Chamic and beyond

studies in mainland Austronesian languages

Pacific Linguistics 569

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, Taiwan, the Philippines, Indonesia, East Timor, southeast and south Asia, and Australia.

Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at The Australian National University. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise, who are usually not members of the editorial board.

Chamic and beyond: studies in mainland Austronesian languages

Edited by
Anthony Grant and Paul Sidwell

Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University

Published by Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University
Canberra ACT 0200
Australia

First published 2005

National Library of Australia Cataloguing-in-Publication entry:

Chamic and beyond : studies in mainland Austronesian languages.

Bibliography.
ISBN 0 85883 561 4.

1. Austronesian languages. I. Grant, Anthony (Anthony Paul), 1962- . II. Sidwell, Paul. III. Australian National University. Research School of Pacific and Asian Studies. Pacific Linguistics.

499

Cover photograph by Paul Sidwell, 2001.
'Ede Longhouses' from Buon Ma Thuot in Central Vietnam.

Typeset by Paul Sidwell
Cover design by Cirils Digital Printers
Printed and bound by Cirils Digital Printers, Fyshwick, Canberra

Contents

Notes on contributors

Marc Brunelle received his Ph.D. in Linguistics from Cornell University in 2005 for a dissertation entitled *Register in Eastern Cham: phonological, phonetic and sociolinguistic approaches*. This dissertation, based on field research near Phan Rang, Vietnam, in 2003 and 2004, explores the issues of registrogenesis and tonogenesis in Eastern Cham and the role of Cham/ Vietnamese bilingualism and cultural contact in these developments. He is currently a lecturer in Linguistics at the University of Michigan, Ann Arbor. His research interests include language contact and the phonetics and phonology of tone and register in Southeast Asian languages, with a special focus on Chamic languages and Vietnamese.

Anthony P. Grant, a native of West Yorkshire, is a Lecturer in English language at Edge Hill College of Higher Education, Ormskirk, Lancashire, UK. His undergraduate work was at the University of York under pioneering creolist Robert Le Page, where he also studied Old English, Swedish and Hindi, and creolistics has remained an active research field for him, together with Native North American languages, Romani, Austronesian languages, historical linguistics and issues in intimate language contact (language intertwining, relexification and issues in typology). He has previously taught and researched at the Universities of Bradford, St Andrews, Southampton, Manchester and Sheffield. He is an Associate Member of *Centre for Research on Language Change* and is on the Editorial Advisory Board of the *Journal of Pidgin and Creole Languages*.

Peter Norquest is enrolled as a Ph.D. student in the Joint Program in Linguistics and Anthropology at the University of Arizona. His main research area is historical phonology, with an emphasis upon the major language phyla of Southeast Asia, and a primary focus on Kra-Dai (Tai-Kadai) and Austronesian. With the aid of a DEL grant and Fulbright award, he spent the academic year 2003-04 on Hainan Island doing linguistic fieldwork with speakers of the Hlai languages indigenous to Hainan. Peter is currently completing his dissertation on the phonological reconstruction of Proto-Hlai.

Pittayawat Pittayaporn received his B.A from Chulalongkorn University in Thailand. He is currently doing his Ph.D Linguistics at Cornell University. Within the sphere of historical linguistics, his interests lie in the issues of language change, language contact, and linguistic prehistory. As for the synchronic, he is interested in tonal phonology, and phonology-morphology interface. His special focus is on Kra-Dai, Austroasiatic, and mainland Austronesian languages. He is currently working on a new reconstruction of Proto-Tai and developing a method for subgrouping its daughter languages.

Paul Sidwell is a comparative-historical linguistic who specialises in the Mon-Khmer languages of Southeast Asia. Since receiving his PhD from the University of Melbourne for his thesis “A Reconstruction of Proto-Bahnaric” Paul has done field work on the Bahnaric and Katuic languages of southern Laos, where he has also been active in environmental and community development activities of the *Green Life Association* (NGO). Paul is a member of the editorial boards of *Pacific Linguistics* and *Mon-Khmer*

Studies, a founding member of the *Centre for Research on Language Change,*and an Associate member of the *Centre for Research in Computational Linguistics* (Bangkok). Presently he is a Visiting Research Fellow at the Australian National University, where his research activies are supported by the Max Planck Institute (Leipzig).

Graham Thurgood is a Professor of English and Linguistics at California State University, Chico. The primary focus of his work is the nature and causes of language change and what it can tell us about non-linguistic (pre-)history and about the nature of language. He has worked in Tibeto-Burman, Chamic, and Tai-Kadai, including work on historical reconstruction, subgrouping, evidentiality, grammaticalization, morphological change, tonogenesis, registrogenesis, migration patterns, contact patterns, fieldwork, and so on. He is also a co-founder and former editor of *Linguistics of the Tibeto-Burman Area.*

Ela Thurgood is an Associate Professor of English and Linguistics at California State University, Chico. Much of her work focusses on instrumental studies of Southeast Asian languages. Her dissertation was a descriptive study of 19th century Baba Malay; she has also published on Papia Kristang, a Portuguese-based creole of Malaysia. Her instrumental work includes a publication on the phonation types in eastern Javanese and an instrumental study of the tonal system of Hainan Cham (Tsat) as well as several instrumental papers on Polish gemminates.

Editors' preface

This volume brings together papers whose common theme is the remarkable processes of linguistic change and transformation that have occurred (and are still occurring) over the last two thousand years in the Austronesian languages of Mainland Southeast Asia.

What are the Mainland Austronesian languages? The description is perhaps at first misleading—some of the languages covered by it are in fact spoken on islands, while Malay, for example, which is spoken on the mainland, is not included. The term characterises those Austronesian languages that have converged typologically to a "mainland' or more specifically, Mon-Khmer type. These languages/language groups are Chamic, Acehnese and Moken/Moklen—not a single genetic sub-grouping but a number of related languages that have undergone parallel typological restructuring away from their Austronesian heritage, converging on a type that places them on the southern periphery of the broader Mainland Southeast Asian Linguistic Area.

Historical interest in Mainland Austronesian languages grew in the 1990s, along with the broader trend to look at the role of contact in linguistic change, and the greater opportunities afforded by improved fieldwork access and an increase in published data. Consequently two substantial and important works appeared independently in 1999: Thurgood's monograph "from Ancient Cham to Modern Dialects" and Larish's (1999) PhD thesis "The position of Moken and Moklen within the Austronesian language family". Both authors took great pains to document and reconstruct the processes of restructuring that had occurred within each family, both reaching the conclusion that in each case it was the prolonged and intimate contact with Mon-Khmer that was the principal cause of change.

Thurgood's work on Chamic has been particularly influential (and is more widely available in published form), with reactions varying across the spectrum from strong approval to strong criticism. It was Thurgood's book that particularly stimulated the papers by Grant and Sidwell appearing here, and it was out of their mutual discussions on the topic that the idea for this volume emerged during 2003. Grant had written drafts of two substantial papers, versions of which appear here, that extensively analyse the history of borrowing in Chamic specifically through the lens of Thurgood's reconstruction. Sidwell was working on the historical reconstruction of Bahnaric and Katuic, the Mon-Khmer groups that Thurgood identified as the most important contact languages for Chamic. At first we planned to put together a successor volume to the Pacific Linguistics (1997) "Chamic Studies", but then later widened of the scope to include Moken/Moklen, and this is reflected in the title we have used for this volume.

This typological restructuring in Mainland Austronesian languages is seen in changes such as more isolating syntax, reduction of the phonological word, and increasing complexity of phonemic distinctions. In the case of Chamic the restructuring has been so great that there were times when even its Austronesian provenance was doubted (e.g.

Schmidt 1906 regarded Cham as a mixed language, while Sebeok 1942 classified Cham as Austroasiatic), although that debate has now long since passed. The issues of great importance now revolve around how we can explain the changes that have occurred, what generalisations can be made, and what specific historical inferences can we draw. Many questions beg investigation, such as: to what extent were all of these changes rooted in a history of prolonged and intimate contact with Mon-Khmer speaking peoples? To what extent were language shifts involved? To what extent were purely language internal processes in train that for whatever reason resulted in striking yet superficial convergences? What heuristic techniques can be most profitably brought to bear on such questions—etymological, comparative, philological?

It is apparent that there were several waves of migration that occurred before the Common Era, in which Austronesian speakers established themselves on the Mainland coasts of the South China Sea and Gulf of Thailand, and at least some of them came into prolonged contact with Mon-Khmer speaking peoples. The nature and extent of those contacts, and their linguistic consequences, were profound and complex. In some cases there was massive and unambiguous lexical borrowing associated with structural changes, in other cases there is far less extensive borrowing yet still remarkably similar restructuring. This raises all sorts of issues about the mechanisms of contact induced change, and the historical inferences we can draw from such linguistic evidence.

Today the Chamic languages are spoken in Indonesia, Vietnam (principally in the southern third of Vietnam), Cambodia and Hainan (People's Republic of China). Acehnese, which has over two million speakers in northern Sumatra, is treated as a Chamic language by Thurgood (1999). Sidwell (a Mon-Khmer specialist), in his paper (this volume) "Acehnese and the Aceh-Chamic Language Family" treats it as a sister of the Chamic family, a view which is perhaps more in step with mainstream Austronesianist views. In any case the general consensus is that the Acehnese reached their present home via a dramatic back-migration to the insular Austronesian world sometime before historical records began (by contrast Dyen 2001 asks us to consider an ancient migration of Chamic from Sumatra to Indo-China). Tragically the Acehnese population was perhaps the most grievously affected by the earthquake and tsunami which struck the Indian Ocean on 26 December 2004.

The most famous Chamic languages is Cham, the coastal language for which the group is named, and the one which has the greatest number of speakers (after Acehnese). Cham has two modern mutually unintelligible dialects, namely Western Cham of Cambodia, which is used by perhaps 250,000 people living around Tonle Sap and in the neighbouring (and largely Khmer-speaking) part of the Socialist Republic of Vietnam, accounting for the majority of the current Cham speaker population. And there is Eastern Cham or Phan Rang Cham of the former Cham city of Panduranga, now Phan Rang in Vietnam, with about 35,000 speakers. There is an outpost of several thousand Western Cham-speakers, refugess from Pol Pot's Cambodia, in and around Bangkok, Thailand (population figures drawn from Grimes 2000). For over a millennium Cham has supported an ancient written and epigraphic literature, including verse narratives and work in some genres typical of Southeast Asian literatures (such as the long poem), much of which has been preserved. (Aymonier 1889 and Aymonier and Cabaton 1906 are the classic works on written forms of Cham, although the latter work especially also explicitly states that it presents a record of contemporary Western and Eastern Cham, as well as a recognition of the continuity of written and contemporary varieties of Cham.

In Brunelle's paper "A phonetic study of Eastern Cham register" (this volume) he discusses the synchronic phonology in detail, complete with spectrographic and other instrumental analyses. Brunelle finds that Eastern Cham has developed a register system that exploits both pitch and voice quality, but that speakers treat these as consonantal features, evidenced from word games, which argues against treating Eastern Cham as a tone language. Careful synchronic studies such as Brunelle's are essential for real progress in the historical analysis of the processes of change in these languages, highlighting as they do the linguistic mechanisms at work, and contributing to our knowledge of the areal typology.

Beyond Cham, other Chamic languages are Jarai, Rade (both of which are spoken in the highlands of Vietnam and are therefore sometimes referred to as Highland Chamic), the coastal language Chru, and the coastal group of Roglai languages. This latter group consists of Northern and Southern Roglai (the native name *ra glay* literally means "people [of the] forest" (Thurgood 1999:2)), and the divergent variety of Cat Gia Roglai. All of these languages are spoken nearer to the Vietnamese coast than are Jarai and Rade. There is also the Coastal Chamic language Haroi (itself an offshoot of Cham whose speakers later moved to the easternmost part of the highlands and settled next to speakers of Jarai and Bahnaric languages). These languages are spoken in a mostly continuous band of speech communities in southern Vietnam (Eastern Cham, spoken to the south of all these, is part of this band).

Poorly documented is the Rai or Seyu language, spoken in Binhtuy and Binhthuan provinces, and also by some people in Tuyenduc province, all of these localities being near the Mekong Delta (Grimes ed. 2000:1:650 sees this language as being an offshoot of Southern Roglai rather than a separate language in the full sense).

The territory of the Chru-speakers is to be found in two discontinuous enclaves near the Vietnamese coast, the southern one being the home of the southernmost speakers of a Chamic language in Indochina. (The name Chru, incidentally should not be confused with that of the neighbouring South Bahnaric language Chrau which has been heavily influenced lexically by Cham: Thomas 1971 is a grammar of this latter language.)

Further afield, at the southernmost tip of Hainan we find the incursive language Tsat (often referred to in the literature by the Chinese name *Huihui*, a reduplication of the traditional Chinese word for 'Muslim'). The Tsat speakers, or *Utsat*, moved to Hainan from Champa—Chinese records note the arrival of refugees in Hainan soon after the Vietnamese sacked the Cham capital Indrapura in 982. Thurgood's historical reconstruction also suggests that the Tsat migration took place from the northern part of the Cham-speaking area.

Given the separation of Tsat during approximately the last period of Chamic political unity, it can be assumed to mark the break-up of what we might call the Common-Chamic or Late Proto-Chamic language. This makes Tsat a crucial witness language for comparative analysis, yet its investigation and documentation has been inadequate, and the structural changes it has undergone so extensive that the comparative work one can conduct is greatly complicated. Thurgood's paper in the present volume brings to bear data from recent work by Chinese scholars and from Thurgood's own fieldwork, illustrating the development of Tsat from non-tonal Proto-Chamic into the fully tonal (and now highly sinisised) language it is today.

Speakers of some Chamic languages are also to be found in migrant communities in Malaysia, Australia, the US and France. Also there are Acehnese among members of the Indonesian student community that is spread throughout institutions of tertiary education throughout the world. The Acehnese, Tsat and most speakers of Cham, including all those in Cambodia and a third of those in Vietnam, are Muslims. This religious affiliation makes them anomalous in Vietnam and Cambodia, and also in Hainan where there are no other long-standing 'native' Muslim communities, apart from speakers of the Tai-Kadai language Jiamao in Hainan. Details of the populations of the various Chamic speech communities are given in Figure 1.

FIGURE 1: *Population Statistics for Acehnese & Chamic Languages*

Language	Where spoken	Speaker pop.	Date of figure
Acehnese	Indonesia	3,000,000	1999
Eastern Cham	Vietnam	35,000	1990
Western Cham	Vietnam	25,000	1990
	Cambodia	220,000	1992
	Thailand	4000	no date given but probably some time in the 1990s
Chru	Vietnam	11,000	1993
Haroi	Vietnam	35,000	1998
Rade	Vietnam	195,000	1993
Jarai	Vietnam	242,000	1993
	Cambodia	15,000	1998
Roglai, Northern	Vietnam	25,000	1981
Roglai, Cat Gia	Vietnam	2000	1973
Roglai, Southern	Vietnam	20000	1981
Tsat	Hainan, PRC	4500	1991

Although minority languages, perhaps only Tsat is immediately endangered. Although none of them are politically significant or dominant in any sense, it appears that all of them are being passed onto children. Speakers of Tsat are bilingual in Hainanese Chinese, and increasingly in Mandarin (while many also know Cantonese, which they use for purposes of trade), while those speakers of Chamic languages who live in predominantly Khmer or Vietnamese-speaking areas are increasingly fluent in these languages. Blood (1962: 113) notes that the Phan Rang Chams were almost completely bilingual in Vietnamese even in the early 1960s, and argued that this had the effect that their language was coming more and more to resemble Vietnamese phonologically. Similarly, a knowledge of Bahasa Indonesia has spread among speakers of Acehnese, and some Tsat-speakers have learnt Bahasa Melayu for commercial and other purposes.

There is little information available about bilingualism or multilingualism involving competence in two Chamic languages, although some better-educated speakers of certain Chamic languages are bilingual in Eastern Cham: this is the case with some speakers of Cat Gia Roglai (Lee 1998: 32). Most Chamic languages are unwritten, although there is a limited written tradition in Rade (using a Vietnamese-based orthography (Tharp et al. 1980). There is a tradition of writing in Cham, sometimes with Arabic characters, but much more often with an Indic-derived alphabet which does not fully represent the phonology of the language. At the time of the Bloods' research in the early 1960s, a Cham scholarly tradition persisted among certain males in Phan Rang, who could read the old Cham script, and such people spoke a phonologically archaising form of Cham which preserved certain phonological distinctions, retained in the written language, which were of Proto-Malayo-Polynesian age, and which the modern language had abandoned (Blood 1962: 113).

Recent material in Cham, in Cambodia as well as in Vietnam, has used an orthography based on that of Vietnamese, though Muslim Chams in Cambodia have apparently sometimes also written their language in Arabic letters. Muslim Chamic-speaking communities once used Malay as their language of religious instruction, but they now have little direct contact with its native speakers or indeed with other Muslims of any linguistic background, though this relative isolation is now changing.

Apart from data on Cham proper and Acehnese, there is rather little material available on any of these languages before about 1870, and what there is from that period is mostly the result of the investigations of French Orientalists and administrators[1]. Consequently there is rather little that can be done on these more recently-recorded languages in terms of philological analysis and diachronic back-projection if one insists on using earlier recordings. Thurgood's book, though, gives an indication of just how much can be done in terms of comparative-historical reconstruction.

Southern Roglai appears to be especially scantily documented, an important consideration if one bears in mind the wide variation within forms of Roglai, and we have seen no material on Rai (whatever its status as a language). Neither language is discussed much in the Chamic works of Graham Thurgood, which in any case give most of their emphasis to phonological developments in modern forms of Cham, Haroi, Tsat and to some extent Northern Roglai, and which bring Highland languages and the other languages into discussion less often, although data from them are provided where relevant. The emphasis of the study by Lee (1974) is rather different, focusing as it does on phonological developments in Northern Roglai, Rade and Jarai, as well as on Cham and Proto-Chamic. Some Chamic languages are extensively documented, but in these cases this documentation has been done mainly in hard-to-find or archaic publications; this is especially true with some of the Highland Chamic languages. On the other hand, a small book on Phan Rang Cham, which was based on fieldwork conducted in 1979 by members of a joint Soviet-Vietnamese linguistic expedition and which contains a general discussion,

[1] Graham Thurgood (personal communication) mentioned that vocabularies of several hundred entries for Rade and Jarai were compiled in the 19th century and written in Thai script. There is also John Crawfurd's wordlist of the 'Malay of Champa' (= Phan Rang Cham) which Thurgood (to appear) discusses and which dates from 1822.

a grammatical sketch, some texts, and a small vocabulary list, has recently been published in Russian (Alieva and Bui Khanh The 1999). The same Soviet-Vietnamese team also conducted fieldwork on Chru, but the results of this have not been published; let us hope that they soon will be.

Chamic languages exhibit quite a bit of internal diversity in regard to their salient structural features. This is manifested in both phonological and morphosyntactic features, though less so (loans apart) in basic vocabulary. Acehnese is more similar to languages such as Malay, and also to less closely related Austronesian languages such as Tagalog, in such matters as its possession of several productive prefixes and infixes which reconstruct back to Proto-Malayo-Polynesian but which are no longer productive in Chamic varieties. Chamic segmental phonology includes some kinds of sounds, such as implosive stops, which are uncommon in Malay and its relatives, and Chamic languages tend to have large numbers of monosyllabic contentives (a feature which is alien to Malay and to Austronesian languages in general except in the form of loanwords) and a wide range of vowel qualities, including (in some languages) nasalised vowels. The range of phonation types that are found among these languages is considerable; some Chamic languages have developed partial or full tone systems and others have developed "restructured register", with contrasting phonation types and complex vowel shifts.

Productive morphology in the Mainland Austronesian languages is minimal and what there is can be classified as being primarily derivational rather than inflectional. Free grammatical morphs are used, but there is also a great deal of zero-marking of many features. The usual element orders are Subject-Verb-Object, Numeral-Classifier-Noun, Noun-Genitive, Adjective-Noun, Preposition-Noun, and Tense/Mood marker-Verb. Bipartite negation using circumfixes is common and indeed characteristic of most Chamic varieties (Lee 1996), as are numeral classifiers. The first feature is alien to Malay, the second is not.

The explanation for this divergence of modern Chamic languages from the Proto-Malayo-Chamic norm, as Thurgood (1999: 251-259) points out, lies to a great extent in the intense linguistic contact which Chamic has undergone from surrounding languages. Many of these languages were once spoken by groups who were politically subservient to the Chams in the period of the Cham Empire. They used Cham as a lingua franca and some of them abandoned their original tongues in favour of Cham. Tsat and Standard Malay stand at opposite poles of a diachronic continuum of change whose major controlling factor is the myriad consequences of language contact or contact-induced language change. But at the same time it would be an unhelpful oversimplification to simply declare that all of the structural changes were the result of borrowing features directly from Mon-Khmer languages. Consider and compare the circumstances of Eastern Cham and Tsat as they are documented in this volume: Tsat has quite clearly assimilated to the phonology of Hainanese Chinese — the systems do not correspond perfectly, but the direction of chance is overwhelmingly towards the model of the local Chinese dialect. On the other hand while it is tempting to suggest that Eastern Cham is developing insipient tone under the influence of Vietnamese, in which most speakers are bilingual, there is little specific structural correspondence between the Eastern Cham and Vietnamese phonological systems, which suggests that more general phonological principals are involved. It is clear that it is premature to making sweeping generalisations about the divergence of modern Chamic languages from Proto-Malayo-Chamic norms due specifically to contact, and that only much more detailed work will be needed to support the convincing reconstruction of the mechanisms of change operating at the Proto-Chamic and immediate post Proto-Chamic times. Instructive in this regard is Peter Norquest's "Word structure in Chamic: prosodic alignment versus segmental faithfulness" in which he argues that the changes which occurred following the shift to word-final stress in ancient times were set in motion by the phonetic lengthening of stressed syllables, and that they continued a trend which began at the Proto-Malayo-Chamic level where less-salient segments were sacrificed. The papers deepens our understanding the oldest phonological processes in Chamic and its closest relatives.

The closest kin to Chamic and Achenese are probably the Malayic languages, Malay and its nearest kin such as Minangkabau and Iban. Blust's (1981) classification puts Malayic, Chamic and Acehnese into a sub-group he calls Malayo-Chamic (MC). The Moken/Moklen languages, which share with Chamic and Acehnese even some very specific phonological developments, such as the diphthongisation of word final **i* and **u*, must have split off well before the phonological and lexical changes that mark Proto-Malayo-Chamic. This suggests a series of stages in the Austronesian penetration of Mainland Southeast Asia well before the beginning of the common era.

Pittayaporn's paper (this volume) on "Moken as a Mainland Southeast Asian Language" challenges, largely by implication, many of the arguments and assumptions made by other contributors to this volume. By investigating in detail the historical origins of many linguistic features of Moken that have been attributed to Mon-Khmer influence by other writers he finds that the evidence suggests a more complicated history. Mon-Khmer loans are actually relatively infrequent in Moken, and they more often do not coincide with the lexicon that shows the changes from insular to mainland typology—for example, virtually all the new vowels and the feature of contrastive length emerged from changes in the phonology of the etymologically Austronesian lexicon. Pittayaporn finds that historically Moken has borrowed from a range of languages, including Burmese, Thai, Malay, Mon and other Mon-Khmer languages, in addition to considerable independent lexical innovation, and no one of these stands out as fundamentally driving the remoldeling of Moken typology.

The Moken and their close relatives the Moklen are also known (inappropriately) as the Sea Gypsies, Sea Nomads or simply the Sea People. The Moken live throughout the Mergui Archipelago in Myanmar and the Moklen further south in Thailand along the west coast and coastal islands of the isthmus of Kra. The small total population of only around 20,000, living in communities which were in some cases devastated by the tsunami of 26 December 2004, is spread out over more than 650 kms (Larish 1999:61), living by subsistence fishing and some modest forest gathering and farming activities. Further to the south is another group of "Sea People", known as Urak Lawoi', who have often been grouped with the Moken/Moklen in the literature, but they are a Malayic group that shares a similar lifestyle to the Moken/Moklen—Larish (1999:53) shows that Urak Lawoi' shares linguistic innovations with Minangkabau, placing firmly within Malayic. By contrast Moken/Moklen as a sub-group clearly diverged before the formation of Malayo-Aceh-Chamic (MAC). The most important evidence lies in the reflexes of Proto-Malayo-Polynesian (PMP) *q, *j and *R, where MAC shows /h/, /d/ and /r/ respectively, while Moken/Moklen show /k/, /y/ and /ʔ/l/n/ respectively (Larish 1999:326-7). Accepting Blust's (1994:47) estimate for the break-up of MAC at between 2300 and 2200 years B.P. would force us to date the separation of Moken/Moklen from pre-MAC to at least 2500 B.P., perhaps as far back as 3000 B.P.

Despite the ancient separation from MAC, Moken/Moklen share various features with Aceh-Chamic which are suggestive that a Sprachbund formed for a time that linked Moken/Moklen, Acehnese and Chamic sometime after the separation of Malayic from MAC but before the Thai and Burmese intrusions into the Moken/Moklen speaking area. Perhaps the most important of these is the shift to fixed wordfinal stress which conditioned various other changes such as the diphthongisation of PMP *i and *u—a strikingly Mon-Khmer type change (cf. PMK *tiiʔ 'hand' > Old Khmer *tai*, Old Mon *tey*, Lawa *taiʔ*, Car Nicobar *-tai*) in the etymologically Austronesian lexicon of Moken/Moklen, Acehnese and Chamic, and also in the Kerinci language, which is closely related to Minangkabau. Careful historical reconstruction which correctly sequences the history of phonological changes and correlates them with the indigenous and borrowed lexicon is needed to determine the specific mechanisms of change and reveal whether a real historical language area is inferred or whether an extraordinary independent parallelism has occurred.

Anthony P. Grant (Ormskirk)
Paul Sidwell (Canberra)
October 2005

References

Alieva, Natalia F. and Bui Khanh The. 1999. *Yazyk Cham: ustnye govory vostochnogo dialekta.* St Petersburg: Institute for Oriental Studies.

Blood, David L. 1962. 'A problem in Cham sonorants.' *Zeitschrift für Phonetik* 15: 111-114.

Blust, Robert A. 1981. 'The reconstruction of Proto-Malayo-Javanic: an appreciation' *Bijdragen tot de Taal-, Land- en Volkenkunde*, 137.4:456-469.

Grimes, Barbara F. (ed.). 2000. *Ethnologue, fourteenth edition.* Dallas: Summer Institute of Linguistics.

Larish, Micheal. 1999. *The position of Moken and Moklen within the Austronesian language family*, Department of Linguistics, University of Hawaii at Manoa: Ph.D. Dissertation.

Lee, Ernest Wilson. 1974. 'South East Asian areal features in Austronesian strata in Chamic.' *Oceanic Linguistics* 13: 643-670.

-----1996. 'Bipartite negatives in Chamic.' *Mon-Khmer Studies* 26: 291-317.

-----1998. 'The contribution of Cat Gia Roglai to Chamic.' Thomas (ed. 1998), *Studies in Southeast Asian languages no. 15: Further Chamic studies.* Pacific Linguistics A-89. Canberra: Australian National University. pp. 31-54.

Schmidt, Wilhelm. 1906. *Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens.* Arch. Anthrop., Braunschweig, 5:59-109.

Sebeok, Thomas A. 1942. 'An examination of the Austroasiatic language family.' *Language* 18: 206-217.

Tharp, James and Y-Bhăm Buôn Yă. 1980. *A Rhade-English dictionary with English-Rhade finderlist.* Pacific Linguistics C-58. Canberra: Australian National University.

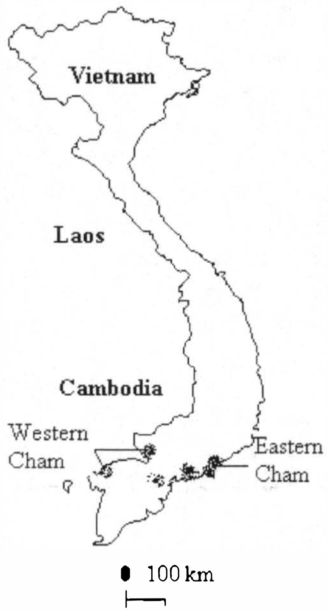
Thomas, David D. 1971. *Chrau grammar.* Oceanic Linguistics Special Publications 8. Honolulu: University of Hawaii Press.

Thurgood, Graham. 1999. *From ancient Cham to modern dialects: two thousand years of change.* Oceanic Linguistics Special Publication 28. Honolulu: University Press of Hawai'i.

-----to appear. 'Crawfurd's 1822 Malay of Champa'. Manuscript, 12 pp., to appear in a Festschrift for P J Mistry.

1 *A phonetic study of Eastern Cham register*[1]

Marc Brunelle



Eastern Cham (also Phan Rang Cham) is an Austronesian language spoken in the provinces of Ninh Thuận and Bình Thuận on the south-central coast of Vietnam. As Chamic speakers have been in contact with Mon-Khmer languages over the past two millennia, it has been claimed that contact has played a major role in the transformation of Cham from a typical Austronesian to a typologically Mon-Khmer language (Thurgood, 1996, Thurgood, 1999, Thurgood, 2002a). Nowadays, the Mon-Khmer language that has the strongest impact on Cham is Vietnamese: after the fall of the kingdom of Champa to the Vietnamese in 1471, the Cham have gradually become a small minority even in the Cham heartland[2] and have lived under the ever-growing sociopolitical dominance of the Vietnamese. For this reason, almost all, if not all, Eastern Cham speakers are bilingual, which significantly affects the structure of their language (Thurgood, 1996, Thurgood, 1999).

A feature of Eastern Cham that is often considered to be contact-induced is register. Registers are complexes of phonetic features such as pitch, voice quality and vowel quality that often accompany vowels in Mainland Southeast Asian languages (Henderson, 1952, Matisoff, 1973). It is well-established that Eastern Cham has two such registers, although their Mon-Khmer origin is difficult to prove (Blood, 1967, Bùi, 1996, Moussay, 1971). More controversially, it has been hypothesized in the past twenty years that these two registers are rapidly evolving into a tonal system under the influence of Vietnamese (Hoàng, 1987, Phú et al., 1992,

[1] The preliminary work that has lead to this paper has been funded by a doctoral grant from the Fonds Formation de Chercheurs Aide Recherche (FCAR) of the Government of Québec. Fieldwork was funded by a doctoral grant from the Social Sciences and Humanities Research Council (SSHRC) of Canada. I would like to thank Pr. Bùi Khánh Thế, Phú Văn Hẳn and Nguyễn Văn Tỷ for sharing their knowledge with me and helping me with fieldwork-related issues and Abigail Cohn, John Wolff, Amanda Miller-Ockhuizen and Tejaswini Deoskar for their insights and their comments on earlier versions of this paper.

[2] Currently, there are 40,000 or 50,000 Cham out of 1,300,000 people in Ninh Thuận and Bình Thuận provinces (Phan et al., 1991).

Brunelle, M. "A phonetic study of Eastern Cham register". In Grant, A. and Sidwell, P. editors, *Chamic and Beyond: Studies in mainland Austronesian languages*.
PL-569:1-36. Pacific Linguistics, The Australian National University, 2005. DOI:*10.15144/PL-569.1*

Thurgood, 1996, Thurgood, 1999). This case of contact-induced sound change can be investigated from three angles: phonetics, phonology and sociolinguistics. In this paper, I look at the phonetic evidence about Eastern Cham registers. I argue that while pitch plays a central role in the Eastern Cham register contrast, registers are still relatively conservative and are not evolving into a full-fledged tone system. A study of sociolinguistic variation in the realization of register and of its phonological status is beyond the scope of this paper, but is addressed in Brunelle (2005a).

In Section I, I discuss the diachronic developments that have led to the formation of Cham register and review the arguments that have been proposed in favor of a tonal analysis of Eastern Cham. In Section 2, I describe the realization of coda consonants and register allophony and show that Eastern Cham cannot be treated as a full-fledged tone language. Finally, in Section 3, I provide the reader with an acoustic description of Eastern Cham register and look at the perception of register by native speakers. I argue that despite the fact that the registers of Eastern Cham "look tonal", there is little phonetic evidence that they have really evolved into tones.

1. Historical developments and previous work

While Ancient Cham had contrastive voicing in onset stops (still reflected in writing), Modern Cham dialects have neutralized this voicing contrast in favor of the voiceless series (except in preglottalized stops). The role of voicing was taken over by a register distinction on the vowels following stops: while vowels following a former voiced stop took on a breathy quality and a low pitch, all other vowels kept a mid-range pitch and a modal voice. There is also evidence that before the split between Eastern and Western Cham, vowel quality might have played a role in the system : in Western Cham, high register vowels are typically realized as more open than low register vowels (Edmondson and Gregerson, 1993, Headley, 1991). We can hypothesize that the original Cham register system had the following features:

(1) Phonetic properties of registers

***Voiceless obstruents, sonorants, implosives >**	***Voiced stops >**
High Register	**Low register**
High pitch	Low pitch
Modal voice	Breathy voice
Lower vowels?	Higher vowels?

It is difficult to date the formation of register, but two 19th century sources list Cham words with voiced stops. The first is a short wordlist compiled by John Crawfurd during a short stay in Vietnam (Thurgood, 2002b). As pointed out by Thurgood, Crawfurd's list is not necessarily reliable, but a later source, Etienne Aymonier's *Grammaire de la langue chame,* contains further evidence that the merger was not complete in the 19th century (Aymonier, 1889). In his grammar, Aymonier does not mention that the voiced onsets of Classical Cham are devoiced in speech, despite a thorough examination of the production of every letter of the Cham script. The only reference to melody is on page 34:

> " Enfin, nous terminerons cette étude de l'alphabet usuel en faisant remarquer la forme particulière, peu usuelle pour ainsi dire, des quatre consonnes aspirées gha, jha, dha, bha. Les mots chames qui les emploient sont assez rares. Au Cambodge, les étudiants lisant l'alphabet chame laissent tomber la voix sensiblement sur ces quatres lettres, comme dans les mots annamites affectés de l'accent grave."
>
> *"Finally, we will conclude this study of the common script by noticing the peculiar, unusual, form of the four aspirated consonants gha, jha, dha, bha. Cham words making use of them are rather rare. In Cambodia, students reading the Cham alphabet let their voice fall on these four letters, as in the Annamese words marked with the grave accent."* (my translation)

In this passage, Aymonier points out that Cambodian Cham speakers pronounce voiced aspirated stops with an intonation reminiscent of the low level tone of Vietnamese. However, there is no mention of devoicing, of a special intonation on plain voiced stops or even of the pronunciation of Eastern Cham speakers, with whom Aymonier did most of his work. This could be interpreted as evidence that the voiced stops of Eastern Cham were not devoiced when Aymonier wrote his grammar and that the characteristic low pitch of the low register was either not clearly audible after unaspirated stops or still masked by the voicing of the onset.

If this interpretation of Aymonier is correct, voicing neutralization in onsets and the resulting registrogenesis could hardly be due to contact. At the time when Aymonier wrote his grammar, Vietnamese, a language that does not have typical Mon-Khmer register, was the only language in contact with Eastern Cham. Some register languages belonging to the Bahnaric branch of Mon-Khmer were spoken in neighboring provinces, but contacts with these languages must have been episodic at best, since the Vietnamese forbade contacts between the Cham and other minority groups after two multiethnic revolts in the 1830's (Po, 1987). Another possibility is that Aymonier and Crawfurd's descriptions are simply inaccurate and that register developed much before the arrival of the French in the late 19th century. This is suggested by the presence of registers in all coastal Chamic languages. Since the common ancestor of the Coastal languages presumably split a few centuries ago, the preservation of the voicing contrast in 19th-century Eastern Cham entails that Coastal languages developed register independently, a rather uneconomical scenario.

In any case, the first modern descriptions of Eastern Cham by Christian missionaries are clear: contrastive voicing had been lost by the 1960's (Blood, 1967, Moussay, 1971)[3]. These sources also emphasize the fact that the two registers of Eastern Cham have different allophonic realizations conditioned by their codas, although the descriptions of these realizations conflict to some extant. Moussay, for example, lists four allophones (p. XIII):

- a level "tone" on vowels preceded by voiceless onsets and followed by all codas but the glottal stop
- a low tone (*ton grave*) on vowels preceded by voiced onsets and followed by all codas but the glottal stop
- a rising tone (*ton quittant*) on vowels preceded by voiceless onsets and followed by a glottal stop

[3] Although this is contradicted by Mr. Lưu Quý Tân, in a personal communication to André Haudricourt (Haudricourt, 1972).

- a falling tone on vowels preceded by voiced onsets and followed by a glottal stop

In contrast, Blood treats the two registers as two pitch "phonemes", but states that "before final stops and the *h* the register of non-low pitch is higher than in syllables ending in the other consonants or silence." (p.29). Despite their different descriptions or register allophony, Moussay and Blood do agree that codas have an allophonic effect on the pitch of the two registers.

Recently, a few scholars have published work in which they treat the coda-conditioned allophones as phonemic or incipiently contrastive (Hoàng, 1987, Hoàng, 1989, Phú et al., 1992). A crucial tenet of these hypotheses is that some final consonants (especially laryngeals) are weakened or dropped, leading to a reinterpretation of the coda contrast as a pitch contrast. While it is uncontroversial that some coda stops have undergone reduction (-p > -wʔ, c > jʔ), other alleged consonantal changes are more speculative. For example, the claim that "... the stops [-p, -t, -k] have fallen together as glottal stop and *h* has been lost altogether" (Phú et al., 1992) is not borne out by the findings presented in Section 2.

Before discussing my own findings about the registers and tones of Eastern Cham, a summary of the only experimental study of these issues to date is in order. Phú et al. (1992) recorded three minimal pairs from one male native speaker of Eastern Cham (following Moussay, I use the subscript dot to mark the low register)[4]:

(2)	High register		Low register	
	/pa/	'where, at'	/p̥a/	'to carry'
	/paʔ/	'four'	/p̥aʔ/	'to walk'
	/păʔ/	'straight'	/p̥ăʔ/	'to tap'

The authors then measured and compared the f0 curves (pitch) of the three pairs. As expected, vowel pitch is consistently lower following /p̥/. The low register also has at least two realizations: a rising pitch before the glottal stop and a low level pitch in open syllable. However, no such split was found in the high register. In that register, the open syllables and the syllables closed by a glottal stop have similar shapes and height. Phú et al. (1992)'s results therefore suggest that Eastern Cham has at least three surface register allophones.

Based on these empirical results, the authors go a step further and propose that coda glottal stops could have become "part of the internal stuff of a given tone" (p.41). The glottal stop would have lost its status of coda consonant to become a tonal element, a part of a glottalized tone. This amounts to saying that the low register would have split into two distinctive tones. In other words, Phú et al. put forward the possibility that Eastern Cham is already a three-tone language. This is to my knowledge the only explicit and refutable scenario for the development of a complex tone system in Eastern Cham. Unfortunately, despite the fact that the authors are careful not to jump to conclusions, their reasons for treating the glottal stop as a tonal property remain unclear.

[4] Besides the fact that /pa/ means 'where' only as an exclamative, which could affect its pitch, there is another problem with the wordlist: according to the first author and subject of the experiment, the word /p̥aʔ/ 'to take a walk' is his "modern rendition" of the Ancient Cham word /kalip̥aʔ/ and is not normally used in speech. Therefore, his pronunciation could be relatively artificial.

Have registers really evolved into tones or are they still a property of onsets? In order to answer this question, I will first look at the phonology and phonetics of registers in Section 2. The realization of codas and their effect on register will then be explored in Section 3.

2. Codas and Tones

In this section, I investigate the status of coda consonants and I describe their effect on the phonetic realization of registers. More specifically, I show that coda consonants condition register allophony, but that this allophony has not been reinterpreted as contrastive tone.

Coda stop weakening is not a recent phenomenon. In the late 19th century, Aymonier already noted that the final graphemes *-p*, *-c* and *-k* were being reduced (Aymonier, 1889):

> "k se prononce faiblement à la fin de beaucoup de mots dont il rend la prononciation brève et saccadée." (p. 32)
> *"k is weakly pronounced at the end of many words and makes their pronunciation short and abrupt." (my translation)*
> "Le p final se reconnaît facilement à l'oreille dans certains mots tels que gâp, mutuel, mais il est bien difficile à un Européen de saisir cette consonne dans d'autres mots tel que hudiêp, femme, vivant." (p.32)
> *"Final p is easily recognizable in some words like gâp, mutual, but it is difficult to perceive it for a European in other words like hudiêp, wife, alive." (my translation)*
> "En somme, à la fin des mots, les consonnes k et p ne se prononcent presque pas et donnent au mot un arrêt un peu brusque de la voix, où une oreille fine et exercée peut seule reconnaître la nature de la consonne." (p. 32)
> *"In short, at the end of words, the consonants k and p are almost not pronounced and give to the word a rather abrupt interruption of the voice, where only a fine-tuned and trained ear can recognize the nature of the consonant." (my translation)*
> "Le ch final du chame est prononcé à peu près comme i ou y dans la plupart des mots. Exemples : lach, dire; ach, incurie; baganrach, grand plateau des sacrifices, sont prononcés laï ou lay ou ay, baganray, etc." (p.33)
> *"The final ch of Cham is pronounced roughly like i or y in most words. Examples: lach, to say; ach, caresslessness; baganrach, large sacrificial tray, are pronounced laï or lay or ay, baganray, etc." (my translation)*

Although they are somewhat impressionistic, these descriptions totally agree with the type of coda reduction that is found today. The first passage describes the modern reflex of written Cham *-k* as a glottal stop, which is still the normal realization of this coda today. The other passages also reflect a state of affair identical to what is found in the modern language: stops are often reduced, especially in high-frequency words, but they are never deleted.

2.1. Experiment

An acoustic study of final consonants and of their effects on vowels was carried out to determine the type of changes that final consonants are really undergoing and to evaluate the claim that the loss or neutralization of final consonants has caused the development of contrastive tone.

2.1.1. Methods

A wordlist designed to test the phonetic realization of register and the effect of codas on pitch was recorded with 43 native speakers of Eastern Cham. The wordlist was composed of all possible monosyllabic words with the vowels /a:/ and /ă/[5], starting with the labial onsets /p, p^h, b, m, w, ʔw/ and combined with all the possible Written Cham codas <p, t, c, ʔ, m, n, ŋ, j, w, h, 0> (Written Cham is transcribed in brackets < >). All possible combinations of these factors were computed, resulting in a list of 252 possible words. I then went through this list with Phú Văn Hẳn, a Cham linguist, and excluded meaningless monosyllables. A few words with dental sonorant onsets were then added to make sure that enough sonorant-initial words would be included in the wordlist, yielding a list composed of 99 real words.

The wordlist was originally designed to be read. However, since very few speakers could read the Cham script fluently, I quickly abandoned the initial idea of working with a wordlist written in this script. To further complicate things, many Cham are hostile to romanization (Blood, 1977, Blood, 1980) and many speakers simply refused to try to read a romanized wordlist. For these reasons, only three speakers read the wordlist. All other speakers were given the target words in Vietnamese and asked to translate them in Cham. The speakers were then instructed to repeat them at least three times in a frame sentence[6]. Whenever speakers were not familiar with a word, it was not recorded[7]. All recordings were made with a Marantz PMD-680 card recorder and an AKG C5900 microphone.

The frame sentence used is given below:

(3) /t̥ahlăʔ dom akhăn ____ ka ɲu păŋ/
I say word ____ for he hear
"I say the word ____ for him"

Minor variations in the frame sentence were allowed (/kăw/ 'informal I' instead of /t̥ahlăʔ/ 'formal I', /aj/ 'brother' instead of /ɲu/ 'he'). Further, most speakers consistently realized /t̥ahlăʔ/ and /akhăn/ as /hl̥ăʔ/ and /khăn/, their colloquial monosyllabic correspondents. A majority of speakers were comfortable with the frame sentence, but a few of them had to be trained for a few minutes before the recording session.

Some target words were realized as sesquisyllables by a few speakers. I cannot discuss this question in detail here, but colloquial Eastern Cham has become almost entirely monosyllabic, except in very formal speech (Brunelle, 2005a, Brunelle, *to appear*). For the purpose of this experiment, whenever a word was realized as a sesqui/polysyllable, only the final stressed syllable was measured.

Since the wordlist just described only includes words with the vowels /a:/ and /ă/, I also recorded a second wordlist consisting of 38 words including all possible other vowels combined with 5 codas that have been claimed to be reduced or dropped by Phú et al. (1992). It is much less systematic than the first wordlist in that it does not exhaust all

[5] Note that short /ă/ is often allophonically realized as /ɛ̆/ before /-j/ and /-t/ and as /ɔ̆/ before /-w/.

[6] With the first three speakers, I recorded the entire wordlist three times, consecutively. However, as this procedure took too long, I made the decision of recording each word three times with the 40 remaining speakers.

[7] Some lexical items vary from village to village. Learned and semi-learned words are not widely known.

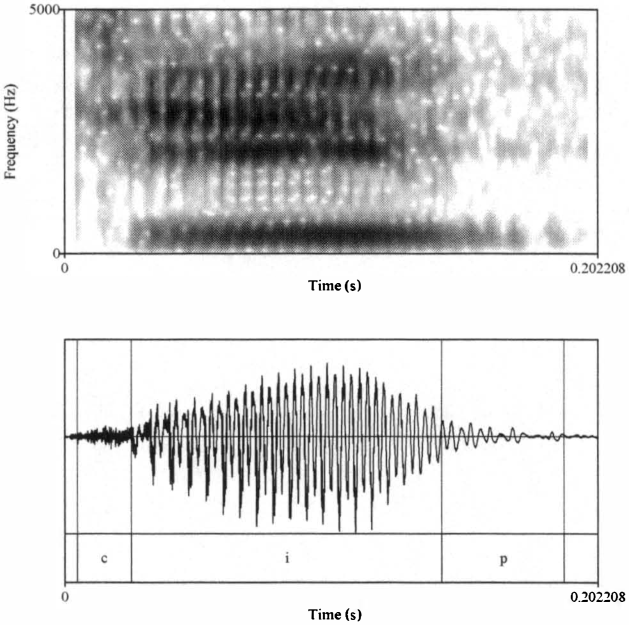
possible rimes and onsets, but it was included for comparative purposes. The target words of the second wordlist were recorded in the same frame sentence as the first wordlist with 14 speakers and in isolation with 26 speakers who showed less patience (these speakers are a subset of the speakers who read the first wordlist). The recording sessions were conducted identically.

The determination of the place and manner of articulation of the various codas was done through a visual and auditory inspection of the waveforms and spectrograms of all target words with the acoustic software *Praat 4.2* (Boersma and Weenink). Coda stops were categorized as either fully realized, debuccalized to a glide + glottal stop sequence, reduced to a simple glottal stop or deleted. Coda <h> was categorized as either fully realized or missing.

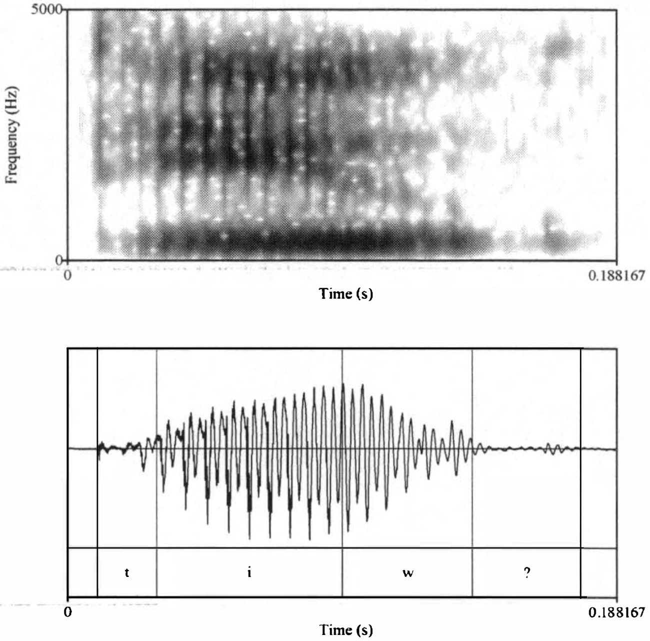
2.1.2. Results

Overall, the results suggest that there is relatively little variation in the realization of written Cham codas in modern Eastern Cham. Individual speakers always realize the coda of a specific word consistently. There is a limited amount of variation across speakers, but it is typically restricted to low frequency words. Coda stops can be either realized as unreleased stops or be debuccalized to a glide followed by a glottal constriction. On the other hand, final laryngeals are almost never deleted except in a few function words.

In order to illustrate what is meant by full realization and debuccalization of coda stops, a short illustration and discussion of the behavior of coda /-p/ follows. This coda was chosen because it is the only one that has two robust variants. Spectrograms of other codas are given in Brunelle (2005a). The modern reflexes of final <-p> are the full stop /-p/ (Figure 4) and the labio-velar glide followed by a glottal stop /-wʔ/ (Figure 5). The word /c̥ip/ 'clear, understandable' is realized with an unreleased [p]. The vowel preceding it has stable formants that are interrupted relatively abruptly by the closure of the coda stop. By contrast, at the end of the vowel of /t̥ĭwʔ/ 'wife', vocal fold vibrations are much more irregular and F2 gradually goes down as the high front vowels turns into a labio-velar glide. Note that in (4), there are still vocal fold vibrations during the [-p], indicating partial voicing. This weak voicing is often visible on waveforms and spectrograms, but is rarely audible.



(4) Waveform and spectrogram of /c̥ip/ pronounced by a man born in 1966



(5) Waveform and spectrogram of /t̪̥iwʔ/ pronounced by a man born in 1966

As is the case for the words /c̥ip/ and /t̪̥iwʔ/, specific lexical items can be realized with either [-p] or [-wʔ], but do not vary. Learned words typically have a [-p], while common words tends to have the debuccalized form. The realization of other codas in Eastern Cham is even less variable as we can see from the modern realizations of Written Cham codas after /a:/ and /ă/ given in Table (6). It gives the realization of each stop (including the glottal stop) after various vowels for 43 speakers. The number of tokens should in theory equal the number of words multiplied by three repetitions and 43 speakers. However, some words were unknown to some speakers and were not recorded, and some tokens had to be excluded because of background noise or other recording problems. Therefore, the total is typically below the possible maximum number of tokens.

(6) Realization of written Cham codas after /a:/ and /ă/ (43 speakers)

Written Coda	<p> (3 words)	<t> (8 words)	<c> (5 words)	<ʔ> (12 words)	<h> (11 words)
Total	200	848	0	1376	1398
Debuccalized	0	0	611 [-jʔ]	[redacted]	[redacted]
Dropped	0	0	0	11	0
% Dropped	*0%*	*0%*	*0%*	*0,799%*	*0%*

First, coda <p> is always realized as a full stop, because the wordlist was not originally designed to test the ways in which codas are reduced, but the effect of codas on the realization of registers. The few items ending in <p> that were included in this list are learned words, which explains the lack of debuccalization of /-p/. In colloquial speech, there are numerous instances of debuccalization of /-p/ after /a:/ and /ă/ (one of them is included in the second wordlist below). Other consonants have a single surface realization: <t> is always realized as [-t], <c> is always surfaces as [-jʔ] and <h> is always fully pronounced. Final <ʔ> seems to be occasionally missing, but since this happens in less than 1% of the tokens, this could hardly be used as evidence of the loss of coda glottal stops.

The realization of coda stops after other vowels is similar, although the idiosyncratic behavior of some words needs to be discussed in some detail. The general results are given in Table (7). This time, the number of tokens in each box of Table (7) should be equal to the number of words multiplied by three repetitions and 40 speakers. However, as for the first wordlist, unknown words were excluded, which results in the total number of words being lower than the theoretical maximum.

To the exception of <-p>, all codas have one very predominant realization that suffers few exceptions. As we have seen before, <-p> has two possible modern reflexes, [-p] and [-wʔ], depending largely on the frequency and status of the word. The experiment further suggests that the choice of one variant over the other is not predictable from voice quality. Only seven words written with final <p> are reported in (7), but my observations of unrecorded speech support this result. There is of course a possibility that some vowels have a probabilistic, non-categorical effect on the choice of a reflex (for example, [-p] could be more common after back vowels than front vowels), but this has not been investigated systematically. As was the case in Table (6), the behavior of other codes is more categorical. Coda <-t> is always realized as [-t] except in the word <hak̥ɛ̆t> 'what', which can be realized as [kɛ̣̆ʔ~ k̥eʔ ~ k̥e], but is a high frequency function word that could be argued to have an underlying coda glottal stop in the modern colloquial language. Final <-c> is very consistently realized as [-jʔ], except for seven cases of [-j], mostly in the word

<ʄɨc> [ʄɨjʔ] 'seaweed'[8]. Finally, although the Cham script does not distinguish /-k/ and /-ʔ/, they are clearly distinct in the modern language, despite the fact that /-k/ is very rare in non-learned words. The only word ending in /-k/ used in this experiment was <tĭk> 'teapot'. It was always realized with a coda [-k]. Words ending in <-ʔ> in written Cham are also consistently realized with a full glottal stop in modern Cham. Only one word out of 1539 has lost its final <-ʔ>, which clearly shows that coda glottal stops are not deleted. The surprising occurrence of six instances of [-h] is due to the word /ç̥eʔ/ 'to knead', which seems to have two variants, a common one with a coda /-ʔ/ and a less frequent one with a coda /-h/[9].

(7) Realization of Written Cham codas after other vowels (40 speakers)

Written Coda	**<p>** (7 words)	**<t>** (10 words)	**<c>** (5 words)	**<k>** (1 word)	**<ʔ>** (14 words)
Modern realization	546/762[wʔ] 210/762 [p] 3/762 [wp]	972/1080[t] 78/1080 [ʔ] 24/1080[∅] 6/1080 [k]	522546[jʔ] 21/546 [j] 3/546 [∅]	120/120[k]	1512/1539[ʔ] 18/1539 [h] 3/1539 [∅] 3/1539 [p] 3/1539 [k]
% Full stop	***27.6%***	***90%***	*0%*	***100%***	***98.2%***
% Debuccalized	***72%*** [wʔ]	*7.2%* [ʔ]	***95.6%*** [jʔ]		
% Dropped	*0%*	*2.2%*	*0.5%*	*0%*	*0.2%*
% Other	*0.4%*	*0.5%*	*3.8%*	*0%*	*1.6%*

Keeping all these exceptions in mind, we can now summarize the facts presented in (6) and (7) in the following way: Written Cham words ending in <p> are realized in the modern language with either /-p/ or /-wʔ/, with only a minimal amount of variation in the realization of individual words which probably reflects linguistic insecurity rather than actual variation in normal speech. Other codas behave even more consistently: If we exclude the word <hak̥ɛ̆t>, already discussed above, coda <t> is almost always realized as a full stop. The final palatal stop <-c> is systematically realized as [-jʔ], and the only word with a final /-k/ that was looked at did not vary either. Finally, laryngeal /ʔ/ is realized as a full glottal stop and shows few signs of being dropped.

2.1.3 Discussion

The results show that the claim that "... the stops [-p, -t, -k] have fallen together as glottal stop and *h* has been lost altogether" (Phú et al., 1992) is not an accurate characterization of the realization of codas in Phan Rang Cham[10]. In the experiment, the laryngeal /-h/ is never dropped and the oral stops, although frequently debuccalized, are never realized as /-ʔ/ except in one word, <hak̥ɛ̆t> 'what', which can be argued to have a final glottal stop synchronically. Of course, the data discussed here come from a relatively formal situation,

[8] This is not sufficient to claim that there is a tendency to reduce [-jʔ] to [-j]. In the word [ʄɨjʔ], the vowel is strongly glottalized due to glottal constrictions in both the onset and the coda. It is possible that some listeners have reinterpreted this glottalization as stemming exclusively from the onset and have lexicalized the word as [ʄɨj].

[9] Final [-h] is not a regular realization of Common Cham *-ʔ.

[10] My observations in Bình Thuận suggest that the same holds for the Cham dialects spoken there.

wordlist recording, where speakers are likely to speak a language variety unaffected by some phonological processes applying only in colloquial speech. However, short interviews carried out with the same speakers do not show coda deletion or neutralization either and, for what it is worth, my impressions of unrecorded running speech go in the same direction.

Even in the case of <-p>, which can be realized as a full stop or as a glide followed by a glottal constriction, the two variants do not seem to occur in different utterances of the same word, even in different speakers. In (6), some words have one divergent speaker out of 43, but it is likely that these unexpected variations are due to affectedness and to the speaker's awareness that the two codas are written with the same grapheme in the conservative Cham script. In the modern language, the debuccalized coda [-wʔ] is possibly not a free allophonic variant of /-p/ anymore, but should be analyzed as a sequence of /w + ʔ/.

The data presented in this section clearly argue against a simplistic description of Eastern Cham in which codas are dropped and the register allophones preceding them become contrastive. However, it does not address two more interesting issues, namely the realization of the coda-conditioned register allophones and the phonological status of codas, more specifically that of the glottal stop, which could have become a part of the tones, while still being realized on the surface. These questions are addressed in Sections 2.2 and 2.3, respectively.

2.2. Coda-conditioned register allophony

Now that I have established that codas have not been deleted and are still realized on the surface, what is their exact effect on the pitch height and contour of the registers? An acoustic experiment was carried out to determine the nature of this effect.

2.2.1 Methods

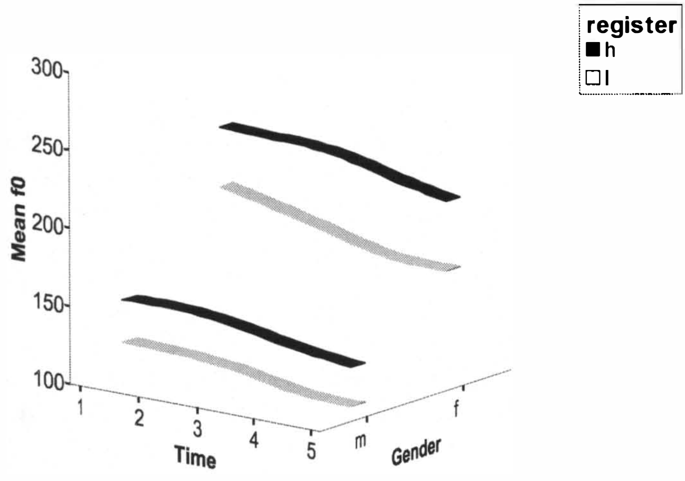
All words containing the vowels /a:/ and /ă/ recorded from 43 speakers for the previous experiment were used to determine the realization of pitch before the various codas. Pitch (f0) was measured with Praat 4.2 at the beginning and endpoint of the vowels and at three equidistant intermediate points.

2.2.2 Results

As we have seen in the introduction, the exact realization of the register allophones is the subject of conflicting descriptions. In order to quantify the data across speakers, the pitch allophones of male and female speakers have been averaged out. As most previous discussion revolves around open syllables and syllables closed by glottal stops, I have plotted the allophones of three pairs of words in Charts (8-10):

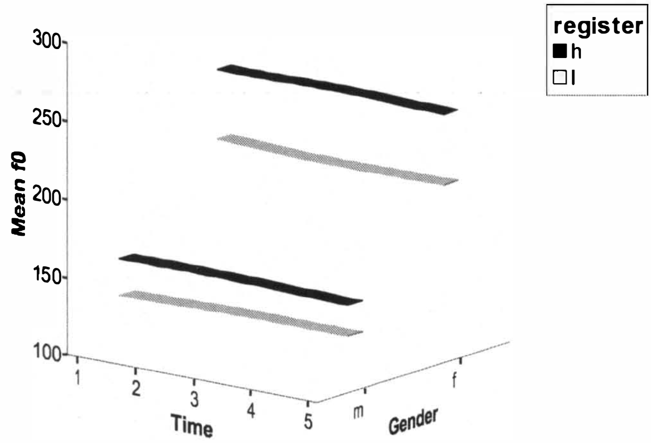
- Open syllable words: /pa/ 'to cross' ~ /p̥a/ 'to carry' (8)
- Syllables with a short vowel closed by a glottal stop /păʔ̣/ 'at' ~ /p̥ăʔ/ 'full' (9)
- Syllables with a long vowel closed by a glottal stop /paʔ/ 'four' ~ /p̥aʔ/ 'to take a walk' (10).

In these charts, the mean duration of long vowels is 155 ms. compared to 79 ms. for short vowels.



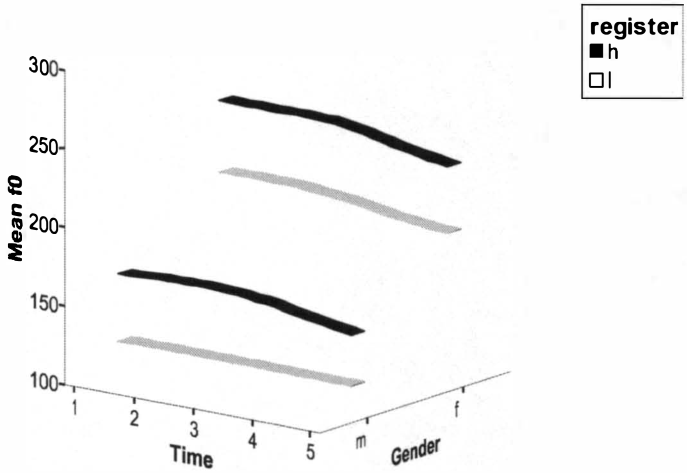
(8) f0 curves of /pa/ and /p̥a/, (20 women, 23 men)

In open syllables (8), the f0 of the two registers is slightly falling for both men and women, the overall pitch range of women being much higher than the pitch range of men. As expected, the pitch of the high register is higher than the pitch of the low register. By contrast, in syllables with a short vowel closed by a glottal stop (9), f0 at the onset of curves is higher by 10-20 Hz and is level instead of falling.



(9) f0 curves of /păʔ/ and /p̥ăʔ/, (20 women, 23 men)

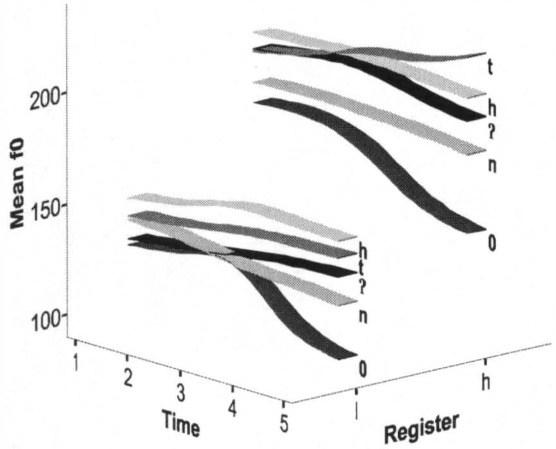
Syllables with a long vowel closed by a glottal stop (10) fall in between: While their overall f0, especially at the beginning of the curve, is about 10-20 Hz higher than in open syllables, their pitch contours are not level like the pitch contours of their short vowel counterparts, but rather slightly falling like the pitch of open syllables. The non-falling pitch of the short vowels could be a consequence of their duration: pitch has a tendency to fall, but the drop does not have time to occur in short vowels.



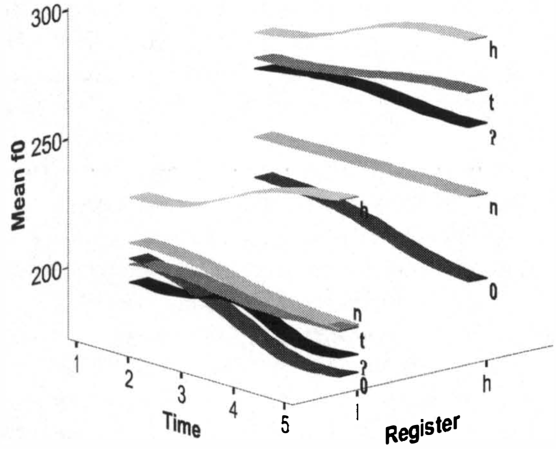
(10) f0 curves of /paʔ/ and /p̥aʔ/ (low register: only three speakers)

Chart (10) requires a caveat: only three educated speakers were aware of the existence of the word /p̥aʔ/, 'to take a walk', and they all insisted that this word is not used in normal speech. It seems to be an artificial colloquial rendition of the written Cham word <kalip̥aʔ>. Therefore, results for the low register on a long vowel closed by a glottal stop may or may not be meaningful, although they go in the same direction as their high register counterpart.

Having compared the pitch of open syllables and syllables closed by glottal stops, we can now look at the effect of other codas on the pitch of the two registers. I give representative data from one male speaker and one female speaker in (11) and (12). To avoid overcrowding the charts, I have chosen to illustrate the f0 curves of vowels belonging to both registers before a limited set of representative codas. Besides open syllables, the two laryngeal codas /-h/ and /-ʔ/ have been included, along with one coda sonorant /-n/ and one coda stop /-t/. We see in both figures that the mean f0 of open syllables has a falling curve that contrasts with the relatively flatter f0 of other allophones. Moreover, open syllables tend to have the lowest overall pitch height. Vowels closed by an /-n/ are a little higher in pitch and also have a slightly falling contour. The other consistent fact is that vowels closed by an /-h/ have a relatively high pitch contour, which is rising for the female speaker (12), but level for the male speaker (11). The remaining two codas, /-ʔ/ and /-t/, are less predictable: their relative pitch is high and flat in the man's speech, but there are inconsistencies between registers in the women's speech. The exact realization of pitch in front of various codas also tends to vary across speakers, a variation in all likelihood due to differences in the exact degree of laryngeal constriction during production of the glottal stop (and possibly /-t/).



(11) f0 curves of the coda-conditioned allophones of the two registers, male speaker born in 1933



(12) f0 curves of the coda-conditioned allophones of the two registers, female speaker born in 1950

2.2.2. Discussion

We have seen in Section I that authors disagree on the exact nature of coda-conditioned register allophony. Their descriptions are summarized in (13). The only syllable types for which all authors agree are the open syllable and the syllable closed by a sonorant. While

Moussay has a symmetrical four-allotone system (2 allophonic contexts X 2 registers), Blood and Phú *et al.* have two allotones in one register and one in the other, the difference being that the register exhibiting allophony is the high register for Blood, but the Low register for Phú *et al.*

(13) Effect of codas on the pitch of the two registers

Register	Coda	Blood (1967)	Moussay (1971) Hoàng (1987)	Phú *et al.* (1992)
High	Sonorants	High	Level	Not tested
	Open syllable			High
	Glottal stop	Higher	Rising	
	-h		Level	Not tested
	Oral stops			
Low	Sonorants	Low	Low	Not tested
	Open syllable			Low
	Glottal stop		Falling	Rising
	-h		Low	Not tested
	Oral stops			

If we exclude the impressionistic descriptions and focus on experimental work, a comparison of the results presented in (11) and (12) with Phú *et al.* (1992)'s findings highlights a few similarities and many differences. As the two studies look at the f0 contours of the *same* Cham words, these differences are puzzling. In both datasets, the pitch height of the register allophones is higher in front of a glottal stop than in open syllables. However, the pitch contours of the various register allophones are very different. The contour of open syllable words in Phú *et al.* is level, while it is falling in our experiment. The same is true of words with a high register long vowel closed by a glottal stop. The way in which these words were uttered could explain these basic differences: Phú *et al.* have recorded the words in isolation whereas they were recorded in a wordlist in the present experiment. Unfortunately, this does not account for the very significant discrepancies in the pitch contour of low register words closed by glottal stops. According to Phú *et al.* (1992), this contour is rising on long vowels and rising-falling on short vowels, a result that contrasts with the level and slightly falling contours found in our present experiment.

One possible explanation for these conflicting descriptions could be variation between speakers and between dialects. However, despite some definite differences between the pitch contours and heights of the different speakers recorded for this experiment, the overall similarities are strong enough to suggest that the mismatch between the different descriptions given in (13) is due to their impressionistic nature rather than to actual production differences. More data on more varieties of the language is obviously needed, but perhaps we also need to abstract away from simple description and to consider the significance of allophony and its variation.

Knowing that codas are maintained in Eastern Cham, would we necessarily expect allophony to be consistent across speakers? Obviously, there should be broad similarities between speakers, but as long as contrast is encoded in the coda rather than the pitch contour or height of the allophones, variation will not lead to confusion and will not hinder communication. Therefore, some variation could easily be maintained. For example, if the transition from modal phonation to glottal constriction is very abrupt and crisp at the end of a vowel, the pitch of that vowel could be kept constant (or even rise slightly because of a tensing of the vocal folds) until the beginning of the glottal stop. By contrast, if a speaker produces glottalization by gradually constricting their glottis at the end of the vowel, then glottal adduction will impede their vocal fold vibration and cause a gradual drop in pitch. If pitch allophony had become contrastive, i.e. if Eastern Cham had tones, this type of variation in the speech signal would not be expected, because it would hinder tone discrimination by listeners.

Taking this in consideration, it becomes meaningless to try to subdivide each register in two or three allotones, as has been proposed by most other authors so far. Obviously, each coda has its own effect on the pitch of the vowel preceding it. Charts (11) and (12) show that the allotones cannot be arbitrarily forced into a small set of discrete categories, as we would expect if there was phonological allotony, but that they rather spread across the whole pitch range, without any cut-off boundaries, a good indicator that the process in strictly phonetic.

Ultimately, the description of coda-conditioned register allophony is an empirical question that would be relevant to the question of tonogenesis only if it could be demonstrated that the register allophones have been phonemicized or that all or some codas are optionally dropped or reanalyzed as suprasegmentals. The frequent variation in the realization of the pitch contour and height of the allotones argues against phonemicization. Further, it has been shown in Section 2.1 that codas are not deleted. Therefore, the only remaining argument in favor of phonemicization of register allotones, i.e. tonogenesis, would be that glottal stops or other consonants are realized on the surface but have been phonologically reanalyzed as suprasegmentals. This argument is evaluated in the next section.

2.3. Evidence against a suprasegmental glottal stop

Since Eastern Cham codas are preserved, the coda-conditioned variants of the two registers are predictable, and we cannot treat them as phonemic tones. However, could some of the Eastern Cham codas be realized on the surface, but be phonologically analyzed as suprasegmentals? More specifically, can a glottal stop be "a part of the internal stuff of a given tone" (Phú et al., 1992), while still being realized on the surface just as if it was a coda?

It is well-known that many Southeast Asian tones have glottalized tones, i.e. tones that are accompanied by a glottal constriction (or creakiness). Standard Vietnamese, to choose a language in close contact with Eastern Cham, has two glottalized tones, called *nặng* and *ngã*, that are respectively a falling tone closed by a glottalization and a falling-rising tone broken by glottalization (Brunelle, 2003, Han, 1969, Hoàng, 1986, Michaud, 2005, Nguyễn and Edmondson, 1997, Phạm, 2001, Vũ, 1982). On the surface, open syllables with the tone *nặng* sound just like words with a low falling tone and a coda glottal stop. However, the phonology of Vietnamese provides ample evidence that this glottalization is a part of the tone. Although Vietnamese can only have simplex codas, *nặng* is found on words ending in sonorants and a few types of phonological processes like reduplication and a word game involve alternations between *nặng* and other tones. No such processes are found Eastern Cham. To my knowledge, there is not a single piece of evidence that coda glottal stops have become suprasegmental in that language: glottal stops are always phased with the end of the rime, they can never be combined with other codas and they are never separated from their codas.

3. Onsets and Registers

We have seen in the previous section that there is no evidence that Eastern Cham already has a full-fledged tone system stemming from the loss of consonantal contrast in codas. I argue elsewhere that there are reasons to believe that Eastern Cham is not even evolving in that direction (Brunelle, 2005a). However, the two registers of Eastern Cham could still have become a simple two-tone system (Blood, 1967, Thurgood, 1999). This possibility in explored in the next few pages. In Section 3.1, I describe the similarities between the registers of Eastern Cham and more typical forms of tones. In Section 3.2, I then proceed to an acoustic analysis of register, followed by a perceptual study in Section 3.3. Based on the results on the results of these two sections, I conclude that, although a two-tone analysis cannot be excluded based on phonetic data only, the registers of Eastern Cham are better treated as a relatively conservative form of register.

3.1. The tonal appearance of Eastern Cham registers

Superficially, Eastern Cham is similar to tone languages in three respects. First, a combination of two diachronic processes, register-spreading and monosyllabicization, has led to the emergence of a number of sonorant-initial minimal pairs distinguished only by their register:

(14)	*Written Cham (reading)*	*Colloquial Eastern Cham*	*Gloss*
	<ini> [ini]	/ni/	‘this, here’
	<p̥ani> [p̥an̥i]	/n̥i/	‘nativized Islam’
	<ala> [ala]	/la/	‘snake’
	<p̥ila> [p̥il̥a]	/l̥a/	‘ivory’
	<talah> [talah]	/lah/	‘lost’
	<t̥alah> [t̥al̥ah]	/l̥ah/	‘tongue’

In many register languages, register is neutralized in sonorant-initial syllables. The register contrast is typically restricted to stop-initial syllables, where it originates. In contrast, co-occurrence restrictions between tone and onsets are rare. The fact that the register contrast of Eastern Cham is found in all types of onsets except implosive stops and preglottalized glides is thus reminiscent of tone.

Second, the phonetic correlates of register are realized on the vowel rather than the onset, with the exception of pitch and amplitude which can be realized on onset sonorants. Further, as emphasized in most descriptions, pitch plays a major contrastive role in the register system of Eastern Cham. As these three characteristics are also found in tone, we could claim that Eastern Cham registers have evolved into a two tones.

However, there is also evidence that Cham is not a tone language. It comes from a word-game called *ɗom k̥ac* ‘inverted speech’ (Brunelle, 2005b). This word game involves permutations of the onset and rime of a phrase to create a comical effect. For example,

/**naw puh**/ ‘to go to the dry rice field’ becomes /**nuh paw**/ ‘to set a trap’. The crucial fact here is that when two monosyllables have different registers, register follows the consonant rather than the rhyme:

(15)	**k̥aj klɔŋ**	**k̥ɔ̆ŋ klaj**
	club	rutting - penis
	club	*erect penis*
(16)	**p̥u klɔh**	**p̥ɔh klu**
	congee - cut, separate	fruit - testicle
	congee with small noodles	*testicle*

The fact that register always moves with the onset is good evidence that register is still a phonological property of onsets, even if it is realized on the rime. An alternative analysis is that register has become a form of lexical tone and that the rules of the word game always force this tone to follow the onset. Since the type of word game presented here is found throughout Southeast Asia (i.e. in Vietnamese) and usually allows the independent movement of the tones, I favor the first analysis. However, as this paper focuses on phonetic evidence, I leave this question open.

3.2. Acoustic experiment

In order to determine the relative importance of factors such as pitch, voice quality, vowel quality and vowel length in the production of Eastern Cham registers, an acoustic experiment was carried out. My results support Phú et al. (1992)’s findings: pitch and voice quality are the main acoustic correlates of register. However, the registers of many speakers also have distinct F1 (vowel height) and intensity (amplitude).

3.2.1. Methods

The acoustic experiment is based on the wordlist containing the vowels /a:/ and /ă/ that is described in Section 2.1.1. These vowels were selected, because they are more reliable for acoustic measurement of voice quality (spectral tilt). A low first formant can boost the amplitude of the lower harmonics on which spectral tilt measurements crucially depend. Since the vowel quality /a/ has a high first formant, it is better suited for these measurements.

