's initial consonants.

	bilabial	alveolar	palatal	velar	glottal
vl unaspirated stops	p	t		k	ʔ
fricatives		s [-ç]			h
nasals	m	n		ŋ	
semivowels	w		j		

Figure 10: So's final consonants.

Regarding to consonants, Suwilai establishes some correspondences with Thavung. He also notes the loss of the common Mon-Khmer trill /r/ and final consonants /-l/, /-ɲ/ and /-c/ (ibid. 169-170).

	front	central	back
close	i	ɨ	u
half close	e	ə	o
half open		ʌ	
open	ɛ	a	ɔ
	ia		ua

Figure 11: So's vocalic phonemes.

Finally, Suwilai (1996: 174-176) identifies the three main suprasegmentals found in "big So". These are (1) distinctive contrast between clear-tense/creaky syllables; (2) distinctive contrast between clear-lax/breathy syllables and (3) pitch. He gives minimal pairs for (1,2) and identifies 3 pitches: (3a) rising, (3b) mid-level and (3c) high-falling. The three phenomena are correlated, as shown in the table below (ibid. 177):

voice quality	**syllable structure**	**final consonant**	**pitch**
creaky	CVV, CVN, CVVN, CVVS	m n ŋ w j p t c k ʔ	rising-falling
breathy	CVV, CVN, CVVN, CVF, CVS, CVVS	m n ŋ w j ç h p t c k ʔ	mid-falling
clear	CVV, CVN, CVVN, CVVS	m n ŋ w j	rising-falling mid-level mid-falling
	CVS CVF	p t c k ʔ ç h	rising

Figure 12: correlation of suprasegmentals and syllable structure in So.

In Laos, Thavung (the name of the village) is spoken by a dozen families in a multilingual area, but by the time Ferlus did his investigation, language transmission was already interrupted and language attrition was already detected (Ferlus, 1979: 71-72). Ferlus' main purpose, as in his other inevstigations on Vietic languages, is to gather data for the reconstruction of proto-Viet-Muong and proto-Mon-Khmer families.

The syllable structure of Laos' Aheu is more reduced than Thailand's Aheu. Only monosyllabic and disyllabic words are found, following the structure (Cv)CV(C) (ibid. 72).

The next tables taken from Ferlus, 1979 (72-73) show the phonemic inventory of the Thavung language:

Consonnes initiales					
ph	th	s	kh	khw	h
p	t	c	k	kw	ʔ
b	d	j			
m	n	ɲ	ŋ		
v	l	y			

Figure 13: Initial consonants in Thavung (Ferlus, 1979: 72).

Consonnes finales				
p	t	c	k	ʔ
m	n	ɲ	ŋ	
		yh		h
w	l			

Figure 14: Final consonants in Thavung (Ferlus, 1979: 72).

ph	th	s	kh	h
p	t	c	k	
b				a
m				

Figure 15: Presyllabic reduced consonantal system (Ferlus, 1979: 73).

The table below shows the vocalic system of Thavung. Ferlus divides it according to the tonal system. He explains that the first vocalic system series, the high series, uses tones 1 to 3,

whereas the low series uses tones 2 to 4. The first series corresponds to old Thavung initial voiceless occlusives and the second to old Thavung initial voiced occlusives which are now devoiced (ibid. 73).

série haute (*1–3*)			série basse (*2–4*)		
i	ɨ	u	i	ɨ	u
e	(ö)	o	E	ə	O
ɛ	ə	ɔ		A	
	a				
ia	ɨa	ua	ia	ɨa	ua

Figure 16: Vocalic system in Thavung and tonal series (Ferlus, 1979: 73).

This interesting change which represents the origin of the two tonal series is explained as follows: after the devoicing of initial voiced stops, a tonal differentiation arose –tonal bipartition–, expanding from a 2-tone to 4-tone system. The old preglottalised initial stops changed to voiced initial stops and remained in the high series, as summarised in the following table:

p t c k ? ⟶ p t c k ? ʼb ʼd ʼj ⟶ b d j	série haute (*1–3*)
b d j g ⟶ p t c k	série basse (*2–4*)

Figure 17: Thavung tonal bipartition (Ferlus, 1979: 74).

Ferlus also analyses phonemic correspondences with PVM, Khamou and especially Modern Vietnamese. He finds systematic correspondences in tones and initial consonants. The table below shows the Thavung tonal system as Ferlus describes it. Thavung presents a four-way tone system, and represents an older stage in the described tonogenesis compared to the full process undergone by Vietnamese or Mường. In this case, we can see that the C column maintains the aspirated finals, with tones 1 and 2. Like Vietnamese and Mường, column D represents closed syllables in obstruents, which only can adopt tones 1 and 2.

	A	**B**	**C**	**D**
High	a¹	a³	ah¹	at¹
Low	a²	a⁴	ah²	at²

Table 6: Thavung tonal system (Ferlus, 1982: 103).

Hayes has focused on Thavung historical phonology. On the register tone system of Thavung, Hayes says that Thavung tones are level and unglided, and that the language has four tonemes, two per register (Hayes, 1982: 112).

In one of his papers, he evaluates the Daic influence on Thavung, which he establishes in 26% loanwords in Thavung from Daic origin, doubling Ferlus' percentage of 13%. These loanwords can be identified by looking at the mutations of initial consonants: some present two mutation processes: (1) voiced-unvoiced and (2) unvoiced-unvoiced aspirated (Lao type, following Ferlus); whereas some others have only undergone the first change (Viet-Muong type, following Ferlus) (ibid.: 101-103).

In his own study of the mutation of /*r/ in Thavung, Hayes identifies up to four strata of Daic loans in Thavung and determines the chronological order of the strata in relation to the register formation layers (before or after the formation) (ibid. 103-106). This latter distinction builds up to Ferlus' interpretation of the mutation of initial consonants. The Viet-Muong type, therefore, corresponds to pre-register formation loanwords, whereas the Lao type explains post-formation loanwords (ibid. 106-107).

