A history of Cham and its speakers

Marc Brunelle, ***University of Ottawa***

Graham Thurgood, ***California State University, Chico***

The final version appeared in a trilingual Khmer-Cham-English book published in Cambodia that is hard to find.

Brunelle, Marc and Graham Thurgood. "The Historical Development of Chamic Languages". Ed. José López Cortina and Alberto Pérez Pereiro. *Rediscovering Cham Heritage in Cambodia*. Phnom Penh: Naga Editions, 2015. 46-97.

Cham and the other modern Chamic languages are distributed over a considerable area in Southeast Asia: from the Tsat (or Hainan Cham)[1] on the outskirts of Sanya City on Hainan island to the Western Cham of Cambodia, and from the Jarai of the Annamite Cordillera to Eastern Cham on the south-central coast of Vietnam. In addition, sizeable immigrant communities exist in Malaysia, France, Australia and the United States.

Cham and its sister languages are unique because, if we exclude the Malay Peninsula, they are the only Austronesian languages spoken in Mainland Southeast Asia. As such, they have long been in contact with languages of the Mon-Khmer family (to which languages like Khmer, Kuy and Vietnamese belong) and they have over time developed various typological features reminiscent of Mon-Khmer. For this reason, the Austronesian character of Chamic languages has not always been recognized. Early scholars did recognize it, using, like Crawfurd, labels such as the "Malay of Champa", but the classification of the Chamic languages came into question at the beginning of the twentieth century, as some researchers were led astray by the large number of loanwords and grammatical similarities with Mon-Khmer languages. An accompanying belief was that Malayo-Polynesian spread out into the islands from the Chamic speaking area, a belief fully repudiated by modern scholarship but still found occasionally in linguistically unsophisticated surveys.

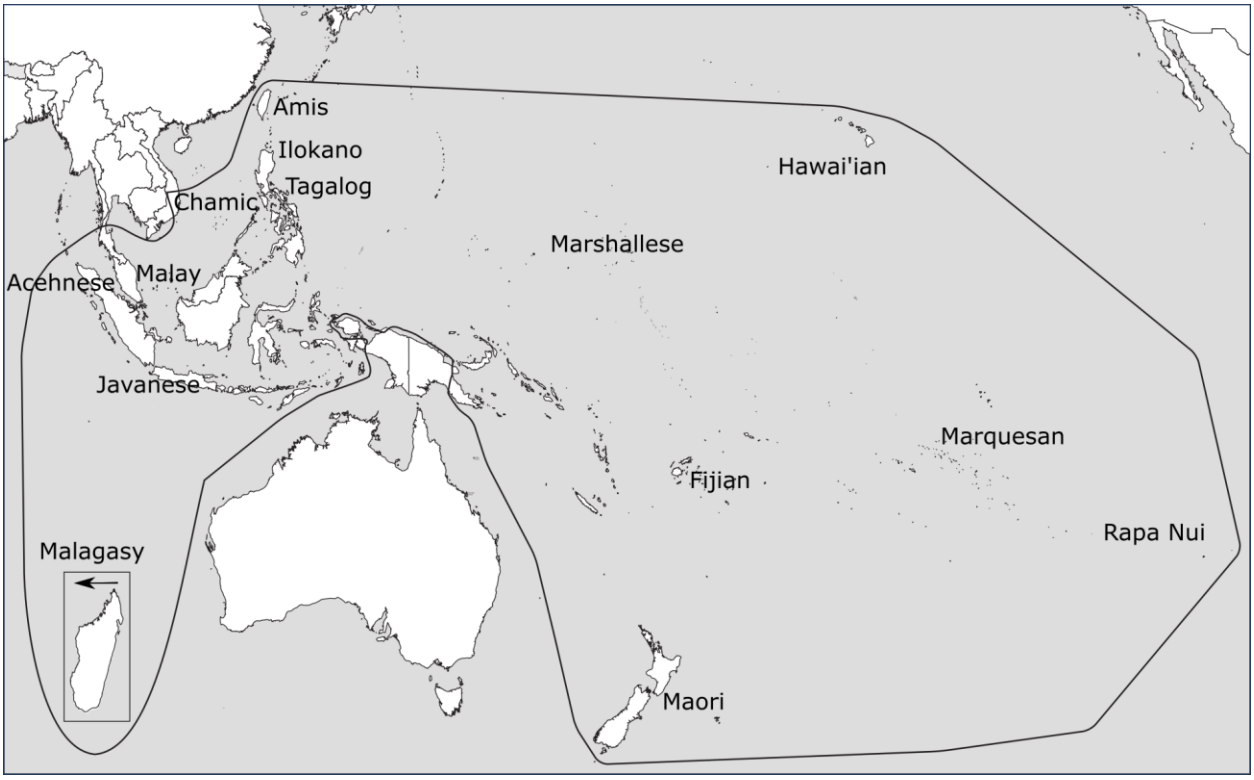
In this chapter, we will review the evidence that establishes Cham as a member of the Austronesian language family. We will also summarize the linguistic history of Cham and of its convoluted migrations from Taiwan to Cambodia, through Insular Southeast Asian and Vietnam. We will conclude with an overview of the modern distribution and characteristics of Chamic languages. This chapter is not meant for an academic audience and will therefore overlook details and controversies; readers interested in furthering their understanding of the topic at hand are invited to start with the references found at the end.

The position of Cham in Austronesian

[1] Hainan Chams refer to themselves as Utsa:nʔ^{33} and to their language as Tsa:nʔ^{33} (simplified as Tsat). This is clearly related to the term *Cham*, which is why they are often called the Chams of Hainan. In this paper, we will use *Tsat* to avoid any confusion: although Tsat is a Chamic language, it is not a variety of the Cham language spoken in Cambodia and Vietnam.

The current consensus on the origin of Austronesian languages is that they all descend from a common mother language, Proto-Austronesian, that was spoken in Taiwan around 5000 BC. For reasons that we will leave to archeologists, but may be related to a population expansion caused by rice cultivation, some Austronesian speakers left Taiwan and sailed to the Philippines around 3000 BC. Over the following millennia, they then expanded over an impressive geographical area ranging from the Easter Island, in the Eastern Pacific, to Madagascar, off the coast of East Africa. This area is illustrated in Map 1.

The Austronesian family now boasts half a million speakers and includes large languages like Javanese, Malay/Indonesian, and Tagalog, but also smaller but well-known languages like Hawai'ian, Maori, Fijian and Malagasy. As can be seen in Table 1, these languages, although they branched out millennia ago, still share important lexical similarities.



