Forward

This proposal is for an international standard of romanization for Eastern (Akhar Thrah) and Western (Akhar Srak/Akhar Ka Kha) variants of the Indic Cham script. Cham language includes a variety of Chinese, Sanskrit, Arabic, Malay, highland, Khmer and Vietnamese loanwords. The term for the script *Akhar* meaning literature 'letters' can be found with variant pronunciations across the Sanskrit and Pali language worlds. There are an estimated 80,000 pages of yet unromanized manuscripts in Vietnam and Cambodia with more than 300 estimated titles. The target population includes 167,000 Cham in Vietnam and 400,000 to 500,000 Cham in Cambodia as well as 50,000 Cham in Malaysia. Even if only a small minority of these populations still use the script it is likely to continue to be the focus of cultural revival efforts amongst the many leaders in the Cham community. This system was first discussed at the CORMOSEA meeting in Philadelphia in 2014. Future work may include the mapping of *Akhar Bani* a variant of Arabic used in Vietnam onto these systems since there is currently an *Akhar Bani* revival underway as well.[1]

I. Justification of ALA-LC romanization

The Cham language can be written best in either Arabic or Indic based script systems. The more popular 'Cham script' is from the Pallava-Grantha sub-family of the Brahmi Indic scripts, called 'Akhar Thrah' in Eastern Cham (Vietnam). Other scripts in this family are: Khmer, Thai, Lao, Burmese, Old Javanese, Balinese and Old Tagalog. There is no current ALA-LC romanization standard for Cham. Previous academic approaches to Cham romanization vary. The long term trend has been toward pronunciation based systems. Accomplished poet and independent scholar Inrasara recently adapted a pronunciation based system in consultation with Lưu Quảng Sang, Nguyễn Văn Tỷ (of the Cham textbook compilation board) and PhD in Education (University of Hawai'i) Quảng Đại Cần for the 2014 release of his *4650 Từ Việt Chăm thông dụng* (Văn Hóa – Văn Nghê publishing house, Ho Chi Minh City, Vietnam). This system is great for those Cham and Vietnamese students and scholars in Vietnam, but only suitable for those students and scholars with an existing knowledge of Vietnamese and therefore cannot be used as an international standard. Another existing model is the EFEO standard, which is script based. Adapting a modified EFEO standard can be used with the *Xalih Akhar Cam* software, which allows one to convert Romanized Cham to Cham script and vice versa.

Other Southeast Asian romanization systems can be used to inform the discussion on the romanization of Cham language. An example of a pronunciation based model is Thai. Khmer is an example of a *mostly* script based model. Revisions to script based systems tend to shift toward pronunciation based systems over time. Still, current ALA-LC standards call for script based proposals. Script based systems have also previously been used by the EFEO for Khmer and Eastern Cham script (Akhar Thrah). There is also a variant of Cham script for Western Cham populations called 'Akhar Srak' or 'Akhar Ka Kha.' Since fluency across Cham scripts is limited

to expert researchers of both Cham heritage and non-Cham heritage, having a comparative model for romanization is helpful since it will help to improve fluency across the scripts, even for experts. A similar argument was used for the recent inclusion of *Akhsar Mul* 'rounded script' common to Khmer language advertising and public signage. It future works it may be necessary to include an *Akhar Bani* local Cham dialectical variant of Arabic script as well.

Advocating an international standard is not a 'must use' for Cham populations. Usage of Cham script is always preferred, since it is the best way to represent the language. However, romanization is useful for: non-script compliant computer systems, text messages, online messages, Facebook and blog posting as well as library catalogues and academic works. Linguists have advocated romanization for the sake of literacy in the past. However, this position has traditionally been met with strong opposition by Cham communal leadership. Regarding the Summer Institute of Linguistics (SIL) romanization system introduced in the 1960s and 1970s Dorris Blood (1980: 42) has written:

> *Fear of the Romanized form of writing centered in a concern that it would push the script into obscurity. Some of the reaction was violent: one group of teachers, scholars and priests decried that any teacher found teaching from the new books would be relieved of his position. Other complaints reversed usual cultural practices of the people. One teacher remarked about his class of five-year-olds, 'They don't want to learn the new letters, they want to learn the script,'...when I asked a middle-aged woman if she would like to learn to read, she looked in disdain at the primer, stating, 'I only want to learn Cham letters.'*

