Doc Type: Working Group Document
Title: Response to proposal to encode Western Cham in the UCS
Source: Alberto Pérez Pereiro, Jorge López Cortina, Leb Ke
Status: Individual contribution
Action: For consideration by UTC and ISO
Date: 2020-01-07

Executive Summary. This document is a response to L2/19-217R3, a proposal to encode Western Cham that removes some characters included in L2/16-198 (N4734). The removed characters are necessary to represent current Western Cham usage in schools and in printed works, and should be included in any proposal to represent Western Cham. Additionally, it is argued that a revision of the collation is needed.

Introduction. The use of Western Cham script has experienced an extraordinary increase in the last decade, expanding from a few dozen users in the Imam San community to several thousand users in several Cambodian provinces. This revival was made possible by the implementation of Western Cham literacy programs and the publication of a series of books and periodicals using Western Cham script (information on those publications is included as an appendix to this document). The fonts used to print these materials were created by Leb Ke in 2011 on the basis of non-Unicode fonts he had been developing since 2005. The source materials for these fonts were extant Western Cham manuscripts. In the absence of a Western Cham block, these fonts were based on the Eastern Cham block. The initial proposal for the encoding of Western Cham, L2/16-198 (N4734), includes the characters from these fonts, which are used in educational materials and other publications, and constitute a reference for a large majority of users of Western Cham script.

Proposal L2/19-217R3, in contrast, seeks to limit the encoding of Western Cham to the characters used in one of the several preexisting spelling systems (historically, there has not been a commonly accepted spelling system in Western Cham). In particular, the proposal follows the system supported by the current leadership of the Krom Kan Imam San, which the proposal seems to equate to the “Western Cham community,” but which in fact is a religious group that constitutes less than a 10% of Western Cham speakers in Cambodia. Leaving aside the political and religious implications of this choice, this is simply not practical, as it ignores current majority usage as well as the totality of Western Cham printed materials currently deposited in the National Library of Cambodia. Note that including all the characters used in current fonts does not represent a problem for those members of the Krom Kan Imam San who choose to use their own spelling system. They can simply avoid using characters not required in their system. Removing these characters, however, prevents Western Cham publishers from switching from the current Eastern Cham block based fonts to Western Cham block fonts when printing in Western Cham. As the teaching of Western Cham script is done with books using the characters that L2/19-217R3 intends to remove, we could find ourselves in a situation where Western Cham speakers trying to type in Western Cham would need to resort to fonts designed using the Eastern Cham block.
As a separate issue, the collation proposed in L2/19-217R3 runs contrary to the one used in the Cham-Khmer dictionary published in 2011 by the same Krom Kan Imam San group that endorses L2/19-217R3. This is briefly discussed in page 22 of this document.

**Reertoire.** L2/19-217R3 removes or omits seven characters (three final consonants, one final consonant sign, two modified nasal consonants and one vowel sign) found in L2/16-198 (N4734). These characters are all in use in contemporary Western Cham writing and should be included in any proposal to encode Western Cham.

What follows is an explanation of the reasons why these characters should be included.

### 1.0 Removales

#### 1.1 Final Consonants

<table>
<thead>
<tr>
<th>Character</th>
<th>Encoding</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ꨄ ꨀ꨻꨽</td>
<td>1E241</td>
<td>Final G</td>
</tr>
<tr>
<td>2 ꨃ ꨂ</td>
<td>1E23C</td>
<td>Final B</td>
</tr>
<tr>
<td>3 ꨪ ꨲ</td>
<td>1E23D</td>
<td>Final M</td>
</tr>
</tbody>
</table>

**1.1.1 Final G (1E241)**

L2/19-217R3’s reasoning in removing this character is:

*U+1E241 (FINAL G) is a borrowed sound into Eastern Cham from the surrounding Vietnamese. There is no such sound in Khmer and so Western Cham has no need for this sound. Thus we remove the character and U+1E240 (ꨄ FINAL K) may be used instead.*

The presence or absence of this sound in Khmer is irrelevant to the writing of Cham, which is a different language, from a different language family, and with a different phonology. The sound is present in contemporary Western Cham pronunciation.

Final G is used in contrast to Final K, with a change of both meaning and pronunciation.

<table>
<thead>
<tr>
<th>Character</th>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
<th>Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final G</td>
<td>ꨄꨪ</td>
<td>[ kəkʰ ]</td>
<td>coin</td>
<td>1E241</td>
</tr>
<tr>
<td>Final K</td>
<td>ꨄꨳ</td>
<td>[ kəʔ ]</td>
<td>to tie</td>
<td>1E240</td>
</tr>
</tbody>
</table>

Examples of the use of Final G are shown in Figures 1, 2 & 5.
1.1.2 Final B (1E23C) and Final M (1E23D)

L2/19-217R3’s reasoning in removing these characters is:

*Of these U+1E23C (FINAL B) and U+1E23D (FINAL M) have been found to be redundant by the Western Cham community in Cambodia. U+1E23C (FINAL B) may be represented by U+1E247 (ꨧ FINAL P) and U+1E23D (FINAL M) by U+1E24C (ꨩ SIGN FINAL M).*

There are three problems with this argument.