Hayes suggests that the Daic family and Thavung language underwent the same tonogenesis process: when certain initial consonants became unvoiced, the pre-existing tones split in two (ibid.: 102).

Register	*Thavung*	*Daic*[25]			
		Proto-tone:	*A	*B	*C	*D
High	Hl	*Siamese:*	Mr	Ll	f	Ll
		Lao:	Hr	Ll	f	Hr/Ll
Low	Ll					
		Siamese:	Ml	f	H	H/f
		Lao:	Ml	Lf	Hf	H/f

Figure 18: Thavung and Daic registers (Hayes, 1982: 102).

Representation of Thavung register tones compared to the Thai and Lao tones developed from the four Daic proto-tones. As Hayes states: "Post-register formation developments have significantly altered the tone system in both Daic languages. The figure is not intended to clearly show the modern tones of either language. H= high, L= low, M= mid, f= falling, l= level, r= rising (Li, 1997)" (Hayes, 1982: 111-112).

Rục:

Rục has been studied mainly by Alves (2003) and Solntsev (1996). Solntsev counts the Rục population to be made up of about 120 members who live in the mountainous jungles in Central Vietnam. They are hunter-gatherers who also practice a slash and burn agriculture. According to him, the Rục and their language were first documented by the Vietnamese government in the 1950s. (Solntsev, 1996: 29).

The author remarks that Rục maintains some remnants of inflexional morphology and also pre-syllables, features that were lost in other languages (i.e. Vietnamese) and not retained in all Vietic languages. He has calculated that simple words to make up 84% and compounds make up 16% of the analysed corpus. Among the simple words, the majority are disyllabic and only a minority are monosyllabic, although "one still can observe the process of monosyllabisation." (ibid.: 29) Usually the disyllabic unit loses the first syllable which is as a weak syllable (a pre-syllable). Solntsev describes the Rục grammatical system as highly variational and unstable (ibid.: 30).

According to Alves, Rục has about 190 speakers who live in an isolated area that has helped preserve certain linguistic features that have been lost in Northern Vietic languages and are more similar to other Mon-Khmer languages. The author describes Rục as being lexically closer to Vietnamese, although its syllable structure and morphology is closer to the typical Mon-Khmer ones. Like other Vietic languages, Rục has 4 tones, an intermediate between the typical Mon-Khmer vocalic register system and the Vietnamese six-tone system. The lower number of tones is related to the preservation of finals /l/, /r/ and /h/. The relation of the finals with Vietnamese helps in describing a more detailed tonogenesis process, as described in the correspondent section (Alves, 2003: 3-13).

None of the two referenced studies make an in-depth phonological description. From Alves (2003) I have extracted the table representing the Rục initial clusters and a table representing the syllable structure compared to Vietnamese, Mường and Khmu:

Ruc Initial Clusters						
pl	tl	kl	ml	p^hl	k^hl	-
bl	-	-	-	-	-	-
pr	tr	kr	-	p^hr	k^hr	sr
br	-	-	-	-	-	-
-	tw	kw	ŋw	-	k^hw	hw
-	-	-	-	-	-	-
p^h	t^h	k^h	-	-	-	-

Figure 19: Rục initial clusters (Alves, 2003: 13)

Language	Monosyllabic	Bisyllabic
Vietnamese	CVC	none
Muong	CCVC	none
Ruc	CCVC	CVCVC
Khmu	CCVC	CVCCCVC

Figure 20: Maximum canonical syllable shapes

Alves also highlights other morphological characteristics, such as reduplication or derivational morphemes. Alves' paper's objectives are more intended to explain Vietnamese historical phonology through a related language rather than describing Rục itself.

Maleng:

There are three main dialects of the Maleng Vietic subgroup according to Ferlus: (1) the variety spoken in Khammouan, Laos, which can be called Maleng (referring to the speakers living in mountainous areas), Pakatan (referring to the speakers living in the village of Pakatan) or Kha Bo, (2) the variety spoken in Vietnam, called Mãliềng (vietnamised form of Maleng and the variety I am going to document and describe) and (3), Kha Phong of Maleng Kari, spoken in Laos close to the border with Vietnam (Ferlus, 1997: 55).

In his 1997 paper, the only one on Maleng, Ferlus briefly analyses one of the varieties of (3) Kha Phong, which he calls Maleng Brô, with 3 speakers left at the time the fieldwork was conducted and now considered extinct. The paper includes references and data from the other Maleng varieties and to Vietnamese and Vietic historical linguistics.

The phonology of Maleng Brô is as follows:

Les consonnes

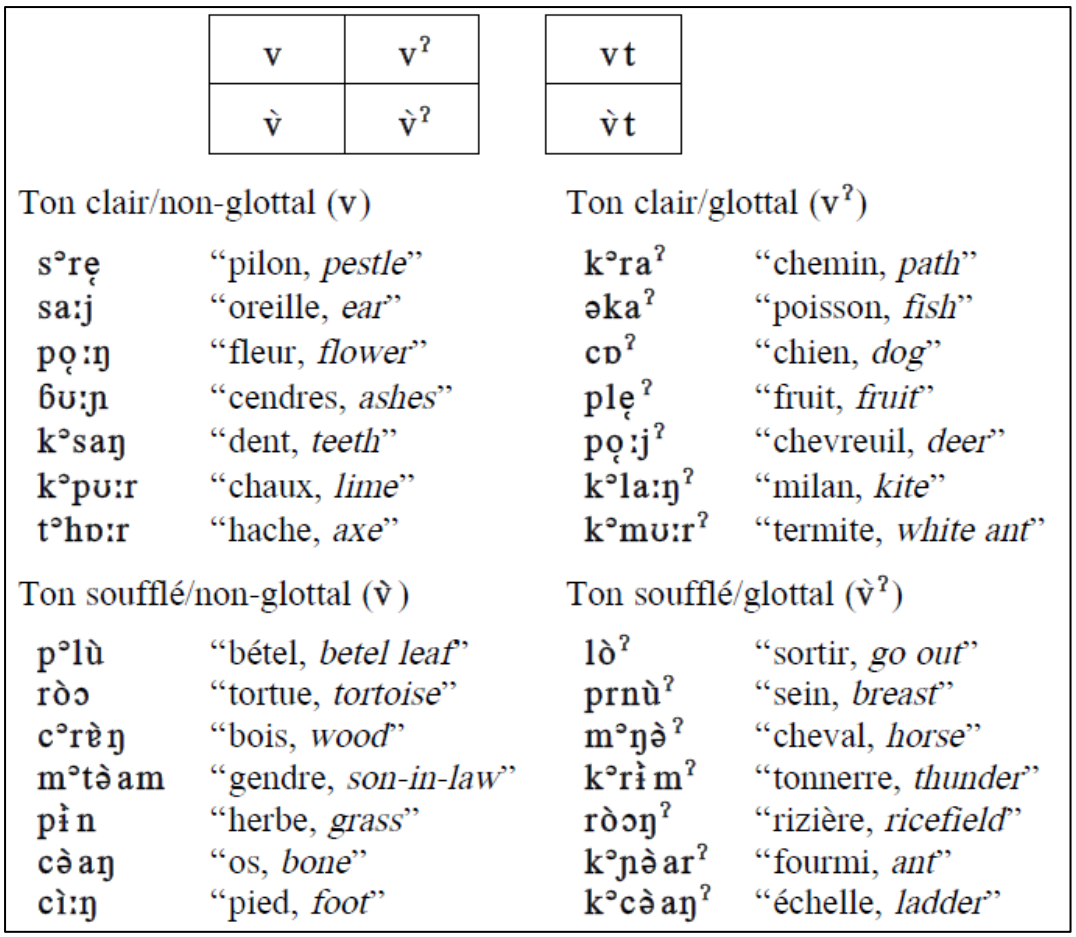
consonnes initiales					consonnes finales			
p	t	c	k	ʔ	p	t	c	k
ɓ	ɗ							
		s		h				
m	n	ɲ	ŋ		m	n	ɲ	ŋ
v		j			w		j	
	r	l				r	l	

Figure 21: Maleng Brô consonantic system. Ferlus, 1997, 57.

The vocalic system is complex: each vocalic timbre can be phonologically long or short and interacts with 3 variables: the opposition between clear/tense, between breathy/lax and between creaky/non-creaky (or glottalized/non-glottalised). Ferlus interprets this as a tonal system:

voyelles claires/tendues						voyelles soufflées/relâchées					
longues			brèves			longues			brèves		
ɪː	ɨː	ʊː	ɪ	ɨ	ʊ	ɪ̀ː	ɨ̀ː	ʊ̀ː	ɪ̀	ɨ̀	ʊ̀
ɐː	ʌː	ɔː		ʌ		èː	ə̀ː	òː		ə̀	
æː		ɒː	ɔ		ɒ	èɛ		òɔ			ɔ̀
	aː			a			ə̀a			ɐ̀	

Figure 22: Maleng Brô vocalic system. Ferlus, 1997: 57.



v	vˀ
v̀	v̀ˀ

vt
v̀t

Ton clair/non-glottal (v)

s°rẹ	"pilon, *pestle*"
sa:j	"oreille, *ear*"
pọ:ŋ	"fleur, *flower*"
ɓu:ɲ	"cendres, *ashes*"
k°saŋ	"dent, *teeth*"
k°pu:r	"chaux, *lime*"
t°hɒ:r	"hache, *axe*"

Ton clair/glottal (vˀ)

k°raˀ	"chemin, *path*"
əkaˀ	"poisson, *fish*"
cɒˀ	"chien, *dog*"
plẹˀ	"fruit, *fruit*"
pọ:jˀ	"chevreuil, *deer*"
k°la:ŋˀ	"milan, *kite*"
k°mu:rˀ	"termite, *white ant*"

Ton soufflé/non-glottal (v̀)

p°lù	"bétel, *betel leaf*"
ròɔ	"tortue, *tortoise*"
c°rɛ̀ŋ	"bois, *wood*"
m°tə̀am	"gendre, *son-in-law*"
pɨ̀n	"herbe, *grass*"
cə̀aŋ	"os, *bone*"
cì:ŋ	"pied, *foot*"

Ton soufflé/glottal (v̀ˀ)

lòˀ	"sortir, *go out*"
prnùˀ	"sein, *breast*"
m°ŋə̀ˀ	"cheval, *horse*"
k°rɨ̀mˀ	"tonnerre, *thunder*"
ròɔŋˀ	"rizière, *ricefield*"
k°ɲə̀arˀ	"fourmi, *ant*"
k°cə̀aŋˀ	"échelle, *ladder*"

Figure 23: Maleng Brô tonal system (Ferlus, 1997: 58).

The syllable structure follows the conservative Mon-Khmer syllabic type, having monosyllables and disyllables–formed by a consonantal or vocalic pre-syllable plus a monosyllable (Ferlus, 1997: 60).

Cuối:

According to Ferlus, the Thổ and Cuối varieties– Thổ from Lâm Lá, Thổ from Quỳ Hợp, Cuối/Thổ from Làng Lỡ and Cuối Chăm from Uy Lô–, together with Mọn and Kẽo, form the Thổ group (Ferlus, 2001: 1). From this subgroup, there are only two scarcely described languages/varieties, both carried out by Ferlus: Cuối from Làng Lỡ (Ferlus, 2001) and Cuối Chăm from Uy Lô (Ferlus, 1994).

Michel Ferlus' study on Cuối Chăm is based on its correspondences with Modern Vietnamese and Mọn. The description of the language is therefore done only in relation to Vietnamese. The Cuối Chăm initials we find in Ferlus' study are: /p, t, c, k, s, b, d, j/,

although we do not know if there can be more. The initial clusters Ferlus has found are: /bl, pl, kl, p^hr and k^hr/ (Ferlus, 1994: 1-2).