Local views have generally not changed the professional opinions of linguistics experts. The complexity of the Cham script is cited as a reason for lack of literacy and the 'simplicity' of romanization is linked to the increase in literacy, or a potential revival of written Cham.[2] In a recent article for the *Journal of Language Preservation and Conservation*, linguistics expert Marc Brunelle (2009: 43) concluded that even though he sought to make no recommendation for Cham communal leaders, they faced a choice, "*Cham communities can either preserve the classical akhar thrah script as an ethnocultural symbol with limited practical purposes, or undertake an orthographic reform (or even more radically, a replacement of their script) aimed at fostering widespread literacy and at reviving written Eastern Cham.*"

Brunelle made two errors. He referred to Akhar Thrah – the script that has been used for writing Cham language from the seventeenth century to the present – as a '*classical script,*' when the term 'classical script' could only truly be applied to the various epigraphic samples that appear dotted along the contemporary coastline of Vietnam. Second, Brunelle left out *entirely* a foreseeable fourth option: that the knowledge of the Cham modern Akhar Thrah script could be revived as a means of improving scholarly knowledge regarding the resources of Cham manuscripts, which provide ample potential source material for understanding the history, traditional medicine, religion and language of the Cham community. In short, Cham manuscripts are a scholarly goldmine that opens resources for the fields of historical linguistics, comparative religious study and medicine. Improving our understanding of these scholarly resources and the

---

[2] See for example: David Blood (1977). A Romanization of the Cham Language in Relation to the Cham Script. Viet Nam Data Microfiche Series No. ND51-17. SIL. Dallas, TX.

Cham language also has the added benefit of treating some of the concerns of the youngest and most active leaders in the Cham community.

In a 2011 interview for *The Word* a popular bilingual magazine in Ho Chi Minh City, Vietnam, Inrajaka (BA in Southeast Asian Studies, excellent English, excellent Thai and son of famous bilingual poet Inrasara) noted that the Cham are "*losing a lot of…heritage…"* as a result of "*…this era of globalization*" and the Cham in Vietnam today are speaking a mixture of 70-80% Vietnamese and only 20 – 30% Cham (Thorton 2011: 43). Fluent literacy in the Cham script is staggeringly low and limited to a very small portion of the population (less than 10%). Nevertheless, the Cham are seen as owing an aspect of their cultural heritage to this script: most popularly referred to as 'Akhar Thrah' in Vietnam and Akhar 'Ka/Kha' in Cambodia. There have also been significant, sustained efforts to revive the script even after Brunelle's conclusions.

Efforts to revive the Cham script and Cham language publications include: a series of post-1975 programs in Vietnam, publication of poetry and essays in Cham script in the *Tagalau* series (14 issues), the use of the Cham script in the mostly Vietnamese language *Journal of Research on Cham Culture* (2 issues), programs sponsored by the American Embassy in Cambodia (post 2010), publications by the EFEO-Kuala Lumpur (4 books and a collection of typed manuscripts), an all Akhar Thrah poetry publication *Ariya Bino* by Sakaya (2013). A collection of Cham manuscripts of nearly one thousand pages was recently published in Cambodia along with the four text books that were used for the teaching of the Cham script in Cambodia. These publications have increased our need to have an acceptable ALA-LC standardized system to input materials published in Cham – intended for use in systems that do not yet support the Cham font.

Target Population:

Cham speakers in Southeast Asia are mostly split between three states: roughly 167,000 people in Vietnam, 400,000-500,000 in Cambodia and 50,000 in Malaysia. In Vietnam the population is divided between Eastern Cham (Ninh Thuận and Bình Thuận provinces as well as Hồ Chí Minh City) as Western Cham (Tây Ninh and An Giang provinces as well as Hồ Chí Minh City). Educated members of Cham society move between the dialects with great ease. In Cambodia the population is entirely Western Cham speaking. In Malaysia, the population is mixed. Speakers of Cham language demonstrate the ability to rabidly acquire knowledge of other Chamic languages (Utsat, Churu, Raglai, Koho, Jarai, Achenese and so on), due to a high percentage of shared vocabulary and shared syntax (Blood 1980, Thurgood 1999, Brunelle 1995, 2008, 2009a, 2009b). The non-Cham 'Chamic' populations are frequently portrayed as having 'lost' their written language, or never having had one. Common perception of the Jarai (who use either Romanization in Vietnam or Khmer script in Cambodia) is that they have 'never' had a script. Based on the oral histories of various Chamic ethnic groups that can be found in the works of Gerald Cannon Hickey (1982a, 1982b), we believe that the 'loss of script' is a more likely scenario, since each group is linked to a legacy of princely lineages. Toshiko Shine (2009) posited that the Cham-Raglai division occurred as an eventual result of a split between classes of Champa society. The 'loss of script' is also commonly articulate among Cham Muslim populations. There are individuals who have therefore specifically sought to 're-learn' Cham script. Cross communal language studies have deep roots. In case of the Cham of Southeast Asia, these can be traced to at least the nineteenth century.

II. Brief History of Cham Language Studies

Lexicons or comparative vocabularies were amongst the earliest studies of language completed by Cham scholars in Southeast Asia. Cham scholars used Akhar Thrah script to write on manuscripts made of Chinese paper, goat skin or other materials. CAM 198 appears to be an important manuscript in the collections of Cham manuscripts in Paris, France. It is likely that this manuscript was copied to the EFEO-Kuala Lumpur in the 1990s. However, the CD backup of selected literature in this collection held at Berkeley University does not appear to contain this manuscript. Po Dharma's cataloguing project on Cham manuscripts suggests that the copyist or author of these piece was a [Po] Gru Su – likely an Awal priest. Portion of the Cham-Malay lexicon can be dated to the twentieth year of the reign of the Vietnamese Nguyễn emperor Minh Mệnh (1839). This was thirteen years before the first 81 word Malay – Cham vocabulary list by John Crawford and A. Morice in their 1852 *A Grammar of the Malay Language* (Po Dharma 1981: 40).[3] Given the prevalence of 'Cham-Malay' vocabulary in Cham Rija ritual ceremonies that have been a mainstay of communal life from the seventeenth century (or earlier) to the present, we can suggest that this manuscript had a role in these ceremonies (Sakaya 2012).

In addition to CAM 198 there are at least five manuscripts, likely dating to between the nineteenth to the early twentieth centuries – all of which are language studies composed in Akhar Thrah script and demonstrate attempts to master other local languages. Three of these manuscripts: CAM 127, 131 and 133 are Koho-Churu lexicons. The Churu owe their origins to a creolization of Cham, Raglai, Rhade and potentially other Chamic groups. The Churu also were the ruling class of the last Champa *negara* kingdom of Panduranga (now: Ninh Thuận, Bình Thuận and Biên Hòa) for much of the seventeenth and eighteenth centuries. Two of these manuscripts (CAM 131 and CAM 133) suggest that they were composed in the region of Đa Lạt and so they can be dated to the twentieth century. The other two manuscripts are a Chu-Ma lexicon (C: 'Cuw-Mak'; CAM 135) and a Cham-Khmer lexicon (C: 'Kir'/'Kur'; CAM 137). (LaFont, Po Dharma & Nara Vija 1977: 87-89).

A final example of language studies can be found on Cham manuscript CM 39 from the library of the Asiatic Society in Paris, France. The manuscript appears to have been penned by a Cham who bore the title 'Prah Balat' – a provincial assistant to the Oknha regents of the Khmer Royalty. Nicholas Weber (2011) dated this manuscript in reference to events that occurred between 1862 and 1867. The reference to the *thun nasak inâ garai* or 'year of the dragon' suggests that some portions were completed in 1868 (LaFont, Po Dharma & Nara Vija 1977: 190-195). A complete reading and translation of the manuscript could be used to determine more about these events. We hypothesized that if we knew whether or not the manuscript was completed in the year of *thun nasak inâ garai jim* this could confirm certain details, since the Cham calendar is commonly combined with the 'Jawi' or Bani-Cham lunar calendar as a method of more accurate dating. Unfortunately, Po Dharma (2014) countered that he never found this

---

[3] An 800 word list of Cham was then published in the *Revue de Linguistique et de Philogie* in 1875 (Moussay 2006: 17-21). This is available at the University of Wisconsin-Madison Memorial Library Microfilm collection: Micro Film 11200 Reel 72. It appears that Crawford may have been aware of Cham as early as 1822, although he referred to it as 'Malay of Champa.'

system in the Cham Royal Chronicles held in the Paris collections.[4] However, Sakaya (personal comm.: 4/4/2014) suggested this may have been used for some turn of the century manuscripts, particularly if they included content of a religious nature. Regardless, Akhar Thrah script was still the dominant mode of written communication among the Cham well into the colonial period.