The duration of the onsets, vowels, codas and rimes of all the target words were measured and corrected for speech rate. Because the overall duration of onset stops is difficult to measure (it is impossible to distinguish the closure from a possible pause between them and the previous word), only their voice onset time was measured. In order to filter out the effect of speech rate on duration measurements, a ratio was calculated by dividing the target segment by the duration of the syllable /khăn/ ‘word’ in the frame sentence. Whenever speakers produced /akhăn/ ‘word’ as the hypercorrect /khărn/ and /khăr/ or as /panoc/, which originally means ‘speech’ but is used for ‘word’ by some speakers, duration measurements were excluded from the results.

All other measurements were made at the beginning, 2/5, midpoint, 4/5 and endpoint of the onsets, vowels, codas and rimes of target words. The following acoustic measurements were made:

- Sonorant onsets:
    - Pitch (f0)
    - Amplitude (intensity)
- Vowels and rimes:
    - Pitch (f0)
    - Amplitude (intensity)
    - Vowel quality (F1 – vowel height - and F2 – vowel frontness/backness)
    - Voice quality (Spectral slope – high coefficients indicate breathiness)
        - H1-H2 (Amplitude of first harmonic – amplitude of second harmonic)
        - H1-A1 (Amplitude of first harmonic – amplitude of peak harmonic of first formant)
        - H1-A3 (Amplitude of first harmonic – amplitude of peak harmonic of third formant)

All f0 measurements had to be visually inspected for doubling and halving. Clear cases were corrected, but ambiguous values were excluded. Since the voice quality measurements were also dependant on pitch measurement (F0 values were used to determine the frequency of the first harmonic in the scripts), all voice quality measurements related to problematic f0 data were excluded.

In order to filter out the effect of codas, onsets and word shape on register, a statistical analysis was run on the acoustic data. For the purpose of the statistical analysis, all target words were divided into the following eight word types:

(17)	pa:C	pa:S	p^ha:C
	paC	paS	p^haC
	Sa:C	Sa:S	

C = stops, laryngeals or #
S = sonorants (except laryngeals)

The reason for breaking down the wordlist into categories is to avoid having an unnecessarily large array of variables to interpret and to avoid comparing word shapes with qualitative rather than quantitative differences. It is also important to note that some word types that are found in the wordlist are excluded because they have too few tokens to have any statistical significance (words with a pha:S shape, for example). When words were realized as disyllables, they were grouped according to their final, stressed syllable. A few trisyllabic realizations of the target words were excluded.

The statistical analysis chosen for this experiment is the General Linear Model (GLM). GLMs determine the effect of a set of categorical or gradient factors on a set of dependant variables. GLMs were run for each speaker and each of the 8 word types in order to determine if register is an appropriate predictor for the variation found in the acoustic measurements. All acoustic measurements listed above were used as dependant variables. The factors that were chosen as potential explanations for the variation are the following:

- Type of onset (consonant used as the onset)
- Type of coda (coda used as the onset)
- Type of syllabic template (monosyllabic with simple onset, monosyllabic with cluster onset, disyllabic)
- Register (Low or High)

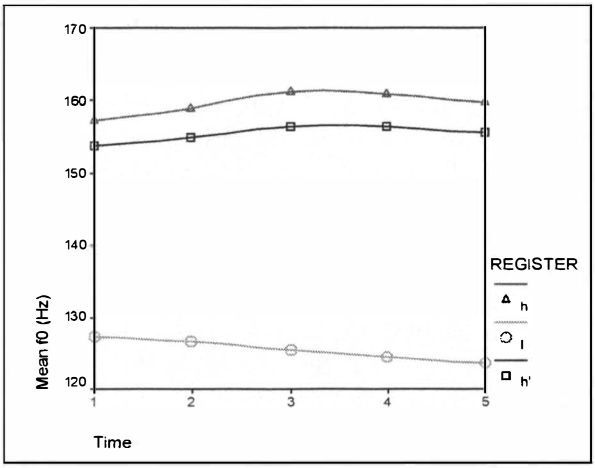
3.2.2. Results

Overall, the acoustic experiment confirms Phú et al. (1992)'s findings. Pitch and voice quality are the most important correlates of the register contrast. Further, the statistical analysis shows that the maximum contrast between registers is timed with the beginning of the rime, not the entire rime, which could either be because register is a phonologically feature of onset consonants or because a suprasegmental register is aligned with onsets.

A more detailed overview of the realization of registers is presented below. Since inter-speaker averages can be misleading, data from a representative man and a representative woman are plotted in charts to give the reader a general idea of the pitch, intensity, formants, voice quality and duration of the two registers. These charts are based on averages of the realization of register on both long and short /a/, in sesquisyllabic and monosyllabic words, and with a wide range of onsets (all possible labial onsets) and codas (stops, laryngeals and open syllables). Therefore, they are only meant to illustrate general tendencies. The statistical significance of the results and a brief overview of the phasing of register with the syllable are given in the text and at the end of the section.

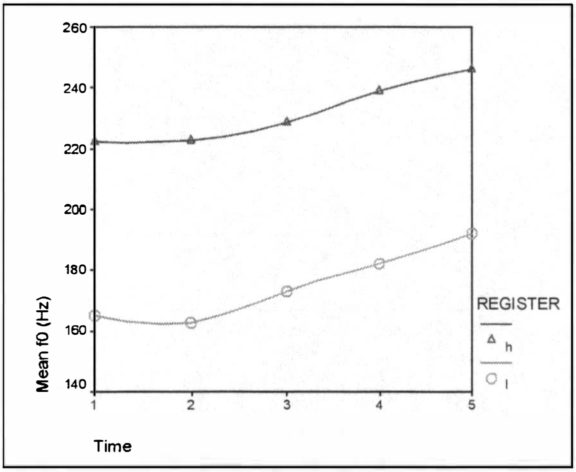
Pitch

Figure (18) shows the average vowel pitch of each register, for a male speaker born in 1977. This speaker is representative of other speakers. F0 is given at five different time points: the onset and the endpoint of the vowel, and three equidistant intermediate points. We see that the pitch of high register words (h) is much higher than the pitch of low register words (l) and that the few words starting with the implosive stop /ɓ/ and the preglottalized glide /ˀw/ pattern with the high register and are therefore labeled (h').



(18) Average f0 during the vowels of a male speaker born in 1977

The same general pattern is found in sonorant onsets, although for historical reasons, words with onset sonorants are never in the neutral register. The behavior of pitch in onset sonorants is illustrated in (19). In this chart, the large f0 difference between registers increases towards the end of the sonorant.



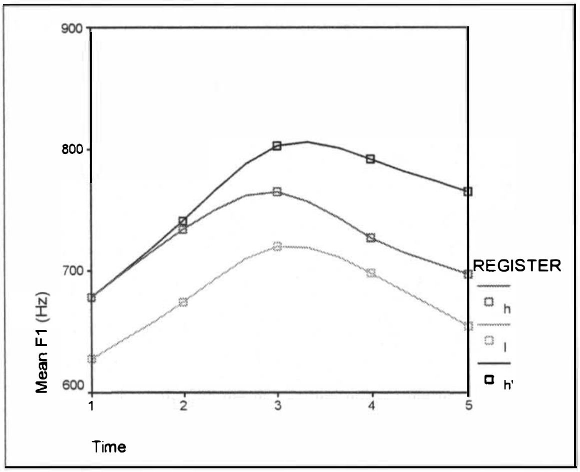
(19) Average f0 during the onset sonorants of a female speaker born in 1950

Amplitude

Amplitude differences vary considerably between speakers: the statistical analysis also shows that amplitude differences between registers are not significant for a majority of speakers. Even for the few speakers that have significant differences, the register that has the highest amplitude is sometimes the low register, sometimes the high one. Because of their limited significance, amplitude results are not plotted here. A more detailed discussion of amplitude results is found in Brunelle (2005a)

Vowel quality

If we now turn to the more interesting question of vowel quality, we find that the differences in vowel height and backness observed between the registers of many Mon-Khmer languages (Huffman, 1976, Miller, 1967, Watkins, 2002) are also present in Eastern Cham, but to a much lesser extent. Overall, as seen in (20), F1 is lower in the low than in the high register. This is expected because of the lengthening of the vocal tract due to the lowering of the larynx during the production of the low register, an articulatory mechanism that will be discussed in Section 3.2.2. The consequence of this lower first formant is that low register vowels should be perceived as more closed than high register vowels. However, the difference between registers is small and we will see in Section 3.3 that it is not used as a perceptual cue.

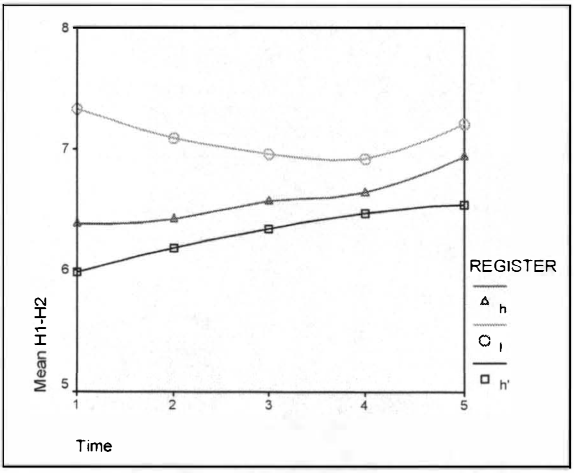


(22) Average F1 of the vowels of a male speaker born in 1977

F2 results are much less coherent than their F1 counterpart. It seems that larynx lowering and vocal tract lengthening do not affect the second formant as much as the first one. Since few speakers have significantly different F2 averages for the high and the low register, these results are not discussed here. A more detailed discussion is found in Brunelle (2005a).

Voice quality

Differences in voice quality are consistent with the results found in many Mon-Khmer languages. When we look at various measures of spectral tilt, the low register is breathier than the high register. The first voice quality measurement, H1-H2, behaves as expected. The low register has consistently higher H1-H2 values than the high and neutral registers, which is an indicator of breathiness. In (21), both subjects have a large difference between their two registers at the beginning of the vowel, but this difference is much narrower towards the end of the vowel, as all tokens become progressively breathier. The fact that the high register following /ɓ, ʔw/ is less breathy than the high register might be due to the glottalization that accompanies the onsets. Since the glottal folds are adducted during the production of these onsets, the vowels following them are produced with a more constricted glottis, which is the opposite of the abduction gesture that accompanies breathiness.

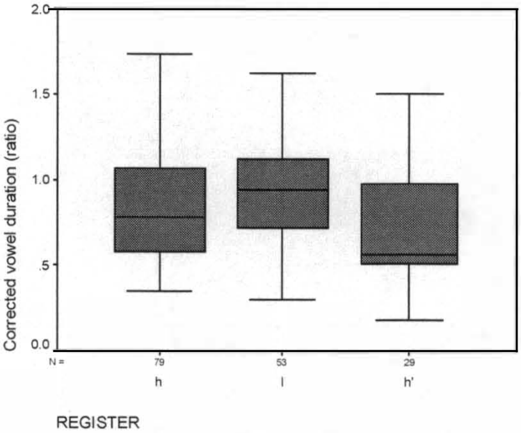


(21) Average H1-H2 on the vowel of a male speaker born in 1977

The same overall tendencies are found for the other acoustic measurements of spectral tilt, H1-A1 and H1-A3 (voice quality), which supports the view that the overall spectral slope, rather than the slope of a specific frequency range, is steeper in breathy vowels than in modal vowels. In fact, the acoustic measurement that seems to capture the voice quality contrast the most consistently is H1-A3, which measures the amplitude difference between the first harmonic and the peak harmonic of the third formant.

Duration

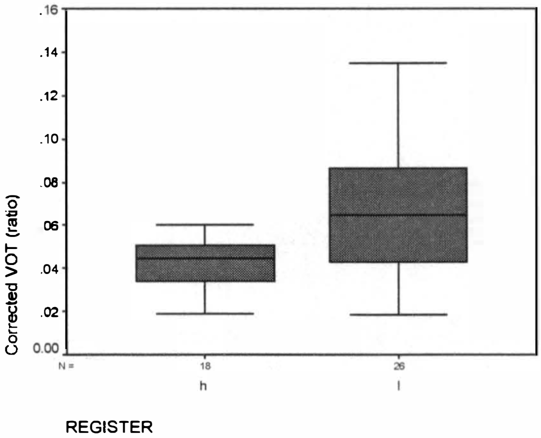
The last type of possible phonetic correlates of register is durational cues. I only present results from words with long /a/'s in this section, but results for words with short /a/'s are similar. There seems to be a tendency for neutral register vowels to be slightly shorter, as can be see in (22). The vowels of the high and low registers are not clearly different.



(22) Vowel duration ratio of the registers of a female speaker born in 1950

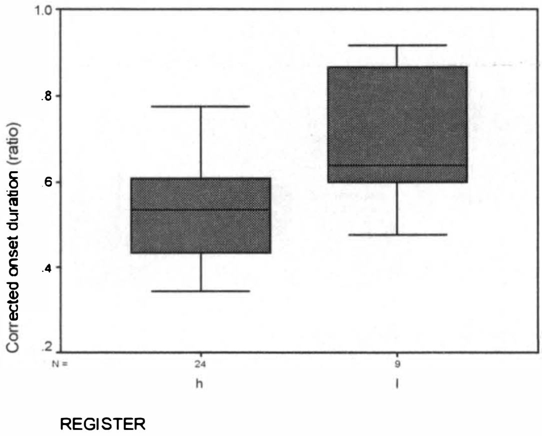
Onsets pattern more differently than vowels depending on the register to which they belong. This is especially true of onset stops, which, in Mon-Khmer languages and in

Javanese, are often slightly aspirated in the low register, but not in the high register (Adisasmito-Smith, 2004, Fagan, 1988, Ferlus, 1979, Hayward, 1993, Maddieson and Ladefoged, 1985). In (23), we see that the VOT of onset /p/ is longer in the low register.



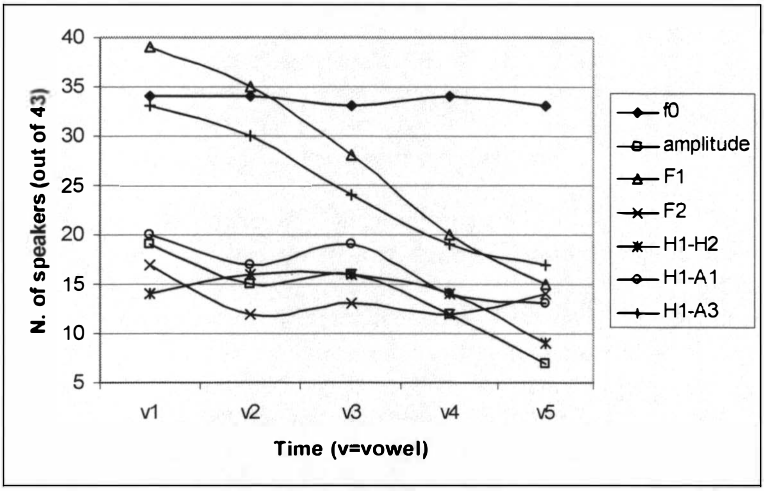
(23) VOT duration ratio of the registers of a male speaker born in 1977 after onset /p/

The durational differences in onset stops also hold for onset sonorants. For example, in (24), low register onset sonorants are longer than their high counterparts. It is therefore tempting to draw parallels with other languages and to claim that onset duration is a crucial feature of Eastern Cham register. Unfortunately, the statistical analysis shows that the durational differences are significant only for a minority of speakers. Therefore, although duration does play a certain role in the production of some speakers, it is not a robust correlate of register.



(24) Onset sonorant duration ratio of the registers of a female speaker born in 1950

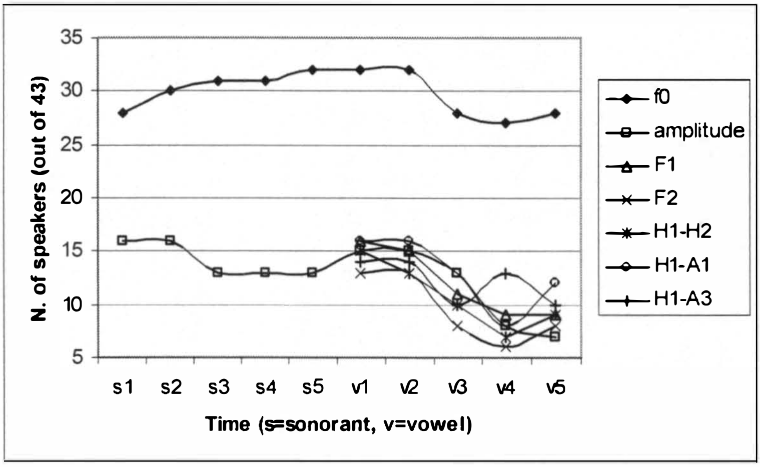
As mentioned during the presentation of these general tendencies, there is a fair amount of between-speaker variation in the acoustic realization of register. This variation is beyond the scope of this paper[11], but it is nonetheless important to emphasize that even if most phonetic correlates of register that were measured in the experiment are used in register production by at least some speakers, three correlates (pitch, voice quality and vowel height) are highly significant in the speech of almost all speakers. To illustrate that point, the results of the statistical analyses run on two representative word types have been plotted as figures. The phasing of register contrast in pV:C words is given in (25) while the phasing of register in SV:C words is given in (26).



(25) Number of speakers who have a significant difference between the High and Low registers for each phonetic correlate in pV:C words

These charts show us the number of speakers who have significantly different mean values for various phonetic correlates in the low and the high register. In the words starting with an onset stop, the pitch (f0), vowel height (F1) and voice quality (H1-H3) of high and low register tokens are distinct for the great majority of speakers at the beginning of the vowel. The F1 and H1-H3 values of two registers are much less distinct at the middle of the vowel and are generally not distinguishable at vowel endpoint. Pitch on the other hand remains clearly distinct throughout the vowel for most, but not all speakers.

[11] A full description of inter-speaker variation, both structural and sociolinguistic, can be found in Brunelle (2005a).



(26) Number of speakers who have a significant difference between the High and Low registers for each phonetic correlate in SV:C words

A similar pattern is found in monosyllables starting in a sonorant (26), but to a different degree. Onset sonorants are interesting because pitch (f0) and intensity (amplitude), which cannot surface on onset stops, can be realized on them. Once again, F0 (pitch) is very distinct at the end of the onset sonorant and the beginning of the vowel. All other phonetic correlates are distinct for one quarter to one third of the speakers at the beginning of the vowel, but then become less significant later on in the vowel.

3.2.2. Discussion

Although speakers mostly make use of pitch, there are different individual strategies to realize the register contrast. In addition to pitch, some speakers also realize the contrast through voice quality, intensity or F1, while others do not. This between-speaker variation in the realization of register was also found in Wa (Watkins, 2002). Should we then claim that each speaker has her own production strategy or should we rather assume that there is a general articulatory mechanism for register (and that each speaker grafts his own idiosyncrasies to it)? The second solution would obviously be more economical, but it can be considered only if we can propose a physiological model of register production. Since the study presented in this section is acoustic in nature, it is difficult to propose an articulatory model. However, I believe that by combining our knowledge of the diachronic formation of register and of the acoustic realization of articulatory gestures, such a model can be proposed.

It seems that the primary mechanism underlying register production in the early stages of registrogenesis is the vertical movement of the larynx. Originally the downwards movement of the larynx is a way of facilitating stop voicing by increasing the subglottal air pressure, but as voicing is neutralized, it is preserved along with its various acoustic consequences (Ferlus, 1979). A full exposition of the arguments in favor of laryngeal movement is beyond the scope of this paper. However, the physiological evidence that the downwards movement of the larynx is originally responsible for register production is the following:

1) The pitch difference between registers: because the spine is curved, the larynx rotates slightly when it moves down, which reduces the tenseness of the vocal folds and lowers pitch (Honda et al., 1999).

2) The voice quality difference between registers: the increased transglottal pressure due to laryngeal lowering causes perturbations in vocal fold vibrations that are perceived as breathiness.

3) The vowel quality difference between registers: the length of the supraglottal tract is lengthened by lowering the larynx, which raises formant values. The expansion of the pharynx that possibly contributes to the original voicing contrast by lowering supraglottal pressure can also enhance vowel quality differences by pushing the root of the tongue forward, and as a result, the tongue tip as well.

However, at a later diachronic stage, an acoustic correlate can be emphasized by adding active control to a specific articulator (Ferlus, 1979). In the present case, the pitch difference between registers can be enhanced by tensing or laxing the vocal folds through direct control of intrinsic laryngeal muscles. Similarly, vowel quality differences can be enhanced by directly controlling tongue muscles, in addition to the original register movement. As a result, some of the enhancement cues can gradually gain importance and the register distinction could eventually be reanalyzed and maintained through one or a few of these originally ancillary phonetic cues even after the neutralization of vertical laryngeal movement.

The fact that all the phonetic cues that were measured in the acoustic analysis are significant for at least some speakers or word types suggest that vertical laryngeal movement is still present. If it had been neutralized, we would expect speakers to retain only one or two features and to lose all other features. That said, it is clear that pitch plays a more central role in the register distinction of Eastern Cham than in most Southeast Asian register systems. This is good evidence that the pitch contrast is enhanced by speakers, but it is also possible that some speakers enhance other phonetic cues as well. Although articulatory evidence is lacking, the articulatory realization of register should at least include vertical movement of the larynx and some laxing/tensing of the vocal folds.

The active role of vertical laryngeal movement in register realization suggests that the register system of Eastern Cham is relatively conservative in that it preserves the articulatory mechanism that has given rise to its register system, despite its enhancement through direct vocal fold control. In that respect, Eastern Cham is different from other languages, where only enhanced features have been preserved and were the original movement of the larynx has been lost (Huffman, 1976 for Khmer, Mundhenk and Goschnick, 1977 for Haroi).

3.3. Perceptual experiment

A perceptual experiment was carried out to determine which phonetic attributes are the most important perceptual cues in register discrimination. Not surprisingly, the main perceptual cues of register are pitch and voice quality, which is congruent with the results of the acoustic study.

3.3.1 Methodology

Five minimal pairs were recorded by Phú Văn Hẳn, a male native speaker of Eastern Cham born in 1963. The first pair (la/l̥a) was chosen because its two stimuli begin in a sonorant.

The next two pairs start with /p/, but while the second pair has an open vowel, the third one has a laryngeal coda /h/. This contrast was chosen to test the effect of final /h/ on the perception of breathiness. Finally, the last two pairs both have /t/ onsets and /ʔ/ codas, but their vowels contrast in length. It was expected that durational cues would not be as relevant to register discrimination for these two pairs. Note that these target words were recorded in their monosyllabic colloquial form.

(27) Stimuli:

/la/	‘snake’	/l̥a/	‘stupid’
/pa/	‘to cross’	/p̥a/	‘to carry’
/pah/	‘to hit with the hand’	/p̥ah/	‘to dust’
/tăʔ/	‘to behead”	/t̥ăʔ/	‘to tidy up’
/taʔ/	‘bean’	/t̥aʔ/	‘pumpkin’

These minimal pairs were uttered in the following carrier sentence, recorded in its colloquial form:

(28) Frame sentence:

/l̥ăʔ poc khăn _____ jwa krɨ khăn ni/
I read word _____ because like word *dem*
‘I read the word _____ because I like it.’

The pitch, voice quality, formant frequencies and durations of the stimuli are given in Appendix 1. Values are given for onsets and offsets although the stimuli were also measured at three other equidistant points.

In order to have a more precise idea of the factors that are playing role in the perception of register, some phonetic cues were modified and some stimuli were resynthesized. Ideally, all the phonetic correlates that have been shown to play a role in the register contrast would have been varied to measure their exact effect on perception. However, the number of resulting stimuli would have been too high to run the perceptual experiment in one session per subject. Further, the original stimuli were modified using Praat 4.2, which limits the types of resynthesis that can be done. My attempts at modifying voice quality, for example, resulted in stimuli that did not sound natural to my listeners. Therefore, only two parameters were modified: pitch and duration. Pitch clearly plays a role in the register contrast. As for duration, it was included because a pilot study led me to believe that it might be relevant to the production and perception of register, a hypothesis that later turned out to be less well-supported than expected, as seen in the previous section. In any case, the stimuli where then modified in three ways:

1) Pitch was *modelized*, i.e., it was smoothed by rounding up the f0 values at vowel onset and offset to the closest multiple of 10 and by interpolating a straight pitch curve between these two points. Words with open vowels were treated differently, because a straight interpolation sounded too artificial to the subjects in a preliminary experiment. They were assigned a third pitch target at their mid value

in order to generate a flat pitch curve on the first half of the vowel followed by a falling curve on the second half.
2) The pitch curve was replaced by the pitch curve of the modelized token of the other member of the pair. For example, the pitch of the *p̥a* was replaced by the modelized pitch of *pa*. These tokens are called *inversed* tokens.
3) Pitch was neutralized by replacing it with an average of the modelized pitch of the two members of the pair. For example, the *neutralized* token of *p̥a* has a pitch value that has been obtained by averaging out the modelized pitch targets of *p̥a* and *pa*. The goal of these pitch manipulations is to determine the role of pitch by measuring the changes in register perception when pitch is modified while keeping other factors constant.

The duration of these three types of tokens was then modified. The onset of high register tokens was lengthened by a factor of 1.5 without changing the duration of the vowel. The duration of the vowel was also lengthened by a factor of 1.5, without modifying the duration of the onset. The opposite was done for low register tokens: both their onsets and vowels were shortened by a factor of 1.5. The aim of these duration manipulations was to weigh the role of duration in register perception and its interaction with pitch, although this effect turned out to be less important than expected, as mentioned above.

Voice quality was kept constant. The initial pairs of stimuli that were chosen for the experiment have markedly different voice qualities at their vowel onsets. The low register member of each pair has a higher H1-H2 value, i.e. a breathier voice at vowel onset. Since the acoustic study showed that the two registers do not always have a robust voice quality contrast at vowel offset, this criterion was not retained in the selection of the natural stimuli. Another factor that was left aside, this time for technical reasons, is the frequency of the first two formants. Attempts at manipulating formants made the stimuli sound artificial to the subjects. By and large, manipulation of one phonetic dimension had little effect on others. Changes in pitch obviously resulted in minor changes in formant frequencies as harmonics frequency was modified. Pitch changes also affected voice quality to some extant, as they modified the frequency of spectral peaks. However, durational manipulation merely stretched or compressed the stimuli without effects on other phonetic dimensions.

In total there were 100 stimuli which were played to the 30 subjects in three separate sub-experiments. The nine female and 21 male subjects were all native speakers of Eastern Cham living in Hồ Chí Minh City. Twenty-four of them were originally from Ninh Thuận province and six from Bình Thuận and all of them were college-educated. This subject sample is less diverse in terms of age and socioeconomic background than the sample used for the production study, but at the same time, it is more diverse in terms of dialectal variation. The choice of subjects was dictated by the fact that I needed subjects who could use a mouse and were not intimidated by computers, and that such subjects were difficult to find in Phan Rang. Further, I was allowed to work freely with Cham speakers on the premises of the University of Social Sciences of Hồ Chí Minh City, whereas every work session in Phan Rang had to be approved by provincial and local authorities, making the enrollment of a relatively large number of subjects for short sessions very difficult.

The experiment was designed as a Praat perceptual setup and was administered to the subjects on a laptop computer. The first sub-experiment included the stimuli based on *la/l̥a*, the second comprised the stimuli based on *pa/p̥a/pah/p̥ah* and the third was made up of the stimuli based on *tăʔ/t̥ăʔ/taʔ/t̥aʔ*. Each stimulus was played three times and the stimuli were played in a random order. Subjects listened to the stimuli through headphones and then selected the word they had heard by clicking a box containing a latin-based transcription and Vietnamese glosses for all possible answers (the two or four possible lexical items used for each sub-experiment). The next token was then played automatically. In the first experiment, there were only two possible answers (*la* and *l̥a*). The second and third sub-experiments had four possible answers each. Subjects did not have the option of not making a choice and were instructed to choose the best possible answer, even if not fully satisfactory. Subjects were allowed to take a short break after every 40 stimuli.

3.3.2 Results

The results of the acoustic analysis suggest that the most salient cues for perception should be pitch and voice quality. F1 could also play a minor role, but duration and amplitude should not have much of an effect on perception. The results of the perceptual analysis largely confirm this. In this section, I present a statistical analysis of the responses of all subjects.

Before starting, a short caveat on terminology is in order: I use the term *correct identification* when subjects identify the register of a stimulus as the register of the natural stimulus from which it was resynthesized. For example, a high-register stimulus that has been resynthesized with a low pitch is *correctly identified* if it is perceived as a high-register token and *misidentified* if it is perceived as a low-register stimulus.

The categorical analysis presented in the previous section cannot account for the effect of small variations in the perception of phonetic cues and determine the relative importance of each phonetic factor in register perception. Therefore, binary logistic regressions were run on the data from each sub-experiment. A regression model takes all the variation in a data set and finds the factor that explains the largest amount of variation in it. It then removes all the variation that can be accounted through this first predictor and finds the factor that accounts for the largest proportion of the remaining variation. This operation, called a stepwise regression, is repeated as many times as required. In the present experiment, the stepwise regression was interrupted when the percentage of cases correctly predicted by the model reached its peak. The factors that were included in the regression model are the normalized duration of the onset and vowel, the f0 and amplitude of the onset, and the f0, amplitude, formants (F1 and F2) and voice quality (H1-H2, H1-A1, H1-A3) of the vowel.

Results for the first minimal pair, /la~l̥a/, are given in (29). Pitch at the midpoint of the vowel (P3), accounts for 60.1% of the variation in the data and, by itself, correctly predicts 84.2% of responses. The next best predictor of the responses given by the subjects is F2 at 4/5 of the vowel. This is unexpected as F2 is not a robust phonetic correlate of register, but note that this factor only explains an additional 2.9% of the variation and 0.4% of the responses, which is at best marginal. Along the same lines, the third significant predictor is F2 at the endpoint of the vowel, which accounts for an extra 0.7% of the variation and 1.2% of responses. Overall, we can conclude that pitch is by far the most important perceptual cue for open syllable with an onset sonorant and that the pitch, intensity and duration of the onset sonorant do not play a role in perception.

(29) la/l̥a

	Cue and time point (P1-P5)	Sig.	Nagelkerke R^2	Percentage of cases correctly predicted by the model
Step 1	f0 P3 vowel	.000	.601	84.2
Step 2	f0 P3 vowel F2 P4 vowel	.000 .000	.630	84.6
Step 3	f0 P3 vowel F2 P4 vowel F2 P5 vowel	.000 .000 .000	.637	85.8

The next pair of words, /pa~p̥a/, also consists of open syllable words, but with onset stops instead of sonorants (30). The same factors have been included in the model, except the pitch and the intensity of the sonorant, which are not relevant in words with onset stop. Moreover, rather than including the normalized duration of the whole onset, only normalized VOT was used. Once again, f0 (pitch) is the most important perceptual cue, but at vowel onset rather than midpoint. It accounts for 54.2% of the variation and 84.8% of responses. At step 2, voice quality at vowel midpoint (as measured by H1-H2) accounts for an additional 5.1% of the variation, but no further responses. Finally, another voice quality measurement, H1-A3 at vowel midpoint, captures another 1.1% of the variation and 0.7% of cases. Once again, f0 alone accounts for a large majority of responses.

(30) pa/p̥a

	Cue and time point (P1-P5)	Sig.	Nagelkerke R^2	Percentage of cases correctly predicted by the model
Step 1	f0 P1 vowel	.000	.542	84.8
Step 2	f0 P1 vowel H1H2 P3 vowel	.000 .000	.593	84.8
Step 3	f0 P1 vowel H1H2 P3 vowel H1A3 P3 vowel	.000 .000 .000	.604	85.5

By contrast, pitch does not play a role in the perception of syllables closed by laryngeals. Results for the minimal pair /pah/p̥ah/ are given in (31). Vowel quality at 2/5 of the vowel, as measured by H1-H2, accounts for 54.5% of the variation and 82.8% of responses. No other perceptual cue correctly predicts an additional proportion of responses.

(31) pah/p̥ah

	Cue and time point (P1-P5)	Sig.	Nagelkerke R^2	Percentage of cases correctly predicted by the model
Step 1	H1H2 P2 vowel	.000	.545	82.8

A similar situation is found for the minimal pair /tăʔ/t̥ăʔ/ shown in (32). Voice quality at vowel midpoint, but this time measured with H1-A3, captures 77.7% of the variation and 93.3% of responses. Again, no other factor accounts for a higher proportion of responses.

(32) tăʔ/t̥ăʔ

	Cue and time point (P1-P5)	Sig.	Nagelkerke R^2	Percentage of cases correctly predicted by the model
Step 1	H1A3 P3 vowel	.000	.777	93.3

Voice quality at vowel midpoint is also the best predictor of responses for the minimal pair /taʔ/t̥aʔ/ (33). This time, however, H1-A1 is the most reliable acoustic cue. It accounts for 57.8% of the variation and 79.7% of responses. Other factors do not increase the proportion of correct predictions.

(33) taʔ/t̥aʔ

	Cue and time point (P1-P5)	Sig.	Nagelkerke R^2	Percentage of cases correctly predicted by the model
Step 1	H1A1 P3 vowel	.000	.578	79.7

In short, while pitch is by far the most important perceptual cue for open syllables, it is not used as a cue when syllables are closed by a laryngeal. Register perception in the syllables closed by a laryngeal depends exclusively on voice quality, although there is variation as to which phonetic correlate of voice quality is chosen.

3.3.3 Discussion

Two acoustic cues are used for perception: voice quality and pitch. Voice quality seems to be the dominant cue in closed syllables (pah, p̥ah, tăʔ, t̥ăʔ, taʔ, t̥aʔ). By contrast, open syllable tokens (la, pa) are distinguished mostly through pitch. Duration does not have a clear effect. Manipulation of the duration of vowels and onsets does not seem to be sufficient to cause misidentification.

Overall speakers rely on the first half of the vowel for perception. All perceptual cues that are used by speakers are timed with the first three measurement points, except F2 in (29), which is timed with the end of the vowel, but accounts for a marginal proportion of correct identification. The timing of relevant perceptual cues with the beginning of the vowel agrees with the results presented in (25) and (26), where we have seen that the acoustic contrast between registers is stronger at the beginning of the vowel.

The fact that different word types are associated with different perceptual cues can be explained by the effect of codas on pitch. We will see in the next section that the laryngeal codas /h/ and /ʔ/ have an effect on the pitch of the preceding vowel. This effect is mostly felt towards the end of the vowel, but it can blur the pitch contrast between the two registers. In short, pitch, the most clearly contrasting phonetic correlate of register, is used as a perceptual cue in open syllables, and possibly in syllables closed by codas that have little effect on pitch (sonorants). However, whenever a coda affects pitch and makes it less reliable for distinguishing register, listeners fall back on the second most salient cue, voice quality.

4. Conclusions

Is Eastern Cham a tone language? Has its register system evolved into a tone system through the interaction of register allophony and coda weakening and deletion? Contrary to what has been claimed by other researchers, the data presented here does not support the hypothesis that codas are dropped or that their pattern of contrast is being modified. There is no evidence that the final glottal stop has become a tonal element either. These conclusions cast serious doubt on the claim that Eastern Cham is undergoing a full-fledged tonogenesis. However, there still exists a possibility that the two registers of Eastern Cham have become two tones, two suprasegmental elements that are distinguished mostly through pitch, but can also have other correlates. In fact, the crucial role of pitch in the register system, along with the monosyllabic character of the colloquial language and the loosening of cooccurrence restrictions between onsets and registers all seem to suggest that Eastern Cham has become a two-tone language. However, there is also synchronic evidence from a word game that register is still a property of onset consonants. This question is addressed in more detail elsewhere (Brunelle, 2005a, Brunelle, 2005b), but the basic facts suggest that Eastern Cham does not have a very developed tone system, if it has tones at all.

Results from the acoustic and perceptual experiments confirm that the register distinction of Eastern Cham is phonetically realized and perceived through pitch and voice quality, the low register having a lower pitch and a breathier phonation than the high register. However, contra Phú et al. (1992), our results suggest that F1 is higher for the low register than the high register. This higher F1 is due to the lower position of the larynx during the production of the low register, a feature that can be traced back to the original voicing contrast that gave rise to register. Another conservative feature of Eastern Cham register is the longer duration of low register onsets in the speech of some speakers, which is reminiscent of incipient register systems elsewhere in Southeast Asia.

The diachronic implications of these findings are beyond the scope of this paper, but models of contact-induced registrogenesis and tonogenesis should be able to integrate these phonetic facts and should be grounded in a sociophonetic investigation of the variation found in the speech community.

References and sources

Adisasmito-Smith, Niken. 2004. Phonetic Influences of Javanese on Indonesian, Linguistics, Cornell University.

Aymonier, Étienne François. 1889. *Grammaire de la langue chame*. Saigon: Imprimerie coloniale.

Blood, David Livingstone. 1967. Phonological Units in Cham. *Anthropological Linguistics* 9:15-32.

Blood, David Livingstone. 1977. A romanization of the Cham language in relation to the Cham script. Dallas: Summer Institute of Linguistics.

Blood, Doris Walker. 1980. The Script as a Cohesive Factor in Cham Society. In *Notes from Indochina on Ethnic Minority Cultures*, ed. Marilyn; Thomas Gregerson, Dorothy, 35-44.

Brunelle, Marc. 2003. Tone Coarticulation in Northern Vietnamese. *Proceedings of the 15th International Conference of Phonetic Sciences.*

Brunelle, Marc. 2005a. Register in Eastern Cham: Phonological, Phonetic and Sociolinguistic approaches, Linguistics, Cornell: Ph.D.

Brunelle, Marc. 2005b. Register and tone in Eastern Cham: Evidence from a word game. *Mon-Khmer Studies* 35.

Brunelle, Marc. *to appear*. Monosyllabicization in Eastern Cham. *Proceedings of the 14th meeting of Southeast Asian Linguistics Society*.

Bùi, Khánh Thế. 1996. *Ngữ Pháp Tiếng Chăm*. Hà Nội: Nhà Xuất Bản Giáo Dục.

Edmondson, Jerold A., and Gregerson, Kenneth J. 1993. Western Cham as a Register Language. In *Tonality in Austonesian Languages*, ed. Kenneth J. Gregerson, 61-74. Honolulu: U of Hawaii Press.

Fagan, Joel L. 1988. Javanese Intervocalic Stop Phonemes. *Studies in Austronesian Linguistics* 76:173-202.

Ferlus, Michel. 1979. Formation des Registres et Mutations Consonantiques dans les Langues Mon-Khmer. *Mon Khmer Studies* V3:1-76.

Han, Mieko. 1969. *Vietnamese Tones*. Los Angeles: University of Southern California Acoustic Phonetics Research Laboratory.

Haudricourt, André. 1972. Bipartition et Tripartition des Systèmes de Tons dans quelques Langues d'Extrême-Orient. In *Problèmes de Phonologie Diachronique*. Paris: Société pour l'Etude des Langues Africaines.

Hayward, Katrina. 1993. /p/ vs. /b/ in Javanese: Some Preliminary Data. *Working Papers in Linguistics and Phonetics*:1-33.

Headley, Robert K. 1991. The Phonology of Kompong Thom Cham. In *Austroasiatic languages essays in honour of H. L. Shorto*, ed. Jeremy Davidson, 105-121.

Henderson, Eugenie. 1952. The main features of Cambodian prononciation. *Bulletin of the School of Oriental and African Studies* 14:453-476.

Hoàng, Cao Cương. 1986. Suy nghĩ thêm về thanh điệu tiếng Việt. *Ngôn ngữ* 69:19-38.

Hoàng, Thị Châu. 1987. Hệ thống thanh điệu tiếng Chàm và các kí hiệu. *Ngôn Ngữ*:31-35.

Hoàng, Thị Châu. 1989. *Tiếng Việt trên các Miền Đất Nước*. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.

Honda, Kiyoshi, Hirai, Hiroyuki, Masaki, Shinobu, and Shimada, Yasuhiro. 1999. Role of Vertical Larynx Movement and Cervical Lordosis in F0 Control. *Language and Speech* 42:401-411.

Huffman, Franklin. 1976. The register problem in fifteen Mon-Khmer languages. *Oceanic Linguistics special publication* Austroasiatic Studies, part 1:575-589.

Maddieson, Ian, and Ladefoged, Peter. 1985. "Tense" and "lax" in four minority languages of China. *Journal of Phonetics* 13:433-454.

Matisoff, James. 1973. Tonogenesis in Southeast Asia. In *Consonant Types and Tone*, ed. Larry Hyman, 71-96. Los Angeles: USC.

Michaud, Alexis. 2005. Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese. *Phonetica* To be published:1-28.

Miller, J.D. 1967. An acoustic study of Brou vowels. *Phonetica* 17:149-177.

Moussay, Gérard. 1971. *Dictionnaire cam-vietnamien-français*. Phan Rang,: Trung-tâm Văn hoá Chăm.

Mundhenk, Alice Tegenfeldt , and Goschnick, Hella. 1977. Haroi Phonemes. In *Papers in Southeast Asian Linguistics no. 4*, eds. David Thomas, Ernest Lee and Đăng Liểm Nguyễn, 1-15. Canberra: Australian National University.

Nguyễn, Văn Lợi, and Edmondson, Jerold. 1997. Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies* 28:1-18.

Phạm, Hoa T. (Andrea). 2001. Vietnamese Tone: Tone is not Pitch, PhD thesis, Department of Linguistics, University of Toronto.

Phan, Xuân Biên, Phan, An , and Phan, Văn Dốp. 1991. *Văn Hoá Chăm*. Thành Phố Hô Chí Minh: Nhà xuất bản Khoa học-xã hi.

Phú, Văn Hẳn, Edmondson, Jerold, and Gregerson, Kenneth. 1992. Eastern Cham as a Tone Language. *Mon Khmer Studies* 20:31-43.

Po, Dharma. 1987. *La Panduranga (Campa) 1802-1835*. Paris: Ecole française d'Extrême-Orient.

Thurgood, Graham. 1993. Phan Rang Cham and Utsat: Tonogenetic Themes and Variants. In *Tonality in Austronesian Languages*, eds. Jerold A. Edmondson and Kenneth J. Gregerson, 91-106. Honolulu: U of Hawaii Press.

Thurgood, Graham. 1996. Language Contact and the Directionality of Internal Drift: The Development of Tones and Registers in Chamic [Mar]. *Language* 72:1-31.

Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects : Two Thousand Years of Language Contact and Change*: Oceanic linguistics special publication ; no. 28. Honolulu: University of Hawai'i Press.

Thurgood, Graham. 2002a. Learnability and direction of convergence in Cham : the effects of long-term contact on linguistic structures. In *Proceedings of the Eleventh Annual Conference on Linguistics*. Fresno.

Thurgood, Graham. 2002b. Crawfurd's 1822 Malay of Champa. Ms.

Vũ, Thanh Phương. 1982. Phonetic Properties of Vietnamese Tones across dialects. In *Papers in Southeast Asian Linguistics*, ed. David Bradley, 55-75. Sydney: Australian National University.

Watkins, Justin. 2002. *The phonetics of Wa: experimental phonetics, phonology, orthography and sociolinguistics*. Canberra: Australian National University.

Appendix I

Phonetic correlates of register in the unmodified stimuli used for the perceptual experiment

	f0 onset (Hz)	f0 offset (Hz)	H1-H2 onset	H1-H2 offset	F1 onset (Hz)	F1 offset (Hz)	F2 onset (Hz)	F2 offset (Hz)	Onset /VOT duration (msec)	Vowel duration (msec)
l̥a	158	135	8,8	8,2	553	793	1696	1502	192	295
la	192	162	6,5	8,5	683	679	1568	1603	86	262
p̥a	154	126	7,8	8,5	590	739	1241	1429	20	334
pa	183	156	5,9	9,7	685	742	1217	1456	13	278
p̥ah	171	171	8,4	10,9	562	721	1255	1410	21	104
pah	208	203	5,1	9,7	638	729	1136	1549	13	114
t̥ăʔ	172	183	10,0	6,2	661	792	1686	1492	12	122
tăʔ	213	214	5,2	6,7	711	802	1274	1489	9	102
t̥aʔ	159	168	11	6,3	696	807	1636	1470	13	232
taʔ	206	197	5,3	8,7	711	802	1274	1489	10	198

2 *The effects of intimate multidirectional linguistic contact in Chamic*[1]

Anthony P. Grant

1. Introduction

Many aspects of the astounding effects of continued and profound linguistic contact which may occur over a millennium or two can be seen from a study of the Chamic languages of South East Asia. This is a group of languages whose role in the 'mixed language' debate, once considerable, has receded in recent decades (though it had played a lively part in this discussion earlier in the 20th century, as the treatment in Haudricourt 1966 indicates).

The similarities between Malay and any Chamic language (for example, a fairly conservative one such as Western Cham) are not only mostly to be found on the surface but also are few and far between. Typologically Chamic languages look much more like Mon-Khmer languages than they resemble modern Malay, although they still look typologically more like Malay than like Ilokano or Tagalog. No Malayic language has diverged from older forms of Malay as much as Cham (or even more so Rade and Tsat) has changed from the Proto-(Malayo-)Chamic norm. Furthermore Proto-Chamic was in turn very similar to Proto-Malayic, to the extent that some scholars have used (admittedly modem standardised) Malay forms as substitutes for Proto-Chamic forms, without having to stretch the facts of linguistic history too far.

The explanation for this divergence of modern Chamic languages from the Proto-Malayo-Chamic norm, as Thurgood (1999: 251-259) rightly points out, lies in the intense amounts of linguistic contact which Chamic has undergone from surrounding languages. Many of these languages were once spoken by groups who were technologically less sophisticated and politically subservient to the Chams in the period of the Cham Empire. They used Cham as a lingua franca and some of them abandoned their original (Mon-Khmer) languages in favour of Cham. Tsat and Standard Malay stand at opposite poles of a diachronic continuum of change whose major controlling factor is the myriad consequences of language contact or contact-induced anguage change. If Rade is included and compared with the rather consercvative Western Cham, then the attested parameters of contact-induced and partially independent change, even within Indochinese Chamic, are even wider.

[1] I would like to thank Sander Adelaar, Philip Baker, Stéphane Goyette, Robert K. Headley, James Matisoff, Russell Murray, Paz Buenaventura Naylor, Peter Patrick, Paul Sidwell, Sally Thomason, and especially Graham Thurgood and Bob Blust for inspiration, advice, encouragement and assistance with an earlier version of this paper. None of them is to be held responsible for any conclusions or errors to be found in this paper, nor is it to be assumed that any of them agree with all the views expressed here.

Grant, A.P. "The effects of intimate multidirectional linguistic contact in Chamic". In Grant, A. and Sidwell, P. editors, *Chamic and Beyond: Studies in mainland Austronesian languages.*
PL-569:37-104. Pacific Linguistics, The Australian National University, 2005. DOI:*10.15144/PL-569.37*

2. Language contact, Acehnese and Chamic: a conspectus of changes.

Graham Thurgood's descriptive and historical work (especially Thurgood 1999) has been definitive in showing to the wider world that Acehnese groups with the Chamic languages in an especially close non-trivial genetic relationship, although he was far from being the first author to make such a connection (as Thurgood himself points out, Niemann 1891 composed the first article on this). Thurgood's diachronic position, which he supports with a large amount of convincing evidence, is that Acehnese is related at a coordinate level with all the Chamic languages, rather than being especially closely related with any one of them, though he suggests that the ancestors of the speakers of Acehnese left from the northernmost part of the chain of Chamic dialects as the result of incursions from the Vietnamese from the north, and that tye went south. He suggests implicitly rather than explicitly that the time-depth of dispersal and division within the Chamic languages (this figure probably including Acehnese) is less than 2000 years, an assumption which is broadly borne out by historical evidence. (Before then the Cham-speaking communities constituted a dialect continuum which stretched along part of the southern Vietnamese coast.).

The similarities between Acehnese and Chamic languages are partly due to their shared history (much of which has been obscured by subsequent contact-induced changes on both sides but from different sources) and partly to their shared context of contact. The differences between them are made up of both retentions on one or another side and innovations on both sides.

There are certain features (such as the productive use of some Malayic affixes) which Acehnese has inherited from Proto-Malayo-Chamic and then from Proto-Malayo-Polynesian and which, through being in renewed and continual contact with Malay lects, it has been able to retain while the other Chamic languages have lost many of these. To this extent, Acehnese is conservative and the mainland Chamic languages and Tsat are innovative. And there are separate clusters of innovations on all levels – phonetic, phonological, morphosyntactic and lexical - which have been caused by prolonged contact between Acehnese and the more dominant Malay on the one hand, and between Chamic and Mon-Khmer languages on the other. Furthermore there are innovations of various kinds, lexical, structural and other, which are found in most or all Chamic languages (including sometimes Acehnese), but for which etymological sources have yet to be found; this gap in our knowledge is especially true of the numerous lexical innovations which are exclusive to Chamic.

These languages have all been in touch with various branches of Mon-Khmer, especially Eastern Mon-Khmer, and more especially the more northerly and central branches of the Bahnaric family. And as our knowledge of the number and content of Mon-Khmer languages has increased, we are able more and more accurately to pinpoint the sources of such influence. The earliest, longest-lasting, most basic and deepest contact has been with Bahnaric languages (including the subgroup represented by Mnong, and principally by North and Central Bahnaric languages or by a protolanguage which is ancestral to one or both these groups), and this is true for all of them. This is especially significant from a historical point of view in the case of Acehnese, which according to Thurgood (1999) additionally contains Katuic elements that are not recorded in other Chamic languages (although Thurgood does not identify these), as well as containing other Mon-Khmer elements which are pan-Chamic.

Subsequently at nation-state level there has been contact with Khmer and/or Vietnamese, which have served as loan sources for the two modern dialects of Cham, and there has been contact with Vietnamese for Jarai, Rade, Roglai and Haroi among other languages. (The speakers of Rade and Jarai who live in Cambodia are also in touch with Khmer, of course, as are those who live in Vietnam's Mekong Delta, which is also home to a sizeable Khmer community. Speakers of Western Cham in Vietnam are bilingual in the regionally dominant Khmer, with Vietnamese as a third language.) All Chamic languages bar Acehnese and Tsat are still in touch with various Mon-Khmer languages, which act as their chief sources of new lexicon. But the significant Mon-Khmer languages with which they are in touch nowadays are not the same ones, spoken by 'Montagnards' (such as Bahnar or its immediate ancestor), which exerted the primary influence upon Proto-Chamic. Instead they are the prestigious Khmer and Vietnamese languages, especially the latter.

Tsat is exceptional in respect of its Mon-Khmer heritage, as it lost contact with members of this family this long ago, and therefore its Mon-Khmer elements go back to a period of Tsat linguistic unity with other Chamic languages. It has been profoundly influenced, not by Khmer or Vietnamese, but by Hainanese Chinese, a Southern Min variety, possibly also by the pre-Chinese Hainanese Hlai languages, which are monosyllabic tonal languages of Tai-Kadai affinity. Most if not all Chamic languages which are spoken by Islamic populations have been influenced by the incorporation of many culturally-oriented lexical items from Arabic relating to Islam (these words are apparently not transferred directly but more probably through an intermediate language such as Malay).

After splitting from Cham, Haroi has also been strongly influenced by the Bahnaric language Hrê in addition to undergoing very strong lexical and other influence from Bahnar proper; in fact Haroi speakers were formerly known as the Bahnar Chams (Thurgood 1996: 14).

Acehnese shows signs of lexical influence from Bahnaric languages, Katuic and probably the Mon-Khmer Aslian languages of Malaya (such as Semang), in addition to receiving strong subsequent influence from Malay (and thereby indirectly from Javanese, Dutch, Portuguese, Arabic and Sanskrit). The Bahnaric elements are also found in other Chamic languages. The Aslian component (for Malay *orang asli* 'people (of) origin' is the name for pre-Malay native inhabitants of the Malay Peninsula; *asli* is not an autonym but derives from Arabic) is of course absent from the other Chamic languages, as indeed are Malay elements (except those few that have been mediated through Vietnamese or Khmer, or those which have been acquired by Muslim Chams as part of an Islamic education).

We should note that the distinctive Mon-Khmer stratum in the Malay vocabulary, which derives from Aslian languages and which manifests itself in a number of fairly high-frequency contentive nouns mostly relating to fauna, is at the moment a recognised but still seriously under-examined subject. Aslian borrowings into Acehnese, if there are any, are apparently independent of those found in Malay, though there may be some commonalities as a secondary result of Acehnese borrowing from Malay. The Aslian impact on Malay is purely lexical in nature and, though it contains a greater number of items of core vocabulary than one might have guessed, it is nowhere near as deep as the impact of Mon-Khmer languages are upon Acehnese. Naturally there are no items of distinctly Aslian origin in the Chamic languages of Indochina and Hainan.

It is the case that there may be some Mon-Khmer components in certain languages of Sumatra and Borneo, but the literature on this is even sparser than that which discusses the Mon-Khmer elements in Malay. Those forms which are of ultimate Mon-Khmer origin and which are widespread in languages of island South East Asia, such as (here I cite Standard Malay forms) *kerbau* 'water-buffalo', *kembar* 'twin', *emas* 'gold', and *perak* 'silver', have all been diffused into Sumatran and Bornean languages, and into others (for instance the first two and the last forms have found their way into Tagalog), through the medium of Malay, together with numerous Malay loans of Austronesian and other vintages.

At least a few dozen high-frequency lexical and other elements which are clearly of Mon-Khmer origin are common to all the languages, including Acehnese, so that we may assert that they belong to Proto-Chamic. Several dozen more such forms, many of them equally high in frequency, are shared by most or all of the Chamic languages apart from Acehnese (which may however have lost some of them and may have replaced them with loans from Malay), and these can be attributed to what we may call Core Chamic. (It is apparent that all these Chamic languages, possibly excepting Acehnese, were still straggling dialects of one language in 982, which suggests that the innumerable changes which have taken place in all directions in these languages have occurred in the space of 1000 years or less. Some of these changes can be dated even more precisely than that, and as Thurgood's work has shown, many of these changes can be ordered sequentially and chronologically.)

At least as early as the period of the Sixth Cham Dynasty, which began its reign in 875, the older form of Written Cham, which was in use as a royal and epigraphic language over a millennium ago with Sanskrit as the language of Champa (see below), had already embodied elements from certain Mon-Khmer languages in addition to adopting numerous loans from Sanskrit (and some of the latter are also found in other Chamic languages of South East Asia). An early observer of Chamic, the Alsatian linguist Himly (in Himly 1890: 326), already noted the fact that Written Cham incorporated both Malayic and non-Malayic elements and that this variety of the origins of Chamic vocabulary could be seen at the most basic level. Although he pointed out that Cham was only as much a mixed language in the sense that English was one, he cited two Cham sentences, one comprising only words also found in Malayic languages, and the other built up (or so he erroneously thought) entirely out of elements which are not found in Malay. Himly provided these examples in order to demonstrate the combination of non-Malayic and Malayic elements in Cham.

The modern (but still archaising) form of this written Cham language, though known to a diminishing number of Chams (mostly male) has been confined to Chams and has not been used by, nor has it exerted influence on other Chamic languages, which did not have written forms until the French came. But nowadays there is a small amount of writing in a modern form of spoken Phan Rang Cham using an orthography based on Vietnamese (a sample of this from a Bible translation appears in Campbell 2000: 327). Literacy work in some other languages has been adumbrated in the past few decades, largely by American and Vietnamese Protestant missionaries.

We can set up a simple chronology for most of the major external developments and movements which have characterised this influence on the various languages; Thurgood (1999: 1-27) is an exemplary guide to this, while a chronology of mostly external and non-linguistic events affecting speakers of Chamic languages is provided in Table 1.

Table 1: *A partial chronology of some external developments in the sociohistory of the speakers of Chamic languages.*

c. 200 BC	(approximate date; the occurrence may be some centuries earlier) proto-Chamic speakers part from other speakers of Malayic languages and migrate to southeastern Indochina from southern Borneo. They set up the empire known as Champa, whose culture is later influenced by Hinduism and also by Mahayana Buddhism, and the language is strongly influenced by the Mon-Khmer Bahnar.
192 AD	Champa is first mentioned (as Lin-Yi) by Chinese chroniclers.
c.350	The first known inscription in Cham, composed in a language which is influenced by Sanskrit and written in an Indic script, is carved at Tra-Kiêu, central Vietnam.
c. 800	Cham inscriptions, some which are bilingual with Sanskrit and others which are monolingual, are produced in greater number from this period onwards. The Cham cities, which are principally ranged along the coast, can be divided into a nothern and a southern empire.
982	The northern Cham empire, with its capital at Indrapura, falls to northern Vietnamese invaders, who themselves were being driven south by the Chinese. Some Chanic speakers flee to the extreme north of Sumatra by way of Malaya (there are some place-names indicative of this along the east coast of the Malacca Peninsula), and thereby give rise to Acehnese, while other Chams go inland. At this time the Chamic languages are still a connected dialect chain.
c. 1100?	Islam comes to Champa and eventually supplants Hinduism among most of the Chams, in addition to influencing the beliefs of some highland Chamic groups. Maybe it is at this time that a merchant branch of the Northern Roglais seeks refuge from political turmoil in Indochina by splitting from its fellow-speakers and emigrating to Hainan, although this event may gave taken place a few centuries later.
1292	Marco Polo mentions meeting Acehnese people in northern Sumatra at this time, this is apparently the first certain record to indicate that Acehnese speakers had reached Sumatra, although we cannot be sure when they arrived and how long their journey took.
1400-1500	The Chinese-Cham vocabulary of words and phrases, written entirely in Chinese characters (therefore presenting numerous problems of the interpretation of presumed sounds) and discussed in Blagden (1940-1942) is produced around this time, probably before the fall of the southern Cham empire in 1471.
1400s	The Khmer empire of Angkor is destroyed by the Chams in retaliation for a series of raids upon Champa by the Khmers.
1471	The southern Cham empire, with its capital at Vijaya, falls to invading Vietnamese from the north, and Cham self-rule is at an end. Chams are subordinated to the Vietnamese (apart from those who flee to Khmer rule after this conquest) and the last remnants of the power are eroded during the following century. Some Chams move west into the highlands among Bahnar-speakers and form the group known as the Haroi. Many Chams are already Muslim by now.
mid-1600s	Islam finally took root among most of the Chams by this time, but some other groups in Vietnam nonetheless retained their Hindu beliefs, but groups such as the Rades practise more syncretic religions. By this time Cham rule over traditoonal territories had weakened from its former might to a state of puppet government.
1960s-1970s	Massive disruption in Southeast Asia as the result of the Vietnam War and the genocidal actions of the Khmer Rouge (in which the Western Chams, as Muslims, were especially heavily targeted). Thousands of speakers of Chamic languages are killed; thousands more are displaced (some migrate to parts of the US such as California and North Carolina, or to Australia or France, others are dispersed to other parts of Southeast Asia).

The split from Proto-Malayic and the move of the speakers of the language which was going to become Cham from Borneo to South East Asia predates the Christian era by maybe a couple of centuries, though we cannot be sure. Champa, the Cham kingdom in what is now Vietnam, which was characterised by its Mahayana Buddhism-influenced Hindu religion and its written language using an Indic alphabet, was first mentioned by Chinese chroniclers in 192 AD; they referred to it as *Lin-Yi*, though the Cham name was *Lemap*. At the time of its greatest extent, Champa stretched from the Vietnamese coast around Danang to the top of the Mekong Delta, encompassing portions of modern northeastern Cambodia and the parts of Laos as far as Pakse. To the south of this area was the Funan empire, the linguistic identity of which (Austronesian, Austroasiatic or otherwise) is still unknown. The first inscription in Cham, a bilingual stone which is also inscribed in Sanskrit (but with both inscriptions written in Cham script) and coming from Tra-Kiêu in Vietnam, apparently dates from c. 350 AD. But most of the 75 or so Cham inscriptions date to the 9th century or after, a period of stele-inscription starting with the Sixth Cham Dynasty.

The northern Cham kingdom crumbled in the period beginning in 982 under the impact of Vietnamese attacks, at which point the Acehnese speakers' ancestors headed south via the Malay Peninsula. At this time they were speaking a language which had already absorbed a considerable number of Mon-Khmer lexical items (though its relatives remaining in Indochina were to absorb far more) and which had adopted the Mon-Khmer syllabic pattern. However, the influence of Mon-Khmer languages upon the Chamic languages which remained in situ was exacerbated in the coming millennium as the power of the Chams declined. Relations between the various groups in Indochina were not peaceful: in response to repeated Khmer attacks on Champa, Thais supported by Chams eventually destroyed Angkor in the 15th century. The southern Cham kingdom, with its capital at Vijaya, was crushed by the Vietnamese in 1471, a couple of centuries or so (we believe) after Islam came into the area. Meanwhile the speakers of Tsat went northeast to Hainan, where they now live near Sanya City. It is possible that the speakers of Tsat, the Utsat, were not yet Muslims when they reached Hainan Island, although we cannot be sure. (It seems likelier that they were Cham merchants who had embraced Islam.) Nor can we be certain that Tsat's presence on Hainan is the result of a single migration from the mainland; there may have been two, one around the 11th century or maybe earlier, and one a few centuries later (Pang 1998, Thurgood 1999: 212-232).

Round about this time, or at least at some time between 1403 and 1511, a Chinese glossary (reproduced and discussed in Edwards and Blagden 1940-1942) listed about 500 Cham words and phrases, using Chinese characters to write them. Edwards and Blagden took the Cham equivalents from the dictionary of Written Cham by Aymonier and Cabaton (1906). This document demonstrates that there were quite a few elements from Mon-Khmer languages which were already in use in the Cham language at this time. Indeed the mix of Malayic, Mon-Khmer and obscure elements in this vocabulary has remained stable, and it is broadly similar to that which is found in modern speech: the same ideas which are expressed in this document by words of Mon-Khmer origin are expreseed likewise in modern Cham, and there has been little if any further relexification of basic Cham vocabulary in the direction of Mon-Khmer languages. Thurgood (1999) does not cite or use this source.

Subsequently the speakers of Chamic languages lost political power and their languages came under ever greater influence from Mon-Khmer (and other) languages, by whose speakers they were surrounded and in which they were often bilingual. The smaller Chamic-speaking groups, such as the Harois, were naturally the ones which were more vulnerable to change by and ultimately to assimilation to neighbouring linguistic communities, whereas bigger and more remote groups, such as the Rades and Jarais, resisted linguistic assimilation and the effects of profound linguistic contact much more strongly. Nevertheless one must allow that the profession of Islam by some groups in areas to which Islam was otherwise alien and where it had no other followers enabled (or compelled) these groups to be endogamous, to resist full absorption through intermarriage, and thus to prevent wholesale linguistic and cultural absorption by surrounding groups. There are linguistic consequences of this. The language through which the Chams learned about Islam (more specifically firstly about the Bani form of Sunni Islam) was Malay, and at least five Cham-Malay glossaries, with all their entries written in Cham script, have been found dating to the 16th and 17th centuries (Blust 1992).

The Mon-Khmer influence upon these languages goes far beyond the effects upon the lexicon; it has affected the phonology and typology as well as the morphemic inventory of these languages. It is unlikely that Acehnese has been in touch with any of the other Chamic languages since their split, and this historical consideration is heuristically useful for further reconstruction of Proto-Chamic features, as an objective correlative: if a featture introduced from Mon-Khmer is found in both Chamic and Acehnese, then it must reconstruct to Proto-Chamic. Indeed Thurgood (1999: 47-58) makes the point several times that Acehnese looks much more like reconstructed Proto-Chamic than one might initially expect. (The same set of circumstances as is found in Acehnese, that of a speech community's early and definitive sundering from the main body of Chamic speakers, is also true of Tsat, and some historical linguistic inferences can be drawn from this.)

In addition Mon-Khmer languages have influenced the morphology and indeed the syntax and word-order patterns of Chamic languages (which, like Mon-Khmer languages, are not heavily inflected, so that there is less scope for grammatical change to occur by means of transfer of elements, though even some of this has taken place). The impact on Highland Chamic languages, which in some cases never lost contact with the same Mon-Khmer languages which had even shaped Acehnese when it was one with the rest of Chamic, has been especially strong. Thurgood (1996) has paid much attention to developments in the phonologies of the Chamic languages, but he omits close discussion of developments in Acehnese.

The possession of a strong word-final stress in disyllabic or longer words (and the allowing of most vowels only in stressed syllables) was a phonological feature which distinguished Mon-Khmer languages from Austronesian ones, in which stress was unpredictable but was often penultimate. The Chamic languages took over the Mon-Khmer stress patterns, applied them to their own language, and therefore shifted the stress of disyllables and longer words in their native language to the final syllable. This was the first step in what resulted in the development of iambic (short-long) syllabification (often with subsequent development to monosyllabism, if the resulting initial consonant cluster was pronounceable) as the unmarked form of syllabic structure in most or all the languages, including Acehnese. The range of possible initial two-element consonant clusters, a phonotactic phenomenon which had not been permissible in Proto-Malayic, was later expanded in many Chamic languages by the borrowing of Mon-Khmer words which contained such clusters, and which were taken over with minimal adjustments. (The expansion of this roster of initial consonant clusters continues in Phan Rang Cham: Blood

1967). The introduction of such clusters licensed their use in loans from other languages, and also in other lexical items which entered the vocabularies of Chamic languages, and which have clear etymologies neither in Austronesian or its subgroups nor in Mon-Khmer languages, nor in other contact languages such as Arabic or Sanskrit.

This process is common to all the Chamic languages and is the cause of several further sweeping phonological changes in Chamic languages, although some later developments are exclusive only to a subset of them. It moved on to the development of register systems in Haroi, to an allophonic high pitch before a final glottal stop in Jarai (Blust 1990: 142), and a partial tonal system, with three or four distinct tones which have developed from two, in Phan Rang Cham in Vietnam. It has also brought about a full five-tone system in Tsat (Hainanese has six tones), with concomitant replacement of disyllables by monosyllables. The development of a partial tone system in Phan Rang Cham, which was originally conditioned by the voicedness status of the initial consonant of a particular syllable, has led to the replacement of phonemically distinctive voicing in stops with a distinction between high and low tone in these words. Low tone is ‘marked’ and is the relic of the presence at the beginning of the syllable of former voiced occlusives, which triggered off a kind of suprasegmental phonation type which in its turn brought about intonational changes. For instance earlier /pa-/ remains /pa-/, but earlier /ba-/ has become /pà-/, while earlier /ɓa-/ has remained /ɓa-/. Forms in Phan Rang Cham which do not exhibit this low tone but which have voiceless obstruents at syllable onset also have slightly aspirated stops. The use of low tone has not spread any further in this language to date, so that we do not find Phan Rang Cham syllables beginning */mà-/ or */sà-/, for instance. Nor do we find Phan Rang Cham incorporating any of the six tones of loans from Vietnamese loans into the phonological forms of these loans into Phan Rand Cham.

Such a tone split, conditioned as it is by phonation changes and register developments involving the interplay of the feature of breathy voice in vowels with that of different kinds of initial consonants in monosyllables, is a South East Asian areal feature. Similar changes have taken place at various times in Tai languages, in some Tibeto-Burman and Sinitic languages, and also in Vietnamese. This change has not happened in Western Cham because there the major contact language, Khmer, is non-tonal, although both Khmer and Western Cham have been prone to the effects of breathy voice phonation in the readjustment of occlusive systems which formerly contained a voiced/voiceless distinction, but where older *p/b* are now *ph/p*.

Also, most of the Chamic languages have acquired typically Mon-Khmer (and atypically Austronesian) features of phonation, such as ptreglottalised forms of /b d/ and often also a preglottalised palatal stop, in addition to the usual exploded voiced forms which have been inherited from Proto-Malayo-Chamic (though the exploded ones have subsequently been devoiced in Phan Rang Cham). However, this acquisition of implosion or preglottalisation has apparently been acquired and subsequently lost in Acehnese as far as we can tell. On the other hand, the Acehnese vowel system has (like the other Chamic vowel systems) expanded in the number of qualities to nine (or in some dialects, ten). These qualities and distinctions have been elaborated in the first instance from the four vowels (*i e a u*) which the Austronesian component of Chamic (more precisely, the Malayic component.

As Blust (1995) has made clear, all Austronesian components that are attested in Chamic languages go back either to Malayo-Chamic or else they are loans from Malay proper, where they may in turn be loans from other languages) inherited directly and without change from its ancestral language. To these there had been added a number of vowel nuclei that had been taken over from Mon-Khmer languages and which were initially brought into Proto-Chamic through borrowed items. Some Chamic languages, including Phan Rang Cham, have secondarily developed contrastive vowel length, especially in the vowel pair /a:/ versus /a/. A few of them also have developed contrastive vowel nasalisation; Haroi is most notable in this respect, having abandoned the register system which it formerly had and having ended up with up to 32 vowel contrasts (including distinctions relating to vowel length and nasalisation). All of these have evolved from the typical Austronesian four-vowel system about two millennia ago, in tandem with the reflexes of the borrowed vowel nuclei mentioned above. This represents an almost 100% increase on the number of phonemic vowels in Cham, from which Haroi diverged about 500 years back (the development of the Haroi system is discussed in Thurgood 1997).

One of the most characteristic changes which has occurred in Chamic languages and which has been brought about by Chamic speakers' contact with speakers of Mon-Khmer languages is a direct result of the conversion of traditional Malayic disyllables into Mon-Khmer-style iambs (namely a light syllable plus a heavy syllable). This result is the consideration that the two parts of the iamb, which Thurgood calls the 'pre-syllable' and 'the syllable', the latter bearing the primary stress, are subject to differing constraints upon the range of initial consonants which are permissible. This feature is also typical of many Mon-Khmer languages, and the constraint within Chamic is probably copied on a principle transferred from Mon-Khmer languages. The Chamic 'syllable' in Thurgood's analysis, that is, the part which was the second syllable in Proto-Malayic or Proto-Malayo-Chamic, may begin with a wider range of consonants (and vowels) than that which the pre-syllable permits. The number of these possible pre-syllable initial consonants is especially limited in Rade, which has only /h k m/ as initial pre-syllable consonants occurring in what were originally consonant-initial disyllables, a process of consonantal assimilation of features which took place after the vowels of the presyllables had been deleted. Additionally certain consonants (for instance liquids and /d/) have merged with following vowels in Rade and this combination has been realised as the vowel /e/, for instance Proto-Chamic **dara* 'girl', a form found in Malay as *dara*, gives Rade *era*. (This matter is discussed further in section 4).

Other Chamic languages have responded to the consequences of the adoption of the Mon-Khmer syllabic structure in several other ways, as Thurgood (1999) amply demonstrates. Many of these reflect the imitation of phonological processes of Mon-Khmer languages whose speakers have exerted power or prestige among speakers of Chamic languages. For the record, Tsat has taken this pattern of syllable reduction and contraction even further, inasmuch as most pre-stressed syllables have been dropped (though not in a completely regular or predictable pattern) and now only the orignal stressed syllables remain.

Furthermore, and this has occurred originally as a result of the development in Chamic of Mon-Khmer syllable types, these languages, Acehnese included, have a very high proportion of lexical and especially contentive stems which are monosyllabic. The proportion of these is increased by the high proportion of such stems in all the form-classes among the Mon-Khmer components in Chamic languages, although many other monosyllables in these languages were once Austronesian disyllables which have first of all lost the first vowel and which have thereafter had their first syllables compacted, or which have contracted previous sequences of contiguous vowels into one vowel. This phonotactic development is especially remarkable when one considers that in the closely related Malayic languages, which are the closest genetic relatives of Chamic, only some few functors and some English, Dutch and Chinese loanwords are monosyllabic and monomorphemic. Indeed all the evidence which we have suggests that Proto-Austronesian did not possess any monosyllabic contentives (whatever the primary nature of many Austronesian disyllabic roots, especially those with primarily verbal senses, may have been: see Blust 1988), and that only some particles were basically monosyllabic in form. Paradoxically, in those Vietnamese Mon-Khmer languages, like Chrau, which borrowed items of (mostly acculturational) lexicon from Chamic languages, such loans are one of the chief sources (together with later loans from French) of monomorphemic contentive disyllables.

The results of this phonological change have been manifold, depending upon the languages which have influenced each particular Chamic language. Some such languages have retained the original system fairly closely: Rade and Jarai, for example, have both undergone many phonological changes from Proto-Chamic, and those in Rade have been especially striking (see above), but they have not developed such register systems (apart from the development of an allophonic high pitch on the vowel before a word-final glottal stop in Jarai). It has led to the development of register systems in Western Cham, of a restructured register system in Haroi (of a kind which is also found in Bahnar), of a four-tone systems from an incipient two-tone system in Phan Rang Cham and of a five tone system (with a possibly concomitant deletion of unstressed syllables and a consequent remodelling of the segmental phonology towards that of Hainanese Chinese) in Tsat. Indeed, Thurgood (1999: 214-232) shows, in his discussion of Tsat tonogenesis, the possession of strong similarity of patterns between the tonal typologies of Tanchou Hainanese (the variety of Chinese with which Tsat speakers would have contact), plus those of two Hlai varieties, and that of Tsat. Both Hainanese and Tsat have three level tones, and each also has a falling tone and a rising tone, while Hainanese has further tonal distinctions not found in Tsat.

2.1. Excursus: Historical issues raised by Dyen's 2000 review of Thurgood (1999).

There have been two major reviews of Thurgood (1999). The first appeared in *Oceanic Linguistics* and was written by Robert Blust (Blust 2000). This was justifiably overwhelmingly positive, and Blust's criticisms centred mostly on points of detail, issues in the phonological reconstruction of features of Proto-Chamic, some potentially misassigned etyma, and so on. The other review is a five-page treatment by Dyen (2000), which appeared in a journal (*Anthropological Linguistics*) with a potentially wider readership than *Oceanic Linguistics*. Dyen's review is much more critical of Thurgood's work, but it is sometimes difficult to see what point Dyen is trying to make which stands in each case in contrast with the analysis of the facts that had been provided by Thurgood.

There is indeed room for criticism of Thurgood's book, although such criticisms are minor and they would more readily reflect the particular tastes of the critic. The treatment of the changes of various kinds that occurred from PMP to Proto-Malayo-Chamic, especially some distinctive innovations (such as the deletion of initial *qa-* in many stems that have been reconstructed as being trisyllabic in PMP), could have been made a little clearer. The Proto-Chamic lexicon could have been checked to ensure that it included all the entries on the Swadesh lists, or better still all the entries on the Blust list, since this is used so much in Austronesian studies. The phonological forms of many Proto-Chamic reflexes in the modern languages include a number of irregularites from one word to another (including irregularities in predicted initials, finals and other parts of the form) which merit further explanation, but which Thurgood does not give. More could have been said about the linguistic sources of the differences that we find between the various modern Chamic languages (especially those spoken in Vietnam) and about the morphosyntactic structure of these languages. More could have been said about the primary distinction between highland and coastal Chamic languages and the extent to which this distinction was genetic and was based on linguistic rather than on cultural or geographical considerations. And there are a few interesting sources (for instance Edwards and Blagden 1940-1942) which could have been cited by Thurgood but which weren't. Nonetheless these criticisms are minor. Dyen's criticisms concern Thurgood's historical approach to his material, and although some of these criticisms are well-founded, the solutions which Dyen proposes are not very helpful.