Cuối Chăm's tonal system has reached the final stage of the described tonogenesis process, like Vietnamese. It has a high and low series organised in 4 categories: A, B and C for open syllables or closed in nasal or sonorant, and D for closed syllables in a voiceless stop. The tonal system is very similar to that of Vietnamese (in italics):

	A	B	C	D
série haute	a^1 *a*	a^3 *á*	a^5 *ả*	a^7 *át*
série basse	a^2 *à*	a^4 *ạ*	a^6 *ã*	a^8 *ạt*

Figure 24: Tonal system of Cuối Chăm in comparison to the Vietnamese one (Ferlus, 1994: 2).

The vocalic system is also very similar to Vietnamese, with more elements. The table below shows the long vowels on the left and the short vowels on the right. The parenthesis indicates the vowels found only in borrowings from Vietnamese. The squared vowels indicate tonal inversions which we will not discuss:

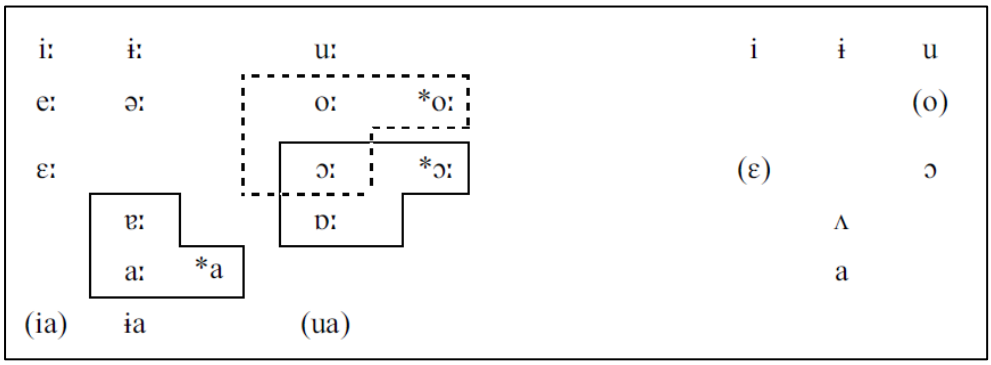


Figure 25: Vocalic system of Cuối Chăm (Ferlus, 1994: 2-3).

With regard to the final consonants, Ferlus highlight that the Proto-Viet-Muong final /-l/ is preserved, and that the Proto-Viet-Muong final /-s/ has become /-l/ and is now associated with tones 5 and 6 (Ferlus, 1994: 3-4).

Michel Ferlus' account of Cuối from Làng Lỡ (2001) carries out an analysis that does not relate to any other language. The following tables present the initial and final consonants and the vocalic system of Cuối:

	tʰ			kʰ	
p	t	c	ʈ	k	ʔ
ɓ	ɗ	ʃ			
f	s		ʂ		h
β	ð			ɣ	
	ʔð				
v		j			
	l		ɽ		
m	n	ɲ		ŋ	

Figure 26: Initial consonants in Cuối from Làng Lỡ (Ferlus, 2001: 2).

p	t	c	k
m	n	ɲ	ŋ
w		j	

Figure 27: Final consonants in Cuối from Làng Lỡ (Ferlus, 2001: 2).

i	ɨ	u
e	ə	o
ɛ	ʌ̆	ɔ
	a ă	
ie/ia	ɨə/ɨa	uo/ua
ea		oa

Figure 28: Vocalic system of Cuối from Làng Lỡ (Ferlus, 2001: 2).

The tonal system differs minimally to Standard Modern Vietnamese and Cuối Chăm, as it collapses tones 5 and 6 in one, just like Central and Southern Modern Vietnamese. The top row in the following table indicates the high tonal series, whereas the bottom row indicates the low one. Tones 1-2, 3-4 and 5-6 are realised on open syllables or closed in nasal or sonorants, whereas tones 7-8 are realised on closed syllables in a voiceless stop:

1 [¹¹]	3 [²⁴]	5	7 [²⁴]
2 [⁵⁵]	4 [⁵³]	6 [³³ʔ]	8 [⁵³]

Figure 29: Tonal system of Cuối from Làng Lỡ (Ferlus, 2001: 3).

Aspirated vowels in Cuối represent a corroboration of Ferlus' theories on the formation of Modern Vietnamese consonantal system (described in Ferlus, 1982; 1992 and discussed in the following sections). Finally, according to Ferlus, Thổ languages are very close to Vietnamese, both for genetic and language contact reasons, but are phonetically more conservative (Ferlus, 1994: 1).

Kri:

Kri is a Vietic language spoken by around 250 people in Laos which has been recently described by Nick Enfield and Gérard Diffloth (Enfield and Diffloth, 2009). 'Kri' is an endonym, being called Arem/Harème or Salang, names also used for other ethnic peoples by neighbouring groups. It is a very different language from widespread national languages such as Thai or Vietnamese, as it lacks lexical tone, it has a very complex phonological structure, it presents derivational–non-productive–morphology and has a fair number of syllable-final contrasts. These syllable contrasts appear in (1) register: heavy–light and in (2) terminance: checked–voiced–voiceless. By 'checked', the authors mean 'with full obstruction of airflow and lack of immediate release'. In fact, register and terminance cross-cut, resulting in six possible syllable types structurally and historically comparable to tones corresponding to a theorised stage of the tonogenesis in Vietic languages (ibid.: 4-9).

Kri's syllable structure may be disyllabic or monosyllabic. Disyllabic structures are formed by a sesqui-syllable (or 'minor syllable', the term used by the Enfield and Diffloth) plus the main syllable, which is allowed to present a complex onset. The structure scheme (without non-segmental features) is as follows: (CV)(C)CV(C) (ibid.: 10). Below I present the tables corresponding to the different syllable elements, taken from Enfield and Diffloth (2009):

	labial	alveolar	retroflex	palatal	velar	laryngeal
stops						
voiced (implosive)	ɓ	ɗ		ʄ		
voiceless aspirated	p^h	t^h	$ʈ^h$		k^h	
voiceless unaspirated	p	t	ʈ	c	k	ʔ
fricatives		s			ɣ	h
nasal sonorants	m	n		ɲ	ŋ	
oral sonorants	ʋ	l r/ɻ		j/ʝ		

Figure 30: Kri's major initial consonants (Enfield and Diffloth, 2009: 11).