A) L2/19-217R3 suggests that there is a unified community of Cham speakers that agrees on the spelling of their language, when in fact there are multiple ways of representing the language in script.

B) Redundancy does not exclude a character from a script. Unicode blocks contain characters no longer in common or contemporary use because they are still needed to represent the language as written by different communities or in earlier stages. For example, the Khmer Unicode block contains the characters ឝ and ឞ although they are no longer in common use.

C) Final B and Final M are not redundant and remain in use. Final B contrasts with Final P and the two are not interchangeable.

Final P can cause diphthongization of a preceding vowel. Final B does not.

<table>
<thead>
<tr>
<th>Character</th>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
<th>Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final P</td>
<td>ꨧꨧ</td>
<td>[ søʔ ]</td>
<td>smoke</td>
<td>1E247</td>
</tr>
<tr>
<td>Final B</td>
<td>ꨧ꨼</td>
<td>[ sap ]</td>
<td>sound</td>
<td>1E23C</td>
</tr>
</tbody>
</table>

Neither are Final M and Final Sign M interchangeable.

Final M is used when the preceding vowel is long, while the Sign Final M is used when the preceding vowel is short. This means that the word ‘Cham’, for instance, is written with a Sign Final M because the ‘a’ is short. However, the Cambodian Province of Kampong Cham is written using the Final M, because that ‘a’ is long.

<table>
<thead>
<tr>
<th>Character</th>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
<th>Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final M</td>
<td>ꨧꨩ</td>
<td>[çaːm]</td>
<td>(Kampong) Cham (Province)</td>
<td>1E23D</td>
</tr>
<tr>
<td>Consonant Final M Sign</td>
<td>ꨧ</td>
<td>[çam]</td>
<td>Cham (ethnicity)</td>
<td>1E24C</td>
</tr>
</tbody>
</table>

This distinction is also made with native Cham vocabulary:

<table>
<thead>
<tr>
<th>Character</th>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
<th>Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final M</td>
<td>ꨧꨩ</td>
<td>[ taːm ]</td>
<td>to meet</td>
<td>1E23D</td>
</tr>
<tr>
<td>Consonant Final M Sign</td>
<td>ꨧ</td>
<td>[ təm ]</td>
<td>to transplant</td>
<td>1E24C</td>
</tr>
</tbody>
</table>

Examples of the use of Final B and Final M are shown in Figures 3 - 10.
1.2 Consonant Sign Final NG (1E243)

L2/19-217R3 removes Sign Final NG (1E243) and proposes that Vowel Sign OE (1E22E) perform both the duties of a vowel and that of a final consonant in order to “avoid confusion”.

Sign Final NG and Vowel OE may be identical in some handwriting but they are not considered the same and are also written distinctly in many styles.

These signs may even co-occur. In the following example, the word [ciəŋ] meaning ‘to be born’, the lower diacritic is the Vowel OE and the upper diacritic is the Final NG.

Examples of the use of concurrent use of Sign Final NG (1E243) and Vowel Sign OE (1E22E) are shown in Figures 10 & 11.
2.0 Omissions

2.1 Modified Nasal Consonants {Figures 10-17}

Everson and Cunningham proposed a total of 8 nasal consonant characters representing 4 nasal sounds followed by either the vowel [ɨʔ] or [aʔ].

<table>
<thead>
<tr>
<th>Character</th>
<th>Pronunciation</th>
<th>Name</th>
<th>Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>ꨊ</td>
<td>[njiʔ]</td>
<td>NGUE</td>
<td>1E20A</td>
</tr>
<tr>
<td>ꨋ</td>
<td>[naʔ]</td>
<td>NGA</td>
<td>1E20B</td>
</tr>
<tr>
<td>ꨐ</td>
<td>[niʔ]</td>
<td>NHUE</td>
<td>1E210</td>
</tr>
<tr>
<td>ꨑ</td>
<td>[naʔ]</td>
<td>NHA</td>
<td>1E211</td>
</tr>
<tr>
<td>ꨒ</td>
<td>[niʔ]</td>
<td>NUE</td>
<td>1E217</td>
</tr>
<tr>
<td>ꨓ</td>
<td>[naʔ]</td>
<td>NA</td>
<td>1E218</td>
</tr>
<tr>
<td>ꨔ</td>
<td>[miʔ]</td>
<td>MUE</td>
<td>1E21F</td>
</tr>
<tr>
<td>ꨕ</td>
<td>[maʔ]</td>
<td>MA</td>
<td>1E220</td>
</tr>
</tbody>
</table>

L2/19-217R3 does not include either NGA or NHA (indicated in grey in the above table). These letters are in use and are necessary for writing Western Cham. The reason for this omission may be that some writers use the sign Takay Klak (1E235) to modify the letters NGUE and NHUE in contexts where other writers would use NGA and NHA. This is not a universal practice. Writers who use NGA and NHA make a distinction between these characters and NGUE and NHUE plus Takay Klak.

Takay Klak is used to modify the tenseness of a consonant (tense consonants become lax and lax consonants become tense). In contrast, the letters NGA and NHA are used not to modify the consonant tenseness but to alter the following vowel sound.