Dyen's main concerns are to do with the historical and other relations between Acehnese (which he spells *Achehnese* throughout) and the other Chamic languages. The presentation of Dyen's arguments is often unclear, as is any alternative hypothesis about the immediate relationships of Chamic, and his claims can be interpreted in more than one way. Furthermore, his assertions do not explain the existence of certain lexical and structural similarities between Acehnese and Chamic languages (such as the change of initial PMP *n—to /l/) that are not shared by Malay. Dyen does not give sufficient account of the shared innovatuons at various levels (lexical, phonological, etc.) which mark Acehnese and the other Chamic languages off from other Austronesian languages, and those further lexical and phonological innovations (including sporadic sound changes which affect particular words both in Malay and Chamic) which go back in their inception to a period of Malayo-Chamic linguistic and genetic unity.

Dyen claims that the present-day Acehnese represent an offshoot of the Chams who were returning to their previous home in Sumatra. This is because he sees the Acehnese and Chams together as having migrated from Aceh to Vietnam and then back again (though he does not say why they would have done this). His ideas do not explain why the numerous loans from Mon-Khmer languages that are common to Acehnese and other Chamic languages (and also those which are found only in Chamic languages) derive specifically from Bahnaric languages of central and southern Vietnam, rather than from the Aslian languages with which the Acehnese would have come into contact on the Malay Peninsula. Indeed his ideas do not explain why distinctively Aslian forms are (potentially) to be found only in Acehnese and not at all in the lexica of all Chamic languages.

In short, Thurgood's book, together with Blust's review, provides the best concise picture of where Chamic fits into Malayo-Polynesian and of the relationship between Chamic and Acehnese.

3. Early lexical strata in Chamic and their historical significance.

Even when we discount the potential genetic relations between Austronesian and Mon-Khmer languages within such tentative concepts as 'Austric' (= Austronesian plus Austro-Asiatic) and after we disregard the shared morphemic items which have been posited as existing within Austric, we can see that lexical influence between Mon-Khmer languages and Chamic languages has been bidirectional. Katuic languages and additionally Bahnar have been especially strong recipients of Austronesian loans from the Chamic languages which abut them, as well as having in common a number of lexical forms which are shared with Chamic languages, but the direction of whose diffusion is uncertain because the origin of these forms is unknown. Most of these borrowings from Chamic into Bahnaric or Katuic (and from Tsat into Hlai languages) are names for introduced concepts or items, and therefore they supplement rather than replacing original Mon-Khmer (or Hlai) lexicon. The converse is not true of Mon-Khmer loans into Chamic. But in this discussion we are only concerned with the various forms of influence exerted on the Chamic languages by the neighbouring and the superordinate Mon-Khmer languages.

The impact of Mon-Khmer languages (and more especially Bahnaric languages) on the lexica of the various Chamic languages has been considerable, to say the least. Thurgood's book provides a lot of useful information on this, although much of the theoretical significance of the quantity and quality of the borrowings has to be gleaned by analysis from the lists which he provides, rather than it being summarised readily. An extensive comparative vocabulary of Chamic languages, including forms from Acehnese, and with etymologies provided where feasible, would be more than welcome. Thurgood (1999) does not provide this, although he does furnish the reader with a great deal of lexical and other information, including etymologies where possible, and he provides parallels with Malay where they are felt to be necessary. It should be noted that two useful lexicographical works relating to earlier stages of Bahnaric languages which would undoubtedly have cast further light upon the etymological composition of Chamic languages, namely Jacq and Sidwell (2000) and Sidwell (2000), were (given their dates of publication) unsurprisingly unavailable to Thurgood when he wrote his book, though it is also true that the Bahnaric component in Chamic is from Northern and Central Bahnaric languages rather than from South Bahnaric languages such as Stieng or Chrau. However, he did have access to the Proto-Katuic material which was reconstructed in Peiros (1996), though he wisely made little use of this since its quality is poor (Peiros took Kuy, an especially aberrant Katuic language, as the starting point for the reconstruction of Proto-Katuic, thereby coming up with a seriously imbalanced and inaccurate reconstrucion). For Acehnese Thurgood had access to a prepublication form of Bukhari and Durie (1999).

From an examination of this, it is quite obvious that at the level of basic and non-cultural vocabulary (let alone for the names of items with which Chamic-speakers would not originally have been familiar) the Chamic languages are among the most heavily 'relexified' languages that one has ever seen. Much of their original Malayic vocabulary has been replaced with elements from Mon-Khmer (or, to a lesser extent, with items from as yet unidentified but possibly Mon-Khmer) sources. The basic Acehnese vocabulary clearly shows the effects of this partial 'relexification' with both confirmed and assumed Mon-Khmer elements, as the report above has suggested, and as we can see from the table. In fact, over 120 certain or probable loans from Mon-Khmer languages have been identified in Acehnese, and they occur at a basic level and in considerable number in almost all the language's form-classes, apart from numerals, which are purely Austronesian.

(A long list of certain and probable elements of Mon-Khmer origin in Acehnese is furnished by Cowan 1948. The list in Thurgood 1999, which makes no attempt at any completeness in regard to the etymological composition of the Acehnese lexicon, gives 45 early Mon-Khmer forms as having entered Acehnese via Proto-Chamic and 25 more as entering Acehnese after the Proto-Chamic stage but as being forms which are still to be found in another Chamic language. For the record, Thurgood provides feasible Mon-

Khmer etymologies, which he takes from Proto-Katuic, Proto-North Bahnaric or Proto-South Bahnaric, for 120 of the 275 forms which he thinks are Mon-Khmer words traceable to Proto-Chamic. He does the same for 83 of the 167 post-Proto-Chamic Mon-Khmer forms occurring in some Chamic languages or for other forms which cannot be traced that far. This reckoning excludes the elements acquired later, which are directly traceable to Khmer or Vietnamese).

But the picture is even clearer if one examines the lexica of other Chamic languages. Sometimes the latter languages preserve Austronesian terms which have been more readily replaced by Mon-Khmer terms in Acehnese, for instance Cham has *minom* (compare Malay *minum*) for 'to drink' whereas Acehnese more generally uses the Mon-Khmer loan *jep*, though it also has a reflex of *minom* too (Cowan 1981: 523). It is unfortunate that the languages which have undergone the greatest degree of phonological adaptation are generally not also the ones whose vocabularies are most fully represented in these lists. Nevertheless we should reiterate that the general high level of borrowing from Mon-Khmer languages is attested throughout the Indochinese Chamic languages (that is, not Acehnese or Tsat; the designation of 'Indochinese' used here is not meant to be a genetic label, though Acehnese and Tsat both contain many Mon-Khmer forms). On average, about 25% of the elements in the Swadesh list lexicon of any of these languages are taken from Mon-Khmer or from other non-Austronesian languages. Unfortunately we have no statistics for the proportion of such borrowed forms as they occur in the lexicon of any Chamic language as a whole, but it is likely to be considerable, and it is probably more considerable in terms of total vocabulary than what is found in the Swadesh list.

In an attempt to see something of the depth of Mon-Khmer influence on lexical fabric in Chamic, I counted Thurgood's listing of the Malayic and other elements, following up from the counts helpfully provided in Blust (2000). Thurgood lists 285 Austronesian (and in truth Malayic) elements in Proto-Chamic as against 277 Mon-Khmer ones, but the comments in Blust (2000) indicate that there have been a couple of misassignments either way. I found that the Mon-Khmer element in the vocabulary increased in proportion from the first language I examined (Acehnese) to the second (Rade), and that it increased, though by a smaller proportion, from the second language to the third (Jarai).

It would be illuminating for us to know how many items from Proto-(Western-) Malayo-Polynesian have been preserved in any Chamic language which has not subsequently been influenced by Malay, and to find out whether Thurgood lists all such attested forms in his lexical lists (he only lists items in any of the lists which are attested in more than one Chamic language). For what it is worth, Tharp et al. (1980) list only 177 stems of Proto-Malayo-Polynesian origin, which they inaccurately style [PAN], for Proto-Austronesian) on the etymological notes in their Rade dictionary, although they list at least twice as many forms going back to 'Proto-Chamic' which cannot be traced back to Proto-Malayo-Polynesian. This figure can be seen in context if we understand that the Rade dictionary under discussion contains approximately 1800 separate morphs, including the Latin letter names which are mostly recent transfers from French, while maybe a couple of hundred more (while other morphs which are listed, especially under *m*-, are actually bimorphemic transitive forms of certain verbs). There may even be a few morphs of Proto-Malayo-Polynesian or Proto-Austronesian origin which are attested in the lexicon of a particular Chamic language but which no longer occur in any Malay varieties (from which they have been lost), but they are very few.

In addition to the Malayic and Mon-Khmer elements there is a considerable lexical element (Thurgood lists 179 such elements) which is as yet unsourced but which probably contains a greater proportion of elements of Mon-Khmer origin than have yet been recognised (though the origns of some may simply be waiting to be recognise: for instance, the widespread form *yɔw* 'yoke' looks to me like a Sanskritism that is based on Sanskrit *yugam*, which would tie in well with the fact that many Chamic words relating to the domestication of animals derive from Indic languages). 26 of these items of uncertain origin are listed by Thurgood as also being attested in Acehnese, though he does not list items of any origin that are found only in Acehnese.

Two things are notable about this stratum of elements which are common to the Chamic languages but as yet unsourced. Firstly, a handful of them (but only a handful) are found in Malay as well as in Acehnese and other Chamic languages, so that they reconstruct to Proto-Malayo-Chamic with a status as lexical innovations there. Secondly, a considerable proportion of these 179 unsourced elements are monosyllabic contentives, including a number of verbs, 'adjectives' and free-standing function words such as personal pronouns. And furthermore these monosyllables include several Swadesh list items (for example they include the pan-Chamic form *thu* 'dry', which is attested in Acehnese and several other languages).

The phonology of these unsourced forms is also interesting, because they represent a post-Malayic stage of segmental and structural elaboration of the phonological system. These words very often incorporate the sounds which were brought into Chamic languages by Mon-Khmer loans, including implosive stops, initial and word-final occurrences of /c-/ (which is a rare phone in any case in Western Austronesian languages, and which is probably not one which can be reconstructed back to Proto-Malayo-Polynesian), and they also include the several non-Austronesian vowels and vocalic nuclei introduced by Mon-Khmer. (This is very rarely the case with unsourced terms attested at the Proto-Chamic level which are represented in Acehnese, however, and this is a consideration which may be historically significant.) Contact with Mon-Khmer languages has also introduced new vocalic nuclei, including some, such as /iaw/, which were built out of vowels or other elements which had previously existed as discrete elements in Malayic and thus in Chamic; these too are found in the 'unsourced' elements.

There are also some 167 elements that are given by Thurgood as being of later origin, they cannot be traced back to Proto-Chamic, though they are found in two or more individual Chamic languages. These words have again been taken in large measure from Mon-Khmer sources or else are of unsourced origin, apart from a couple which derive from French (maybe from Tay Boi or Vietnamese Pidgin French, which was known to many 'Montagnards'). Thurgood also lists a couple of dozen items which have been traced to Indic or Arabic sources and which are common to several Chamic languages and sometimes to neighbouring Mon-Khmer languages too (they are widespread cultural loans: Thurgood 1999: 346-349). Their phonological forms and their distribution among most or all the Chamic languages show that they are not recent direct borrowings from these donor languages (in any case, Sanskrit has been out of the linguistic frame in this part of the world for several centuries) but that they have been inherited from Proto-Chamic or from an early descendant of this. (There are additionally a number of contentive elements in Chamic languages which derive from Chinese languages, but these are indirect borrowings of a cultural nature which have entered Chamic languages via Vietnamese, Khmer or Malay, except of course in the case of the innumerable Hainanese forms in Tsat.)

Below I have presented a figure, labelled Figure 2, that compares the sources of the usual glosses for the traditional combined 215 and 100-item Swadesh list elements for the relevant languages, but which only uses lexical items that are cited above in Thurgood's book. It should be noted that the data which are available for Acehnese and Tsat, for example, are much fuller than those which are provided in Thurgood (1999), and the number of Mon-Khmer forms on the full Swadesh wordlist is closer to 45 than the 16 for which Thurgood gives equivalents. The 'Samples' column gives the total of such items in the relevant categories in the lists in Thurgood's book; Thurgood's lists are not to be taken as complete. (Grant 2005 in this volume uses the Blust list for a similar study.)

	Aceh.	**Rade**	**Jarai**	**Haroi**	**Chru**	**Cham**	**N.Roglai**	**Samples**
Austronesian	93	93	95	97	94	94	99	285
Early Mon-Khmer	13	48	56	44	47	46	43	277
Unknown - MK?	6	14	12	17	19	17	13	179
Other	1	3	1	3	3	3	3	24
Later Mon-Khmer	3	23	18	17	18	13	19	167
Total items	**116**	**181**	**182**	**178**	**181**	**173**	**177**	**936**

Figure 1: *Swadesh list elements in Chamic langages according to strata*

The 'Cham' variety which is documented in the table is the modern Western Cham variety of Cambodia (Phan Rang Cham would provide very similar figures), while the Roglai variety documented here is Northern Roglai. (Collins 1969 provides a fuller Swadesh list for Northern Roglai but he does not furnish the etymologies which are needed for me to be able to interpret it for the purposes of the above table.)

For the record, the equivalent figures for the Swadesh list elements in Tsat, according to the rather scanty data that were then available to Thurgood and thereafter until recently to myself (Thurgood 1992 states that he had only about 500 Tsat lexical items at his disposal despite using all available English, French and Chinese-language sources), are as follows: items deriving from Austronesian or its daughter languages (that is, Proto-Malayo-Chamic) 82, early Mon-Khmer items 19, later Mon-Khmer items 4, unknown 8, other sources 1 (this is a single loan, 'person', from Indic, which is common to the Chamic languages).

Zheng (1997), written in Chinese, provides a form-class/semantically-ordered Chinese-Tsat vocabulary of some 2428 items, with parallels in Rade (from Tharp et al 1980) and from Lee's reconstruction of Proto-Chamic where such forms were available. Zheng (1997: 54) points out that of these 2428 items, 211 of the 1005 Tsat nouns (21%) derive from Chinese, as do 120 of the 843 verbs (14.2%), 40 of the 267 adjectives (15%), 57 of the 182 classifiers (31.3%), 3 of the 38 pronouns (7.9%) and 42 of the 93 conjunctions, prepositions and adverbs (especially the conjunctions; this proportion constitutes 45.2% of such forms in Tsat). But here again the lower numerals are non-Chinese but reconstruct back to Proto-Chamic. On the other hand, only a few forms in the Zheng list are derived from Hlai, the Kam-Tai language which the Tsat-speakers first encountered (Graham Thurgood, personal communication, April 2003).

To recapitulate, Proto-Chamic had already absorbed lexical elements from North and Central Bahnaric and had begun to undergo the phonological changes which gave rise to sesquisyllables, all in the period before Acehnese and later Tsat split away from the

Chamic-speaking area and became de facto languages on their own. In addition it had acquired a considerable number of lexical and other items whose origins may lie in Mon-Khmer languages, but which remain to be etymologised in terms of them. All these developments had occurred in the period before Acehnese (which participated in these changes) split off and returned to island Asia by way of the Malacca Peninsula, where it acquired words from the Aslian languages. Before Acehnese split off, Proto-Chamic had already absorbed a large number of 'basic' lexical and other elements from Mon-Khmer languages, to the extent that it had partially relexified, replacing many Austronesian or Proto-Malayic forms with items transferred or copied from Bahnaric languages. (One word which is possibly of Mon-Khmer origin, a form of *putao* 'king', occurs in the earliest inscription, which is incidentally the earliest recorded shred of Austronesian linguistic material. However, *putao* may be of Austronesian origin, as Marrison (1975: 53) suggests, offering as origin the Cham etymology *pu tao* 'lord-person'; this word occurs in Chrau, although the second part of Marrison's Cham etymon, a well-know Austronesian form (for instance Tagalog *tao* 'person'), does not otherwise occur in Cham.[2]) Acehnese later acquired further lexical elements of Mon-Khmer origin in the course of its speakers' peregrinations through Malacca, as we know; we recall that the Acehnese attacked and controlled much of Malaya in the 13th century.

And in absorbing those words, Proto-Chamic had also expanded its vocalic inventory, as we have seen before. This inventory now included open forms of /e o/, and schwa (vowels which also occur in inherited items as vowels preceding semi-vowels in words which ended in /-i/ and /-u/ in Proto-Malayo-Polynesian, for instance the occurrence of schwa plus /-y/ at the end of words which once ended in /-i/). It also included the complex nuclei /ia ua iaw/ (these latter nuclei consisting of elements which previously existed in Proto-Chamic though not as nuclei except maybe in one or two inherited words as the result of crasis; some Chamic languages, such as Acehnese, developed further vowels). The above nuclei are elements which it acquired through taking over words from Mon-Khmer languages which contained them. The presence in some Chamic languages of contrastive vowel length, a feature which was absent from Proto-Austronesian and Proto-Malayo-Chamic but which is one source of the 'heavy' syllables that are so characteristic of the second part of Mon-Khmer (and by.extension, also many Chamic) roots, is also due in no small measure to the impact upon Chamic of Mon-Khmer languages, but these features are also found in the Chamic reflexes of a small number of inherited words, where they may have arisen as the result of the operation of earlier phonological rules.

All this expansion must have happened at a time before Acehnese split off, since Acehnese shows signs of undergoing these developments. This was a time when Proto-Chamic was in a position vis-à-vis Mon-Khmer languages that made it easy for Proto-Chamic to absorb words from them essentially phonologically unaltered. Thurgood notes that these typically Mon-Khmer sounds had only spread to a very few words of Austronesian origin even by the time Acehnese split off, and it appears that they do not seem to have spread further into the Austronesian (or Malayic) stratum. An examination of the Proto-Chamic and post-Proto-Chamic lexical material which he presents shows one instance (plus another more problematic one) of an initial imploded /ɓ/ (/ɓuk/, a word

[2] Another feature which makes this etymology rather unusual is that Western Austronesian languages (and specifically languages such as the one which gave rise to Malayic and Chamic languages) rarely have monosyllabic contentives as part of their inherited vocabulary.

meaning 'hair'; see Blust 1973 for a discussion of similarly implosive reflexes of this selfsame word in Bintulu and some other Bornean languages, deriving as it does from Proto-Austronesian **buSek*). But there were no clear instances of initial imploded /ɗ/ in the Malayic stratum of Chamic languages, although there are 19 word-initial occurrences of both these sounds (meaning 19 occurrences for /ɓ/ and 19 others for /ɗ/) in the non-Malayic strata of Chamic lexica as this is presented by Thurgood, and although the Chamic reflex of Proto-Austronesian *Zauh 'far' has /ɗ/ in many Chamic languages.

Nonetheless, a very important point about Thurgood's lists of Mon-Khmer elements in Chamic needs to be made. Thurgood often posits a Mon-Khmer origin for a word for which he is unable to provide a Mon-Khmer etymon from the available resources. But he assumes these words to be taken or copied from a Mon-Khmer language because of the phonological nature of the word (for example as a consequence of its possession of certain vowels, or because of the occurrence of non-final glottal stops or whatever). Headley (1976) suggested that about one-tenth of the forms which were then reconstructible for Proto-Chamic derived from Mon-Khmer languages, that is, some 72 out of about 700 items came from Mon-Khmer languages, and he provides numerous examples of these with supporting evidence from Chamic and Mon-Khmer languages. There are certainly more than 72 elements from Mon-Khmer languages which can be pointed out in Proto-Chamic, but the actual number of such elements, and secondarily the proportion of these within the reconstructed Proto-Chamic lexicon, is not yet certain (it appears to be over 200). But we may assume that the number of such identified loans will rise as our understanding of the histories of Austroasiatic mid-level proto-languages increases.

But whatever the final number and proportion of elements of Mon-Khmer origin within the various Chamic-language lexica, what we find on looking at the two lists of Mon-Khmer elements which Thurgood provides is that they exhibit two interesting characteristics which are crosslinguistically very unusual. Firstly, they include a very high proportion of items belonging on the Swadesh lists, or on the Blust list (and therefore indicating 'core vocabulary' of the kind which is supposed not to be replaced frequently through borrowing, or indeed is thought not to be able to be borrowed at all). The proportion of such borrowed forms reaches to almost 40% on the Jarai list (for which Thurgood's material provides 185 out of the 207 standard Swadesh list items; we find 74 such Mon-Khmer elements listed for Jarai on the Swadesh list, plus many more Jarai lexical elements which derive from Mon-Khmer but which do not appear on the list).

By comparison, some 96 out of the 185 Jarai items are attributed to Austronesian (or at least Malayo-Polynesian) sources and are therefore part of Jarai's 'genetic inheritance' from Proto-Chamic. These forms are unlikely to be later Malay loans into Jarai, since the Jarais are not Muslim and therefore have no direct contact with Malay. For the rest, 12 forms are attributed to the 'unknown but possibly Mon-Khmer' stratum, and one (a form for 'person, human being' that is exemplified by Malay *manusia*) is attributable to Indic. (Details of the loan element contents of the Swadesh and Blust lists for Jarai, with the English glosses of the relevant loan forms are given in Table 2.)

Table 2: *The glosses of elements on the Jarai versions of the Swadesh and Blust lists which derive from Mon-Khmer, other non-Austronesian, or unknown sources*

EARLY MON-KHMER BORROWINGS (OR ASSUMED BORROWINGS FROM MON-KHMER LANGUAGES) WHICH ARE THUS SHARED WITH OTHER CHAMIC LANGUAGES:	
Nouns:	Head*; neck*; husband/male; man/male; wife (2 forms*); firewood*; leaf; grass*; sand; mountain range*; bird (also means: animal); a fly; rope; river; meat*
Adj. equivs:	Old*; correct/right; other; black; white*; dry; warm; big*; narrow*; small*; good; round (2 forms)
Verbs:	To hold; to sing; to hit or strike (2 forms, 1 of uncertain origin); to vomit; to swell; to scratch*; to eat*; to bum (2 forms); to dig*; to stand; to lie down; to sleep (2 forms); to climb; to pull (2 forms*); to flow or run; to split; to bite*; to break*; to spit (2 forms); to yawn; to steal; to say; to swim; to cut; to choose; to open (2 forms); to wash (2 forms, 1 of them uncertain in origin); to weep*; to tum
Other classes:	Not; what?*; near; and/with*
LATER MON-KHMER LOANS (OR POSSIBLE LOANS) WHICH ARE NOT COMMON TO MOST CHAMIC LANGUAGES BUT WHICH ARE FOUND IN MORE THAN ONE SUCH CHAMIC LANGUAGE:	
Nouns:	The back (anatomical); hom*; thunder; spider
Adj. equivs:	Dull or blunt; right side*; left side; dirty; heavy*; cold
Verbs:	To see; to smell; to squeeze*; to hide*; to flow; to hear; to fear; to pull; to wipe away
Other classes:	we; thou
LOANS FROM OTHER LANGUAGES:	
Nouns:	Seed (2 forms); person/human being; salt (all of these are from Indic sources and all are pan-Chamic or just post-pan-Chamic)
ITEMS OF UNKNOWN ORIGIN WHICH ARE ALSO ATTESTED IN AT LEAST ONE OTHER CHAMIC LANGUAGE AND WHICH CAN BE USED IN ARGUMENTS FOR SUBGROUPING:	
Nouns:	Female; roof; branch; root, water, forest; earthworm; cloud; night (or else the form is <Proto-Austronesian); house
Adj.:	Full; dry (2 forms); thin; much
Verbs:	to hold; to blow; to laugh; to suck; to cook
Other classes:	below; I (polite); that; thou; because

Jarai is the single Chamic language which apparently contains the greatest number of non-Austronesian elements in its basic vocabulary, as far as we can tell; it has therefore been chosen for discussion here. These forms are taken from Thurgood (1999: 279-370); the list was completed with the inclusion of a few forms from Lafont (1968). Asterisked forms in the relevant sections are those whose Mon-Khmer credentials are not secure but are assumed to exist by Thurgood for reasons of their un-Austronesian phonological characteristics. The words listed here are the glosses for the usual Jarai words for these concepts.

This proportion of borrowed items is astoundingly high, and is almost unparalleled in the record of the world's languages. (Robert Blust has pointed out that Tiruray of South Mindanao, a member of the Bilic sungroup within the Philippines, contains an impressively high proportion of basic vocabulary items which are known to be loans, but in this case the borrowed items come from other languages of Mindanao such as those of the Danaw group, and, unlike the loans into Jarai, are often Austronesian or at least Malayo-Polynesian items in origin. The Tiruray case is discussed in Blust 1993b.)

It early occurred to me, on examining these lists, that it might be the case that, were one to add to the Jarai Mon-Khmer Swadesh list the elements in other Chamic languages which came from Mon-Khmer languages, one would find that more than half of these items are represented in at least one Chamic language by an item taken from a Mon-Khmer language. This is a state of affairs that may well be unparalleled in the world's languages. I put these claims to the test by examining the relevant data. In Thurgood's 1999 materials I counted 106 Swadesh list glosses, out of 207, which were represented in at least one Chamic language by a form which is certainly or purportedly of Mon-Khmer origin, and sometimes more than one Mon-Khmer form in Thurgood's list had the same or a similar gloss as was used by another form. The number for forms of Malayic origins would naturally be somewhat higher. I counted 114 such Swadesh list items in Thurgood's materials, many of which were represented in some languages by Mon-Khmer forms and in others by Malayic forms, and here again there were a few cases where two different Malayic forms shared the same gloss, and sometimes both were found in the same language. The number of Swadesh list glosses which were represented by at least one common Chamic form which is as yet of unknown origin in at least two Chamic languages, on the other hand, was 26, while three Swadesh list items occurred in the list of Proto-Chamic loans from Arabic or Sanskrit, and a small number of Swadesh-list items were not presented in Thurgood's materials. The forms of 'unknown' origin are often common to more than one Chamic language, as we can see, and indeed several of them are found also in Acehnese; they include words as common as *sa:ng* 'house', a pan-Chamic item which is also found in Tsat. This has a form which is also attested in Acehnese though with the meaning 'tent'; in modern Acehnese orthography it would be spelt *seueng*.

The second feature which makes these loan strata so surprising or even anomalous from a crosslinguistic perspective is the nature and variety of the form-classes which they cover. With the sole exception of the lower numerals (at least those going up to '1000'), which are solidly Austronesian in origin and which actually show some traces of secondary innovative formations which support the special affinities of Chamic to the Malayic languages, they represent all structural form-classes, pretty much however these are classified. (In fact, several of the papers in Diffloth and Zide eds. 1976 indicate that language contact has gone in the other direction in the case of numerals, and that many Indochinese Mon-Khmer languages have borrowed some higher numerals from Proto-Chamic.). Indeed the borrowed items include even such rarely-transferred elements as deictic adverbs or demonstratives and some semantically empty adpositions, to say nothing of nouns which represent all major subject classes: kinship terms, topographical terms, names for implements, abstract nouns, a few body parts, and so on.

It is significant that these form-classes with Mon-Khmer members include numerous instances of the form-class of verbs. In fact, the number of verbs and other elements (and these include some but not all members of such form-classes as pronouns, particles, adpositions, place adverbs, negators, etc.) in this 275-item list of Mon-Khmer-derived Chamic forms is 138. This is very slightly more than the number of nouns and adjectives (which in any case are realised in these languages as a kind of verb). Were the Mon-Khmer-derived adjectives to be included in the non-noun category set up above, then the total of non-nouns would be almost twice the total of nouns of Mon-Khmer origin. The old folk-linguistic idea that 'languages don't borrow verbs' is seen to be commonly-held but erroneous, and this tranche of Jarai evidence as it has been presented in Table 2 disproves it completely (an examination of lexical evidence from other Chamic languages would show this just as well). More than half the Mon-Khmer elements which have been identified in Acehnese are not nouns, and again, most of these are stative verbs which are used as adjectives, or else they other kinds of verbs, although other form-classes, with the significant exception of numerals, also figure here. Crosslinguistically this is most remarkable. Only in the course of an examination of the lexicon of the apparently creole language Berbice Dutch (Kouwenberg 1994) have I otherwise come across such a situation.

The reason why it has been possible for so many verbs to be transferred into Chamic languages from Mon-Khmer languages, running as it does against the common crosslinguistic prediction about the unlikelihood or notional impermissibility of widespread verb transfer, can be traced to the parallels within the structure of the donor and recipient languages. Verb stems in these donor languages are readily isolatable and identifiable to the speaker or learner, since these languages lack bound inflectional verbal morphology (and have rather little in the way of bound derivational verbal morphology). As a result, they can also be inserted for use from such donor languages into a language in which inflectional verbal morphology is sparse, a criterion which the Chamic languages fit well. (We may also note the transfer of a free-standing Mon-Khmer anterior marker *jræy* into the free and productive morphological apparatus of traditional written Cham; this marker has no Austronesian source but has cognates in at least Khmer and Vietnamese: the form is given in Campbell 2000: 325-327. Work has yet to be carried out into the origin of many of the preverbal TMA markers in Chamic languages, though Thurgood [to appear a] points out that those in Phan Rang Cham are grammaticalised forms of preexisting verbs.)

The preservation of Austronesian (and more directly Malayic) smaller numerals in Chamic (Thurgood 1999: 37-38) is an especially interesting case. (Chamic and Malay share the innovations for '7, 8, 9' but the Malay innovation for '3', a possible loan from Sanskrit which is discussed in Dyen 1946, 1953, is shared only with Iban, Chamic preserving the Proto-Austronesian form.) This preservation has probably been assisted by the fact that many of the Mon-Khmer languages with which most Chamic languages were in contact used quinary or quaternary rather than decimal numeral systems. In contrast, the systems of Malay, Acehnese and Chamic languages (and that of Vietnamese, and, by virtue of borrowing the system from Thai, the system of the Mon-Khmer language Maleng) are decimal. Austronesian languages have rarely abandoned a decimal system to replace it with a quinary or other additive system. It is more often the case that non-Austronesian languages have borrowed higher decimal numerals from Austronesian sources. (The origins of lower numerals are generally, but not inevitably, a good guide to the genetic affinity of a language, but no known language anywhere in the world retains only higher numbers from its ancestral language but has replaced lower numnerals with loans.)

We should also recognise that the phenomenon of reversal of direction of contact has also occurred here, although the nature of Chamic contact upon Mon-Khmer languages is not as well documented. Bahnaric languages exerted influences upon Chamic languages in the early days, and we can tell the relative level of the profundity of their shaping influences and of the impact of the lexicons of these Mon-Khmer languages by examining the earlier materials in what is called Inscriptional Cham (and the later but still pre-modern Written Cham language) and the current materials in Acehnese, a language which has important testamentary value as a Chamic language whose speakers left the Austronesian-Mon-Khmer contact scene early. If the forms which contain implosives in the source languages and which consistently do so in the mainland or Indochinese Chamic languages do not do so in Acehnese, it may be that Acehnese had implosive consonants once but has since replaced them with their equivalent exploded counterparts. (For its part, Malay lacks imploded consonants, though voiceless stops are unreleased). But in other cases Thurgood (1999: 86-93) points out that sometimes original initial preglottalised consonants /ɓ ɗ/ became simple glottal stops in Acehnese (rather than becoming /b d/ [b d], as one might have expected.).

The impact of such highland Mon-Khmer languages upon an intrusive language such as Cham, and even more so upon those languages such as Haroi which were not buttressed by the possession of a national status as Cham once was and which therefore were therefore not in the sociolinguistic position to be the target of language shift or extensive second language acquisition, would have been earlier and stronger than the impact of languages of state (or at least that of languages of prestige) such as Angkorian Khmer or Vietnamese would be. This would most especially be the case if there had been a considerable degree of intermarriage between the first wave of male Cham-speakers and female speakers of Mon-Khmer languages. The latter would then acquire a form of the ancestor of Cham as the household language, which they passed onto their children. (Unfortunately we do not know if there was a gender imbalance, with male predominance, in the earlier waves of settlement which gave rise to the Cham nation.) Given that the Chams, a people intrusive to Indochina, had settled the more easily cultivable and more prosperous parts of coastal Indochina and were developing an agricultural and commercial empire there, while the indigenous peoples were living hunter-gatherer existences in the highlands, we may assume that the Chams took over the coastal area (probably not without bloodshed) and increased their numbers by absorbing many highlanders into Cham society through intermarriage, in addition to exacting tribute from many such groups. Consequently a process of language shift, with people changing dominant languages from various Bahnaric languages towards rapidly diversifying varieties of Cham, may be assumed to have taken place. The presence of some Mon-Khmer loans among the Chamic kinship terms, including some Acehnese ones, may be a sign of this. Members of smaller, less centralised groups would have been powerless to resist.

Thus as time went on the speakers of Chamic languages would have become more powerful, and once they had established their power bases the direction of linguistic influence (and that of language acquisition and shift) would have been in their favour: Chamic languages would influence the Mon-Khmer languages which had formerly influenced them, and we know that to some extent this has happened, because of the presence of Chamic loans in some Mon-Khmer languages of southern Vietnam (apparently including a few such loans in Vietnamese itself, there deriving from Cham: we should remember that the Vietnamese are later entrants to southern parts of Vietnam than the Chams are). So it is more than likely that many speakers of some less prestigious Mon-Khmer languages, who were politically subordinate to speakers of Cham during the 1300 or so years while Champa was in power, may have shifted in large numbers to dominance in Chamic languages in previous centuries.

The same question, that of early sociolinguistic practices and the patterns of bilingualism which they imposed, is touched upon by Pang (1998) in her article about the name of the Utsat. It is believed that the wave of settlement which gave rise to the Tsat-speaking community was overwhelmingly male, and that on their arrival in Hainan they intermarried with native Hlai-speaking women, who underwent language shift to a Chamic language. However, we cannot be sure whether there was more than one wave of settlers from Hainan who helped give rise to the Tsat language and the community which spoke it. And we do not know the period of time which may have elapsed between migrations if there had been more than one. Nor yet do we know whether the members of the first or any subsequent wave of settlement were Muslims when they reached Hainan or whether they adopted the religion later. (The present-day Northern Roglai, for instance, are not Muslim, and there is no evidence that they ever have been, although some concepts deriving from a knowledge of Islam have been found in the religions of certain Highland Chamic groups.) Since shari'a permits Muslim men to marry non-Muslim women (whom they are supposed to convert to Islam by the example of their devotion to the Five Pillars), there would have been no prima facie reason why an exclusively male party from the northern Chamic area could not have intermarried with local women, whom they then converted, and having produced children there is no reason why they could not then have founded a new, Chamic-speaking Islamic society in Hainan.

One area where further work is needed, although it may end up absorbing more energy than it repays in output, is that of sourcing the items of unknown origin which occur in Proto-Chamic (and to some extent in Acehnese) and in its descendant languages at several levels. Forms of unknown origin can be reconstructed to several historical levels and, as in other languages (the example of Romani springs especially to mind here) they can cast light upon other aspects of the history of a language's development. There will be some forms which reconstruct to Proto-Malayic (or rather Proto-Malayo-Chamic) and which are shared with some forms of Malay, although the literature is silent about these. Nonetheless, these provide exactly the kind of evidence for linking Chamic especially closely to Malayic which is invaluable and the most clearly illustrative sort in the case of pairs of languages which have minimal inflectional or derivational morphology (and in which whatever morphology there is is either clearly inherited from a common ancestor, or has been borrowed from a third source which can be identified).

More notable are the forms which are attested or reconstructed for Proto-Chamic, although they are not found in Malayic lects. The proportion (not to mention the number) of Proto-Chamic unsourced items is bigger than that of the Proto-Chamic forms for which a secure Mon-Khmer etymology has been found so far, and yet the proportion of the etymologically secure Mon-Khmer forms is far from negligible, and their status within Proto-Chamic is even more significant. They constitute a smaller number of elements in the basic vocabulary and a smaller proportion of that basic lexicon, though even here their number is not inconsiderable. In fact, if Thurgood has managed successfully to identify the entirety of the Mon-Khmer elements in Proto-Chamic, and if the 155 or so forms at the Proto-Chamic level which he imputes to a possible but unidentified Mon-Khmer source because of certain of their phonological features are not in fact from Mon-Khmer languages (and in fact this may be the case for the majority of such forms), then the number of etymologically as yet unsourced elements in Proto-Chamic exceeds by some way the number of elements of Malayic origin in Proto-Chamic. A list of glosses of Acehnese forms which are certainly or probably of Mon-Khmer origin is given in Table 3.

Table 3: *Glosses for words of Mon-Khmer origin in Acehnese*

Nouns:	
Body parts:	cheek; nostril; neck; stomach/guts; jaw/chin, arm, urine.
Kin terms:	nephew/niece; grandchild; old man; stranger; parents; older sister; older brother; baby; father; person; great-grandchild.
Natural phenomena:	hill; swamp; river; tree; coals/embers; noon; dawn; mountain; ditch.
Flora and fauna:	citrus; cotton; eggplant; lizard; a bear; python; bird; straw; hawk; deer; a frog; a duck; a bird's beak.
Manufactured items:	a match; a harrow; ladle; a stable; a mat; a card for a loom; rope; pillar/post; handle; bowl (<Khmer<Malay<Arabic<Farsi); ladder.
Other terms:	yard/court; top/extremity; a drop; size; dirt; a grip; fame, renown; a piece; meaning/sense; drought.
Verbs:	to yawn; to break; to sink; to peel (2 forms); to open the mouth; to climb; to arrive; to drink; to chop; to bail water; to drop; to catch; to bite (of snakes); to hunt; to fly; to graze; to pluck; to hew; to kill; to grind; to stay overnight; to stand; to scratch; to say; to dig; to wink; to hold; to see; to wash; to hit; to bum; to excrete; to rub; to return or go home (2 forms); to urinate; to be asleep; to swallow; to stir/mix; to wear; to do; to build; to come; to take; to hold; to lend/borrow; to cover; to laugh; to love; to pinch; to call/summon; to pull faces; to open; to throw away; to bend; to bore through; to open up; to cut; to button, fasten; to throw away; to loosen, let go; to dip, dye; to enter; to extinguish; to use; to grasp; to close the eyes; to let go; to get rid of; to hang; to swallow; to turn; to wrap up; to forget; to remove; to fall down.
Adjectives:	good; small; hungry; many; left side; left-handed; shallow; crooked; all; a few; empty; evil-smelling (3 terms); hot; stupid; dry; genuine/just; flickering; piercing, sharp; fine in texture; pointed; shapely; little/not many; flaming; tired; strong; submerged under water; dumb, mute.
Pronouns:	he/she; every/each; yonder; that one.
Other free grammatical forms:	at/in; don't; never; tomorrow; let X do Y; so, then; also, then; (call for a dog); more; very, extremely; ever.

On the other hand, the forms of unknown origin constitute a much smaller proportion of the core vocabulary of any Chamic language, and they are proportionally more widely situated around the periphery of the high-frequency lexicon, although they are not clearly marked out as representing some kind of special 'cultural lexicon' stratum in the way that Arabic loans into Cham are so marked. (The fact of the numerical prominence of unsourced items in many non-nominal form-classes is, however, unchanged. In this respect they pattern similarly to the borrowed Mon-Khmer elements.)

Thurgood does not go into details about the number or nature of the unsourced elements in Chamic languages after the break-up of Proto-Chamic, except indirectly. Slightly under half the later post-Proto-Chamic forms of Mon-Khmer or other origin which Thurgood provides are not given an etymology, although we know that they are common to at least one branch of Chamic and often additionally to a stray language outside that branch. What is not discussed, and in truth we would not expect this topic to be detailed much in a comparative study such as Thurgood (1999), is the number and nature of the unsourced elements that are exclusive to a single Chamic language, say to Rade, or to a closely-knit subgroup of Chamic, for instance items which are only found in Highland languages such as Rade and Jarai, or in individual languages.

What does seem to be clear, however, is that the number of such elements in the basic lexicon of each Chamic language (that is to say, the number of elements which are of unknown origin and which are exclusive to a single language) is rather small. This is what one might expect from a longitudinal lexical study of a group of languages which have only been diversifying internally for about a thousand years and which are largely exposed to the same languages in situations of close contact. (Because of the criteria which he applies to the stratification and examination of Chamic lexica, Thurgood 1999 also does not provide information about any post-Proto-Chamic loans which have come into individual Chamic languages from Mon-Khmer languages, that is, loans from local Mon-Khmer languages which are confined to a single Chamic language. In this respect their position within our knowledge base is similar to that of the forms of uncertain origin which I alluded to above. They may be numerous as a whole, but there are few of them in the basic lexicon. There has been rather little basic lexical differentiation among most Indochinese Chamic languages, even if shared forms do sometimes look different between languages.)

Incidentally, mention should be made of the special lexical registers to be found in some Chamic languages, which are characterised by vocabulary replacement and phonological disguise. (Similar registers are apparently found in neighbouring Mon-Khmer languages.) A special register is used among Chams who are engaged in gathering camphor, and in this register every normal Cham word is replaced by another word with a disguised form. Deliberate lexical change of another sort occurs among the Northern Roglai, who taboo the names of dead relatives; in any case, Roglai names are obliged not to be the same in phonological form as actual words in the Roglai language (such matters are discussed in Simons 1982).

3.1. Some aspects of 'basic lexicon' and the relationship between Malayic and Chamic: a study based on norm-referenced lexicostatistics.

The proportion of Austronesian lexicon in any Chamic language or in the sum of Chamic languages, and its role as a basic (not to say genetic) layer within the lexicon of such languages, can be seen from an analysis of the Blust lists in Malayic lects and in Chamic languages. I have used Malayic data from Blust (1988) and Chamic data from Thurgood (1999; I refer to his Proto-Chamic work), Moussay (1971; this documented Phan Rang Cham) and Collins (1969; this source provided data from Acehnese and Northern Roglai), only a fraction of which is cited or presented here. (A more finely-tuned examination of Chamic Blust lists is given in Grant 2005.)

The topic is large; given space constraints, my field of concern was narrowed to the issue of the linguistic position of those elements occurring on the Malayic Blust lists which are not inherited from Proto-Malayo-Polynesian (or which at least cannot be traced back to it), and the means with which the concepts which they encode are expressed in Chamic languages. About 80 forms, or 40% of the glosses, are affected, and I compared the Malay reflexes of these affected glosses with the realisations in Proto-Chamic where this was possible, and with data in Northern Roglai and Phan Rang Cham in the few cases where comparison with Proto-Chamic was impossible because of the lack of a form in the latter language.

In this study I was especially interested to see how many Malayic-Chamic shared innovations there were on such lists. Blust (1988) provided equivalents for the 200 items on his list (originally evolved in 1967) for eight Malayic lects, namely Standard Malay, Deli Malay of Medan, Iban, Minangkabau, Salako, Banjarese, Jakarta Malay or Jakartanese (Betawi), and Ambonese Malay (Bahasa Ambon), and in only one case, that of Salako, were more than two items missing from the list (27 of the 200 forms were missing from the Salako list; Iban and Minangkabau are often regarded as languages which are separate both from one another and from Malay). I compared these forms with the reconstructed Proto-Malayo-Polynesian forms presented in Blust (1993), and with the Proto-Chamic forms presented by Thurgood, with a sideways glance (but no more than that) at the Northern Roglai and Cham datasets which were alluded to above. It appears that at least in terms of basic lexicon the especially conservative Malayic dialects here, when compared with Standard Malay, are Iban, Selako and also Banjarese, each of which includes some forms which are traceable back to Proto-Malayo-Polynesian but which have been replaced in most Malayic lects (including Standard Malay) by internal innovations, often of uncertain origin. (Some further dialectal Malay 200-item lists can be found in Adelaar 1992, and yet others are available elsewhere. Taken together, these constitute a fine basis for Malayic dialectal classification, especially so since the amount of inflectional morphology available for reconstruction within Proto-Malay is slight.)

In addition it is necessary to take the differing patterns of diffusion into account before checking for possible shared innovations. Sometimes individual Malay lects have borrowed an item from another language as a means of expressing a concept, while Chamic seems to be conservative and to use an inherited term, and vice versa. We should note that the various Malay 200-item lists which Blust has provided include in their contents not only elements inherited from Proto-Malayo-Polynesian (or indeed from Proto-Austronesian) and some Malay-internal innovations, but that there are diffused elements which have been borrowed from Mon-Khmer, Sanskrit, Tamil, Arabic, Batak, and (in Jakartanese alone) also from Javanese and Hokkien Chinese, and (in Ambonese Malay alone) also from Portuguese.

For their part the equivalent Chamic lists examined (mainly Western Cham, Jarai, Rade, Northern Roglai and Tsat) include older and more recent elements drawn from Mon-Khmer languages of various branches, a few other elements from Sanskrit, and numerous ones from as yet unidentified sources, plus (probably) some in Acehnese which derive from Malay. The diachronically most interesting forms are those few which occur exclusively in Proto-Chamic and in Proto-Malayic.

This examination of Malay and Cham forms is a study in norm-referenced lexicostatistics, because the items on the lists are each being separately compared with those from a predetermined dataset (the reconstructed Proto-Malayo-Polynesian forms in Blust 1993) which has been chosen because of its diachronic significance, and which serves as the norm. In this respect the approach differs from the pair-referenced lexicostatistics which underpin the Austronesian classification in Dyen (1965), and which involves pairs of languages being compared with one another, without each of them being compared to a standard. (Dyen's failure to do this – his failure for instance to compare the glosses on the test-list for individual Austronesian languages with the reflexes for these words in Proto-Malayo-Polynesian inasmuch as they are provided in the works of Otto Dempwolff – is an important factor in Dyen's ambitious, dramatic and methodologically erroneous reconstruction of 40 separate sub-branches, each supposed to be of equal epistemological status, which subtend from Proto-Austronesian. Had Dyen used norm-referenced lexicostatistics instead of relying solely on inferences from results from cross-comparisons of living Austronesian languages, the resulting picture of interrelationships within the family which he developed from such an analysis would have been very different and much more sharply nuanced, and it would probably have prevented him from coming to his odd conclusions about the cradle of Austronesian being in New Guinea.)

The issues at hand here can be illustrated by an example from the Philippines. Zorc (1974) demonstrated the importance of evaluating and classifying the various kinds of shared similarities which two languages exhibit, with his examination of the strata in the basic lexicon of Kagayanen. This is a Manobo language (and thus an Austronesian one) of the Western Philippines whose speakers moved there from their original home among other Manobo-speakers in Mindanao, and who have borrowed a large amount of core vocabulary from Hiligaynon and other Bisayan languages. The largest element of the basic lexicon consisted of items which were common to most or all Philippine languages and certainly both to Manobo and Bisayan languages. Several other forms on the 100-word Swadesh list which Zorc used did not have a certain etymology, so that the material to be used to determine the closest affinity of Kagayanen was contained in the remaining items of lexicon. And an examination of this, combined with the analysis of some non-linguistic features relating to the geographical location of the Kagayanen-speakers, demonstrated that its true affinities were with Manobo languages, despite its location in the midst of Bisayan languages.

A similar situation arises with the examination of basic vocabulary in Malayic and Chamic varieties. Both languages possess lexical elements which have been taken from (or which have been inherited from) the same sources – Austronesian and its subgroups, Mon-Khmer and Sanskrit, and latterly also Arabic and Chinese languages. So there are both inherited elements and loan elements which are common to the two sets of languages. But this does not mean that the same lexical elements are going to be found in both languages: if a particular Sanskrit word is found in Malay, it may or may not also be found in Cham. We have firstly to identify and secondarily separate out the various kinds of lexical commonalities, expunge or set aside from these those elements which are clearly loans, and examine what remains. Some of the commonalities which we come across will be shared inherited elements, items which will perforce be found in languages outside those which we are examining. Some will be shared innovations, which may or may not indicate a special relationship between Malay and Chamic. There may also be elements which are clearly loans from a third language but which nonetheless reconstruct back to the period when Malay and Chamic were one languages, and there will be later loans which are found in both languages but which have been borrowed separately by the two languages. (This latter category applies to the Arabic adstrate in Cham, since most of the Chams, especially the more westerly ones, embraced Islam in the early centuries of the second millennium AD, maybe more than 1000 years after Proto-Chamic had split from Proto-Malayic. Nevertheless some of the coastal Chams may have embraced Islam at the same time as the Malays or as a result of contact, in Champa or beyond, with Muslim Malay traders. Additionally Malay has served as the language of Islamic learning among the Chams.)

Every stratum of the vocabulary of Cham (or of any other language, for that matter) has its own significance within the history of a language. This is true whether the stratum in question serves as an attestation of the ultimate genetic origin of a language, or as a sign that this language is a sister-language of others of the same ultimate origin. But it is also true if it is the case that this speech community has had social interactions at various levels with speakers of other languages, or even that the language has (for sociopolitical or other reasons) been insulated from being influenced as the result of contact with other languages, and that it has consequently expanded its resources through the use of extensive internally-driven innovations. Periods of intense internal innovation, and the new morphs which result from this, are attestations to periods when the effects of linguistic contact did not disturb the social peace of a particular group. But we should remember that the full effects of linguistic contact may take centuries to be bedded into a language. For example Old Norse was more or less extinct in England by the time most of the Norse elements that replaced original Old English elements came into general use, even if the forms themselves had been taken over as synonyms or whatever some centuries earlier[3]. Norse forms came into standard English largely through the influence of non-standard varieties whose speakers fled south after depredations under William I. And unlike the prestige position of Norman French (about which similar chronological remarks to those about Norse may justly be made) there was by that time no Norse cultural 'support system' to enable the continued borrowing and propagation of Norse elements in English, once Norse lacked native speakers in England.

An examination of the Blust list data for Malay and for Chamic languages reveals the following details. According to Blust (1988: 15), Standard Malay has 112 elements on the 200-word Blust list which are directly inherited from Proto-Malayo-Polynesian and which require no further comment in this case. (A few more items of Proto-Malayo-Polynesian origin have been retained in non-standard Malay dialects, and are exemplified as such in the lists which Blust provides, but they do not remain in the standard language, and a few further ones are only retained in certain fixed expressions in Malay.) The number of clear loans from other languages in the Malay list (indeed, the number of loans on the lists for any of the eight dialects which Blust provides) is rather small. There are 18 loans on the Standard Malay version of the 200-item list: three from Arabic, one each from Tamil (actually a loanblend) and Batak, and the rest from Indic.

Some Malay lects include a greater number of borrowed elements on their Blust lists than others (Jakartanese and Ambonese, with forms for 'you singular' that have been borrowed from Hokkien and Portuguese, *lu* and *ose* respectively, spring to mind.). Conversely some loans are common to all eight Malayic lects (the Sanskrit-derived *kepala* 'head' is a good example of this, but the Proto-Austronesian-derived *hulu* is also in use in Malay in figurative senses. This form is replaced by an element of Mon-Khmer origin in Cham (*ako'*), although the Cham form *dihlôw* 'at first, formerly' incorporates the Austronesian stem; compare Malay *di-hulu, dulu* 'at the start', literally 'at-head'). The relative lexical conservatism of Iban, Selako and Banjarese has been mentioned above.

In contrast, the number of loans on the Chamic list (and this statement is intended to apply to Proto-Chamic but is in fact true of any Chamic list, including that for Acehnese) is much higher, maybe four times as high. Most of these are assumed to be derived from Mon-Khmer languages. Yet it is true that all but one of the Sanskritisms which occur in the modern Cham version of the Blust list ('to smell', if this is indeed an Indic form and not Mon-Khmer in origin, one of the forms for 'person/human being', and the word for 'seed') are also found and used in Malay (where they appear as *cium*, *manusia*, *biji*). Only *dhul* 'dust', a Sanskritism used in several Chamic languages and also

[3] That old chestnut from History of English classes, Caxton's story about the mercer Sheffield asking for *egges* at a shop on the Thames estuary when the local word for eggs was *eyren*, springs immediately to mind.

in Khmer, is missing from the Malay lexicon. By comparison, Jarai has retained 82 items out of the PMP 200 reconstructed forms on the Blust list (Robert Blust, personal communication, December 2001.)

In addition, some lexical forms which occur in Malay but which cannot be reconstructed as far back as to Proto-Malayo-Polynesian can be found among the pre-Mon-Khmer elements of Chamic languages (which are mostly listed in Thurgood 1999: 280-308). In the absence of evidence from instances of innovated bound morphology, lexical forms such as those constitute the best evidence for the existence of an ancestral language from which both the various Malay lects (and also Iban, Minangkabau, etc.) and the Chamic languages are descended.

We find the following forms in Malayic and Chamic as shared lexical innovations on the Blust list, for which I have here provided the Malay equivalents: 'rat' (*tikus*), 'to sit' (*dudok*), 'and, with' (*dan*, *dengan*), 'tooth' (*gigi*, a PAn form for 'barb' that is also found with a changed meaning as Madurese *ghighi* 'tooth'), 'green' (*hijau*) and the older Malayo-Chamic word for 'person, human being' (*orang*). These did not occur in PMP as far as we can tell, but are innovations of a later period. Most of these can also be found in the Acehnese and Northern Roglai lists provided in Collins (1969).

Chamic and Malay also share the semantic shift of PMP **malem* 'afternoon, evening' to 'night', though this development is a crosslinguistically common one, and it could have occurred independently in the two groups. Chamic preserves the PMP word for 'to cook', the form of which in Malay means 'to staunch blood, to act as a styptic' (of the forms are the same in origin, then presumably they are linked by the concept of cauterisation of wounds). The inherited Chamic form meaning 'sea', as it did in PMP, has shifted to meaning 'saltwater' in Malay (where the form is *tasi*), which has innovated another word for 'sea' (*laut*) from a word which was originally a directional term meaning 'towards the sea'. Metatheses, and a number of forms which amalgamate two or more earlier morphs into one synchronically unanalysable form, and which suggest a period of shared development, are common to Chamic and Malay in the case of 'to drink' (Malay has *minum*; compare the Tagalog stem *inom*), but in contrast to Malay, the Proto-Chamic form for 'tongue', *dilah*, and 'to live' (Proto-Chamic *hudip*), are phonologically conservative. These words have not undergone the metatheses found in Malay *lidah* and *hidup*, forms shared by all the Malay dialects in Blust's lexical sample and (as loans) also in some languages now used in Indonesia.

There are very few instances, on the Blust list or elsewhere, of Proto-Malayo-Polynesian forms which continue to be employed in Chamic while being replaced by loans or other forms in Malayic (although some other inherited forms have retained their original meaning in Chamic but have shifted their primary sense in Malayic). The items on the 200-item list which fall into this category are as follows:

'three' (Chamic languages preserve reflexes of PMP **telu*, as do most other Western Malayo-Polynesian languages, though this has been almost completely replaced in Malay, and also in Iban, by the Middle Indic form *tiga*),

'shoulder' (PMP **qabarah* is preserved in Chamic as *bara*, with predictable loss of the first syllable's laryngeal plus accompanying vowel, according to a Malayo-Chamic rule, but this form is replaced in Malay by a loan from Sanskrit),

'name' (Malay has replaced this form with a Sanskrit loan *nama*, although Jakartanese Malay uses a form *ngaran* which is borrowed from Ngoko Javanese, where in turn it is inherited from Proto-Malayo-Polynesian, while Chamic is conservative), 'mouth' (Malay has replaced PMP **baqbaq*, which it has lost, by the innovation *mulut* but Chamic has preserved the Proto-Malayo-Polynesian form), and

'to go, to walk' (Cham preserves a reflex of Proto-Malayo-Polynesian **panaw* but Malay has not done so, instead verbalising the noun *jalan* 'path', a form of Proto-Austronesian vintage, as *berjalan*; Malay preserves Proto-Malayo-Polynesian **lakaw* 'to walk' as *laku* 'behaviour', though this verb is not preserved in Chamic, while the widespread Chamic verb *laba:t* 'to go' corresponds to Malay *lewat* 'over, past', a form that appears to have undergoone grammaticalisation).

The taboo word for 'dog' has replaced the older form; 'dog' is nowadays expressed in Standard Malay by *anjing* (except in the phrase *gigi asu* 'canine tooth'; it is also preserved as the common word for 'dog' in Iban and Selako), but Cham preserves the reflex of PMP **qasu*.

We can see two different trends of lexical change at work here. There is one in which an original form has been replaced by a loan in Malayic or Chamic. In the other an original form has dropped out and is replaced in one set of languages but not in the others by an innovation which dates from the period after Malay and Cham separated.

There are more forms of Proto-Malayo-Polynesian vintage in the Malayic lists than there are in the Chamic lists, as the latter include several elements which were innovated at the Proto-Malayo-Chamic level, a stratum which I have excluded from my count of the 120 Proto-Malayo-Polynesian elements on the list which are attested for Malayic and which are mentioned above. Some 40 items (at least) on the 200-item list for Proto-Chamic derive from Mon-Khmer languages (or at least they may be claimed as possible Mon-Khmer elements because of some phonological characteristics which they possess), and 11 forms are of unknown origin in the current state of knowledge but are still common at least to most or all of the Indochinese Chamic languages.

A comparative count of the Blust list forms in Proto-Chamic and in Standard Malay shows that the two languages have 85 items in common out of 200. This total is exclusive of commonly-shared loans from a third party (in this case from Sanskrit), of forms which have undergone a semantic shift in one of the languages, which has resulted in giving the form a meaning which does not correspond to one found on the Blust list (although the same form in the other language retains a Blust list meaning), and of items which have been borrowed from another Austronesian language in one language (for instance the Malay borrowing of a form meaning 'yellow', *kuning*, from the Batak word for 'turmeric', where **kunik* or **kunit* would have been expected had the term been inherited from Proto-Malayo-Chamic) but which are directly inherited in the other (for instance Phan Rang Cham has *kunit* 'yellow'). But this number of shared forms includes the small number of lexical innovations which are not found in other Western Austronesian languages - for instance they are absent from Tagalog - and which are characteristic of, or are confined to, Malayic and Chamic languages (but they are words which secondarily may have been transmitted to languages which have borrowed such terms from these languages).

The task of reconstructing the phonological and other paths of development which distinguish Proto-Malayic from Proto-Chamic and those which distinguish Proto-Malayo-Chamic from other subgroups within the amorphous construct that is Western-Malayo-Polynesian has yet to be carried out fully. It is significant that Proto-Malayic and Proto-Chamic have identical, regular and non-trivial reflexes for several diagnostic sounds or groups of sounds, such as **Z, *R, *c, *q, *ñ, *w-, *qVC-, *hVC-, b-*, which are realised both in Proto-Chamic and in Proto-Malayic as as*j, r, c, h, ñ, Ø-, C-, C-, b-* in both (while the last sound becomes *w-* in Javanese), while both Proto-Malayic and at least the earliest stages of Proto-Chamic kept all four Proto-Austronesian vowels, including schwa, intact and distinct. Locating such features, more than tracking down shared lexical commonalities, is the first step to proving the existence of an exclusive subgrouping between Chamic and Malayic.

4. Languages in contact: the Chamic languages as mixed languages? Combining, integrating and productive continuation of elements of diverse sources.

The histories of the Chamic languages, including Acehnese, are good examples of the importance to diachronists of separating out and thereby understanding the various complexities of the results of language contact, especially the facets of contact-induced language change. The results which are obtained from an examination of the documentation of the remarkable developments which they have undergone through the effects of contact-induced change also underline the importance of applying both the philological method and the evidence of whatever data sources are available to us. (And this is not just so in the case of Chamic.) All of these are things which we do in an attempt better to understand the historical developments of these languages. Once this preliminary spadework has been done we may build up a nuanced picture of the consequences of linguistic contact. We do not know everything that we would hope to know about this linguistic scenario (or rather, this chronological series of scenarios), and we probably never will. But we can find out a surprisingly large amount from the information available to us.

Thurgood is exactly right in suggesting (Thurgood 1999: 251-259) that external influences have shaped the Chamic languages so significantly, causing them to become the way they are now, and that they have done this to a much greater degree than internal influences have. The amount of change through externally-induced contact which they have undergone is impressive, especially in the case of Tsat. In terms of the impact of external contact Chamic languages belong to levels 4 and 5, the highest points on Thomason and Kaufman's five-point scale (Thomason and Kaufman 1988: 74-76). Different parts of the structure and lexicon of the Chamic languages rate being posited on different levels of the Thomason-Kaufman scale of contact, however, and in addition some Chamic languages have been influenced more directly through contact with specific languages in certain respects than others have been.

Many of the changes which we find in Rade seem to be internally-driven and without a clear parallel in Mon-Khmer languages which surrounded and which might have influenced Rade, whereas most of the changes in Haroi in the past 500 years appear to be the results of Haroi dominance by speakers of Bahnar. We may note especially the relevance of Thomason and Kaufman's level 5 for the nature and depth of Hainanese Chinese contact with Tsat. This is a level which is especially and extremely clear when one examines the patterns, canons, features and segments of Tsat phonology, which have come more and more to resemble those of Hainanese Chinese.

But we should not shrink from admitting the existence of some logistical problems in applying the Thomason-Kaufman scale to languages which are without much visible morphology, since so many of the features which these authors discuss in their scale relate to the stepwise transferral of morphological elements. And while Chamic languages have certainly done some of this transferral, and while they have also through time lost some of the sparse morphology which they originally had, the degree of high morphological density has never been as strong in languages deriving from Proto-Malayic as it has for (say) the native languages of the Philippines. What is more, we should not underplay the role, quantitative and qualitative, in the various Chamic lexica of elements which may be borrowed from Mon-Khmer or wherever but whose origins are as yet shrouded in uncertainty.

Other questions may be asked about the scale, especially in relation to the implied order or concomitance of adoption of some of the transferred features. It is a false assumption that the taking over of features in one stratum of a language necessarily implies the simultaneous or contemporaneous taking over of features in another part of a language's structure at the same level. To manufacture an example, one may say that the borrowing of an adjectival comparison marker from a donor language into a recipient language does not imply or actuate the borrowing of (let us say) rules for the palatalisation of velar consonants from the same donor language at the same time. Nor does it imply that other features of the recipient language's adjectival morphosyntax will also be modified in the direction of those of the donor language. (For example, Urdu borrowed the free-standing morph *zya:da:* to express adjectival comparison from Farsi, but it did not abandon its marking of adjectival number and gender concord within such comparative constructions, even though Farsi adjectives are invariable in form and Farsi has no grammatical gender and does not mark plurality in attributive adjectives.)

There is also the question of the grading of some of the phenomena in relation to one another on the Thomason-Kaufman scale. From the perspective of a crosslinguistic examination of natures and states of borrowing, some items (for example certain kinds of conjunctions) seem to be placed too high on the scale, and some others (borrowed basic vocabulary which has come into replacive use in a language through partial relexification, for instance) seem to have been placed too low. The large-scale borrowing of subordinating conjunctions often occurs in languages which have undergone a greater use of hypotactic constructions in subordinate clauses (and a greater use of such clauses) than their uninfluenced relatives use.

Typological questions of systematic congruity come into play here too. The collocation of structural facts, namely that Mon-Khmer languages and Malayic languages have the same form-classes of polymorphemic words (and that they have many similar kinds of free grammatical morphs, and additionally that their bound morphology is rather sparse in any case) may have more significance than we had previously realised, as a fuller contact history of Chamic languages might show. It does seem to have made the borrowing of 'unborrowable' items such as verbs more easy.

But we have not written more than a fragment of the linguistic histories of any of these languages. For example, we have said nothing substantive about the morphosyntax (typological as well as formal) of the Chamic languages and the ways in which these structural systems may have been affected by contact with (or by any previous typological or structural similarity to the structure of) Mon-Khmer languages. For this reason, and in attempt to start filling this gap, some broad-brush typological comparisons (including details of verb phrases structure) involving Cham, Malay, Tagalog, Chrau, Khmer and Vietnamese are presented in Table 4.

I have used Chrau structural data from Thomas (1971) as an example of the structural features of a South Bahnaric language of the kind with which many Chamic languages were in close contact. I used Western Cham data from Baumgartner (1998) as a sample of Chamic structural data because this is the non-Acehnese Chamic variety for which I had the greatest amount of structural information at the time.

The more salient structural similarities which have been found between Cham and Mon-Khmer languages are italicised (NegC = negative plus circumfix; MC = main clause, Cl – numeral classifier or measure word; X = the feature is missing).

Table 4A: *Some typological features of morphosyntax in Western or Cambodian Cham (Baumgartner 1998, where attested) and other relevant South East Asian languages (Malay from Hamilton 1997; Chrau from Thomas 1971, Cambodian from Jacob 1966, Vietnamese from Đình-Hoà 1997).*

FEATURE	CHAM	MALAY	CHRAU	KHMER	VIETN.
Element order	SVO	SVO	SVO	SVO	SVO
NG	NG	NG	NG	NG	NG
PossN	N Poss	N Poss	N Poss	N Poss	NGen Person
NA*	*N A*	A N	*N A*	*N A*	*N A*
NNum	*N Num Cl*	Num Cl N	Num Cl N	*N Num Cl*	Num Cl N
NDef	X	X/N Def	X	N Def	X
NIndef	N Indef	X	X	N Indef	'one'/zero
NDet	N Det	N Det	N Det	N Det	N Cl Det
AdposN	Prep N	Prep N	Prep(Prep)N	Prep N	Prep N
NegN	?	Neg N	Neg N	Neg N	Neg N
NegAdj	?	Neg Adj	Neg Adj	Neg Adj	Neg Adj
NegV	*V Neg (C)*	Neg V	Neg V/*Neg C*	*V Neg(C)*	Neg V; *V Neg*
TMAVerb	TMA Verb	TMA Verb	TMA Verb	TMAVbTMA	TMA Verb
AdjModifier	Adj Mod	Adj Mod	Adj Mod	Adj Mod	Adj Mod
AdjCompar.	?[4]	Compar Adj	?	Adj Compar	Adj Compar
AdjSuperl	?	Superl Adj	?	Superl Adj	Superl Adj
CopulPredic.	Cop Pred	Cop Pred	X – no copula	Cop Pred	Cop Pred
Subrd-Main cl.	Subd MCl	Subd MCl	Subd MCl	Subd MCl	Subd MCl
Copula?	<'stand'	absent	absent	yes	yes
Cop=Loc**?	yes	no	no	no	no
Cop = 'have'?	no, separate	loc='have'	no	no	no
Existent=have	no	yes	yes	yes	no?
Tes/No QMC.	QMC	MCQ	MCQ	?	?
QInversion	no	repetition	no	no	no
4B: SOME BROADER TYPOLOGICAL FEATURES					
Pro-drop?	No	Yes	No	No	No
NPluralisation	(Pl particle) N	zero	no	no	Plur Noun
Case-marking	none	none	no	no	no
Inflections?	none	none	no	no	no
Bound deriv?	Slight	yes	some	some	not now
Numerals	dec-subtr****	dec-subtr	decimalquinary	decimal	
Num classif?	Yes	yes	yes	yes	yes
Prefixes?	Some	some	some	some	no
Infixes?	*Some*	no	*some*	*some*	no
Suffixes?	*No*	some	*no*	*no*	*no*

[4] In present-day Eastern Cham such a form is expressed by *hon* (from Vietnamese) plus the adjectives (Alieva 1999).

4C: STRUCTURE OF THE BASIC VERB PHRASE AS A TWO-PLACE PREDICATE:

Eastern Cham:	*Subj (TMA) Verb Obj*
Malay:	(tma) Subj (TMA/Modal) (Prefix) Verb Obj
Tagalog:	(TMA) (Voice) Verb Subj (Object Marker) Obj
Chrau:	*Subj (Preverb) (TMA) (Auxiliary) Verb Obj*
Khmer:	*Subj (TMA) Verb Obj*
Vietnamese:	*Subj (TMA) Verb Obj*

4D: SOME PHONOLOGICAL FEATURES

No. of vowels	10	(4>)6	11 long,7 short	10	2 short, 9 long
Vowel length?	No	no	no	no	yes
High cent v.?	*yes*	no	*yes*	*yes*	*yes*
Nasal vowels?	No*****	no	no	no	no
Voiced stops?	*Yes>no*	yes	yes	*yes>no*	yes
Implosives?	*Yes*	no	*yes*	*yes*	*yes*
Final sibilant?	*(yes>) no*	yes	*no*	yes	*no*
/-s/ > /-ih/?	*Yes*	no	*yes*	*no > yes*	no
Final palatals?	*Yes*	no	*yes*	*yes*	*yes*
/ng-/ present	only loans?	yes	yes	no	yes
/n-/> /l-/?	Yes	mostly	no	no	no
/ñ-/?	Yes	yes	yes	yes	yes
CC-?	*Yes*	no (>yes)	*yes*	*yes*	*yes*
Tone system?	No***	no	no	no	5, 6
Registers?	(Yes>)No	no	yes	yes	no (< yes)
Stress	final	varies/penult	final	final	final
Stems 1-syll?	*Often*	no	*yes*	*generally*	*yes*

NOTES:

* the form for 'my three big houses' in Western Cham is expressed as 'house 1sg three big CLASSIFIER' (Baumgartner 1998: 15).

** 'Loc' = this is the locative 'to be' verb, that is 'to be at' as distinct from the copula.

*** Although Western Cham lacks tones, Phan Rang Cham has three or four tones, which have developed from a two-tone system which itself developed from a registral system (Thurgood 1996). Furthermore, Western Cham has preserved /s-/ in cases where Phan Rang Cham has shifted to /th-/ in imitation of a similar phonological change which is documented for Vietnamese.

**** the basic numeral system in Cham and in Malay is essentially decimal, but it is one in which the earlier Austronesian form for '7' has been replaced in both languages by a form deriving from the name for the index finger, while the forms for '8' and '9' gave been replaced by subtractive constructions, the same ones being used in both Malay and Cham.

***** nasalised vowels are not found in either variety of Cham but are attested in abundance for Haroi and Northern Roglai (where they have developed under separate circumstances in each language).

Most of these languages share the strong areal characteristic of a paucity of bound inflectional morphology (and of the possession of few productive bound derivational morphs). Since, according to Ludolf's Law, the morphology of a language is to be taken as a better guide to the genetic affinity of a language than the lexicon is, the task of demonstrating genetic affinity among South East Asian languages is made much harder.

This is especially so in a region where the practice of borrowing and subsequent productive use of free grammatical morphs from one language to another (even of those which relate to a tense-aspect system) is far from being unknown.

The Chamic languages have quite a few free grammatical morphs, next to no inflectional morphology, and rather little bound derivational morphology, and some of the latter derives from Mon-Khmer sources (as pointed out in Thurgood 1999: 237-250). Other morphological processes are encoded by the use of free grammatical morphemes, and many such processes which Western observers take for granted (such as subject-verb agreement, noun-adjective concord, often also tense or aspect marking in the verb phrase, or the presence of case-systems in nouns) are not marked at all. By comparison, Tagalog, another Western Malayo-Polynesian language, has abundant bound inflectional and derivational morphs (see the discussion in the relevant section of Table 4). It should be understood that in this respect Malay has innovated over the past two millennia, in that it has discarded many inflections while Tagalog, Malagasy, Toba Batak and several other major Western Malayo-Polynesian languages are conservative in this respect (exhibiting a conservatism which is reflected by the occurrence of these affixes in many of the Formosan languages), and these conservative morphological structures more closely represent the state of affairs in Proto-Malayo-Polynesian.

A considerable number of the features listed in Table 5, involving phonological, morphological and syntactic differences, differ in their patterning or structure between Tagalog and Malay on the one hand and Cham on the other (in which instances the Cham features usually parallel those of Chrau or Khmer). Several more are similar in construction in Malay and Cham, where they represent shared South East Asian areal features, but they are realised differently in Tagalog. For a few features I had no information about the mode of their realisation in Western Cham. The only features among those listed which seem to show the retention in Cham and Tagalog of any morphological features which have been lost in Malay relate to the presence in both languages of infixes (which are retained in Cham, although the most productive infix in Cham is loaned from Mon-Khmer). There is also a negative feature (and therefore one that is useless for subgrouping!) which is shared between Tagalog and Cham, namely the disinclination to use pro-drop, this being something which Malay also employs.

The main reason for this discrepancy between the occurrences or otherwise of these features in what are all Western Malayo-Polynesian languages is an areal one. Malay has not been integrated into the South East Asian Sprachbund (partially outlined and mapped in Henderson 1965, and discussed in much more detail in Alieva 1984 and 1992, which draw in part upon Alieva's work on Phan Rang Cham) as strongly as Cham has. But Malay is still more of a part of this network of areal phenomena than Tagalog is. For example Malay and Cham have both developed numeral classifiers (also known as numeral coefficients, or measure words), a form-class of items which are typical of a range of East Asian languages from Mandarin to Khmer, but which are not found in Tagalog and which are not reconstructible, either as a form class or in terms of individual forms, for Proto-Malayo-Polynesian. Cham, like Malay and like other Chamic languages, uses some classifiers which also have a full lexical meaning in the language, while other classifiers have no separate existence in the lexicon of the respective languages. And, just as Malay has done with *biji* (with its meanings of 'seed' and its role as a classifier for small grain-like objects, a word which has the status of a loan from Sanskrit into both Malay and Cham, and which exists in both Malay and Cham as both classifier and full lexical item), it has borrowed the words which are in use for some classifiers from other languages (in the case of Cham, though, these come mostly from Mon-Khmer ones).

TABLE 5: *Structural differences between Western Cham, Malay and Tagalog (the last representing a more structurally conservative form of Western Malayo-Polynesian): a typological survey.*

FEATURE	CHAM	MALAY	TAGALOG
Element order	S V O	S V O	V S O
NG	N G	N G	N Lig G
NA	N A	A N	A Lig N
NNum	N Num Cl	Num Cl N	Num N
NDef	X	X/N Def	Def N
NIndef	N Indef	X	X
NDet	N Det	N Det	Det N
NegN	?	Neg N	Neg N
NegAdj	?	Neg Adj	Neg Adj
NegV	V Neg	Neg V	Neg V
AdjModifier	Adj Mod	Adj Mod	Modif Adj
AdjCompar.	?	Compar Adj	Compar. Adj
AdjSuperl	?	Superl Adj	Superl-Adj
Copula?	<’stand’	absent	late development
Cop=Loc?	yes	no	No
Cop = ‘have’?	no, separate	loc=’have’	no
Existent=have	no	yes	yes
Tes/NoQMC.	Q MC	MC Q	MC Q
Pro-drop?	No	Yes	No
NPluralisation	(Pl particle) N	zero	Pl-particle N
Case-marking	none	none	yes
Inflections?	none	none	yes
Bound deriv?	Slight	yes	yes
Numerals	dec-subtr	dec-subtr	decimal
Num classif?	Yes	yes	no
Prefixes?	Some	some	yes
Infixes?	Some	(relics)	yes
Suffixes?	No	some	yes
No. of vowels	10	(4>)6	(3>)5
Vowel length?	no	no	tied in with stress
High cent v.?	yes	no	no
Voiced stops?	Yes>no	yes	yes
Implosives?	Yes	no	no
Final sibilant?	(yes>) no	yes	yes
/-s/ > /-ih/?	Yes	no	no
Final palatals?	Yes	no	no
CC-?	Yes	no (>yes)	via loans
/ng-/	only in loans	yes	yes
/n-/ > /l-/?	Yes	mostly	no
/ñ-/?	Yes	yes	no
Tone system?	No	no	no

Registers?	(Yes>)No	no	no
Stress	final	varies/penult	varies
Stress phonemic	no	no	yes
Stems 1-syll?	Often	no	no

In Table 4 I have italicised those features in Western Cham morphosyntax which show parallels with forms in non-Austronesian languages but which are not areal features throughout South East Asia, to the extent that they have no diagnostic significance, whether or not these are found in some other Austronesian language. It will be seen that Malay has acquired fewer South East Asian areal featuires than Cham has, although the number in Malay is significant. Some of these areal similarities may be the secondary consequence of the acquisition of other areal features at a previous stage in the languages' histories. A particularly significant case is that of 'basic word order' in Malay, Cham and Tagalog. Tagalog preserves the general verb-initial pattern which is thought to be typical of Proto-Austronesian and Proto-Malayo-Polynesian, and Tagalog has also preserved a case system which operates in tandem with the (inherited and elaborated) focus system and which allows one to distinguish morphologically between agents and patients even when the noun phrases or pronominal phrases containing them are adjacent in the sentence. Malay has lost such morphological features, as has Cham, and in both these languages the basic order is SVO, with the verb sandwiched between the (pro)nominal phrases.