III. A closer look at the French colonial period and Romanization

Even though the French began Romanizing Southeast Asian languages in the seventeenth century, it took two centuries for romanization to become popular amongst even the Vietnamese population. By the late nineteenth century, French scholars advocated Romanizing Cham as well. The first system was published in 'Grammaire de la langue Chame' in the journal *Excursions et Reconnassance* XIV, 31, 1889 by Etienne Aymonier.[5] However, it appears that Aymonier and several Cham scholars had been working on the system for years. CAM 169 is a Romanized version of the manuscript *Ariya Po Pareng* by Cham scholar Hợp Ai of Palei Hamu Tanran. A contemporaneous manuscript, penned by Ja Mul Cak was composed in the *thun nasak mânuk* 'year of the rooster' (or 1885). This was on the 6th day of the crescent moon in the 12th month of the Cham calendar. CAM 169 includes a Cham-French lexicon and a romanization of travel accounts, perhaps recording the movements of Etienne Aymonier (Po Dharma 1981: 23). Other Romanized manuscripts include: CM 74 [transcription into romanization of a manuscript by Thông Bình Phú Sơn], CM 76, 77, 78, 79 and 80. All of these appear to be the accounts of journeys of Etienne Aymonier or other Frenchmen into Romanized Cham (Po Dharma 1981: 203-204).

Aymonier's system of romanization was modified by Antoine Cabaton for *Nouvelle Recherches sur les Chams* (1901) and 'La Transcription du Cam' (*Memoire de la Societe Linguistique de Paris*, XIII, 4, 1905).[6] Aymonier and Cabaton merged their systems in the 1906 Cham-French dictionary. Although the dictionary opted to use the Eastern Cham script for all entries, it included the Western Cham variants spellings and pronunciations. This system closely resembled the existing system used for Sanskrit with slight modifications for French pronunciation of consonant, vowel and diphthong-like formations (Moussay 2006: 17-21).

Another project contemporaneous with Aymonier and Cabaton's studies was that of MP EM Durand. Durand completed a series of BEFEO articles on the Cham. Manuscripts used for his collection are held in the archives of the Foreign Missions in Paris, France. Although MEP 1190/5, /6, /7 and /8 were all made from this collection and appear to include Romanized Cham,

---

[4] Po Dharma (3/31/2014). Trả lời độc giả: nguồn gốc của Suk Yeng và vấn đề lịch Chăm [Answers to readers: The origins of the Suk Yeng ritual and issues regarding the Cham calendar]. *Champaka.info*. Available Online @ http://champaka.info/index.php?option=com_content&view=article&id=1061:traige&catid=53:quan-im-vn-hoa&Itemid=60

[5] This report was also published as a monograph in Saigon as by the Imprimerie Coloniale in 1889. Despite his detailed work Aymonier seems to have been missing the two extra rows of nasals that occur in the Cham system – by comparison to Khmer, which has only one row of nasals. He also appears to have Romanized one of these nasals as *za* which he then approximated as an extra form of *nhoe* (Aymonier 1889: 11).

[6] For those who don't read French this was translated from the French by Basil Guy. Human Relations Areas Files. New Haven, 1955. In this file Cabaton noted that "apocope and apheresis are frequent in Cham" giving the examples: kok for akok = head; cin for kacin=button; ra for urang=man; and lan for bilan=month/moon (Cabaton 1901: 192). Cabaton's 1901 system still appears to be missing at least one row of nasals.

it does not seem that Durand published any changes to the systems of Aymonier and Cabaton. Of these collections: MEP 1190/6 is notable. It may be the earliest datable Cham-Arabic-romanization lexicon outside of Aymonier and Cabaton's 1906 dictionary.