The Cham Orthography Guide published in 2011 makes a point that NGA and NHA are not simply NGUE and NHUE with the addition of Takay Klak. There are many examples of Takay Klak co-occurring with these modified nasal consonants, demonstrating that one is not a substitute for the other.

Examples of the use of NGA, NHA and Nasals Consonants co-occurring with Takay Klak are shown in Figures 12 - 21.

2.2 Vowel Sign U (1E22D) {Figure 18}

L2/19-217R3 also omits Vowel Sign U, which was present in Everson and Cunningham (16-198). This vowel sign is present in Eastern Cham and is used when transcribing Eastern Cham into Western Cham for the purposes of language learning or comparison. It is also used by speakers of some dialects of Western Cham to reflect their own pronunciation of certain words. The Cham Language Advisory Committee (CLAC) believes this vowel sign should be retained in order to facilitate these activities. Because the original shape of the vowel sign is similar to Takay Klak, the committee has proposed an alternate shape in order to avoid confusion.

Figure 22 shows the CLAC’s decision as published in Mukva #11 in 2018.
Figure 1: Final G (1E241) (Script Primer).
Figure 2: Final G (1E241) (Rediscovering Cham Heritage).

Figure 3: Final B (1E23C) The scribe’s writing style creates some ambiguity as this word initially appears to be written with the symbol Takay Kak. However, this would mean the word is pronounced [pha:paʔ], but this is not a word. Instead it is pronounced [pha:p], meaning ‘people’ – written with a Final B. These ambiguities must be resolved on a case by case basis by the researcher studying the manuscript. (Buraq Manuscript).

Figure 4: Final M (1E23D) The scribe’s writing style creates some ambiguity as this word initially appears to be written with the symbol Takay Kak. However, this would mean the word is pronounced [sam:supha:], but this is not meaningful in this context. Instead it is pronounced [samsupha:], meaning ‘foundation’ – written with a Final M in medial position. (Buraq Manuscript).
Figure 5: Final G (1E241) in red, Final B (1E23C) in yellow and Final M (1E23D) in blue (Mukva #10).
Figure 6: Final B (1E23C) (Mukva #5).
Figure 7: Final B (1E23C) in yellow and Final M (1E23D) in red (Discovering Cham Heritage).

Figure 8: Final M (1E23D) (Mukva #1).
Figure 9: Final M (1E23D) (Mukva #12).
Figure 10 - Co-occurrence of Sign Final NG (1E243) and Vowel Sign OE (1E22E) in green and Final B (1E23C) in yellow (Mukva #9).
Figure 11 - Co-occurrence of Sign Final NG (1E243) and Vowel Sign OE (1E22E) (Rediscovering Cham Heritage).
Figure 12 – Comparison of Nasal Consonants with vowel UE and modified Nasal Consonants with vowel A. The bottom line indicates that the Takay Klak symbol is not a substitute for the use of the modified nasal. Character NGA (1E20B) (Cham Orthography Guide).
Figure 13 - Character NGA (1E20B) (Folk Tales 1).

Figure 14 - Character NGUE (1E20A) with Takay Klak (1E235) (Rediscovering Cham Heritage).
Figure 15 - Character NHA (1E211). This word meaning ‘soul’ is pronounced [ɲɔːʋə] with a lax consonant. It is not pronounced [ɲɑːʋə] with a tense consonant. Although this appears at first glance to be a Takay Klak symbol, this is an artefact of the scribe’s writing. (Adab Safi Manuscript)

Figure 16 - Character NHA (1E211) This consonant is lax. (Rediscovering Cham Heritage).
Figure 17 - Character NHA (1E211) The consonant is lax. (Mukva #10).
Figure 18 - Character NA (1E218) (tense) with Takay Klak (1E235) and Character NHA (1E211) in red (Mukva #10).

Figure 19 - Character NUE (1E217) with Takay Klak (1E235) (Rediscovering Cham Heritage).
Figure 20 - Character NUE (1E217) with Takay Klak (1E235) in yellow and Character NHUE (1E210) (tense) with Takay Klak in red (Mukva #5).
Figure 21 - Character NHUE (1E210) with Takay Klak rendering it tense. (Rediscovering Cham Heritage).
Figure 22 – Vowel Sign U (1E22D) (Mukva #11).
**Other issues.** The collation proposed in L2/19-217R3 is presented as following Eastern Cham. It does not mention the existence of a traditional ordering, which is used in the 2011 Cham-Khmer dictionary published by the Krom Kan Imam San group and subsequently followed by other publications, teacher training materials, and textbooks. Note that this is a traditional order used by the Krom Kan Imam San group independently of any recent developments in Western Cham education or publications, even if recent publications follow it. It is also important to note that the Krom Kan Imam San are the same group cited in L2/19-217R as support for the proposal.

To give a sense of the relevance of this change it could be noted that the traditional name of Western Cham script is Ka-Kha. The symbols representing those sounds, which are the first two in the dictionary, appear in ninth and tenth place in the collation proposed in L2/19-217R3.