Most of the structural or typological differences between Tagalog and Cham represent one of two things. Either they are losses on the part of Cham as against retentions from Proto-Malayo-Polynesian in Tagalog, or else they point to the Sprachbund-driven absorption of features into Cham which were never taken into Tagalog. Two exceptions to this trend are noteworthy: first of all, the preservation in Cham and Malay of a rare initial palatal nasal consonant /ñ-/ which has been replaced by /n-/ in Tagalog is an example of the rare conservatism of Cham as against Tagalog. Furthermore, the use in Tagalog of a free-standing pre-adjectival form derived from Spanish as the usual means of expressing the comparative degree with adjectives is a rare example of a structural-typological feature in Tagalog which is loaned from another, non-Austronesian language (though superlation in Tagalog is expressed with a verbal prefixal complex *pinaka-*, a prefix with an infix embedded in it, whose elements are of Austronesian vintage).

The Chamic languages must be some among the very few in the world which have productively borrowed some infixes from other sources; the main nominalising infix *–ən-* ~ *-an-*, which is productive in Chamic languages, is a Mon-Khmer infix which is of Proto-Chamic vintage (Thurgood 1999: 308; Blust 2000 demurs and sees the form as being equally likely to be of Austronesian origin). But it does somewhat resemble in form an Austronesian infix *-in-*, a voice and focus marker which sometimes has similar nominalising uses to the borrowed Mon-Khmer infix.

And we should not forget the possibility that the numerous and remarkable contact-induced changes have overshadowed the various internally-driven and internally-induced changes which the Chamic languages have undergone. (Not all change in Chamic languages has been externally-actuated, although parallel influence from Mon-Khmer languages of power continues and can extend to fairly minor changes which are shared with the dominant language.

For instance the replacement of /s-/ in Phan Rang Cham by the aspirated stop /th-/ (rather than by the voiceless interdental fricative which one might have expected on more universalist phonetic grounds) might at first seem to be independent of any developments in the phonological histories of Khmer and Vietnamese. Yet further investigation and use of comparative evidence shows that something similar, indeed an identical change, has happened syllable-initially in the relevant morphs in Vietnamese. (Similarly, in extremely allegro forms in Phan Rang Cham, a former Cham /ph-/ has become /f-/, just as it has done in Vietnamese: Blood 1962: 11; we note the allegro Phan Rang Cham form *frèw* 'new' (or the less allegro form *pihrèw*), from Proto-Malayo-Polynesian **baqeru.*) This development can be seen more clearly when the Vietnamese forms are compared with their cognate forms in Katuic languages, which have been controversially suggested as being the languages that are most closely related to the Vietic subgroup (Diffloth 1991). The change from /s-/ to the aspirated stop /th-/, incidentally, has not taken place in Western Cham, probably because this phonological change has not occurred in Khmer.

The change from the earlier /-l/ to /-n/ in more modern forms of Phan Rang Cham, the 'Cham sonorant problem' which was discussed by Blood (1962), is a problem which is diachronic and sociolinguistic in its terms of reference more than anything. The speech of older men who have had a traditioal education in written Cham retained the distinction whereas the speech of younger men and of women who had not received this education lacked it and used only /-n/. But the impetus for this change has much to do with the fact that /-l/ is impermissible in the dominant Vietnamese, while /-n/ is allowed. (But original /-l/ remains unchanged in Western Cham; Khmer permits /-l/ and /-n/).

This case illustrates the fact that the more powerful Mon-Khmer languages can still exert constraining and shaping structural and typological influences upon Chamic languages. This is especially so when we consider that the speakers of Phan Rang Cham are largely bilingual in Vietnamese (in any case Phan Rang has long had a large Vietnamese element in its population, an element which is now so large that it now outnumbers the Cham sector.)

There do appear to be some highly marked changes in Chamic languages which have arisen independently or which have become independent of changes in dominant languages, even if the original impetus for such changes was from neighbouring Mon-Khmer languages.

A striking example of this is the development in Rade which has arisen from the bipartition of reflexes of Proto-Chamic initial consonants according to whether they belong to the syllable proper or the pre-syllable (which was the former first syllable when Proto-Chamic was disyllabic). Over time the number of consonants which may occur in Rade at the beginning of the pre-syllable, and therefore at the beginning of most Rade words, has shrunk from over a dozen to three, /h k m/ (the initial clusters involving which are exhaustively listed in Shintani 1981). Zero is also permitted as the reflection of certain voiced stops which find themselves in pre-syllables; the coalescence of zero and the first vowel results in /e-/. Consequently very many disyllabic words in Rade commence with /h k m/ (as do an unusually high number of monosyllables, since contraction of the vowel that occurred between these consonants and the major syllable had already occurred before the initial consonantal change was implemented).

On the other hand, the original monosyllables which have not been contracted from original disyllables show a greater range of initial consonants. Thurgood (1999: 76) demonstrates that the three consonants /h k m/, which do not constitute a clearly defined phonological subset, are used as specifically pre-syllabic reflexes of numerous Proto-Chamic consonants which are much better preserved, and much more clearly differentiated, within the Rade syllable proper. For instance /k-/ is the reflex in these circumstances of all original voiceless stops apart from /p-/, while the labials which occurred as the first consonant of the presyllable, including /p-/, are now represented here by /m-/.

Examples of these forms are given as follows in Figure 3 (all words that have been chosen are spelt phonemically where possible, and all of them reconstruct to Proto-Malayo-Polynesian):

	PMP	Proto-Chamic	Rade
'rat'	**tikus*	**tikus*	*kĕkuih*
'damp	**basah*	**basah*	*mĕsah*
'salt'	**qasiRa*	**sira*	*hra*
'thorn'	**duRi*	**durɛy*	*erue*

Figure 3: *Development of some Rade presyllabic onsets (Thurgood 1999).*

Now it is not unusual for a language to adopt new syllable canons; it is rather more unusual for a language to adopt the same general principles of constraints upon the structure of that syllabic canon as the donor language had. It is even more remarkable that a language such as Rade should reconfigure the principles pertaining to consonant-initial presyllables in such a drastic way as it has done. This is especially notable since the more sweeping of these phonological changes appear to have been carried out independently in Rade, and not as the reflection of contact-induced processes of phonological change (though one could certainly maintain that they have been carried out as a *consequence* of these contact-induced processes). The importance to historical phonologists and Austronesian diachronists of recognising the very fact of this un-Austronesian change, and then of understanding the ordering of the steps which brought about this change can certainly be imagined. Developments brought about by these changes can be understood more clearly if the historically-motivated rules are applied in the relevant order. This is another reason for linguists to apply processes of 'top down' reconstruction. Since they already know the answer' to the historical riddle, they can reconstruct the stages obtaining between the proto-language and the modern language in the correct sequence.

This massive reduction of possible presyllabic onsets is a change within Rade which has no parallel within a neighbouring Mon-Khmer language, nor even with a Chamic language such as the neighbouring language Jarai. Thurgood (1999: 78) draws parallels with a similar change in the Mon-Khmer language Chong, a Pearic language spoken in eastern Thailand in which /k-/ has become the only permissible pre-syllabic consonant, though Chong does not neighbour Rade territory. It is as though a trend which was already present in the language as the result of contact with Mon-Khmer, and which had begun its operation in Rade and other languages too, has been independently extended within Rade phonology. The effect of this is to develop within Rade a morphological pattern which has affected the structures of syllabic and word-level Rade phonology. This has happened in much the same way as a non-Semitic language such as Farsi would have been affected, if the extremely strong impact of Semitic languages had caused a redesigning of Farsi polysyllabic elements into forms imitating the traditional triconsonantal Semitic canon.

Since the speakers of Rade were numerous enough and strong enough to resist wholesale influence from surrounding Mon-Khmer (or other) languages, we may note some of the developments in Rade phonology as indicating that after a given period of externally-actuated change, Rade was able to develop phonological changes which were both internally-driven and which seem to be crosslinguistically startling and unparalleled in neighbouring languages. This gives a glimmer of an indication of some of the directions in which Chamic languages might have changed had they been relieved of influence from external forces a millennium ago. And it is rare indeed that such a specific phonological template as what we may call 'the withered presyllable template' has been borrowed from one language and has then been so thoroughly implemented throughout the lexicon of the recipient language. Yet in Rade it is even used with inherited forms.

Had Rade or its Proto-Chamic ancestor never been in contact with Mon-Khmer languages, it is probable that such a range of phonological changes, from the development of sesquisyllables to the restrictions upon the consonantal presyllabic onsets, would never have taken place. But nonetheless the changes which are exclusive to Rade, striking though they are, are internal developments - even though the initial impetus towards syllable contraction and dissimilation of the pre-syllabic consonant was external, deriving from the influence of Mon-Khmer languages. The example of Rade is an interesting illustration of the fact that striking changes may occur even in languages which (relative to their geographical area) are dominant, or which have been dominant rather than subservient languages and which have not borrowed massively from their neighbours after the breakup of Proto-Chamic. (But then Chamic is a linguistic group in which the splits into new languages have occurred most strikingly among languages spoken in the northern area, the area from which most invasions have come, with more southerly languages the last to be riven apart by northern invaders. In addition, since more southerly Chamic languages have been in closer contact with one another, it has been easier for innovations to diffuse among them.).

The effects of internally-driven grammaticalisation in Chamic, meaning in this case the development of structures or semantic changes which are not replicated in or predicated on Mon-Khmer models, can also be seen in a number of cases, some of which instantiate the essentially random nature of transfers into Chamic languages from Mon-Khmer languages. For example, the verb *dok* 'to sit', a stem which is of at least Malayo-Polynesian vintage and which is shared with Malay *dudok* 'to sit', secondarily becomes used as the existential verb 'to be' in Western Cham (Baumgartner 1998), a usage which is not paralleled in Khmer. On the other hand, in most Chamic languages the verb meaning 'to stand' is not inherited from Proto-Malayo-Chamic but derives from Mon-Khmer and has the form of *dêng* (Thurgood 1999: 316).

Another interesting example, in this case an instance combining internal development, calquing and transfer of borrowed material, is that of the series of bipartite negatives, which can best be described as circumfixes since more often than not they go at either side of the verbal piece. These are to be found in Cham and most other Vietnam Chamic languages, and which are described in Lee (1996). It is possible that *ôh*, the negator which is found in Northern Roglai, Rade, Jarai and Eastern Cham, and which serves as the second, post-verbal negator, derives from Mon-Khmer, but this is not certain. (There does not seem to be any trace in Mainland Chamic languages of the Malayic negator that is represented by Standard Malay *jangan* 'don't'.) However, whatever the actual forms in use may be, bipartite negatives as a pattern are commonly found and are used for emphasis ('not in the least') in a number of Mon-Khmer languages, including Vietnamese, Chrau and Northern Khmer. What has been transferred from Mon-Khmer to

Chamic is not so much the form of the negator which is used but rather the bipartite pattern of negation. (The borrowing of a Vietnamese form *đừng* by speakers of Chamic languages such as Northern Roglai to express the negative imperative is a separate matter, but after all, Proto-Chamic took over *bè'* 'don't' from Mon-Khmer, and this form is old enough within Chamic for it to occur even in Acehnese as well as in other Indochinese Chamic languages.)

If Lee's surmise is correct, then there is a further feature in the structure of Chamic bipartite negation which as far as I know cannot be traced as a calque from Mon-Khmer languages, and that is the construction of the first negator in Roglai from a form which is phonologically identical to the Roglai (and indeed Common Chamic) verb meaning 'to see', and which may indeed be derived from this. As yet we cannot explain everything about the channels of origin and development of bipartite negators in Chamic simply by reference to predictions from certain contact phenomena. But what we have here in Roglai, as in so many cases in Chamic, is an independently-composed riff on a theme donated by the result of contact with Mon-Khmer.

But we can use a combination of social factors, which explain the ways in which contact and more importantly transfer was made possible, and (secondarily) various structural-typological factors, in order to unravel some of the contact history of this and other constructions. Thereafter we may avail ourselves of the opportunity (which is enhanced by the availability of comparative linguistic and philological materials) to see these factors operating on linguistic material whose previous history is well-understood. As such they will enable us to see something of the possibilities and effects of a remarkably strong degree of linguistic contact, driven by migration and apparently enhanced by numerous instances in history and prehistory of communal language shift, which has operated across numerous genetic boundaries (Chinese, Tibeto-Burman, Tai, Kam-Sui, Hmong-Mien, Mon-Khmer-Austroasiatic, and Austronesian) in Southeast Asia. The area south of the Yangtze and east of the Irrawaddy is a geographical region which has previously received relatively little attention in the general run of language contact literature. But it is one in which areal forces have been remarkably strong in effecting typological change and in incorporating 'new' languages (languages originating outside the area, or arriving from outside) into membership in typological networks. (This typologically-charged state of affairs is what brought into being the earlier development of tones in Vietnamese and Mường, for example).

It is certainly true that the effects of various waves of Southeast Asian areal contact (and also the effects of the influence of individual Mon-Khmer languages) upon Chamic languages, both as a unit and even more as individual languages, have been astoundingly strong. The borrowing of numerous Mon-Khmer forms into these languages, with their distinctive and very 'un-Austronesian' phonological features, is only the most obvious and easily-spotted manifestation of this influence. These contacts lead us to recognise the different kinds of effects of contact, direct and indirect, which we can find here. The impact of Mon-Khmer can modify the shape of forms which in their origin are purely Austronesian. And we need to recognise that aside from a batch of contact-induced changes which all Chamic languages (or later, batches of changes which all of them save Acehnese) have undergone, several further structural changes, often very striking ones, are confined to one Chamic language or to just a small group of them. (To take an example from the most easily diffused stratum of a language, quite a few words of assumed Mon-Khmer origin are found only in the Highland Chamic languages and secondarily in Haroi, which we know to be a displaced Coastal Chamic language now used at the edges of the Southern Highlands.)

Typologically the Indochinese Chamic languages are coming to look more and more like the Bahnaric and other Mon-Khmer languages with which the bulk of their speakers are in contact (or have been in previous centuries). Meanwhile Acehnese has retained many features which it inherited with Malay from their common ancestor, and which the Indochinese Chamic languages lost or permitted to atrophy as the result of exposure to Mon-Khmer languages. This continuing typological convergence towards Mon-Khmer languages is still the case for Chamic languages in Cambodia and Vietnam, even though they are probably no longer absorbing elements from the Bahnaric languages that shaped them.

Meanwhile Tsat (as Pang 1998 shows, the name derives from *Cham*, with phonological changes which show the effects of the phonological canonical syllabic constraints of both a Chamic language similar to Northern Roglai and Hainanese) has undergone perhaps the strongest and most radical set of changes of them all. It has relinquished the feature of voicing in stops for a distinction between aspirated and non-aspirated voiceless stops (as certain other Chamic languages have done, though independently, becoming rampantly monosyllabic in its stem form to an extent unparalleled in other Chamic languages. In both instances Tsat has assumed the phonological features characteristic of Hainanese Chinese, a Southern Min language. Indeed the phonological inventory, the tendency towards monosyllabicity, and the strongly marked and very Southern Chinese constraints on syllabic canons and on final consonants in Tsat are very similar to those of Hainanese (although Hainanese does not have preploded nasals as Tsat and Northern Roglai do). Tsat has five phonemic tones to Hainanese Chinese's six, though the five Tsat tones resemble five of the Hainanese tones perfectly (in the case of the level tones) or very closely (in the case of the falling and rising tones); they also resemble tones in varieties of Hlai. It is unfortunate that because of the paucity of relevant information in the literature (*pace* Zheng 1997) we cannot say very much specific about the possible linguistic influence of Hlai upon Tsat, since Hlai itself, as a Tai-Kadai language, is, like Hainanese, polytonal and monosyllabic, nor can we be certain that it rather than Hainanese provided the initial impetus towards tonality and monsyllabicity. But we should never forget that Tsat, like its sister language Roglai, had already become indelibly impregnated with Mon-Khmer typological features and basic lexicon before it came into contact with Kadai and Chinese languages. Perhaps a closer examination of Mon-Khmer lexical elements in Tsat, and an analysis of those which is shares uniquely with one or another Chamic language, would enable us to see whereabouts in Chamic it derives from.

Despite the fact that their period of divergence from the immediate ancestral language probably does not exceed a thousand years, the Chamic languages nonetheless show such a startling range of linguistic systems, especially phonological systems, that we have to reconstruct from the top down, as Thurgood (1999) cheerfully admitted to doing, in order to reconcile the features of the many and divergent systems to the framework of a coherent and cohesive historical pattern. The existence of material from earlier stages of Cham, and the parallel example provided by (modern) Acehnese, are invaluable in this respect, and they help to indicate that a top-down approach is the correct method to employ. But even these materials cannot solve all the problems for us, because they present problems themselves which are mostly related to the narrowness of their scope (in the case of Cham) or to later forms acquired as the fruits of their contact histories (in the case of Acehnese).

We have to use a certain degree of diachronic foreknowledge in order to avoid building traps for ourselves by reconstructing proto-forms which do not really go back to Proto-Chamic. For example, Acehnese evidence cannot be used as a failsafe guide to the extent and content of the Austronesian or Malayic stratum in Chamic languages because it has been in strong subsequent contact with Malay, nor can its Mon-Khmer stratum be wholly attributed to the same Mon-Khmer languages which influenced other Chamic languages. There are more Katuic elements in Acehnese than occur in other Chamic languages (which however do appear to have a very few forms of Katuic origin, some of which are shared with Acehnese). And there may also be some borrowed Aslian elements in Acehnese, and these latter are naturally enough completely alien to Chamic languages, which have never been in contact with Aslian languages. There are also hundreds of loans from Malay which are found in Acehnese and which do not occur in other Chamic languages, and a number of post-Proto-Chamic loans from Malay, especially in those Chamic languages which were used by Muslims. These forms are often plentiful, but they have to be discounted before one can begin to reconstruct Proto-Chamic in any detail with any hope of achieving the comparatist's dream of reconstructing a proto-language which is as similar to (or better yet, which is identical with) the ancestral language which people actually used as one can make it.

And yet the very fact of historical separation of speakers of Acehnese from speakers of other Chamic languages can be of some use to us. If an archaic feature is not found in Chamic or in Malay, but is retained in Acehnese, then we can confidently project it back to the Proto-Chamic era. This is true of certain kinds of infixation, specifically those involving reflexes of Proto-Malayo-Polynesian *–um-* and *–in-*. There are a very few embalmed relics of both of these as parts of individual words in Malay and in written Cham (a language which is considerably more archaic than modern Cham dialects are, and which thus reflects the Cham language as it was used in previous centuries, before the split into Eastern and Western Cham), but these infixes are fully productive in Acehnese, even though there is no neighbouring language which has influenced Acehnese to such an extent that speakers of Acehnese could have borrowed them from a language that had retained them; they must be inherited.

All this means that these infixes must have been vital and productively-used forms in the language ancestral to Acehnese and the other Chamic languages, since Acehnese could not plausibly have borrowed them from any other language after Acehnese-speakers arrived in Sumatra. Therefore Acehnese must have retained a feature which has been more or less lost in the other languages under inspection.

4.1 Finding and exploiting theoretical frameworks concerning mixed languages: ideas and underpinnings – and some observations.

Three factors, two of them astoundingly obvious but still overlooked, have to be borne in mind when one is examining the results of a situation of language contact. Firstly, we should recognise that languages are systems of behaviour which are created by people and as such, they are changeable by people, even if this change is automatic, teleologically blind, and non-predictable in the chronology of its changes (though the outcomes of such changes can often be predicted). Whatever else it may be (for it is seen as being many things, and its status as a symbolic system is not ruled out by what follows), language is something that people do (see an illustration of this in Le Page and Tabouret-Keller 1984), and it is people who make language change, and who sometimes attempt to keep it it the same as it used to be.

Bradshaw (1995) is an exemplary discussion of the discourse of contact-induced language change and of the way in which people as agents of such change have been lost sight of as a result of the reification of behavioural systems as constructs called 'languages'. (These matters can be kept at the back of one's mind when writing about contact linguistics, but their essential veracity and crucial importance must never be forgotten. They provide a covert theoretical backdrop without which any further discussions would be meaningless.)

Secondly, people inherit these constructs called languages as behavioural systems which they learn from other people, and in so doing they inherit the changes, including those driven by contact, which have accreted in these languages, changes in which neither they nor their recent ancestors may have had a part. (An example of what we may call this 'principle of unrecognised inheritance' in the Chamic languages would be the large-scale incorporation of elements from particular Mon-Khmer languages with which the speakers of some Chamic languages may not have been in direct contact for a millennium or more. These elements are now firmly part of the Chamic language in question, although their origin in Mon-Khmer languages will be unknown to speakers of the Chamic languages, since they are no longer in contact with the domor languages. Most of the overt knowledge of the history of one's language is acquired externally, rather than it being part of any language acquisition faculty.)

Thirdly, there are two kinds of language contact (or rather, we may say that active language contact results in the transfer of two kinds of features). These are the *transfer of fabric* and the *transfer of pattern*. Transfers of both kinds of these features have happened frequently in Chamic languages, sometimes with one occurring as a consequence of the other. And both of these kinds of transfer can result in typological change in a language, if the transferral of a pattern includes the transferral of the relevant morph in order to actuate pattern transfer. Transfer of fabric involves the transmission or copying of a morph from one language, which we may call the donor language, to another language, which is called the recipient language. Borrowing an affix or a lexical item into another language involves transfer of fabric. These borrowings can bring about the transfer of patterns if the item in question includes (for example) a phone which did not previously occur in the phonetic system of the recipient language, but which is brought over into that language from such a word. Such phones can in time come to modify considerably the phonological system of the recipient language and thetypological features which this contains.

One can also argue for the borrowing of phonological features (such as aspiration or nasalisation, which often occur first of all in borrowed items and sometimes as secondary developments in a very few inherited items) as being a kind of borrowing of fabric, which results in the modification of patterns, if this occurs by borrowing words containing these features. Nevertheless it makes more sense to see such borrowing (for instance the taking over of iambic syllable pattern in Chamic, which did not previously have these) as a kind of transfer of pattern which may originally have been brought about in the first place by the transfer of lexical fabric.

The transfer of pattern (for this is what it was named in Heath 1984; the term 'transfer of fabric' is my own coining) involves the addition to the grammar of a language of a rule which introduces previously unfamiliar patterns in the ordering of pre-existing (or indeed borrowed) morphs. The moving of the cardinal numerals in Western Cham (Baumgartner 1998: 15) from their position before the noun, as they occur in Malay, to a place after the noun but before the classifier, as they occur in Khmer, is an example of the transfer of a pattern without the transfer of the actual relevant morphs taking place from one language to the other. People who gain familiarity with a second language from which they are disinclined or unable to borrow much lexicon (or who do not feel the need to borrow much lexicon, because they already have names for all the relevant cultural features and concepts) may indulge in a considerable degree of transfer of patterns, and they may do this without acquiring many morphs from the language from which they have absorbed these patterns. It is apparent that at least during the period of existence of the southern Cham empire, the Chams were able to dominate other groups, and the Cham language served as a source of loans with which Bahnars and others could name previously unfamiliar concepts. But it also seems likely that many speakers of Cham at that time were Cham-Other bilingual descendants of Mon-Khmer-speaking people who had adopted Cham as their major language, and who were able to exert a surprisingly large amount of influence upon the language to which they were to shift.

Both these kinds of pattern are significant in language contact, but the so-called 'mixed languages' (I refer to them as 'so-called' because there are many differing definitions of them, and because as a consequence no two investigators' lists of mixed languages coincide exactly) rely more on issues in transfer of fabric than on transfer of pattern. Fabric identification is especially important when one goes about identifying linguistic systems as 'mixed languages' (see Bakker and Mous eds. 1994 for an important discussion of several mixed languages).

For these authors the default model of mixed language (and it is certainly the model which explicates the largest number of cases) is the *intertwined language*, a speech variety in which the lexicon derives from one language and the morphological apparatus derives from another, and in which neither lexicon nor morphology have been significantly reduced in form or content. This model accounts for languages such as Media Lengua of Ecuador, Ma'a of Tanzania, and Amarna-Akkadian of the ancient Near East, all of them discussed and exemplified in Bakker and Mous (eds. 1994). Well-known mixed languages such as Michif (with Cree verbal stems and morphology and with French nominal stems and morphology), or Mednyj Aleut (with Western Aleut stems and nominal morphology, and Russian verbal morphology applied to Aleut stems) fit the pattern less readily. This less-than-perfect fit into the 'classical' intertwining model is also true, though for slightly different reasons, of Callahuaya, the secret language of a group of itinerant male native curers in Bolivia, which uses a morphosyntactic system with its origins in several forms of Quechua together with a lexicon based on the extinct Andean language Puquina, but also incorporating elements from Tacana, Quechua, Aymara and Spanish, in addition to using many lexical forms of unknown origin.

Such a model of genesis or analysis applies even less well to the Chabacano Creole Spanish variety of Zamboanga City and adjacent areas in the Philippines (Forman 1972), in which the blending of Spanish and Bisayan elements involves replication of some Philippine structural and semantic subsystems using either wholly Spanish elements or else a combination of Spanish and Bisayan elements with Spanish elements being in the majority. (The Zamboangueño plural pronominal system has preserved the transferred Hiligaynon plural pronominal paradigms almost intact and without undue simplification. But the singular elements in the personal pronominal paradigm, which are taken from

Spanish, show the effect of modification of an original system; the traditional Spanish direct and indirect object forms are not preserved in Zamboangueño.)

Furthermore, in the case of Berbice Dutch of Guyana (Kouwenberg 1994), although all the bound inflectional morphology which the language possesses is drawn from Eastern Ijo while Dutch comprises the largest element in the lexicon, a great deal of basic vocabulary derives from Eastern Ijo too. But the structural subsystems which have been taken over from Eastern Ijo represent only a small portion, and that simplified, of the inflectional morphology of Eastern Ijo – most Eastern Ijo morphology has never been taken over into Berbice Dutch. In addition, both these elements have been modified in terms of their phonological representation, so that Berbice Dutch is not truly an intertwined language in the strict sense because intertwined languages do not radically simplify either of their major components).

Can we examine the Chamic languages profitably in this light? We can try, but there are severe limitations to the application of the standard or classical 'language intertwining' formula to any or all Chamic languages. We need to separate out the lexicon from the morphology, and then we need to source the contents of these two bundles of elements. In doing so, we find that the division of Chamic forms between lexicon and bound morphology is almost exclusively in favour of forms with 'structural' meanings (personal pronouns, etc.) counting as lexicon, since there is so little bound morphology. Most of the rather few morphological processes which are overtly expressed in Chamic languages are expressed by free morphs. If the Chamic languages were mixed languages in the full 'language intertwining' sense, we would expect them to involve Austronesian lexicon being employed in a framework of Mon-Khmer morphology, and to a very small extent, this is what we find.

But we immediately encounter two problems. Firstly, the amount of Austronesian or even Malayic lexicon in Chamic languages is a small and static proportion of the total morpheme list of any Chamic languages. We do not have precise figures for the number of morphs which derive from Proto-Malayo-Chamic sources, since Thurgood only discusses those words which have been reconstructed to Proto-Chamic or to a cluster of its daughter-languages. There may be some lexical orphans of Malayo-Polynesian vintage which are still lurking in the vocabularies of less-exhaustively documented Chamic languages, for all we know. (I have not come across any such in my search through the data.) But if the total number of forms in Chamic languages which have been inherited (rather than borrowed) from Proto-Malayo-Chamic is much more than 300, including both bound and unproductive morphs, we may justifiably express surprise. For the record, Thurgood lists 285 such forms, and with a few exceptions, his assignments of these to a descendant of Proto-Malayo-Polynesian are correct, and any bookkeeping mistakes found there are cancelled out by the tiny number of unrecognised forms of Austronesian origin which he misclassifies elsewhere. (Data from the observations of Blust 2000 would raise the total of Proto-Malayo-Chamic forms to 292.)

Naturally, not all these Malayo-Chamic forms will go back to Proto-Malayo-Polynesian or even further back, and in fact there is a small battery of shared innovated lexical forms (Blust 1992 lists 21 such forms) which indicate a special relationship between Malayic and Chamic. However, there is no similar battery of items which indicate that Chamic has a special and cladistically exclusive relationship with, say, Barito languages or with Philippine languages. All forms which are of Malayo-Polynesian origin and which occur in Chamic will either reconstruct back to Proto-Malayo-Chamic and they may thus be used as evidence for the earlier presence of forms of Austronesian origin which Malayic has shed, or else are later loans from Malay. (It should be noted, though, that differences in basic vocabulary in Chamic languages are rarely due to the possession of larger tranches of Mon-Khmer loans or unsourced elements in some Chamic languages than in others, even though the contents of the tranches may differ somewhat from one language to the next. Most of the reconstructions of elements of Proto-Chamic lexicon which Thurgood 1999 provides can be found in most or all the Indochinese Chamic languages, and are not just to be found in Highland ones or Coastal ones.)

(However, we may note that the discussions in Blust 1999, 2000a provide only 199 and 285 forms respectively as being reconstructible to Proto-Austronesian for the lexicon of Pazeh of Taiwan, and reconstructible to Proto-Malayo-Polynesian in the case of his work on Chamorro of the Marianas. The total number of 'Austronesian reconstructibles' for Proto-Chamic, which stands at almost 300, may be higher than these totals. But it must be pointed out that the requisite forms for Proto-Chamic include those reconstructible to Proto-Austronesian, to Proto-Malayo-Polynesian, and to a putative Proto-Western Malayo-Polynesian, and also those which are only reconstructible to Proto-Malayo-Chamic, as well as those which may reconstruct to any intervening but as yet unassured subgroups such as Proto-Malayo-Javanic. And furthermore not all Proto-Chamic forms of Malayo-Chamic origin are perpetuated in all its daughter languages, as Thurgood's listing shows.)

This total of 285 inherited forms (give or take ten forms) compares with a little over 200 items which have been demonstrated to have Mon-Khmer affinities and which are also attested at the Proto-Chamic level or at a cross-subgroup level within Chamic. The number of forms which are of Mon-Khmer origin, and which are not recent loans from Bahnar, Vietnamese or Khmer (all of which have donated large amounts of lexicon to individual Chamic languages), may yet rise in the light of our increased knowledge of the proto-lexica of subgroups and sub-subgroups within Mon-Khmer. Eventually the total of pan-Chamic items which are assuredly of Mon-Khmer origin may even surpass the number of pan-Chamic forms which have been inherited from Proto-Malayo-Chamic.

The number of unsourced items and the number of possible but as yet unproven Mon-Khmer forms are both quantities which are large enough, and represented significantly enough in the basic vocabulary of Chamic languages, to be statistically notable and therefore it is necessary for them to be taken into account in a historical study of Chamic languages. Unsourced items include some personal pronouns, some interrogative pronouns, and a number of high-frequency verbs. We must remember that even if we exclude from the total of unsourced elements those forms which may actually be from Bahnaric, but the trajectory of whose diffusion cannot be verified, we still have over a hundred unsourced forms which are nearly or wholly pan-Chamic (that is to say, some of them also occur in Acehnese), and which include some of the commonest and most polyvalent words in these languages. And our sources for these languages are not so sparse that we must have missed out many obviously Mon-Khmer words occurring in Chamic.

Probably only a quarter or less of the morphs which occur in any Indochinese Chamic language can be provided with a secure etymology from any language or proto-language, be it Austronesian, Mon-Khmer or otherwise. (The influence of the various Mon-Khmer languages on individual Chamic languages is the theme of Table 6.) But even this number is itself merely guesswork.

Secondly, morphology of any sort, especially bound morphology, is in short supply in Chamic languages. And not even the origins of free grammatical morphs, such as personal pronouns, give a very clear picture of the origins of the languages themselves. By no means all the free grammatical morphemes in Chamic languages derive from Proto-Austronesian. Some of them certainly do; others, including many common ones, are taken from Mon-Khmer languages; yet others are of uncertain origin, even if some exhibit the characteristically Mon-Khmer sounds such as implosives and low mid vowels. This much is true of personal pronouns and of prepositions, both of these form-classes being matters which Thurgood discusses in some detail. (Singular personal pronouns in Chamic languages tend to be Malayo-Chamic in origin, while plural ones have more diverse origins. There are some pronouns that are inherited from Proto-Malayo-Javanic and there from Proto-Austronesian, others which are loans from Mon-Khmer languages and others whose origin is as yet unknown, and there is also a great deal of use of pronouns which are not number-specific (the distinction between these pronouns being whether they are informal or polite) and which can be construed either as singular or as plural pronouns.)

As to the bound morphology, which is derivational rather than inflectional in nature, Thurgood points out that the infix *–um-* (which is productive only in Acehnese, and otherwise only found in a few fossilised forms in Cham) is certainly Austronesian. But the productive infix *–ən-/-an-* is from Mon-Khmer despite its resembling an Austronesian infix of similar shape and broadly similar meaning. (As early as the ninth century AD, the infix was integrated strongly enough into Cham for it to be applied to Cham stems which were themselves Sanskrit loans*: ś-an-āpa* 'a curse' from Sanskrit *śāpa* 'curse': Marrison 1975).

The productive causative *pa-* is more likely to be from Mon-Khmer than from Malayic (the form of a causative prefix commencing with *pa-* is attested in Austronesian, for instance in Philippine languages, where it is an inheritance from Proto-Austronesian, but is unknown in Malayic at any stage). Meanwhile the productive Chamic verbal prefix *mĕ-* derives from Austronesian and is shared with Malayic (where it is *meng-*; in both Malay and Chamic there are also embalmed relics of the Austronesian infix *–um-* in a few verbs and deverbative nouns). The non-productive 'inadvertent' prefix *ta-* is found in both families (it occurs in Malay as *ter-*: *tertawa* 'to laugh'), though both the senses and the forms differ slightly from family to family and from one member to another within Mon-Khmer. And the non-productive individuative particle *sôh* is certainly from Mon-Khmer (the sources of these are discussed in Thurgood 1999: 237-250). In short, most of the few productive items of derivational morphology that are found in most Chamic languages derive from Mon-Khmer, while much of the morphology which also occurred in Proto-Malayo-Polynesian is only found in Chamic languages in a few items, in which it now forms part of the stem.

Nonetheless, it is important to make clear that in Chamic languages both Malayic and Mon-Khmer elements (and also other loans, such as Arabisms and Sanskritisms, and of course the unsourced elements) make use of the same small set of morphs for grammatical purposes. Chamic languages do not have parallel Malayic and Mon-Khmer morphological systems into which forms of the same origin (Malayic forms into Malayic structuires, etc.) are inserted. Mon-Khmer elements are integrated into what there is of Chamic morphology, and as such they can and do take a Malayic verbal prefix such as *m-*. So can verbs of unknown origin, since the prefix is productive at this period. Similarly, verbs of Malayic origin can and do take the Mon-Khmer prefixes and infixes which have been taken over into Chamic languages. There is one and only one morphological system in use in any one of the Chamic languages, even if the origins of its various elements are diverse. There are no morphological features in Cham grammar that are used only with loans.

Once the free grammatical morphs have been analysed and etymologised, what we are left with in terms of Chamic morphosyntactic structure, to a very large extent, is simply a bundle of element-order rules, and these by their very nature cannot be used to prove genetic affinity. Typology can tell us nothing about the genetic source of a language, and we should never assume that it can do so. A typological profile is simply the aggregation of certain salient structural characteristics which happen to be present in that language at any one time. Some characteristics which find their way into this profile may be inherited while others are acquired through borrowing or through intrernal innovation, and yet others may once have been present in the language but have since been replaced or shed. Typological features can be lost or modified as a result of other changes taking place in the language, and when this happens, the typological profile of the language will then be reclassified (and can then be equated with profiles for quite a different selection of languages, to none of which it may happen to be related) without this suggesting that the language has departed further from its genetic inheritance. The typological change of essential features in a language does not imply the concomitant adoption of linguistic fabric from the language which influences its typology, and it does not impugn the validity of its genetic affinities. Both of these are facts which Ross (1996) astutely demonstrates for the Austronesian Takia language of northern New Guinea, which has copied much of the syntax of the non-Austronesian language Waskia without borrowing the morphs needed to carry this operation out – or indeed without borrowing many morphs from Waskia at all. (In fact Waskia has borowed a greater amount of vocabulary from Takia, and has done so at a more basic level than Takia has done from Waskia.)

Typological affinities are subordinate in importance to genetic ones, although they can be extremely informative about historical contacts and about potential patterns and directions of grammaticalisation. But even then they have their limitations, and they cannot be relied upon excessively. For example, the matter of verb-placement aside, there is no special historical or typological link which unites all verb-initial languages (for instance) in ways which separate them substantively from all languages which are not verb-initial. And what is more, the possession of verb-initial word order is a sign of membership of a club which can be joined at a late date (as can be seen from the history of the Insular Celtic languages when they are compared with the material from Continental Celtic). But history shows that it is also a group which also can later be departed from (as the history of Malay shows: the ancestor of Malay was VSO but Malay is now SVO).

What typological features do come in useful for, however, is to demonstrate typological allegiance in areas in which similar morph orders are shared across genetic boundaries, which enables one to map linguistic areas, and which allows one to predict the likely pathways of instances of grammaticalisation. As I have shown in Table 5 and as Henderson (1965) demonstrated with her discussions and maps, Cham is an even surer and more solidly confirmed member of an areal Sprachbund than Malay is, and this affinity is therefore one which cuts across genetic boundaries. (Many of the features which I have listed, especially the more 'marked' ones such as the use of numeral classifiers, could be paralleled in Thai, Lao, Burmese, Hmong, Mien and various Chinese languages, to name just some of the more obvious languages, just as they are found in Mon-Khmer languages.) Cham's membership of this Sprachbund was brought about by, and is based firstly upon, the presence of those features which might have been acquired by Proto-Malayo-Chamic from intimate contact with South East Asian languages, if there are such features. But it has been massively reinforced by two millennia or more of strong contact with Mon-Khmer languages, languages which also have precisely such features.

Settling the question of whether the Chamic languages are mixed languages is made somewhat easier by the fact that we have material on Chamic languages from a sufficient number of periods, and from far enough back, for us to be certain of the broader paths of development of Chamic languages from a language which itself had undergone numerous contact-induced changes before diversifying, but which in its earlier form was once very similar to Malay. (We can see the very thorough absorption of Mon-Khmer elements into Chamic as it took place from the ninth century or before) It is clear that the lexical forms in Chamic which are not found in earlier Chamic materials, and which cannot be traced back to Proto-Malayo-Chamic because they are shared with other languages in the area, are the ones which are intrusive from other languages. It is therefore clear that they are not relics of some lost language which has been submerged under an inundation of Austronesian morphemes, thereby giving rise to Chamic. Whether or not they outnumber the elements that have been inherited from Chamic's proto-language is strictly irrelevant to the question of the genetic origins of Chamic, although Malayo-Chamic elements do have a slight numerical edge in the realm of basic vocabulary.

Table 6 presents a summary of major retentions, innovations and losses in the phonological, morphological and other strata of the Chamic languages. I discuss various stages of the histories of the Chamic languages in an appendix at the end of this paper.

Table 6: *Conspectus of retentions, innovations and losses in Chamic (in certain languages, and in Chamic in general) which have occurred since its separation from Proto-Malayo-Chamic.*

The four periods listed here are as follows:

Period 1: Malayic and Chamic are a single language.
Period 2: Chamic splits off from Malayic and begins to come into contact with Mon-Khmer languages.
Period 3: Chamic is strongly modified by the effect of Mon-Khmer languages, and the historical records of Cham begin.
Period 4: Chamic splits, Tsat and Acehnese go their separate ways, and the various other Chamic languages undergo secondary influence from other languages.

Retentions

- A few hundred lexical (and principally contentive) stems of Austronesian, Malayo-Polynesian or Malayo-Chamic origin, with their original disyllabic forms retained to a greater or lesser extent
- A couple of partially productive derivational prefixes with broad but originally verbal ranges of meanings

Losses

- Loss of many contentive morphs. Many Malayo-Chamic stems, perhaps more than 50% of those which would have been inherited from Proto-Malayo-Chamic, have been replaced by forms of Mon-Khmer, other, or uncertain origin ('partial relexification'). This loss applies also to many free grammatical morphs.

- Loss of the focus system and of the aspectual features associated with it, of the ligatures within phrases, and of ergative features of syntax
- Loss over time of most prefixes and suffixes, together with their uses, and the loss (in all but Acehnese) of the productive use of infixes
- Reduction of most pre-stressed syllables with the concomitant loss of the vowels in these syllables

Innovations

- Gradual shift of the standard Chamic word-shape from disyllable to monosyllable by way of sesquisyallabic forms (under the influence of Mon-Khmer languages), with the effect of introducing initial consonant clusters into these languages (This change takes place in periods 2-4).
- Development under Mon-Khmer influence of pre-syllables as a separate phonological entity with their own sets of constraints (period 2)
- Development (under Mon-Khmer influence, though not always identically in all details) of a new (yet smaller) phonological class of consonants which can occur at the beginning of a pre-syllable (periods 2-3)
- Introduction of phonation types from Mon-Khmer with far-reaching effects for Chamic language phonologies, most markedly in Haroi and Western Cham (periods 3-4).
- Introduction of the consonantal distinction (separately, manifested in different ways, and in several Chamic languages) between aspirated and unaspirated voiceless consonants, which begins to supplant the inherited distinction between voiced and voiceless obstruents, though glottalised obstruents remain voiced (period 4)
- Acquisition of numerous simple and complex vowel nuclei from Mon-Khmer languages and from words that were taken from such sources, many of which are also found in the unsourced element of the Chamic vocabulary. The complex nuclei are usually built up of elements which already occurred in the PMP element of Chamic. (Periods 1-4).
- Acquisition (and sometimes subsequent loss) of a set of nasalised vowels in some languages. These are first found in words of Austronesian origin (where they would originally have occurred allophonically) as well as in borrowed or innovated forms and they have developed in the environment of original nasal consonants (presumably period 2.)
- Acquisition of some preglottalised stop consonants (usually as a result of borrowing Mon-Khmer words which contained these) (Period 2-4).
- Acquisition (and licensing) of a final palatal stop (brought into Chamic first of all through words from Mon-Khmer, although the parallel word-final palatal nasal which also occurs in Mon-Khmer languages has not been transferred in that position into Chamic) (Periods 2-4). (Acehnese formerly had this palatal stop, which it nowadays realises as /-t/, although an original /-c/ is still reflected in the Arabic orthographical spelling of some Acehnese words.)
- Acquisition (which is separately executed) of the first stages of a tone system in Phan Rang Cham (under Vietnamese influence) and Tsat (under the influence of Hainanese, and maybe also originally Li) (Periods 3 and 4).
- Replacement of final voiceless stops by one of several outcomes (replacement with the glottal stop, development of preploded nasals, total erasure) (Periods 3-4).
- Devoicing of final voiced stops (this is an early change, possibly pre-Chamic and therefore belonging to Period 1)
- Development, under Mon-Khmer influence, of numeral classifiers (these are also found in Malay) (Period 3-4 or maybe earlier).

- Development of a series of phrase-, clause- or sentence-final discourse particles, which themselves are of varied origin (although some derive from Malay). (Periods 3-4).
- Acquisition and implementation of many contentive lexical loans from Mon-Khmer languages, which are often replacive of pre-existing forms (Period 2 onwards if not already within Period 1.)
- Borrowing and assimilation of a small number of prefixes or infixes from Mon-Khmer languages (Period 2 onwards?)
- Development, from at least common Chamic times, of a significant proportion of elements of pan-Chamic vocabulary, of uncertain origin, which is found in almost all form-classes and which outnumbers by several hundred percent the amount of innovated lexicon which is exclusively shared by Malayic and Chamic languages. (Presumably from Period 2 onwards.)

The absorption of morphemic material from other languages has been of most significance here. This is because it presents a sort of surprise when it is compared with the more quotidian and more easily-found effects of language contact. This is because there is no prima facie reason why a language, many of whose speakers acquired this language as an L2 and who speak the language with a strong L1 accent and sound system, should not absorb (say) phonological constraints from a more dominant language without taking over large amounts of morphs from these languages.

What we have as a result of cultural and social changes in Champa is a situation of pendular bidirectional diffusion. This is one in which elements have first gone from Mon-Khmer languages to Chamic and have influenced Chamic languages strongly, while afterwards a large number of elements have gone from Chamic languages to Mon-Khmer languages (and they are still doing so, since Cham is an important source of loans into modern Bahnar and Chrau). And although they may be more numerous and their effect in Chamic languages has lasted longer, they have not penetrated or influenced the core of the language half as much.

It would be stretching several points for us to describe the Chamic languages as mixed languages which incorporate a basically Austronesian or Malayic lexicon with a basically Mon-Khmer typology. The Malayic component of the Chamic lexicon is, as I have said, numerically outweighed by that portion which is of uncertain or Mon-Khmer origin. Even so, these strata are less germane to the etymologising of the contents of a Chamic-language Swadesh list, or to the sourcing of the items on the list that had been drawn up for the investigation of Bornean languages by Alfred B. Hudson (Hudson 1967) and popularised by Robert Blust, than the Malayic elements are. The discussion in section 3 has already shown this. But the testamentary evidence of those rather scarce elements in Chamic languages which are Austronesian or Malayo-Polynesian in origin and which do not occur in Malayic should also be recognised. The existence of such forms in Chamic languages will normally point to their existence in the parent language, even if they are lacking from the other daughter of that parent language.

But in a part of the world in which the practice of conducting linguistic classification according to the sources of the bound morphology in a language is a non-starter, specifically because there is no such morphology to analyse and classify, this kind of lexically-based classification (with comments on the occurrence or non-occurrence of certain typological features) may have to suffice. After all, such a kind of classification uses the most genetically diagnostic material that the languages can still provide. Lexical material is the least reliable kind, but we have next to no morphological material to go on, while the usual phonological strategies that historical linguists use in order to reconstruct languages are problematic when applied to Chamic languages, since such strategies usually begin by reconstructing the initial consonants of proto-forms, and this is not easy to do when working with languages in which presyllables have retained only a subset of original consonants and accompanying vowels.

We may quietly dispose of any idea that the Chamic languages are creoles deriving from previous pidgins, despite their paucity of inflection. There is no evidence of pidginisation at any stage of Chamic (although in the earliest materials we find numerous instances where Malay would have used an affix but where Inscriptional Cham zero-marks a particular grammatical relation, using apposition of elements instead, and this has occurred in texts which do not show wholesale borrowing of Mon-Khmer elements). There being no evidemce of pidginisation, nor do we find any evidence of subsequent creolisation. Nor is there any evidence of interrupted transmission of linguistic material from the earliest Chamic records to their lineal and genetic descendants in Acehnese, Cham and beyond. Nonetheless, the impact of Bahnaric languages on earlier stages of Cham suggests that many users of Cham who were living a millennium or more ago were actually L1 Mon-Khmer language speakers who shifted to using the language of the empire which controlled them, and whose shift to Cham culture, religion and mores enabled the intrusive Malayo-Polynesians to get a firmer foothold in the territory.

The unusual concatenation of acquired features in Chamic languages also raises the question of what constitutes an Austronesian language if morphology rather than lexicon is to be the definitive determiner of genetic affiliation. Can a language with no (or next to no) productive morphology of Austronesian origin seriously be classified as an Austronesian language? Is Cat Gia Roglai, for instance, truly an Austronesian language in any meaningful sense, what with its sprinkling of very partially productive bound morphology remaining as its only structural and non-lexical elements which are of Austronesian (or Malayo-Chamic) origin, and with its expanded and very un-Austronesian (and even rather un-Chamic) segmental and canonical phonology and syntax? (We need hardly mention the contents of its lexicon with its few hundred items of Malayo-Chamic vintage, its large amounts and equally large proportions of elements of non-Austronesian origin, and the complex and internally-driven phonological rules which disguise the essential shapes of many of the forms which it has inherited from Proto-Chamic and often from Proto-Austronesian.) We may wonder aloud just how much Austronesian material a particular language needs to have retained, how 'basic' (whatever that means) the material is meant to be, and what kind of material this has to be (lexical, morphological, syntactic), in order for it to be regarded as an Austronesian language.[5]

We need to decide which parts and subsystems of a language - indeed of any language - are definitive in our quest for the genetic affiliations of a language, and which ones are not. This is a complicated matter, and it is one that provides us with rather few options in Chamic languages. Here we are dealing with languages which do not afford us the benefit of preserving much irregular morphology or sets of suppletive lexical items, reflexes of which can be looked for in other languages with which they are assumed to be

[5] As a reductio ad absurdum of this principle, we should note that Kaulong, a language belonging to the Pasismanua branch of Oceanic which is spoken in inland New Britain, preserves less than 6% of PMP cognates among the forms which are reconstructed and presented on the Blust 200-item list (Blust 1993a).

especially closely related. And Austronesian languages, with their sparse morphology and their consequent dearth of morphological irregularity, are often diachronically unrevealing languages of just this kind.

My principle when asked to define this matter is that only linguistic fabric – material that has morphemic substance, such as lexicon and derivational and inflectional morphology, can be used to trace genetic affinities between languages. This is the same classic position which Antoine Meillet embraced (Meillet 1921, 1925) and there is no reason to abandon it. In contrast to this, the characteristics of phonology, morphological processes rather than morphological forms, syntactic patterns at phrase-, clause-, sentence- or paragraph-level, and the structure of semantic fields, are not usable in attempts to prove genetic affinity. However, such patterns are invaluable for filling in features of the history of a language after its speakers have begun to separate from any other bodies of speakers of the same language. For instance, one cannot be said to transmit syntactic patterns genetically within a language in the same way as we can observe that a lexical morph is transmitted from generation to generation of speakers.

Furthermore, there are a limited number of possible orders for subject-verb-object strings (and some of these six possible orders are rarely used or encountered in the world's languages, which reduces even more the choice or possibility of different orders being used in two or more languages being compared). As a result, the fact that two adjacent languages shared one of these six basic constituent orders is of little moment in classifying them genetically, and it tells us nothing about a language's genetic history, although the fact of a language's typological affinity may be more illuminating about its contact history.

In a context such as this one George Grace's concepts of 'aberrancy' and 'exemplariness' (which were discussed for instance in Grace 1990) come into play in an interesting way. The terms are of course relative ones rather than absolutes, but nonetheless it is possible for us to invoke and utilise these concepts quite fruitfully in this investigation, after one has interrogated the materials in Proto-Chamic and on the subgroups from which Proto-Chamic has evolved. (The chief point of reference here is of course the reconstruction work on Proto-Austronesian and its daughter languages which has been carried out by Robert Blust, reconstructed forms from whose ongoing work are extensively cited in Thurgood's works. Languages which are 'exemplary', it is implied, would have a lot to contribute to the reconstruction of a proto-language, and furthermore, the process of incorporating and demonstrating these findings is assumed to be simpler to carry out if one is using 'exemplary' language data. Aberrant languages are rarely also languages which are full of archaic features; rather, they tend to have retained plenty of well-known features which' are well attested in other languages but which happen to have evolved in startlingly anomalous ways in the particular aberrant language under scrutiny. And it need hardly be said that two aberrant languages may manifest their aberrancies by bringing about changes, often even on the same morphs or sounds, which have gone in very different directions both from the ancestral language and from one another.).

It is therefore fortunate that Thurgood examined developments in Chamic from a 'top down' perspective, since this approach enables one to seem more clearly, and to demonstrate more forcefully, the paths of development both of Chamic as a unit and of individual Chamic languages. The extent to which this large degree of historical revelation would have been possible from the employment of a bottom-up approach, something which would have involved investigators piecing Proto-Chamic together from the evidence of modern languages and then tying it into further relationships within Austronesian, is something of a matter for wonder.

Of course aberrancy can occur at several levels in a language, and it often does. A language is often aberrant in several respects all at once. It is the combination and constellation of aberrancies at several levels and in several parts of a language (though especially those which relate to the perpetuation of actual morphs) which makes some languages stand out, and which makes them of minimal use in the task of reconstructing proto-languages. On the other hand, aberrancies in a language are supposed to be unravellable and explicable in terms of the structure of the proto-language as we know them. Aberrancies are not themselves caused by the possession in a language of features which otherwise are not allowed for in the reconstruction of the proto-language, and which therefore have to be incorporated into the structure of the proto-language, even if the language possessing such archaisms isanomalous in other ways when compared with the rest of the family. (The existence of laryngeals in Anatolian languages, for example, was unusual among Indo-European languages, but this did not make them aberrant in terms of Indo-European languages, because the possession of laryngeals provided information about the structure of Indo-European which had previously been largely unavailable. Laryngeals, after all, were a feature of an earlier stage of Indo-European and one that had largely been lost from other Indo-European languages, although the effects of their loss were not the same in all Indo-European languages. But the small proportion, and indeed the small amount, of lexicon in our admiitedly imperfect and gap-riddled records of Anatolian languages which can be traced to Indo-European makes them seem much more aberrant.)

A language such as Cat Gia Roglai is aberrant in the light of Proto-Austronesian in terms of its segmental, suprasegmental and canonical phonology, its (paucity of) inflectional and derivational morphology, and also because of the small amount of Malayo-Chamic items in its lexicon (which themselves make up only a small part of the total Cat Gia Roglai lexicon). Many of these changes date from Proto-Malayo-Chamic and ar especially shared with other Chamic languages, others (for example the major syntactic patterns and some of the lexicon) date from Proto-Chamic, and yet others have entered (or have developed within) the language over the last millennium. This is especially the case with those phonological and other changes which do not appear to be externally-motivated inasmuch as they are not paralleled by the presence of the same changes in languages which are known to ave been in contact with (and to have influenced) some or all Chamic languages.

We may compare the contact-driven aberrancy of Chamic with the internally-driven aberrancy of Nauruan, a Micronesian (though not Nuclear Micronesian) language which has undergone sweeping and often unique phonological changes in tandem with large-scale lexical replacement both by borrowing (apparently from Kiribatese in the period preceding European contact, and latterly from English), by compounding in many cases where other languages use monomorphemic words, and by circumlocution (Nathan 1973). These changes presumably happened to Nauruan at the same time as it elaborated certain features of its structure, such as the 39 separate sets of numerals which it developed for use with specific types of nouns (a feature now in decline).

Closer to Chamic, both geographically and genetically, we have the case of Kerinci of Sumatra, a language which is very similar to Minangkabau (and thus to Malay), to which it is clearly also very closely related, but which has undergone a number of striking phonological changes. These changes have not been brought about as the result of heavy contact by speakers of Kerinci with external linguistic forces (both groups are Muslim, for instance), but instead they are internally driven (and for that matter, they are rule-governed). And these changes are not paralleled by equally sweeping changes in the phonology of the very closely related Minangkabau (this case is discussed in Prentice and Usman 1978, while further sound-changes in Kerinci are discussed and exemplified in Steinhauer 2002.) Such aberrancy in the historical phonology of Kerinci is not paralleled by any similar aberrancy of, say, basic Kerinci lexicon or morphology from the viewpoint of Minangkabau, Some cognates in Kerinci which are historically related to forms which are also found in Minangkabau are hard to recognise at first, because of the effects of multiple cyclically-applied sound-changes on the original Kerinci forms, but they are cognates noneteheless.[6] (In this respect it is similar to Cat Gia Roglai or to Tsat.) But even so, the sound-changes which have taken place in Kerinci are not as dramatic as those which characterise many Chamic languages, and it appears to contain little vocabulary (or bound morphology) which is alien to Minangkabau. And there are other examples in Austronesian of clusters of co-occurring internally-motivated innovations which have combined to make certain languages seem hard to classify.

Grace (1990: 109-110) pointed out the near-impossibility of reconstructing Proto-Austronesian, or indeed of inferring shared genetic affinity, from three aberrant languages such as the Formosan language Atayal (in which the aberrancy is not caused by borrowing), Yapese (in which borrowing has played a large part in making the language seem aberrant, though this is far from being the entire explanation) and a language of Southern New Caledonia, each of which are aberrant in their own different ways. Such aberrancy is also found in Chamic languages such as Rade and especially Tsat. Tsat, Nauruan and a Formosan language such as Tsou would be another trio of languages for which a common origin would be very difficult to prove, while the subsequent task of reconstructing any inferred proto-language based solely upon evidence from these three languages would face insuperable problems. In contrast, a language such as Malay is much more ‘exemplary’, at least on phonological and lexical levels, than Tsat or Cham (or maybe even than Tagalog, with its relatively low proportion of inherited Proto-Malayo-Polynesian vocabulary), even if it has shed or fossilised much of the heritage of Proto-Malayo-Polynesian bound morphology that Tagalog (or instance) retained.

What is interesting and significant, of course, is the fact that we know something of the internal and external histories of Chamic languages. We know that Chamic languages have developed in their wide variety of atypically Austronesian ways from a language which looked a lot like the language which has given rise to the various forms of Malay (although it seems to have been somewhat more innovative in terms of its phonological development). We know that this diversity of development has happened as a result of the effects of various waves of contact, we know that this change was effected in large measure, at least at first, by the gradual spread of a number of phonological rules, and we know something of how they may have looked, say, 2200 years ago, how they did look 1600 and even 1100 years ago, as well as 500 years ago. We can do the latter investigation courtesy of the data in Edwards and Blagden (1940-1942), despite the numerous philological problems inherent in extrapolating from the Chinese transcription which it

[6] Indeed Blust (1981) shows that Kerinci has retained 100 out of the 200 PMP forms that Blust reconstructed on his list, which makes it one of the most lexically conservative Austronesian languages of all

used, and we can compare the forms which the Chinese vocabulary uses with how they look now. The blend of Mon-Khmer and Malayic elements (and indeed of common Chamic elements of unidentified origin) that are to be found in that vocabulary shows that the absorption and full integration, into an as yet undivided Cham, of basic Mon-Khmer elements, complete with their phonological characteristics (inasmuch as this can be conjured out of the clues provided by the Chinese character transcription), had already taken place more than half a millennium ago and had probably occurred much earlier. It also suggests that Cham proper no longer borrows from the (Bahnaric) Mon-Khmer languages which originally wrought such great changes on its lexicon and structure; others, especially Vietnamese, have taken their place as the major sources of external loans.

The historical continuity between these various forms of Chamic languages is quite clear, even though the individual changes which are demonstrated are often striking. And we should remember that the same changes in Chamic languages have sometimes occurred independently more than once. This is especially clear in the sphere of segmental and canonical phonology. For instance both Haroi and Northern Roglai have developed batteries of nasalised vowels in the course of the period of their development which began after the break up of Proto-Chamic, but since they are separated geographically by Chamic languages and by other languages which have not evolved these, they have done this independently of one another. Other changes have operated more as the result of drift, for instance the gradual loss of /-p/ (replaced word-finally by zero) which has occurred in most Indochinese Chamic languages apart from Rade and Jarai. (This is a change which cuts across linguistic boundaries between the highlands and the coast: Headley 1991.).

And between the testamentary power of the materials in Inscriptional Cham (which includes the first data ever written down in any Austronesian language), the literary material in written Cham, unwritten material in the two varieties of modern Cham and in the offshoot Haroi, and the evidence of Acehnese, both earlier and more modern, and the evidence of Tsat (not to mention the evidence from modern Malay lects), we can adduce a great deal more about the history and courses of development of Chamic languages than one might expect.

5. Conclusions, and some priorities for further research.

As the result of two millennia of linguistic contact with Mon-Khmer languages (contact which has picked up strongly in the last millennium after the decline in power of the Cham empires), and with concomitant separation from their Malayic kin, the Chamic languages have absorbed more overt features (such as lexical loan elements, including those which replaced previously-existing words for long-familiar concepts) and more typological characteristics (including a whole range of phonological features which are highly marked in terms of their occurrence in the world's languages) from the languages of their immediate Mon-Khmer-speaking neighbours. Many of these Mon-Khmer speakers, especially those who were speakers of 'small' and territorially-constrained languages, may have come to be dominant in the ancestral form of modern Cham, which they acquired chronologically as a second language but which they used more frequently than their native, ethnic or first language. This absorption of elements has been taking place at least since the ninth century and probably since a much earlier period (if we are to judge by the fair number of Mon-Khmer elements which are to be found in Acehnese, a language absent from Indochina since at least the eleventh century, and which are common to other Chamic languages). The overall effects of various Mon-Khmer languages upon assorted Chamic languages (an issue which is discussed in an excellent paper, Sidwell 2002) are summarised in Table 7.

Table 7: *A table of Mon-Khmer languages and language groups which have influenced individual Chamic languages.*

Language /source of elements	Earlier Bahnaric languages	Bahnar proper	Hrê	Khmer	Vietnamese
Malayic	No	No	No	A handful of forms	No
Proto-Chamic	Yes	Uncertain	Unlikely	Unlikely	No
Acehnese	Yes	No	No	No, unless there were some widely distributed loans	No
Tsat	Yes	No	No	No	No
Rade	Yes	No	No	No	Later on
Jarai	Yes	No	No	No	Later on
Northern Roglai	Yes	No	No	No	Later on
Haroi	Yes	Yes, much	Yes, much	No	Later on
Written Cham	Yes	Yes, a little	No	Yes?	No?
Western Cham	Yes	Yes, a little	No	Yes, plenty	(yes, but recently and in Mekong Delta variety)
Phan Rang Cham	Yes	Yes, a little	No	(maybe yes, if it includes forms inherited from pre-1471 Cham)	Yes

Chamic contact with other Mon-Khmer languages continues apace, and features from these are still being transferred into Chamic languages, and especially into the lexicon and the segmental phonology. This transferral of such material into Chamic languages has been aided by the fact that Mon-Khmer languages had minimal affixal morphology which might impede the transfer of elements, especially verbs and free grammatical morphs, to Chamic languages. There were few typological barriers which might inhibit the transferral of just about any kind of Mon-Khmer morph into a language such as early Cham, in which there was little affixal morphology as much of it had dropped away.

Consequently Chamic languages are highly atypical when compared with Western Malayo-Polynesian languages, but they bear a strong typological similarity at many levels to Bahnaric languages. One might misuse metaphors from another scientific field and say that in terms of their morphemic mitochondrial DNA the Chamic languages are Austronesian, but according to their adaptations and typological e-fits they are very much like Mon-Khmer languages. And they are more like Mon-Khmer languages in this respect than they are even like Malay, which itself has been brought (partly by chance, partly through the imitation of certain salient features such as numeral classifiers) into the fringes of the Southeast Asian typological network.

Another important factor in the typological approximation of Chamic languages to the salient features of their Mon-Khmer neighbours is the gradual loss in Chamic languages of most of the productively-employed bound morphs which had been attested in Proto-Malayo-Chamic and which have been retained in some conservative forms of Malay. Although the loss of morphology is a negatively-weighted feature in typological terms since by its very nature it does not involve the transfer of morphs, and is therefore of very limited heuristic value in assessing the depth of language contact, such loss (which is an areal feature and which predates intensive Mon-Khmer contact) has been an important consequence of, and has acted as an aid to, contact between Chamic and Mon-Khmer languages.

The overall result is that the Chamic languages have come to resemble Mon-Khmer languages ever more closely in terms of their phonological systems and phonotactics (and in terms of their suprasegmentals, in those cases where Chamic languages were in touch with tonal languages) and also their syntax. This increasing similarity can be shown to have occurred in several stages over time, but also to have been quite obvious by c. 1000 AD. In addition, these languages have acquired or developed a surprisingly large proportion of lexical elements (belonging to most form classes) which have yet to be supplied with etymologies, although many of these show phonological characteristics (including borrowed segments) which are typically Mon-Khmer and which are atypical of Austronesian languages and of the PMP stratum which provides the genetic background of Chamic.

The two Chamic languages which departed Indochina (both quite early) and which therefore missed out on the secondary waves of the effects of the influence of Mon-Khmer languages, namely Acehnese and Tsat, have gone in separate directions as regards languages which they have been in contact with, and phonological developments. Acehnese has preserved much of the structure of 10th century Cham (and quite a bit of its lexicon, including numerous forms of uncertain or Mon-Khmer origin which provide important historical evidence for the historical development of the language). But in the past several centuries it has also absorbed much lexicon from Malay, some of which will have replaced Cham-internal developments and Mon-Khmer loans which were present in earlier stages of Acehnese.

On the other hand, Tsat has come to resemble Hainanese Chinese (and latterly Mandarin Chinese: Thurgood to appear, c) more and more in terms of its segmental, suprasegmental and canonical phonology, as well as in the formation of certain kinds of noun phrases such as those involving demonstratives or possession.

The evidence of certain features of Acehnese morphology and lexicon makes it clear that the Chamic languages emerged from a language which had retained some of the complex inflectional patterns of earlier Malayic languages (for instance the productive use of infixation). In addition this language had absorbed many features of all kinds from Mon-Khmer languages, and had acquired later (mostly lexical) developments that were post-Mon-Khmer and exclusive to Chamic languages. The language which was ancestral to Indochinese Chamic, Tsat and Acehnese was probably more complex morphologically than the Malayo-Chamic proto-language because it had acquired many new features through borrowing and had retained many others; its descendants were to lose many of these features, of whatever origin. What happened then is that different Chamic languages shed different structural features from this amalgam, usually as the result of areal influence from the more powerful languages which shaped them. For instance Acehnese did not acwuire infixation from Malay, because Malay no longer had it to give. In this instance Acehnese had tretained something that fell into greater and greater disuse in other Chamic languages.

There is much work still to be done on Chamic languages, and the amount of time to do it may not be as long as we think. We may enumerate some tasks for the future in regard to diachronic (and also synchronic) Chamic language research. These include (but are not restricted to):

1) Integration into Chamic studies of the new findings about Proto-South Bahnaric and Proto-West Bahnaric (and Proto-Bahnaric) reconstructions, in an attempt to reduce the sizeable number of items of 'unknown' origin in Proto-Chamic and in the sublevels beyond.

2) the creation of more grammatical descriptions and more widely-available text collections and lexica of Chamic languages, these being needed especially strongly for Chru and Southern Roglai, though all Chamic languages warrant being described more fully, given the patchy if often excellent material available.

3) More integration is needed with work that has been carried out on the reconstruction of various levels of Austronesian. How many Proto-Malayo-Polynesian elements which are NOT Malay loans but which are directly inherited elements occur in any or all Chamic languages? How many other attested post-Proto-Malayo-Polynesian forms are exclusive to Malayic languages and Chamic languages? Are there any Austronesian forms which are found in Chamic languages but not in other Malayic ones, and if so, what are the heuristic significances of these forms? Are they Austronesian or at least Western Austronesian retentions in Chamic which have been replaced by loans or internally-coined forms in Malayic? Are there any post-Proto-Western Austronesian forms in Chamic languages that are also not found in Malayic? (Probably not.)

4) More work needs to be done on the Austronesian and especially on the Chamic components in what geographically may be classed as (non-Chamic) Vietnamese languages, especially on those which are found in the Vietnamese Mon-Khmer language Katu (which is supposed to contain some morphological material from Austronesian languages, at least according to Reid 1994).

5) More work could be done on the analysis of that stratum of forms which is common to most or all Chamic languages (including Acehnese) but which is of unidentified origin.

6) We require a diachronic examination of Cham structure, lexicon and phonology, from 350 AD onwards, using the inscriptional, classical and modern written and dialectal data; such a longitudinal examination is a unique opportunity to be taken in Austronesian historical linguistics.

7) Further work could be done on analysing the dialectology within Cham, on seeing what genetic justification there may be for positing Highland and Coastal Chamic divisions, and on understanding where Chru, Roglai varieties and Haroi fit into this picture.

References

Adelaar, Karl Alexander. 1992. *Proto-Malayic: the reconstruction of its phonology and parts of its lexicon and morphology*. Pacific Linguistics C-119. Canberra: Australian National University.

Alieva, Natalia F. 1984. 'A language union in Indo-China.' *Asian and African Studies* [Bratislava] XX: 11-21.

1992. 'Malay and Cham possession compared.' *Oceanic Linguistics* 30: 73-91.

---, and Bui Khanh The. 1999. *Yazyk Cham: ustnye govory vostochnogo dialekta*. St Petersburg: Institute for Oriental Studies.

Aymonier, Etienne. 1889. 'Grammaire de la langue chame.' *Excursions et reconaissances* XIV 31: 1-92.

Aymonier, Etienne, and Antoine Cabaton. 1906. *Dictionnaire-cham-français*. Paris: Leroux. (Publications de l'Ecole française de l'Extrème-Orient 7).

Bakker, Peter, and Maarten Mous (eds.). 1994. *Mixed languages: 15 case studies in language intertwining*. Amsterdam: IFOTT.

Baumgartner, Neil I. 1998. 'A grammar sketch of Western (Cambodian) Cham.' Thomas (ed., 1998), 1-20.

Blood, David L. 1962. 'A problem in Cham sonorants.' *Zeitschrift für Phonetik* 15: 111-114.

Blood, Doris Walker. 1962. 'Reflexes of Proto-Malayo-Polynesian in Cham.' *Anthropological Linguistics* 4 (9): 11-20.

-----1978. 'Some aspects of Cham discourse structure.' *Anthropological Linguistics* 20: 110-132.

Blust, Robert A. 1973. 'The origins of Bintulu *b', d'*.' *Bulletin of the School of Oriental and African Studies*, 47: 603-620.

-----1988. *Austronesian root theory*. Philadelphia: John Benjamins. (Studies in Language Volumes.)

-----1990a. 'Malay historical linguistics: a progress report.' *Rekonstruksi dan cabang cabang bahasa Melayu induk*, edited by Mohd. Thain Ahmad and Zaid Mohamed Zaidi, 1-33. Kuala Lumpur: Dewan Bahasa dan Pustaka.

-----1990b. 'Patterns of sound change in the Austronesian languages.' *Patterns of change, change of patterns*, edited by Philip Baldi, 129-163. New York and Berlin: Mouton de Gruyter.

-----1992. 'The Austronesian settlement of mainland Southeast Asia.' Karen L. Adams and Thomas John Hudak (eds.), *Proceedings of the Second Annual Meeting of the South East Asian Linguistics Society*, 25-83. Tempe: Arizona State University Press.

-----1993a. 'Central and Central-Eastern Malayo-Polynesian.' *Oceanic Linguistics* 32: 243-292.

-----1993b. 'On speech strata in Tiruray.' *Papers in Western Austronesian Linguistics*, edited by Hein Steinhauer, 1-52. Pacific Linguistics A-91. Canberra: Research School of Pacific and Asian Studies, Australian National University.

-----1999. 'Notes on Pazeh phonology and morphology.' *Oceanic Linguistics* 38: 121-165.

-----2000. Review of Thurgood 1999. *Oceanic Linguistics* 39: 435-445.

-----2000a. 'Chamorro Historical Phonology.' *Oceanic Linguistics* 39: 83-121.

Bochet, Gilbert, and Jacques Doumes. 1953. *Lexique polyglotte: Vietnamien, koho, roglai, français*. Saigon: Editions France-Asie.

Bradshaw, Joel. 1995. 'How and why do people change their languages?' *Oceanic Linguistics* 34. 191-201.

Bukhari, Daud, and Mark Durie. 1999. *Kamus Basa Acèh - Kamus Bahasa Aceh – Acehnese-Indonesian - English Thesaurus.* Pacific Linguistics: C-151. Canberra: Australian National University.

Campbell, George L. 2000. *Compendium of the world's languages*. London: Routledge.

Collins, [Ira] Vaughn. 1969. 'The position of Atjehnese among Southeast Asian languages.' *Mon Khmer Studies* 3: 48-60.

Cowan, H. K. J. 1948. 'Aanteekeningen betreffende de verhouding van het Atjehsch tot de Mon-Khmer talen.' *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië*. 104: 429-514.

-----1981. 'An outline of Atjehnese phonology and morphology.' *Bulletin of the School of Oriental and African Studies* 54: 522-549.

Diffloth, Gérard. 1991. 'Vietnamese as a Mon-Khmer language.' *Papers from the first annual meeting of the South-Easr Asian Linguistic Society*, edited by Eric Schiller and Martha Ratliff, 125-139. Tempe: Northern Arizona University.

Diffloth, Gérard and Norman H. Zide (eds.). 1976. *Austroasiatic numeral systems.* (= *Linguistics* 174.)

Đình-Hoà, Nguyễn. 1997. *Vietnamese.* London: School of Oriental and African Studies.

Durie, Mark. 1990. 'Proto-Chamic and Acehnese mid-vowels: Towards Proto-Aceh-Chamic.' *Bulletin of the School of Oriental and African Studies* 53: 100-114.

Dyen, Isidore. 1946. 'Malay *tiga* '3''. *Language* 22: 131-137.

-----1953. 'Malay *tiga* '3' once again.' *Language* 29: 465-473.

-----1962. 'The lexicostatistical classification of the Malayopolynesian languages.' *Language* 38: 38-46.

-----1973. 'The Chamic languages.' *Current Trends in Linguistics, volume 8*, edited by Thomas A. Sebeok, 110-120.

-----2000. Review of Thurgood (1999). *Anthropological Linguistics* 43: 390-394.

Edwards, Evangeline Dora, and Blagden, Charles Otto. 1940-1942. 'A Chinese glossary of Cham words and phrases.' *Bulletin of the School of Oriental and African Studies* 10: 53-91.

Egerod, Søren. 1978. 'An English-Rade Vocabulary.' *Bulletin of the Museum of Far Eastern Antiquities* 50: 49-106.

-----1980. 'To what extent can genetic-comparative classifications be based on typological considerations?' *Typology and Genetics of Language: Proceedings of the Rask-Hjelmslev Symposium held at the University of Copenhagen, 3rd-5th September 1979*, edited by Torben Thrane, Vibeke Winge, Lachlan Mackenzie, Una Canger and Niels Ege, 115-139. (Travaux du Cercle Linguistique de Copenhague XX). Copenhagen: Linguistic Circle of Copenhagen.

Forman, Michael L. 1972. *Zamboangueño texts with grammatical analysis.* Unpublished Ph. D. dissertation, Cornell University.

Grace, George W. 1990. 'The 'exemplary' (vs. 'aberrant') Melanesian languages.' *Patterns of change, change of patterns*, edited by Philip Baldi, 109-127. New York and Berlin: Mouton de Gruyter.

Grant, Anthony P. 2005. 'Norm-referenced lexicostatistics and the case of Chamic.' 43 pp. In this volume.

Grimes, Barbara F. (ed.). 2000. *Ethnologue, fourteenth edition.* Dallas: Summer Institute of Linguistics.

Hamilton, A. W. 1997. *Easy Malay vocabulary: 1001 Essential Words*. Singapore: Times Publications.

Haudricourt, André-Georges. 1966. 'The limits and connections of Austroasiatic in the northeast.' *Studies in Comparative Austroasiatic Linguistics*, edited by Norman H. Zide, 44-56. The Hague: Mouton.

-----1984. 'La tonologie des langues de Hai-Nan'. *Bulletin de la Société linguistique de Paris* 79 (1): 385-394.

Headley, Robert A. 1976. 'Some sources of Chamic vocabulary.' *Austroasiatic studies,* edited by Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta, 453-476. Honolulu: University Press of Hawai'i.

-----1991. 'The phonology of Kompong Thom Cham.' *Austroasiatic languages*: *essays in honour of H. L. Shorto*, edited by Jeremy H. C. S. Davidson, 105-121. London: School of Oriental and African Studies.