After the turn of the century collections there seems to have been nearly a generation long lag of interest in Cham manuscripts before the collections of JY Claeys and Paul Mus were made at the turn of the 1920s and into the 1930s. A critical figure at this time was Etienne Aymonier's Cham son Bố Thuận, who appears to have been a research partner of Paul Mus. On the topic of language study there are at least seven manuscripts that are related to the study of French, Vietnamese, Malay and Cham language. CAM 45 is a Cham French word list from the collection of JY Claeys. Meanwhile, CAM 160, 182, 183, 184 and 185 are from the Paul Mus collection. CAM 160 for example is a Cham-French lexicon from the collection of Bố Thuận. CAM 182 and 183 are Cham-Vietnamese vocabularies that were composed in the region of Phú Yên – and thus, may represent important windows into the linkages between the Cham and the minority Bahnaric 'Cham' – Hroi population that has represented a significant minority community that province. CAM 184 is a manuscript that includes a Cham-Vietnamese vocabulary from an Awal Po Acar priest and CAM 185 includes a lexicon of Cham, Malay and French that was collected by Paul Mus. Hence it is likely that these studies informed Paul Mus's modification to Etienne Aymonier and Antoine Cabaton's system. Paul Mus' system was then first reproduced in a BEFEO article (XXXI: 44) and then solidified in his article 'Transcription alphabetique de la language Chame' (BEFEO XXXVIII: 507 – 510; Po Dharma 1981: 29-31; Moussay 2006: 17-21).[7]

IV. The Introduction of Pronunciation based systems, 1950s through the 1970s

A generational gap combined with the Second World War seems to have led to a decrease in publications on Cham language studies. However, beginning in the 1950s, the Summer Institute of Linguistics (SIL) increased interest in Romanization. SIL programs emphasized oral communication and, hence, adapted different Romanization based on pronunciation models for the Eastern and Western Cham dialects in Vietnam.[8] The impact may have further divided the populations, since the models were different. Since most Western Cham who were literate in Arabic script by this point in time, this could have been an opportunity to promote interreligious fluency in Cham language. However, the two Romanization systems operated upon different understandings of spoken Cham. The two systems likely divided the

---

[7] Again this Romanization appears to have been missing two rows of nasals.

[8] The SIL original orthography is missing the extra nasal rows it seems. Field notes indicate the Romanization advocated by the SIL for Eastern Cham was heavily influenced by Quốc Ngữ David & Dorris Blood (1977). E. Cham Language Material on Phonemic Distribution. SIL. Dallas, TX. Originally from Ong Canh, D&D Blood, Ervie Lee, Dave Thomas and Dick Pittman from 4/18/1968 Field Session. UW-Madison Microfiche: 3793.

A 1962 primer seems to indicate a script based romanization only. See Lữu Quý Tân (1962). *Sram Akhar Cham Tapuk 1-3* [Học Tiếng Chàm Tập 1-3]. Bộ Quốc Gia Giáo Dục Xuất Bản. in Revised Sample Lesson of Teacher's Guide by David & Dorris Blood. SIL (1977). UW-Madison Microfiche: 3791.

A 1970 primer series: *Aday Bach Akhăr Chăm Birau*. Tapuk 1-3. Tăl Birau Bach. Gilăng Pato-Pakai. Ba Tabiak. Baigol. Uses Romanization first, based upon popular pronunciation and then follows with the script and Vietnamese translation. The primers were prepared by David and Dorris Blood for the Highlander Education Project for the use of the Cham people living in the vicinity of Phan Rang and Phan Rí. David & Dorris Blood (1977). Eastern Cham Primers and Guide. SIL. Huntington Beach, CA. UW-Madison Microfiche: 3792.

Cham further, so we suspect that this was a significant factor in the rejection of Romanization or the 'suspicion' that Blood (1980) recorded when she noted that throughout the 1960s and 1970s Eastern Cham leaders had a negative view of Romanization strategies. It is no surprise that the dominant form of Romanization used by Gerard Moussay (EFEO) and the Cham scholars of the Center for Cham Culture (CCC): Nại Thành Bộ, Thiên Sanh Cảnh, Lưu Ngọc Hiến, Đàng Năng Phương, Lưu Quang Sang, Lâm Gia Tính and Trương Văn Tốn in their 1972 Cham-French dictionary were actually two fold. The first was pronunciation based – aimed at appeasing French linguists – while the other was script based using the *Inâ Akhar* system of Akhar Thrah script.