Map 1: The current geographic distribution of Austronesian languages, with location of languages mentioned in the chapter (see maps below for Chamic languages)

Table 1: Comparative table of Austronesian vocabulary

gloss	**Amis**	**Tagalog**	**Ilocano**	**Western Cham**	**Malay**	**Malagasy**	**Javanese**	**Maori**	**Hawai'ian**
one	cecay	isa	maysa	sa, ha	satu	iray	siji	tahi	kahi
two	tosa	dalawa	dua	ṭwa	dua	roa	loro	rua	lua
three	tolo	tatlo	tallo	klăw	tiga	telo	telu	toru	kolu

four	sepat	apat	uppat	pa:t	empat	efatra	papat	mhă	hă
five	lima	lima	lima	limɨ	lima	dimy	limo	rima	lima
six	enem	anim	inem	năm	enam	enina	nem	ono	ono
seven	pito	pito	pito	tuc̥uh	tujuh	fito	pita	whitu	hiku
eight	falo	walo	walo	t̪apăn	lapan	valo	wolu	waru	walu
nine	siwa	---	siam	salipan	sembilan	sivy	songo	iwa	iwa
ten	poloʔ	---	sangapolo	pluh	sepuluh	folo	sepuluh	tekau	-ʔumi
house	luma	bahay	balay	sa:ŋ	rumah, balai	trano	omah	whare	hale
dog	wacu	aso	aso	săw	anjing	alika	asu	kuri	īlio
road	lalan	daan	dalan	c̥alan	jalan	lalana	dalan	ara	ala
pig	fafoy	baboy	baboy	p̪a̪puy	babi	kisoa	babi	poaka	pua'a

By comparing the shared and divergent sounds of the words of various Austronesian languages, linguists were able to establish their family tree. The basic idea is that if languages share a regular sound change (or an innovation), this change is more likely to have happened once in their ancestor language before it split up than multiple times in all its daughters after it split up. It is therefore probable that languages that share an innovation are more closely related than languages that do not share this innovation. This allows us to establish, for instance, that Malay, Javanese and Moken are more closely related than, say, Hawai'ian. The family tree of Austronesian languages, according to one among many recently proposed models, is given in Figure 1. They are subdivided into several sub-groups, or sub-families that roughly match geographical areas. Except for the Formosan languages, all other Austronesian languages belong to the Malayo-Polynesian subgroup, which includes all the languages that descend from the language that left Taiwan. As shown in Figure 1, Malayo-Polynesian then split into Nuclear Malayo-Polynesian and Borneo-Philippines languages, with Western Malayo-Polynesian containing a myriad of languages including various languages of Sumatra, Java, Malayic (e.g., Malay and Iban), and our Chamic languages.

When sub-groupings are well-established, it becomes possible to use a series of techniques called the comparative method to reconstruct the ancestor language of a subgroup. We can, for instance, reconstruct with a good level of confidence the ancestor of Polynesian languages by comparing the words of Hawai'ian, Maori, Fijian, Trukese and their sisters, even if this language was never written. A language that has been reconstructed but for which we have no written records is usually labeled a proto language. Thus, Proto-Austronesian is the reconstructed ancestor of all Austronesian languages.

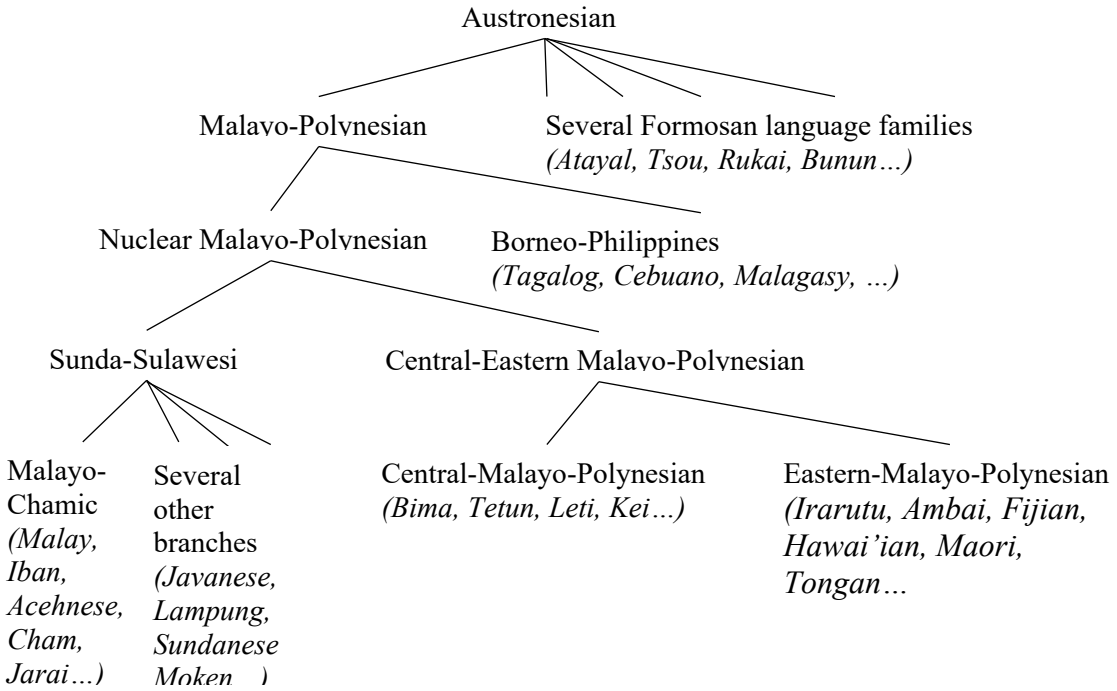


Figure 1: The position of Chamic in Austronesian (the detailed subgroupings are still debated)

Not marked in Figure 1 is the fact that the Malayic languages and the Chamic languages have a sister relationship with each other and constitute a subgroup called Malayo-Chamic. The similarity of forms between the two subgroups can be seen in the numbers, shown in Table 2. The first column is the reconstructed PMP (Proto-Malayo-Polynesian); the * indicates that the form is reconstructed, rather than attested. The second shows modern Malay, a Malayic language, the third shows the reconstructed PC (Proto-Chamic), while the fourth and fifth show two attested modern Chamic languages, Western Cham and Rhade. The similarities between the oldest stage represented by PMP and the modern languages Malay (representing the Malayic languages) and Proto-Chamic (representing the oldest Chamic forms), along with E. Cham and Rhade (representing the modern Chamic languages) are quite striking.

Table 2: Numerals in Malayo-Chamic

Proto-Malayo-Polynesian	Malay	Proto-Chamic	Western Cham	Rhade	
*esa; *isa	sa-, satu	*sa	sa, ha	sa	'one'
*duha	dua	*dua	ṭ̥wa	dua	'two'
*epat	ĕmpat	*pa:t	pa:t	păt	'four'
*lima	lima	*lima	limɨ	ema	'five'
*enem	ĕnam	*nam	năm	năm	'six'

Within the Chamic languages, such cross-linguistic similarities are found throughout. A caveat is in order here; for the skilled historical linguist, it is the regular correspondences between the sounds, more than phonetic similarity in forms that provide the strongest evidence of a genetic relationship; in the case of Chamic we have both regular correspondences and phonetic similarity.

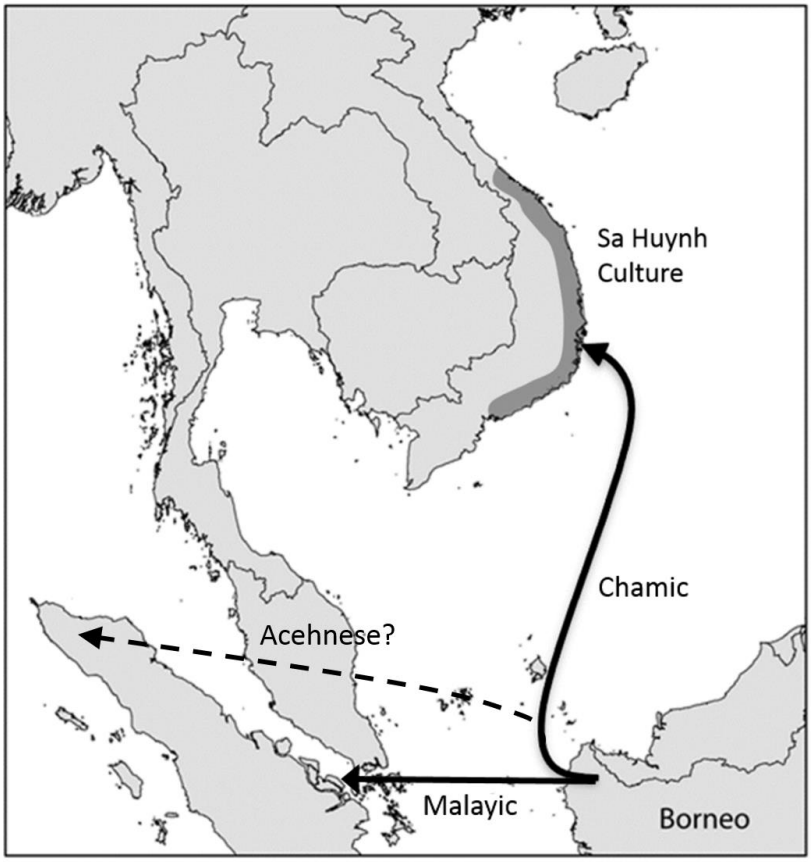
Table 3: Some corresponding Chamic words