Heath, Jeffrey. 1984. 'Language contact and language change.' *Annual Review of Anthropology* 13: 367-384.

Henderson, Eugenie J. A. 1965. 'The topography of certain phonological and morphological characteristics of certain South-East Asian languages.' *Lingua* 15: 400-434.

-----1980. 'Discussion'. *Typology and Genetics of Language: Proceedings of the Rask-Hjelmslev Symposium held at the University of Copenhagen, 3rd-5th September 1979*, edited by Torben Thrane, Vibeke Winge, Lachlan Mackenzie, Una Canger and Niels Ege, 145-152. (Travaux du Cercle Linguistique de Copenhague XX). Copenhagen: Linguistic Circle of Copenhagen.

Himly, K. 1890. 'Sprachvergleichende Untersuchungen des Wörterschatzes der Tscham-Sprache'. *Sitzungsberichte der kaiserlichen-königlichen Akademie der Wissenschaften* 322-456.

Hudson, Alfred B. 1967. *The Barito dialects of Borneo, a classification based on comparative reconstruction and lexicostatistics*. Ithaca, New York: Cornell University, Department of Asian Studies.

Jacob, Judith M. 1966. *Introduction to Cambodian.* Cambridge: Cambridge University Press.

Jacq, Pascale, and Paul J, Sidwell. 2000. *A Comparative West Bahnaric Dictionary.* London and Munich: Lincom-Europa.

Kouwenberg, Silvia. 1994. *A grammar of Berbice Dutch Creole.* New York and Berlin: Mouton de Gruyter.

Lafont, Pierre-Bernard. 1968. *Lexique jarai parler de la province de plei ku.* Paris: Publications de l'Ecole Française de l'Extrème-Orient LXIII.

Larish, Michael D. 2005. 'Moken-Moklen.' *The Austronesian Languages of Asia and Madagascar*, edited by Alexander Adelaar and Nikolaus Himmelmann, 513-533. London: Curzon.

Lee, Ernest Wilson. 1974. 'South East Asian areal features in Austronesian strata in Chamic.' *Oceanic Linguistics* 13: 643-670.

-----1996. 'Bipartite negatives in Chamic.' *Mon-Khmer Studies* 26: 291-317.

-----1998. 'The contribution of Cat Gia Roglai to Chamic.' Thomas (ed. 1998), 31-54.

Le Page, Robert B. and Andrée Tabouret-Keller. 1984. *Acts of identity*. Cambridge: Cambridge University Press.

Marck, Jeffrey H. 2000. *Topics in Polynesian language and culture history*. Pacific Linguistics 504. Canberra: Australian National University.

Marrison, Geoffrey E. 1975. 'The early Cham language and its relationship to Malay.' *Journal of the Malaysian Branch of the Royal Asiatic Society* 48: 2: 52-59.

Meillet, Antoine. 1921. *Linguistique historique et linguistique comparée*. Paris: Champion.

1925. *La méthode historique en linguistique historique*. Oslo: Aschehoug.

Moussay, *Père* Gérard. 1971. *Lexique cam-vietnamien-français*. Phan Rang (Vietnam): Centre culturel cam.

Nathan, Geoffrey S. 1973. 'Nauruan in the Austronesian language family.' *Oceanic Linguistics* 12: 479-501.

Niemann, K. G. 1891. 'Bijdrage tot de Kennis der Verhouding van het Tjam tot de Talen van Indonesië.' *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 40: 27-44.

Pang, Keng-Fong. 1998. 'On the Ethnonym 'Utsat.' Thomas (ed.), 55-60.

Peiros, Ilia. 1996. *Katuic Comparative Dictionary*. Pacific Linguistics C-132. Canberra: Australian National University.

Pittman, Richard S. 1959. 'Jarai as a member of the Malayo-Polynesian family of languages.' *Asian Culture* 1 (4): 59-67.

Prentice, D. John, and A. Hakim Usman. 1978. 'Kerinci sound changes and phonotactics.' Stephen A. Wurm and Lois Carrington (eds.), *FOCAL 1: Proceedings of the Second International Conference on Austronesian Languages, Volume 1: Western Austronesian*, 121-163. Canberra: Australian National University.'

Reid, Lawrence A. 1994. 'The morphological evidence for Austric.' *Oceanic Linguistics* 33: 323-344.

Ross, Malcolm. 1996. 'Contact-induced change and the comparative method: cases from Papua New Guinea.' *The comparative method reviewed: regularity and irregularity in language change*, edited by Mark Durie and Malcolm Ross, 180-218. New York: Academic Press.

Sebeok, Thomas A. 1942. 'An examination of the Austroasiatic language family.' *Language* 18: 206-217.

Shintani, Tadahiko L. A. 1981. 'Etudes phonologiques sur la langue Rhadé (I)'. *Journal of Asian and African Studies* 21: 120-129.

Sidwell, Paul J. 2000. *Proto South Bahnaric*. Pacific Linguistics 501. Camberra: Australian National University.

-----, 2002. The Mon-Khmer substrate in Chamic, and the history of Chamic, Bahnaric and Katuic. Paper presented at the XIIth Annual Meeting of the South East Asia Linguistic Society, Tempe, Arizona.

Simons, Gary. 1982. 'Word taboo and comparative Austronesian linguistics.' Amran Halim, Lois Carrington and Stephen A. Wurm (eds.), *Accent on variety: Papers from the Third International Conference on Austronesian Linguistics*, 157-226. Canberra: Pacific Linguistics C-76.

Steinhauer, Hein. 2002. 'More (on) Kerinci sound-changes.' K.Alexander Adelaar and Robert Blust (eds): *Between Worlds: linguistic papers in memory of David John Prentice*, 149-176. Canberra: Pacific Linguistics 529.

Tharp, James and Y-Bhăm Buôn Yă. 1980. *A Rhade-English dictionary with English-Rhade finderlist.* Pacific Linguistics C-58. Canberra: Australian National University.

Thomas, David D. 1971. *Chrau grammar.* Oceanic Linguistics Special Publications 8. Honolulu: University of Hawaii Press.

-----, (ed.) 1998. *Studies in Southeast Asian languages no. 15: Further Chamic studies.* Pacific Linguistics A-89. Canberra: Australian National University.

-----, Ernest W. Lee and Nguyen Dang Liem (eds.). 1977. *Papers in Southeast Asian languages, no. 4. Chamic studies*. Pacific Linguistics A: 48. Canberra: Australian National University.

Thomason, Sarah Grey, and Terrence Kaufman. 1988. *Language contact, creolization and genetic linguistics.* Berkeley and Los Angeles; University of California Press.

Thurgood, Graham. 1992. 'From atonal to tonal in Utsat (a Chamic language of Hainan).' *Proceedings of the Eighteenth Annual Meeting of the Berkeley Linguistics Society. Special Session on the Typology of Tone Languages*, edited by Laura A. Buszard-Welcher, Jonathan Evans, David Peterson, Lionel Wee and William F. Weigel, 145-156. Berkeley: Berkeley Linguistic Society.

-----1996. 'Language contact and the directionality of internal drift: the development of tones and registers in Chamic.' *Language* 71: 1-31.

-----1997. 'Restructured register in Haroi: reconstructing its historical origins.' *Southeast Asian Linguistic Studies in Honour of Vichin Panupong*, edited by Arthur S. Abramson, 283-295. Bangkok: Chulalongkorn University Press.

-----1999. *From ancient Cham to modern dialects: two thousand years of change.* Oceanic Linguistics Special Publication 28. Honolulu: University Press of Hawai'i.

-----2005. 'A preliminary sketch of Phan Rang Cham.' *The Austronesian Languages of Asia and Madagascar*, edited by Alexander Adelaar and Nikolaus Himmelmann, 489-512. London: Curzon.

-----To appear a. 'Learnability and direction of convergence in Cham: the effects of longterm contact on linguistic structures.' Manuscript, 21 pp.

-----, and Fengxiang Li. To appear. 'Contact-induced variation and syntactic change in the Tsat of Hainan.' Manuscript, 15 pp., to appear in *Proceedings of the Berkeley Linguistic Society.*

Zheng Yiqing. 1997. *Huihuihua yanjiu.* Shanghai: Shanghai Yuandong Chuban She.

Zorc, R. David. 1974. 'Towards a definitive Philippine wordlist: the qualitative use of lexicon in identifying and classifying languages.' *Oceanic Linguistics* 13: 409-455.

Appendix.

An approximate and partial chronology of major phonological and other contact-induced changes giving rise to phenomena in chamic languages, drawing upon Thurgood (1999).

c. 100BC +/- 100 years. Proto-Chamic splits from Proto-Malayic (as the term is used in the broader sense) or from Proto-Malayo-Chamic, on the occasion when the speakers of Proto-Chamic move to the Indochinese mainland: The Proto-Chamic language is structurally, typologically and lexically similar to Proto-Malayic, its closest relative, and in many respects is little different from what has been constructed for Proto-Austronesian. It has four vowels, a basically disyllabic and occasionally trisyllabic word-structure with a generally penultimate stress pattern, an embargo against initial consonant clusters and with a restriction upon the nature and kinds of medial consonant clusters which are permitted morpheme-internally, and a small battery of bound morphological items including prefixes, infixes and some suffixes.

Since the Mon-Khmer lexical elements in Malay mostly differ from those in Chamic (the few exceptions may be loans which were transmitted from Malay into Chamic, or which were borrowed separately in each language), we may assume that the latter language was a *tabula rasa* at this time as far as Mon-Khmer loans were concerned. (Although some of the Sanskrit loans in Malay are also shared with Chamic and especially with written Cham, not to mention Khmer, this is more because in both languages they are cultural borrowings taken over to express innovations than for any diachronic reason). On the other hand, several lexical, phonological and other innovations which are common to Malayic and Chamic languages and which mark them off from other Western Malayo-Polynesian languages will have been formed by this time. Proto-Malayic or Pre-Malayic **q* consistently became /h/ in Malay and Cham (though it did not do so in **qaqay* 'leg', where it became /k/ in both instances (Malay *kaki* and Cham *kakey* 'leg') and in both languages), but it became /k/ in the Moken and Moklen language of the Mergui Archipelago, Burma, and of surrounding islands belonging to Thailand; this pair of languages is another displaced Malayic offshoot (Larish 2005).

After this period the list of items of Austronesian or Proto-Malayo-Chamic origin is closed for the rest of the course of the development of the Chamic languages. Therefore the reservoir of Proto-Malayo-Chamic morphs is to be seen as the source of all forms of Austronesian origin in these languages except in the case of those languages (such as Acehnese, and to some extent written Cham) which have had later connections with Malay.

2) c. 350 AD. A Chamic language is first recorded in the period before dialectal diversity. Inscriptional Cham is noted down, apparently in the 4th (in one short bilingual inscription) and latterly in the 9th centuries AD, the latest one which has been securely dated being carved in 1401 (though there may be some later ones which are undated). This material (at least that which is provided in Marrison 1975 and which was reproduced in Thurgood 1999: 3) shows that the process of contraction (and indeed in some cases the deletion) of the first vowel in disyllables had already taken place in many words by the 4th century. This is especially the case when the first vowel is schwa (this contraction predates a similar contraction in Malay varieties) or /a/, and this contraction happened when the resulting consonant cluster was easily pronounceable. /i/ and /u/ were still retained in many words. (Rade and Jarai later deleted the first vowel of disyllables in all cases, producing many more initial two-member, and in the case of Rade often three-member, consonant clusters.)

The morphosyntax of the inscriptional language (certainly that of the 4th century inscription) is characterised by an absence of bound inflectional morphs, although free grammatical morphs abound, many of them being shared with Malay such as the relative clause marker *ya* (compare Malay *yang*, a form which combines PMP **ia* 'he, she' and *ang 'focus marker'. The first inscription in Old Malay is a few centuries younger than the oldest Cham inscription (the date on it is 683), but has preserved more morphological features than the Chamic inscription has. The lexicon of these inscriptions contains a large amount of Sanskrit material, some elements of which later passed to the spoken Chamic languages, although most of this did not pass further (except into classical written Cham), and in any case the mode of expression of these inscriptions follows Indic formulaic patterns. Many of the later ones, which come from the ninth century onwards, contain lexical elements of Mon-Khmer origin. There are some 75 such inscriptions. It is possible that the merger of /n-/ into /l-/ word-initially in Chamic is a reflection of a similar phonemic merger which is to be found in some southern Vietnamese Mon-Khmer languages, but we cannot be sure; in any case /n-/ was rare to begin with. The source of Mon-Khmer influence at this time is probably Bahnar, a Bahnaric language spoken in southern Vietnam which has itself already undergone some influence from the Katuic languages (Paul Sidwell, p. c.), which are situated to the north of Bahnaric languages and which belong to a separate branch of Mon-Khmer.

3) After 982 AD. Acehnese splits from Chamic. Acehnese has been said (Thurgood 1999) to have a larger proportion of elements from Katuic languages than other Chamic varieties have, and its earlier form was probably the most northern variety on the Chamic dialect chain. The externally-motivated separation of Acehnese from the other Chamic languages (the result of attacks from the north) may have been the catalyst for the gradual unravelling of the Cham dialect chain, much as when, in the history of Polynesian, the departure of Maori-speakers for Aotearoa/New Zealand may have actuated the split up of Proto-Tahitic (Marck 2000: 139).

Subsequently Acehnese goes furrther south via Malacca to the extreme north of Sumatra, where it maintains ties with Champa for a few centuries, and where, profoundly islamised, it dominates the surrounding groups. The major and increasing source of new lexicon in Acehnese (including later borrowings from Tamil, Chinese, Portuguese, Dutch and English) is Malay.

By this stage Chamic has already begun to absorb Mon-Khmer words, which have undergone little in the way of phonological adaptation to Malayic phonological norms, rather the reverse has happened. This has the result that several new segments, including vowels and vocalic nuclei (but not yet implosive consonants) are borrowed, integrated and used productively. This integration includes their being found in elements which cannot be attributed easily to Austronesian or to Mon-Khmer. Even by the time of the first known Cham inscription the language has begun to turn inherited (but not borrowed) disyllables into iambs, and to begin to reduce (to /a/ or to schwa) or drop the first unstressed vowel. This change results in the creation of a number of initial consonant clusters (in words of Malayic origin) which are not tolerated in other Malayic languages, and the number of these is added to by the absorption of Mon-Khmer words with their frequent and often new initial consonant clusters. The effect is that the number of canonical syllable shapes, and the number of possible shapes for a phonological word, are both greatly increased.

Loans from Mon-Khmer languages are first reliably attested and documented in Chamic materials in the late ninth century, and the items which are borrowed are (as far as our records tell) already at this time replacive of preexisting Austronesian forms which were found in Malayic languages (such as the first recorded example, borrowed Cham *dom* 'all' from Khmer rather than older PMP *amin*), rather than simply only being cultural borrowings. Mon-Khmer elements which are shared between a Chamic language and Acehnese, and which can be shown to come from the same branch of Mon-Khmer (Northern or Central Bahnaric, see Cowan 1981), will have entered the ancestors of these languages in the period before the speakers of Acehnese left the mainland and will therefore be reconstructible to Proto-Chamic. The borrowing and integration of Mon-Khmer infixes such as the denominative /-an-/ has already taken place by this time, as the Inscriptional Cham data and the evidence from Acehnese both show.

4) After 986 AD. Tsat splits from Northern Roglai and thus from further contact with other forms of Chamic: Northern Roglai was probably the language which was spoken immediately south of that variety on the Chamic dialect chain which became Acehnese, and when the speakers of what became Acehnese left the area, speakers of Northern Roglai were briefly exposed. This language has, with its sister-language Northern Roglai, undergone the change of original /-a:s/ to /-a:/ (rather than the combination becoming /-aih/ as has happened in some other Chamic languages such as Eastern Cham), and both these have also seen the development of phonetic final preploded nasals.

After this separation Tsat is no longer in contact with Mon-Khmer languages, with the result that borrowing from these languages comes to an abrupt end, and therefore any Mon-Khmer elements in Tsat will of necessity have been shared with an earlier version of Northern Roglai. Speakers of Tsat are later in contact with a more southerly form of Chamic (possibly because some speakers of this language migrate and integrate with the more northerly Chamic community on Hainan whose speech gave rise to Tsat in the first place), and borrow some words from this. Instead Tsat comes into contact with Li/Hlai for some time (though these languages are not in contact with Tsat nowadays), with Hainanese Chinese (with which it is still in daily contact), and with sources of Islamic linguistic materials as well (namely Malay and Arabic, which some members of the Tsat community have recently begun to learn). Latterly speakers of Tsat come into increasing contact with Cantonese Chinese (the major trade language in the area) and various forms of Mandarin Chinese, with which latter Tsat is currently being swamped.

5) After 1471 AD. Cham proper splits into Eastern and Western Cham and Haroi. The subsequent fates of Chamic languages. In this instance the primary division took place after 1471, with the fall of the southern Cham empire. This division was exacerbated to some degree by religious differences between the groups, since Western Chams in Cambodia became (or remained) Muslim and adopted Arabic names, while two out of three Eastern Chams practise the modified version of the form of Hinduism which had been the state religion of Champa. In addition the religious contexts of the two communities were somewhat different, since Cambodia practised Theravada Buddhism and Vietnam mostly practised forms of Mahayana Buddhism.

Cambodian Western Cham comes into contact with Khmer as its dominant language and absorbs a huge number of loans from it. Mekong Delta Western Cham acquires some loans from Vietnamese but Khmer remains the major language in contact and it provides far more material, even in those areas which belong politically to Vietnam. Phan Rang Cham speakers are eventually outnumbered even in their own city by speakers of Vietnamese, and Phan Rang Cham has absorbed many allophonic features of Vietnamese phonology (the rise of tone systems based on the nature of initial obstruents and word endings, /s-/ becoming /th-/, /-l/ (< former /-r/ and /-l/) becoming /-n/, an increasing trend towards monosyllabism) by introducing them into previously conservative Cham phonological forms.

Speakers of Haroi, meanwhile, have split from speakers of the then regionally undifferentiated Cham at the time of the 1471 disruptions and have come into increased contact with Bahnar and also with the North Bahnaric language Hre, which leads to the development of restructured register and the absorption of numerous Hre and Bahnar loans.

The later (and very different) histories of Tsat and Acehnese have been discussed above. Speakers of Chru stayed in contact with (firstly) 'Common' Cham and later Eastern Cham, although Chru has not undergone the strong phonological changes in the direction of Vietnamese that Eastern Cham has experienced. Speakers of Roglai have been in constant contact with Vietnamese, and to some extent, with speakers of Eastern Cham, although the parallel vocabularies of Sre and Roglai in Bochet and Doumes (1953) show that these two languages share a lot of vocabulary, much of it of Mon-Khmer rather than of Austronesian origin. Speakers of Jarai and Rade had split off from the other Chamic communities before 1471; these languages have not subsequently been strongly influenced by other languages (although there appears to be a fairly sizeable Bahnar component in Jarai). Jarai has been relatively conservative in terms of phonology, apart from innovating a final low tone on vowels preceding a glottal stop, but Rade has strongly innovated phonologically.

3 *Norm-referenced lexicostatistics and Chamic*[1]

Anthony P. Grant

1. Norm-referenced lexicostatistics: introduction, history and methodology.

The lexicostatistical techniques that are used for analysis of materials in historical and comparative linguistics, which were first developed in their modern form by the American structuralist Morris Swadesh (and which were first made readily available in Swadesh 1950, see also Swadesh 1955 for a protracted exposition) have enjoyed mixed fortunes in the last half-century of historical linguistic work, although they are currently enjoying a certain degree of revival. (Glottochronology, with which lexicostatistics is often used and sometimes confused although the use of neither technique of necessity entails use of the other, is currently much less popular. Yet glottochronological dates of separation between languages and within proto-languages are still cited with reverence by non-linguistic specialists in other fields such as archaeology and anthropology, who impute to them a degree of methodological accuracy and overall reliability which few linguists would now agree with.)

The 100-item and 200-item lists (and to a lesser extent the older 215-item list) that were drawn up by Swadesh in the 1950s are still those which are used most frequently. This remains the case half a century on, even though it has long been recognised that they are not equally appropriate for all languages. Sometimes this is because of 'cultural gaps' in some languages. Often, however, it is because of differing semantic patterns, in certain fields at least, from those which were promulgated and incorporated onto the lists by Swadesh on the basis of his firsthand experiences of particular languages. Up to the time when Swadesh was assembling this list (a little before 1950[2]) this involved languages of Europe, North America, Mexico and (in part) the Far East, more specifically Mandarin and Burmese, both of which he had worked upon for the US military during WWII.

Consequently, a number of scholars have elaborated somewhat different gloss lists which are better suited to capturing certain of the semantic characteristics of a particular family of languages. This has been done on at least two occasions for the historical investigation of interrelationships within Austronesian languages. The renowned work of Dyen (1962 and especially Dyen 1965), which attempted to present a genetic classification of the Malayo-Polynesian languages by using lexicostatistical materials, used a 196-item list, namely the Swadesh 200-item list minus 'that' (the demonstrative adjective, which is not always distinguished from 'this' in these languages, though often split into different

[1] I would like to thank Bob Blust, Robert K. Headley, Russell Murray, Peter Patrick, Graham Thurgood and David Zorc and the staff of the Special Collections Reading Room at the School of Oriental and African Studies, University of London, for their assistance with aspects of the production of this paper. Any infelicities are of course my own responsibility.

[2] The first mention of Swadesh's use of this technique was in 1948, at a Viking Fund Supper Club presentation which he gave in New York that year.

Grant, A.P. "Norm-referenced lexicostatistics and Chamic". In Grant, A. and Sidwell, P. editors, *Chamic and Beyond: Studies in mainland Austronesian languages.*
PL-569:105-146. Pacific Linguistics, The Australian National University, 2005. DOI:*10.15144/PL-569.105*

forms depending upon the distance from the speaker, the visibility of the object referred to, and so on), and the tropically inappropriate 'ice', 'freeze', and 'snow'. Similarly-structured searches among the overtly-expressed morphological features of Malayo-Polynesian languages were not carried out in extenso. Nevertheless, on the basis of the findings from this lexicostatistical experiment Dyen posited the existence of 40 primary groups of Austronesian, with their area of greatest diversity (according to the findings of this lexicostatistical experiment) being in New Guinea, which he therefore proposed as the Austronesian *Urheimat*. In contrast, one of the 40 groups, the Malayopolynesian Linkage, accounted in Dyen's scheme for more than half of the languages surveyed, including practically all those languages which are now regarded as Western Malayopolynesian.[3] Dyen's vision was a view which has won remarkably little acceptance, despite Dyen's eminence in Austronesian linguistics. The reason for this is that Dyen was wrong in the inferences which he had drawn from the use which he had made of lexicostatistics (a point which was first made clear in Grace 1966, although Grace's valid reasons for his criticisms did not include an analysis of the faultiness of Dyen's lexicostatistical methodology).

In terms of the technique employed, what Dyen had used in his comparisons was *pair-referenced lexicostatistics*. In Dyen's investigation, each gloss in each Malayo-Polynesian language was compared by computer with the same gloss in every other Malayo-Polynesian language, so that each gloss in Itbayaten of the northern Philippines was compared with the appropriate gloss in Chru of Vietnam[4], Atayal of Formosa, Nauruan of Micronesia, and hundreds of other languages. What the glosses in these languages were not compared with, however, was the equivalent forms in any kind of a reconstructed proto-language at any level.

In the methodology underpinning this work Dyen was comparing Language A with Language B, Language B with Language C, Language C with Language D, and so on. This strategy is interesting in itself and can bring forth fascinating intimations of lower-level linguistic relationships (for pair-referenced lexicostatistics is very useful in certain spheres), and Dyen's concept of the 'critical percentage' (the greatest percentage of cognates which one language that is being surveyed has with any other language which is being surveyed) is valuable. But it is the wrong kind of lexicostatistical methodology to be used for what Dyen was trying to achieve, and without firstly using the right sort of methodology, his wider aims for his research and such findings as emerged from them were futile.

What Dyen did not attempt to do in the course of his lexicostatistical studies was make us of any information which would have enabled him to indicate which of the elements in these languages went back to a proto-language and which other elements were borrowings from current or previously surrounding languages (both Austronesian and non-Austronesian), later internally-driven lexical developments, or forms confined to sub-

[3] This is paradoxical and counterfactual because Western Malayo-Polynesian is not a proven subgroup, as it is not distinguished by the possession of any shared innovations, and therefore has to be defined negatively as being that subset of Malayo-Polynesian languages which does not possess the shared innovations of Oceanic for instance, or of Central Malayo-Polynesian. (Nevertheless Western Malayo-Polynesian does contain several well-defined subgroups of its own: Malayo-Chamic is one such.) I call such negatively-defined large groups 'antigroups'.

[4] Chru was the only Chamic language, apart from Acehnese, for which Dyen had access to a lexicostatistical list, and Dyen's findings did not pick up on the special historical connection between these two.

branches of Malayo-Polynesian (MP) or whatever. Furthermore he was interested in the number of cognates which were to be found between pairs of languages, but he was concerned with absolute figures and not with forms. The actual cognates, and the degree in each instance to which they were replicated in the vocabularies of one language or another, did not enter the picture and they were not exemplified. The result is an internally-enclosed and self-referential analysis, which has the potential to give observers a misleading picture of the relevant genetic linguistic relationships.

In short, Dyen was using an approach which was too purely quantitative, whereas the nature of the task required recourse to more qualitative methods. These methods took note of the quantitative findings which could be gathered fairly quickly, but did not confine themselves to them, going instead beneath the surface to analyse the kinds and the relative historical statuses (PAn, PMP, Proto-Malayo-Chamic, etc.) of the forms which two languages shared.

If it had been the case, for example, that in a hypothetical family Languages A and B shared 25% of the cognates on the list, and that Languages B and C shared 25%, and that Languages C and D shared 25%, but that none of the actual shared cognates were to be found in more than any two of these languages or in any more than one of the pairs listed above, then this highly significant fact, which might at least superficially cast serious and reasonable doubt upon the ultimate unity in origin of A, B, C, and D, would not have been clear from the tables of percentages presented in Dyen's study.[5] Looking at these tables of percentages of forms which are common to any two particular Austronesian languages in each case, we cannot tell from such figures which items among the commonly-shared forms are inherited from Proto-Austronesian, which other forms reconstruct back only to Proto-Malayo-Polynesian, and which other of those forms are first found in a daughter-language of Proto-Malayo-Polynesian, such as what we now call Proto-Oceanic. And we may assume that on certain occasions those words which are common to two contiguous languages and which are taken by Dyen as being cognates jointly inherited from a parent language may actually have been introduced from one to another, and it is sometimes possible that they may even have come into both languages from a third language.

Dyen was an admirer of the achievements in the Malayo-Polynesian reconstruction work of Otto Dempwolff (as, to a large extent, am I). In Dyen's published work he has given little indication that the doubts the essential correctness of the visible fruits of Dempwolff's remarkable intellectual achievements; though he does revise and improve many of the spellings of Dempwolff's PMP reconstructions, he does not doubt that they are correct and valid. Yet crucially he did not compare the gloss list for any language with those available for each item in the three volumes of Dempwolff (1934-1938). This was a lost opportunity which had considerable consequences for much later work on Malayo-Polynesian subgrouping.

Had Dyen referenced the items on each list to their occurrence or non-occurrence on (and their cognacy with) a list of equivalents which used elements derived from Dempwolff's list, he would have been practising a kind of *norm-referenced*

[5] It is always theoretically possible for two languages which are descended from the same parent language, but which belong to different subgroups and which are both low scorers in regard to lexical retention from the parent language, to have a cognacy rate of 0%, although I do not know of any certain examples of this

lexicostatistics[6], a technique in which the forms in each language are compared with the forms in the same control language, control case, or 'norm language'. This type of lexicostatistics can be seen as a development from the lexicostatistical principle which is also used as an essential part of traditional glottochronology, namely that a wordlist from one historical state of a language, which is taken as the control case or norm, is compared with a wordlist from a later historical state of the same language or with several such states of the same language (which are each compared the forms from the earlier stage of this language). When this has been done, then the number and proportion of forms remaining in the later state (or states) and that have been perpetuated from the former state, that is from the control case language, is calculated. In this particular study, however, glottochronological techniques are not going to be used.

In such a scenario as one using Dempwolff's reconstructed Malayo-Polynesian proto-forms (for want of better reconstructions), the ideal norm (or the language) against which the forms in each language were being compared, one language after another, would be an assumed and reconstructed proto-language which had been arrived at independently of the investigation of any of the daughter-languages under discussion. Using such a method early in his examination would have enabled Dyen to spot numerous recurrences of the same widespread but non-Proto-Malayo-Polynesian (and therefore not directly inherited) morphs in various languages, and this might have led to the earlier reconstruction of such important subgroups as Oceanic. Such a technique, measuring the proportion of forms which a particular language has retained from a list of forms from its proto-language, is something which Robert Blust has done in certain of his papers (for instance Blust 1993). Most importantly, Blust has shown that the number and proportion of retentions varies from one set of Malayo-Polynesian languages to the next (see also Blust 2000b for an illustration of this.).

Epistemologically at least such a comparison would have been something of a risky exercise, since one is dealing with an abstraction (namely Dempwolff's inductive reconstruction of Proto-Malayo-Polynesian), the degree of whose similarity to an assumed but unrecorded entity is uncertain (and was even more uncertain at that time). Furthermore, one is comparing elements of this abstraction with data from attested languages. Nevertheless, as one attempts to do this kind of historical reconstruction of the linguistic manifestations of actual speech-community splits, the use of such a technique demonstrates the similarities (admittedly both retentions and innovations of various sorts, including those borrowings found in more than one language) of different languages to a particular reference point, and is a valid approximation to the facts.

The findings of norm-referenced lexicostatistics are best seen displayed overtly, for instance in the form of a grid. This has been done by Miller (1984), using a modification of the Swadesh 100-word list, in an attempt to subgroup a couple of dozen Uto-Aztecan languages in North and Central America, and a modification of Miller's model (a model which is closer to the technique used in Miller, Carpenter and Foley 1971) is the one which I have pursued here. The primary purpose in such comparisons is to spot similar forms, and

[6] This term was introduced in Bennett (1998), to describe a kind of lexicostatistics that he applied to Semitic languages, in which the number of forms, inherited from a proto-language (which was the norm against which each of the modern languages was referenced), that remained in the lexicon of a modern language, was counted for each language that was surveyed. For instance it might be the case that out of 5 forms reconstructed back to Proto-Semitic, Language A retained 4 but Language B only retained 2 while Language C retained 3.

more specifically, to spot cognates between languages which represent retentions, and thereafter to distinguish them from those which represent innovations. The result is a kind of 'multilateral comparison' (a term which was made famous by Greenberg 1987), but it is one in which there is a norm language used in the comparisons, a language which may have true historical or other non-trivial significance to the project. (A similar technique was used at about the same time by Hooley 1971 in his classification of the Austronesian languages of Morobe Province in present-day Papua New Guinea, although Hooley used numerals rather than letters to separate out words belonging to different cognate sets, and he did not use special indicators in his tables of forms for missing glosses, unique forms, or loan elements as Miller did.)

Although Miller had previously published a long list of 'formulist' reconstructions of Proto-Uto-Aztecan forms (Miller 1967), he did not employ the results of this in his 1984 work. Consequently a PUA (= Proto-Uto-Aztecan) column is not provided as a norm language in his table of cognates and similarities, which is presented in grid form, and the reflexes of the forms in Miller's list are not compared with those which had already been reconstructed for PUA. In fact, Miller does not cite the actual forms used for the expression of each gloss in each language. Instead, what Miller did was to start from the leftmost and most northerly languages in his table, the Numic languages of eastern California and the Great Basin, and to assign the letter 'a' to the word which is used in this language, so that the reflex of each word in this leftmost language is always marked with 'a'. If the next language used a form of a different word to express the same concept, then 'b' is used, and if a further language uses a form of a word which is different from both of these then 'c' is used, and the process continues this way.

When drawing up his table Miller used the symbol 0 for cases in which a form for a particular gloss in a particular language was not available to him, so that a particular slot or cell had to be left empty, while he capitalised the letters in cases representing words in the list which had been borrowed from another language. Instances in which the gloss for a particular item was represented by a form which was exclusive to that language and which was found in no other language in the sample, were represented with 'x'; there could be more than one 'x' in each line of the list (sometimes there were half a dozen or more). If a loanword was only attested in one language, it too could be capitalised as X. We may call the grid which results from these procedures a *cognate grid* or *cognacy grid.* Cognate grids may not necessarily result from the application of principles of norm-referenced lexicostatistics (and we have seen that Miller was not using such norms), but they can be developed for use in data regression after the application of this kind of lexicostatistical discovery procedure.

For Austronesian languages the default lexicostatistical list used nowadays is that drawn up in Blust (1981), a brilliant and still unpublished paper. The list draws upon the work of another Borneanist, Alfred B. Hudson (Hudson 1967), which used a 203-item list, but goes beyond it in terms of its range of applicability, and versions in English and Malay have been widely circulated. In addition, Blust (1993), a paper which uses this list as a basis, provides reconstructed Proto-Malayo-Polynesian translations or equivalents for every item on the 200-gloss semantically-arranged list, and well over half of these forms (at least 116: Robert Blust, personal communication, 1997) are also attested in some or all Formosan languages and can thus be reconstructed back to Proto-Austronesian, with appropriate phonological adjustments. Almost 85% of the items on Blust's list are to be found listed on either the 100-item or 200-item Swadesh lists, while the remaining forms are (with a couple of exceptions) well-suited to one's expectations of the assumed semantic primes of the lexica of Austronesian languages. Blust's list is much better suited to this particular task and to these particular languages than the one which Dyen used or adapted (though of course it would also have been perfectly feasible to practise norm-referenced lexicostatistics using Dyen's 196-item list), and Blust's list is the one which I have used below.

Blust (2000b) has made a terminologically useful distinction between *horizontal lexicostatistics* and *vertical lexicostatistics*. The former technique is the one which is more widely used nowadays (though this was not always so). This technique compares lexical data from languages which are supposed to have been attested in the same time period and to be roughly contemporaneous. Meanwhile the latter technique compares lexical data from an earlier stage of a particular language with data from other languages which are assumed to be descendants from this language. Comparisons between material from Classical Latin on the one hand (Latin being the control case or norm) and French, Spanish, Italian and so on, on the other, would be an example of the use of vertical lexicostatistics. Comparisons between French, Spanish and Italian would be instances of horizontal lexicostatistics. The study offered in this paper uses horizontal lexicostatistics as a point of departure, since most of the languages compared are contemporaries of one another, but additionally it incorporates the findings which vertical lexicostatistics (and more specifically, which the use of the Blust 200-item list) can give us.

For Blust (2000b: 320) horizontal lexicostatistics is characterised by a known retention rate (which Swadesh had long since set at 0.81 per millennium, or 81/100 items are supposed to be retained from the word list after a thousand years), an unknown period of divergence between the two or more contemporary languages that were being surveyed (indeed we may say that the time when these diverged was the question to which we were seeking an answer), and an ability to calculate these figures horizontally. With vertical lexicostatistics the rate of retention was unknown, but the time of divergence between the control case language and the descendant language(s) was supposed to be known, and the figures could be calculated vertically. The unspoken assumption is that in vertical lexicostatistics all the languages concerned diverge from the ancestral language to approximately the same degree. But this is not the case with horizontal lexicostatistics, and this is supposed to enable us to subgroup languages (and then to construct family trees) according to their depth or recency of split from one another.

Combining the strengths of historical investigation and of the use of a cognate grid in norm-referenced lexicostatistics in which the norm comprises items from a reconstructed language allows one to take advantage of the strengths of the various subfields: the use of a well-selected lexical sample (a choice of material which is especially germane in the case of languages which have minimal inflectional morphology of the sort relied upon for historical linguistic purposes by diachronists), and the ability more clearly to see patterns of lexical distribution within a chosen sample of languages.

There is also the benefit that can be provided by working from a set of reconstructed forms, which (if we have enough historical background information to make assumptions secure) allows one to recognise whether the equivalent form in a modern language which is being surveyed is an inherited form that the proto-anguage contained, or whether it is one or another kind of innovation. Different kinds of such innovations would include borrowing (including the borrowing of a form which is cognate to one which might have been found in the lexicon of the proto-language under discussion, and thus a 'false cognate'), internally-developed form, or whatever. Indeed Blust (2000b) pointed out that it is the inability of horizontal lexicostatistics to be able to let us distinguish between inherited forms and other forms which are innovations shared between two or more languages (but not between all the languages that are being surveyed), which vitiates this technique. With the use of vertical lexicostatistics this confusion of the historical status of elements does not happen.

2. The Chamic languages in their historical and contact setting.

The Chamic languages, long overlooked or misclassified as Austroasiatic by linguists as recently as Sebeok (1942), have received considerable recent attention in the Austronesian linguistic literature, thanks very largely to the work of Graham Thurgood over the past decade (for instance in Thurgood 1996; the work which he has carried out is encapsulated in Thurgood 1999; see latterly also Thurgood to appear a, b, c, Thurgood and Li 2003). Thurgood's work, rooted as it is in historical phonology and the use of 'top-down' and also 'bottom-up' modes of reconstruction[7], and with its copious references to parallel forms in Malay (which shares a number of non-trivial phonological developments of Proto-Malayo-Polynesian sounds with those which are found in Chamic languages), demonstrates beyond reasonable doubt that the speakers of the ancestor of the Chamic languages left Borneo (where its sister-languages were spoken) a few centuries before Christ. This is also what Proto-Malayic had done, although the movement of the speakers of Proto-Malayic from Borneo took place probably some centuries after the departure of the speakers of Proto-Chamic (or maybe Pre-Proto-Chamic).

The linguistic evidence which can be gleaned from the responses to the Blust list and from other sources also shows the skeptical observer that the speakers of Chamic languages, like those of the Malayic languages which are its closest genetic relatives, have returned to mainland Asia after their ancestors spent millennia in the islands, rather than having remained in Asia in situ for millennia.[8]

The most widely-spoken Chamic language is Acehnese of extreme northern Sumatra, with over 2 million speakers (its relationship with other Chamic languages, which is beyond doubt, is discussed in part in Durie 1990). It is one of two Chamic languages which has left Indochina, the other being Tsat or (as it is called in Putonghua) Huihui, a language spoken by a Muslim minority of a few thousands in two villages on the extreme south coast of Hainan, China, who descend at least in part from Chamic-speakers who

[7] 'Top-down' reconstruction, starting with forms which can reasonably be assumed to have occurred in a proto-language and then tracing their phonological histories in the various daughter languages, is preferable in Chamic languages, because it is certain that they are all related to one another, and because many of the customary reconstructional techniques of historical linguistics are difficult to apply to items in Chamic languages as a result of the varying but often dramatic effects of changes in the forms of the syllable, especially in the presyllable segments. For example Cham, just like Malay, has _lima_ but Jarai has _rema_, Rade has _ema_ and Tsat has _ma33_ for proto-Chamic *_lima_ 'hand, five' (PMP *_qalimah_). These changes are perfectly in accordance with the developments of Proto-Chamic historical phonology in each language, even though in other phonological environments PMP *l would become /l/ in all the languages concerned.

[8] Proto-Chamic and Proto-Malay share the same diagnostic reflexes of PMP sounds such as *_q_, *_Z_, *_R_, *_D_, *_b-_ and also the same innovative shapes of PMP words such as *_wahiR_ 'water', and *_qaqay_ 'leg, foot', features which allow them to be subgrouped together against languages such as Moken of the Mergui Arcipelago, Burma, and Javanese.

migrated from what is now Vietnam maybe a millennium ago[9]. Of the other languages, the most widely known is Cham, which was formerly the language of Champa, a series of kingdoms of Hindu/Buddhist cultural affiliation, part of the East Asian Indosphere, the southern part of which was finally brought to its knees in 1471 by Khmer invasions (the northern kingdoms had succumbed to the incursions of the Vietnamese in 982, when the Vietnamese were themselves responding to pressure from the Chinese to the north). Cham survives in two differentiated dialects which now have the status of separate languages. These areEastern Cham or Phan Rang Cham of Phan Rang, formerly known as Panduranga, in coastal Vietnam, and the emigrant Western Cham of the area around Tonle Sap in Cambodia, and of Chau Đoc and other Khmer-speaking areas in the Vietnamese part of the Mekong Delta. An earlier form of the language, as it was spoken before the dialectal division and before the strong impact of Vietnamese on Eastern Cham, was (and to a slight extent still is) used as a written language by male Chams, employing a distinctive alphabet of Indic origin.

Other Chamic languages are Jarai and Rade/Rhade/Ede, which are spoken in the Vietnamese highlands, Haroi, which has moved to the highlands from coastal Vietnam, and two other languages or language groups spoken in areas near the Vietnamese coast, namely Chru and Roglai (the latter includes several forms of speech, notably Northern Roglai, which is the best described form, Southern Roglai, and the aberrant Cat Gia Roglai, all of them used in coastal regions of Vietnam). These languages are all clearly related, as even a cursory inspection of wordlists shows, but just as clearly they exhibit an impressive array of variation and diversity, especially in regard to the developments in each language of features of Chamic historical phonology. Nevertheless there are phonological developments from Proto-Malayo-Polynesian (hereafter PMP), such as the change of initial PMP *n-* (itself a very rare sound word-initially) to *l-*, which are common to all Chamic languages including Acehnese and which, regionally at least, mark them out from other Malayo-Polynesian languages in the area (including Malay in this instance). It should be understood, though, that these changes are not exclusive to Chamic throughout the whole Austronesian world.

To the best of my understanding, almost none of the languages listed above are mutually intelligible. Phan Rang Cham and Western Cham may be a partial exception, as these may be interintelligible, although Western Cham has absorbed a large amount of lexicon from Khmer, including epistemic particles and other grammatical morphs, and none of this is found in Eastern Cham. Meanwhile male speakers of Cat Gia Roglai are bilingual in Phan Rang Cham: the situation is discussed in Lee (1998), but this societal bilingualism does not constitute mutual intelligibility of the languages involved.

Phonologically the most aberrant Chamic languages are Tsat (this aberrancy has come about as a result of influence from non-Chamic languages such as Hlai and

[9] But there may have been more than one wave of migrants from Champa to Hainan, and it is further possible that several centuries may have elapsed between migrations to Hainan (Pang 1998). Nor need the different waves of migrants of necessity have come from the same region in Champa. Indeed, as Graham Thurgood pointed out (personal communication, 22 March 2002), there is evidence of some dialect mixture within the Chamic component of Tsat, with some southern forms (for instance the numeral ‘hundred’) being mixed at a later period into the basically more northern language which gave rise to Tsat as we now know it (the lower numerals show more distinctly Northern Chamic traits, insofar as diagnostic forms are available for inspection).

Hainanese), Rade, and Cat Gia Roglai (or Cac Gia Roglai). In the last two cases there is no obvious external linguistic motivator for the striking and surprising – and, it must be noted, very different – sets of developments in their historical phonology. We cannot state that they have modified their phonologies in order for the resulting system to resemble more closely the phonological system of any particular neighbouring language. The fact that the remarkable presyllabic phonological constraints in Rade resemble those of the Mon-Khmer language Chong of Laos and eastern Thailand, which is a Pearic language, is almost certainly a coincidence. This is because there are several Mon-Khmer and other languages separating the areas populated by speakers of Rade and Chong, and these separating languages do not share these highly marked presyllabic constraints (see Thurgood 1999). Matisoff (2001) has, however, pointed out that there are some apparent shared innovations, in terms of the kinds of massive erosion that the forms undergo, between the construction of presyllabic onsets in Rade and those found in Tsat.

There are published and unpublished descriptive materials available for all of these languages, but only Acehnese and to a lesser extent Tsat and Phan Rang Cham are well-described in regard to lexical coverage. The provision of text collections, and grammatical descriptions are rare for these languages, and only Acehnese and (rarely) Phan Rang Cham are used in writing. I refer to the Chamic languages apart from Acehnese and Tsat as Indochinese Chamic languages; I would point out that I use this term as no more than a geographical expression and I would assert that no genetic considerations, suggesting that Indochinese Chamic languages constitute a single genetic subgroup, should be read into it. It is simply that they are Chamic languages which remained in Indochina throughout.

Typologically and especially phonologically Chamic languages resemble Mon-Khmer languages (including the Bahnaric languages with which Proto-Chamic was in prolonged and intimate contact, as well as Khmer and Vietnamese, with at least one of which most speakers of Chamic languages have been in contact).[10] In fact they look superficially like Mon-Khmer languages much more than they resemble the Western Malayo-Polynesian languages of Borneo, including such languages as Proto-Malayic, from which they have derived. Even more so than what has happened with many Malay dialects, the Chamic languages have been integrated into the Southeast Asian linguistic area more and more over the past couple of thousand years. The result of this is that they now exhibit such Southeast Asian areal characteristics as numeral classifiers, which are also found in

[10] The customary classification of Mon-Khmer languages within Austroasiatic recognised eleven groups organised into four larger branches: Northern Mon-Khmer, Southern Mon-Khmer, Eastern Mon-Khmer and Viet-Muong or Vietic. The first branch includes Khasi, Palaungic and Wa, and these languages have not been involved with Chamic languages. Southern languages are Monic, Nicobarese, and the Aslian languages of Malaya (the latter have influenced Acehnese but have not otherwise been involved with Chamic languages). Eastern Mon-Khmer groups are Pearic, Khmer (the closest relative of Pearic), Bahnaric languages (with two major divisions) and Katuic. (Southern and Eastern Mon-Khmer languages are themselves regarded as being the two branches of South-Eastern Mon-Khmer, a grouping which is parallel to Northern Mon-Khmer and to Vietic.) Vietic languages constitute the fourth branch, although it is possible that they are most closely related to Katuic languages. Eastern Mon-Khmer languages, specifically Bahnaric and Katuic, and in many cases latterly Vietic (specifically Vietnamese) have been the languages which have exclusively exerted the Mon-Khmer influence on all languages including Acehnese, though the latter has, as previously stated, been in later contact with Aslian languages (though not with Vietnamese). Paul Sidwell (personal communication) indicates that Katuic languages exerted strong influence upon Bahnaric languages.

Malay and in some North Sarawak languages but which are not part of the structures of many other Austronesian languages. This absence is true even those languages which contain a considerable stratum of loans from Chinese languages: this is the case for instance of Tagalog, which lacks numeral classifiers (although it does contain numerous loans from Hokkien Chinese).

The Chamic languages have furthermore adopted or acquired many of those salient typological characteristics of Mon-Khmer languages which are not also pan-Southeast Asian typological features which cut across genetic lines. (It is a reasonable asusmption that Mon-Khmer languages are the major source for Southeast Asianisms in the Chamic languages.) These features include the prevalence in the vocabulary of monosyllabic and sesquisyllabic contentive stems, the presence and widespread use of glottalic consonants and of many vowel nuclei alien to most Austronesian languages, the use of derivational infixation (rather than the more primarily inflectional infixation found in many Austronesian languages), and most significantly, distinctive registral patterns – patterns which have sometimes (as also with Mon-Khmer languages such as Vietnamese) led to the development of partial or full tone systems. This development has happened independently in Tsat, Phan Rang Cham and to a slight extent in Jarai.

The vocabulary of most of the Chamic languages contains a greater number of lexical items of Mon-Khmer origin than there are those of Austronesian, Malayo-Polynesian or even Malayic origin (these latter numbering a few hundred at most). Even the proportion of *undoubted* Mon-Khmer elements in the reconstructed vocabulary of Proto-Chamic is well over 15% (I counted 205 assured Mon-Khmer-derived items out of 755 Proto-Chamic and post-Proto-Chamic forms that are listed in Thurgood 1999, and there are over 200 further forms which may be of Mon-Khmer derivation). It is certain that all Chamic languages have been recipients of this ‘partial relexification’, as many core items that are of (say) Bahnaric origin are also found in Acehnese, as are a number of basic pan-Chamic forms which are of uncertain origin. And most of what few productive (or even unproductive) elements of bound morphology there are either derive from Mon-Khmer languages or else are very close in both form and meaning in Mon-Khmer and Austronesian languages. By contrast, most of the small battery of inherited Western Malayo-Polynesian affixation has either been lost completely, or at best is preserved in a few frozen stems and is no longer productive. There may be Austronesian languages that have retained fewer elements from their Proto-Austronesian lineage than the Chamic languages have, but there cannot be many of them (such languages are found in Papua New Guinea and the Solomons).

A fairly strong case could be made for claiming that the Chamic languages are mixed languages (and that they are to some extent even intertwined languages in the sense in which the term is used in Bakker and Mous eds. 1994, since some of what little bound morphology they have is taken from Mon-Khmer languages). It is possible that such ‘linguistic mixture’ has taken place here because the earliest Cham communities were built up mostly by exogenous men (the Chams were notorious pirates), who were in a position socially, politically and technologically to dominate the members of the communities upon which they had intruded, and who intermarried with indigenous Mon-Khmer-speaking women, upon whom they imposed their Austronesian language once they had established coastal communities.

For its part, Tsat, a language which started out being very similar to Northern Roglai, has become typologically more and more like Hainanese Chinese and latterly more like Mandarin Chinese as time has progressed, and this direction of change is manifested both in lexicon and morphosyntax (Thurgood and Li 2003). The salient features of Tsat segmental and canonical phonology look like a subset of those of a modern Southern Chinese language, and this has extended to the development of a full five-tone system whose origins Thurgood (1999) reconstructs on the basis of changes in Tsat historical phonology since the language's separation from other Chamic languages. The relics of PMP prefixes and infixes which can be found in other Chamic languages have been completely lost from sight in Tsat, since the words which contained such forms have undergone far-reaching sound changes, to the extent that the syllabic canons and prevocalic consonantal forms which are now permitted in Tsat are a subset of those permitted in Hainanese. Only the use of internal reconstruction and subsequent comparison with corresponding forms in other Chamic languages can shed light on the underlying phonological forms of Tsat words, so that only reconstruction from the top down could show the clear Austronesian origin of more than a small number of Tsat forms.

By contrast, what makes it possible for us to classify the Chamic languages genetically as Austronesian or even Malayo-Chamic is their possession of a few hundred morphs, very few of them bound (such inherited bound morphology as Chamic languages have is no longer productive and much of it has been lost completely) and the bulk of them lexical items which centre in the most frequently-used and generally the most culturally-neutral items of the vocabulary of Chamic languages. Yet even this most basic lexical element is not exclusively a Malayo-Chamic domain, as the table below makes clear. Much of the Chamic lexicon of all kinds, including very many high-frequency verbs, derives from Mon-Khmer languages, and this includes numerous forms which are found in most or all Chamic languages, and with the impact of (especially) Vietnamese on modern Chamic languages, this proportion is growing even more. There is an ineluctable Mon-Khmer element (over 10% of the total at a conservative estimate) in the portion of vocabulary which is common Chamic, which is reconstructible back to Proto-Chamic and which appears on the Blust list. This percentage is surprisingly large for such a loan stratum which can be found in a securely-reconstructed proto-language.

Furthermore a considerable proportion of the lexicon of any Chamic language (and this is a stratum which is less well-represented in the most basic lexicon, but certainly far from absent even here) is made up of forms which have not been properly etymologised, but which have no cognates in any Austronesian languages (nor yet have clear etyma for these any been found in Mon-Khmer languages). But at the same time these very words often possess certain surface phonological characteristics, such as implosive stops or particular vocalic nuclei, which are typical of Mon-Khmer elements in Chamic languages but which are rarely if ever found in items belonging to the slender yet genetic Austronesian stratum in Chamic. The presence of such phonological features in these items suggests that these words entered Chamic languages either at or some time after the period of intense Chamic contact with Mon-Khmer languages, and after the rise to prominence of the monosyllabic contentive. There is a small but nonetheless significant stratum, smaller than that deriving from Mon-Khmer languages, of forms which are reconstructible to proto-Chamic and which are also found on the Blust list.

A considerable proportion of common free grammatical morphs in Chamic languages are as yet of uncertain origin (and a number of these are common to Acehnese and other Chamic languages, so that they must reconstruct back to Proto-Chamic), and some others derive from Mon-Khmer languages. This latter group of forms comprises both those forms which are common to all or most Chamic languages, and a later but sizeable number of free elements, for instance certain negators and some modal verbs, which have been borrowed into individual Chamic languages from Khmer (in the case of Western Cham) or Vietnamese (in the case of all the Indochinese Chamic languages spoken in Vietnam) since the split up of the Chamic languages about two millennia ago. The incursion of Vietnamese and Khmer elements into Chamic languages is apparently a matter of only a few centuries' age. Yet other such free morphs, including the numerals, are inherited from Proto-Malayo-Chamic (and some higher numerals have been diffused from Chamic into neighbouring Mon-Khmer languages).

Many of the same Bahnaric elements are common to all Chamic languages and therefore reconstruct to Proto-Chamic, into which they are loans, and this much could be demonstrated many times over by the employment of Venn diagrams or by using other demonstrations of the principles of set theory. A second set of Mon-Khmer forms is found in most or all Indochinese Chamic languages (that is, all save Acehnese and Tsat). Having split from the other languages more than a millennium ago, having lost all contact with other Chamic languages and with its speakers having migrated to Sumatra by way of eastern Malaya, Acehnese contains elements from Aslian Mon-Khmer languages, which were probably dominant in that part of Malaya at the time, but it contains an especially large number of loans from Malay (these including some forms which have replaced the more characteristic and inherited Proto-Chamic forms and which therefore count as instances of relexification), Sanskrit and Arabic. There are also (fide Thurgood 1999) a number of loans in Acehnese which derive from Katuic languages, and which are not found elsewhere in Chamic languages.[11]

The other migrant Chamic language, that is Tsat, contains a few stray elements of Hlai (a pre-Chinese language group of Hainan which is distantly related to the Tai languages) and many more from Chinese languages. Thurgood and Li (to appear) note the presence of four layers of loans into Tsat. Chronologically the first layer derived from Hlai. The second layer was taken from Hainanese Min Chinese, the third layer was taken in the course of the 20th century from the Mandarin spoken by the military personnel who were settled near the Tsat villages, and which was acquired by Tsat from contact of its speakers with these personnel, and the fourth and most pervasive layer derives from standard Mandarin (Putonghua) as taught in all Chinese schools. This final layer has wrought strong typological changes upon Tsat (some of which are exemplified in Thurgood and Li 2003), though Tsat may already have developed a tone system even under influence from the multitonal Hlai. This influence has been actuated by the spread of universal Putonghua-medium state-education among the Tsat rather than by intermarriage with native speakers of Putonghua, since Tsat speakers are endogamous Muslims whereas Han Chinese are, officially at least, atheist and therefore Chinese men at least are not permitted to marry Muslims. (The incursion of Vietnamese and Khmer elements into the lexica of the Indochinese Chamic languages rather unsurprisingly postdates the separation of Tsat and Acehnese from other Chamic languages, since the lexica of Acehnese and Tsat contain no unambiguously Khmer or Vietic forms.)

Haroi has borrowed heavily from Bahnar and Hrê, both of them being Bahnaric languages, Haroi and Bahnar have both developed restructured register, and Haroi-

[11] Paul Sidwell (personal communication, April 2003) assures me that earlier claims that there is a pan-Chamic component that is of exclusively Katuic origin are largely incorrect.

speakers have most in common culturally with speakers of Bahnar – indeed Haroi culture *is* Bahnar culture, and the Haroi have sometimes been known as the 'Bahnar Cham'. Western Cham contains added elements from Khmer, which are not found in Eastern Cham, while other Chamic languages have borrowed heavily from Vietnamese, and through this have recently acquired elements originally from French and English. Cham in both its modern forms (and additionally in the traditional written form) also contains a number of elements from Malay, since all Western Chams and many Eastern Chams are Muslims who used Malay as a liturgical language after their conversion to Islam. These are not usually to be mistaken for inherited Malayo-Chamic elements, however, because of the semantic fields in which these Malay borrowings enter (namely religious and similar cultural considerations). Jarai, Rade and Northern Roglai do not appear to have been especially adlexified or even relexified by the absorption of innumerable words from neighbouring Mon-Khmer languages; the main source of new words in these languages is Vietnamese. Rade had had some role as a lingua franca in part of the Vietnamese highlands (Tharp 1980) and may have been a donor language to some (Mon-Khmer) languages rather than being a recipient language.

All the forms which are of Malayo-Polynesian origin and which occur in Chamic languages have either been inherited from Proto-Malayo-Chamic, which is the common ancestor of Malay languages and Chamic languages (and have sometimes subsequently been lost in Malay), or else they are secondary possessions. More specifically, they are borrowings into these languages from Malay, and therefore are loans but not true cognates. In addition some forms which derive from Proto-Malayo-Chamic are found in Malay and in Chamic languages but cannot be reconstructed further back, which suggests that they are innovations within Proto-Malayo-Chamic. (A couple of dozen Malayo-Chamic lexical innovations are given in Blust 1992.) The vast majority of forms on the Blust list which have been retained in Malay lects are also found in Chamic languages, and vice versa. Together Malayic and Chamic have retained some 60% of the 200-item Blust list PMP reconstructions, and the bulk of these retentions are found both in Malayic (which retains 116 of the 200 forms) and Chamic, as indeed are most of the small set of phonologically modified retentions, such as *kaki* 'leg, foot' from PMP **qaqay*.

Despite some superficial coincidental similarities, there is absolutely no lexical or other linguistic evidence in the inherited component of Chamic languages to suggest that Chamic languages subgroup especially closely with Formosan languages, or with one or another subset of Philippine languages, much less with Oceanic or other Central or Eastern Malayo-Polynesian languages. Such similarities in phonological developments as we sometimes find occurring between Chamic and (say) Oceanic are coincidental and of independent development, and do not indicate a special non-trivial historical relationship or period of shared development.

As far as we are aware, Malay has not borrowed any items from Acehnese or Chamic languages, though Vietnamese (though to a very slight extent) and some other Mon-Khmer languages have done so; for instance the Vietnamese word for 'island' *cù lao* is probably a loan from Cham *pulaw*. (Malay *pulau* is also a possibility as a source, though.) The phonological form of the Vietnamese word shows that it was probably borrowed at a time before /p-/ was a permissible or legitimate syllable-initial consonant in Vietnamese (where original /p/ had apparently changed into /f-/), as it was to become in the 19th century with the incursion of borrowings, especially nouns, from French.

3. The use of the Blust list for historical explorations in Proto-Malayo-Polynesian, Proto-Malayic and Chamic languages: aims and operations.

Given the primary consideration - or the primary obstacle - that bound inflectional morphology, which is the kind of evidence which is most prized by diachronists who are attempting to prove the genetic relationship of two or more languages, is almost absent in the Chamic languages, and that much of what little bound morphology there is appears to be borrowed, the best that we can do is to explore some of the possibilities inherent in comparing basic lexicons. (If there were more inflectional morphology available for us to work with, then we would give that part of the languages preferential treatment in a study like this.)

The 200-item Blust list is well-suited to the purpose of comparing basic portions of these languages, although longer lists could also be used and these would tell us even more about the history (and especially the external history) of Chamic languages. It should be noted that evidence from the Chamic languages and Acehnese played little or no part in the original elaboration of the Proto-Malayo-Polynesian forms which are displayed in Blust (1993), so that these can be analysed objectively using this method. There is no risk of circular reasoning in this regard.

What we are trying to do is to see what can be gained from employing a combination of several techniques which are being employed sequentially, in order to put into practice a kind of multilateral comparison. We are employing lexicostatistics (though not glottochronology), and we are referencing each entry to its occurrence or non-occurrence on the equivalent wordlist for the norm which we are using (in this case the norm being used is Blust's reconstruction of Proto-Malayo-Polynesian, against which the reflexes of the glosses in the various modern Chamic languages are mapped). What is more, we are indicating the cognacy of each item to the norm or to other equivalents by the alphabetic code which was detailed above in the discussion of Wick Miller's work.

In this case, though, I am not using 'x' as a marker of lexical singularity; instead I am giving a separate letter to every discrete form, whether it is unique to one language or used among two, more or among all the sampled languages. Whichever gaps remain after my strenuous and studious attempt to fill them will be marked with 0, and loans, which in the case of the Chamic languages are mostly from Mon-Khmer languages (while in Malay they are mostly from Arabic or Sanskrit), will be indicated in a special column at the right of the table. I am comparing the cognacy of these Chamic forms (including the Acehnese forms), wherever possible, with the equivalent forms in Proto-Malayo-Polynesian as reconstructed by Blust, and with those in standard Malay, in an attempt to derive a more nuanced picture of the interrelationships within Chamic. Where possible, plausible loans into Acehnese from Malay (for instance, those which do not follow the sound correspondences that have been drawn up as obtaining between Proto-Chamic and Acehnese in Thurgood 1999 but which are nonetheless similar in shape to elements to be found in Malay) are also indicated. This is because these forms do not count as proper cognates but need to be recognised, somewhat paradoxically, as 'non-cognate' because they have entered Acehnese from Malay as loans.

I have also, for the sake of interest, sampled and surveyed a few further forms across Chamic languages, over and above the Blust list gloss forms. All of these are taken from the traditional Swadesh lists. These spare forms are 'to sing', 'five' (which in Malayo-Chamic languages is generally distinct from the form for 'hand'), and 'to play' (a form which I selected specifically as it is one of the most lexically diverse forms or 'characters' in Indo-European: Ringe at al. 2002. It certainly does not share that distinction in Chamic, since most Chamic languages use a reflex of PMP *maqin). I have also collected both the inclusive and exclusive forms of the pronoun 'we'; this last is a distinction that is of Proto-Austronesian vintage, and one which is perpetuated in very many of the modern languages, including the Chamic ones apart from Tsat (Mon-Khmer languages often make this distinction too whereas Chinese does not, and this typological parallel may account for its preservation in most Chamic languages). The inclusive 1pl form is represented as item 204. (The items numbered above 200 have not been further included in my calculations, though the distribution of forms within them and the variety of forms to express them within Chamic are facts duly noted.)

There are certain considerations and certain desiderata to be taken note of when using the Blust list in this operation. The desiderata constitute the aims and objectives of this project. I wished to see whether there was a valid Malayo-Chamic grouping within Malayo-Polynesian. I also wanted to see whether Chamic constituted a subgroup on its own, whether subgroups within Chamic could be identified and defined on the basis of lexical evidence, and where Acehnese fitted into all of this (and an important if secondary consideration was the extent to which Acehnese basic lexicon might have been influenced by later contact with Malay). I was also interested in seeing whether there were any PMP forms that were still preserved in Chamic which were not to be found in Malay, and if there was such a set of forms, I wanted to attempt to see why they were missing from Malay – had they been replaced in Malay by internal creations, or by external diffusions (lexical borrowings)?

In presenting my findings in the table I have started off with providing a code for the Proto-Malayo-Polynesian forms, which are uniformly logged here as 'a', because they come first on the chart. Next to the right come the letters indicating the cognacy or otherwise of these forms with those for Standard Malay (with loans into all languages indicated where known), and after this follows the column for the cognacy firstly with PMP, and secondarily with Malay, of the equivalent forms in Acehnese. I have continued to do this for the equivalents in several other Chamic languages: Western and Phan Rang Cham, Jarai, Rade, Northern Roglai, Tsat, Haroi and Chru. The sample of languages which I have surveyed is purposely limited, not least because of space constraints, and I have not provided comparable lexicostatistical information on other potentially interesting and relevant Western Malayo-Polynesian languages such as Madurese, Javanese or Tagalog, most of which, incidentally, appear to have preserved fewer of the 200 PMP forms in the Blust list than Phan Rang Cham has.

An important consideration in this study is the relative availability of the relevant kinds of lexical data. My sources were fullest for Phan Rang Cham (Moussay 1971), Acehnese (Daud and Durie 1997), Rade (Tharp et al 1980, also Egerod 1978 and Shintani 1981) and Northern Roglai (for which I used the list in Collins 1969 and some data from Bochet and Dournes 1953), and I have all the forms available for the lists for Jarai (Lafont 1968) and for Western Cham and Tsat as well, the latter thanks to the kindness of Robert K. Headley and Graham Thurgood respectively. Lexical data on Haroi were taken from Thurgood (1999) and Tegenhardt-Mundhenk and Goschnick (1977), and those for Chru come from Thurgood's book, from Fuller (1977) and also from Tín (1955), which also provides a Jarai glossary, together with lists in French and (the language of alphabetisation, and the source of the orthography for entries in Jarai and Chru) also Vietnamese. Thurgood's book was the main source for my data on Western Cham, together with papers in Thomas (ed., 1977, 1997) and Headley (1991), plus a few forms which Robert K. Headley gave me in personal communication, while for Tsat I used Zheng (1997, a source which was unavailable to Thurgood at the time of writing his book) plus one datum from Benedict (1941) for a single Tsat form which I was unable to find in Zheng's book.

The table could have been fuller. But I have reluctantly omitted a column of forms from the earlier stage of Written Cham (which shares a very high degree of lexical similarity with the two modern Cham languages) because I have too many gaps in my data, and I have available even fewer forms which are attested for Inscriptional Cham or Old Cham. I have also desisted from including a column of Proto-Chamic forms, whether they be those reconstructed by Lee or Thurgood, because I feel that an analysis of the distribution of particular forms across individual Chamic languages is the best first step towards reconstructing this portion of Proto-Chamic lexicon. It should be noted that the data which I use in this study have almost all been gathered by investigators working within the last 50 years or less, so that this exercise is a comparison of materials of roughly contemporary vintage. Many further forms that were not otherwise available to me were graciously provided by Graham Thurgood in personal communications.

The lexical material in Thurgood's book was the starting point for this work, and the basis and source for most of the entries on the grid. Since Thurgood's concerns there are primarily comparative rather than purely descriptive, it means that certain lexical forms which occur only in one Chamic language or which are not otherwise historically interesting are not going to be listed in his lexical lists, no matter how high the forms' text frequency may be. Such forms would include for instance the so-called 'lexical orphans' which may have been present in the proto-language but which are attested only in one modern daughter language. Others would be forms which have developed independently, which are unique to a particular language and are recorded for no other, or alternatively Mon-Khmer or other borrowings which no other Chamic language has taken up. On the other hand, it is unlikely that very many comparative Chamic cognate forms which are essential to this study, especially those of Austronesian origin, cannot be found in the lists in Thurgood's book, just as long as the forms for the relevant English glosses have been included in his lexical appendices in the first place.

We should also remember that, Phan Rang Cham apart (see the dictionary by Bui Khanh The 1995), we do not have voluminous lexica for any Chamic language of the kind which is available for Malay, and that indeed it may be the case some forms whose presence is alluded to in the table may have widely-known cognates in other Chamic languages, but that these cognates have simply not come to our attention because they are not noted in the available literature. All the columns in the table are at least 85% complete; the one with the most gaps is the Haroi. By contrast, the columns for Phan Rang Cham, Western Cham, Rade, Jarai, Northern Roglai, Chru and Tsat, and of course those for Malay and Acehnese, are complete and most of the rest are nearly so. Gaps in the Chamic lexical data which are available to me at the moment are infuriating, as they always are, but here they are not serious enough to distort or impugn the validity of the use of the particular methodology employed and the overall findings of this study.

4. Identifying some problems in Chamic lexicostatistical investigation.

Another consideration in this study was the suitability of the Blust list as manifested in the problems inherent in getting good forms for glosses. The semantic spaces of Chamic languages and that which is assumed for reconstructed PMP did not always coincide, although I did not substitute any of Blust's forms. Decisions sometimes had to be made as to what kinds of 'cutting' (chopping, hewing, splitting, slicing, etc.), 'lying down' (full length, prone or supine), throwing' (hurling, releasing an arrow discarding, throwing underarm as distinct to overarm throwing, or whatever) or 'turning' (spinning, revolving, flipping over, all of these either intransitive or transitive) were involved. There is also the issue of whether 'to smell' is intransitive (in which case the form is most likely Proto-Malayo-Polynesian) or transitive, in which case one chooses between a form meaning 'to sniff, snuffle' from PMP, or 'to sniff at, to kiss' from Mon-Khmer (but reflected also in Malay[12]). Furthermore, the semantic distinction between 'long in distance' and 'far', which is retained in Malay, seemed to be encapsulated in the same word in some (though not all) Chamic languages, while the distinction between 'wide/broad' and 'thick' does not seem to be made lexically in all Chamic varieties.

One or two forms are apparently compounds involving one or more forms which are attested elsewhere on the list. This is the case with the form for 'lake' in some language ('big water'), while in some languages 'to kill' is expressed by a form analysable as 'CAUSATIVE-to.die', thus involving a form which had already been found on the list. Furthermore, one or two forms which reconstruct to PMP are still recorded both for Chamic and Malay, but have developed new semantics in both languages. For instance the widespread Proto-Austronesian and PMP stem **qulu* 'head', which is realised as *hulu* in Malay, has been replaced by a Mon-Khmer loan, namely *'akó',* in the whole of Chamic and by a Sanskrit loan in Malay (and for that matter in Khmer), at least as far as the name of the anatomical part is concerned. Yet it still occurs in certain kinds of compounds in both languages (and it is used as 'head' in most metaphorical senses in Malay). For instance there is Chamic *dihlau*, Malay *d(ih)ulu,* both of these being forms with the meaning 'formerly', literally 'at+head' in Proto-Malayo-Polynesian (PMP **di* + *qulu*). Malay *dulu* has subsequently been grammaticalised as an indicator of completive aspect.

Other distinctions which are less easy to capture using the Blust list are those which involve pronouns, especially personal and interrogative ones. In Chamic languages the interrogative pronouns are often bimorphemic words involving a general interrogative form and a specifier which indicates such a sense as 'place', 'time', 'manner' or whatever. Consequently the same morpheme occurs in several glosses on this list, and this replication of the same interrogative morph happens in several Chamic languages. The Blust list glosses provide for only two demonstrative positions, namely proximal and distal, yet many of the languages here have at least three such forms in both pronouns and adverbs. As to personal pronouns, the Blust list assumes a system which involves a two-way distinction of number and a three-way distinction of person, without special reference being made to a distinction between inclusive and exclusive senses of 'we'. The system in Chamic languages is rather different. Except in the first person plural (where an inclusive/exclusive distinction is regularly made), number in pronouns is of secondary importance, although three persons are regularly distinguished. The primary division in

[12] A catalogue and analysis of the Mon-Khmer component in most or all varieties of Malay, which is not massive but not negligible either, is long overdue.

most Chamic languages is between polite or formal versus ordinary pronouns – and this is a distinction which is by no means unknown in Southeast Asia. In addition, the ordinary word for 'I' in Chamic languages is the normal PMP one, whereas the formal word for 'I' is derived from the PMP form meaning 'slave', thereby perpetuating a trope which is also found (inter alia) in Vietnamese *tôi* and Malay *sa(ha)ya* (this last being a borrowing from Sanskrit).

An exception to this patterning is provided by Tsat, which has developed new 2pl and 3pl pronouns by combining the relevant singular pronouns with a suffixed *za:ng,* a Malayo-Chamic form meaning 'person' (cf. Malay *orang,* Phan Rang Cham *uraang* 'person'). This is exactly what many forms of Min Chinese (including Hokkien and Hainanese) have done. Coincidentally it is also what has happened in those varieties of Malay which have also been in touch with Hokkien, or which have developed at a later date from such varieties, such as Betawi of Jakarta in the first instance, and Baba Malay, Sri Lanka Malay and Cocos Malay in the second instance (each of which are developments from Betawi; the observations are based on personal communication from Graham Thurgood in the first case, and Adelaar 1991, 1996 for Sri Lanka Malay and Cocos Malay) in the second. Such dialects have, for instance, *dia orang* '3sg-person' for 'they'. Although some speakers of Tsat have contact with formal Malay through Islamic teaching, we cannot assume that this particular structural parallelism has developed or been percolated through the effects of Tsat contact with Malay, because this construction is not typical of the particular formal Malay lect to which Tsat speakers have been exposed through religious work, which would use the Standard Malay 3pl personal pronoun *mereka.* Instead, what we have here is the development of the same structure within a pronominal system as the result of influence from the same kind of Chinese language upon two related languages, but we see that it developed separately in two areas and on two occasions where the same kind of language (in other words, a Western Malayo-Polynesian one) happened to have been influenced by varieties of Min Chinese.

The results of this investigation are presented in the table below. What then are the outcomes of this experiment? We can imagine a set of outcomes each being displayed on a number of occasions in the result in the table. The first outcome shows the Chamic languages retaining forms inherited from PMP. The second shows them retaining forms inherited from Proto-Malayo-Chamic, in which these forms had developed. The third outcome shows Acehnese having the same form for a particular gloss as other Chamic languages do, but Malay differing from these (and maybe also from PMP). The fourth scenario would have the Indochinese Chamic languages (with or without Tsat) sharing forms which are not also found in Acehnese, and which we assume to have developed at a time when Acehnese had split away from the other Chamic languages, which were all still in contact with one another and which were in a position to diffuse items to one another. Some of these innovations may be loans, as may some forms which bind Acehnese and other Chamic languages together against Malay and PMP. Another series of outcomes would indicate the development of subgroupings within Chamic, say a Jarai-Rade subgroup, which are marked out by the development of shared lexical innovations (including loans), which have replaced forms which have otherwise been conserved in other varieties. Another set of outcomes would show Tsat as being either conservative or, more probably, especially lexically innovative (as the result of borrowings) against the consensus of the evidence of the Indochinese Chamic languages. If it had conserved forms whereas Indochinese Chamic languages had all shared in the introduction of an innovated form, this might have some historical significance. And there is the possibility that for a certain period of time all the Chamic languages had gone their separate ways and were still doing so (though latterly many were following some of the same paths of conformity as a result of sharing cultural borrowings from Vietnamese, which was the language of power in most or all the communities under discussion). We can find instances of all of these scenarios in the table below, although I should point out that the direct impact of Vietnamese on the contents of the basic Blust list lexica of any of these languages is negligible.

5. Some observations on the results.

How does the use of this bundle of techniques work out in practice? What can we learn from its application? For a start, the rows of straight ‘a’s which run through Malay, Acehnese and the other Chamic languages (with occasional interruptions due to lexical replacement in one or more languages) show that the Austronesian (and more certainly the Malayo-Polynesian) affinities of Chamic languages are manifested very clearly in the lexicon (and in most of what remained of the inflectional morphology of these languages). Indeed there are even a few cases in which the lexicon of modern Malay has replaced or shed a preexisting PMP form, which has nonetheless been retained in Chamic languages (and in these instances sometimes Acehnese has borrowed the Malay form, while on other occasions it has retained the same inherited form as the Chamic languages). This is the case, for instance, with the word meaning ‘to bathe’.

These instances of lexical replacements of inherited forms will have occurred at some time after the split of Chamic and Malayic, a split which occurred a few centuries before the birth of Christ. In this respect it is significant that some of the forms which have been lost in Malay have been replaced there by words which derive from languages with which Proto-Malayo-Chamic could not have been in contact, namely Sanskrit and Arabic.[13]

In an analysis of the items entered on the grids I counted 108 forms (out of 200 glosses), occurring in one or all the Chamic languages (Acehnese apart) which trace back to Proto-Malayo-Polynesian, and three of these forms consistently show phonological irregularities which accord with those for the same cognate forms in Malay[14], giving some credence to the establishment of a special Malayo-Chamic group. As far as I can tell, all but one of these forms (the exception is the form for ‘flesh, meat’ deriving from PMP **hesi*) also occur in Phan Rang Cham, while one further inherited PMP form (the reflex of PMP **naŋuy* ‘to bathe’, above) also occurs only in Acehnese but has been replaced by other forms in the remaining Chamic languages.[15] In addition, there are a number of forms

[13] The lexica of the Chamic languages have provided minimal evidence for the reconstruction of Proto-Austronesian and Proto-Malayo-Polynesian, so that they have not been explored much, and indeed there are rather few inherited forms which are preserved in Chamic languages which cannot also be found in Malayic lects.