Enfield and Diffloth treat terminance as a non-segmental feature. Thus, final oral sonorants have a three-way distinction, whereas final nasal sonorants present a two-way one. The following tables present the inventory of possible final consonants–where the capital V stands for 'vowel'–and the distribution of terminance distinctions across the three classes of final segments:

		bilabial	alveolar	palatal	(post) velar
stops	checked	p̚	t̚	c̚	k̚
sonorants	nasal	m	n	ɲ	ŋ
	oral	w	l r	j	V

Figure 31: Final consonants in Kri (Enfield and Diffloth, 2009: 16).

Realized with	Final stops	Final nasals	Final oral sonorants
Checked Terminance	✓	✓	✓
Voiced Terminance	*	✓	✓
Voiceless Terminance	*	*	✓

Figure 32: Distribution of terminance distinctions in Kri (Enfield and Diffloth, 2009: 19).

The vocalic system of Kri shows a fundamental distinction between long and short vowels. Long vowels present a greater number of quality distinctions than short vowels do. The following tables present the vocalic system of Kri. The first figure shows the long vowels paired by register, showing the opposition between heavy and light phonation types. The second figure shows the short vowels, with lower type frequency and more restricted phonotactics. In sesqui-syllables, only the short vowels /a, i, u/ can occur. There is no register distinction or length distinction:

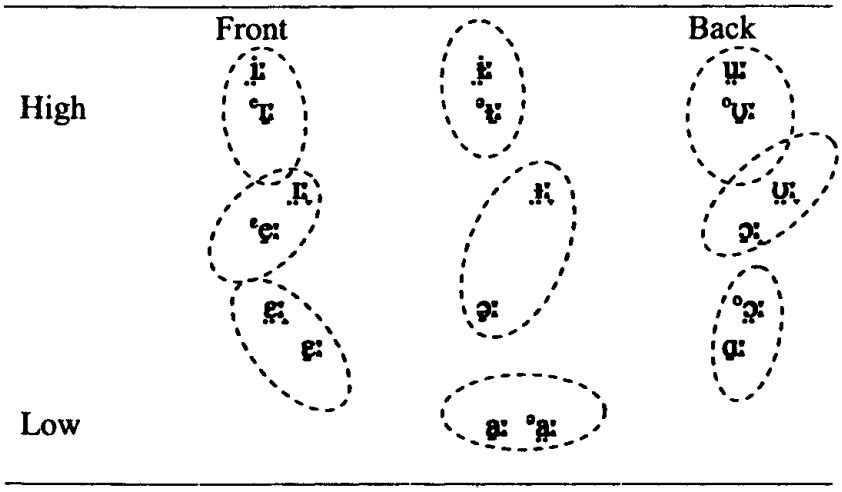


Figure 33: Long vowels in Kri, paired by register (Enfield and Diffloth, 2009: 28).

		Front	Central	Back
Non-Low	*Heavy*			
	Light			
Low	*Heavy*			
	Light			

Figure 34: Short vowels in Kri (Enfield and Diffloth, 2009: 31).

As for the morphology, Kri only presents derivational morphology, which is unproductive. All of the derivational morphemes are infixes, which add a sesqui-syllable to a major syllable. Kri also presents reduplication strategies (ibid.: 46-49).

7.2.2 Vietic documented data

Vietic languages–except for Vietnamese–are under-documented and under-researched. The research community has focused on other language families and subgroups in Vietnam, neglecting the closest relative Vietnamese still has. Moreover, all Vietic languages–again except for Vietnamese and to a minor extent, Mường–are severely endangered, although there is almost no information on their sociolinguistic status.

The only data gathered on Vietic languages available on-line is the information compiled by Michel Ferlus. These data are available on the collection Pangloss[4], belonging to the LACITO lab (CNRS, France). The data is organised according to the classification of Vietic language made by Ferlus (1996). The table below, adapted from the Pangloss collection and with information I have added, summarises the collected data present on Pangloss. The language variety I am going to document appears in bold. Note that not all the varieties of each listed language are documented. The data on Vietnamese corresponds to 2 peculiar central-north varieties. The last column indicates if there is any paper published related to the available data of each particular language. An overview of the Vietic languages and its classification deriving from the collected data is developed in Ferlus, 1996.

Language	**Language variety**	**Vocabulary list**	**Narrative**	**Comments and related papers available**
Arem		4 audio files + transcription (2x47' and 2x37')	1 Audio file (10')	Ferlus, 2014
Maleng	Maleng Brô	8 audio files (of about 20' each)		Ferlus, 1997
	Pakatan	4 audio files (3x47' and 1x21')		–
	Malang	4 audio files (3x47' and 1x15')		
	Malieng	3 audio files (2x46' and 1x39')		
	Kha Phong	4 audio files (of		

[4] https://lacito.vjf.cnrs.fr/pangloss/languages/Maleng_en.php [consulted on May, 2019]

		about 45’ each)		
Chứt	Rục	3 audio files (2x47’ and 1x28’)		
	Sách	3 audio files (2x47’ and 1x27’)		
	Salang	6 audio files (5x47’ and 1x28’)		
Aheu	Thavung	1 audio file (15’)	1 audio file (10’)	Ferlus, 1979
	Phôn Soung	6 audio files (3x32’, 2x19’ and 1x5’)		(poor audio quality)
Hung	Access restricted, available on contact (Ferlus)			
Thổ	Cuối Chăm Cuối	11 audio files + transcription (of about 15’ each)		Ferlus, 1994; 2001
Mường	Broken link			Nguyen M.C. 2015
Vietnamese	Phong Nha, broken			Michaud, Ferlus, Nguyen, 2015
	Cao Lao Hạ, broken			Ferlus, 1995

Table 7: Vietic documented data on Pangloss.