We believe that at least some additional discussion is in order.
Appendix A

Publications in Western Cham

This is a list of publications in Western Cham script since 2011. Some of them also include text in Khmer or in Khmer and English. Except for the dictionary, all of them use Leb Ke’s fonts and a spelling system that includes the characters L2/19-217R3 seeks to remove. The dictionary uses a different collation than the one proposed in L2/19-217R3. All of the books published by Emerging Markets Consulting and Naga Editions have ISBNs and are deposited in the National Library of Cambodia and in the library of the Buddhist Institute of Phnom Penh. The list includes only the first editions. The textbooks have been reprinted several times and there are over 5,000 copies in circulation, not counting photocopies. This may seem a modest number but it represents the most widely distributed publication in Western Cham by several orders of magnitude.

The book Rediscovering Cham Heritage in Cambodia: Language, Script, and Community gathers a series of articles on Western Cham language, script, and culture written by leading scholars in their respective fields and offers them in a careful, culturally sensitive graphic presentation. The book is published in a trilingual version, with side-by-side Khmer, English, and Cham text. A second edition was published in 2017. A chapter of this book is included in this document as Appendix B.

Mukva is a seasonal magazine distributed in Cham villages in nine Cambodian provinces since 2015. The run is usually 1,000 copies, higher for Ramadan editions.

Naga Editions is a publishing house that was established in 2014 to reprint old titles from Emerging Markets Consulting and to continue developing Western Cham materials. The authors of this document have collaborated with Naga Editions since its inception. The books published by Naga Editions include work by over two dozen Western Cham writers and translators.

Dictionary


Textbooks


Teacher Training


Literature


Academic


Periodical


Some publications by Naga Editions.
Appendix B


We believe it is interesting to include this article here because it shows Western Cham being used

a) in print,
b) side by side with the national language, Khmer, and with the international language of academic communication, English,
c) for purposes beyond the preservation of traditional religious texts,
d) to discuss Cham culture and history, and
e) to discuss topics that require modern terminology.

The production of this article, and other publications like this, shows a level of dedication and maturity on the part of Western Cham writers and translators that goes beyond anything that could be achieved through any individual effort.

The article includes several examples of use of the characters L2/19-217R3 seeks to remove.
Rediscovering Cham Heritage in Cambodia

LANGUAGE, SCRIPT, AND COMMUNITY
Rediscovering Cham Heritage in Cambodia

LANGUAGE, SCRIPT, AND COMMUNITY
Rediscovering Cham Heritage in Cambodia: Language, Script, and Community
Edited by Jorge López Cortina and Alberto Pérez Pereiro

naga® EDITIONS
ស្ថានីយ៍ភាសា - អន្តរបទ - Phnom Penh
www.nagaeditions.com

© Breogán Southeast Asia Co., Ltd.

ISBN: 978-9924-9041-0-6

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
This work may be copied and shared in its entirety and without modification, provided it is distributed without charge and credit to the authors is given. The articles and photographs included in this work are published by courtesy of their respective authors, who retain copyright.
Jorge López Cortina, PhD

Jorge López Cortina is an associate professor at Seton Hall University. He is interested in language education and revitalization and has worked on Spanish, Khmer, Chavacano, and Cham.

Alberto Pérez Pereiro, PhD

Alberto Pérez Pereiro is the general manager of Breogán Consulting. He has lectured in anthropology and linguistics at Arizona State University and Thammasat University, and conducts ethnographic research on Cham culture and Cambodian Islam.

Marc Brunelle, PhD

Marc Brunelle is an associate professor in Linguistics at the University of Ottawa, Canada. His research deals with the phonology and phonetics of tone and register in Mainland Southeast Asia, and with issues of language contact.
Graham Thurgood, PhD
Graham Thurgood is a specialist in the history of the languages of Southeast Asia and the author of From Ancient Cham to Modern Dialects. He is a professor of Linguistics at California State University, Chico.

Thành Phần, PhD
Thành Phần is a Cham scholar and native of Ninh Thuận Province in Vietnam. He is an associate professor at the University of Social Sciences and Humanities of Ho Chi Minh City and has conducted extensive research on Cham language, culture and history, including studies of Cham epigraphy.

Mohamed Effendy, PhD
Mohamed Effendy is a lecturer at the National University of Singapore. His research focuses on Cham social history as seen through its literature and documentary tradition.
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rediscovering Cham Heritage in Cambodia</td>
<td>Jorge López Cortina</td>
<td>14</td>
</tr>
<tr>
<td>A History of the Cham People and their Present Situation in Cambodia</td>
<td>Alberto Pérez Pereiro</td>
<td>16</td>
</tr>
<tr>
<td>The Historical Development of Chamic Languages</td>
<td>Marc Brunelle, Graham Thurgood</td>
<td>46</td>
</tr>
<tr>
<td>The Importance of the Cham Literary Heritage</td>
<td>Thanh Phan</td>
<td>98</td>
</tr>
<tr>
<td>Cham Traditional Literature Through its Manuscripts: Dalukal and Ariya</td>
<td>Mohamed Effendi</td>
<td>120</td>
</tr>
<tr>
<td>The Cham Heritage Expansion Program</td>
<td>Alberto Pérez Pereiro, Jorge López Cortina</td>
<td>148</td>
</tr>
</tbody>
</table>
ការអភិវឌ្ឍនភាសាចាម្ប៍ក្នុងបវត្តិសាស