[14] For instance they may be stems which in both sets of languages incorporate the form of an infix (present at PMP level, but obsolete as a productive morphological device by the time of the first attestations of Malay in the late 7th century AD) into a newly metanalysed stem. The form for ‘to dream’ is an example of this.

[15] Compare this tightness of bunching with the situation in Oceanic, in which the vast majority of forms which have been reconstructed for the Blust list for PMP are attested as inherited forms in at least one Oceanic language (and the Samoan list has almost half of these, involving 84 of the 200 forms reconstructed in the list for PMP and an even greater proportion of those

(I counted 30 such) which do not occur in PMP but which also occur in Malay as well as in Chamic languages, and the existence of this cluster of lexical innovations bolsters the claims for a Malayo-Chamic group too.[16] 22 forms on the list are shared between Acehnese and some or all of the other Chamic languages, but do not occur in Malay or in PMP, although several of these are loans from Mon-Khmer languages rather than being innovated forms that were first generated at the Common Chamic level. Still, they strengthen the evidence for a historical genetic link between Acehnese and the Indochinese (and Tsat) Chamic languages (while in one further case, Acehnese and Tsat have preserved a PMP form which has been lost in Indochinese Chamic).

By contrast, at least 44 Chamic forms, many of them pan-Chamic, certainly or probably derive from a Mon-Khmer language. Another pan-Chamic form ('dust') derives from Sanskrit by way of its having previously been borrowed into Mon-Khmer languages such as Khmer, and at least 10 further glosses have equivalents which are pan-Chamic in spread, but for which an origin has yet to be found in any known language. Meanwhile 2 further Blust list glosses are variously expressed in Chamic languages, sometimes being expressed by Mon-Khmer forms and sometimes by widespread forms, which are found in several Chamic languages, and which are of unknown origin.

reconstructible back to Proto-Oceanic), but where most Oceanic languages lack most of these forms, while some of the forms which can be reconstructed back are found only in a few Oceanic languages.

[16] These 108 forms inherited from Proto-Malayo-Polynesian constitute an unknown but certainly high proportion (at least one third of such forms and probably much higher) of the total of such forms which any or all Chamic languages have inherited from their ultimate genetic ancestor (Proto-Austronesian) or which have been acquired from this ancestor's descendants (Proto-Malayo-Polynesian, Proto-Malayo-Chamic) which are nonetheless antecedents of Proto-Chamic. By comparison, on the Blust list for Standard Malay 112 items out of 200 derive from PMP – and this is the highest proportion of such forms which has so far been recorded for any Malayo-Polynesian language (Blust 1990). This compares with 89/200 retained PMP forms for the Blust list for Chamorro, for example, and with a miserable 10 retained PMP forms out of 194 attested equivalents in the Blust list for Kaulong of New Britain, a language whose Austronesian affinities (within the Pasismanua languages of the Oceanic branch) have never been in doubt. This last figure is less than ¼ the number of *attested Mon-Khmer loans* which are to be found among the Chamic-language equivalents of the Blust list! (The estimate of Headley 1976, to the effect that Mon-Khmer loans accounted for about a tenth of the reconstructed Proto-Chamic lexicon, is set too low.) 3 further forms deriving from PMP, which are replaced in Standard Malay by loans from other languages, occur on Blust lists for some non-standard Malay varieties. According to my calculations the comparable score for Acehnese is 110/200, though some of these 'inherited forms' may actually be unrecognised Malay back-borrowings into Acehnese. The bulk of the recognised Mon-Khmer elements in the Acehnese list are also found in other Chamic languages and reconstruct back to Proto-Chamic, and this is also true of some of the forms which are as yet of 'unknown' origin. As speakers of Acehnese never returned to Champa, the presence of such forms in Acehnese can only be explained by reference to a previous period of common development between Chamic languages and Acehnese, during which contact with Mon-Khmer languages occurred, leading to lexical transfer. On the Blust list some 7 forms which are of PMP origin but which are not recorded in Malay are attested in at least one Chamic language; in Malay these have either been replaced by loans (*nama* 'name' from Sanskrit, expressed in Jakartanese by the Javanese loan *ngaran*, a form which is cognate to the lost Proto-Malay form) or by innovated forms. Blust (1981a) provides scores for two Chamic languages; according to his calculations he assesses Acehnese as retaining 81 items out of 200 and Jarai as retaining 73 out of 200.

What is most striking is that the Chamic languages show a very high degree of lexical similarity and internal lexical homogeneity, no matter what the origin of the individual lexical items may be, and this is especially true when Acehnese and Tsat are removed from the picture. I found 162 instances out of 200 in which either all the Chamic languages in the table from Phan Rang Cham onwards shared the same form (of whatever etymological origin), or else all these languages bar one for which I had an attestation of a gloss for the particular form used forms of the same origin. Proto-Chamic equivalents, which are reconstructible at least to the period after which Acehnese had split off from the other languages and often much further back, could be reconstructed for at least 85% of the items on the Blust list simply by using traditional methods and by drawing upon traditional kinds of evidence for proving the existence and shape of lexical reconstructions.

In addition to the forms on the Blust list which go back to PMP and for which reflexes can be found in at least one Chamic language, there are 13 further forms which are post-PMP but which are found in Malay and in Indochinese Chamic languages as well as in Acehnese, so that they reconstruct back to Proto-Malayo-Chamic. There are 19 further forms which are common to Acehnese and other Chamic languages but which do not occur outside this subgroup so that they are not found in Malay, and there are at least 45 further forms which are common to two or more Chamic languages outside of Acehnese, and many of these have Mon-Khmer etymologies, as have some of the 19 forms which are common to Acehnese and other Chamic varieties. Indeed the stratum of forms of Mon-Khmer origin which are found in all Chamic languages is bigger than the stratum of common Malayo-Chamic innovations, and the stratum of forms of common Chamic heritage but of unknown origin is also broader than that. A few further forms probably reconstruct back to Proto-Chamic on the basis of their widespread distribution in modern Chamic languages, but they are not found in Cham proper or in Acehnese. And it is possible that these numbers are themselves underestimated, and that the lexical uniformity within Chamic (and especially within those varieties still spoken in Indochina) may be greater than these suggest. By contrast, there are few forms on the Blust list which it would be almost impossible to reconstruct using the judicious application of standard historical linguistic methods. But there are also some glosses (the verb 'to throw' being an especially vivid example, in Chamic as in many other language families) which exhibit a very large number of different forms among the dataset for this form for the modern Chamic languages. Indeed, if we had data for Haroi forms meaning 'to throw', it is likely that there would be more than the six separate forms listed which have so far been attested for the Chamic languages surveyed (let alone the other forms that have been noted for PMP, Malay and Acehnese, which all differ from one another). But we need to be mindful of the fact that 'to throw' is one of those forms for which many languages have more than one equivalent, depending upon the nature of the item thrown, the trajectory of the throwing action, the question of whether the item projected hits its target or not and so on. We cannot blandly assume that the semantic range covered by any, most or all of the forms meaning 'to throw' in the various Chamic languages is identical in any or all the languages.

To some extent this widespread core lexical similarity within Chamic languages is a continuation of the manifestation of other clear cognacies. It is beyond doubt that the various Malay lects and Chamic lects subgroup with one another against other Malayo-Polynesian languages, and that they share some common and irregular developments of inherited forms which other MP languages do not. It is beyond doubt that Acehnese and

Chamic fit together in a subgroup against Malay and with one another (we may state this securely despite the presence of a few high-frequency Malay loans in Acehnese, though there are none to be found in corresponding lexical strata in Chamic. This is unless the word for 'green' in Phan Rang Cham and Western Cham is an unrecognised borrowing from Malay rather than a retention from Proto-Malayo-Chamic, in which group it would be an innovation against the inherited PMP form). It is clear that Acehnese has gone its own way in matters of lexical change, loss and replacement for a time, at least before being 'swamped with Malay loans' (Blust 2000a)[17], and it is clear that Tsat fits in lexically with Chamic despite the wide typological and the lesser lexical differences between Tsat and even Northern Roglai, its probable closest genetic relative – in which Tsat is the innovating partner. The presence in the Tsat lexicon of a subset of the same Mon-Khmer-derived loans which one finds in Northern Roglai (and which also generally occur in other Chamic languages) is highly significant here as an indicator of Tsat's Chamic affinities and origins.

What makes this considerable lexical uniformity within Chamic (a uniformity which is somewhat underplayed by the lexicostatistical results presented in Tables 2a and 2b) so remarkable is the fact that it is accompanied by an impressive degree of contact-induced phonological diversity from one Chamic language to the next. (There is less internal diversity in regard to Chamic morphological systems, apart from the conservative features of Acehnese morphology.) It is highly unlikely that Rade and Jarai, or Haroi and Phan Rang Cham, or whatever, are mutually intelligible, despite the similarities of their basic lexica, and much of this is due to the different outcomes of each language's reflexes from Proto-Chamic.

It is fitting that Thurgood had to reconstruct Proto-Chamic phonology from the top down (and also from the bottom up), working from hypothesised Proto-Chamic forms which more often than not bore a strong resemblance to those which are still found in more conservative dialects of Malay. This is because any attempt at reconstructing Proto-Chamic simply by working from the bottom upwards, using only the evidence of the modern Chamic languages (even if Tsat were excluded and if Acehnese data were mined solely for their conservative features) would have made the task immeasurably more difficult. This is especially true of anything affecting the reconstruction of the shapes of presyllables. It is also true that numerous phonological irregularities remain in the forms of Thurgood's Proto-Chamic reconstructions; we simply do not know everything about the phonological history of Chamic languages. Many loose ends still remain at the level of the reconciliation of troublesome facts about individual word histories in these languages (for instance the wide range of disparate and 'irregular' word-final consonants and vowels which Thurgood lists for many of his Proto-Chamic reconstructions).

The degree of morphological diversity among Chamic languages, especially as far as the use of bound inflectional morphology is concerned, is less than that which is found

[17] Part of this lexical self-direction on the part of Acehnese has involved the absorption of Mon-Khmer loans (presumably Aslian ones from languages of eastern Malaya, but maybe also some further Katuic ones) which are not found in other Chamic languages. These loan strata have yet to be identified or worked upon fully, although a good place to start would be among the large number of monosyllabic contentives found in Acehnese which have no PMP, Malay or Chamic parallels. Further attention also needs to be paid to the Mon-Khmer lexical stratum in Malay, which is not inconsiderable in size or in centrality to the everyday Malay vocabulary (though it is especially rich as a source of ecological terms), but which has yet even to be listed comprehensively.

in the Chamic segmental, canonical and other phonological systems, but this apparent uniformity is largely a result of the overall paucity of inflectional morphology in these languages to begin with. The Chamic language which stands out the most from the others in the realm of morphology is Acehnese, which looks incongruous when it is compared with other Chamic languages or with Malay. But this is because in many respects (for example in its possession and productive use of verbal infixation or derivation) Acehnese has been conservative, and as such resembles non-Malayo-Chamic but nonetheless Western Malayo-Polynesian languages such as llokano and Tagalog, whereas Malay and the Chamic languages have innovated in shedding this morphology.

The loss of productive use of inflections is a process which has happened extensively but separately in Chamic and in Malay; it naturally dates after the split-up of Malayic and Chamic, and occurred under separate sets of social circumstances and as the result of intense contact from different sets of languages. (Thurgood 1999: 43 finds another structural parallel of this post-split typological difference between very closely related languages within the realm of Western Malayo-Polynesian. He points out that Malagasy, which historically and genetically is a Bornean language of the Southeast Barito subgroup which was removed c. 400 AD from that island to Africa and thereby from the full-scale morphological effects of intensive and submissive contact with Malay[18], preserved the inherited morphological feature of infixation. In contrast, Malagasy's unrelocated relatives among the Southeast Barito languages of Borneo, that is to say languages such as Ma'anyan which were all much more heavily exposed to direct and continued influence from Malay than Malagasy was, eventually lost their infixes and simplified their morphology). A table illustrating this situation and comparing the morphological typologies of a number of relevant South East Asian languages can be found as Table 3.

Some representative scores for the percentage and number of shared forms (of whatever origin) between particular pairs of Chamic languages include the following sets of results:

Malay/Acehnese: 135 items out of the 199 discrete forms which were recorded in the available data (although 4 of the shared forms may actually be loans from Malay into Acehnese),

Phan Rang Cham (henceforth PRC)/PMP: 107/198[19];
PRC/Acehnese: 133/198;
PRC/Standard Malay: 102/198;
PRC/Western Cham: 194/197;
PRC/Jarai: 177/198;
PRC/Northern Roglai: 177/198,
PRC/Tsat 171/198,
PRC/Rade: 168/198,
PRC/Haroi: 171/182,

[18] The lexical impact of Malay upon Malagasy, though, could be found in some surprisingly basic realms of vocabulary, for instance the names given to body-parts (a number of such examples are given in Adelaar 1995).

[19] Despite the fact that we have complete 200-item Blust lists for PMP, Malay, Achenese, PRC, Jarai, Rade, Chru, Tsat and Northern Roglai, the numbers of compared forms add up only to 199 (where Acehnese is involved) or 198 (if any other Chamic language is involved) because of the duplication of certain stems in the system of plural pronouns; we cannot count the same stem twice.

PRC/Chru 187/198.
Western Cham/Malay: 102/198;
Western Cham/Acehnese: 131/198;
Western Cham/Haroi: 158/180;
Jarai/Rade: 175/198;
Rade/Northern Roglai: 173/198;
Haroi/Chru: 178/180;
Chru/Jarai: 180.198;
and Tsat/Northern Roglai: 143/198.[20]

There are also 89 forms out of 200 which meet two conditions: they are shared by Malay and PRC, and they reconstruct back to PMP. 3 such forms show Chamic phonological modification of the original PMP form in a way which is also shared with the cognate form in Malay (for instance we have Acehnese *lumpeuy*, Malay *men-**impi***, PMP *h-in-ipi*, an infixed form of earlier PMP **hipi*, all of these meaning 'to dream'; the Malay form involves the addition of a modern Malay prefix to a stem which includes an infix which is no longer identifiable as such to modern Malay speakers), and 30 forms are Common Malayo-Chamic, inasmuch as they are found in Chamic, and Malay, and sometimes Acehnese, but are not among the PMP forms. I have used PRC data here in this comparison since this is the Indochinese Chamic variety for which my data were fullest and clearest at the time when I first did my calculations. In addition it is the Chamic variety which has strayed the least geographical distance from the historic centre of Champa.

By contrast, there are at least 21 unique items (items with no cognates in any other Chamic language or elsewhere) out of the 200 Tsat forms which were available to me for completion of the Blust list, though rather few of these unique items are taken from a Chinese language (nor, as far as I know, do they derive from Hlai). Indeed the origin of most of these forms which are unique to Tsat is uncertain and there are no clear instances among them of unique retentions, within the range of exemplified Chamic languages, of forms from PMP which have been replaced elsewhere within Chamic by borrowed or innovated forms. One item from Western Cham (the word for 'spider') apparently derives from or is influenced by the form in Khmer (Robert Headley, personal communication).

The number of items that have been retained from PMP in the lists in the various Chamic languages is given in the table below. Cases where a PMP descendant and a form of other origin coexist have been marked as though they were pluses. Cases where the same form is used in more than one gloss (for instance where the same morpheme is used in both the singular and plural pronouns, or cases where 'short' and 'small' are expressed by the same root) are only counted once, however, which explains why some languages with full lists show totals under 200. This is also the practice where the form in question in a particular language is a compound of two elements, both of which are already separately logged in the table. The figures are as follows:

[20] The proportion of cognates on these lists which can be found between several of these pairs of languages (which are of course instances of pair-referenced lexicostatistics!) are several percentage points higher than those cited in the Tables 2a and 2b. But since different lists have been used in the present study from the ones used to calculate the percentages in Tables 2a and 2b (which themselves are based on the results gleaned from slightly different lists), no direct comparison of these sets of percentages is possible. In those cases where one language has two equivalents for the same gloss, one cognate with another form and the other not so, the cognate form is the one taken notice of in my calculations.

PMP/Standard Malay: 112/200;
PMP/Acehnese: 114/199 (but this total possibly minus a couple of as yet undetected Malay loans);
PMP/PRC 107/198,
PMP/Western Cham 105/198,
PMP/Haroi: 105/180,
PMP/Chru: 105/198,
PMP/Jarai: 100/198,
PMP/Rade: 98/198,
PMP/Northern Roglai. 98/196,
PMP/Tsat: 101/198.

In all these cases those forms which have been retained from PMP account for over 50% of the forms on each Chamic Blust list.

If we want to track uniquely shared lexical innovations within subgroups of Chamic as a means of ascertaining the scope of any subgroups (say Highland versus Coastal Chamic) we need to assess which forms are common to all the Chamic languages, or which have been replaced by loanwords in one or more cases. We then need to identify and discount these loanwords, and we also need to establish and set aside any bodies of what we may call *uniques*. The number of 'uniques' found in the Blust lists for other Chamic languages is much smaller. By 'unique' I mean what is sometimes (albeit pedantically inaccurately) referred to as a *hapax legomenon*, namely a phonological form which is only found in a single language and which is assumed to be an innovation within that language. For instance the verb 'jump', for which no etymology is known, is a unique within English.

Rade, which stands out from other Chamic languages in a number of linguistic respects, including the phonology of its presyllables, has only five uniques in the list (plus maybe another one), including a form for 'eye' which refers to the yolk of an egg in other Chamic languages (Rade has lost PMP **mata* in the sense of 'eye'), and the number of uniques in the other languages is even smaller. There is only one unique form given for Blust list glosses in the (admittedly incomplete) data for Haroi, for example, there are only two uniques each for the same bodies of data for Phan Rang Cham and Jarai, and there are none for the Blust list items for Chru or for Western Cham. There also do not appear to be any forms on this list which are lexical innovations (rather than borrowings) that are exclusively found in Northern Roglai and Tsat (which has 21 unique forms of its own), though there is an abundance of inherited Proto-Chamic forms which are common to these languages.

Given the understandably large role which Thurgood (1999), a volume with an admittedly comparative approach to Chamic, has played in the assembling of these data, it is probable that, if we had fully-recorded lists for all the above languages, that there would not be an appreciably greater number of shared cognates than we already find, and that consequently the overall percentage of cognate forms between any pairing of two Chamic languages would be lowered accordingly, even though the cognates which have so far been recognised between various Chamic languages would remain.

This leaves open the question of how (if at all) one should interpret the silence of our information on certain languages (especially Haroi and Chru) in regard to the potential existence there of words which are found in many or most other Chamic languages. Since the Haroi and Chru equivalents for many glosses have not been made available in Thurgood's comparative Chamic lists (which is the source for most of my Haroi and Chru forms), are we to assume that Haroi and/or Chru express each of these ideas by using words which are not found anywhere else in Chamic? Are the words which these languages do use to express these concepts recent borrowings from Mon-Khmer languages such as Vietnamese or Bahnar, or are they sometimes forms whose origins are as yet unknown and which may have originated within the languages themselves? Or is it simply the case that Haroi and Chru cognate forms of well-known Common Chamic words have not been recorded in the materials available to us? It is impossible for us to say, given the information currently available to us. We can only work with what we have and we cannot employ *argumenta a silentio* to help us out.

The lexical forms in Acehnese which are not shared with other Chamic languages mostly fall into two groups. There are those which are similar to forms in Malay and which look as though they may have been borrowed from Malay, and there are those whose origins are uncertain, though some of them may derive from Aslian languages (however, no convincing etymologies for these latter have been found yet). Among the 200 forms on this list only the Acehnese form for 'to swim' preserves a PMP form, in this case a reflex of PMP **naŋuy*, a form which has been by chance replaced (albeit by different words) both in Malay and in other Chamic languages, and which in Acehnese shows the distinctive Chamic change of *n-* to *l-*: Acehnese *languy* 'to swim'. The replacement word for 'swim' in the other Chamic languages is pan-Chamic, and is most probably borrowed from a Mon-Khmer language. The Malay form for 'swim' is pan-Malayic in distribution but I do not know its origin. One further form, a retention from PMP, is shared between Tsat and Acehnese but not with the other Chamic languages.

A rough and ready indication of the relative degree of linguistic diversity in Chamic can be provided by simply counting up the number of different forms used in the aggregation of Chamic languages in expressing the 200 glosses on the Blust list and then expressing it as a ratio. Acehnese apart, eight lists have been used, those for the Phan Rang and Western varieties of Cham, for Jarai, Rade, Northern Roglai, Tsat, Haroi and Chru. Although the Chamic-language material available to me has serious lexical gaps for Haroi (and there are more gaps here for Haroi than there are for the comparable Malayic lists), I have found that the number of forms used for expressing the 200 concepts on the sum total of the Chamic lists, apart from Acehnese, is 300, that is, a ratio of 1.5 different forms per gloss across a sample of eight languages. (A ratio of 1.00 would indicate to us that all the languages were identical isolects with nothing to distinguish one from another; a ratio of 8.00 would highlight to us that all indications suggested that the eight languages were completely unrelated to one another.) I have full lists for Phan Rang Cham, Western Cham, Jarai, Rade, Chru, Tsat and Northern Roglai; the list for Haroi has 17 omissions.

Now these 300 forms cover 1568 filled slots. The number of slots is arrived at as follows: Ideally I would have 200 forms from the 8 sampled non-Acehnese Chamic languages, making 1600 slots on the grid for these languages. But I have 17 gaps on my table for glosses for which I lack a form in one language. Additionally there are gaps in the columns for most Chamic languages for the 3pl pronoun form, since it is identical to the 3sg form in nearly all Chamic languages, and the same is true of most 2pl pronouns, while some languages also use the same form for 'short' and 'small'. These slots could potentially be filled by 1568 different items, if it were the case that the languages in question bore no lexical resemblances between each other whatsoever. But in fact only 300 separate items are used (excluding a handful of cases in which one language uses two different unique forms to express the meaning of a particular gloss – only one unique is counted for such slots in each case). This makes this a ratio of 5.26667 slots per individual glossed item (for whatever this ratio may actually be worth; please note that this figure is the **reciprocal** of the figure for the average number of cognate sets per word across the eight languages surveyed).

If one adds into this total the forms on the list which are only found in Acehnese among the Chamic languages (whether or not they are also retained from PMP or are also found in Malay, or whether they are innovations within Acehnese), the total of different forms rises to 361 and the ratio of unrelated forms per gloss therefore rises to 1.85 forms per gloss across a sample of nine languages, exhibiting a total of 1756 slots (for we have a full list for Acehnese), making this a ratio of 4.8642659 slots (out of nine slots available for each gloss) which are occupied on average by each individual glossed item.

This overall very high degree of congruity and commonality of the basic lexicon in Chamic is, we should point out, in marked contrast to the very wide degree of phonological variation (if one views the matter diachronically) which is found across these languages and which is even amply exemplified in the various phonological shapes of those forms which have been inherited from PMP, but which is especially vivid in fully-tonal Tsat.

By comparison, the number of different forms which are used for the equivalents on the eight wordlists which were given for various Malay isolects in Blust (1988)[21], a dataset which has fewer overall gaps than the Haroi list has, and one which represents a group of isolects whose genetic unity has never been in doubt, is 490, or 2.45 forms per individual glossed item across a sample of eight isolects. Were Blust's Salako (or Selako) Dayak Malayic list fuller for our purposes (but unfortunately it is not, as it contains only 173 of the Blust list forms out of a target of 200 (though Sander Adelaar has provided me with the Salako forms for the missing entries), while one form is missing from his Ambonese Malay list), the number of discrete items in use here (and the proportion of items to each gloss) would certainly be higher and it might push the average figure for the number of cognate sets per gloss above 2.50. This is because the material which Blust presents, though incomplete, nonetheless shows that Salako is lexically innovative when contrasted with other Malayic isolects.

So what do we find when we look for shared innovations in an attempt to subgroup the Chamic languages? Not a lot, really. We can make a solid start at answering this question, since we know which forms are inherited from PMP or Proto-Malayo-Chamic and which other widespread forms in Chamic are actually innovations within Chamic (including or excluding Acehnese). We also know which forms on the lists are 'uniques' and which forms are loanwords from various sources. There are also a few cases in which one or more language has two forms to express one gloss, and one or both of these forms are uniques, a fact which also inflates the figures slightly. If we subtract these strata, then

[21] The Malay lects which are surveyed in that article are Bahasa Indonesia, Banjarese, Medan Malay, Salako, Iban, Minangkabau, Jakartanese (Betawi) and Ambonese Malay (Bahasa Ambon). Adelaar (1991) uses five of these lists and also provides a directly comparable wordlist for the Middle Malay language Seraway, providing equivalents for 188 out of the 200 items on the Blust list. He additionally reconstructs Proto-Malayic forms from this evidence wherever possible. Neither author provides a list for Kerinci, which is usually classified as a phonologically divergent dialect of Minangkabau, although we do know that it retains 100 out of the 200 PMP forms that are used on the Blust list (Blust 2000b: 329). I have only recently had access to Blust's list for Kerinci (Blust et al. 2005) and have therefore not used it in the above work.

what remain should be the clusters of exclusively shared lexical innovations. The problem is that so little material is left to us after these subtractions. Even some assured historical linkages, such as that of Northern Roglai and Tsat, are not manifested by large bundles of shared lexical innovations in our data; Thurgood (1999) shows us that the strong evidence for this history of shared development is actually phonological. In order for us to have strong evidence, from the basic lexicon, of substantive subgroups within Chamic (apart from Acehnese, which stands somewhat on its own, by virtue of having retentions, innovations and numerous loans from Malay) we would need to find clusters of lexical forms which have developed independently among two or more Chamic languages at a period after at least the beginning of dialectal differentiation within Chamic, and this we do not find to any striking extent.

In regard to a possible Highland versus Lowland Chamic division, Jarai and Rade seem to share a couple of forms on their translations of the Blust list which are not also found in PMP, Malay, Acehnese, or Western and Eastern Cham (forms standing for 'dust' and 'to spit', for instance Rade *ɓruih* and *bah* respectively; I cite these forms from Egerod 1978). But this 'Highland Chamic' group is weakly supported, and there is no innovatory lexical evidence in the basic vocabulary for a similar coastal group comprising (say) Chru, Roglai and Haroi.

On a final note, it should be mentioned that the origins of the various forms are not easily stratifiable by form class. Mon-Khmer borrowings into Chamic languages include verbal and several pronominal forms in addition to nouns, while the stratum of forms of uncertain origin also includes some pronouns. The form class which is most homogeneous in terms of its origin is that of the numerals, which are robustly Austronesian or at least Malayo-Polynesian in origin.

6. Summary of findings

The distinction between horizontal and vertical lexicostatistics has been discussed above. In this study both techniques are used, firstly the horizontal and then the vertical, together with norm-referenced lexicostatistics in which the norm used provides the vertical element in the study, and the various techniques tell us different kinds of things. (It is therefore essential to employ the several methods in the correct sequence, otherwise the end result is pseudo-statistical nonsense.) Comparison of lists for various modern Chamic languages is an example of horizontal lexicostatistics, whereas the use (as the cross-referencing 'norm') of Proto-Malayo-Polynesian reconstructions against which to compare the occurrence or non-occurrence of such forms in modern Chamic languages is an instance of vertcial lexicostatistics. The inclusion in the study of tabulated findings from lists from modern Standard Malay and from modern Acehnese allow us to examine diachronic issues which have to do with Chamic languages, and they allow us to appreciate that Malay is the most closely related language grouping to Chamic and that Acehnese is equidistantly similar to all the (more conservative) Chamic languages. This is just what one might expect from a language which derives from a Chamic variety whose speakers left Indochina in a period before the Chamic lects had had opportunity to separate into different languages.

The status of Acehnese as a historical witness is reinforced by the fact that it retains morphological features, for instance the use of productive infixation, which were common to Proto-Chamic and Proto-Malayic and further back in time, to Proto-Malayo-Chamic, but which were subsequently lost (or reduced to lexicalised vestiges) in all the other Chamic languages and in Malayic ones too. Its status as a lexical witness is somewhat diluted by its wholesale absorption of words from Malay. Malay, Acehnese and the other Chamic languages have all lost many features which their common ancestor had retained and which it had often retained from PMP or even earlier, but they have not always lost the same things.

The use of the informative but still undervalued technique of norm-referenced lexicostatistics makes the degree of similarity between pairs of languages much clearer than the normally-used technique of pair-referenced lexicostatistics does. It also enables us to see what kinds of forms are shared between languages, which other forms differ in any or all languages, and we can also see whether there are any forms which seem to buck otherwise prevalent linguistic trends – the presence of stray retentions from the common proto-language in one language when all other languages in the sample have shared an innovation, for instance. If the norm which is used as the point of comparison with the other languages is an earlier stage of an attested language, or a reconstructed proto-language, **but only if it is a proto-language which has been reconstructed without reference to the particular languages which are under discussion in the sample being examined**, the findings can be far more informative and they may give a much clearer historical picture. Such information, often regarded as too cumbersome to present in part (as I do here with the grid) or in whole (as would be done by reproducing the exact forms) can shed light on what lies behind the bleak tables of unannotated percentages which Dyen and his followers have offered.

This 'criterion of primordial objectivity' is clearly met here, because Proto-Chamic and its descendants have played little or no part in the reconstruction of Proto-Austronesian (PAn) or of its daughter proto-languages, so that the process does not involve an excess of application of circular reasoning. PMP or PAn reconstructions can this be used as much more objective yardsticks to cast certain sorts of light on Chamic linguistic history. The advantage of using forms from an actually-attested, or at least well-reconstructed, earlier stage (a 'parent language') of the languages under examination in such a sampling is that one can mark up the rows on the grid to show which forms are maintained from the parent language and which are replaced by innovations (or borrowings) in each of the 'daughter' languages under observation. Having done that, it is then possible for us to plot patterns of occurrences of these innovations, and to see to what extent these correlate across and within the daughter languages.

Lexical material has been privileged in this study because of the paucity of elements of bound morphology in the Chamic languages, and the list which I have used is one which is supposed to be especially suitable for the exploration of the histories of Austronesian languages. Other language families would require the use of other lists, but there is no reason why norm-referenced lexicostatistics should not be used as part of the battery of tests used to determine the genetic affiliation of 'troublesome' languages, and in the case of the Chamic languages, where the usually diagnostic bound morphology is so sparse (and is sometimes clearly borrowed), it happens to be especially useful.

Bound morphology is the kind of material which would normally be looked upon as providing firmer evidence for the Austronesian affinities of Chamic and for placing Chamic in the right niche within the Austronesian family tree than would normally be provided by lexical evidence. The fact that most of the rather few productive bound morphs in Chamic languages are typologically, semantically, formally and functionally very similar to those found in Mon-Khmer languages, not least the Central Bahnaric ones, is a factor which has probably supported and assisted their continued use in the structures of Chamic languages.

It is also the case that at least the Indochinese Chamic languages have borrowed (or, more precisely, it is true that the descendants of women who shifted from Mon-Khmer languages to Chamic ones have perpetuated) a very large proportion of their free grammatical morphs from Mon-Khmer languages, a proportion which, viewed crosslinguistically, is unusually high and which includes personal pronouns, semantically-blank adpositions such as *baʔ* 'at', discourse particles, and possibly some negators such as *bɛʔ* 'irrealis negator' (to say nothing of the presence of some very common Chamic verbs which are of Mon-Khmer origin). Some of these forms are included on the Blust list, which in any case was not drawn up primarily with Chamic languages in mind. The same remarks also apply, though to a smaller degree, to the presence in these form-classes of a number of elements of (at present) uncertain origin which are pan-Chamic in distribution and which therefore reconstruct back to Proto-Chamic. In fact, the only form-class in Chamic languages whose contents are purely Austronesian in origin is the system of numerals.

The Proto-Chamic (or at least pan-Chamic) material which the Blust list provides and draws upon can be shown to contain many elements inherited from PMP, a band of elements modified from their PMP prototypes, further elements which are shared with Malayic lects, and yet other elements which are derived from Mon-Khmer languages and some which are pan-Chamic in distribution but of unidentified origin. The contents of these bands rarely overlap, there are rather few cases in which some languages have adopted words for a particular items from one source while other Chamic languages have retained words for the same concepts which were to be found in an earlier historical stage of the language. It can only rarely be shown (for instance in the case of the additional Bahnaric forms which come from Hrê and which occur in Haroi but nowhere else in Chamic) that one Chamic language has taken a greater share or a bigger number of forms from a particular group of Mon-Khmer languages, and this donor group being a group which has provided forms that are reflected throughout the Chamic languages, than any other Chamic language has.

This relative discreteness of the various bands suggests that the contents of each new band of elements had largely consolidated in the period before the next wave of elements entered Proto-Chamic, and this implies that productive contact with each set of donor languages had largely ceased, before the next wave of loans or innovations came in from a different direction. Periods of borrowing, from whatever source, appear to have been succeeded by periods of consolidation of these borrowings (and of other external influences). Proto-Chamic included elements from both North and Central Bahnaric languages, and these spread into the modem languages from Proto-Chamic rather than from fortuitously coincidental borrowing of such forms from adjoining languages.

The main features of the picture are clear enough, and they fully support Thurgood's historical hypotheses in Thurgood (1999). The Chamic languages derive from a Western Malayo-Polynesian language which in origin is very similar to modem Malay, with which it shares a number of phonological and lexical innovations which set them apart from other Malayo-Polynesian languages. (Nonetheless the Chamic and Malayic branches have both subsequently gone their separate ways, and it is evident that they had already done so even at the periods of first attestation of Old Cham and Old Malay.)

Acehnese is clearly a part of the Chamic subgroup, rather than being coordinate with Chamic and Malay.[21] It is coordinate with all the other Chamic languages (which had not differentiated much before the departure of the Acehnese to Kelantan and latterly to Sumatra, and which may in fact have diffused several more innovations and loans among its dialects after Acehnese's departure). Acehnese shares with the other Chamic languages a number of basic loans from Mon-Khmer languages and additionally a number of pan-Chamic words of unknown etymology, although an examination of the contents of its basic Blust list vocabulary indicates that later Acehnese contact with Malay and (probably and to a lesser extent) with other, as yet unidentified, languages has also taken place. The striking parallels (which are caused mostly by shared retentions of Proto-Chamic forms) which are to be found within the basic lexicon of Chamic languages belie the first impressions of immense diversity among them. These first superficial differences have resulted from several different series of phonological changes which have affected particular Chamic languages (or which have sometimes affected groups of them together, we note for instance the development of word-final preploded nasals which is shared by Tsat and Northern Roglai, and which caused Graham Thurgood to link them together historically). Probably none of the Chamic languages discussed here are nowadays interintelligible, but a millennium ago this mutual unintelligibility may not have been the case.

This study also shows that the specifically Chamic affinities of Tsat are similarly historically secure, as Tsat contains elements from Mon-Khmer languages and a portion of the aforementioned lexical 'unknowns', in addition to containing a few later loans from Hlai and an especially large number of loans from Chinese languages, which are not found in other Chamic languages and which mostly do not figure in the materials on the Blust list.[22] The retention in the Tsat lexicon of a number of common Chamic forms, which are

[22] Not everyone agrees with this view, and Sidwell (2004) discusses some objections to it. In his view, which draws upon some descriptive and historical work on Moken-Moklen by Michael Larish (Larish 1999), Malayic, Moken-Moklen of the Mergui Archipelago in Burma and coastal Thailand, plus the language of the Orang Laut of eastern coastal Sumatra, plus Acehnese and Chamic, are all members of a genetic subgroup of Malayo-Polynesian that was centred on mainland Southeast Asia, and whose members were strongly influenced (at least at a lexical level) by Mon-Khmer languages and also by other substrate languages which have no known cognates and which have left no other trace, and which I call the 'submerged substrate' language(s). (Part of the evidence for this is that exceedingly few of the alleged Bahnaric loans into Chamic languages are found in West Bahnaric languages; they are much more common in Central and South Bahnaric languages. Consequently, Sidwell suggests that both South and Central Bahnaric and Chamic languages have borrowed these forms from the submerged substrate.) According to this hypothesis, Acehnese, which shares only a small proportion of the Mon-Khmer and 'submerged substrate' elements which are found in all (other) Chamic languages (including the 'submerged substrate' loans which Thurgood (1998) and others took to be loans into Chamic from Bahnaric), has subsequently acquired many features which make it appear Chamic as a result of the migration of many Chams to northern Sumatra in the Middle Ages. Additionally, Acehnese has borrowed massively and often at a very basic level from Malay in the last few centuries, a practice which would have diluted the number of Mon-Khmer loans in the language in any case. The full implications of these controversial claims have yet to be worked out.

[23] It is also feasible that the Utsat, being Muslims, have also borrowed philosophical vocabulary and other Islamically-focussed lexicon from Malay, but I do not know of any such examples in the available Tsat material.

found in Indochinese Chamic languages (and which are often though not always also evident in Acehnese[23]) and which are of Mon-Khmer origin, easily gives the lie to any vain idea that Tsat reflects the evidence of an early separate migration to Hainan which is coeval with the date of dispersal of Acehnese and the other Chamic languages, and which would suggest that Tsat is a primary and primordial subdivision of Chamic, with Acehnese on the one hand and the remaining Chamic languages on the other, comprising the two other branches. But Tsat and Acehnese do not appear to share any special lexical innovations (nor do they preserve many sole lexical retentions, for that matter) against the other Chamic languages which would permit us to unite them in a special subgroup. If they had shared some lexical retentions, then it would most probably be the case that the forms for the relevant glosses in other Chamic languages would be lexical innovations which had passed through Indochinese Chamic languages after the departure of what were to become the Acehnese and Tsat speech communities.

There is little in the way of strong or plentiful innovatory lexical evidence for any particular subgroupings within Indochinese Chamic, although Jarai and Rade seem lexically to be slightly more similar to one another than they are to the other Chamic languages, and they seem to share a few (but only a few) lexical innovations which are unknown elsewhere. This admittedly small degree of shared lexical innovation occurs despite their very different phonological histories, with the relative phonological conservatism of Jarai pitted against the extreme degree of Rade phonological innovation, something which is especially marked in Rade presyllables. But we should always remember that the Jarai and Rade speech communities neighbour one another in the southern Vietnamese highlands and the neighbouring parts of Cambodia. It is also true that they are more similar to one another in most respects of typology and in fabric (that is, in the morphemes which they possess) than they are to any other neighbouring language, so that some words may have diffused from one of the languages to the other one. For the rest, the basic vocabularies and Blust list responses of the two modern Cham (rather than Chamic) languages, and those of Chru, Haroi and Northern Roglai, especially the first two mentioned there, are very similar to one another (at least as far as what we can adduce from what we have available), so that we have to look elsewhere other than the basic lexicon in order to find differences between the languages.

The evidence of Blust-style lexicostatistics supports the picture which Thurgood gives in his book, that of what was originally the northernmost Chamic language (or the northernmost link in the Chamic dialect chain) peeling off, migrating to the south and forming the basis for modern Acehnese. Meanwhile the next most northerly language also moves out of what is now northern Vietnam and its speakers cross to Hainan and form the nucleus of the modern Tsat speech community (a community which is to be expanded with the later arrival of speakers of a more southerly Cham dialect).

We may date the split of Acehnese to the late tenth century AD, with the downfall of the northern Cham empire, and that of Tsat maybe a little later. Jarai and Rade are the next to split away, but if they share a period of unity it is a brief one, with few shared lexical innovations and not many shared phonological ones either. The remaining Cham lects then diffuse along the coast and their speakers go somewhat further into the hinterland, presumably in the course of the first half of the second millennium, while

[24] Some of these forms may once have existed in Acehnese but they may have been replaced (possibly by loans from the more prestigious Malay). Absence of evidence is not evidence of absence.

speakers of Western Cham are parted from their stay-at-home Eastern Cham fellows after 1471 and retreat to Khmer-speaking territories. Acehnese went with its speakers from Indochina to Sumatra as a Western Malayo-Polynesian language which had come under strong influence from Mon-Khmer languages, an influence which were already beginning to reshape its phonology and which had already done this to its lexicon, although its morphological system remained typically and conservatively Austronesian.

The Chamic languages which remained in Indochina gradually absorbed more and more of the areal features of general (and later on, of individual) Mon-Khmer languages, and they absorbed more and more vocabulary from this source too. Much later – mostly within the last century – all of the Indochinese Chamic languages except Western Cham, which has been as strongly influenced by Khmer as the others have been by Vietnamese, have absorbed huge amounts of vocabulary from Vietnamese (and there are even a few of these Vietnamese loans present in Mekong Delta Western Cham too, since the official language there, though not the regional majority language, is Vietnamese; Headley 1991 provides a couple of examples of these loans).

The Chamic languages are unusual among Austronesian languages inasmuch as a high proportion of the elements in the extensive non-Austronesian parts of the basic vocabularies can be etymologised. Furthermore a very high proportion of the inherited elements in the Chamic language lexica that derive from PMP can be found in the most basic strata of the vocabulary (at a rough guess, maybe almost 40% of the inherited morphs in Chamic languages which serve to make these languages lexically Austronesian can be found among the Blust list responses).

Tsat's process of change, which had progressed further from the inherited Western Malayo-Polynesian norm than Acehnese had, was interrupted at a tome when it had already shed such features as infixation (involving a feature and elements which Acehnese never lost) and had absorbed plentiful amounts of Mon-Khmer lexicon. Firstly weak influence from Hlai, then much stronger influence from the Hainanese form of Southern Min (Minnan) and finally two waves of influence from Mandarin, the second much stronger than the first, shaped and shape Tsat. With the very partial exception of Acehnese, all Chamic languages show in every way the marks of profound influences from non-Austronesian languages, but none show these more so than Tsat, where the influence, which comes especially from Chinese but which also includes earlier influence from Mon-Khmer languages, is massively clear at all levels if one knows what to look for.

And the effects of these languages on Tsat's basic vocabulary are no exception to this generalisation. At first glance Tsat appears to be anomalous among the Chamic languages because it contains a higher proportion of forms on the Blust lexicostatistical list which are unique and which are not found elsewhere in Chamic. But some of these un-Chamic forms are simply recent loans that have been acquired from Chinese languages, and the origins of others may yield themselves up to us after further investigations are carried out among the languages spoken in southern China and especially on Hainan Island.[24] Tsat's genetic relationships with other languages can still best be seen by the

[25] The impact of Tsat on the Hlai lexicon is probably not to be underestimated either, although this topic requires further investigation – and yet there is not as much information available on Hlai as one might wish for. It is almost certain that Hlai was the first tonal language with which Tsat-speakers were in contact, and that it was spoken by people who were living around and maybe among Tsat-speakers, and it is further probable that this Tsat-Hlai contact began long before any Chinese language came to be used in that part of Hainan. However, so far only a handful of

etymological examination of its basic vocabulary, since typologically it doers not look Chamic at all, having lost all its distinctive bound inflectional morphology, while in its possession of tones and other features it rather resembles the typology of other languages of Hainan (such as the Tai-Kadai languages Hlai and Ong-Be, and of course the non-Tai Minnan Chinese), whatever the genetic origins of these languages are.

We may note that Thurgood (1999) points throughout his book that there are frequent problems with demonstrating that the various reflexes in the daughter languages of Proto-Chamic forms are perfectly lautgesetzlich. Quite often there are phonological irregularities of realisation simultaneously in the initial, vowel and final phone in some Chamic language's reflex of a particular Proto-Chamic form, even though we can be all but certain (or we put faith in the hope) that the form derives from (or reconstructs back to) Proto-Chamic. And many of the problems which these almost certainly cognate but phonologically aberrant forms present have yet to be solved, just as is the case with many other issues in the internal and external history of the Chamic languages.

References

Adelaar, Karl Alexander. 1991. 'Some notes on Sri Lanka Malay.' *Papers in Western Austronesian languages,* edited by Hein Steinhauer, 23-37. Pacific Linguistics: A-81. Canberra: Research School of Pacific Studies, Australian National University

-----1992. *Proto-Malayic: the reconstruction of its phonology and parts of its lexicon and morphology.* Pacific Linguistics C-119. Canberra: Research School of Pacific Studies, Australian National University.

-----1995. 'Borneo as a crossroads for comparative Austronesian linguistics.' *The Austronesians: historical and comparative perspectives*, edited by Peter Bellwoof, James J. Fox and Darrell T. Tryon, 75-97. Canberra: Australian National University.

-----1996. 'Malay in the Cocos (Keeling) Islands.' *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen,* edited by Bernd Nothofer, 167-198. Hamburg: Abera.

-----2001. 'Malayic, Chamic and Bali-Sasak-Sumbawa: the demise of Malayo-Javanic.' Paper presented to the Fifth International Symposium on Malay/Indonesian Linguistics.

Alieva, Natalia F. 1984. 'A language union in Indo-China.' *Asian and African Studies* [Bratislava] XX: 11-21.

Bakker, Peter, and Maarten Mous (eds.). 1994. *Mixed languages: 15 case studies in language intertwining*. Amsterdam: IFOTT.

Benedict, Paul K. 1941. 'A Cham colony on the island of Hainan.' *Harvard Journal of Asiatic Studies* 4: 129-134.

Bennett, Patrick R. 1998. *Comparative Semitic linguistics: a manual.* Winona Lake, Illinois: Eisenbrauns

Blust, Robert, Russell D. Gray and Simon Greenhill. 2005. *Austronesian Basic Vocabulary Database*. Department of Psychology, University of Auckland. Accessible at: http://language.psy.auckland.ac.nz/index.php

loans have been identified as coming from Hlai into Tsat, and a few words have gone the other way, notably a form for 'six' which derives from Chamic **nam*, which presumably replaced an earlier Hlai numeral (Graham Thurgood, personal communication, April 2003).

Blust, Robert A. 1981a. 'Variation in the retention rate in Austronesian languages.' Paper presented at the Fifth International Conference on Austronesian Languages, Denpasar, Bali.

-----1981b. ' The reconstruction of Malayo-Javanic: an appreciation.' *Bijdragen tot de Taal-, Land- en Volkenkunde* 137: 456-469.

-----1990a. 'Malay historical linguistics: a progress report.' *Rekonstruksi dan cabang cabang bahasa Melayu induk*, edited by Moh[amme]d Thain Ahmad and Zaid Mohamed Zaidi, 1-33. Kuala Lumpur: Dewan Bahasa dan Pustaka.

-----1992. 'The Austronesian settlement of mainland Southeast Asia.' Karen L. Adams and Thomas John Hudak (eds.), *Proceedings of the Second Annual Meeting of the South East Asian Linguistics Society*, 25-83. Tempe: Arizona State University Press.

-----1993a. 'Central and Central-Eastern Malayo-Polynesian.' *Oceanic Linguistics* 32: 243-292.

-----2000a. Review of Thurgood (1999). *Oceanic Linguistics* 39: 435-445.

-----2000b. 'Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages.' *Time depth in historical linguistics*, edited by Colin Renfrew, April McMahon and Larry Trask, 311-331. Cambridge: McDonald Institute for Archaeological Research.

Bochet, Gilbert, and Jacques Dournes. 1953. *Lexique polyglotte: Vietnamien, koho, roglai, français.* Saigon: Editions France-Asie.

Bui Khanh The. 1995. *Tu-Dien Cham-Viet.* Ho Chi Minh City: Nha Xuat Ban Khoa Hoc Xa Hoi.

Collins, [Ira] Vaughn. 1969. 'The position of Atjehnese among Southeast Asian languages.' *Mon Khmer Studies* 3: 48-60.

Daud, Bukhari, and Mark Durie. 1999. *Kamus Basa Acèh - Kamus Bahasa Aceh – Acehnese - Indonesian - English Thesaurus.* Pacific Linguistics: C-151. Canberra: Australian National University.

Dempwolff, Otto. 1934-1938. *Vergleichende Lautlehre des austronesischen Wortschatzes.* Hamburg: Beiheft zur Zeitschrift für Eingeborenensprachen.

Durie, Mark. 1990. 'Proto-Chamic and Acehnese mid-vowels: Towards Proto-Aceh-Chamic.' *Bulletin of the School of Oriental and African Studies* 53: 100-114.

Dyen, Isidore. 1962. 'The lexicostatistical classification of the Malayopolynesian languages.' *Language* 38: 38-46.

-----1965. *A lexicostatistical classification of the Austronesian languages.* International Journal of American Linguistics Memoir 19. Baltimore.

-----1971. 'The Chamic languages.' *Current Trends in Linguistics, volume 8: Oceania*, edited by Thomas A. Sebeok, 110-120.

Egerod, Søren. 1978. 'An English-Rade Vocabulary.' *Bulletin of the Museum of Far Eastern Antiquities* [Stockholm] 50: 47-108.

Fuller, Eugene. 1977. 'Chru phonemics'. Thomas et al (ed., 1977): 77-86.

Grace, George W. 1966. 'Austronesian lexicostatistical classification: a review essay.' *Oceanic Linguistics* 5: 13-31.

Greenberg, Joseph H. 1987. *Language in the Americas.* Stanford: Stanford University Press.

Headley, Robert A. 1976. 'Some sources of Chamic vocabulary.' *Austroasiatic studies*, edited by Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta, 453-476. Honolulu: University Press of Hawai'i.

-----1991. 'The phonology of Kompong Thom Cham.' *Austroasiatic languages: essays in honour of H. L. Shorto*, edited by Jeremy H. C. S. Davidson, 105-121. London: School of Oriental and African Studies.

Hooley, Bruce A. 1971. 'The Austronesian languages of Morobe District, Papua New Guinea.' *Oceanic Linguistics* 10: 79-151

Hudson, Alfred B. 1967. *The Barito dialects of Borneo, a classification based on comparative reconstruction and lexicostatistics*. Ithaca, New York: Cornell University, Department of Asian Studies.

Jacob, Judith M. 1963. 'Prefixation and infixation in Old Mon, Old Khmer and Modern Khmer.' *Linguistic comparison in South East Asia and the Pacific*, edited by H. L. Shorto, 62-70. London: Luzac.

Lafont, Pierre-Bernard. 1968. *Lexique jarai parler de la province de plei ku.* Paris: Publications de l'Ecole Française de l'Extrème-Orient LXIII.

Larish, Michael David. 1999. *The position of Moken-Moklen within the Austronesian language family*. Unpublished doctoral dissertation, University of Hawai'i at Manoa.

Lee, Ernest Wilson. 1998. 'The contribution of Cat Gia Roglai to Chamic.' Thomas (ed., 1998), 31-54.

Matisoff, James A. 2001 'Genetic versus Contact Relationship: Prosodic Diffusability in Southeast Asian Languages', in Alexandra. Y. Aikhenvald and R.M.W. Dixon (eds.), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, Oxford University Press, Oxford, 291-327.

Miller, Wick R. 1967. Uto-Aztecan cognate sets. *University of California Publications in Linguistics 46.* Berkeley and Los Angeles: University of California Press.

-----1984. 'The classification of the Uto-Aztecan languages based on lexical evidence.' *International Journal of American Linguistics* 50: 1-24.

-----, James L. Tanner and Lawrence P. Foley. 1971. 'A lexicostatistical study of Shoshoni dialects.' *Anthropological Linguistics* 13: 142-164.

Pang, Kang-Feng. 1998. 'On the Ethnonym 'Utsat.' Thomas (ed.), 55-60.

Ringe, Don, Tandy Warnow and Ann Taylor. 2002. 'Indo-European and computational cladistics.' *Transactions of the Philological Society* 100: 59-127.

Shintani, Tadahiko L. A. 1981. *Boh blu Êdê-Yuan-Za pô nê = Tù v.ung Êdê-Vi.êt-Nhât.* Tokyo: Viên Nghiên Cúu Ngôn Ngụ Và Van Hóa á Phi. [Japanese, Rade and Vietnamese dictionary.]

Sidwell, Paul. 2004. 'On the origins of the Proto-Chamic lexicon and significance of Acehnese.' To appear in *Mon-Khmer Studies.*

Swadesh, Morris. 1950. 'Salish Internal Relationships.' *International Journal of American Linguistics* 17: 257-267.

-----1955. 'Toward Greater Accuracy in Lexico-Statistic Dating.' *International Journal of American Linguistics* 21: 121-137.

Tegenhardt-Mundhenk, Alice, and Hella Goschnick 1977. 'Haroi phonemes.' Thomas et al. (ed., 1977): 1-15.

Tharp, James and Y-Bhăm Buôn Yă. 1980. *A Rhade-English dictionary with English-Rhade finderlist.* Pacific Linguistics C-58. Canberra: Australian National University.

Thomas, David D. (ed.) 1998. *Studies in Southeast Asian languages no. 15: Further Chamic studies*. Pacific Linguistics A-89. Canberra: Research School of Pacific Studies, Australian National University.

-----, Ernest W. Lee and Nguyen Đang Liem (eds.). 1977. *Papers in Southeast Asian languages, no. 4. Chamic studies*. Pacific Linguistics A-48. Canberra: Research School of Pacific Studies, Australian National University.

Thurgood, Graham.1996. 'Language contact and the directionality of internal drift: the development of tones and registers in Chamic.' *Language* 71: 1-31.

-----1999. *From ancient Cham to modern dialects: two thousand years of change*. Oceanic Linguistics Special Publication 28. Honolulu: University Press of Hawai'i.

-----To appear a. 'A preliminary sketch of Phan Rang Cham.' 25 pp., to appear in *The Austronesian Languages of Asia and Madagascar*, edited by Alexander Adelaar and Nikolaus Himmelmann. London: Curzon.

-----To appear b. 'Learnability and direction of convergence in Cham: the effects of longterm contact on linguistic structures.' Manuscript, 21 pp.

-----To appear c. 'Crawfurd's 1822 Malay of Champa'. 12 pp., to appear in a Festschrift for P J Mistry.

-----, and Fengxiang Li. 2003. 'Contact-induced variation and syntactic change in the Tsat of Hainan.' David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds., *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 185-200. Canberra: Pacific Linguistics 555.

Tín, Pham Xuan. 1955. *Đạ ngữ tiêu tự diển: Lexique polyglotte.* Dalat [Vietnam]: Langbian. [Vietnamese, Jarai, Chru and French glossary.]

Zheng Yiqing. 1997. *Huihuihua yanjiu.* Shanghai: Shanghai Yuandong Chuban She.

TABLE 1: *A norm-referenced lexicostatistical grid for PMP, Malay and Chamic languages (including Acehnese), with comments.*

(The languages surveyed are: Proto-Malayo-Polynesian, Standard Malay, Acehnese, Phan Rang Cham, Western Cham of the Mekong Delta in Vietnam, Haroi, Chru, Jarai, Rade, Northern Roglai, and Tsat). Comments are provided for some entries.

No.	Gloss	PMP	Mal	Ach	PRC	WeC	HA	CR	JA	RA	RO	TS	Comments
1	Hand	A	A	A, B	A	A	A	A	A	A	A	A	
2	Left	A	A	B	C	C	C	C	C	C	C	C	C<MK
3	Right	A	A	B	C	C	C	C	C	C	C	D	D<UNK
4	Foot	A	A	A	A	A	A	A	A	A	A	A	
5	To walk	A	B	C	A	A	0	A	A	A	A	A	
6	Road	A	A	A+	A	A	A	A	A	A	A	A	A+<MAL
7	To come	A	B	C	A	A	A	A	A	A	A	A	
8	To turn	A	B	C	C	C	C	C	C	D	C	C	C<UNK
9	To swim	A	B	A	C	C	C	C	C	C	C	C, D	C<MK (often compounded with 122), D<UNK
10	Dirty	A	B	B+	C	C, D	D	C	D	E	C	C	B+<MAL
11	Dust	A	B	C	C	C	0	C	D	C, D	C	D	C<SKT
12	Skin	A	A	A	A	A	A	A	A	A	A	A	
13	Back	A	A	B	C	C	C	C	C	C	C	C	C<MK
14	Belly	A	B	B	A	A	A	A	A	A	A	A	B<MK?= Chamic 'guts'

15	Bone	A	A	A	A	A	A	A	A	A	A	A	
16	Guts	A	B	B	B	B	B	B	B	B	B	B	
17	Liver	A	A	A	A	A	A	A	A	A	A	A	
18	Breast	A	A	A	A	A	A	A	A	A	A	A	
19	Shoulder	A	B	A	A	A	A	A	A	A	A	A	The A form occurs with a different sense in Malay
20	To know	A	A	A	A	A	A	A	A	A	A	A	
21	To think	A	B	B	C	C	C	C	D	C	C, E	C?	B<AR
22	To fear	A	A	A	B	B	B	B	B	B	B	C?	
23	Blood	A	A	A	A	A	A	A	A	A	A	A	
24	Head	A	B	C	C	C	C	C	C	C	C	C	B<SKT, C<MK
25	Neck	A	A	B	B	B	B	B	B	B	B	B	B<MK?
26	Hair	A	B	A	A	A	A	A	A	A	A	A	B= 'root' in PMP
27	Nose	A	A	A	A	A	A	A	A	A	A	A	
28	Breath	A	A	A	A	A	A	A	A	A	A	A	
29	To smell	A	B	B	B	B	B	B	B	B	B	B	B<MK
30	Mouth	A	B	A	A	A	A	A	A	A	A	A	
31	Tooth	A	B	B	B	B	B	B	B	B	B	B	
32	Tongue	A	A~	A	A	A	A	A	A	A	A	A	
33	To laugh	A	A	B	C	C	C	C	C	C	C, D	C	C<UNK
34	To weep	A	A	B	C	C	C	C	C	C	C	C	C<MK?
35	To vomit	A	A	A	A, B	B	A	B	A, B	A, B	B	A	B<MK
36	To spit	A	B	A	C	D	C	C	C, D	C, D	C	C, E	C, D<MK
37	To eat	A	A~	A~	A~	A~	A~	A~	A~	A~	A~	A~	
38	To cook	A	A	A, B	A	A	A	A	A	A	A	A, B?	
39	To chew	A	A	A	A	A	A	A	A	A	A	A	
40	To drink	A	A~	A~	A~	A~	A~	A~	A~	A~	A~	A~	
41	To bite	A	B	C	C	C	C	C	C	C	C	D?	C<MK
42	To suck	A	B	C	D	D	D	D	D	D	D	C	D<MK
43	Ear	A	A	A	A	A	A	A	A	A	A/B	A	
44	To hear	A	A	A	B	B	B	B	B	B	B	B	B<UNK
45	Eye	A	A	A	A	A	A	A	A	B	A	A	B= 'eggyolk' in other Chamic lgs
46	To see	A	B	C	D	D	D	D	D	D	D	D	D<MK
47	To yawn	A	A	B	C	C	B	B	B	B	B	C	B<MK, C<UNK
48	Sleep	A	A	B	B	B	B	B	B	B	B	B	B<MK?
49	To lie down	A	B	B	C, D	C, D	0	D	C	C, D	C	D	
50	To dream	A	A~	A~	A~	A~	A~	A~	A~	A~	A~	A~	
51	To sit	A	B	B	B	B	B	B	B	B	B	B	
52	To stand	A	B	C	C	C	C	C	C	C	C	C	C<MK
53	Person	A	B	B	B	B	B	B	B	B	B	B	
54	Male	A	A	A	A	A	A	A	A	A	A	A	
55	Female	A	B	C	A, D	A, D	A, D	A, D	A, D	A	A, D	A, D	D<UNK
56	Child	A	A	A	A	A	A	A	A	A	A	A	
57	Husband	A	A	B	C	D	C	D	C	C	E	A?	C<MK, E= 'master of house'
58	Wife	A	B	A	C	C	C	C	C	C	C	C	B<SKT
59	Father	A	B	C	B	B	B	B	B	B	B	B	
60	Mother	A	B	A	A	A	A	A	A	A	B	A	
61	House	A	A	A	B	B	B	B	B	B	B	B	B in ACH = tent
62	Thatch	A	A	A	B	B	B	B	B	B	B	B	B<UNK
63	Name	A	B	A	A	A	A	A	A	A	A	A	B<SKT
64	To say	A	B	C	D	D	D	D	D	D	D	D	B<SKT, D<MK
65	Rope	A	A	A	A	A	A	A	A	A	A	A	
66	To tie	A	A	A	A	A	A	A	A	A	A	A	
67	To sew	A	A	B	A	A	A	A	A	A	A, C	A	
68	Needle	A	A	A	A	A	A	A	A	A	A	A	
69	To hunt	A	B	B	C	D	0	D	C	D	E	F	C<UNK
70	To shoot	A	A	A	A	A	A	A	A	A	A	A	
71	To stab	A	B	C	D	D	E	E	F	E	E, G	E	F<BAH
72	To hit	A	B	C	C	C	C	C	C	C	C	C	C<MK
73	To steal	A	B	C	C	C	0	C	C	C	C	C	B<SKT, C <MK

74	To kill	A	A	B	C	C	C	D	E	D	C	D	D= 'CAUS + die', E= 'CAUS + ?'
75	To die	A	A	A	A	A	A	A	A	A	A	A	
76	(To) live	A	A~	A	A	A	A	A	A	A	A	A	
77	To scratch	A	B	C	D, E	E	E	F	D	D, F	F	F	D<MK?, F<MK
78	To cut	A	A	A	A	A	A	A	C	A	A	A, B	B, C <UNK
79	Wood	A	A	A	A	A	A	A	A	A	A	A	
80	To split	A	A	A	A	A	A	A	A	A	A	A	
81	Sharp	A	A	A	B	B	0	B	B	C, D	B	E	B<MK?
82	Dull	A	A	A	B	B	0	B	B	B	B	B	
83	To work	A	B	B, C	D	D	D	D	D	D	C	C	
84	To plant	A	A	A~	A	A	A	A	A	A	0	A	
85	To choose	A	A	A	B	B	B	B	B	B	B	B, C	B<MK, C<UNK
86	To grow	A	A	A	A	A	A	A	B	C	A	A	
87	To swell	A	A	A	B	A, B	A, B	A	B	B	A, B	C	B<MK
88	To squeeze	A	A	B	A	B	A	B	A, B	B	B	B	B<MK
89	To hold	A	B	C	D	D	E	D	D, E	E	D, E	D	D<UNK?, E<MK
90	To dig	A	A	B	A	A	A	A	A	A	A	B?	
91	To buy	A	A	A	A	A	A	A	A	A	A	A	
92	To open	A	A	B	C	C	C	C	D	D	D	E	
93	To pound	A	B	C	C	C	C	C	D	E	C	A	C<MK; D, E < UNK
94	To throw	A	B	C	D	E	F	F	D	G	D	H?	
95	To fall	A	B	C	A	A	A	A	A	A	A	A	
96	Dog	A	B	A	A	A	A	A	A	A	A	A	
97	Bird	A	B	C	C	C	C	C	C	C	C	C	C<MK
98	Egg	A	A	A	A	A	A	A	A	A	A	A	
99	Feather	A	A	A	A	A	A	A	A	A	A	A	
100	Wing	A	A	A	A	A	0	A	B	A	C	=1	
101	To fly	A	B	C, D	D	D	D	D	D	D	D	D	D<MK
102	Rat	A	A	A	A	A	A	A	A	A	A	A	
103	Meat	A	A	A	B	B	A	B	A, B	A, B	B	A	B<MK
104	Fat	A	A	A	A	A	A	A	A	A	A	A	
105	Tail	A	A	A	A	A	A	A	A	A	A	A	
106	Snake	A	A	A	A	A	A	A	A	A	A	A	
107	Worm	A	A	A	A	A	A	A	A	A	A	A	
108	Louse	A	A	A	A	A	A	A	A	A	A	A	
109	Mosquito	A	A	A	A	A	0	A	B	C	A, D	A	
110	Spider	A	A~	B	C	D	E	E	E	E	F	A	D<Khmer, E<MK, F<BAH
111	Fish	A	A	A+	A	A	A	A	A	A	A	A	A+<MAL?
112	Rotten	A	A	A	A	A	A	A	A	A	A	A	
113	Branch	A	A	A	A	A	A	A	A	A	A	*	* compound using item 79
114	Leaf	A	A	B	C	C	C	C	C	C	C	C	C<MK
115	Root	A	A	B	B	B	B	B	B	B	B	B	
116	Flower	A	A	A	A	A	A	A	A	A	A	A	
117	Fruit	A	A	A	A	A	A	A	A	A	A	A	
118	Grass	A	B	B	C	C	B, C	B, C	B, C	B, C	B, C	B	C<MK?
119	Earth	A	A	A	B	B	B	B	B	B	B	B	B<UNK
120	Stone	A	A	A	A	A	A	A	A	A	A	A	
121	Sand	A	B	B	C	C	C	C	C	C	C	C	C<MK
122	Water	A	A	B?	B	B	B	B	B	B	B	B	B<PMP?
123	To flow	A	B	C	C	C	C	C	C	C	C	C	C<MK
124	Sea	A	B	B+	A	A	A	A	A	A	A	A	B+<MAL (= PMP 'sea-ward'
125	Salt	A	A	A	A	A	A	A	A	A	A	A	
126	Lake	A	A	A	A	A	A	A	A	A	A	*	*=compound from PMP elements
127	Forest	A	B	B	B, C	C	C	C	C	C	C	C	C<UNK
128	Sky	A	A	A	A	A	A	A	A	A	A	A	
129	Moon	A	A	A	A	A	A	A	A	A	A	A	
130	Star	A	B	A	A	A	A	A	A	A	A	A	
131	Cloud	A	B	B	C	C	C	C	C	C	C	C	B<UNK, C<MK

132	Fog	A	A	B, C	C, D	C, D	0	C, D	C, D	C, D	C, D	E?	C<MK<SKT, D<MK
133	Rain	A	A	A	A	A	A	A	A	A	A	A	
134	Thunder	A	B	C	D	D	D	D	D	D	D	D	B, C, D all similar, possibly cognates
135	Lightning	A	A	A	B	B	0	B	B	B?	B	C	
136	Wind	A	A	A	A	A	A	A	A	A	A	A	
137	To blow	A	A	A	A	A	A	A	A	A	A	A, B	B<UNK
138	Hot	A	A	B	C	C	0	C	C	C	D	C	
139	Cold	A	B	C	C	C	C	C	C	C	C	C	C<MK
140	Dry	A	A	B	B	B	C, D	B	B, C	D	B	B	B<UNK, C<UNK
141	Wet	A	A	A	A	A	A	A	A	A	A	A	
142	Heavy	A	A	A, B	C?	C	C	C	C	C	C	C	
143	Fire	A	A	A	A	A	A	A	A	A	A	A	
144	To burn something	A	B	C	D	D	0	D	D	C	B	C	
145	Smoke	A	B	B	B	B	B	B	B	B	B	B	
146	Ash	A	A	A	A	A	A	A	A	A	A	A	
147	Black	A	A	A	A, B	A, B	B	B	B	B	B	A	B<MK
148	White	A	A	A	A	A	A	A	A	B	B	A	B<UNK
149	Red	A	A	A	A	A	A	A	A	A	A	A	
150	Yellow	A	B	A	A	A	A	A	A	A	A	A	B<BTK
151	Green	A	B	B	B	B	C	C	C	C	C	C	A and C both <PMP
152	Small	A	B	C	B?	B?	D	D	D?	D	E	F	
153	Big	A	B	A, C	A, C	A, C	A, C	C	C	C	C	C	
154	Short	A	B	B	C	C	0	C	D	=152	B	B	
155	Long	A	A	A	A	A	A	A	A	A	A	A	
156	Thin	A	A	A	A	A	A	A	A	A	A	A	
157	Thick	A	A	A?	A	A	A	A	A	A	A	A	
158	Narrow	A	B	C	D	D	D	D	D	D	D	C	D<MK?
159	Wide	A	A	B	C	C?	0	C	C	C	D	C	D<MK
160	Sick	A	A	A	A	A	A	A	A	A	A	A	
161	Shy	A	A	A	A	A	A	A	A	B	A	B	B<UNK, it means 'fear' too in Rade
162	Old	A	A	A	A	A	A	A	A	A	A	A	
163	New	A	A	A	A	A	A	A	A	A	A	A	
164	Good	A	B	C	C	C	C	C	C	C	C	D	C<MK?
165	Bad	A	A	A	A	A	A	A	A	A	A	A	
166	True	A	A	A	B, C	B, C	B, C	B, C	C	C	C	D?	C<MK
167	Night	A	B	B	B	B	B	B	B	B	B	B	B=PMP 'evening'
168	Day	A	A	A	A	A	A	A	A	A	A	A	Form A in Malayic and Chamic is irregular in shape
169	Year	A	A	A	A	A	A	A	A	A	A	A	
170	When?	A	B	C, D	E	E	0	0	C	E	F	E?	
171	To hide	A	A	B, C	D	D	D	D	D	D	D	D	
172	To climb	A	A	A	A	A	A	A	A	A	A	B?	
173	At	A	A	A	A	A	A	A	A	A	A	A	
174	In	A	A	A	A	A	A	A	A	A	A	A	
175	Above	A	A	A	B	B	0	B	C	D	E	D	E<UNK
176	Below	A	A	B	B	B	B	B	B	B	B	B	B<UNK
177	This	A	A	A	A	A	A	A	A	A	A	A	
178	That	A	A~	A~	A~	A~	A~	A~	A~	A~	A~	A~	
179	Near	A	B	C	D	D	D	D	D	D	D	D	D<MK
180	Far	A	A~	B	B	B	B	B	B	B	B	B	
181	Where?	A	B	C	B	B	0	B	D	E	F	F	
182	I	A	A, B	A	A	A	A	A	A	A	A	A	B<SKT 'slave'
183	Thou	A	B	C	D	D	D	D	D	A, D	D	D	
184	S/he	A	B	C	A~	A~	A~	A~	A~	A~	A~	A~	
185	We	A	A	A	A	A	A	A	A	A	A	A	
186	You	A	A, B	=183	=183	=183	=183	=183	=183	=183	=183	=183	
187	They	A	B	C	=184	=184	=184	=184	=184	=184	D	=184	

188	What?	A	A	A	A	B	0	B	C	B	D	A~	
189	Who	A	A	A	A, B	A, B	B	B	B	B	B	B	B<UNK
190	Other	A	A	A, B	A, B	A, B	A	A, B	A, B	B	A	B	B<MK
191	And	A	B	B	C	C	0	C	C	C	C	B	C<MK
192	All	A	B	B	B	B	B	B	B	B	C	B	
193	If	A	B	B	C	D	0	D	E	F	E	D	B<SKT, E<CHI
194	How?	A	B	C	D	E	C	C	C	C	D	D	B< com-pound: TAM+PMP
195	No	A	B	C	C	C	A	A	C	C	C	C	C<UNK
196	To count	A	B	A	A	A	A	A	A	A	A	A	
197	1	A	A~	A	A	A	A	A	A	A	A	A	
198	2	A	A	A	A	A	A	A	A	A	A	A	
199	3	A	B	A	A	A	A	A	A	A	A	A	B<SKT
200	4	A	A	A	A	A	A	A	A	A	A	A	
201	5	A	A	A	A	A	A	A	A	A	A	A	
202	To sing	A	A	B	C	C	C	C	C	C	C	D	C<MK, D<CHI
203	To play	A	A	A	A	A	A	A	A	A	A	A	
204	We incl	A	A	A	A	A	A	A	A	A	A	A	

LEGEND: ACH (= Acehnese), AR(abic), BAH(nar), BTK (= Batak), CAUS(ative), CHI (Min Chinese), MA(lay), MK (Mon-Khmer, usually North or Centtral Bahnaric), SKT (Sanskrit), TAM(il), UNK(nown as to origin but usually reconstructible to an immediate proto-language such as Proto-Chamic). The use of the symbol ~ indicates that the language uses a morphologically aberrant development of a form which is nonetheless cognate with the PMP form. The use of + (in the Acehnese column) indicates that the form is related to the form whose letter it bears, but that it is actually a loan of this form from Malay, rather than being an inherited element. The sign 0 indicates that an equivalent for this gloss and in this language was not available to me. The cognacy of those items which are marked with a letter followed by ? with other items that are marked out with the same letter is indicated as yet being uncertain.

Table 2a. *Dyen's lexicostatistical percentages for selected Indochinese Chamic languages, using the Swadesh 200-item list and horizontal lexicostatistical techniques (Dyen 1971: 111).*

Cham				
73.0	Chru			
68.0	73.0	Roglai		
66.0	71.5	66.5	Jarai	
60.0	68.5	64.5	83.5	Rade

Table 2b. *Lexicostatistal percentages for certain Chamic languages using the Swadesh 200-item list and horizontal lexicostatistics (Thomas 1977: viii).*

Western Cham							
82	Eastern Cham						
75	76	Chru					
77	77	77	Southern Roglai				
71	71	72	71	Northern Roglai			
64	67	69	65	67	Haroi		
62	62	64	60	64	73	Jarai	
61	61	63	59	61	66	72	Rade

Table 3. *Selected morphological properties of Chamic and certain other relevant languages.*

Feature	**Tagalog**	**Proto-Malayo-Chamic**	**Bahasa Melayu**	**Acehnese**	**Old/Inscriptional Cham**	**Written Cham**	**Phan Rang Cham**	**Tsat**	**Modern Chinese**	**Modern Khmer**
Bound inflection	Yes	No?	No/yes	No	No	No	No	No	No	No
Prefixes	Yes	No?	No/yes	No	No	No	No	No	No	No
Infixes	Yes	No?	No	No	No	No	No	No	No	No
Suffixes	Yes	No?	No	No	No	No	No	No	No	No
Bound derivationals	Yes	Yes	Yes	Yes	Yes	Yes	Hardly	No	Emerging	Yes
Prefixes	Yes	Yes	Yes	Yes	Yes	Yes	Not productive	No	No?	Yes
Infixes	Yes	Yes	No	Yes	Yes	Yes	Not productive	No	No	Yes, non-productive?
Suffixes	Yes	No	Yes, but few	No	No	No	No	No	Emerging ?	No
Lexical tones	None	none	none	None	None	none	two	five	Six in Hainanese	none

The Proto-Malayo-Chamic language has not been reconstructed in detail and no descriptions of how it may have looked exist in the linguistic literature. The presence of certain kinds of morphological features in this language is inferred from the evidence of retentions of actual inherited morphemic forms (which are what I call 'fabric') in our records of Old Malay, Old Cham, modern Chamic languages, and in modern Malay and Acehnese. Prefixes and especially infixes were used more productively in Old and Middle Khmer than they are in Modern Khmer, which uses more free grammatical morphemes, though suffixes have never been used in Khmer (this issue is discussed further in Jacob 1963).

4 *Word structure in Chamic: prosodic alignment versus segmental faithfulness*

Peter Norquest

0. Introduction[1]

Chamic is an Austronesian sub-group which was originally spoken on on the Eastern coast of the Southeast Asian peninsula in what is modem Vietnam. A shift from light, disyllabic feet to heavy, monosyllabic feet occurred in Chamic diachronically between Proto-Malayo-Chamic and Proto-Chamic, under the influence of Mon-Khmer languages which surrounded Proto-Chamic (Thurgood (1999)).

This paper argues that the changes which occurred following this stress-shift were set in motion by the phonetic lengthening of stressed syllables, and that they continued a trend which began at the Proto-Malayo-Chamic level where less-salient segments were sacrificed in favor of aligning the edges of prosodic categories. It is also argued that the sesquisyllabic forms of Proto-Chamic were not truly iambic, but were suboptimal heavy trochees with left-edge appendages maintained through faithfulness to segments bearing place features. It was the maintenance of these segments which prevented the prosodic drive to align the edges of the foot and the prosodic word, producing tension between these two different areas of the phonology.