The Mon-Khmer Languages Project, from the SEAlang project[5], also has compiled data from Austroasiatic languages, including some data from Vietic languages. Most of these data, however, is the same data available on Pangloss and collected by Michel Ferlus. The scarce data found in the *Mon-Khmer Studies* publication is also mostly based on Ferlus’ fieldwork and publications.

[5] http://sealang.net/mk/vietic.htm# [consulted on May, 2019]

7.2.3 Vietic tonogenesis

One of the most important milestones in Asian comparative and historical linguistics has been the description of the process of tone adoption in Vietnamese, and by extension to the other Vietic languages which have also developed tone. This process was called tonogenesis by James Matisoff (1970; 1973: 73). The first scholar to detect signs of this historical process was Henri Maspero in 1912, who already detected the correlation between initial consonants and tone height in Chinese (Haudricourt, 1954: 70-71).

His classification of Vietnamese as a Tai language and his view that 'a language without tone cannot develop it and it can only be explained by its genetical affiliation' prevented him from completing the description. Przyluski at the same time defended that the conservation or loss of tones in a particular language is not relevant to determine the genetic affiliation of a language (Haudricourt, 1954: 69; Parkin, 1991: 90).

André-Georges Haudricourt completed his account of the tonogenesis in his 1954 publication *De l'origine des tones en Vièt*namien, which at the same time helped in classifying Vietnamese as an Austroasiatic language. His description was revised and expanded in 1961, when he published a comparative work on the tonogenesis processes in East- and Southeast-Asian languages (Haudricourt, 1961; see Haudricourt, 1972 for the English translation).

His description starts by analysing the correspondences between voiceless initial–high tone and voiced initial–low tone found by Maspero in Chinese, but also found in other languages. Ancient Chinese, which had 3 tones, at some point developed another 3 tones when a phonological change turned voiced initials into voiceless initials. The affected words adopted a high or low tone, according to the original consonant, in order to be distinguishable (Haudricourt, 1954: 71-72). Haudricourt applies this theory on Old Thai and finds correspondences, explaining the development of the Thai tonal system: from 3 tones to 6. Maspero also applied the theory to Vietnamese, explaining the bipartition of the tonal system also from 3 tones into 6 (Haudricourt, 1954: 72-73).

Haudricourt expands Maspero's theory by explaining the development of tones from the time at which the language had no tones. He does this by comparing data from several languages from different families (Miao-Yao, Sino-Tibetan, Austroasiatic and Tai-Kadai). Thus, the appearance of a 3-level tonal system is due to internal changes which led to having a middle tone, a low tone and a high tone. The low tone first appears as a result of the relaxation of the

vocal chords in words with final aspirations. When the aspiration disappears, this relaxation is maintained in the form of a descending tone. At the same time, the high or ascending tone appears in words with final glottals or glottalisations. These glottalised sounds make the vocal chords increase the tension in order to prepare for the following glottal sound. When the glottal sounds disappear, the tension increase is maintained in the form of an ascending tone (Haudricourt, 1954: 80-81).

The middle tone, according to Haudricourt, is developed in opposition to having both low and high tones, in order to be distinguishably not high and not low. It therefore appears together with the other two tones at the same time when the mutations of the initials occur (Haudricourt, 1954: 79).

After the development of a 3-tone system, the mutation of the initial consonants from voiced to voiceless prompted a tonal distinction, a bipartition of the existing 3 tones into 6 in order to distinguish the words previously disambiguated by the voicing of the initial consonant.

The following table, adapted from Haudricourt (1954: 81), explains the origin of the Vietnamese tonal system[6]:

Beginning of the 1st millenium		VI century		XII century		At present	
	pa sla hla		pa hla		pa la		ba la
N O	ba la	3	ba la	6	pà là	6	bà là
T O N E S	pas pah slas hlah bas bah las lah paX paʔ slaX hlaʔ	T O N E S	pà hlà bà là pá hlá	T O N E S	pả lả pã lã pá lá	T O N E S	bả lả bã lã bá lá
	baX baʔ laX laʔ		bá lá		pạ lạ		bạ lạ

Table 8: Vietnamese tononegesis according to Haudricourt, 1954.

[6] The 6 tonal system follows the Modern Vietnamese writing convention (see section 7.2.1)

Gage (1985) validates Haudricourt's theory and presents data to argue that historically the disappearance of final glottal stops preceded the disappearance of final aspirations. He also argues that Haudricourt's consonantal finals alone are insufficient to account for the described tonal developments. In a study published in 2005, Honda compares the tonal systems of Vietnamese, Ruc and Arem, corroborating Gages' demonstration and revising Haudricourt's model. Assuming Gage's and Honda's demonstrations, he proposes that the tonal genesis should be 0 > 2 > 4 > 6 or 0 > 4 > 6 instead of Haudricourt's 0 > 3 > 6. He also argues that the distinction of breathy–modal already existed in the toneless proto-language, and that 'both breathy voice and post-vocalic laryngeal constriction became responsible for pitch height and pitch contour respectively' (Honda, 2005: 185).

This contrast of breathy-modal is also accepted by Diffloth. According to the author, the breathy-modal contrast is historically posterior and unrelated to the creaky-clear voice contrast, attested in Proto-Katuic and Proto-Pearic, and as the French author proposes, in Proto-Austroasiatic, making these proto-languages register (not yet tonal) languages (Diffloth, 1982; 1989; 1990).

Graham Thurgood published a paper in 2002 in which he updates and extends Haudricourt's analysis by replacing its segmentally-driven model by a laryngeally-based model, incorporating the effects of voice quality distinctions. He explains that Mon-Khmer specialists have had to modify Haudricourt's theory, in that the Vietnamese developments are adequately explained only if the laryngeal effects of voice quality distinctions are recognized as central. He reanalyses the role of both initials and finals taking these affects into consideration (2002: 1-3). He concludes that

> "It is argued that distinctive laryngeal gestures are the primary mechanism of tonogenesis and that in most, if not all cases, these gestures have developed in the context of voice quality distinctions. Such a laryngeally-based model helps provide phonetically plausible explanation for the widely-attested correlation of pitch height and initial voicing and for correlations between voice quality and vowel quality." (Thurgood, 2002: 32).