Malay	Proto-Chamic	Acehnese	Chru	Northern Raglai	Formal Eastern Cham	
darah	*darah	darah	drah	darah	ṭarăh	'blood'
dara	*dara	dara	dra	dara	ṭara	'girl'
bulu	*bulɔw	bulɛə	bləu	biləu	p̥ilŏw	'hair'
bulan	*bula:n	bulɯən	ea bla:n	ia bila:t	p̥ilan	'moon'
baharu	*bahrɔw	baro -f	bərhəu	bahrəu	p̥irŏw	'new'
tahun	*thun	thon	thun	thut	thŭn	'year'
tangan	*taŋa:n	---	təŋa:n	taŋãn	taŋĭn	'hand'

Table 3 illustrates words with the regular correspondences and phonetic similarities. The initial d- of Acehnese, Chru, and Northern Raglai regularly corresponds to the ṭ- in E. Cham, a correspondence that makes phonetic sense; in a parallel way, the initial b- corresponds to p̥- in Eastern Cham, and so on. It is not of course just the presence of a handful of such forms but instead several that constitutes the evidence of a genetic relationship. Proto-Chamic reconstructed on the basis of such patterns scholars represents a scholar's reconstruction of the oldest Chamic forms.

The move from the islands to the mainland: early Chamic

Comparison with other languages suggests that Malayo-Chamic languages were originally spoken in Borneo, probably in the Kapuas river basin. In the last centuries BC, Malayo-Chamic languages seem to have expanded out of west Borneo with the Malayic speakers moving to Sumatra, while the Chamic speakers sailed off to the coast of Vietnam. Several authors have speculated that Malayo-Chamic speakers might also have migrated to other areas of Mainland Southeast Asia, perhaps forming a string of settlements going from the Malay Peninsula to Central Vietnam and including the Mekong delta (i.e. the polity referred to as Funan in Chinese contemporary sources). While this is certainly a possibility, there is no clear evidence in support of such a scenario. Map 2 therefore sticks to migrations for which we have actual evidence.



Map 2: The early migration to the Mainland (a few centuries BC). The other possible migration path for Acehnese appears in map 3.

We know very little about the establishment of Chamic people on the coast of south-central Vietnam, but a number of archeological and genetic facts are often used as a basis for educated linguistic speculation. First of all, there seems to be a chronological overlap between the departure of Malayo-Chamic speakers from Borneo and the development of the Sa Hùynh culture on the coast of Central Vietnam. As the geographical distribution of Sa Huỳnh archeological sites roughly match the historical distribution of Cham polities, many scholars equate the Sa Huỳnh culture and Chamic speakers. While this is probably a gross over-simplification, it is likely that Chamic speakers played an active part in Sa Huỳnh culture. As for genetic evidence, it shows that current Eastern Cham speakers share most of their mitochondrial DNA haplogroups with their Mon-Khmer neighbors, but most of their Y-chromosome lineages with Austronesian groups. There are many possible interpretations of these facts, but the simplest would be to assume that upon arrival on the coast of Vietnam, Chamic groups were mostly, but not exclusively, composed of young men who established families with Mon-Khmer women (contact with a Mon-Khmer group in Borneo is another).

The linguistic evidence seems to point in the same direction. It suggests that Chamic evolved out of the contact of an Austronesian-speakers with a Mon-Khmer speakers. Even in the

earliest forms, those reconstructed for Proto-Chamic, the influence of Mon-Khmer can be seen throughout the language. The restructuring under the influence of Mon-Khmer languages is strikingly obvious in the development of a more Mon-Khmer-like phonology, in the massive Mon-Khmer lexical borrowings, and in the borrowing of bits and pieces of Mon-Khmer morphology.

If we put all this evidence together and look at the later distribution of Chamic-speaking groups, the most likely scenario is that Chamic was originally a dialect chain stretching along the coast of Vietnam, and possibly extending into the Highlands relatively early on. As the Cham lost control of the sea lanes along the coast, the settlements became more isolated from one another allowing the once unified dialect to differentiate, which led to the development of the modern Chamic languages.

The structure of Proto-Chamic

Although we normally expect linguistic systems to become simpler under contact, the original Austronesian sound system of Chamic became more complex under contact with Mon-Khmer. There was for instance an increase in the number of vowels and consonants, including the introduction of length distinctions, and typologically rare sounds were borrowed. This may be interpreted as further evidence that a large number of Mon-Khmer speakers shifted language and adopted Proto-Chamic.

A first example of this is that the basic Chamic word went from an Austronesian disyllabic word with first syllable stress to a Mon-Khmer-like iambic system, that is, from forms with stressed first syllable and an unstressed second syllable (or possibly no word stress) to an unstressed first syllable and a stressed second syllable. As the first syllable became unstressed, it was even lost in some environments. In Table 4, the first five examples show words that, like all other words of Chamic, have stress on their final syllable. In Chru, unstressed vowels are reduced to the vowel /ə/, which is similar to what happens in English (e.g. the unstressed vowel in saxophone [ˈsæksəˌfoʷn]). In Tsat, the reduction process goes even further and the unstressed vowel is deleted. The second group of examples in Table 4 shows that the stress shift has other effects. In these words, the reduction of the unstressed vowels led to the development of the aspirated stops that are now pervasive in Chamic languages.

Table 4: Stress shift of Proto-Chamic

Proto-Malayo-Polynesian	**Malay**	**Proto-Chamic**	**Acehnese**	**Chru**	**Northern Raglai**	**Tsat**	
*mamaq	mamah	*mamah	mamʌh	məmah	mumãh	ma^{55}	‘chew’
*qumah	huma	*huma	umʌŋ	həma	humã	ma^{33}	‘dry field’
*lima	lima	*lima	limʌŋ	ema	lumã	ma^{33}	‘five’

*panaq	panah	*panah	panah	mənah	panãh	na^{55}	‘shoot [bow]’
*baseq	basah	*basah	basah	məsah	pasah	sa^{55}	‘wet; damp’
*paqit	pahit	*phit	phet	phi:ʔ	phi:ʔ	phiʔ^{24}	‘bitter’
*taqu	tahu	*thɔw	thɛə	thəu	thəu	tiauʔ^{42} -i	‘know’
*paqa	paha	*pha	pha	pha	pha	pha^{33}	‘thigh’
*taqun	tahun	*thun	thon	thun	thut	thun^{33}	‘year’

Compared to Proto-Malayo-Polynesian, the number of second syllable vowels of Chamic also exploded through splits and borrowings. To illustrate this, we give in Table 5 the vowel system of Proto-Malayo-Polynesian and vowel system of Eastern Cham, which derives from it. Note that we exclude diphthongs, i.e. vowels that are composed of two sequential vowels qualities.