Three stages of change will be examined in this paper: (1) the period between Proto-Malayo-Polynesian and Proto-Malayo-Chamic, (2) the period between Proto-Malayo-Chamic and Proto-Chamic, and (3) the period between Proto-Chamic and a selected group of its daughter languages. An Optimality Theoretic (OT: Prince & Smolensky 1993) analysis will be offered for each stage of change.

I attempt to show that the changes in speakers' grammars between each stage are due to reanalyses of language structure which result from changes in output at the phonetic level. The resulting changes in prosodic structure may be modeled through the interaction between prosodic alignment constraints, faithfulness constraints, and structural markedness constraints. In addition, a new alignment constraint crucial to the present analysis, ALIGNSYLL(ABLE), is posited which is modeled on the traditional alignment constraints ALLFEETLEFT and ALLFEETRIGHT (Prince & Smolensky 1993).

[1] I would like to thank several people who at different times have discussed different aspects of this analysis with me and contributed to its development: Diana Archangeli, Dick Demers, Mike Hammond, Bob Kennedy, Diane Ohala, Joe Pittayaporn, Paul Sidwell, Graham Thurgood, and Adam Ussishkin. Thanks also to Marcel den Dikken for sharing with me his unpublished manuscript on the Rotuman Noun Phrase. All data on Chamic is taken from Thurgood (1999); I have incorporated specific changes to the Proto-Chamic reconstructed therein based on Blust (2000), specifically *-əy*, *-əw* and *tl-* for *-ɛy*, *-ɔw*, and *kl-*, respectively. Any mistakes are mine.

Norquest, P. "Word structure in Chamic: prosodic alignment versus segmental faithfulness". In Grant, A. and Sidwell, P. editors, *Chamic and Beyond: Studies in mainland Austronesian languages.*
PL-569:147-188. Pacific Linguistics, The Australian National University, 2005. DOI:*10.15144/PL-569.147*

In order to strengthen the points made in this paper and further test the proposed analysis, two other typologically similar cases of this kind of change are also examined. The first example is a pair of Oceanic languages, Rotuman and Kwara'ae, where a stress-related shift from disyllabic feet to monosyllabic feet has occurred in the informal register of discourse, while forms with original, disyllabic feet are still preserved in careful citation. The second example is Hlai, a subgroup of Kra-Dai (Tai-Kadai), where comparison with Austronesian cognates reveals that as a result of stress-shift, the optimal foot became a heavy syllable, resulting in changes quite similar to those which have occurred in Chamic and its daughters.

1. From Proto-Malayo-Polynesian to Proto-Malayo-Chamic

This section examines the larger context of prosodic change which has been ongoing since Proto-Malayo-Polynesian (PMP) and the drive toward limiting phonological words to disyllabic trochees, which had been largely achieved by the time of Proto-Malayo-Chamic (PMC), the immediate ancestor of Proto-Malayic (PM) and Proto-Chamic (PC). Examples are provided below of exactly how these changes occurred, and it is argued that there was a tension between prosodic requirements faithfulness to segmental material. It is asserted that there is a hierarchy in the segment inventory dependent on the saliency of the segments in question, and that this hierarchy is crucial in understanding the changes between one stage of the language and the next. An OT analysis is offered in the second half of this section, which lays the foundation for the formal analyses offered in subsequent sections of this paper.

1.1 *Reduction to disyllabic trochees*

One of the most salient changes between PMP and PMC is the reduction of the prosodic word. While the most common type of word in PMP (the ancestor of all non-Formosan Austronesian languages) was a disyllable, it also contained a number of trisyllabic and quadrisyllabic forms. The majority of these forms underwent reduction to a disyllabic template by the time of PMC, which can be understood formally as a drive towards aligning the edges of the foot with the edges of the prosodic word. The ways in which this happened seem to have been largely predictable, as will be shown below.

There were two changes which occurred between PMP and PMC which are relevant to the present discussion. The first was the change *h* → *Ø*, and the second was the subsequent change of *q* → *h*, which reintroduced *h* into the segmental inventory following the loss of original *h*. Under specific conditions to be discussed below, the change *h* → *Ø* then reoccurred.

1.1.1 PMP h → PMC Ø

Figure (1) illustrates the general loss of PMP *h* in all positions of disyllabic words: initial position (1a), medial position (1b), and final position (1c). Two exceptions where *h* is retained, and two exceptions in medial position where *h* has been reanalyzed as *ʔ*, are given in (1d):

(1)	Gloss	PMP[2]	PMC
(a)	the wind	haŋin	áŋin
	fire	hapuy	ápuy
	to tie	hikət	îkət
	snake	hulaR	úlər
(b)	water (fresh)	wahiR	áir[3]
	count	ihap	îap
	do; work	buhat	búat
	tree; wood	kahiw	káyu
(c)	chest	dahdah	dáda
	claw; fingernail	kuhkuh	kúku
	rope; string	talih	táli
	sugarcane	təbuh	tə́bu
(d)	green; blue	hijaw	hîjaw
	stench	bahu	báhu
	head hair	buhuk	búʔuk
	knee	tuhud	túʔut

This deletion of *h* is regular and fairly uninteresting in and of itself. However, when forms longer than two syllables are examined, an additional pattern emerges:

(2)	Gloss	PMP	PMC
	younger sibling	huaji-q	ádi(k)
	breath, soul, air	nihawa	ɲáwa
	to winnow	tahəpi	támpi
	pestle	qahəlu	hálu
	drunk	ma-buhək	mábuk
	after; behind	(ma-)udəhi	húdi[4]
	woman	b-in-ahi	bînay

The deletion of *h* in the examples in (2) is always accompanied by the additional deletion (always *ə* in a two-vowel sequence) or reanalysis of a neighboring vowel, with the result being a disyllabic form.

[2] I do not assign stress to PMP forms as there is not yet a full general concensus about how it is to be reconstructed. I believe it is safe to assume that by PMC there was a fixed trochaic pattern in place, which I mark here; however, even this is potentially controversial (but not in any way which should seriously affect the present analysis). For discussion related to this problem, see Pittayaporn (this volume).

[3] The PC form for 'water' (*íar*) (is problematic (as noted by Thurgood); Blust (2000: 441) states that it 'cannot be associated' with the PMP form, but I consider vocalic metathesis in PC at least a possibility.

[4] The *h* appearing at the beginning of this word is in a non-etymological location, and may have resulted from either irregular metathesis or epenthesis.

1.1.2 PMP q → PMC h (→ Ø)

The phenomenon in section 1.1.1 can be shown to be even more robust when examples where PMP *q* → *h* are examined. This was another change which happened regularly in initial (3a), medial (3b) and final (3c) position. An irregular development *q* → *k* is found in three examples (3d) :

(3)	Gloss	PMP	PMC
(a)	liver	qatay	hátay
	black	qitəm	hîtəm
	taro; tuber; yam	qubi	húbi
	worm	quləj	húlət
(b)	branch, bough	daqan	dáhan
	sew	zaqit	jáhit
	bitter; bile	paqit	páhit
	know; can; able	taqu	táhu
(c)	move (residence)	aliq	álih
	tongue	dilaq	dîlah
	raw; green; unripe	məntaq	mə́ntah
	shoot; bow	panaq	pánah
(d)	fart	qə(n)tut	kə́ntut
	leg	qaqay	kákay
	younger sibling	huaji-q	ádi(k)[5]

In forms longer than two syllables, it can be seen that there is a regular deletion of Pre-MC[6] *h*, along with *ə* (either original or derived) (4a). As in (2) above, the ultimate result of this is a reduction to a disyllabic foot. Three other counterexamples are given in (4b):

(4)	Gloss	PMP	Pre-MC	PMC
(a)	shoulder	qabaRa	habara	bára
	salt	qasiRa	hasira	sîra
	egg	qatəluR	hatəlur	tə́lur
	water leech	qali-mətaq	hali-mətah	lîntah
	salted; salty	ma-qasin	ma-hasin	másin
	withered, faded	laqəyu[7]	lahəyu	láyu
	bone	tuqəlaŋ	tuhəlaŋ	túlaŋ

[5] The *k* in this form appears only in PM and not in PC, indicating variants at the PMC level.

[6] I define Pre-MC as the stage of the language which is directly ancestral to PMC, but which post-dated PMP.

[7] This form is more strictly Proto-Western-Malayo-Polynesian.

(b)	centipede	qalu-hipan	halu-ipan	hə̀lVípan[8]
	new; just now	baqəRu	bahəru	baháru
	red	ma-qiraq	ma-hirah	mahírah

1.1.3 PMP V → PMC Ø

In addition to these changes in the PMP consonantal system, there were two kinds of positional vowel deletion which occurred between PMP and PMC. In disyllabic words, there was no deletion of either word-initial vowels (5a) or vowels in hiatus across syllables (5b):

(5)	Gloss	PMP	PMC
(a)	child	anak	ának
	tail	ikuR	íkur
	vein, tendon	uRat	úrat
	one	əsa	ə́sa
(b)	fruit; egg	buaq	búah
	far, distant	zauq	jáuh
	water (fresh)	wahiR (> wair)	áir
	count	ihap (> iap)	íap

In forms longer than two syllables, deletion occurred in initial position (6a). As there is a small body of evidence in both PM (Adelaar 1992: 52-3) and PC that *a* reduced to an *ə* in the unstressed position of trisyllabic forms, it can be assumed that words in which the initial syllable was *ha* reduced to *hə* and underwent regular deletion (6b). This lenition to *ə* may also be assumed for words which had an etymological *ma-* prefix (6c)[9]. Word-internally, *ə* was always deleted when in hiatus with another vowel or when preceded by *h* (which may have been lost first, creating new vowel hiatus) as in (6d). There are two examples of vowel deletion in (6e) when the result would be an NC cluster, and one where the vowel *i* was reanalyzed as part of the initial (6f). (6g) gives examples of three trisyllabic forms where the conditions for vowel deletion were not met, either because there was no *ə* involved and no vowels were in hiatus, or because the deletion of *ə* would have led to an illicit cluster. Finally, (6h) provides one example of a form which (based on the PM evidence, see fn. 5) was not reduced to a disyllabic form, possibly because the initial *hə* bore secondary stress.

(6)	Gloss	PMP	Pre-MC	PMC
(a)	to drink	um-inum	um-inum	(m)ínum
	come	um-ari	um-aray	máray
	younger sibling	huaji-q	uadi-(k)	ádi(k)

[8] This word is somewhat complicated. It shows a quite regular development into disyllabic PC *limpá:n, but the PM form is *həlilipan (Adelaar 1992: 53), implying the persistence of a quadrisyllabic form which persisted at least to the level of PMC.

[9] The one notable exception to this being *red*, PMC *mahirah < PMP *ma-qiraq, see (6g) below.

(b)	shoulder	qabaRa	habara	(> həbára)	bára
	salt	qasiRa	hasira	(> həsíra)	síra
	egg	qatəluR	hatəlur	(> hətə́lur)	tə́lur
	water leech	qali-mətaq	halimətah	(> həlíntah)	líntah
(c)	salted; salty	ma-qasin	ma-hasin	(> mə-hásin)	másin
	die	ma-atay	ma-atay	(> mə-átay)	mátay
	weave; trill	ma-aɲam	ma-aɲam	(> mə-áɲam)	máɲam
(d)	withered, faded	laqəyu	lahəyu		láyu
	bone	tuqəlaŋ	tuhəlaŋ		túlaŋ
	to winnow	tahepi	taẽpi		támpi
	pestle	qahəlu	haəlu		hálu
	sour; vinegar	ma-əsəm	ma-əsəm		(m)ásəm
	drunk	ma-buhək	mabuək		mábuk
	after; behind	udəhi	hudəi		húdi
(e)	water leech	qali-mətaq	hali-mətah		líntah
	ghost; corpse	qanitu	hanitu		hántu
(f)	breath, soul, air	nihawa	niawa		ɲáwa
(g)	red	ma-qiraq	ma-hirah		mahírah
	crocodile	buqaya	buhaya		buháya[10]
	ear	taliŋa	taliŋa		təlíŋa
(h)	centipede	qalu-hipan	halu-ipan		hə̀lVípan

1.1.4 Prosodic motivation for deletion

As argued in Norquest (2003)[11], the predictability in which segments are deleted in the reduction from PMP to disyllabic forms in PMC lies largely in the featural specification and inherent salience of the segments themselves. In the case of consonants, deletion is limited to the glottal fricative *h* (and in some PM forms without PC cognates also to glottal stop *ʔ*), and in the case of vowels, the most commonly deleted segment is *ə* word-internally, although there are specific environments above which affect *i*, and *u* , as well as general deletion of unstressed *a* at the beginning of trisyllabic forms, which as argued above may have first lenited to *ə*.

By the time of PMC, the lexicon consisted overwhelmingly of disyllabic trochees. There were no word-internal clusters, and word-final codas were optional but apparently non-moraic. Prosodic change from Pre-MC to PMC therefore progressed as in the example in (7) below, with the optimal PMC prosodic word having the shape in (7b).[12] In formal

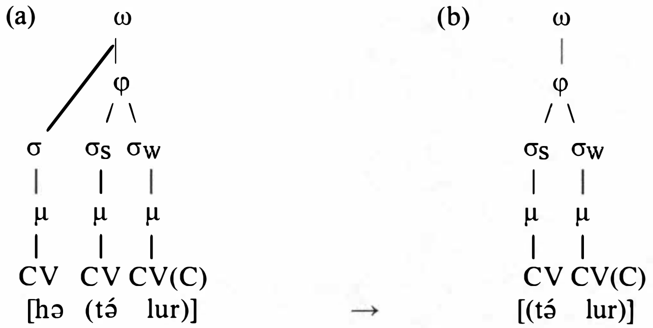
[10] By the time of PC, this form can be constructed as the expected disyllabic form *buyá:*; it remained *buháya* in PM.

[11] See this paper for a formal analysis of the changes from PMP to PMC in line with the analysis offered below for PMC to PC.

[12] The following symbols are used here for the respecitive prosodic categories: ω = prosodic word, φ = foot, σ = syllable, ς = nonmoraic sesquisyllable (Cho & King 2003), μ = mora; on syllables and morae, subscript S = 'strong' and W = 'weak'.

terms, the left edge of the foot became aligned with the left edge of the prosodic word, allowing these two prosodic categories to be fully coterminous:

(7) Reduction of Pre-MC form to PMC optimal prosodic word template



Given the amount of evidence above that disyllabic trochees were so strongly preferred in PMC, one may ask why any exceptions such as those in (6g) remained at all. It is argued below that success or failure to reduce to a disyllabic form hinged entirely on the segmental composition of the word in question -- not all segments were equally robust. Segments with place were salient enough to be reproduced at the PMC level in violation of the optimal prosodic word; segments without place were not salient enough to be reproduced, and their deletion allowed the parsing of the optimal prosodic word. However, this latter class of segments was deleted only in the case that the word was more than two syllables – if the word was already a disyllabic trochee, then these weaker segments were retained.

1.2 Optimality Theoretic analysis of PMC grammar

Before beginning the analysis proper, I would like to briefly discuss my assumptions about how an account of historical change should be cast within Optimality Theory (hereafter OT).

There has been a tendency in OT analyses of historical sound change to use the same model which is normally adopted for language learning. The latter normally assumes that an individual begins with a certain set of constraints, ranked in a certain order which explains their utterances at some stage of acquisition. The changes in the individual's grammar are normally explained by maintaining the same set of constraints, and allowing successive reranking (optimally in such a way that neighboring constraints are re-ranked by adjoining pairs, without some constraint 'skipping' over several others).

While this is a reasonable hypothesis of how to model the evolving grammar of a single individual, there is no justification for treating the transmission of language X from an earlier generation Y to a later generation Z in the same way. Since every learner will start from a null point and generalize their categories accordingly over more and more data, not only will constraint rankings have the potential to be very different from one generation to the next, but the constraints themselves may even differ (the same can be said for features, prosodic categories, etc.). What this ultimately means is that it will often be the case that the grammars of successive generations of speakers will resemble each other in many ways, because their input data over which they formed generalizations was similar.

However, they need not be *required* to be similar, nor will their differences need to proceed systematically from one grammar to the next as happens in models of language acquisition of a single grammar.

That being said, the differences in grammars between PMP and PMC, between PMC and PC, and between PC and its descendants (as well as the two extra-Chamic examples I draw on for comparison, the Oceanic languages Rotuman and Kwara'ae, and Proto-Hlai) *do* appear to draw on a pool of similar constraints, which in turn allows for informative typological comparison. Moreover, similar changes occur in each of these cases which allow a postulation of similar tendencies in sound change. They thus collectively comprise an interesting typological microcosm from which broader generalizations may be drawn.

1.2.2 Proto-Malayo-Chamic footing and prosodic constraints

PMC, like its PMP ancestor, had a trochaic rhythm timed at the level of the syllable, with a one-to-one correspondence between syllables and morae. The constraints required for an analysis of PMC footing are given below:

(8) PMC Footing Constraints

(a) ALLFEETRIGHT: Align (φ, Right, ω, Right)
'The right edge of every foot must be aligned to the right edge of some prosodic word which contains it' (Prince & Smolensky 1993)

(b) FTBRANCH(S-W): Branch (φ, Strong-Weak)
'Feet must branch into a strong-weak (trochaic) pair at some level'
(based on Ussishkin 2000; this merges the functions of traditional FOOTBINARITY and FOOTFORM (trochaic))

(c) ALLFEETLEFT: Align (φ, Left, ω, Left)
'The left edge of every foot must be aligned to the left edge of some prosodic word which contains it' (Prince & Smolensky 1993)

(d) ALIGNSYLL: Align (σ, Left, φ, Left; σ, Right, φ, Right)
'Both edges of every syllable must be aligned to both edges of some foot
which contains it' (based on ALLFEETLEFT/ALLFEETRIGHT, merging their directionality)

For both PMP and PMC, the constraints above must have the ranking in (9):

(9) ALLFEETRIGHT, FTBRANCH(S-W) >> ALLFEETLEFT >> ALIGNSYLL.

In tableau (10) below, the relevance of ALLFEETRIGHT can be seen by comparing (10a) and (10d) – it ensures that all feet will be aligned to the right edge of the prosodic word. (10d) also fails because there is a single strong constituent of the foot with no weak counterpart, violating FOOTBRANCH. (10c), an iambic foot, violates FOOTBRANCH since it has a weak-strong pattern, and (10b) is suboptimal because of the failure to parse the first syllable:

(10) Proto-Malayo-Chamic Constraint Hierarchy (*kulit* 'skin')

/kulit/	ALLFEETRIGHT	FOOTBRANCH	ALLFEETLEFT	ALIGNSYLL
☞ a. [(kú$_{\mu}$.lit$_{\mu}$)]				σσ
b. [ku.(lī́t$_{\mu\mu}$)]			σ!	
c. [(ku$_{\mu}$.lī́t$_{\mu\mu}$)]		W!S		σσ
d. [(kú$_{\mu}$).lit]	σ!	*		

1.2.3 Faithfulness and Markedness constraints

The necessity of considering segmental and/or featural faithfulness constraints is shown through the consideration of two additional candidates:

(11) Problems with a constraint inventory governing only prosody

/kulit/	ALLFEETRIGHT	FOOTBRANCH	ALLFEETLEFT	ALIGNSYLL
☞ a. [(klī́t$_{\mu\mu}$)]				
☞ b. [(lī́t$_{\mu\mu}$)]				
☜ c. [(kú$_{\mu}$.lit$_{\mu}$)]				σ!σ
d. [ku.(lī́t$_{\mu\mu}$)]			σ!	
e. [(ku$_{\mu}$.lī́t$_{\mu\mu}$)]		W!S		σσ
f. [(kŭ́$_{\mu}$).lit]	σ!	*	σ	

The inclusion of these new candidates indicates that there is something lacking in the present analysis, as the constraint inventory and ranking chosen thus far selects forms which may have a deleted vowel (11a) and even a deleted consonant (11b), at the expense of parsing all syllables into feet and aligning the edges of foot-internal syllables with the edges of the feet which contain them – in other words, there is nothing which requires segmental faithfulness. As this optimal prosodic alignment between syllable and foot generally fails to occur at the level of PMC, a more sophisticated analysis is required. In order to do this, as well as capture the change between PMP and PMC of forms larger than two syllables, the following additional faithfulness and markedness constraints are necessary:

(12) Faithfulness and Markedness constraints necessary in the PMC inventory

(a) MAX-C: Every consonant in the input has a correspondent in the output. (Prince & Smolensky 1993)

(b) MAX-V: Every vowel in the input has a correspondent in the output. (Prince & Smolensky 1993)

(c) MAX-C[PLC][13]: The place feature of every oral consonant in the input has a correspondent in the output (where *ʔ* and *h* do not have place). (modeled on MAX-C)

(d) MAX-V[PLC]: The place feature of every vowel in the input has a correspondent in the output (where *ə* does not have place). (modeled on MAX-V)

(e) *CC: There are no consonant clusters in the output. (Prince & Smolensky 1993)

(f) CONTIGUITY: Segments which are contiguous in the input remain contiguous in the output.(Prince & Smolensky 1993)

While the constraints in (12a-b) and (12c-d) may look very similar, they are qualitatively quite different.

MAX-C and MAX-V target segments, regardless of their featural make-up, and act to preserve all segments equally. MAX-C[PLC] and MAX-V[PLC], in contrast, target the features which inhere in segments. For example, MAX-V will enforce faithfulness of the full inventory of PMP vowels *i*, *u*, *ə*, and *a* because they are all segments in an equal sense. MAX-V[PLC], on the other hand, will only enforce faithfulness of the vowels *i*, *u*, and *a*, because they can be defined by the features [high, front], [high, back], and [low] respectively, but not *ə*, which lacks specific features.

1.2.4 The PMP and PMC constraint inventories and rankings

At the PMP level, the constraints in (12) are all undominated, with MAX-C and MAX-V making the more specific constraints MAX-C[PLC] and MAX-V[PLC] redundant, with no reason to assume their place in the grammar at all[14]:

(13) The full PMP constraint inventory and ranking (*qatəluR* 'egg')

/qatəluR/	MAX-C	MAX-V	*CC	CONTIG	ALLFTRIGHT	ALLFTLEFT
☞ a. [qa.(tə́.luR)]						σ
b. [(qá.tə).luR]					σ!	
c. [(qá.tluR)]		ə!	tl	tl		
d. [(qá.luR)]	t!	ə		al		
e. [(tə́.luR)]	q!	a				

However, in the grammar leading to PMC, it is necessary to assume that MAX-C and MAX-V were replaced by MAX-C[PLC] and MAX-V[PLC] at the top of the constraint hierarchy, leaving the glottal consonants and *ə* vulnerable to deletion, forcing violations of CONTIGUITY. The constraint ranking which is consistent with the forms in PMC is:

[13] MAX-C[PLC] and MAX-V[PLC] should not be confused with IDENT constraints, which target features within a segment but take the ultimate parsing of that segment in the output for granted; in the case at hand, the parsing of the segment itself is dependent on the targeting of its features by these constraints.

[14] In the following tableaux, FTBRANCH(S-W) is implicitly assumed as undominated and ALIGNSYLL as completely dominated, but these are excluded for ease of presentation

(14) MAXC[PLC], *CC, ALLFEETRIGHT >> ALLFEETLEFT, MAXV[PLC] >> CONTIG (>> MAX-C, MAX-V):

Under this inventory and ranking, it can be seen how reduction to a disyllabic trochee is achieved:

(15) The full PMC constraint inventory and ranking (Pre-MC *ħətəlur* → PMC *tə́lur* 'bone')

/ħətəlur/	MAXC[PLC]	*CC	ALLFTRIGHT	ALLFTLEFT	MAXV[PLC]	CONTIG
☛ a. [(tə́.lur)]						
b. [ħə.(tə́.lur)]				σ!		
c. [(ħə́.tə).lur]			σ!			
d. [(ħə́.tlur)]		t!l				tl
e. [(ħə́.lur)]	t!					əl

This constraint inventory and ranking is also sufficient to avoid the unwanted reduction in (11) above:

(16) Faithfulness to segmental material in PMC (Pre-MC *kulit* → PMC *kúlit* 'skin')

/kulit/	MAXC[PL C]	*CC	ALLFTRIGHT	ALLFTLE FT	MAXV[PL C]	CONTIG
☛ a. [(kú.lit)]						
b. [ku.(lît)]				σ!		
c. [(kú).lit]			σ!	σ		
d. [(klît)]		k!l			u	kl
e. [(lît)]	k!				u	

Tableau (17) is an example where there is a placeless vowel in the first syllable which would be expected to be deleted, but where the initial and following consonants cannot form a licit cluster (17b); the vowel must therefore be retained to break up the cluster:

(17) Segmental material blocking reduction to disyllabic trochee (Pre-MC *təliŋa* → PMC *təlíŋa* 'ear')

/təliŋa/	MAXC[PLC]	ALLFTRIGHT	*CC	ALLFTLEFT	MAXV[PLC]	CONTIG
☛ a. [tə.(lî.ŋa)]				σ!		
b. [(tlî.ŋa)]			t!l			tl
c. [(tə́.li).ıŋa]		σ!		σ		
d. [(lî.ŋa)]	t!					

Tableau (18) shows the resolution which occurs when two vowels are in hiatus within a word and one of them is *ə*:

(18) Deletion of *ə* when in hiatus with another vowel (Pre-MC *mə-atay* → PMC *mátay* 'die')

/mə-atay/	MAXC[PLC]	ALLFTRIGHT	*CC	ALLFTLEFT	MAXV[PLC]	CONTIG
☞ a. [(má.tay)]						ma
b. [(mə́.tay)]					a!	ət
c. [mə.(á.tay)]				σ!		
d. [(mə́.a).tay]		σ!		σ		

The following tableau shows the importance of CONTIGUITY, which selects between (19a) and (19b):

(19) Deletion of the leftmost vowel in a vowel-initial form (Pre-MC *uadi(k)* → PMC *ádi(k)* 'y. sibling')

/uadi(k)/	MAXC[PLC]	ALLFTRIGHT	*CC	ALLFTLEFT	MAXV[PLC]	CONTIG
☞ a. [(á.di(k))]					u	
b. [(ú.di(k))]					a	u!d
c. [u.(á.di(k))]				σ!		
d. [(ú.a).di(k)]		σ!		σ		

And finally, tableaux (20) and (21) show that even though they are very low-ranked, MAX-C and MAX-V still play a role in the grammar (*CC and CONTIGUITY are omitted to conserve space):

(20) Retention of *h* in an optimal prosodic word (Pre-MC *hatay* → PMC h*á.tay* 'liver')

/hatay/	MAXC[PLC]	ALLFTRIGHT	ALLFTLEFT	MAXV[PLC]	MAX-C
☞ a. [(há.tay)]					
b. [(á.tay)]					h!
c. [ha.(táy)]			σ!		
d. {(há).tay]		σ!			

(21) Retention of *ə* in an optimal prosodic word (Pre-MC *ənəm* → PMC *ə́.nəm* 'six')

/ənəm/	MAXC[PLC]	ALLFTRIGHT	ALLFTLEFT	MAXV[PLC]	MAX-V
☞ a. [(ə́.nəm)]					
b. [(nə́m)]					ə!
c. [ə.(nə́m)]			σ!		
d. [(ə́).nəm]		σ!			

2. From Proto-Malayo-Chamic to Proto-Chamic

Upon the break-up of PMC, speakers of pre-Chamic relocated to the Southeast Asian mainland, where their language underwent intense contact with speakers of Mon-Khmer (MK) languages (Thurgood (1999)). This contact had the effect of shifting the main stress of the prosodic word from the penultimate syllable to the final syllable, which then became heavy and could include a moraic final consonant.

Along with the shift to word-final stress, certain consonant clusters also became permissible for the first time. This presumably resulted not just from the change in stress pattern, but also because this type of phonotactic pattern became increasingly accessible to Chamic speakers as they were exposed to greater and greater volumes of MK vocabulary. Below is a list of some words with clusters which are reconstructible at the Proto-Chamic (hereafter PC) level that are loan words from various MK sources[15]. Words in (22a) are stop + laryngeal clusters/ implosives[16], and those in (22b) are stop + liquid clusters:

(22)	Gloss	PC	Source of loan	
(a)	different	pha:	PNB:	pha
	cloth; blanket	khan	PNB:	khan
	face	ɓɔ:ʔ	Bahnar:	bɔ̀k
	lie on back	ɗa:ŋ	PNB:	qdla:ŋ
(b)	skirt	blah	PNB:	blah
	boa, python	klan	PSB:	klan
	squirrel	prɔ:k	PSB:	prɔ:ʔ
	eggplant	trɔŋ	PNB:	troŋ

After contact with MK and the resulting shift to word-final stress, there remained essentially two classes of words: those which remained disyllabic (and rarely trisyllabic), and those which reduced to heavy, monosyllabic forms. I will argue shortly that the difference was due entirely to the tension between prosodic pressure to reduce disyllabic forms to monosyllabic forms in order to further align the edges of prosodic categories on the one hand, and the continued pressure to retain segmental material on the other.

2.1 New monosyllabic forms in PC

There were three classes of words which became monosyllabic in PC (see Thurgood 1999: 60-66), detailed below.

2.1.1 Deletion of initial vowels

The first class of words which became monosyllabic were those which began with an initial *ə* (23a) or where an initial *i* could be reinterpreted as part of the initial (23b). Vowel-initial words with other vowels did not reduce (23c), with one exception given in (23d). In

[15] PNB = Proto North Bahnaric, PSB = Proto-South Bahnaric. For references, see Thurgood (1999).

[16] I include words with initial implosives here, since there are two words (given below) with PMP etymologies which appear to contain *ɓ* as a result of the coalescence of *b* + *ʔ*.

the data below, both Proto-Malayic[17] (PM) and PC forms will be provided after the PMC form so that it can be observed upon what data the PMC is being reconstructed:

(23)	Gloss	PMC	PM[18]	PC
(a)	one	ə́sa	əsaʔ	sá:
	master; lord	ə́mpu	*əmpu*	pɔ́:[19]
	four	ə́mpat	əmpat	pá:t
	six	ə́nəm	ənəm	nám
(b)	blow nose; mucus	íŋus	iŋus	ɲús
	count	íap	----	yá:p
(c)	tail	íkur	ikur	ʔikú:
	fish	íkan	ikan	ʔiká:n
	snake	úlər	ulər	ʔulár
	person, someone	úraŋ	uraŋ	ʔurá:ŋ
	the wind	áŋin	aŋin	ʔaŋín
	fire	ápuy	api	ʔapúy
(d)	I	áku	aku	kə́w

(23d) may have lost its intial vowel through an irregular development of *a* to *ə*, although this must remain conjectural based on present evidence.

2.1.2 Stop + laryngeal clusters

The second class of words which contracted to monosyllables were those with an initial stop and a medial laryngeal consonant, either *h* (24) or *ʔ* (26). Stop + *h* contraction occurred when *h* was preceded by the vowel *a* (24a); there is an example of contraction when there was a sequence of identical high vowels (24b) (this is the only example with *h* flanked by identical high vowels in PMC). The palatal stop (or affricate) *j* could not form a cluster with *h* (24c):

(24)	Gloss	PMC	PM	PC	
(a)	bitter; bile	páhit	*pahit*	phít	
	chisel; to plane	páhət	pahət	phá:t	< pahat
	thigh	páha	paha(ʔ)	phá:	
	know; be able	táhu	tahu(ʔ)	thə́w	
	year	táhun	tahun	thú:	

[17] I do not mark stress on PM in order to remain faithful to the original reconstructions, the vast majority of which are taken from Adelaar (1992). I assume that it was the same as stress in PMC, which I do indicate.

[18] Occasionally when there is no PM form available, a form from Malay will be substituted; Malay words will always be placed in italics.

[19] The vowel in this form is irregular (the expected form is *pə́w*).

	forehead	dáhi	dahi	dhə̆́y[20]
	branch; bough	dáhan	dahan	dhá:n
(b)	trunk; log; stem	púhun	puhun	phún
(c)	sew	jáhit	jahit	jahít
	bad; wicked	jə́hat	jah[a]t	jəhá:t

It is possible that there was something about *h* which interfered phonetically with the cues distinguishing a preceding *a* from a preceding *ə* (25a); in an identical high-vowel sequence, the first vowel may have been reanalyzed as *ə* on the assumption that it was a carry-over from the second vowel (a case of dissimilation) (25b). In these cases, the following scenarios are possible:

(25)	Original PC form		Perceived as		Phonologized as		Reduction
(a)	/tahú:/	→	[tə̆hú:]	→	/təhú:/	→	[thú:]
(b)	/puhún/	→	[pə̆hún]	→	/pəhún/	→	[phún]

Although the number of examples is limited to two, there is also reason to believe that voiced stops could coalesce with *ʔ*, ultimately leading to an implosive (26a). There are no examples in PMC of a *d* + *ʔ* sequence, and therefore no examples of a resulting implosive *ɗ*. There is one exception to the expected pattern (26b), and evidence in (26c) that implosives could not result from a voiceless stop + *ʔ* sequence:

(26)	Gloss	PMC	PM	PC
(a)	head hair	búʔuk	buʔuk	ɓúk
	stench	báhu	bahu	ɓə́w[21]
(b)	paper; book	----	----	baʔár
(c)	armpit	----	----	paʔá:k
	knee	túʔut	tuʔ[u]t	tuʔút

2.1.3 Stop + liquid clusters

The third class of words to reduce was those where the medial consonant was a liquid which could form a phonotactically permissible cluster with a preceding stop. Generally, disyllabic forms were retained when the first of the two vowels was any vowel other than *ə*, before both *l* (27a) and *r* (27b):

[20] Thurgood (1999) reconstructs this as *ʔadhə̆́y, with an initial syllable; since it is supported by only a single language (Rade), I prefer to see that as an independent development and reconstruct *dhə̆́y.

[21] The sequence of changes in this word was presumably bǎhú: → bə̆hú: → bə̆ʔú: → ɓú: → ɓə́w.

(27)	Gloss	PMC	PM	PC
(a)	tongue	dílah	dilah	diláh
	skin	kúlit	kulit	kulít
	moon	búlan	bulan	bulá:n
	dig	káli	kali	kalə́y
(b)	tortoise; turtle	kúra	*kura*	kurá:
	thorn	dúri	duri(ʔ)	durə́y
	shoulder	bára	bara	bará:
	blood	dárah	darah	daráh

However, contraction of stop + liquid occurred regularly when the first vowel was *ə* before either *l* (28a) or *r* (28b); the one exception to this being the word for *ear* (28c) -- the form *tliŋá:* would be otherwise expected:

(28)	Gloss	PMC	PM	PC
(a)	chop; split	bə́lah	bəlah	bláh
	buy	bə́li	bəli	blə́y
	three	tə́lu	təlu	tlə́w
	egg	tə́lur	təlur	tlú:[22]
(b)	(husked) rice	bə́ras	bəras	brá:s
	give	bə́ri	bəri(ʔ)	brə́y
	fast; short time	də́rəs	*dəras*	drás
	monkey (chatter)	kə́ra	kəra	krá:
(c)	ear	təlíŋa	təliŋa(ʔ)	təliŋá:

In addition, words in which there were identical high vowels and where the second consonant was *r* underwent contraction as well (29a). There is one example of this contraction in the case of *l* (29b), but otherwise it seems not to have occurred in *l*-medial forms (29c):

(29)	Gloss	PMC	PM	PC
(a)	self; body	díri	diri	drə́y
	rotten	búruk	buruk	brúʔ
	descend	túrun	turun	trún
(b)	ten	púluh	puluh	plúh
(c)	twist	bə́lit	bəlit	bilít[23]
	body hair	búlu	bulu	bulə́w
	to roll	gúluŋ	*guluŋ*	gulúŋ

[22] The reduction in this word may be a post-PC development, since there is one language (Jarai) which has been recorded as disyllabic: *tə̆́lu* (also noted in Blust 2000: 440)

[23] The first vowel in this form is irregular (bəlít → blít would be expected).

It is unclear if the voicing of the initial consonant was important, as the initial of *ten* is voiceless but the initials of the other words are all voiced; an alternative explanation may lie in the fact that *ten* was a word of higher frequency and thus more prone to reduction.

In the case of the forms in (29a), I suggest that there was something about the phonetic implementation of *r* (articulatorily and/or acoustically) which may have caused confusion in the learner about whether the first of the two identical high vowels was intentional or whether it was underlyingly *ə*. If this is so, then a scenario similar with that in (25) can be suggested:

(30)	Original PC form		Perceived as		Phonologized as		Reduction
	/turún/	→	[tə̆rún]	→	/tərún/	→	[trún]

It can therefore be suggested that *r* interfered with the perception of a preceding high vowel in the same way that *h* interfered with the perception of a preceding low vowel.

2.2 OT analysis: The shift to PC rhythm

As mentioned above, the shift to the PC stress pattern was triggered by contact with speakers of sesquisyllabic Mon-Khmer languages. This contact involved the absorption of a large number of loan-words into the Chamic lexicon, all of which presumably bore word-final stress; words with foreign phonotactic patterns were borrowed as well. As contact intensified, there was a wholesale shift to a word-final stress pattern in Chamic:

(31)	Former PMC rhythm		New PC rhythm
	kú.lit	→	ku.lît

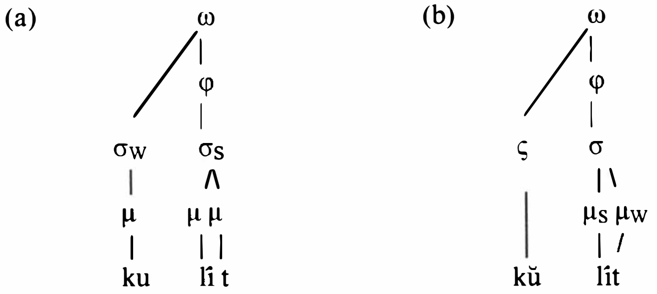
I argue here that PC feet were trochaic heavy syllables, which could be preceded by one (and rarely two) unfooted sesquisyllables. If the optimal foot had been iambic, it seems that there should have been more systemic pressure to retain the fully-footed iambic forms; instead, I attempt to show below that wherever a word could be reduced to a monosyllable it was, and the retention of all di- (and tri-)syllabic forms was merely a response to segmental faithfulness, not an instantiation of true iambic rhythm.

An important change must have occurred initially at the phonetic level, where the Weight-to-Stress principle became active in the language and stressed final syllables lengthened (and were analyzed as heavy) in correlation with their prominence; initial syllables were likely shortened compensatorily. Although at first being a strictly phonetic effect, and not represented phonologically in the lexicon of the speaker, at some point this length became phonologized:

(32)	Original Form		Spoken as		Phonologized as
	/ba$_{\mu}$.lú$_{\mu}$/	→	[bă.lú:]	→	/ba.lú$_{\mu\mu}$/

The difference in underlying structure between a true iamb (33a) and a sesquisyllabic trochee (33b) is shown below. Note that unlike in PMC, a final consonant is assigned a mora, a crucial difference between the two:

(33) A true iamb vs. a heavy syllabic trochee preceded by unfooted sesquisyllable



To express the change to PC foot structure, a reversal of ALLFEETLEFT and ALIGNSYLL is required in the new PC grammar:

(34) *Proto-Chamic Foot Structure*

/kulit/	ALLFEETRIGHT	FOOTBRANCH	ALIGNSYLL	ALLFEETLEFT
☛ a. [kŭ.(lît$_{\mu\mu}$)]				σ
b. [(kú$_{\mu}$.lit$_{\mu}$)]			σ!σ	
c. [(ku$_{\mu}$.lît$_{\mu}$)]		W!S	σσ	
d. [(kú$_{\mu}$).lit]	σ!	*		

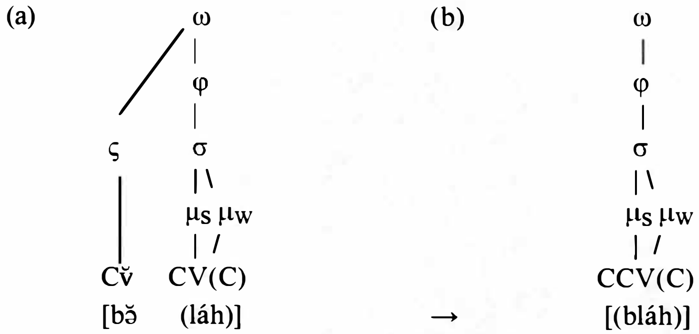
In this tableau, (34d) fails for the same reasons as its PMC counterpart above. (34c) fails because it is not a strong-weak pair, violating FOOTBRANCH(S-W). The failure of (34b) may be formally expressed as being due to a violation of ALIGNSYLL, which requires the alignment of both edges of all parsed syllables within the foot containing them (making their edges coterminous), resulting in the PC pattern. Since it is no longer possible at the level of the syllable, this requires that FOOTBRANCH(S-W) be satisfied at the level of the mora; it also entails the inability to parse the initial syllable into a foot[24].

2.2.1 Faithfulness and Markedness Constraints

Despite appearances, the foot in PC remains trochaic, with FOOTBRANCH(S-W) being satisfied at the moraic instead of the syllabic level. As in the case of the period between PMP and PMC, there was once again a drive to align the edges of the foot and the prosodic word. Optimally, this occurred in the following way:

[24] While ternary feet are a logically possible option, I will not treat them here.

(35) Alignment of the edges of feet and prosodic words in PC



The following constraints are necessary to complete the analysis of the change from PMC to PC:

(36) PMC → PC Faithfulness and Markedness Constraints

(a) MAX-C: Every consonant in the input has a correspondent in the output.
(b)MAX-V[PLC]: The place feature of every vowel in the input has a correspondent in the output.
(c) *CC: Syllables do not have complex onsets or codas.

*CC prevents all complex codas and most complex onsets. This must be exploded into a constraint family:

(37) *CC[25] (all types except for OR and OH) >> *OR, *OH

Without ranking *CR and *CH low in the grammar, stop + laryngeal and stop + liquid clusters would be impossible, which is obviously not the case in PC. The best explanation for this more complex ranking lies outside of the OT grammar proper. As the frequency of MK borrowings with CR and CH clusters increased and Chamic speakers became more practiced, these specific types of clusters would have eventually become nativized (driving down *CR and *CH). This in turn forced the learner to choose whether or not a perceived form [bə̆.láh] was derived from one of two possible underlying representations:

(38)

Original Representation	Spoken	Possible Representations	
/bə.láh/	[bə̆.láh]	(a)	/bə.láh/
		(b)	/bláh/

Since words with *bl* clusters also existed, it may have been easily assumed that the [ə̆] in this word was merely a phonetic transition between the two consonants of a cluster, and assume (38b) accordingly as the new underlying representation. This is because *ə*, like

[25] C = any consonant, O = non-palatal obstruent, R = liquid, H = laryngeal.

the transition, lacks vocalic features which would otherwise provide evidence that it constituted an individual (and intentional) phonological entity. Eventually, the representation in (38a) failed to be an option, and that in (39b) became the only choice[26].

Tableau (39) shows a normal case of stop + liquid contraction. The loss of a consonant is not possible due to MAX-C (39d). (39b) is prosodically suboptimal because it contains an unparsed syllable, and (39c) fails because a parsed syllable exists which is not aligned with the right edge of the foot. (39a) emerges as the winning candidate since [ə] is a placeless vowel and therefore doesn't incur a violation of MAX-V[PLC]:

(39) PMC bə́lah → PC bláh 'chop; split'

/bəlah/	MAX-C	MAX-V[PLC]	*CC	ALIGNSYLL	ALLFTLEFT
☞ a. [(bláh)]					
b. [bə̆.(láh)]					σ!
c. [(bə.láh)]				σ!σ	
d. [(láh)]	b!				

Tableau (40) shows a situation where there is a featureless vowel in the first syllable, but where the initial and final consonants cannot form a licit cluster; (40c) therefore incurs a violation of *CC:

(40) PMC də́pa → PC də̆pá:

/dəpa/	MAX-C	MAX-V[PLC]	*CC	ALIGNSYLL	ALLFTLEFT
☞ a. [də̆.(pá:)]					σ
b. [(də.pá:)]				σ!σ	
c. [(dpá:)]			d!p		
d. [(pá:)]	d!				

Tableau (41) and (42) show that deletion of vowels with place features (41c), (42b) is unacceptable. The PMC form must preserve segmental information at the expense of an unparsed syllable at the left edge of the prosodic word:

(41) PMC kúlit → PC kŭlît 'skin'

/kulit/	MAX-C	MAX-V[PLC]	*CC	ALIGNSYLL	ALLFTLEFT
☞ a. [kŭ.(lît)]					σ
b. [(ku.lît)]				σ!σ	
c. [(klît)]		u!			

[26] There are languages where both kinds of representation are possible, for example the Kammu minimal pair *klɔ́:k* 'bamboo bowl' versus *kəlɔ́:k* 'slit drum' (Pittayaporn, in preparation)

(42) PMC ȋkan → PC ĭ.kán (→ ʔiká:n) ‘fish’

/ikan/	MAX-C	MAX-V[PLC]	*CC	ALIGNSYLL	ALLFTLEFT
☞ a. [ĭ.(kán)]					σ
b. [(ká:n)]		i!			

Finally, (43) shows that an initial ə, when not protected by an onset, will be lost:
(43) PMC ə́nəm → PC nə́m (→ nám) ‘six’

/ənəm/	MAX-C	MAX-V[PLC]	*CC	ALIGNSYLL	ALLFTLEFT
☞ c. [(nə́m)]					
a. [ə̆.(nə́m)]					σ!

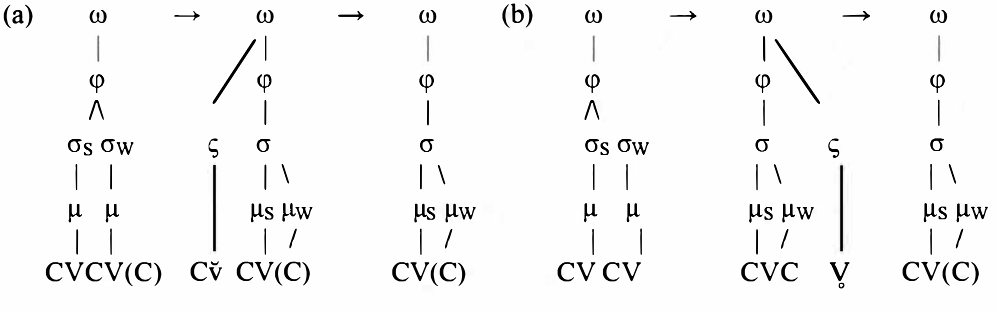
2.3 Rotuman and Kwara’ae: syllable alignment at the right edge of the foot.

All examples of prosodic alignment so far have involved left-edge (PMC and PC) and word-internal (PMC) readjustments. The question arises of whether or not there are examples of right-edge readjustment which are driven by the same principles as in the cases discussed so far. This may be answered affirmatively, and this section takes a detour in Austronesian away from Chamic to Oceanic, within which two languages, Rotuman and Kwara’ae, show just these effects.

The results are quite similar: segmental material which is parsed is preserved, while unparsed material is ultimately lost, the difference being that the loss occurs on the left edge of the prosodic word in Chamic (44a), but on the right edge of the prosodic word in Rotuman and Kwara’ae (44b):

(44) Alignment of prosodic categories in Chamic and the Oceanic languages Rotuman and Kwara’ae

(45)



2.3.1 The Rotuman and Kwara’ae register distinction

Rotuman and Kwara’ae share in common the use of two distinct sociolinguistic registers, one which is more conservative and referred to here as the *citation form*, and the other which is used in informal day-to-day conversation, and is referred to here as the *discourse form* (for more specific information on the usage of these registers, see Churchward (1940) for Rotuman and Watson-Gegeo & Gegeo (1986) for Kwara'ae, along with the rest of the references pertaining to these languages in the bibliography).

In their conservative citation forms, Rotuman and Kwara'ae have in common the fact that all syllables are light (monomoraic) and lack codas. In the innovative discourse register, both of these facts change; footed syllable nuclei are universally heavy, and codas are not only possible but common. Most importantly, both edges of syllables are aligned with the edges of the feet which contain them.

In the shift from disyllabic to monosyllabic feet, there are four possible outcomes in both Rotuman and Kwara'ae depending on the nature of the input segments. The Rotuman and Kwara'ae forms below are given with the conservative citation register on the left and the innovative discourse register on the right:

2.3.1.1 Vowel tautosyllabification

If the final syllable has no onset, then *vowel tautosyllabification* results, where the final two vowels are parsed together into the nucleus of a single, heavy syllable:

(45) Vowel tautosyllabification in Rotuman and Kwara'ae

(a) Rotuman			(b) Kwara'ae		
Citation	Discourse	Gloss	Citation	Discourse	Gloss
(ké.u)	(kéu)	'to push'	(gé.o)	(géo)	'megapod'
pu.(pú.i)	pu.(púi)	'floor'	a.(bú.i)	a.(búi)	'to climb'

2.3.1.2 Metathesis

If the final consonant has an onset, then three outcomes are possible. If the last two vowels are compatible in a closed nucleus, then *metathesis* occurs.

(46) Metathesis in Rotuman and Kwara'ae

(a) Rotuman			(b) Kwara'ae		
Citation	Discourse	Gloss	Citation	Discourse	Gloss
(hó.sa)	(hóas)	'flower'	(sé.lo)	(séol)	'sail'
se.(sé.va)	se.(séav)	'erroneous'	da.(lú.ma)	da.(lúəm)	'bailer, to bail'

2.3.1.3 Coalescence

If they are not, but if some feature (usually [front] and/or [hi]) of the second vowel may be preserved, then *coalescence* occurs.

(47) Coalescence in Rotuman and Kwara'ae

(a) Rotuman			(b) Kwara'ae		
Citation	Discourse	Gloss	Citation	Discourse	Gloss
(fú.ti)	(fýt)	'to pull'	(mó.li)	(mǿ:l)	'lemon'
fa.(mó.ri)	fa.(mǿr)	'people'	a.(lá.ge)	a.(lǽ:ŋg)	'seaweed'

2.3.1.4 Deletion

Finally, if the two vowels are identical or if they are otherwise incompatible, complete *deletion* is observed.

(48) Deletion in Rotuman and Kwara'ae

(a) Rotuman			(b) Kwara'ae		
Citation	Discourse	Gloss	Citation	Discourse	Gloss
(sú.lu)	(súl)	'coconut-spathe'	(sá.ta)	(sá:t)	'name'
fe.(ʔé.ni)	fe.(ʔén)	'zealous'	sa.(tá.da)	sa.(tá:nd)	'their name'

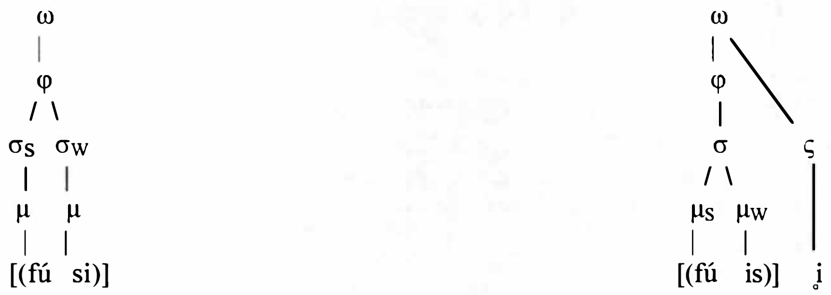
Although the register distinction seems to be quite discrete in Rotuman, it is a bit less so in Kwara'ae, and there is evidence for how the distinction actually came about. Both Blevins and Garrett (1998) and Heinz (2005) describe their respective work with native speakers of Kwara'ae, and describe similar conditions in the discourse register under which there is a short, devoiced vowel where it would be expected in the citation form. Examples from both sources are given below:

(49) Voiceless final vowels in the Kwara'ae discourse register

(a) Kwara'ae (Blevins & Garrett 1998: 530)			(b) Kwara'ae (Heinz 2005: 29)		
Gloss	Citation	Discourse	Gloss	Citation	Discourse
cat	fúsi	húisi̥	fear	máʔu	máŭʔu̥
thin	kádo	káodo̥	wife	ʔáfe	ʔáĕ.he̥
name	sáta	sá:tḁ	to burst	búsu	bú:su̥

This seems to indicate a past situation in which phonetic lengthening occurred, allowing the transition between the first vowel and second vowel to begin before the implementation of the second consonant within the foot. This was eventually phonologized as a metathesized segment, and in the end the final portion of the second vowel was lost altogether, most likely due to perceptual difficulty:

(50)	Original	Pronounced	Perceived	Phonologized
	/fú.si/	[fũĩsi̥]	[fũĩs(i̥)]	/fuis.(i̥)/



These facts, although of interest in and of themselves, are relevant to Chamic because they show the same change (light syllables becoming heavy under stress with concomitant effects on foot structure), but with a twist, since the effect occurs with trochaic forms in which stress remains on the first syllable – in both cases, the change seems to be completely language-internal, and not triggered by language contact as in the case of Chamic.

2.3.2 Optimality Theoretic Analysis

I assume *a priori* that both Rotuman and Kwara'ae feet are universally trochaic. The footing pattern is slightly different between the two, in that the Rotuman prosodic word only has a single head foot aligned to its right edge, whereas Kwara'ae also has a head foot aligned at the right edge of the prosodic word, but additionally has secondary feet which iterate inward from the left edge of the prosodic word. Although these differences are interesting in and of themselves, to treat them in detail here would exceed the relevance of these two languages to the present topic; I shall focus on Rotuman generally hereafter, with the exception of one case in which Kwara'ae provides some significant information.

2.3.2.1 Rotuman Footing

The prosodic constraints necessary to capture the Rotuman footing pattern are given in (48) below:

(51) Rotuman Prosodic Footing Constraints

(a) FTBRANCH(S-W): Branch (φ, Strong-Weak)
'Feet must branch into a strong-weak (trochaic) pair at some level'

(b) ALLFEETRIGHT: Align (Ft, Right, PrWd, Right)
'The right edge of every foot must be aligned to the right edge of some prosodic word which contains it'

(c) ALLFEETLEFT: Align (φ, Left, ω, Left)
'The left edge of every foot must be aligned to the left edge of some prosodic word which contains it' (Prince & Smolensky 1993)

(d) ALIGNSYLL: Align (σ, Left, φ, Left; σ, Right, φ, Right)
'Both edges of every syllable must be aligned to both edges of some foot which contains it'

In the grammars under consideration, FOOTBRANCH is undominated, and requires that feet branch at some level (syllable or mora); it rules out monosyllabic, monomoraic candidates like (50c). ALLFEETRIGHT disallows any foot which is not aligned to the right edge of the prosodic word (50b). Finally, ALIGNSYLL must be violated in any prosodic word of more than two syllables if all segments in the input are to be faithfully parsed (50a).

(52) Rotuman Prosodic Constraint Hierarchy (*seseva* 'erroneous')

/seseva/	FOOTBRANCH	ALLFEETRIGHT	ALLFEETLEFT	ALIGNSYLL
☞ a. [se$_\mu$.(sé$_\mu$.va$_\mu$)]			σ	σσ
b. [(sé$_\mu$.se$_\mu$).va$_\mu$]		σ!		σσ
c. [se$_\mu$.se$_\mu$.(vá$_\mu$)]	*!		σσ	

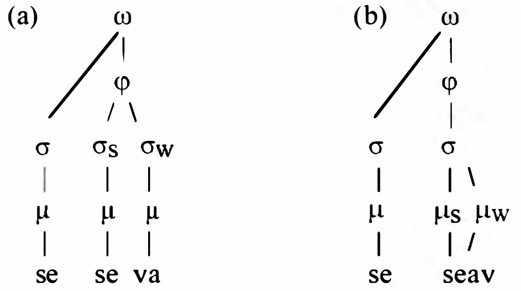
In the discourse register, ALIGNSYLL is promoted, which leads to the optimal form of the multisyllabic prosodic word in (53):

(53) Rotuman Discourse Register

/seseva/	FOOTBRANCH	ALLFEETRIGHT	ALIGNSYLL	ALLFEETLEFT
☞ a. [se$_\mu$.(séav$_{\mu\mu}$)]				σ
b. [se$_\mu$.(sé$_\mu$.va$_\mu$)]			σ!σ	σ
c. [(sé:s$_{\mu\mu}$).va$_\mu$]		σ!		
d. [se$_\mu$.se$_\mu$.(vá$_\mu$)]	*!			σσ

The optimal shape of the Rotuman foot in the citation register is given in (54a) and in the discourse register in (54b):

(54) Rotuman Prosodic Word (Citation and Discourse Registers)



2.3.2.3 Faithfulness vs. Markedness Constraints

In order to clearly understand the entire picture, the following additional faithfulness and structural markedness constraints must be taken into account:

(55) *Rotuman Faithfulness and Markedness Constraints*

(a) MAX-C: Every consonant in the input has a correspondent in the output.
(b) MAX-V: Every vowel in the input has a correspondent in the output.
(c) MAX-V[PLC]: Every vowel feature in the input has a correspondent in the output.
(d) NOCODA: Syllables do not have coda consonants.

(e) LINEARITY: S$_1$ is consistent with the precedence structure of S$_2$ and vice versa.

The Rotuman constraint ranking for the citation register must be that in (56), exemplified in (57):

(56) MAX-C, LINEARITY, NOCODA, MAX-V >> ALLFEETLEFT >> ALIGNSYLL
(57) Rotuman constraint inventory and hierarchy (Citation Register)

/seseva/	MAX-C	MAX-V	LINEARITY	NOCODA	ALLFTLEFT	ALIGNSYLL
☞ a. [se.(sé.va)]					σ	σσ
b. [se.se.(vá:)]					σ!σ	
c. [se.(séav)]			a!v	v	σ	
d. [se.(sév)]		a!		v	σ	
e. [se.(séa)]	v!				σ	

All segments are parsed faithfully, disallowing (57d) and (57e), and are in the correct linear order with no codas, throwing out (57c). The foot at the right edge of the word extends leftwards as far as possible (two syllables) dispensing with (57b), leaving (57a) the winner, which fails to align the edges of its syllables and the foot which contains them, violating ALIGNSYLL.

The Rotuman ranking for the discourse register must be that in (58), exemplified in (59):

(58) MAX-C, ALIGNSYLL >> MAX-V[PLC], ALLFEETLEFT >> LINEARITY, NOCODA

(59) Rotuman constraint inventory and hierarchy (Discourse Register)

/seseva/	MAX-C	ALIGNSY LL	MAXV[PL C]	ALLFTLEF T	LINEARIT Y	NOCOD A
☞ a. [se.(séav)]				σ	av	v
b. [se.se.(vá:)]				σ!σ		
c. [se.(sév)]			a!	σ		v
d. [se.(sé.va)]		σ!σ		σ		
e. [se.(séa)]	v!			σ		

In tableau (59), the winning candidate violates both LINEARITY and NOCODA, while parsing as many feet as possible and aligning its syllable edges with the foot which contains it, satisfying ALIGNSYLL.

Although in this case MAX-V[PLC] (which replaces MAX-V in (57) is completely satisfied, it is evident from the examples in (48) that it cannot be undominated. A more detailed analysis would require the explosion of the constraint MAX-V[PLC] into specific place features. While this constitutes an intriguing part of the overall analysis, it lies beyond the scope of this section.

2.4 Interim discussion

As in the case of Chamic, the cause for the difference between the Rotuman and Kwara'ae register distinctions lies outside the grammar. Just as in Chamic, it was brought about by the phonetic effect of lengthening stressed syllables. And once again, it resulted in the promotion of ALIGNSYLL in the formal grammar of the learner, leading to the alignment of syllable edges with foot edges. The tension between faithfulness to segmental material and the new prosodic pattern can also be observed – although the latter ultimately wins out in Rotuman and Kwara'ae whenever a conflict occurs, the former is enforced when possible – either fully (as in the case of tautosyllabification and metathesis) or partially (coalescence).

3. Continuing changes in the Chamic daughter languages

After the breakup of PC into daughter languages, the prosodic pressures which were at work on PC intensified in some cases as a result of continuing contact with MK languages; these were often fed by vocalic neutralization in sesquisyllables, resulting in the disappearance of place features in vowels which occupied these prosodically weak positions. This section selects three of the most affected languages and discusses how to understand these changes in light of the present analysis. In each case, the more affected language will be compared to a closely-related language which is less-affected.

3.1 Coastal Chamic

Further reduction to monosyllables has occurred in the Coastal Chamic languages Western Cham (WC) and Phan Rang Cham (PRC), but is more advanced in the former. This seems to be directly correlated with the fact that the first vowel in WC disyllabic forms (that is, the unstressed vowel) was neutralized to *ə*, which later underwent either deletion or lowering to *a* (see below). This process is less advanced in PRC, with a rather large degree of variation between original vowels, neutralized vowels subsequently lowered to *a*, and deleted vowels.

The largest category of words in which this has occurred are those which are vowel-initial. This reduction has occurred almost without exception in WC, and there is a high degree of variation in PRC between forms which maintain or delete the initial vowel. Examples are given where the initial consonant of the final syllable are obstruents (60a) and sonorants (60b). The one exception to deletion in WC is given in (60c):

(60)	Gloss	PC	WC	PRC
(a)	fish	ʔiká:n	kan	(i)kan
	nose	ʔidúŋ	t̤uŋ	(a)t̤ŭŋ
	root	ʔughá:r	k̤ha	(u/a)k̤ha
	dog	ʔasə́w	saw	(a)thɔ̆w
(b)	father	ʔamá:	mɯ	amɯ
	snake	ʔulár	la	(u/a)la
	person; someone	ʔurá:ŋ	raŋ	uraŋ
	blow; whistle	ʔayúp	yŭʔ	(a)yŭʔ
(c)	ghost; corpse	ʔantə́w	ataw	atɔ̆w

The second class of words in which reduction has occurred regularly is in those which began with *h* (< PC *h or *s) before obstruents (61a), with one exception in (61b). There is variation in PRC between forms with and without an initial syllable if the main syllable initial was voiced, but apparently not if it was voiceless (61c). *h*-initial syllables were retained before sonorants (6d), but there were novel *h* (and *s*) + liquid clusters which became possible (61e):

(61)	Gloss	PC	WC[27]	PRC
(a)	ashes	habə́w	p̤aw	(ha)p̤ɔ̆w
	(a)live	huɗip	t̤iwʔ	(ha)t̤ĭwʔ
	rain	hujá:n	ç̤an	(ha)ç̤an
	ant	sidə́m (> hidəm)	t̤ɔ̆m	(ha)t̤ăm
(b)	after; behind	hudə́y	hat̤ay	(ha)t̤ɛ̆y
(c)	liver	hatáy	tay	hatay
(d)	cultivated field	humá:	hamɯ	hamu
	rattan	hawáy	haway	hawɛ̆y
(e)	pestle	halə́w	hlaw	hlɔ̆w
	slave; servant	hulún	hlŭn	halŭn
	day; sun	hurə́y	hray	harɛ̆y
	write; letter	surát (> hurát)	hrăʔ	harăʔ

There was a continuing development towards stop + *h* clusters in forms with original intial voiced consonants (62a), although apparently not with initial voiceless consonants (62b):

(62)	Gloss	PC	WC	PRC
(a)	sew	jahît	ç̤hiʔ	ç̤hiʔ
	bad; wicked	jəhá:t	----	ç̤haʔ
(b)	old (people)	tuhá:	taha	taha

Finally, there was variation in WC in the case of stop + liquid clusters (no such reduction occurred in PRC). Those forms in which it occurred are given in (63a); those in which it did not in (63b). The one absolute ban on clustering is upon alveolar and palatal stops clustering with *l*:

[27] I replace the single underdot used in Thurgood (1999) which indicates breathy phonation (or vowel quality induced thereby) in WC and PRC with the IPA double underdots indicating the same.

(63)	Gloss	PC	WC	PRC
(a)	palm; sole	palá:t	plaʔ	palaʔ
	girl	dará:	t̤ra	t̤ara
	needle	jarúm	c̤rum	c̤arŭm
	skin	kulît	kliʔ	kaliʔ
(b)	swell; swollen	baráh	p̤arah	----
	rope, string	talə́y	talay	talɛ̆y
	road; path	jalá:n	c̤alan	c̤alan
	to hatch	karə́m	karɔ̆m	karăm

The exceptions above might be explainable if we make certain assumptions about the vocalic development in WC initial syllables. In the WC examples here, all initial syllable vowels are *a*. If it were the case (as mentioned above) that presyllables were first neutralized to *ə*, and then subsequently lowered to *a*, then we could posit that two distinct developments – the deletion of initial syllable *ə* and the lowering of *ə* to *a* – crosscut each other, so that *ə* might be deleted before it could lower to *a* (64a), but if it did first lower to *a*, it would resist deletion (64b):

(64) Variation in the development of unstressed vowels in WC

	Gloss	PC	V-neutralization	Deletion~ Lowering
(a)	fish	ʔiká:n	əka:n	ka:n
	skin	kulît	kəlit	kliʔ
	sew	jahît	jəhit	c̤hiʔ
(b)	ghost	ʔantə́w	ətəw	ataw
	tongue	diláh	dəlah	dalah
	old (people)	tuhá:	təha	taha

One final point of interest involves the participation of the Coastal Chamic languages in the resolution of the few remaining trisyllabic forms in PC. With only two discernible exceptions (Rade, which did not allow metathesis, and *ear* in N. Roglai), these were reduced via metathesis to the disyllabic forms *mɚriah* and *təŋia* respectively:

(65)	Gloss	PC	WC	PRC
	red	mahiráh	mareah(< məriah)	mɯryăh
	ear	təliŋá:	----	taŋi (< təŋia)

3.1.1 OT Analysis of WC grammar

The constraint inventory and ranking necessary to capture the grammar of WC is the following:

(66) MAX-C[PLC], *CC >> ALLFTLEFT >> MAX-C >> MAX-V[PLC] >> LINEARITY

In addition, the exploded ranking of *CC must be altered -- *HR and *JH[28] must be demoted, indicating a probable increase in contact with languages having these cluster types:

(67) *CC (all types except OR, OH, JH, HR) >> *OR, *OH, *JH, *HR

Tableau (68) shows the normal development of vowel-initial forms:

(68) Pre-WC əká:n → WC ká:n 'fish'

/əka:n/	MAXC[PLC]	*CC	ALLFTLEFT	MAX-C	MAXV[PLC]	LINEARITY
☞ a. [(ká:n)]						
b. [ə̆.(ká:n)]			σ!			

Since ALLFTLEFT dominates MAX-C, there is nothing enforcing the preservation of *h* in *h*-initial forms:

(69) Pre-WC hətáy → WC táy 'liver'

/hətay/	MAXC[PLC]	*CC	ALLFTLEFT	MAX-C	MAXV[PLC]	LINEARITY
☞ a. [(táy)]				h		
b. [hə̆.(táy)]			σ!			
c. [(htáy)]		h!t				

Since *JH and *HR have been demoted, (70a) and (71a) do not incur violations of *CC:

(70) Pre-WC jəhît → WC jhît (→ c̤hîʔ) 'sew'

/jəhit/	MAXC[PLC]	*CC	ALLFTLEFT	MAX-C	MAXV[PLC]	LINEARITY
☞ a. [(jhît)]						
b. [(jît)]				h!		
c. [jə̆.(hît)]			σ!			
b. [(hît)]	j!			j		

[28] J = palatal obstruent.

(71) Pre-WC hələ́w → WC hlə́w (→ hláw) 'pestle'

/hələw/	MAXC[PLC]	*CC	ALLFTLEFT	MAX-C	MAXV[PLC]	LINEARITY
☛ a. [(hlə́w)]						
b. [(lə́w)]				h!		
c. [hə̆.(lə́w)]			σ!			

CR clusters (71) also remain licit, and more became possible because of unstressed vowel neutralization:

(72) Pre-WC bəlá:n → WC blán (→ plán) 'moon'

/bəla:n/	MAXC[PLC]	*CC	ALLFTLEFT	MAX-C	MAXV[PLC]	LINEARITY
☛ a. [(blán)]						
b. [(lán)]				h!		
c. [bə̆.(lán)]			σ!			

Finally, the tension resulting from a trisyllabic form such as *red* can be observed in tableau (73). Even the winning candidate (73a) incurs three different kinds of violation. ALLFEETLEFT, outranking LINEARITY, forces a reduction in the word by one-syllable, but is unable to completely align the left edge of the foot and prosodic word due to the preservation of *m* through MAXC[PLC].

(73) Pre-WC məhiráh → WC məŕiah (→ maréah) 'red'

/məhirah/	MAXC[PLC]	*CC	ALLFTLEFT	MAX-C	MAXV[PLC]	LINEARITY
☛ a. [mə̆.(ríah)]			σ	h		ir
b. [mə̆.(ráh)]			σ	h	i!	
c. [mə̆.hĭ.(ráh)]			σ!σ			
d. [(mríah)]		m!r		h		ir
e. [(ríah)]	m!			mh		ir

3.2 Highlands Chamic

Rade and Jarai are closely related members of one of the two Highlands Chamic subgroups. While both have undergone additional reduction towards monosyllables, Rade has in general progressed much further than Jarai. Rade generally preserved initial consonants including *h*, but allowed several new cluster types at the cost of deletion of the first vowel. Like Western Cham, unstressed vowel neutralization has occurred in both Rade and Jarai, with later developments occurring in Rade conditioned by the preceding consonant if one existed.

The first set of forms in which this has occurred are stop + liquid clusters. There has been uniform reduction in Rade in voiceless stop clusters, and in most cases in Jarai (74a). Reduction has also occurred in voiced bilabial stop + liquid clusters (74b) as well as bilabial stop + glide sequences (74c):

(74)	Gloss	PC	Rade	Jarai
(a)	palm; sole	palá:t	plăʔ	----
	rope, string	taləy	klɛy	tə̆ləy
	dig	kaləy	klɛy	klə̆y
	to hatch	karə́m	krăm	krə̆m
(b)	moon; month	bulá:n	mlan	blan
	hair; feathers	buləw	mlăw	bləw
	shoulder	bará:	mra	bra
	night; evening	malám	mlam	mlăm
(c)	crocodile	buyá:	mya	bĭa

The second group of words in which reduction has occurred is in those with *h* (< PC *h or *s) + liquid sequences (75a); this has also occurred in Jarai. The same is true in clusters of h (< PC *r) + *l* sequences in Rade, but not in Jarai (75b). Finally, this process has occurred in the Rade word *red* as a result of the irregular deletion of the initial syllable, although the more normal Chamic development (*mahirah* > *məriah*) has occurred in Jarai (75c):

(75)	Gloss	PC	Rade	Jarai
(a)	pestle	halə́w	hlăw	hləw
	worm	hulát	hluăt	hlăt
	day; sun	hurə́y	hrue	hrəy
	salt	sirá: (> hirá:)	hra	hra
(b)	grass, thatch	ralá:ŋ (> halá:ŋ)	hlaŋ	həlăŋ
	candle wax	ralín (> halín)	hlin	hə̆lin
(c)	red	mahiráh (> hiráh)	hrah	mriăh

As in Coastal Chamic, clusters have formed from stop + *h* sequences (including palatal stops):

(76)	Gloss	PC	Rade	Jarai
	otter	buháy	kəmhe[29]	pə̆hay
	sew	jahít	jhĭt	(cɛt)
	bad; wicked	jəhá:t	jhat	săt
	old (people)	tuhá:	khua	tha

Vowel-intial words have generally been reduced (77a), with some exceptions similar to WC (77b):

[29] The initial syllable in this word is unexpected.

(77)	Gloss	PC	Rade	Jarai
(a)	fire	ʔapúy	puy	puy
	far; above; long	ʔatá:s	tayh	atayh
	fish	ʔiká:n	kan	akan
	sharpen	ʔasáh	sah	ăsah
(b)	ghost; corpse	ʔantə́w	atăw	----
	flesh; meat	ʔasə́y	asɛy	(ăsar)
	dog	ʔasə́w	asăw	asəw

The following forms have unexpected deletion of their initial syllables (expected to be retatined because they have intial consonants). They all have voiceless main-syllable intials, which were preceded by either original voiced stops (78a), or initial *h* (78b):

(78)	Gloss	PC	Rade	Jarai
(a)	stone	batə́w	tăw	pətəw
	calf (leg)	bətíh	tih	pə̆tih
	armspan	dəpá:	pa	tə̆pa
(b)	liver	hatáy	tey	hə̆tay

Finally, the common development of *ear*, from PC *təliŋa:* to Jarai *təŋia* can be observed; in Rade, on the other hand, metathesis involving *i* does not seem to have occurred (the situation with *u* is quite different, and will not be treated here; see below for further description of the difference between *i* and *u* metathesis):

(79)	Gloss	PC	Rade	Jarai
	ear	təliŋá:	kəŋa	tə̆ŋia

3.2.1 OT analysis of Rade grammar
The constraint inventory and hierarchy necessary for Rade is the following:

(80) *CC, LINEARITY >> ALLFEETLEFT >> MAX-C >> MAX-V[PLC]

As in WC, a word-intial *ə* is generally expendable:

(81) Pre-Rade əká:n → Rade ká:n 'fish'

/əka:n/	*CC	LINEARITY	ALLFTLEFT	MAX-C	MAX-V[PLC]
☞ a. [(ká:n)]					
b. [ə̆.(ká:n)]			σ!		

New JH and HR clusters are also licit:

(82) Pre-Rade jəhît → Rade jhĭt 'sew'

/jəhit/	*CC	LINEARITY	ALLFTLEFT	MAX-C	MAX-V[PLC]
☞ a. [(jhît)]					
b. [(jît)]				h!	
c. [jə̆.(hît)]			σ!		

(83) Pre-Rade hələ́w → Rade hláw 'pestle'

/hələw/	*CC	LINEARITY	ALLFTLEFT	MAX-C	MAX-V[PLC]
☞ a. [(hláw)]					
b. [(láw)]				h!	
c. [hə̆.(láw)]			σ!		

Again, thanks to unstressed vowel neutralization, CR clusters continue to form:

(84) Pre-Rade bəlá:n → Rade blán (→ mlán) 'moon'

/bəla:n/	*CC	LINEARITY	ALLFTLEFT	MAX-C	MAX-V[PLC]
☞ a [(blán)]					
b. [(lán)]				b!	
c. [bə̆.(lán)]			σ!		

Finally, the interaction between every constraint in the present hierarchy can be observed in (85):

(85) Pre-Rade təliŋá: → Rade təŋá: (→ kəŋá:) 'red'

/təliŋa/	*CC	LINEARITY	ALLFTLEFT	MAX-C	MAX-V[PLC]
☞ a. [tə̆.(ŋá:)]			σ	l	i!
b. [tə̆.lĭ.(ŋá:)]			σ!σ		
c. [tə̆.(ŋîa)]		i!ŋ	σ	l	
d. [(ŋîa)]		i!ŋ		tl	
e. [(tŋîa)]	t!ŋ	iŋ		t	

3.3 Northern Cham

Tsat and Northern Roglai form the Northern Cham subgroup within the other major branch of Highlands Chamic. While N. Roglai has continued the prosodic structure of PC relatively unchanged, Tsat has gone through a fundamental evolution. The reason for this is that at some point, speakers of Tsat left the mainland and moved to Hainan island, where monosyllabic, fully tonal languages (Chinese, Hlai, Lingao and Mien) are the norm.