The Historical Development of Chamic Languages

ការអភិវឌ្ឍនភាសាចាម្ប៍និងភាសាចាម្ប៍បានរីកសាយភាយនៅក្នុងតំបន់អាស៊ីអាគ្នយ៍យ៉ងធំទូលាយដលលាតសន្ធឹងពីចាម្ប៍ខាងលិចនៅកម្ពុជារហូតដល់កម្មចាយនៅតាមតំបន់ជួរភ្នំអាណ្មមក៏ដូចជាទៅលើចាម្ប៍ខាងកើតនៅតាមបណ្ដយឆ្នរកណ្ដលឆៀងខាងត្បូងប្រសវណមសព្វថ្ងធមិនស្ថិតក្នុងចង្កមភាសាចាម្ប៍ដលកំពងបានស់នៅកម្ពុជានិងវៀតណមសព្វថ្ង។ លើសពីនេះទៀតគក៏ឃើញមានសហគមន៍អន្តបវសចាម្ប៍នៅសកមា៉កមា៉លបាារបានអូស្តលីនិងសហរដ្ឋអាមរិចផងដរ។
Cham and the other modern Chamic languages are distributed over a considerable area in Southeast Asia: from Western Cham in Cambodia to Jarai in the Annamite Cordillera to Eastern Cham on the south-central coast of Vietnam to Tsat on the outskirts of Sanya City on Hainan island (we will use Tsat here to avoid any confusion the term Hainan Cham might create: although Tsat is a Chamic language, it is not a variety of the Cham language spoken in Cambodia and 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 Crawford, labels such as the “Malay of Champa,” but the classification of the Chamic languages came into question at the beginning of the 20th 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
Cham and its sister languages have long been in contact with languages of the Mon-Khmer family.
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.

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
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.

<table>
<thead>
<tr>
<th>Amis</th>
<th>Tagalog</th>
<th>Ilocano</th>
<th>Western Cham</th>
<th>Malay</th>
<th>Malagasy</th>
<th>Javanese</th>
<th>Maori</th>
<th>Hawai’ian</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>cecay</td>
<td>isa</td>
<td>maysa</td>
<td>sa, ha</td>
<td>satu</td>
<td>iray</td>
<td>siji</td>
<td>tahi</td>
</tr>
<tr>
<td>two</td>
<td>tosa</td>
<td>dalawa</td>
<td>dua</td>
<td>twa</td>
<td>dua</td>
<td>roa</td>
<td>loro</td>
<td>lua</td>
</tr>
<tr>
<td>three</td>
<td>tolo</td>
<td>tatlo</td>
<td>tallo</td>
<td>klāw</td>
<td>tiga</td>
<td>telo</td>
<td>telu</td>
<td>toru</td>
</tr>
<tr>
<td>four</td>
<td>sepat</td>
<td>apat</td>
<td>uppat</td>
<td>paːt</td>
<td>empat</td>
<td>efatra</td>
<td>papat</td>
<td>mhã</td>
</tr>
<tr>
<td>five</td>
<td>lima</td>
<td>lima</td>
<td>lima</td>
<td>limi</td>
<td>lima</td>
<td>dimy</td>
<td>limo</td>
<td>rima</td>
</tr>
<tr>
<td>six</td>
<td>enem</td>
<td>anim</td>
<td>inem</td>
<td>nâm</td>
<td>enam</td>
<td>enina</td>
<td>nem</td>
<td>ono</td>
</tr>
<tr>
<td>seven</td>
<td>pito</td>
<td>pito</td>
<td>pito</td>
<td>tuçuḥ</td>
<td>tujuh</td>
<td>fito</td>
<td>pita</td>
<td>whitu</td>
</tr>
<tr>
<td>eight</td>
<td>falo</td>
<td>walo</td>
<td>walo</td>
<td>tāpān</td>
<td>lapan</td>
<td>valo</td>
<td>wolu</td>
<td>waru</td>
</tr>
<tr>
<td>nine</td>
<td>siwa</td>
<td>---</td>
<td>siam</td>
<td>salipan</td>
<td>semblan</td>
<td>sivy</td>
<td>songo</td>
<td>iwa</td>
</tr>
<tr>
<td>ten</td>
<td>poloʔ</td>
<td>---</td>
<td>sangapololo</td>
<td>pluh</td>
<td>sepuluh</td>
<td>folo</td>
<td>sepuluh</td>
<td>tekau</td>
</tr>
<tr>
<td>house</td>
<td>luma</td>
<td>bahay</td>
<td>balay</td>
<td>saːŋ</td>
<td>rumah, balai</td>
<td>trano</td>
<td>omah</td>
<td>whare</td>
</tr>
<tr>
<td>dog</td>
<td>wacu</td>
<td>aso</td>
<td>aso</td>
<td>sāw</td>
<td>anjing</td>
<td>alika</td>
<td>asu</td>
<td>kuri</td>
</tr>
<tr>
<td>road</td>
<td>lalan</td>
<td>daan</td>
<td>dalan</td>
<td>çalan</td>
<td>jalan</td>
<td>lalana</td>
<td>dalan</td>
<td>ara</td>
</tr>
<tr>
<td>pig</td>
<td>fafoy</td>
<td>baboy</td>
<td>baboy</td>
<td>çapuy</td>
<td>babi</td>
<td>kisoa</td>
<td>babi</td>
<td>poaka</td>
</tr>
</tbody>
</table>