Table 5: The dramatic expansion of the Chamic vowel inventory

PMP vowels		Eastern Cham vowels in stressed syllables		
*i	*u	i/i:	ɨ/ɨ:	u/u:
*ə		e	ɤ/ɤ:	o
*a		ɛ/ɛ:	a/a:	ɔ/ɔ:

Finally, a number of other changes also occurred, likewise making the system more Mon-Khmer-like. For example, the typologically rare glottalized consonants ɓ, ɗ and · (as in Cham ɓuʔ ‘hair’, ɗih ‘to lie down’ and ·ăwʔ ‘correct’) and aspirated consonants entered Chamic both through borrowing and in part through internal paths of change (see the bottom four rows of Table 5).

The Mon-Khmer influence on the Chamic lexicon is also important. A well-known fact in historical linguistics is that basic vocabulary (body parts, kinship terms, common animals, simple meteorological phenomena) is less likely to be borrowed than more complex or more specialized vocabulary. A large proportion of Proto-Chamic basic vocabulary is indeed Austronesian:

*kɔw ‘I (familiar)’
*ʔadhĕy ‘forehead’
*dalam ‘in; inside’
*ʔana:k ‘child’
*ʔadĕy ‘younger sibling’
*-ən- ‘instrumental infix’
*ʔama ‘father’
*kulit ‘skin’

*kukɔw ‘fingernail, claw’ *ʔapuy ‘fire’, *bula:n ‘moon’

*danaw ‘lake’ *hitam ‘black’

*hurɛy ‘day; sun’

However, many Proto-Chamic pronouns, kinship terms and basic vocabulary have been borrowed from Mon-Khmer, which could suggest a massive influx of non-native speakers:

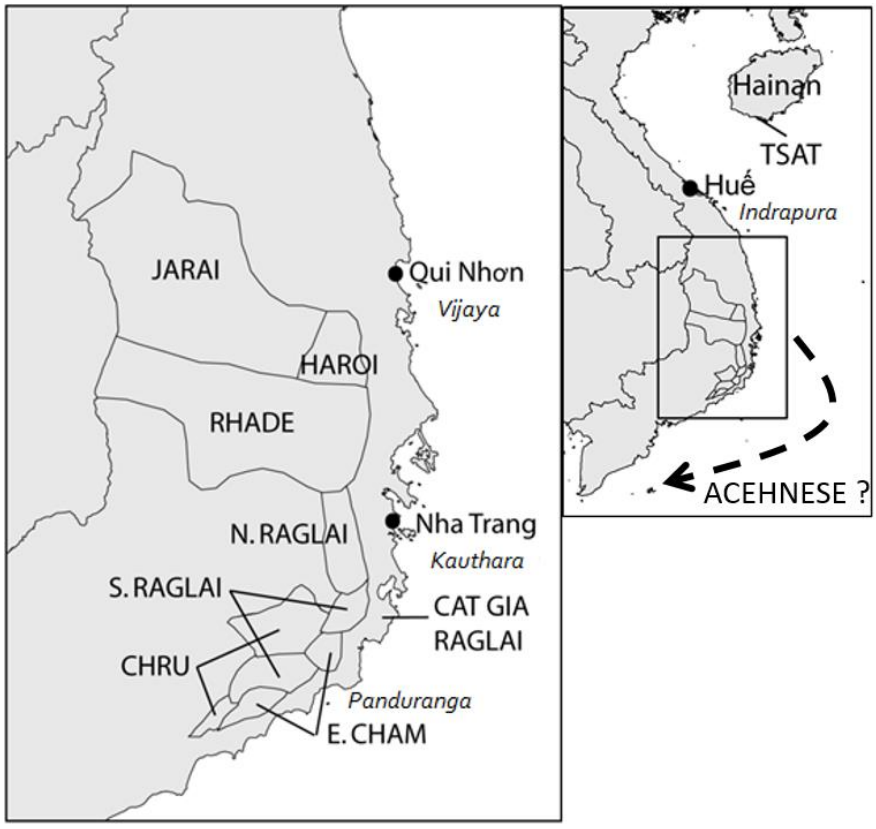
*gəp ‘other; group’ *cuco ‘grandchild’

*ʔakɔʔ ‘head’ *ɓɔ:ʔ ‘face’

*cadiaŋ ‘finger’ *ka:ŋ ‘chin’ jaw’

*sapal ‘arm’ *takuay ‘neck’

Cham enters history



Map 3: The Cham homeland: Central Vietnam with current location of Cham linguistic groups (in uppercase) and rough location of regions of Champa (in italics)

As mentioned above, it is likely that in the first century AD, varieties of Cham were roughly concomitant with vestiges of the Sa Huỳnh culture, from Thừa Thiên to Bình Thuận

provinces along the coast Central Vietnam. Although we have no direct linguistic evidence at this early stage, Cham was probably not very differentiated at that time. Central Vietnam was probably still a patchwork of mutually intelligible Cham dialects interspersed with Mon-Khmer languages.

Cham enters history at the end of the 2nd century. The first attested Chamic polity, Linyi, first shows up in the Chinese sources in 192 AD in the area around modern Huế, after a local revolt against Chinese rule. Although Linyi was under Chinese cultural influence, this left no trace on Chamic languages. There is a sizeable number of words of Chinese origin in the modern Chamic languages spoken in Vietnam, but the large majority of were borrowed through Vietnamese.

There is limited evidence of state formation further south, but we know, thanks to a Sanskrit inscription at Võ Cạnh, in Khánh Hòa, that Indian influence had reached Cham-speaking areas by the 3rd century AD. The Võ Cạnh stela is the first Sanskrit inscription in Southeast Asia and is written in a script derived from Pallava, the South Indian script on which all Southeast Asian Indic scripts (including Khmer, Thai and Lao scripts) are based. Sanskrit had an influence on Chamic languages. This influence is mostly felt in learned vocabulary, but frequent colloquial words also have Sanskrit origins, like *raṭeh* 'cart' from *ratha* 'charriot', *aseh* 'horse' from *aśva* and *manujh* 'person' from *manuṣya*. However, this does not mean that there was a strong Indian presence in Chamic-speaking polities. The current consensus is that there were significant commercial and cultural ties with southern India, but that the actual number of Indians present in the area was probably small, and limited to traders, cultural and religious specialists and artisans. Recent genetic evidence reveals that Indian Y chromosomes are found in the Cham communities of south-central Vietnam, suggesting that some Cham have Indian male ancestors, but the prevalence of these chromosomes is relatively low.

From the 4th to the 15th centuries, we have extensive historical evidence of the existence and development of Cham polities in Central Vietnam. A large proportion of this evidence comes from Chinese and Vietnamese chronicles, but more than 200 inscriptions in Sanskrit and Cham from that period have also been found all over Central Vietnam. The northernmost Cham inscription was recently discovered in the Phong Nha cave in Đồng Hới province, while the southernmost was found on a Vishnu statue in Biên Hòa in Đồng Nai province (Figure 3). These inscriptions are mostly commemorating military victories and establishing donations from the powerful to temples and religious foundations.

Until the 9th century, the large majority of inscriptions are in Sanskrit, and bring us no evidence about the linguistic structures of Cham. A possible exception is the Đông Yên Châu inscription found in Quảng Nam province and first published in 1939 by Georges Cœdès (Figure 2). Although it is not dated, a date can be estimated from the middle of the 4th century based on similarities with the script of neighbouring Sanskrit inscriptions. The Đông Yên Châu inscription contains five Sanskrit loanwords (underlined in (1)), but no Mon-Khmer loanwords, which could very well be due to its brevity. It also contains a few words have been lost in Cham but still exist

in other Malayo-Polynesian languages (like *spūy*, which is no longer found in Cham, but seems related to Malay *sepoi* ‘gentle’).