The CCC's initial Romanization systems appeared in the series *Roh Tuah No. 1* (see: CHCPI 2). Meanwhile, both CHCPI 7 and CHCPI 9 appear to have had this system recorded on them as well. CHCPI 9 is a manuscript that was initially composed at the CCC in Phan Rang, Ninh Thuận, Vietnam before it was deposited into the CHPI in 1974 by the late Father Gerard Moussay. On July 5, 1974 of the same year, the *Akhar Cam Krun Dahlau* series was released in two parts and also used the initial CCC system. This system was also likely preserved on Po Dharma's PO DHARMA CAM 6 and CAM 7 'fascicule's' in the Paris collections. Most importantly it is these systems that have been the existing international standard for the Cham Romanization as of the 1998 publication of the 'Equipe de Recherces Cam: Transcription du Cam Modern' and were hence adapted by the EFEO (Moussay 2006: 17-21).

V. Post 1975

Since the topic of Romanization in Cambodia has traditionally been heated, no current standard for the Cham script in Cambodia has been published. In Vietnam, the post – 1975 linguistic condition of the Cham community has been one of almost total bilingualism and increasing polylingualism. Through contact induced change Cham has become increasingly mono-syllabic in its spoken form and the conceptions of 'tonalization' have been used to teach proper pronunciation. This model likely led to certain modifications of the Romanization system in Búi Khánh Thế's Cham-Vietnamese dictionary that demonstrate the continued use of Vietnamese Quốc Ngữ style diacritics, modifying the EFEO's system for the use of the community living in Vietnam.[9] Accomplished poet and independent scholar Inrasara then recently adapted this system in consultation with Lưu Quang Sang (who was involved in the CCC's dictionary project), Nguyễn Văn Tỷ (of the Cham textbook compilation board) and PhD in Education (Hawai'i) Quảng Đại Cần in his 2014 release of the *4650 Từ Việt Chăm thông dụng* (Văn Hóa – Văn Nghệ publishing house, Ho Chi Minh City, Vietnam). The romanization system of this most recent dictionary is specifically designed to teach pronunciation for individuals who are already familiar with Vietnamese language. It is therefore incredibly useful bilingual youth in the Cham community seeking to re-connect with their heritage and Vietnamese speakers who wish to learn Cham. However, it does not comply with the current ALA-LC regulations, or account for the Cham script used by the Cham populations in Cambodia, nor is it useful for Cham populations that are not already familiar with Vietnamese. Hence it is not acceptable for an international standard. Using a script based model has the added benefit of internationalizing communication in the Cham language, suiting the needs of cataloguers, who will sooner or later

---

[9] See: Bùi Khánh Thẻ (chủ biên). (1995). Từ Điển Chăm-Việt. [nhóm biên tập: Đinh Lê Thư, Nguyễn Văn Lịch, Phú Trạm; tham gia biên soạn: Thành Phần, Phú Văn Hẳn, Lương Đắc Thắng]. NXB Khoa Học Xã Hội: TP Hồ Chí Minh.

be likely be requested to deal with Cham language materials regardless of their familiarity with the language and finally, can be used with the *Xalih Akhar Cam* software system, which allows one to type Romanized Cham or Akhar Thrah script and allows conversion.

VI. Pronunciation notes

Pronunciation generally varies, and romanization generally does not.

1. The /Ppa/ romanization for the second /pa/ phoneme does not differ in pronunciation from the first.
2. The /sa/ initial phoneme in Eastern Cham is frequently shifted to a hard /tha/ as in 'Thomas' e.g. *sang* house is pronounced *thang* .
3. The /xa/ special phoneme can marked by an /xa/ or /sa/ in Romanization is pronounced as a soft /s/ - although this is also frequently shifted to a /sha/ or /tha/.
4. In Western Cham, if the /ra/ phoneme appears in the middle of a word, it is pronounced as /ga/.
5. In both Eastern and Western Cham the initial /wa/ is often hardened to a /va/.