The Historical Development of Chamic Languages
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,
The Historical Development of Chamic Languages

Figure 1: The position of Chamic in Austronesian (the detailed subgroupings are still debated).
Cham and Rhade (representing the modern Western Cham and Rhade.

The similarities between the oldest stage of the Malayic languages, W estern Cham and Rhade, 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.

Maori, Fijian, Trukese and their sisters, 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.
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 illustrates words with the regular correspondences and phonetic similarities. The initial d- of Acehnese, Chru, and Northern Raglay regularly corresponds to the t- in Eastern Cham, a correspondence that makes phonetic sense; in a parallel way, the initial b- corresponds to p- 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.
**Table 2: Numerals in Malayo-Chamic.**

<table>
<thead>
<tr>
<th>Proto-Malayo-Polynesian</th>
<th>Malay</th>
<th>Proto-Chamic</th>
<th>Proto-Chamic</th>
<th>Western Cham</th>
<th>Rhade</th>
</tr>
</thead>
<tbody>
<tr>
<td>*esa; *isa</td>
<td>sa-, satu</td>
<td>*sa</td>
<td>sa, ha</td>
<td>sa</td>
<td>‘one’</td>
</tr>
<tr>
<td>*duha</td>
<td>dua</td>
<td>*dua</td>
<td>tва</td>
<td>dua</td>
<td>‘two’</td>
</tr>
<tr>
<td>*epat</td>
<td>ěmpat</td>
<td>*paːt</td>
<td>paːt</td>
<td>păt</td>
<td>‘four’</td>
</tr>
<tr>
<td>*lima</td>
<td>limi</td>
<td>*lima</td>
<td>limi</td>
<td>ema</td>
<td>‘five’</td>
</tr>
<tr>
<td>*enem</td>
<td>ěnam</td>
<td>*nam</td>
<td>nǎm</td>
<td>nǎm</td>
<td>‘six’</td>
</tr>
</tbody>
</table>

**Table 3: Some corresponding Chamic words.**

<table>
<thead>
<tr>
<th>Malay</th>
<th>Proto-Chamic</th>
<th>Acehnese</th>
<th>Chru</th>
<th>Northern Raglai</th>
<th>Formal Eastern Cham</th>
</tr>
</thead>
<tbody>
<tr>
<td>darah</td>
<td>*darah</td>
<td>darah</td>
<td>drah</td>
<td>darah</td>
<td>‘blood’</td>
</tr>
<tr>
<td>dara</td>
<td>*dara</td>
<td>dara</td>
<td>dra</td>
<td>dara</td>
<td>‘girl’</td>
</tr>
<tr>
<td>bulu</td>
<td>*bulɔw</td>
<td>buleso</td>
<td>blou</td>
<td>bilou</td>
<td>‘hair’</td>
</tr>
<tr>
<td>bulan</td>
<td>*bulaːn</td>
<td>buluan</td>
<td>ea bлаːn</td>
<td>ia bilaːt</td>
<td>pilan</td>
</tr>
<tr>
<td>baharu</td>
<td>*bahrɔw</td>
<td>baro</td>
<td>bɔrɔu</td>
<td>bahru</td>
<td>piriːw</td>
</tr>
<tr>
<td>tahun</td>
<td>*thun</td>
<td>thon</td>
<td>thun</td>
<td>thut</td>
<td>thûn</td>
</tr>
<tr>
<td>tangan</td>
<td>*tangaːn</td>
<td>---</td>
<td>таŋːаːn</td>
<td>таŋːn</td>
<td>таŋːn</td>
</tr>
</tbody>
</table>
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
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 Huỳnh 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,
Marc Brunelle and Graham Thurgood

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.

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.
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.