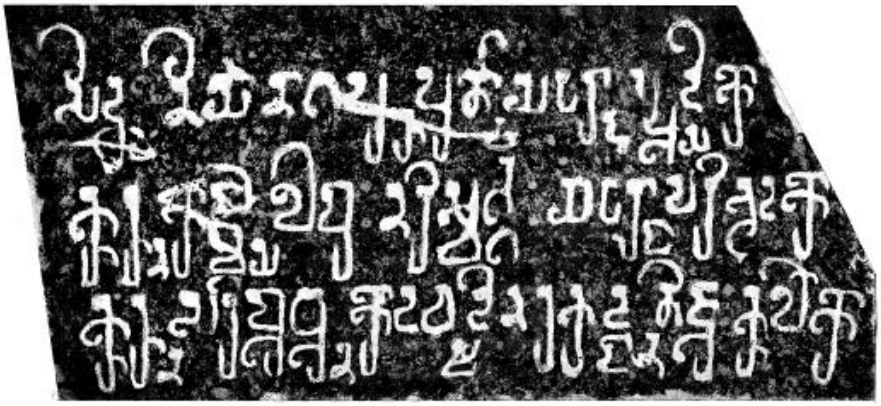


Figure 2: Moulding of the Đông Yên Châu inscription, 4th century (from Coèdes 1939)

(1) Siddham ni yāŋ nāga puŋ putaw ya urāŋ spūy di ko
Success! This god naga possession king who person gentle at 3p.s.

kurun ko jmāy labuh nari svarggah ya urāŋ paribhū di ko
? 3p.s. jewel fall from sky who person insult at 3p.s.

kurun saribū thūn ko davan di naraka dŋan tijun kulo ko.
? 1000 year at ?? at hell with seven family 3p.s.

‘Success! This is the holy naga of the king. Whoever treats it gently … jewels will fall from the sky; whoever insults it… a thousand years in hell with seven generations of his family’

From the 9th century to the 11th century, Sanskrit and Cham coexist in stone inscriptions. From the 11th century on, Sanskrit inscriptions become rare and Cham becomes increasingly codified, despite significant variation in spelling conventions. The script itself also evolves during this period (compare Figures 2 and 3). A form of modern Cham script still in use in Vietnam for amulets and religious texts, akhăr rĭk ‘the old script’ is derived from the script used in inscriptions (Figure 4). To our knowledge, the last known Cham inscriptions are engraved at Po Rome temple in Ninh Thuận province.

Figure 3: The Biên Hoà inscription (1421) EFEO estampage 263

http://isaw.nyu.edu/publications/inscriptions/campa/inscriptions/C0001.html

Figure 4: An example of akhăr rĭk (from Aymonier et Cabaton 1906)

There is growing evidence that Champa was not a centralized kingdom, but a loose confederation of small coastal states with networks of influence in the Annamite cordillera. This confederation included people not only speaking Chamic, but also Mon-Khmer languages. Chamic linguistic groups living on or near the coast and thus likely to have been an integral part of Cham polities include the Cham, Chru, Raglai, Haroi. Mon-Khmer groups like the Koho, Sre

and Hre were possibly in similar situation. Chamic Rhade and Jarai, and Mon-Khmer Bahnar and Sedang, spoken farther in the Highlands, may have been part of Cham political and economic networks, but were probably not under direct Cham control. It is difficult to evaluate the degree of direct contact between speakers of these diverse languages, but there is historical evidence that Cham kings married women from neighboring ethnic groups, suggesting that there were at least important political alliances. A case in point is the legendary king Po Klong Garai, who is said to have been of mixed Cham, Chru and Koho background.

During the heyday of classical Champa, there was a Cham presence throughout Southeast Asia. Cham traders regularly visited the major ports of the region, from Southern China to Java, possibly leading to the formation of more or less permanent diaspora, and matrimonial alliances with royal families of Insular Southeast Asia and Vietnam are well-documented. There is also evidence of networks of alliance and influence between Cham and Khmer factions, especially between the 10th and the 13th centuries, and it is likely that Cham artisans and mercenaries were present in Cambodia during that period. The existence of these diaspora could have opened the door to the establishment of Cham refugees outside the traditional territory when Cham polities started losing ground to the Vietnamese. The first significant Tsat settlements on Hainan could be dated to the fall of Indrapura, the most important political center However, the fall of Vijaya, a Cham political center located in Bình Định province, in 1471, seems to have been a more important turning point. A sizeable number of refugees escaped to Cambodia, leading to stable Cham-speaking settlement in that country. It has also been proposed that the Acehnese may have migrated from Champa at that time: although there is no evidence of a large scale Cham migration to Aceh, the *Sejarah Melayu*, or Malay chronicles, mention the foundation of the Acehnese dynasty by a Cham prince. In Central Vietnam, the fall of Vijaya seems to have caused Indianized culture to recede significantly: a direct linguistic consequence of this cultural change is the abandonment of Sanskrit.

After the fall of Vijaya, Cham political centers moved south to Kauṭhāra (Khánh Hòa province) and Pāṇḍuranga (Ninh Thuận province). In the following centuries, the de-Indianization of these polities was accompanied by a rise of Malay influence, as attested by a new type of Cham historical epics, the *akayet*, that closely mirror the style and some elements of the plot of the Malay *hikayat*. Malay influence also manifested itself through the growing influence of Islam, which may have been marginally present in Cham communities from the 10th century on. A similar cultural shift took place in Cambodia, where the Cham were influenced by local Muslim populations of Javanese and Malay origin.

From a linguistic point of view, Islamization led to the introduction of Arabic loanwords in Cham. Besides numerous religious loanwords, Arabic loans include more colloquial words like *ală?* 'alcohol' from *arak*, *katap* 'book' from *kitab*, *rapap* 'lute' from *rabāb*. It also led to the gradual adoption of the Arabic script. Although there is fluctuation in the terms used, one should distinguish the script Arabic scripts used to write Arabic for religious purposes (like akhăr pani), which can be more or less stylized and seem to have developed early on, from the Arabic-based script used to write Malay, akhăr ça̱wa (or jawi), which was adapted to use Cham more recently (Figure 5).