Ferlus also adds a new contrast to Haudricourt's analysis to take into consideration: the *tense* vs *lax* contrast. His proposal for a new tonogenesis theory is based on the assumption that Viet-Muong languages were heavily influenced by Chinese and underwent similar processes which were involved in tonogenesis: monosyllabification and the adoption of tense-lax contrast, which lead to the known glottal-non glottal oppositions in finals (Ferlus, 2004).

Finally, Alves also proposed a reanalysis of Haudricourt's tonogenesis hypothesis which also argues for the consideration of different phonetic and phonation features that play an essential role: height, contour, breathiness, creakiness or tonal duration. His analysis also assumes that a toneless proto-Viet-Muong had creaky vowels and that pharyngeal and glottal phonetic phonation features played an essential role in creating a phonemically lexical pitch. He concludes that there are more stages involved than Haudricourt proposed in his tonogenesis account (Alves, 1995).

7.2.4 Proto-Vietic and Proto-Viet-Muong

Most of the research done on proto-Vietic is in fact on proto-Viet-Muong, as Vietnamese and Mường are the most well-documented languages. The studies on Vietic historical linguistics are scattered, inconsistent and very concrete, so it is very difficult to follow a research track on them. Vietnamese, and to a minor extent Mường, have been well-studied in terms of historical linguistics. Furthermore, tonogenesis and the history of tone has taken most of the attention drawn to Vietic languages, and other linguistic aspects have been understudied. Despite these issues, I will try to expose the work done on Vietic and Viet-Muong historical linguistics.

Michel Ferlus is the scholar who has worked most on Vietic historical linguistics and proto-Vietic. His work on the Vietic syllable structure is essential to comprehend his comparative studies (2014: 2). According to him, Vietic languages are divided up into monosyllabic and sesqui-syllabic languages in a continuum that goes geographically North (more monosyllabic) to South (more sesqui-syllabic). According to this author, a sesqui-syllabe is a composed syllable formed by a short, non-stressed pre-syllabe with no distinctive vowel and a reduced consonant system that precedes a monosyllable. The following table (adapted from ibid.) shows this continuum:

Vietnamese	100% monosyllabic
Muong/mường	100% monosyllabic
Thổ	100% monosyllabic
Toum-Liha	100% monosyllabic
Pong	10% of sesqui-syllabic structures
Thavung	35-40%

Maleng Brô	35-40%
Sách-Rục	35-40%
Arem	55-60%

Table 9: syllable structure of Vietic languages, adapted from Ferlus, 2014: 2.

Therefore, Ferlus believes that comparing the sylabically more conservative Vietic languages with the more innovative Vietnamese and Mường will shed light on the historical changes and the appearance of the proto-language. His comparison of Thavung with Vietnamese is a clear example of the successful application of the comparative method following this premise (Ferlus, 1979; 1996: 10). Ferlus' view that more conservative Vietic languages help reconstruct the proto-language is shared by other scholars: Alves (2003) used Rục (more conservative) to reconstruct older phases in the Vietnamese language, which strengthen the arguments that it is a Mon-Khmer language.

Michel Ferlus has made attempts to reconstruct Proto-Vietic, scattered throughout his publications. His reconstruction of Proto-Vietic initials is as follows:

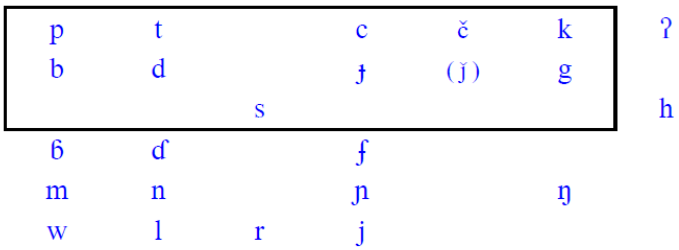


Figure 35: Proto-Vietic initials (Ferlus, 1982: 84, revised in 2014).

Departing from this reconstruction, Michel Ferlus has described a set of phonological changes, developed in Ferlus, 1982 (revised in 2014) and Ferlus, 1992 (also revised in 2014). In other papers exposing data on dialects of Vietnamese and other Vietic languages, he has made references to these rules, supporting them with the presented data.

Approximation of medial obstruents

This approximation occurred at some point between Proto-Vietic and Modern Vietnamese. As Ferlus notes, Modern Vietnamese consonants v, d, r, gi and g/gh (in Vietnamese ortography: quốc ngữ) are a result of an approximation of medial obstruents reconstructed in Proto-Vietic. By medial consonants we understand the initial consonant of the second syllable in a typical disyllabic word: sesqui-syllable + major syllable. Vietnamese was a

disyllabic language when this change occurred. When the sesqui-syllables disappeared, the medial consonants remained approximant as initials in monosyllables (Ferlus, 1982: 87-88).

The following table shows the Modern Vietnamese consonant inventory with the dialectal differences (North, South and Central). The squared consonants are the ones which, according to Ferlus, underwent this phonological change. These consonants are a result of the approximation of the squared consonants presented in the previous figure with Proto-Vietic initials:

ph /f/	*th* /th/	*x* /s/	*s* /s/N /ʂ/CS	*kh* /χ/	*h* /h/
	t /t/	*ch* /c/	*tr* /c/N /ʈ/CS	*k/c/q* /k/	# /ʔ/
b /ɓ/	*đ* /ɗ/				
m /m/	*n* /n/	*nh* /ɲ/		*ng/ngh* /ŋ/	
v /v/NC /j/S	*d* /z/N /j/CS	*gi* /z/N /j/CS		*g/gh* /ɣ/	
	r /z/N /j/CS				
	l /l/				

Figure 36: Vietnamese consonant inventory, with aspirated consonants signalised (Ferlus, 1982: 84, revised in 2014).