Table 4: Stress shift of Proto-Chamic

<table>
<thead>
<tr>
<th>Proto-Malayo-Polynesian</th>
<th>Malay</th>
<th>Proto-Chamic</th>
<th>Acchinese</th>
<th>Chru</th>
<th>Northern Raglai</th>
<th>Tsat</th>
</tr>
</thead>
<tbody>
<tr>
<td>*mamaq</td>
<td>mamah</td>
<td>*mamah</td>
<td>mamah</td>
<td>mumah</td>
<td>ma₃⁵</td>
<td>'jēga' 'mē' 'chew'</td>
</tr>
<tr>
<td>*qumah</td>
<td>huma</td>
<td>*huma</td>
<td>huma</td>
<td>humā</td>
<td>ma₃³</td>
<td>'hōng' 'mā' 'dry field'</td>
</tr>
<tr>
<td>*lima</td>
<td>lima</td>
<td>*lima</td>
<td>lima</td>
<td>limā</td>
<td>ma₃⁴</td>
<td>'līng' 'nō' 'five'</td>
</tr>
<tr>
<td>*panaq</td>
<td>panah</td>
<td>*panah</td>
<td>panah</td>
<td>panāh</td>
<td>na₅⁵</td>
<td>'pēm [m]' 'pē' 'shoot [bow]'</td>
</tr>
<tr>
<td>*baseq</td>
<td>basah</td>
<td>*basah</td>
<td>basah</td>
<td>basah</td>
<td>sa₃⁴</td>
<td>'tēd [m]' 'sē' 'wet; damp'</td>
</tr>
<tr>
<td>*paqit</td>
<td>pahit</td>
<td>*phit</td>
<td>phet</td>
<td>phiː?</td>
<td>phiʔ²⁴</td>
<td>'pī' 'pōk' 'bitter'</td>
</tr>
<tr>
<td>*taqu</td>
<td>tahu</td>
<td>*thaw</td>
<td>theo</td>
<td>thau</td>
<td>tiau²³</td>
<td>'ti' 'no' 'know'</td>
</tr>
<tr>
<td>*paqa</td>
<td>paha</td>
<td>*pha</td>
<td>pha</td>
<td>pha</td>
<td>pha³³</td>
<td>'pō' 'mō' 'thigh'</td>
</tr>
<tr>
<td>*taqun</td>
<td>tahun</td>
<td>*thun</td>
<td>thun</td>
<td>thun</td>
<td>thun³³</td>
<td>'tū' 'mō' 'year'</td>
</tr>
</tbody>
</table>
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 [ˈsæksəˌfɔːn]). 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.

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-
Finally, a number of other changes also occurred, likewise making the system more Mon-Khmer-like. For example, the typologically rare glottalized consonants ɓ, ð, and ʃ (as in Cham ɓu? ‘hair’, ðih ‘to lie down’ and ʃaw? ‘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
<table>
<thead>
<tr>
<th>Proto-Chamic</th>
<th>Vietnamese</th>
<th>Khmer</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>kow</em></td>
<td>'T' (singular)</td>
<td>'្ល្ិ'</td>
<td>'I' (familiar)</td>
</tr>
<tr>
<td><em>ʔadhlį’</em></td>
<td>'្ពេស'</td>
<td>'្ពេស'</td>
<td>'forehead'</td>
</tr>
<tr>
<td><em>dalam</em></td>
<td>'្ម៊ី; ្អាប៊ី'</td>
<td>'្ម៊ី'</td>
<td>'in; inside'</td>
</tr>
<tr>
<td><em>ʔama:</em></td>
<td>'្ឈ្ម្យេ'</td>
<td>'្ឈ្ម្យេ'</td>
<td>'child'</td>
</tr>
<tr>
<td><em>ʔapuy</em></td>
<td>'្សំ'</td>
<td>'្សំ'</td>
<td>'fingernail; claw'</td>
</tr>
<tr>
<td><em>ʔana:</em></td>
<td>'្ទ្ុក'</td>
<td>'្ទ្ុក'</td>
<td>'lake'</td>
</tr>
<tr>
<td><em>ʔadhu</em></td>
<td>'្ឈ្ម្យេ'</td>
<td>'្ឈ្ម្យេ'</td>
<td>'day; sun'</td>
</tr>
<tr>
<td><em>ʔadhuy</em></td>
<td>'មើងមូល'</td>
<td>'្មើងមូល'</td>
<td>'younger sibling'</td>
</tr>
<tr>
<td><em>-on</em></td>
<td>instrumental infix</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>ʔama</em></td>
<td>'្សំបូល'</td>
<td>'្សំបូល'</td>
<td>'father'</td>
</tr>
<tr>
<td><em>ʔapuy</em></td>
<td>'្រជ្ជ'</td>
<td>'្រជ្ជ'</td>
<td>'skin'</td>
</tr>
<tr>
<td><em>ʔana:</em></td>
<td>'្ឈ្ម្យេ'</td>
<td>'្ឈ្ម្យេ'</td>
<td>'fire'</td>
</tr>
<tr>
<td><em>ʔapuy</em></td>
<td>'្មេ'</td>
<td>'្មេ'</td>
<td>'moon'</td>
</tr>
<tr>
<td><em>ʔana:</em></td>
<td>'្មេ'</td>
<td>'្មេ'</td>
<td>'black'</td>
</tr>
</tbody>
</table>

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:
The Historical Development of Chamic Languages

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).
Cham enters history

As mentioned above, it is likely that in the 1st 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.

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
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
Figure 2: Moulding of the Đông Y ên Châu inscription, 4th century (from Cœdès 1939).

[1] Siddham ni yān nāga puñ putaw ya urāŋ spūy di ko
   Success! This god naga possession king who person gentle at 3p.s.
kurun ko jmāy labub nari svarggah ya urāŋ paribhū di ko
   ? 3p.s. jewel fall from sky who person insult at 3p.s.
kurun saribhā thān ko davan di naraka dyan tiyun 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'
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 Coedès (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 that 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 3: The Biên Hòa inscription (1421) EFEO estampage 263.
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, akhär rik ‘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 Thuân province.