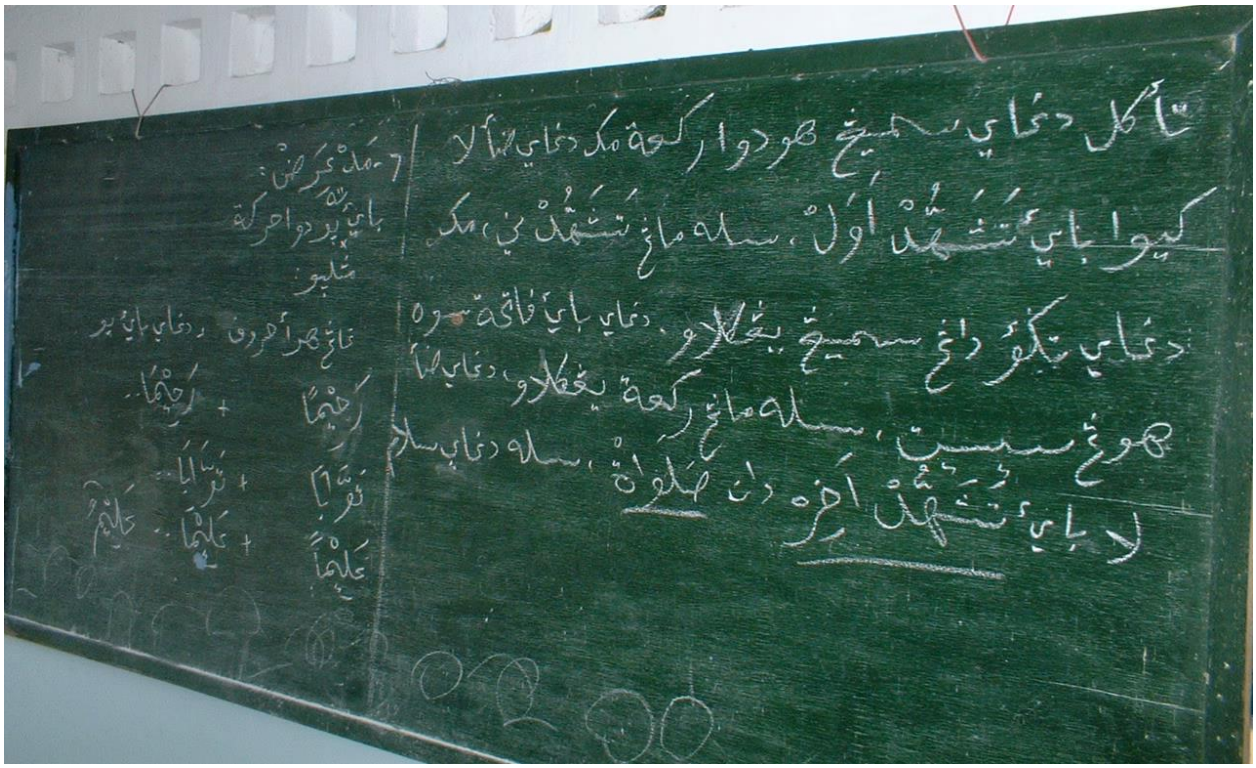


Figure 5: Example of jawi script, Châu Đốc, Vietnam

Arabic-based scripts did not supersede the Indic script, however. A new script called *akhăr srah* seems to have developed in that period. It is first attested in a 16th century inscription at Po Rome temple, but could have been in use for longer. From the 17th century, the modern Cham latan leave manuscripts compiling historical events, contracts, rituals and traditional medicine that started circulating in Cham territories were written in *akhăr srah*. They are still preserved (and recopied in notebooks) by contemporary Cham communities and some are consigned in French libraries. Interestingly, *akhăr srah* has evolved in slightly different ways in Vietnam and Cambodia (compare Figures 5 and 6).

Figure 6: Akhăr srah manuscript, Scarborough collection (Cornell University), manuscript 47a, Akayet Inra Patra

Figure 7: Akhăr srah sample from Kompong Chhnang province, Cambodia: The Qur'an with Cham commentaries

From the 15th to the 19th century, Cham gradually lost ground to the Vietnamese. Kauṭhāra (Khánh Hòa province) fell in 1651 and the remaining Cham state of Pāṇḍuranga, located in Ninh Thuận province, was gradually vassalized by the Nguyễn dynasty, to be finally integrated into their kingdom in 1832. There is good evidence of frequent movement between Cham communities in Vietnam and Cambodia during that period (and even all the way to Siam). One example of contact between Cambodian and Vietnamese Cham is the revolt of the Katip Sumat, a Cambodian Cham religious leader who led a jihad against Nguyễn rule in 1832-1834. After this movement and the ensuing Ja Thak Wa uprising were crushed in 1835, many Cham fled to Cambodia. Until the establishment of the French in Cochinchina in 1862, the Vietnamese court then used Cham networks to extend its influence in Cambodia and established Cham military settlements to control areas of the Mekong Delta (Tây Ninh and Châu Đốc). We do not have a full picture of Cham population movements in Cambodia and Southern Vietnam at that time, but what seems clear is that the relatively important dialectal (and religious) diversity found among the Cham of Cambodia is likely due to the various waves of Cham refugees and settlers that established communities in Cambodia from the 15th to the 19th century.

A characteristic of modern Cham dialects in the large number of words that have been borrowed from the national languages spoken around them. The Cham of Cambodia have borrowed words from Khmer, a state of affair that was already well under way at the beginning of the 20th century according to Aymonier and Cabaton 1906's dictionary. Khmer loanwords in Cambodian Cham dialects now include words as basic as *ɲ̊it* 'he/she' from គាត់ *koat*, *mɨjɛj* 'to say' from *nijɛj* និយាយ, or *tanot* 'palm sugar' from ត្នោត *tnaot*. Vietnamese loanwords into Eastern Cham were less numerous than Khmer loanwords in Cambodian Cham at the beginning of the 20th century, but the integration of Cham speakers into the Vietnamese state since World War II has changed this. Eastern Cham now has countless Vietnamese borrowings, including words as basic as *co* 'place' from Vietnamese *chỗ*, *jadin* from *gia đình*, or *pʰaj* 'ought to', from *phải*. Cham dialects have also borrowed a certain number of French words, like *aŋklɛ* 'English' from *Anglais* or *plih* 'police' from *police*.

Chamic migrations overseas

We must add to this survey two languages that are closely linked to Chamic languages but are not spoken in Mainland Southeast Asia proper: Tsat, which is spoken in Hainan, and Acehnese, which is spoken on the northern tip of Sumatra.

It is difficult to establish when the first Chamic speakers reached Hainan and how much coming and going there was between the coast of Champa and Hainan, but according to Chinese dynastic sources a sizeable group of Cham landed there in 986 and paid tribute to the emperor. As this date roughly coincides with the conquest of the northernmost Cham territories by the Vietnamese, it is possible that these were refugees from the northern polity of Indrapura. Chinese records also indicate that another group of Chamic speakers arrived in Hainan in 1486, which could have happened in the wake of the fall of Vijaya in 1471.

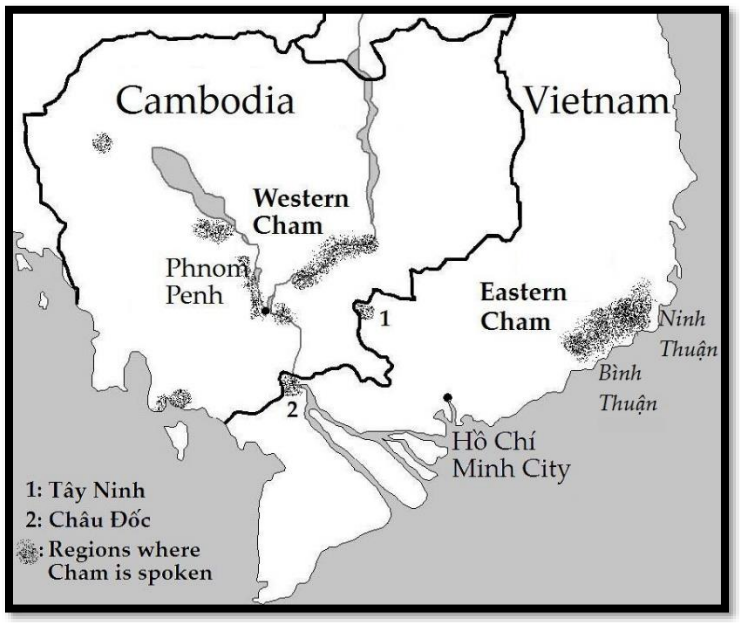
In any case, Tsat is a classic case study in convergence under the influence of intense multilingual contact. The language that has the most impact on Tsat is Chinese (probably a variant of Southwestern Mandarin) but there was also some linguistic contact with Hlai (or Li) a

Tai-Kadai indigenous language of Hainan. The linguistic evidence most obvious to the non-specialist is the inundation of Chinese loans. Not only has roughly a quarter of the lexicon been borrowed, but the borrowings are found in every category: nouns, verbs, adjectives, classifiers, adverbs, and prepositions. Even more revealing about the intensity and nature of the contact is the phonology and the syntax. The language spoken by the Chamic speakers arriving on Hainan was disyllabic and lacked tones, yet in contact with the languages of Hainan in general and with Chinese in particular, it became monosyllabic and fully tonal. The syntax was also greatly restructured: for instance, while the basic word order of Mainland Chamic languages is Subject-Verb-Object, an order that is still possible in Tsat, the Subject-Object-Verb order of Chinese is now prevalent.