The table below shows the described process, from Proto-Vietic medials to Middle Vietnamese approximants. The letters ‘k’ and ‘a’ represent voiceless and voiced pre-syllables respectively. Note that there is also a correspondence with the tonal system, indicated in both Mường and Vietnamese. This scheme represents the approximation of labials, but it can be applied to the other places of articulation affected, as seen in the previous figures.

proto formes	évolution en vietnamien		séries tonales	
	bipartition tonale			
	spirantisation	*chute des présyl-labes et voisement*	muong	viêt
1 **k*-p	*k*-ɸ = k-ɸ	ɸ > β	h	h
2 **a*-p	*a*-ɸ > a-β	β = β	h	b
3 **a*-b	*a*-β = a-β	β = β	b	b

Figure 37: Approximation of medial obstruents (Ferlus, 1982: 98, revised in 2014).

The approximants resulting from this change are represented below:

β	ð	ɾ́	j̨	ɣ
w	j	r		

Figure 38: Approximants in Middle Vietnamese (Ferlus, 1982: 100, revised in 2014).

Formation of Modern Vietnamese initials

After this phonological change, approximants underwent other changes according to the Modern Vietnamese variety, until conforming the Modern Vietnamese phonemic inventory (figure 36).

pʰ	p	b	ɓ		m	w	
tʰ	t	d	ɗ	s	n	r	l
	c	ɟ	ƒ		ɲ		
	tʃ	(dʒ)		(ç)			
kʰ	k	g			ŋ		
	ʔ			h			

Figure 39: Proto-Vietnamese consonant inventory (Ferlus, 1992: 111).

The figure above presents the reconstructed Proto-Vietnamese consonant inventory. The figure below shows the current inventory for each major Modern Vietnamese variety resulting from the Middle Vietnamese approximants:

PV		spirantisation			*graphies* *XVIIe*	*graphies* *actuel*	exemples
p - b	>	ɸ/β	>	β	*ƈb/ƈbě*	*v*	*vôi* "chaux" (thavung **kpoːl**)
t - d	>	θ/ð	>	ð	*d/dě*	*d*	*dố* "mentir" (rục **ptoːj**)
c - ɟ	>	ç/j̨	>	j̨	*gi*	*gi*	*giết* "tuer" (rục **kcit**)
tʃ - (dʒ)	>	ç/j̨	>	j̨	*gi*	*gi*	*giụm* "aider" (cf. *xum* "se réunir")
k - g	>	χ/ɣ	>	ɣ	*g/gh*	*g/gh*	*gạo* "paddy" (thavung **əkoː**)
s	>	ř/r	>	r	*r*	*r*	*răng* "dent" (thavung **ksaŋ**)

Figure 40: Evolution of Middle Vietnamese approximants in Modern Vietnamese (Ferlus, 1982: 101, revised in 2014).

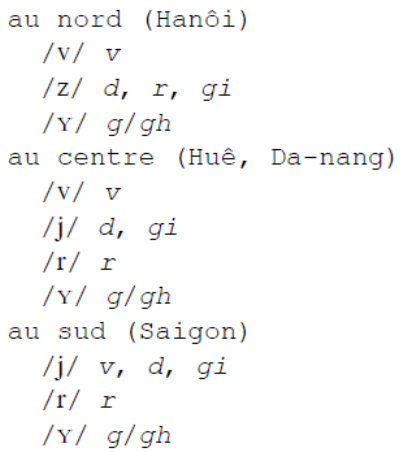


Figure 41: Modern Vietnamese approximant diversity (Ferlus, 1982: 101, revised in 2014).

As the scheme above shows, (1) in the North */r/ collapsed with */j/ into */r/ (now pronounced [z]), (2) in Central Vietnamese */j/ collapsed with */ð/ into */ð/ (now pronounced [j]), (3) in all Vietnamese varieties we can observe the following change: /β/ > /v/ (pronounced [j] in the South). Central Vietnamese is the variety which conserves the most distinctions (Ferlus, 1982: 100-101).

Vietnamien moderne	ɓ *(b)*	ɗ *(đ)*	t *(t)*	th *(th)*	c *(ch)*	s *(x)*	k *(k/c/q)*	ʔ *(#)*
XVIIe (centre) *Dictionarium*	ɓ *(b)*	ɗ *(đ)*	t *(t)*	th *(th)*	c *(ch)*	ś *(x)*	k *(k/c/q)*	ʔ *(#)*
XV-XVIe (nord) *Hua-yi yi-yü*	ɓ *(p)*	ɗ *(t)*	t *(t)*	(ś) *(sh/ss ch'/ch)*	c *(ch)*	č *(ts' ch'/sh)*	k *(k)*	ʔ *(#)*
Dévoisement	p	t	s	(ś)	c	č	k	ʔ
1 Obstruantes du Proto viêt-muong	p b	t d	s	(ś)	c ɟ	č (ǰ)	k g	ʔ
2 Spirantisation	ɸ β	θ ð	ř ŕ		ç j̊	χ ɣ		
3 Voisement	β	ð	r		j̊	ɣ		
4 XV-XVIe (nord) *Hua-yi yi-yü*	β *(p)*	ð *(t)*	r *(sh/ch'/j)*		j̊ *(ch)*	ɣ *(k)*		
5 XVIIe (centre) *Dictionarium*	β *(ɓ/ɓě)*	ð *(d/dě)*	r *(r)*		j̊ *(gi)*	ɣ *(g/gh)*		
6 Vietnamien N moderne C S	v v j *(v)*	z j j *(d)*	z r r *(r)*		z j j *(gi)*	ɣ ɣ ɣ *(g/gh)*	ʔ ʔ ʔ *(#)*	
Muong (khến)	p	t	th	th	c	s	k	ʔ

Figure 42: Evolution medial obstruents from PV to MV (Ferlus, 1982: 101, revised in 2014).

The table above summarises the changes exposed by Ferlus (1982). Ferlus relates these changes to the influence of Chinese on northern Vietic languages, which has enriched their consonant systems and lead them to undergo monosyllabification (Ferlus, 1982: 103; 1992: 120). In his 1992 paper, Ferlus explores in more depth the historical changes of the Vietnamese consonant system in relation to Chinese influence.