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

\[\text{ni \ ti–k–u–h \ k–u–ba–v \ ri–om–n}\]

\[\text{ti–pa–y \ no–ga–ra–y \ u–la}\]

\[\text{a–a–i \ n \ a–ai–s \ pa–ai–b}\]
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 this 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 Kauthāra (Khánh Hòa province) and Pânduranga (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 *alā* 'alcohol' from *arak*, *kata* 'book' from *kitab*, *rap* 'lute' from *rabab*. It also led to the gradual adoption of the Arabic script.

Although there is fluctuation in the terms used, one should distinguish the Arabic scripts used to write Arabic for religious purposes (like *akhir pani*), which can be more or less stylized and seem to have developed early on, from the Arabic-based script used to write Malay, *akhir cawa* (or *Jawi*), which was adapted to use Cham more recently (Figure 5).

Arabic-based scripts did not supersede the Indic script, however. A new script called *akhir 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 *akhir srah*. They are still preserved (and recopied in notebooks) by contemporary Cham communities and some are consigned in French libraries.
Figure 5: Example of Jawi script, Châu Đốc, Vietnam.
The Historical Development of Chamic Languages

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 Pānduranga, 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

Interestingly, akhār srah has evolved in slightly different ways in Vietnam and Cambodia (compare Figures 5 and 6).
Figure 6: Akhär srah manuscript, Scarborough collection (Cornell University), manuscript 47a, Akayet Inra Patra.
Among the Cham of Cambodia, diversity found in the large number of words that have been borrowed from the mainland languages spoken around them. The Cham of Cambodia, having been a state of affair that was already well under way at the beginning of the 20th century, is likely due to the various waves of Cham refugees and settlers that established communities in Cambodia from the 15th to the 19th century. Anglais or Polish police from police.
រូបភាព ៧ ៖ សំណកអក្សរស់ មកពីខត្តកំពង់ឆ្នំង កម្ពុជា ៖ គួរអាន ជាមួយអតាដូវបាយជាភាសាចាម្ប៍។

Figure 7: Akhârrah sample from Kompong Chhnang province, Cambodia: The Qur’an with Cham commentaries.
A characteristic of modern Cham dialects in the large number of words that have been borrowed from the national languages spoken around them.

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
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
The Historical Development of Chamic Languages

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...
Marc Brunelle and Graham Thurgood

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.
The Historical Development of Chamic Languages

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 Cham 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.
Map 4: Current geographic location of Cham dialects.

1. Tây Ninh
2. Châu Đốc

Regions where Cham is spoken
The Historical Development of Chamic Languages

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>älêmˎ</td>
<td>411,275</td>
<td>26,335</td>
</tr>
<tr>
<td>oyuˊ</td>
<td>161,729</td>
<td>204,080</td>
</tr>
<tr>
<td>oyuˊ</td>
<td>105,000</td>
<td></td>
</tr>
<tr>
<td>oyuˊ</td>
<td>30,000</td>
<td></td>
</tr>
<tr>
<td>oyuˊ</td>
<td>25,000</td>
<td></td>
</tr>
<tr>
<td>oyuˊ</td>
<td>331,194</td>
<td></td>
</tr>
<tr>
<td>oyuˊ</td>
<td>122,245</td>
<td></td>
</tr>
<tr>
<td>oyuˊ</td>
<td>19,314</td>
<td></td>
</tr>
</tbody>
</table>

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

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 as Cham.
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.

A systematic dialectal survey of Cambodian Cham is urgently needed.
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)tom ‘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
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 тəhàʔ/ʔhàʔ in Eastern Cham, and “body” пətən is ḫupapap/ṛup phap in Eastern Cham. Second, some words have developed different meanings in different dialects. Thus, пərj is a classifier for animals in Western Cham, but has come to mean “we, us” in Eastern Cham, and акхəɬ, which only means ‘script’ in Western Cham, has also taken the meaning of ‘language’, along with сəp, 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 th- in Eastern Cham. Thus, saŋ ‘house’ and asɬ ‘horse’ are pronounced as thəŋ and theɬ in Eastern Cham. Another characteristic of Eastern Cham is that its diphthongs (its complex vowels), tend to
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 pha ‘thigh’ and the last syllable of papha–rapha ‘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.
Marc Brunelle and Graham Thurgood

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 limo tami san ‘the cow enters the house’ is normally produced as mo mi than in casual speech.
have to deal with the massive settlement of the Central Vietnamese Highlands also difficult to predict. Some communities of growing proportion of Chamic speakers in the spread of formal education in Khmer in larger cities, where daily contact with the happening normally. Assimilation to transmission to younger generations is systematic character as in Eastern Cham, but it does not have the same Monosyllabization is also common in Western Cham, but it does not have the same systematic character as in Eastern Cham. 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 settlement 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 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.
The Historical Development of Chamic Languages

Corpus of the Inscriptions of Campā.