We have very little evidence to speculate about the nature of the contact between Tsat speakers and the native and Chinese populations of Hainan, but the genetic evidence (both mitochondrial DNA and Y chromosomes) shows that modern-day Tsat speakers uncontroversially pattern with local Hainanese population rather than with the Cham of Vietnam. This suggests either a dramatic assimilation of local populations to Tsat or, more likely, heavy intermarriage over the centuries.

The similarity between Acehnese and Chamic languages has first been noted by Niemann in 1891 and has since been the subject of much scholarly debate. Although the vocabularies of these languages have borrowed extensively for their neighbors, making the job of the historical linguist especially difficult, several innovations show that they belong to a single branch. Furthermore, there are no innovations that allow us to group Acehnese with a specific subgroup of Chamic, or to set Acehnese apart from the rest of the family. In short, regular sound correspondences do not allow us to date precisely the geographical separation of Chamic and Acehnese, which could have happened any time after the split between Malayic and Chamic languages in Borneo. Because of this, scholars have turned to historical evidence and loanwords, and have developed two migration theories: the first is that the ancestors of the Acehnese left central Vietnam for northern Sumatra in the 15th century, while the other is that the geographical split between the two groups probably occurred at the time of the Malayo-Chamic migrations out of Borneo or shortly thereafter. Proponents of the first view argue that there are shared Mon-Khmer loanwords in Acehnese and Chamic and that Malay chronicles, the *Sejarah Melayu*, mention that the Aceh dynasty was founded by a Cham prince in 1471, the year the Cham polity of Vijaya fell to the Vietnamese. Proponents of the other view argue that the loanwords shared by Acehnese and Chamic are only attested in the Bahnaric branch of Mon-Khmer. As Bahnaric languages are in direct contact with Chamic, these loanwords could thus be Chamic borrowings in Bahnaric. They further point out there are no historical records of a large-scale Chamic migration to Aceh in the 15th century.

Contemporary Chamic languages



Map 4: Current geographic location of Cham dialects

Nowadays, there are close to 1.3 million speakers of Chamic languages in Southeast Asia. Detailed populations figures are given in Table 6. In Cambodia, this includes 330,000 speakers of Cham and Jarai. While the large majority of Chamic speakers, live in Vietnam, there are more speakers of Cham proper in Cambodia than Vietnam. There is also an estimated 4,500 speakers of Tsat on the island of Hainan in China. Finally, if Acehnese is Chamic, it is by far the largest Chamic language, with more than 4 million speakers.

Table 6: Population of Chamic speakers in the 2009 Vietnamese census and the 2008 Cambodian census

	Vietnam (2009)	Cambodia (2008)
Jarai	411,275	26,335
Cham	161,729	204,080
Eastern	*105,000*	
Western	*30,000*	
Haroi[2]	*25,000*	

[2] Haroi speakers are classified as Cham in the Vietnamese census, but their language is actually a Chamic language distinct from Cham.

Rhade	331,194	
Raglai	122,245	
Chru	19,314	

The Cham living in Vietnam speak two main dialects: Eastern Cham, which is spoken on the South-Central coast, and Western Cham, which is used by communities established in Tây Ninh and Châu Đốc, in the Mekong delta. Since the latter is very similar to Cham varieties spoken in Cambodia, it is customary to group them together and to designated them all as Western Cham. However, Cambodian Cham is understudied and seems to be much more diverse than previously assumed. A systematic dialectal survey of Cambodian Cham is urgently needed. We should also emphasize that besides the 200,000 speakers of Cham, Cambodia is also home to about 50,000 other Muslims who speak Khmer and are the descendants of Malay and Javanese who have been living in Cambodia for centuries and of Cham linguistically assimilated to Khmer.

Several features distinguish modern Cham dialects from other Chamic languages. The most conspicuous is the use of *register*, the phonetic property that distinguishes syllables like *tom* 'to meet' and (*ha)ţom* 'how many'. Register is a combination of pitch, vowel quality and voice quality realized on vowels, that serves to distinguish syllables that were previously distinguished by consonant voicing (i.e. the vocal fold vibrations that distinguish p from b, or t from d). While Cham dialects all have registers, Highland Chamic languages, like Jarai and Rhade, have preserved their original voicing contrast (although some Jarai dialects are reported to also have register). Some Chamic languages spoken close to the coast of central Vietnam, like Southern Raglai and Haroi, do have registers, but they take very different forms: the registers of Haroi, for instance, are realized as dramatic differences in vowels similar to the difference between the a-series and the o-series in Khmer. At the other extreme, Tsat has gone even further than Cham dialects and has transformed its registers into a full-fledged tone system similar to those of Thai, Chinese or Vietnamese.

Even within Cham itself, there are important differences between dialects. A first major difference between Cham dialects is their lexicon. Some basic words used in Western Cham dialects are different from those used in Eastern Cham. For instance, the formal word for "I" is *hulĭn/lĭn* in Western Cham, but *ţahḷăʔ/hḷăʔ* in Eastern Cham, and "body" *p̥aţăn* is *rup paphap/rup phap* in Eastern Cham. Second, some words have developed different meanings in different dialects. Thus, *ţrĕj* is a classifier for animals in Western Cham, but has come to mean "we, us" in Eastern Cham, and *akʰăr,* which only means 'script' in Western Cham, has also taken the meaning of 'language', along with *săp*, in Eastern Cham. Moreover, there are, as mentioned above, important lexical differences caused by the introduction of loanwords from different national languages (Khmer and Vietnamese) in the two communities, a phenomenon that is accentuated by the massive introduction of Arabic borrowings in more orthodox Muslim communities.

Significant dialectal differences are also found in sound patterns. For instance, the consonant s- of Western Cham (both in Cambodia and Mekong delta) is often pronounced as t^h-

in Eastern Cham. Thus, *saŋ* 'house' and *asɛh* 'horse' are pronounced as *t^haŋ* and *t^hɛh* in Eastern Cham. Another characteristic of Eastern Cham is that its diphthongs (its complex vowels), tend to be reduced to long monophthongs (simpler vowels). Examples include words like *seam* 'beautiful' or *juon* 'Vietnamese' that tend to be produced as *sa:m* and *ju:n* in Eastern Cham. Mergers between final consonants also distinguish the two major dialects. While Western Cham preserves the difference between final –l, -n and –r (although final –r is often lost), Eastern Cham merges these three consonants in word final position. Thus, Western Cham *ak^{h}ăr* 'script', has become *k^{h}ăn* in Ninh Thuận and *k^{h}ăl* in Bình Thuận.

The difference between Western and Eastern Cham dialects that has received the most attention in the literature is the phonetic realization of register (the above-mentioned combination of pitch, vowel quality and voice quality that distinguishes the syllable *p^{h}a* 'thigh' and the last syllable of *pap^{h}a~rap^{h}a* 'share, distribute'). While initial work suggested that Western Cham was mostly marking register with vowel quality whereas Eastern Cham was relying on pitch, recent work suggest that differences may not be as great as previously assumed: all dialects make use of pitch, vowel quality and breathiness to some extent, but to varying degrees. Eastern Cham seems to make more use of pitch than the Western dialects, which might be due to contact with tonal Vietnamese, but pitch is also an important property of Western Cham register. What is crucial here is that different Cambodian Cham dialects exhibit more variation in register than previously assumed, which deserves more research. There are also some differences in sound patterns that do not match the Western and Eastern Cham divide. For example, in most Cham dialects, the prevocalic r- of Ancient Cham, which is preserved in Ninh Thuận, has become a velar sound (g or ɣ) in other dialects. Thus, *prah* 'rice' is pronounced *pgah* in Bình Thuận and in Western Cham.

A last, important difference between Eastern and Western Cham is the degree of linguistic variation within communities. Western Cham seems more diverse that previously assumed, but within each community, the gap between formal and informal speech seems relatively limited. In Eastern Cham, by contrast, there is an enormous difference between formal speech, which is more conservative and preserves characteristics of the written language, and informal speech, in which the realization of various sounds has evolved in new directions. The most salient feature of informal Eastern Cham speech is the systematically monosyllabic character of words. The formal sentence *limɔ tamɨ saŋ* 'the cow enters the house' is normally produced as *mɔ mɨ t^{h}aŋ* in casual speech. Monosyllabization is also common in Western Cham, but it does not have the same systematic character as in Eastern Cham informal speech.

Overall, these differences have caused significant divergence between Cham dialects, to the point that mutual intelligibility is not always easy, even if similarities between varieties are obvious.

The future of Chamic languages

We can be relatively optimistic about the prospect of Chamic languages in the short term. In all Chamic-speaking communities of Vietnam and Cambodia, language transmission to younger generations is happening normally. Assimilation to national languages is rare, and

mostly limited to families who have chosen to live in larger cities, where daily contact with the heritage language is difficult. However, with the spread of formal education in Khmer and Vietnamese and the establishment of a growing proportion of Chamic speakers in urban centers, long term scenarios are more difficult to predict. Some communities of the Central Vietnamese Highlands also have to deal with the massive arrival of ethnic Vietnamese settlers that forces them into a minority status.

Large Highlands Chamic languages, like Jarai and Rhade, have well-established latin-based alphabets that are widely used within the community. The status of written Cham proper is more precarious, but efforts to promote the Cham script (akhăr srah) in educational programs in both Vietnam and Cambodia, and its growing use in electronic media are positive elements that could have a major impact on language maintenance. A crucial factor in the revitalization of the Cham script is the development of written standards shared by speakers of various Cham dialects, not only within each country, but also throughout Southeast Asia.

Selected references

Linguistic sources

Aymonier, É. F. (**1889**). *Grammaire de la langue chame* (Imprimerie coloniale, Saigon).

Aymonier, É. F., and Cabaton, A. (**1906**). *Dictionnaire cam-francais* (E. Leroux, Paris).

Blood, D. L. (**1967**). "Phonological Units in Cham," Anthropological Linguistics **9**, 15-32.

Blood, D. W. (**1980**). "The Script as a Cohesive Factor in Cham Society," in *Notes from Indochina on Ethnic Minority Cultures*, edited by M. T. Gregerson, Dorothy M., pp. 35-44.

Blust, R. (**2010**). "The Greater North Borneo Hypothesis," Oceanic Linguistics **49**, 44-118.

Brunelle, M. (**2009**). "Contact-induced change? Register in three Cham dialects," Journal of Southeast Asian Linguistics **2**, 1-22.

Durie, M. (**1985**). *A grammar of Acehnese on the basis of a dialect of North Aceh* (Foris Publications, Dordrecht; Cinnaminson, NJ).

Edmondson, J., and Gregerson, K. (**1993**). "Western Cham as a Register Language," in *Tonality in Austronesian Languages*, edited by J. Edmondson, and K. Gregerson (U of Hawaii Press, Honolulu), pp. 61-74.

Headley, R. K. (**1991**). "The Phonology of Kompong Thom Cham," in *Austroasiatic languages essays in honour of H. L. Shorto*, edited by J. Davidson, pp. 105-121.

Lee, E. W. (**1966**). "Proto-Chamic phonologic word and vocabulary," (Bloomington).

Moussay, G. (**1971**). *Dictionnaire cam-vietnamien-français* (Trung-tâm Văn hoá Chăm, Phan Rang).

Moussay, G. (**2006**). *Grammaire de la langue Cam* (Les Indes Savantes, Paris).

Phú, V. H., Edmondson, J., and Gregerson, K. (**1992**). "Eastern Cham as a Tone Language," Mon Khmer Studies **20**, 31-43.

Shorto, H. L. (**1975**). "Achinese and Mainland Austronesian," Bulletin of the School of Oriental and African Studies **38**, 81-102.

Sidwell, P. **(2006).** Dating the separation of Acehnese and Chamic by etymological analysis of the Aceh-Chamic lexicon. Mon-Khmer Studies 36.187-206.

Thurgood, G. (**1999**). *From Ancient Cham to Modern Dialects : Two Thousand Years of Language Contact and Change* (University of Hawai'i Press, Honolulu).

Thurgood, G., Thurgood, E., and Li, F. (**2015**). *A Grammatical Sketch of Hainan Cham: History, Contact, and Phonology* (Mouton De Gruyter, Berlin-Boston).

Wouk, Fay and Malcolm Ross **(2002)**, *The history and typology of western Austronesian voice systems.* Pacific Linguistics : Australian National University.

Inscriptions and scripts

Corpus of the Inscriptions of Campā. http://isaw.nyu.edu/publications/inscriptions/campa/about.html, retrieved 17-04-2015.

Brunelle, M. (**2008**). "Diglossia, Bilingualism, and the Revitalization of Written Eastern Cham," Language Documentation and Conservation **2**, 28-46.

Coedès, G. (**1939**). "La plus ancienne inscription en langue cham (Inscription rupestre de Đông Yên Châu, province de Quảng Nam, Annam)," in *New Indian Antiquary: A volume of Eastern and Indian Studies in honour of F.W. Thomas, C.I.E.*, edited by S. M. Katre, and P. K. Gode (Karnatak Publishing House, Bombay), pp. 46-49.

Schweyer, A.-V. (**1999**). "Chronologie des inscriptions publiées du Campā: Études d'épigraphie cam I," Bulletin de l'École française d'Extrême-Orient **86**, 321-344.

Historical and genetic evidence

He, J.-D., Peng, M.-S., Quang, H. H., Dang, K. P., Trieu, A. V., Wu, S.-F., Jin, J.-Q., Murphy, R. W., Yao, Y.-G., and Zhang, Y.-P. (**2012**). "Patrilineal perspective on the Austronesian diffusion in Mainland Southeast Asia," PloS one **7**, e36437.

Li, D.-N., Wanguan-Chao, Yang, K., QinZhen-Dong, Lu, Y., LinXue-Jing, and Li, H. (**2013**). "Substitution of Hainan indigenous genetic lineage in the Utsat people, exiles of the Champa kingdom," Journal of Systematics and Evolution **51**, 287-294.

Peng, M.-S., Huy Ho, Q., Khoa Pham, D., An Vu, T., Wang, H.-W., Yao, Y.-G., Kong, Q.-P., and Zhang, Y.-P. (**2010**). "Tracing the Austronesian Footprint in Mainland Southeast Asia: A Perspective from Mitochondrial DNA," Molecular Biology and Evoluation **27**, 2417-2430.

Southworth, William A. **(2001)**. *The origins of Campa in Vietnam: a preliminary overview*. PhD thesis, London: SOAS.

Weber, N. (**2014**). *Histoire de la diaspora cam* (Les Indes Savantes, Paris).