With one class of exceptions to be discussed below, there seems to have been a rapid reduction of the prosodic word in which unstressed syllables were simply deleted. Examples are given in (86) of words with new initials resulting from this wholesale loss, including voiceless stops (86a), voiced stops (which later devoiced to aspirated stops) (86b), fricatives (86c), nasals (86d), and liquids/glides (86e):

(86)	Gloss	PC	Tsat	N. Roglai
(a)	thick	kapá:l	pa:n^{11}	kapan
	hundred	ratús	tu^{33}	ratuh
	sick, painful	sakít	kiʔ^{24}	saki:ʔ
	play	maʔín	ʔin^{33}	maʔin
(b)	sugarcane	təbə́w	phə^{11}	tubəw
	unripe, young	mudá:	tha^{11}	mida
	rain	hujá:n	sa:n^{11}	huja:t
	tooth	gigə́y	khay^{11}	digəy
(c)	iron	basə́y	say^{11}	pisəy
	dog	ʔasə́w	saw^{33}	asəw
	old (people)	tuhá:	ha^{33}	tuha
	thirst, desire	mahãw	hawʔ^{24}	mahāw
(d)	we (excl)	kamí	mi^{33}	kamīn
	pus	lanáh	na^{55}	lanăh
	oil	miɲá:k	ɲaʔ^{24}	maɲa:ʔ
	the wind	ʔaŋín	ŋin^{33}	aŋīn
(e)	bone	tulá:ŋ	la:ŋ^{33}	tula:k
	red	mahiráh	za^{55}	mariah
	ginger	liyá:	za^{33}	riya
	rattan	hawáy	va:yʔ^{42}	haway

The only instances where original word-initial consonants were preserved were those in which licit clusters could be formed as a result of reduction. These included stop + liquid (87a-b) and stop + *h* clusters (87c) (note that in (87a-b), the liquids *l* and *r* vocalized to y). It is an open question whether this reduction occurred in Tsat before or after the arrival of its speakers on Hainan:

(87)	Gloss	PC		Tsat	N. Roglai
(a)	village	palə́y	(> plə́y)	piay^{33}	paləy
	palm; sole	palá:t	(> plá:t)	pieʔ^{24}	pala:ʔ
	moon; month	bulá:n	(> blá:n)	phian^{11}	bila:t
	hair; feathers	bulə́w	(> blə́w)	phiə^{11}	biləw
(b)	shoulder	bará:	(> brá:)	phia^{11}	bara

	new; just now	bahrə́w(> brə́w)	phiə^{11}	bahrəw
	blood	daráh (> dráh)	sia^{55}	darah
	needle	jarúm (> jrúm)	sun^{11}	jurup
(c)	sew	jahít (> jhít)	siʔ^{24}	chi:ʔ
	bad; wicked	jəhá:t (> jhá:t)	sa:ʔ^{24}	----

Although there does not seem any way to know for sure, it is likely that Tsat unstressed vowels went through a stage of neutralization before the changes described above ultimately occurred:

(88)	Gloss	PC	V-neutralization	Deletion/Clustering
	flower	bŭŋá:	bə̆ŋá:	ŋá:^{11}
	moon	bŭlá:n	bə̆lá:n	blá:n^{11} → phían^{11}

3.3.1 OT analysis of Tsat grammar

In the emergent Tsat grammar, prosodic and markedness constraints absolutely dominated faithfulness constraints:

(89) ALIGNSYLL, ALLFTLEFT, *CC >> MAX-C[PLC] >> MAX-V[PLC]

Normally, no trace of the initial syllable survived (90), unless the consonants could form a licit cluster (91):

(90) PC buŋá: → Tsat ŋá:^{11} 'flower'

/buŋa:/	ALIGNSYLL	ALLFTLEFT	*CC	MAX-C[PLC]	MAX-V[PLC]
☞ a. [(ŋá:)]				b	u
b. [(bŋá:)]			b!ŋ		u
c. [bŭ.(ŋá:)]		σ!			
d. [(bu.ŋá:)]	σ!				

(91) PC bulá:n → Tsat blá:n (→ b̤já:n^{11} → phían^{11}) 'moon'

/bula:n/	ALIGNSYLL	ALLFTLEFT	*CC	MAX-C[PLC]	MAX-V[PLC]
☞ a. [(blá:n)]					u
b. [(lá:n)]				b!	u
c. [bŭ.(lá:n)]		σ!			
d. [(bu.lá:n)]	σ!				

3.4 Hlai: another example of complete prosodic alignment

Before concluding this section, I wish to examine one more typologically relevant case outside of Chamic which will reenforce the data and the concepts underlying them, discussed above.

Chamic is not the only family in which there has been a shift in prosodic rhythm with consequences for the segmental structure of the language. It has been recognized for some time now that there is some kind of relationship between Austronesian and Kra-Dai (also known as Tai-Kadai). While the question of whether this relationship is genetic or one of contact remains as yet unresolved, this is unimportant for the purposes of the present discussion.

What is important is that there are cognates between the two languages which provide evidence that Proto-Kra-Dai, or its immediate daughters, shifted to a word-final stress pattern which led to reduction of unparsed syllables at the left-edge of the phonological word, and ultimately to an absolute alignment of prosodic categories (syllable with foot with prosodic word).

Since Proto-Kra-Dai itself has not yet been fully reconstructed, I provide examples here from one of its daughters, Proto-Hlai (Norquest (in preparation)[30]), the modern languages of which are spoken in Hainan, China (also the home of Tsat speakers, as mentioned above).

Some of the better comparisons between Proto-Austronesian[31] (PAn) and Proto-Hlai (PH) are given in (92). In the table below, the Hlai forms on the left side either show no evidence of a former presyllable, or the evidence is ambiguous; forms on the right side definitely had an original presyllable. The PH forms below are organized by initial and medial stops (92a), affricates (92b), nasals and laterals (92c), and rhotics and approximants (< intervocalic voiced stops) (92d):

(92) PAn-PH cognates

	Gloss	PAn	PH[32]	Gloss	PAn	PH
(a)	ancestor	a(m)pu	páw^{C}	hut; village	lə+paw	C-báw^{C}
	seven	pitu	tə́w	black	qitəm	C-dám
	rib	tak(ə)ʀaŋ	ká:ŋ^{C}	crocodile	buqaya	C-gáy^{C}
(b)	fire	Sapuy	pfə́y	tooth	nipən	C-pfján
	eye	maCa	ʈʂá:	head louse	kuCuh	C-ʈʂwú:
(c)	five	lima	má:	you	kamu	C-mə́ɰ
	six	ənəm	nóm	this	(qa-)ni[H]	C-nə́y^{B}
	child	ałak	lɯ́:k	fish scales	quSəłap	C-lʌ́:p
(d)	buy, sell	saliw	ŕi:w^{C}	shoulder	qabaʀaH	C-βá:^{B}
	eight	walu	rə́w	shrimp	qudaŋ	Cɾwá:ŋ
	to plant	mula	rwá:	thigh	paqaS	C-ɣá:

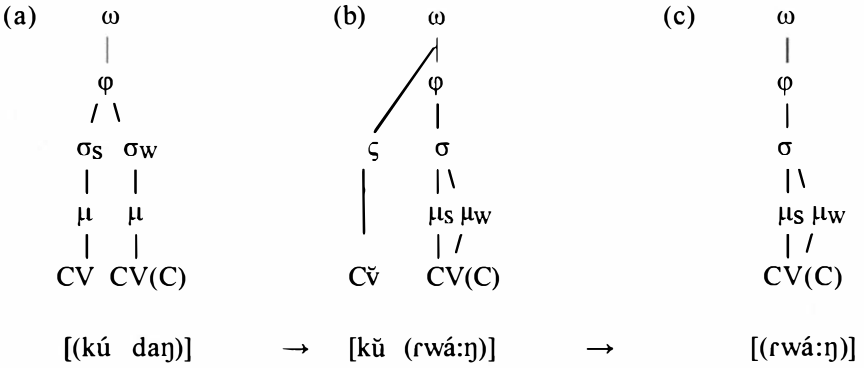
[30] The Proto-Hlai reconstruction offered here is a work in progress; although the parts of the reconstruction relevant to this paper are secure, it should be understood that some details may change.

[31] PAn forms are primarily from Zorc (1995).

[32] The superscripted letters B and C on PH forms refer to tone categories.

The same general phenomenon occurred in Hlai as in Chamic, with an original trochaic PAn form (93a) undergoing stress-shift to a sesquisyllable form (93b) and ultimately reduction to a single syllable, as in Tsat (93c):

(93) The evolution of the Hlai prosodic word



3.4.1 OT analysis of Hlai prosodic evolution
The OT analysis below will be divided into grammars which model each of the stages in (93) above.

3.4.1.1 Pre-Hlai with Penultimate Stress
The initial stage of Pre-Hlai grammar can tentatively be borrowed from that for PMP in (10) above, as it is assumed that words at the earliest stage of Pre-Hlai bore the same trochaic stress pattern with the PAn words with which they are cognate.
(94) Pre-Hlai Foot Structure with Penultimate Stress (*qudaŋ* 'shrimp')

/kudaŋ/	ALLFEETRIGHT	FOOTBRANCH	ALLFEETLEFT	ALIGNSYLL
☛ a. [(kú$_{\mu}$.daŋ$_{\mu}$)]				σ
b. [kŭ.(dáŋ$_{\mu\mu}$)]			σ!	
c. [(ku$_{\mu}$.dáŋ$_{\mu}$)]		W!S		σ
d. [(kú$_{\mu}$).daŋ]	σ!	*		

3.4.1.2 The Pre-Hlai shift to word-final stress
Although the details are far from clear, it seems a reasonable assumption that pre-Hlai (or perhaps even Proto-Kra-Dai itself) participated in some language area which included word-final stress as one of its salient aspects. Like Chamic, this led to the instantiation of word-final stress in the native Hlai lexicon.

One of the differences between the PC and Pre-Hlai forms, as shown with the form 'shrimp' below, is that high vowels from the first syllable underwent metathesis under certain conditions, being reanalyzed as a coarticulation on the following consonant. This is reminiscent of the metathesis which occurred regularly in Rotuman and Kwara'ae at the right edge of the foot, and in fact also occurred sporadically in the Chamic daughter languages (sometimes only in one language).

In the Chamic daughter languages, there was a fundamental difference between *i* and *u* metathesis. The former seems to have been completely regular, was restricted to trisyllabic forms and occurred in all languages except Rade and Tsat (95a). The latter, by comparison, seems to have been much more sporadic, and could occur in either disyllabic or trisyllabic forms, with no consistent pattern in any language (95b). In the following table, metathesized forms are shown in bold type:

(95) High vowel metathesis in the Chamic daughter languages

English	PC	Rade	Jarai	Chru	N. Roglai	Tsat	Haroi	WC	PRC
(a) *i*-metathesis									
red	mahiráh	hrah	**mriăh**	**məriah**	**mariah**	za^{55}	**məreah**	**mareah**	**mɯryăh**
ear	təliŋá:	kəŋa	**tə̆ŋia**	**tənia**	**riŋiā**	ŋa^{33}	**cəŋea**	----	**taŋi**
(b) *u*-metathesis									
yesterday	[m/t]uburə́y	**məbrue**	----	**kəbruəy**	tubrəy	----	**məcəpruy**	**maprɔy**	**paproy**
thorn	durə́y	**erue**	drə̆y	**druəy**	**daruəy**	----	cərɯ̆y	**ṭaruay**	**ṭaroy**
turtle, tortoise	kurá:	**krua**	**krŏa**	kra	kura	----	**kroa**	----	kara
taro; yam	hubə́y	həbɛy	hə̆bə̆y	həbəy	**habuəy**	phay^{11}	**aphuy**	p̣ay	(ha)p̣ɛ̆y
cultivated field	hɯná:	həma	**həm(u)a**	həma	humā	ma^{33}	həmɯa	hamɯ	hamu
day; sun	hurə́y	**hrue**	hrəy	hərəy	hurəy	zay^{33}	hərɯy	hray	harɛ̆y
old (people)	tuhá:	**khua**	tha	tha	tuha	ha^{33}	cəha	taha	taha
vein; tendon	ʔurát	**aruăt**	arăt	araʔ	uraʔ	zaʔ^{24}	arăʔ	răʔ	(u/a)răʔ
worm	hulát	**hluăt**	hlăt	həlaʔ	----	----	----	hlăʔ	halăʔ

Both *i* and *u* metathesis occurred in Hlai as well. The coarticulation in Hlai most likely occurred at first as a purely phonetic effect, and was phonologized by language learners; it is as yet unclear exactly how regular Pre-Hlai metathesis was, but when it occurred it seems to have done so as in (96):

(96)	Original form	Uttered as		Phonologized as
	/kuɾáŋ/ →	[kŭɾwáŋ]	→	/kuɾwáŋ/

This perhaps should most properly be considered a production violation of LINEARITY in the grammar of the speaker as it likely began as a purely phonetic deviation from the underlying representation; it could not be a violation in the grammar of the listener, since the listener's underlying representation was constructed faithfully according to what was perceived. I therefore do not consider LINEARITY in the tableau below.

To capture the shift to Pre-Hlai, it is again necessary to posit a grammar where ALIGNSYLL dominates ALLFEETLEFT:

(97) Pre-Hlai Foot Structure with Final Stress (*kŭɾwáŋ* ‘shrimp’)

/kuɾwaŋ/	ALLFEETRIGHT	FOOTBRANCH	ALIGNSYLL	ALLFEETLEFT
☛ a. [kŭ.(ɾwáŋ$_{\mu\mu}$)]				σ
b. [(kú$_{\mu}$.ɾwaŋ$_{\mu}$)]			σ!	
c. [(ku$_{\mu}$.ɾwáŋ$_{\mu}$)]		W!S	σ	
d. [(kú$_{\mu}$).ɾwaŋ$_{\mu}$]	σ!	!		σ

3.4.1.3 The Pre-Hlai shift to monosyllables

Finally, at a stage of reduction parallel with Tsat, a grammar must be constructed where ALLFEETLEFT is moved into a position, along with *CC, above MAX-C and MAX-V, so that prosodic words consist of single heavy syllables which are necessary if a strong-weak binary pattern (as enforced by FOOTBRANCH) is to be maintained within the foot. These heavy monosyllables are the result of complete alignment between syllable, foot, and prosodic word:

(98) Reduction to Proto-Hlai monosyllables

/kuɾwáŋ/	*CC	ALIGNSYLL	ALLFEETLEFT	MAX-V	MAX-C
☛ a. [(ɾwáŋ)]				u	k!
b. [kŭ.(ɾwáŋ)]			σ!		
c. [(ku.ɾwáŋ)]		σ!			
d. [(kɾáŋ)]	k!ɾ			u	

5 Conclusion

The unifying theme in this paper has been that prosodic change tends to proceed in such a way that (1) the edges of prosodic categories become aligned, (2) unparsed segmental material is vulnerable to loss, and (3) the degree of vulnerability of a segment depends on its featural specification. This has been shown in the following cases:

(99)	Language	Example			Categories aligned
(a)	PMP → PMC	[qa.(tə́.luR)]	→	[(tə́.lur)]	Foot, L, PrWd, L
(b)	PMC → PC	[(bə́.ras)]	→	[(brá:s)]	Syllable, Foot
(c)	Rotuman CF → DF	[se.(sé.va)]	→	[se.(séav)]	Syllable, Foot
(d)	PC → Tsat	[bŭ(ŋá:)]	→	[(ŋá:^{11})]	Foot, L, PrWd, L
(c)	Pre-Hlai → Proto-Hlai	[kŭ.(ɾwáŋ)]	→	[(ɾwá:ŋ)]	Foot, L, PrWd, L

(99d-e) represent fully optimal prosodic alignment, since the edges of all prosodic categories are in alignment with each other at both ends of the word: syllable with foot, and foot with prosodic word.

These changes have all come about as a result of changes which occurred outside of the formal grammar. In the case of Chamic (and most likely Hlai), there was a shift to word-final prosody induced by contact with other languages which already possessed this word-final prosody. In Chamic, Hlai, Rotuman, and Kwara’ae, a common phenomenon occurred where stressed syllables become phonetically longer; this prompted a reanalysis of the weight of those syllables.

In all cases, unparsed segmental material was lost, either at the left edge or the prosodic word (PMC, PC, the Chamic daughter languages, and Hlai) or the right edge (Rotuman, Kwara'ae). In Chamic, the class of segments which were the most susceptible to loss were those segments which lack oral place features: the laryngeal consonants *ʔ* and *h*, and the vowel *ə*. This loss can be considered to be due at least partly to perceptual difficulties outside of the grammar, but the fact that these segments were retained when parsed into prosodic structure indicates that there is another part of the explanation which involved the formal grammar, with segmental material parsed into feet being privileged in comparison with unparsed material.

The constraint family which is crucial to an OT analysis of these changes is the following set of prosodic alignment constraints:

(100) ALLFEETLEFT: Align (φ, Left, ω, Left)
ALLFEETRIGHT: Align (φ, Right, ω, Right)
ALIGNSYLLABLE: Align (σ, Left, φ, Left; σ, Right, φ, Right)

These alignment constraints ensure that the edges of nested prosodic categories be aligned; in concert with FTBRANCH(S-W), they militate for a heavy monosyllabic trochee as the optimal prosodic word. When in tension with faithfulness and markedness constraints, however, they result in a mixed lexicon of forms meeting these prosodic requirements with various degrees of success, ranging from those words which conform perfectly, to longer words with segments that cannot be deleted or phonotactic constraints which cannot be overcome, which are ultimately larger than the optimal prosodic word.

References

Adelaar, K. A. 1992. *Proto-Malayic, the reconstruction of its phonology and parts of its lexicon and morphology*. Pacific Linguistics, Series C-119. Canberra: The Australian National University.

Besnier, Niko. 1987. 'An Autosegmental Approach to Metathesis in Rotuman.' *Lingua* 104: 147-186.

Blevins, Juliette. 1994. 'The Bimoraic Foot in Rotuman Phonology and Morphology.' *Oceanic Linguistics* 33, 2: 491-516.

Blevins, Juliette & Andrew Garrett. 1998. 'The Origins of Consonant-Vowel Metathesis.' *Language* 74, 3: 508-556.

Blust, Robert. 2000. Review of Thurgood (1999). *Oceanic Linguistics*, 39:2. 435-45.

Cho, Young-mee Yu & Tracy Halloway King. 2003. 'Semisyllables and Universal Syllabification'. In Féry & van de Vijver, eds. *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press. 183-212.

Churchward. C. M. 1940. *Rotuman Grammar and Dictionary*. Sydney: Methodist Church of Australia.

Dikken, Marcel den. *The structure of the noun phrase in Rotuman*. Ms. (to be published by Lincom Europa).

Hale, Mark & Madelyn Kissock. 1998. 'The Phonology-Syntax Interface in Rotuman.' *UCLA Occasional Papers in Linguistics 21: 'Recent Papers in Austronesian Linguistics: Proceedings of the third and fourth meetings of the Austronesian Formal Linguistics Association,'* Matthew Pearson (ed.). Los Angeles 1996-97.

Hale, Mark, Madelyn Kissock & Charles Reiss. 1997. 'Output-Output Correspondence in Optimality Theory.' In *WCCFL* 16, Emily Curtis, James Lyle & Gabriel Webster (eds.). 223-236.

Heinz, Jeffrey. 2005. *CV metathesis in Kwara'ae.* M.A.. thesis: UCLA.

Matisoff, James A. 1988. Proto-Hlai initials and tones: a first approximation. In *Comparative Kadai: Linguistic studies beyond Tai.* Edited by Jerold A. Edmondson and David B. Solnit. Summer Institute of Linguistics and The University of Texas at Arlington Publications in Linguistics No. 86. 289-321.

McCarthy, John & Alan Prince. 'Prosodic Morphology I: Constraint Interaction and Satisfaction.' Ms., University of Massachusetts, Amherst, & Rutgers University, New Brunswick, N.J.

Norquest, Peter. 2001. *The Collapse of the Foot in Oceanic.* Presented at WECOL 2001: University of Washington, Seattle, 10/28/01.

Norquest, Peter. 2002. *Rotuman, Kwara'ae, and Diachrony.* Presented at the LSA 2002: San Francisco, 1/5/02.

Norquest, Peter. 2003. *From Multisyllabic to Sesquisyllabic in Malayo-Chamic.* Presented at the LSA 2003: Atlanta, 1/3/03.

Norquest, Peter. Forthcoming. *A reconstruction of the Proto-Hlai phonological system.* Ph.D. dissertation: University of Arizona.

Ostapirat, Weera. 2004. Proto-Hlai Sound System and Lexicons. In *Studies on Sino-Tibetan Languages: Papers in Honor of Professor Hwang-cherng Gong on His Seventieth Birthday.* Edited by Ying-chin Lin, Fang-min Hsu, Chun-chih Lee, Jackson T.-S. Sun, Hsiu-fang Yang, and Dah-an Ho. Institute of Linguistics. Academia Sinica, Taipei, Taiwan. 121-175.

Ouyang, Jueya & Zheng Yiqing. 1983. *Liyu diaocha yanjiu* [*A survey of the Li languages*]. Beijing: Zhongguo Shehui Kexue Chubanshe: Xinhua shudian Beijing faxingsuo faxing.

Pittayaporn, Pittayawat. 'Moken as a Mainland Southeast Asian Language'. This volume.

Pittayaporn, Pittayawat. In preparation. 'The Kammu Minor Syllable: Syllable Structure, Tone, and Infixation'.

Prince, Alan & Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar.* Ms., Rutgers University, New Brunswick and University of Colorado, Boulder.

Sohn, Ho-Min. 1980. 'Metathesis in Kwara'ae'. *Lingua* 52: 305-323.

Thurgood, Graham. 1991. Proto-Hlai (Li): A look at the initials, tones, and finals. *Kadai: Discussions in Kadai and SE Asian Linguistics* III: 1-49.

Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change.* Oceanic Linguistics Special Publication No. 28. Honolulu: University of Hawai'i Press.

Ussishkin, Adam. 2000. *The Emergence of Fixed Prosody.* Ph.D. dissert., UC Santa Cruz.

Watson-Gegeo, Karen Ann & David W. Gegeo. 1986. 'Calling-out and repeating routines in Kwara'ae children's language socialization.' In *Language Socialization across Cultures*, Bambi B. Schiefflin & Elinor Ochs (eds.), 17-50. Cambridge: CUP.

Zorc, David. 1995. 'A glossary of Austronesian reconstructions.' In *Comparative Austronesian Dictionary: An introduction to Austronesian studies*, Darrell T. Tryon (ed.). Part 1: Fascicle 2. Mouton de Gruyter. 1105-1197.

Norquest, P. "Word structure in Chamic: prosodic alignment versus segmental faithfulness". In Grant, A. and Sidwell, P. editors, *Chamic and Beyond: Studies in mainland Austronesian languages.*
PL-569:147-188. Pacific Linguistics, The Australian National University, 2005. DOI:*10.15144/PL-569.147*

5 *Moken as a Mainland Southeast Asian Language*

Pittayawat Pittayaporn

1. Introduction

The Moken are one of the three sea-oriented groups scattered in the Andaman Sea of Southern Thailand. Their languages are now spoken from southern Burma on the Mergui Archipelago to the west coast of Southern Thailand to the Malaysian border. These 'sea people' are known as Urak Lawoi, Moklen, and Moken in the literature. Nowadays the Moken life is still very sea-oriented, but the Moklen have settled on the mainland and become land-based agriculturalists (Larish 1999). It is clear that Moklen and Moken are closely related while Urak Lawoi does not belong to the same group. However, the history of this Moken is still as much a mystery as the history of this Austronesian (AN)-speaking people themselves. For the classification of Moken within Austronesian, see Blust (1992) and Larish (1999).

The Moken language shows phonological characteristics strikingly similar to other mainland SEA languages but absent from insular AN. While previous researchers (particularly Larish 1999) have recognized the importance of Mainland SEA languages, especially Mon-Khmer influence in the diachronic development of Moken, in many cases the exact processes by which these developments took place have not been systematically explored. Larish explicitly says that Moken "may have adopted word-final stress under MK influence, and this single change could have served as a catalyst for a complete typological shift (Larish 1999:381)."

I argue that attributing the mainland features found in Moken to Mon-Khmer influence is too hasty and that the stress shift is not necessarily responsible for the drastic typological shift. This paper will focus on (1) how some salient characteristics that are Mainland SEA features developed from the original Austronesian system, and (2) whether they can simply be attributed to Mon-Khmer influence. After outlining the phonology of the Moken, I will discuss the Mainland features in Moken, which are divided into (1) loan-induced features, (2) features resulted from internal restructuring and (3) features that may reflect earlier PAn stress. To conclude, I will attempt to characterize the contact situations within the framework presented by Ross (2003), which is, in turn, built on Thomason and Kaufman (1988).

2. Phonology of Moken

Different dialects of the Moken-Moklen group are now scattered along the Andaman SEA coast of Thailand and Myanmar. Although these varieties have not been extensively studied, a few phonological descriptions of different dialects are available (Chantankomes 1980; Larish 1999; Makboon 1981; Naw Say Bay 1995; Swastham 1982, Lewis 1960).

Pittayaporn, P. "Moken as a mainland Southeast Asian language". In Grant, A. and Sidwell, P. editors, *Chamic and Beyond: Studies in mainland Austronesian languages.*
PL-569:189-210. Pacific Linguistics, The Australian National University, 2005. DOI:*10.15144/PL-569.189*

The variety analyzed in this paper is that of Rawai Beach, Phuket, Thailand. The data come from two sources. The first is an excellent description by Chantanakomes (1980), who carefully describes the sound system of the language, as well as its grammatical characteristics. The second is data from the fieldwork conducted by John Wolff and myself during July 14-24, 2004 in Rawai District, Amphur Muang, Phuket Province, Thailand. Since the preliminary analysis of the sound system based on data from our fieldwork agrees with that described by Chantanakomes (1980), forms from both sources have been used to cross-check with each other.

Labial	Alveolar	Palatal	Dorsal
p	t	c	k
ph	th	ch	kh
b	d	ɟ	g
	s		h
m	n	r	ɲ
w	l	j	

Figure 1: *The consonant inventory of Moken*

Unlike Chantanakomes, our analysis does not posit *ʔ* as a phoneme because its distribution is predictable. It is an epenthetic consonant that is inserted initially in words beginning with a vowel, and medially to break hiatus. According to Chantanakomes, final *ʔ* occurs only after short vowel, suggesting that vowel length is neutralized in this environment. Since in our data Vʔ and V: alternate freely, all instances of Vʔ are analyzed and transcribed as V:, e.g. *mata:* 'eye' is pronounced as *mataʔ* or *mata:*. This is consistent with Larish (1999)'s description of 'long' vowels followed by *ʔ* as being half-long.

i, i:		u, u:
e, e:	ə	o, o:
ɛ, ɛ:	a, a:	ɔ, ɔ:
iə		uə[1]

Figure 2: *The vowel system of Moken*

It is important to note that a major difference between Chantanakomes's and our analysis is the nature of the vowel quality distinction. Chantanakomes analyzes Rawai Moken

[1] Veena posits a contrastive long *uə:* but explains that it has been found only in two words. One of these, *buə:k* 'fruit' are recorded as *buwa:k* in our data, therefore the proposed phoneme does not exist in our analysis.

(henceforth Moken) as having contrasts between high lax, high tense, and mid vowels, while in our analysis the language has a three-height distinction. That the so-called "tense high vowels"[2] pattern with low vowels in vowel harmony suggests that they are non-high. In this paper, data taken from Chantanakomes are re-transcribed to conform to our system. Like other mainland languages, Moken has a strictly iambic word template. In other word, the canonical shape of Moken words is disyllabic, with a stressed second syllable: CVCV́(:)(C). Although monosyllabic forms are found, they are rare and are mostly restricted to function words. In casual speech, however, the first syllable is often dropped, leaving the root monosyllabic. Interestingly, most cases where the first syllable is dropped are verbs, i.e. *dɔt~mədɔt* 'to cook', and *jaj~mijaj* 'to think (that)'. This phenomenon is also reported in Larish (1999).

Chantanakomes (1980) divides syllables into three types; pre-syllable, minor syllable, and major syllable. The major syllable takes the primary stress and always occupies the right edge of the word. It is the head of the word, its presence is obligatory. This syllable can be either open or closed. The minor syllable always receives secondary stress and always precedes the major syllable. This type of syllable is optional. The last class of syllable is the pre-syllable, whose difference from the minor syllable lies in the vowel quality, the stress, and the possible vowel that can occur in this position. Specifically, this type of syllable has a very weak, short and neutralized vowel. Both the pre-syllable and the minor syllable are invariably of CV shape.

To determine whether the distinction between the presyllable and the minor syllable is phonologically supported[3], a careful study of their phonological behavior is needed. That the short unstressed syllables with *ə* is treated as a separate category seem to come from the practice of Mon-Khmer describing roots as consisting of syllable and a half (Matisoff, 1973). This practice is well accepted but the precise nature of the phonological distinction is still unclear. Therefore, in this paper the term 'minor syllable' and 'major syllable' will be used to refer to the position of the syllable within the word without making any phonological claim about whether there is a distinction between what Larish (1999) and Chantanalomes (1980) call "presyllable" and "minor syllable". Specifically, in this paper "minor syllable" refers to the unstressed first syllable and "major syllable" refers the stressed second syllable in disyllabic words.

Although it has been claimed that the degree of mutual intelligibility between speakers of different dialects is low (Larish), Moklen-Moken dialects comprise a substantially homogenous group. They share a huge number of lexical entries, most of which are almost identical in form. They also have similar sound inventories and phoneme distributions. I assume that Proto-Moken-Moklen must have been similar enough to modern dialects not to affect the analysis of specific cases. Therefore, it is reasonable to trace the development of Moken from PAn directly to Moken.

Since Dempwolff (1934-8), various aspects of PAn phonology have been addressed, both in terms of phoneme inventory and prosody.[4] Although different opinions on the consonant system exist, there is a general consensus about the vocalism. It was certain that PAn had two contrastive stop series: voiced and voiceless. The distinction is found initially,

[2] ĭ and ŭ in Chantanakomes (1988)'s notation.

[3] For the distinction to be "phonologically supported", the two types of syllables must behave differently phonologically, e.g. they do not follows the same pattern of affixation etc.

[4] An overview of different PAn reconstructions can be found in Ross (1992).

medially and finally. Its vowel system is a simple one with only 4 vowels. The canonical shape of PAn roots was CVCVC while trisyllabic and monosyllabic roots also existed (Ross 1992, Wolff 1999). Both Ross (1992) and Wolff (1993) agree in reconstructing contrastive stress in PAn while in other author's reconstructions stress placement is ignored. In this paper, the PAn reconstruction used is that of Wolff (2002, and in progress).

These phonological characteristics distinguish descendents of PAn from languages of Mainland SEA, which has converged into a linguistic area with its own phonological characteristics (Matisoff 2001). However, Moken appears to have diverged from the PAn norms and come to have striking Mainland SEA characteristics. These features include a three-way contrast among stop consonants, neutralizations in the coda, a rich vowel system and a strict prosodic template of words. These characteristics are found across languages of Mainland SEA but are absent from insular PAn languages (Bennett 1995; Green 1995; Ratliff 1992; Svantesson 1983; Teoh 1994). Another group of Mainland Austronesian group, Chamic, is also very similar typologically to other Mainland languages. Note that, these 'so-called' Mainland SEA features on their own are not unique to the area. It is the pervasive co-occurrence of these features in the area that makes them as areal features. These features are summarized in Table 1.

Table 1: *Mainland Southeast Asia areal features compared to Malay*

Features	Thai (TK)	Kammu (MK)	Burmese (TB)	Hmong (HM)	Malay (AN)
Three-way stop contrast	✓	✓	✓		
Neutralization in coda position	✓	✓	✓	✓	✓
Three-height vowel contrast	✓	✓	✓		
Vowel-length contrast	✓	✓		(✓)	
Contrastive diphthongs	✓	✓	✓	✓	
Phonological word = Foot		✓		✓	
Bimoraicity of foot head	✓	✓	(✓)		
Iambicity	✓	✓	✓	✓	

TK = Tai-Kadai HM = Hmong-Mien
MK = Mon-Khmer AN = Austronesian
TB = Tibeto-Burman

(✓) represent high-tendency but not absolute requirements.

3. Contrastive aspirated stops as a loan-induced feature

Among the Mainland features outlined above, one striking characteristics of Moken from an Austronesian perspective is the presence of a full series of contrastive voiceless aspirated stops. An overwhelming majority of forms that have aspirated consonants cannot be identified as having Austronesian affinities. The most obvious source of loanwords is Thai (presumably Southern Thai).

Table 2: *Some Thai loanwords with aspirated stops in Moken*[5]

məchay ‘to use’	< chay	məkha:m ‘to cross’	< kha:m
məthu:n ‘to carry on head’	< thu:n	thuʔ ‘sorrow’	< thuʔ (< Pali)
məkhɔʔ ‘to strike’	< khɔʔ	məchaŋ ‘to weigh’	< chaŋ
kathaŋ ‘to arrive at’	< thɤŋ	khiŋ ‘half’	< khrɯŋ
phu:ŋ ‘herd’	< fu:ŋ	phəlu:ŋ ‘hole’	< phroŋ

However, some forms found with aspirated consonants are inherited words. The first set contains words with aspirated *ch* of Austronesian origin, i.e. *cxchɔy* ‘milk breast’ < **cucu*, *macham* ‘sour’ < **qaləcam*, and *mxchɔŋ* ‘to carry’ < **qucuŋ*. These etyma all go back to PAn **c* in Wolff’s reconstruction.[6] Larish (1999) shows successfully that *ch* and *s* in Moken are free variants of the phoneme *ch*. It is not clear what the original reflex was but that the **c* is also reflected in some forms as *c*. Originally, **c* may have become *ch* which later developed to have *s* as a variant. The on-going change from *ch* to *s* must be internally motivated. Such change is not necessarily connected with any particular language family; it is common in the world’s language and also found among Tai languages in the Shan and Lao groups, which have no contact with speakers of Moken.

The second group of Austronesian etyma with aspirated stops consists of forms with aberrant aspiration, including *khuɟa:n* ‘rain’ < **quɟaɲ*, *phəla:* ‘husked rice’ < **bəlac*, and *thuwa:* ‘two’ < **dusa*. Normally, PAn voiceless stops are reflected as unaspirated; such forms are sporadic and no explanation can be offered at this point. However, it is to be noted that ‘two’ and ‘husked rice’ show variation both within the speaker and cross-dialectally. The other variants are the regular reflexes *duwa* and *bəla:* respectively. Whatever the source of this aberrance is, the aspiration in these forms must be relatively recent given its limited geographical distribution and its status as a variable in speech of individual speakers. Therefore, it cannot be due to contact in the remote past.

As shown above, the aspiration contrast was imported from the languages that the speakers of Moken have been in contact with or else is a sporadic and recent development. Crucially, it is unlikely that Mon-Khmer was the source of such heavy borrowing since only a very small number of the few forms of certain Mon-Khmer origin, if any at all, show aspirated stops. The only solid case of word of Mon-Khmer affinity is *kəthiəm* ‘onion, garlic’, cf. Proto-South Bahnaric **diəm* (Sidwell 2000) but this etymon was borrowed via Thai, cf. *krəthiəm* ‘garlic’. This is in general agreement with Lewis (1960)’s preliminary estimates that out of 1430 significant entries of a Moken dialect spoken in Myanmar, 365 are Austronesian, 69 are Thai, 36 are Burmese, 914 are of unknown origin, and only 46 are Mon-Khmer.

[5] Thai tones are omitted from the transcription.

[6] *t’ in Dempwolff’s and *s in other authors’ (Wolff, personal communication).

4. Vowel contrasts as internal restructuring

One striking feature of Moken is its large vowel inventory, in contrast with the compact PAN system. In the Rawai Moken vocalism, three height distinctions are found along with diphthongs, and length contrast. The origin of these distinctions is hypothesized by Larish (1999:318, 394-403) as being of Mon-Khmer influence. However, I show that they should be viewed as internal changes within Moken.

4.1 Vowel-height contrast

The full vowel contrast occurs only in the major syllable, which is the head of the word. The most apparent change in quality is that of PAN **ə*, which becomes *a* regularly in both closed and open syllables. Unlike the central vowels, the two PAN high vowels **i*, and **u* change dramatically according to their phonological environment. This leads to the present-day quality distinctions in Moken.

PAn **-i-*, and **-u-* in closed syllables are lowered to *-ɛ- and -ɔ-* in most cases. Larish (1999) mentions that these changes are conditioned by segmental and suprasegmental low-pitched environments without explaining how these environments are defined and how they affect the development of the Moken vocalism.

Table 3: *Reflexes of PAn *-i- and *-u-*

	*-i- > -ɛ-		*-u- > -ɔ-
a. *kulit	kɔlɛt ‘bark of tree’	e. *likud	lɛkɔt ‘behind’
b. *lilin	lɛlɛn ‘wax, candle’	f. *gaɣut	ŋalɔ:t ‘to scrape’
c. *nasik	ɲaʔɛk ‘to ascend’	g. *ɣatuc	latɔh ‘100’
d. *paqit	pakɛ:t ‘bitter’	h. *manuk	manɔk ‘chicken’
	*-i- > -i-		*-u- > -u-
i. *biɣbiɣ	bibi:n ‘lip’	k. *buɲuq	munu:k ‘to kill’
j. *butəliɣ	buti:n ‘cyst’	l. *butuq	butu:k ‘penis’
		m. *ikuɣ	ʔiku:n ‘tail’

In fact, the lowering seems to have applied pervasively. Relevant conditioning environments blocked this lowering process, rather than enforcing it. The most obvious conditioning environment is the final consonants. Specifically, **-q*, and **-ɣ* blocked lowering, creating allophonic alternations between high vowels before **-q*, and **-ɣ* and low vowels elsewhere. Subsequently, **-q*, and **-ɣ* merged with **-k* and **-n*, resulting in a new contrast between low and high vowels. This pattern is illustrated by the contrast between examples (i-m) in Table 3 which ended with *ɣ and *q and whose modern forms show long high vowel on the one hand, and examples (a-h) where the PAn etyma ended with other consonants and where modern reflexes show low vowels on the other.[7]

[7] However, there are some cases that the vowels unexpectedly failed to lower. These forms are sporadic and the failures to lower might possibly be related to PAn stress. That is, PAn ultimate stress prevented the major syllable vowel from lowering, cf. *ɲipih ‘thin’ < *ɲisəbic*, *kudip < *kudip*. This explanation remains a speculation since the reconstruction of PAn stress is still in its infancy.

Chantanakomes (1980) analyzed the mid vowels as a tense high vowel, not in terms of height contrast, although she does not discuss what is meant by ‘tense-lax”. Larish (1999; 154) considers such distinction to be of MK-type register distinction and thus reconstructed it for Proto-Moken-Moklen.[8] In fact, my data suggests that the distinction in Rawai Moken should be viewed as rather similar to the distinction between lax *ɪ* and *i* and between *ʊ* and u in English. In this paper, these vowels are simply labeled “mid vowel”, as I assume that the distinctive feature is height and not tenseness.

In any case, it is clear that the occurrence of these vowels is limited compared to the high and low vowels (Chantanakomes 1980;17). Those few that do occur go back to the PAn diphthong _*-iw_. Examples include *kaʔe:* ‘tree’ < _*kásiw_, *male:* ‘flee’ < _*laɣiw_. Note the final diphthongs in Moklen *kaʔɛ:w* ‘tree’. This suggests that the vowels are recent innovations in individual dialects. Other cases of Moken mid vowels seem to be of non-AN origin, e.g. *lase:* ‘book’< *naŋsɯ:* (Thai), *phe:* ‘to be defeated’ < *phɛ:* (Thai), *yiʔo:* ‘radio’. None of them seems to be of Mon-Khmer origin. An exhaustive list of forms with mid vowels recorded by Chantanakomes (1980) is provided in the Appendix.

4.2 Length distinction

Another Mainland feature which can be seen as an internally driven is the development of the simple PAn vocalism into Moken complex system with quantity contrast, which is, like other aspects of Moken historical phonology, full of complexities, not all of which have been solved. Although a vowel length distinction is present in Moken, it does not seem to be a robust one. This is seen in the non-occurrence of some expected combinations, such as *-ep*, *-u:p*, *-uk*, *-i:m*, *-u:y*, and *-ɔy*. In addition, short vowels can also become long before pause or when emphasized (Chantanakomes 1980).

The most apparent and reliable source of vowel length distinction is from the earlier contrast between PAN _*ə_ and _*a_. In the major syllable the quality contrast transformed into a one of quantity. That is, _*ə_ became _*a_ while _*a_ became _*a:_.

Table 4: *Reflexes of PAN *ə and *a.*

*ə > a		*a > a:	
*ipən	*lɛpan ‘tooth’	*mata	mata: ‘eye’
*pukət	pukat ‘dragnet’	*bəɣəhat	baʔa:t ‘heavy’
*kəp	məŋap ‘to catch’	*gap	maŋa:p ‘to grobe’
*baɣəq	balak ‘swell’	*bəlaq	məla:k ‘split’

Another source of long vowels is the lengthening conditioned by the two post-velar consonants _*q_ and _*ɣ_. Specifically, high vowels are lengthened when followed by these two consonants. That the combinations of high vowels and the post-velar consonants are precisely lowering-blocking environments suggests that the lengthening occurred first and

[8] Larish's evidence for tense/lax distinction in PMM is also based on data from the Dung dialect (Naw Say Bay). I, however, suspect the tense/lax differences to be allophonic. In any case, the supposed tense/lax distinction cannot have resulted from MK influence as the development in Dung Moken shows the same type of internal development as in Rawai Moken.

then the lowering subsequently occurred to the high vowels that stay short[9]. Once the height is established, _*q_ and _*ɣ_ could then easily have merged with _*k_ and _*n_. The task of filling the vowel space to have length contrasts for every vowel can then be left to borrowing.

Table 5: _Lenghtening before *q and *ɣ_

*Vq > (*V:q) > V:k		*Vɣ > (*V:ɣ) > V:n	
*tubuq	numu:k 'to grow out'	*biɣbiɣ	bibi:n 'lip'
*tuduq	tudu:k 'leak'	*ikuɣ	ʔiku:n 'tail'
*butuq	butu:k 'penis'	*qitəluɣ	kəlu:n 'egg'

4.3 Diphthongs

In addition to its relatively rich inventory of monophthongs, Moken also has a variety of diphthongs in its vowel inventory. The full three-height contrast occurs only in the closed major syllables and not in open syllables. PAN _*i_, and _*u_ regularly diphthongized and merged in open final syllables but are retained in closed syllables (Larish 1999:323). The resulting diphthong may differ among dialects but Rawai Moken shows _ɔy_ and _uy_ regularly for both _*i_, and _*u_.

Table 6: _Moken reflexes of PAN *-i, *-u, *-ay, and *-aw._

*-i > -ɔy, -uy		*-u > -ɔy, (-uy)	
*gali	ŋalɔy 'to dig'	*cúcu	cochɔy 'milk'
*qəti	katɔy 'finish'	*batu	batɔy 'stone'
*wáɣi	ʔalɔy 'day'	*búbu	bubɔy 'fish trap'
*buni	munuy 'to hide'	*kuku	kɔkɔy 'fingernail'
*-ay > -ay		*-aw > -aw	
*balay	balay 'large open house'	*baɲaw	maɲaw 'to wash'
*ɣuqáɲay	kanay 'man'	*lakaw	lakaw 'to walk
*lantay	latay 'platform'	*láŋaw	laŋaw 'to fly'
*qatay	katay 'heart, liver'	*talaw	talaw 'coward'

Note that some forms in Rawai Moken that go back to PAn _*i_ and _*u_ show _uy_, instead of the expected _ɔy_, i.e. _diluy_ 'thorn' < _*diɣi_. The most likely scenario is that these vowels diphthongized into _uy_ and then regularly lowered to _ɔy_. However, some _uy_—both original and secondary—unexpectedly did not lower. The existence of _uy_ may be the same phenomenon as the non-lowering found in monophthongs discussed earlier. The resulting diphthongs adds to the inventory of diphthongs inherited from PAN _*ay_, _*aw_, and _*uy_ that did not lower.

[9] A few forms with low vowels have long vowels, i.e. _ŋalɔ:t_ 'to scratch' and _pakɛ:t_ 'bitter'. They might be exceptions. Again, the length contrast is still not very robust.

These forms are clearly of Austronesian origin and no conditioning environment can be identified. Therefore, I hypothesize that these non-lowered forms resulted from dialect-mixing since some other Moken-Moklen dialects show *əy* as regular reflex of **u* and **i*. According to my experience in the field, intermarriage between different groups is also common; this sociolinguistic fact gives support to the dialect-mixing analysis.

In addition, some cases of diphthongs *iə* and *uə* can also be said to result from a conditioned change within Moken. As shown earlier that high vowels followed by *-*q* did not lower. Some of these, however, show diphthongal reflexes instead of high vowels, cf. *bituək* 'start' < **bituq,* and *miliək* 'to choose' < **piliq*. One possibility is that PAn stress conditioned the split, that is, stressed syllables with *-*q* did not lower but, unlike their unstressed counterparts, went through another process of diphthongization, in which *-*íq* and *-*úq* became *iək* < *(*iəq)* and *uək* < *(*uəq)* respectively. Contrast *bituək* 'star' < **bituq* and *butu:k* 'penis' < *butuq. This hypothesis still need furthers investigations.

4.3 Vowel distinction in the minor syllable

Unlike the major syllable, the minor syllable lacks length contrast and does not allow mid vowels other than the reduced vowel *ə*. Larish (1999:321) notices that PAN high vowels are retained in both syllables when the two vowels share high vowels and that vowel lowering is a common process in both type of syllable. He also provides some cases of exceptions to the retention of the high vowels without any explanation.[10]

Ignoring the length distinction and the mid vowels, the cases of identity of the vowel of the minor syllable and that of the major syllable should rather be viewed as the vowel in the minor syllable harmonizing with that of the major syllable. That is, the vowel height of the minor syllable is determined by the major syllable. The minor syllable must agree in height with the major syllable, except for *a* and *ə* which do not have high counterparts and thus are not raised or lowered.

Table 7: *Height harmony of the vowel of the minor syllable and that of the major syllable*

[+high] – [+high]	[-high] – [-high]	[-high] – [+high]	[+high] – [-high]
ɲulu:k 'to shine'	kɔlɛt 'bark of tree'	babuy 'pig'	miɟak 'to tread'
ʔuɟuŋ 'end'	ʔɛkɔy 'elbow'	kaʔu:n 'bamboo'	kuɟa:n 'rain'
gilin 'to roll up'	ʔɛnɔŋ 'mother'	ɲapu 'to sweep'	bula:n 'moon'
miliək 'to choose'	phəla: 'husked rice'	ɟalum 'needle'	binay 'woman'
kudip 'life'	mɛla:k 'red'	kabut 'cloud'	bulɔy 'body hair'
ɲuli:t 'to slit'	ʔɔma:k 'house'	ləpu:k 'lion fish'	duwa: 'two'
midu:n 'to sleep'	pɔcat 'navel'	pənuk 'full'	dulaŋ 'k.o. basket'
lipuy 'to dream'	ɟana:t 'child'	məɲup 'to blow'	gutɔy 'louse'

The generalization is shown clearly in column 1, 2, and 3 above. Forms in column 4, though, seem to be exceptions at first glance but a blocking environment can be identified. Initial *ɟ* of the major syllable blocks lowering of the preceding vowel, cf. khuɟa:n 'rain',

[10] Contrary to this analysis, the diachronic lowering of the high vowels is, in fact, confined to the major syllable only.

miɟak ‘to tread’. In addition, a voiced initial in the minor syllable disallow lowering of the vowel following it. These blocking environments are attested by the gaps in the distribution of low vowel in minor syllable. That is, *b*, *d*, and *g* in the minor syllable never precede low vowels (Chantanakomes, 1980: 23).[11]

As shown above, the striking Mainland features in Moken can be accounted for without appealing to influence from MK or other Mainland languages, aside from borrowing from Thai in the case of aspirated stops. In addition, the minor syllable shows a six-vowel contrast expanded from the PAn four-vowel system, in contrast with the minor syllable in MK and some Chamic languages which only have a neutral vowel in the minor syllable (Thurgood 1999). Therefore, it is clear that the shift is a result of gradual sequence of internal processes that result in rich vowel inventory commonly found in Mainland Southeast Asia.

5. Strict word-template as reflexes of PAn

5.1 mə- prefixation

As in the case of the vocalism, some other Mainland SEA features in Moken can be viewed as instances of internal restructuring. That is, PAn contrasts are reflected in Moken but the total organization of the system has changed. The restructuring is clearly Moken-internal but that the language opts for these particular paths is suggestive of the areal influence on Moken. The development of Moken toward a language with a strict word template is a case in point.

The strict word-template in Moken is instantiated in two aspect of the grammar: the frequency of monosyllabic forms and the process of affixation. Lewis (1960) observed that the Moken word is predominantly two syllables of the form CVCV(C) and that trisyllabic roots are very rare. All most all that occur are loanwords. She also shows that only a total of 172 out of 1430 entries are monosyllables, many of which are loan words or grammatical items. This observation is consistent with the data from Rawai Moken. In addition, Moken morphology also demonstrates that the iambic disyllabic word-template is very strict.

Table 8: *Forms that show mə- prefixation and their PAn roots*

*baɲaw	maɲaw ‘to wash’	*bəli	məlɔy ‘to buy’
*pacək	masak ‘to nail’	*piliq	miliək ‘to choose’
*tawa	nawa: ‘to laugh’	*tubuq	numu:k ‘(for beard) to grow out’
*culuq	ɲulu:k ‘to shine’	*capus	ɲapu ‘to sweep’
*qubaɲ	ŋɔba:n ‘(for hair) to be grey’	*qucuŋ	mɔchɔŋ ‘to carry on pole’
*sikət	mɛkat ‘to tie’	*satəd	matat ‘to deliver’

As shown above, forms that show the nasality alternation are all verbs. Larish (1999) rightly suggests that the alternation is morphologically conditioned by attributing it to a non-productive nasal replacement, through which verbs are derived. A full prefix *mə-*, in contrast, is added to monosyllabic roots, yielding a disyllabic verb. However, he

[11] Phonetically, voiced stops have a lowering effect on F1. The F1 onset of the following vowel starts relatively low (Jessen 2001). Therefore, this explains the blocking of lower as a reanalysis of the following low vowel with low F1 as high vowels, i.e. **bulan* > (**bɔlan*) > *bulan*.

wrongly treats this nasal replacement as separate from *mə-* prefixation[12]. These two processes are, in fact, a single process, that is *mə-* prefixation. Moreover, that *mə-* prefixation is still required for incorporation of Thai words, cf. *məplɛ:* 'to translate', in contrast, suggests that the affixation is still productive. Although the precise morphological and semantic function of this prefix is still not fully understood, its phonological behavior is clear.

In cases of monosyllabic roots, the *mə-* prefix is simply added to the root, making it disyllabic, i.e. *məcay* 'to row (a boat)', *mədɔ:k* < **duk* 'to sit', and *məku:n* < Thai *ku:n*. It takes the form of *m-* when attached to a vowel-initial disyllabic root, keeping the verb disyllabic. Ultimately, if the root is disyllabic and begins with a consonant, the initial consonant becomes nasal homorganic to the original initials. The *mə-* prefixation of disyllabic roots is illustrated in Table 8. Synchronically, this prefixation points out to a non-violable word-maximality constraint that requires that every Moken phonological word be an iambic foot. It is strongly suggestive that this characteristic is a truly mainland feature that Moken has adopted since similar processes also found in other Mainland SEA language, i.e. the causative infixation in the MK language Kammu spoken in northern part of Southeast Asia (Pittayaporn ms.; Svantesson 1983).

5.2 Syncopation and PAn stress

The strict disyllabic word-template found in Moken is not characteristic of Austronesian languages. It is the end product of a long process of syncopation. which is highly likely to have been conditioned by PAn stress. Since Zorc (1983; 1992) showed that contrastive stress has to be reconstructed for what he calls Proto-Philippines, linguists working on PAn (Ross 1992; Wolff 1993; Zorc 1978) have presented evidence in support for the contrastive stress in PAn; others (Blust 1997) still express doubts. The development of Moken from PAn presented in this paper argues at least partially in support of the existence of unpredictable accent in the Proto-language. The hypothesized PAn stress is reflected in Moken in various ways.

The main piece of evidence comes from syncope of PAn trisyllabic roots to form a canonical Moken iambic disyllable. Like in other mainland Southeast Asian languages, a Moken word must be an iambic foot of shape CVCV́(:)(C), as opposed to PAn canonical (CV)CVCV with stress on either of the syllables. For PAn disyllabic roots, the path to the observed Moken canonical shape is simply shifting the stress to the final syllable, thus not providing any evidence for the earlier stress pattern. However, the development of trisyllabic etyma is more complicated and very suggestive of syncopation in Moken as reflexes of PAn stress pattern.

Larish (1999:369-70) suggested that unstressed syllables in PAn trisyllabic forms, such as **tuqəlaɲ*, **buqaya*, * *baqəɣu*, and **taliŋa* are dropped to yield Moken *kəla:n* 'bone', *kaya:* 'crocodile', *kəlɔy* 'new', and *tɛŋa:* 'ear'. In these specific cases, the vowel of the antepenult is lost and the resultant cluster is simplified. However, there are cases where the antepenult is retained, e.g. *kaʔɔy* 'pestle' < *qasəlu, and *kapaw* 'gall' < **qapəgu*.

Such loss of unstressed syllables is well-attested in languages all over the world and across the range of the Austronesian languages as well (Wolff, personal communication). The most famous example is the development from Latin to Romance

[12] A possible PAn candidate is **maN-*, as reconstructed by Wolff (1996).

languages (Posner 1996). That either the ultimate, the penultimate or the antepenultimate can be dropped suggests that stress may have been placed in different position in different roots (Zorc 1993).

The relationship between syncopation in Moken and PAn stress can be clarified by examining the similar pattern of syncopation in Proto-Malay (Adelaar 1992) and Proto-Philippines (Zorc 1978). Previous researchers (Zorc 1978, Wolff 1993, Ross 1992) have discussed this relationship; in Table 9. I have presented Proto-Malayic (PM) and Proto-Philippines (PPh) data together with the corresponding Moken forms[13].

Table 9: *Syncopation of PAn in Proto-Malayic, Proto-Phillipine, and Moken*

	PAn[14]	Proto-Malayic	Proto-Philippine	Moken	Gloss
1	*báɣəhat[15]	*bərat	bigat (Tg)	baʔat	heavy
2	*qásəlu	*halu	*haqlu	kaʔɔy	pestle
3	*búɣəsu	cəm-buru (Ml)	pani-bughoʔ (Tg)	mɔlɔy	to be jeaolous
4	*qáləcəm	*m-asəm	ásim (Tg)	masam	sour
5	*tuqəláɲ	*tulaŋ	*tuqlaŋ	kəla:n	bone
6	*talíŋa	*taliŋa	*tali:ŋa	tɛŋa:	ear
7	*buqáya	*buhaya	*buqa:ya	kaya:	crocodile
8	*ɟuɣámi	*jərami	*daRa:mi[17]	-	straw
9	*qaɲítu	*hantu	*qani:tu	katɔy	evil spirit
10	*baqə́ɣu	*baharu	*baqRuh	kəlɔy	new
11	*qapə́gu	*hampədu	apdo (Tg)	kapaw[18]	gall
12	*qitə́luɣ	*təlur	itlog (Tg)	kəlu:n	egg
13	*sapəgíq[16]	*pədih	hapdiq (Tg)	pəyiək	to smart

[13] According to Wolff (1993), the PAn stress is preserved in many Philippine languages in form of vowel length in most cases and stress is predictable in term of length. Tagalog prominence is realized both as length and stress but phonological evidence suggests that it should be considered stress (French 1988).

[14] PAn/PMP roots cited in Adelaar (1992) are substituted by Wolff's reconstructions (in progress). Some forms presented here may not go back to PAn but only to PMP but all languages being compared are Malayo-Polynesian. Therefore, it is justifiable to include MP forms in the analysis. Reconstruction of stress is my own and is based on the arguments presented above.

[15] Wolff (personal communication) suggests that PAn 'heavy' might have to be reconstructed as *bəɣəhat to account for the *ə* in Ml and *i* in Tg.

[16] Tg. *hapdiq*, there are two possible reconstructions for PAn that are consistent with the retention of the penult in PM and the syncope of the **ə* in PPh: *sapəgíq or *sapə́giq. However, the loss of the antepenult is unexpected because the initial is not a laryngeal. Wolff (in progress) notes that the **sa-* in this case might turn out to be a prefix. The PM and the Moken forms are derived from the unaffixed disyllbic form of the root. The issue is then not related to the question of syncope.

[17] Zorc's *R correspond to *ɣ in Wolff's system.

Examples (1-5) show that the penults of some roots were syncopated before the PM stage. This suggests that the penult was unstressed in the proto-language. This interpretation is strengthened by the Philippine cognates, which also show syncope of the penult. The PM and Philippine data alone cannot be used to determine whether stress was on the ultimate or the antepenultimate syllable. It is the contrast in Moken between (2) *kaʔɔy* < **qasəlu* and (5) *kəla:ŋ* < *tuqəlaɲ* in Moken forms that shows that stress fell on the antepenult in (1-4) but on the last syllable in (5).

Unlike (1-5), PM in (6-8) show retention of the penult, agreeing with the Philippine forms, which not only retain the syllable but also show the predicted stress. This correlation strongly suggests that these roots had penultimate stress in PAn. However, the PPh form in (9) shows stressed penult while the PM form shows syncopated penult. It is possible that there was a stress shift either in PM or PPh due to the taboo nature of the etymon. The Moken forms also retain the penults, giving further support to PAn penultimate stress in these roots.

At first glance, forms in (10-12) seem to present a problem for stress reconstruction since there are disagreements between PM and the Philippine cognates. Specifically, PM roots retained the penult suggesting it was stressed while the syncopated forms in the Philippine languages suggest non-stressed penult. However, these cases all involve **ə* in the penult, suggesting the possibility of a stress shift conditioned by **ə*. According to Zorc (1992:89), **ə* cannot be stressed unlike PPh **i*, **u*, and **a*. This lends support to the stress-shift speculation since there is a gap of stress distribution in PPh. In other words, I hypothesize that PAn had stressed penultimate **ə* in these cases and that PM retains the original pattern while PPh innovated by shifting the stress to avoid accented **ə*. Note that loss of antepenults as is the case for (13) is a well-attested change in PM (Adelaar 1992). Moken consistently dropped the antepenult, suggesting that it agrees with PM in having stressed antepenult in these etyma.

Focusing on the Moken forms, Table 9 shows that Moken only disagrees with PM in cases of (5) and in the case of (6-10) where the PAn penultimate is retained. That Moken *kəla:n* 'bone' corresponds to PM **tulaŋ* suggests that the antepenult of this form was also unstressed in PAn, leaving the last syllable as the only candidate for accentuation. The rarity of ultimate stress may explains why PM unexpectedly shows *-ŋ* instead of *-n* for PAn *-ɲ in this etymon. That is, it is possible that final *ɲ is reflected as *ŋ only in PAn forms in whose last syllable was accented. Assuming this analysis, it becomes clearer that Moken systematically syncopated the antepenult in roots with accented penultimate or ultimate syllable. Cases in (1-4) can then be taken as evidence for antepenultimate stress since Moken agrees with PM, as well as PPh, in retaining the first syllable of the roots. The syncope rule in Moken is then that the antepenultimate vowel is syncopated unless accented.

The generalization about PM and PPh syncope is then that the penultimate syllable of PAn trisyllabic roots was lost regularly unless it was stressed. The retention of the penult in PM can then be used as a diagnostic for PAn penultimate stress. It may not be

[18] Moken *kapɑw* 'gall' (11) shows a seemingly problematic retention of the unstressed antepenult **a*. This is because the change from *-*g*- to -∅- is regular in this environment. That is, the root must have already been reduced to disyllabic before the syncope took started to operate.

amiss to anticipate the argument that these reconstructions implicating stress risk circularity. That is, it seems to attribute certain otherwise unexplained patterns of syncopation in Moken to PAn stress and then proceed to reconstruct stress in the relevant items. However, the crucial point is that idiosyncratic patterns of syncopation in the three languages PM, PPh and Moken strongly reinforce each other and suggests that some kind of prominence, if not stress per se, played an important role in such processes.[19]

The process of canonical reduction discussed shows that importance was given to the last syllable of the roots. However, this pattern of syncopation strongly suggests that in Pre-Moklen-Moken, the stress pattern was still not predictable, in contrast with the hypothesis that it was the stress shift that triggered the syncopation. Only after the syncopation had taken place could the stress be shifted to the ultimate syllable, as evidenced by the retention of the stress in PAn antepenultimate vowel, e.g. *baʔat* 'heavy', *kaʔɔy* 'pestle' in Table 9. The strict word-template resulting from such canonical reductions, though suggestive of MK influence, may be viewed simply as an areal feature that cannot be attributed to a single source. This is because iambicity is found through out Mainland SEA, not just in MK and the word-maximality constraint can also be viewed as epiphenomenal.

5.3 Cluster resolution

Larish (1999) shows how syllables in trisyllabic roots are dropped to yield strict disyllabic template in Moken. However, he does not mention the fact that * *talíŋa* gives *tɛŋa:*, not the expected *taŋa* or *lɛŋa:*. The onset of the antepenult is preserved while it is the penultimate vowel that is retained. This paradox suggests that it is not the whole unstressed syllable but only the vowel that is lost, resulting in a complex cluster. The cluster was then resolved according to the sonority sequencing. That is, the less sonorous element is retained while the more sonorous one is lost. If a cluster of two stops is created, the one further back is retained, as shown in Table 10. These tendencies for syncope are quite regular. Note that in *lɛta:k* 'leech' the liquid is unexpectedly retained. This is because the fourth syllable fron the end **qa-* was lost early on as it was also in Ml *lintah* (Wolff, personal communication).

[19] Adelaar (1992) summarizes that syllable reduction in PM occurred in roots of more than two syllables through any of the three processes: vowel contraction, syncope of penultimate syllable, and loss of PMP laryngeal initials. Assuming stress in PAn, these three processes can also be explained in terms of unpredictable stress placement.

Table 10: *Syncope of some PAn trisyllabic and quadrisyllabic forms*

*talíŋa	> *tlíŋa	> *tíŋa	> tɛŋa: ‘ear’
*buqáya	> *bqáya	> *qáya	> kaya: ‘crocodile’
*baqə́ɣu	> *bqə́ɣu	> *qə́ɣu	> kəlɔy ‘new’
*tuqəláɲ	> *tuqəláɲ	> *qəláɲ	> kəla:n ‘bone’
*ɣuqáɲay	> *ɣqáɲay	> *qáɲay	> kanay ‘man’
*qitə́luɣ	> *qtə́luɣ	> *qə́luɣ	> kəlu:n ‘egg’
*sabáɣat	> *sbáɣat	> *báɣat	> bala:t ‘west wind’
*isəkan	> *iskan	> *ikan	> ʔɛka:n ‘fish’
*qasulipan	> *qsulpan	> *qupan	> kɔpa:n ‘centipede’
*(qa)ɲimatáq	> *ɲimtáq	> *ɲitáq	> lɛta:k ‘leech’

Among the language families of Southeast Asia, MK languages are well-know for having relatively large number of clusters, as opposed to other languages, such as Thai and Burmese, whose clusters are rather scarce and usually subject to simplification. That is, once again, the phenomenon of cluster resolution does not provide evidence in support of Moken having been influenced by any particular language group. Rather, it supports the view that the absence of clusters is a constraint continued from PAn[20].

6. Conclusion

Thomason and Kaufman (1988:37-39) distinguish two main types of change induced by contact: borrowing and shift-induced interference. Borrowing is defined as a situation in which the native language is maintained with addition of the incorporated features while shift-induced change is defined as interference from an imperfect learning of the target language by the shifting speakers. The main diagnostic is the relative degree of influence within the subparts of the grammar. Since the influence of Mainland SEA languages is pervasive through out the phonology, the lexicon[21] and the morphosyntax[22], it becomes unclear in which part of the grammar the interference started. However, the radical phonological shift toward Mainland SEA type has been shown here to involve sequences of changes, suggesting that the typological change, or “metatypy”, was gradual. If the generalization that language shifts occur rapidly holds (Thomason and Kaufman 1988:41), the case of Moken must have been a moderate to heavy borrowing situation.

Ross (2003) proposes that contexts of contact situation be analyzed in terms of internal and external relationships of the speech communities. In this framework, the

[20] There are cases of Moklen clusters corresponding to Moken simple consonants, cf. Moklen *caplɔh ‘10’* vs. Moken *capɔh*. Larish (1999: 151, 325, 477) notices this phenomenon but does not attempt to explain the dialectal differences. I propose that it results from two related processes: deletion of Moklen minor syllable *ə* and Moken *ə* epenthesis.

[21] Basing calculations on “the Matisoff 200-word list”, approximately 45% of the basic vocabulary is of AN affinity. However, only 25% can be identified as having AN affinity when both basic and non-basic vocabulary is taken into account, cf. Lewis (1960).

[22] As a speaker of Thai, my impression is that Moken is morpho-syntactically very similar to Thai although some obviously Austronesian features are still retained. Systematic comparison is still needed.

Moken community may have been open, tightknit and multilingual. Open communities are ones that have numerous external links, while tightknit communities are characterized by having a strong social network and by associating their primary language with high emblematic value. In this view, the present day Moken is essentially an Austronesian language that has gone through a metatypy, in which the native language was restructured on the model of the secondary languages, that is, the Mainland SEA languages it has been in contact with.

Although more socio-historical investigation is needed in order to be certain how the contact situation really was, this hypothesis is in general agreement with the observed sociolinguistic situation in the present-day Moken speech community. The Moken at Rawai beach live in a tightknit community which outsiders do not frequent. Although they live directly adjacent to the Urak Lawoi village, the relationship between the two villages can be characterized as segregation. Although it is a close-knit speech community, it is a considerably open speech community. They do have relationships with the Thai majority as they are hired by local Thai to dive and fish. Women also sell sea products, such as fish and shells, to Thais and tourists. Although the primary language of communication in the village is Moken, they are all bilingual in Southern Thai. In addition, children are now going to school where the only language of instruction is Standard Thai. There is a tendency for young children not to speak the Moken language actively.

In this paper, I have shown systematically how Moken, an Austronesian language, has become typologically similar to Mainland SEA languages. These processes are results of both borrowing and internal development. Such convergence has most likely been gradual and involved a complex series of changes that cannot be attributed to a MK or any single source. Rather, it should be viewed as being propelled by an internal mechanism, which is, in turned, accelerated and directed by the languages it has been in contact with. These languages may include Thai, Burmese, one or more MK languages and possibly an unknown language. It has also been hypothesized that the present stage of Moken results from a prolonged borrowing interference that the open tightknit and multilingual Moken community has been subject to. Such contact situation can be viewed not as the providing directionality to internally-motivated changes which lead to a convergence towards Mainland SEA typology.

Acknowledgement

I am greatly indebted to Professor John Wolff of Cornell University who taught me what I know about Historical Austronesian and who not only permitted me to learn from him in the field and to use his Austronesian database but also gave me valuable comments on this paper. I am highly thankful to my advisor Professor John Whitman for his time, insights, and encouragement. I would like to thank Professor Abigail Cohn, and Marc Brunelle for their valuable comments. I would also like to dedicate this first paper of mine to Professor Pranee Kullavanijaya, Professor Theraphan L-Thongkum, and Professor Kalaya Tingsabadh, who have always been my support and who inspired and prepared me to work in Southeast Asian Linguistics. I would also like to thank the Department of Linguistics, Cornell University, without whose support my graduate study is possible and the Anandamahidol Foundation, Thailand, for providing me initial financial supports to start my study at Cornell. I would like to thank the Department of Linguistics, Southeast Asia Program, and the Graduate School at Cornell for their travel grants.

References

Adelaar, Alexander. 1992. *Proto Malayic: the reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Deparment of Linguistics, Research School of Pacific Studies, the Australian National University, 1992.

Bennett, Jefferson Fraser. 1995. *Metrical foot structure in Thai and Kayah Li: Optimality-theoretic studies in the prosody of two southeast Asian languages.* Department of Linguistics, University of Illinois at Urbana-Champaigne: Ph.D. Dissertation.

Blust, Robert. 1992. The Austronesian settlement of Mainland Southeast Asia. *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society 1992*, ed. by Karen L. Adams and Thomas John Hudak. Tempe, Arizona: Arizona State University.

—. 1997. Rukai Stress Revisted. *Oceanic Linguistics*, 36.398-403.

Chantanakomes, Veena. 1980. *A description of Moken: A Malayo-Polynesian language*, Institute of Language and Culture for Rural Development, Mahidol University: M.A. Thesis.

Dempwolff, Otto. 1934-8. Vergleichende Lautlehre des Austronesischen Wortschatzes. *Zeitschrift fur Eingeborenen Sprachen.* Supplements 15, 17, 19. Berlin: Dietrich Reimer.

French, Koleen Matsuda. 1988. *Insights into Tagalog: Reduplication, Infixation, and Stress from Nonlinear Phonology*. Dallas, Texas: The Summer Institute of Linguistics and the University of Texas at Arlington.

Green, Anthony. 1995. Word, Foot, and Syllable Structure in Burmese. *Working Papers of the Cornell Phonetics Laboratory,* 10. 67–96.

Jessen, Michael. 2001. Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. *Distinctive Feature Theory*. ed. by Allan T. Hall. 235-294. Berlin: Mouton de Gruyter.

Larish, Michael. 1999. *The position of Moken and Moklen within the Austronesian language family*, Department of Linguistics, University of Hawaii at Manoa: Ph.D. Dissertation.

Lewis, M. Blanche. 1960. Moken text and word-list. A provisional interpretation. *Federation Museum Journal,* M. C. ff Sheppard (ed.). Kuala Lampur: Museum Department, Federation of Malaysia.

Makboon, Sorat. 1981. *A survey of sea people's dialects along the west coast of Thailand*, Institute of Language and Culture for Rural Development, Mahidol University: M.A. Thesis.

Matisoff, James A. 1973. *Tonogenesis in Southeast Asia. Consonant Types & Tones*, ed. by Larry M. Hyman. Los Angeles: The Linguistic Program, University of Southern California.

-----. 2001. Genetic versus contact relationship: Prosodic diffusability in South-East Asian languages. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. 291-327. Oxford: Oxford Univeristy Press.

Naw Say Bay. 1995. The phonology of the Dung dialect of Moken. Papers in Southeast Asian Linguistics No. 13: *Studies in Burmese Languages*, ed. by David Bradley, 193-205. Canberra: Department of Linguistics, Research School of Pacific and Asian Studies, The Australian National Univeristy.

Pittayaporn, Pittayawat. ms. *Kammu minor syllable: syllable structure, tone, and infixation.* (unpublished term paper).

Posner, Rebecca. 1996. *The Romance Languages.* Cambridge: Cambridge University Press.

Ratliff, Martha. 1992. *Meaningful tone: a study of tonal morphology in compounds, form classes and expressive phrases in White Hmong.* DeKalb, Ill.: Northern Illinois University, Center for Southeast Asian Studies.

Ross, Malcolm. 1992. The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics*, 31.23-64.

-----. 2003. Diagnosing prehistoric language contact. *Motives for Language Change.* ed. by Raymond Hickey. 174-198. Cambridge: Cambridge University Press.

Seong, Teoh Boon. 2003. *The Sound System of Malay Revisted.* Kuala Lampur: Ministry of Education, Malaysia.

Sidwell, Paul J. 2000. *Proto South Bahnaric: A reconstruction of a Mon-Khmer language of Ind-China.* Canberra: Pacific Linguistics.

Svantesson, Jan-Olof. 1983. *Kammu Phonology and Morphology.* Lund, Sweden: Liber Forlug.

Swastham, Pensiri. 1982. *A description of Moklen: A Malayo-Polynesian language.* Institute of Language and Culture for Rural Development, Mahidol University: M.A. Thesis.

Teoh, Boon Seong. 1994. *The Sound System of Malay Revisited.* Kuala Lumpur: Dewan Bahasa dan Pustaka, Ministry of Education, Malaysia

Thomason, Sarah, and Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics.* Berkeley; Los Angeles, Oxford: University of California Press.

Thurgood, Graham. 1999. *From ancient Cham to modern dialects : two thousand years of language contact and change: with an appendix of chamic reconstructions and loanwords.* Honolulu: University of Hawai'i Press.

Wolff, John. 1993. *Proto-Austronesian stress. Tonality in Austronesian languages*, ed. by Jerold A. Edmondson and Kenneth J. Gregerson, 1-15. Honolulu: University of Hawaii Press.

-----. 1995. *The development of the passive verb with pronominal prefix in Western Austronesian Languages. Reconstruction, Classification, Description: Festschrift in honor of Isidore Dyen*, ed. by Bernd Nothofer. 15-40. Hamburg: Abera.Jerold A. Edmondson and Kenneth J. Gregerson, 1-15. Honolulu: University of Hawaii Press.

-----. 1999. The monosyllabic roots of Proto-Austronesian. *Selected Papers from the Eighth International Conference on Austronesian Linguistics.* ed. by Elizabeth Zeiton and Paul Jen-kuei Li. Taipei, Taiwan: Institute of Linguistics (Preparatory Office), Academia Sinica.

-----. 2002. The sounds of Proto-Austronesian. *The 9th International Conference on Austronesian Linguistics.* Australian National Univerity Canberra, Canberra Australia.

-----. in progress. The PANSORT Database for the reconstruction of PAn.

Zorc, David R. 1978. Proto-Philippine word accent: innovation or Proto-Hesperonesian retention? *Papers of the Second International Conference on Austronesian*

Linguistics, Fascicle 1, ed. by S. A. Wurm and Lois Carrington, 67-119. Canberra: Department of Linguistics, Research School of Pacific and Asian Studies, The Australian National Univeristy.

-----. 1993. Overview of Austronesian and Phillippine accent patterns. *Tonality in Austronesian languages*, ed. by Jerold A. Edmondson and Kenneth J. Gregerson, 17-24. Honolulu: University of Hawaii Press.

Appendix

Forms with mid vowels

e		o	
bate:	kind of sea-shell	baso:	bad-smelling
bateʔ	female's skirt somewhat like a tube skirt	bikoŋ	kind of shark
bileʔ	room	boy	part of boat
biʔe:ŋ	you (second person singular)	caho:m	(of a tree) shade, shadow
kate:	chair	cakoy	digging tool
kaʔe:	piece of wood	chəlo:ŋ	(of wood) stick
kəle:	good friend	chəloy	kind of wild animal
khiem	beside	coy	I (first person singular)
lale:	in vain	gayo:ŋ	tall, high
lase:	book	ho:ŋ	money
le:	wheel	kakoŋ	kind of vegetable
mane:	to talk in one's sleep	kaso:t	pair of shoes
məle:	to move to a new place, to migrate	kɔloŋ	red sea-slug
mɛle:t	to move to another place	kəbon	place, a garden
mɛse:t	to move slightly	kho:m	as someone pleases, depend on somebody
ŋɛlep	to step aside, to make way for	khoy	used to - a modal
pace:	to whisper	laŋo:ŋ	bittern
pade:	to hiccup	laʔo:	hot
pɔle:	cot	lɔboy	the Moken spirit posts and houses
pɛle:	squint-eyed	ləŋoy	to raise one's head
phage:	(of time) next	ləpho:ŋ	to charge, to accuse
phe:	to be defeated	mɔŋo:	kind of rock
tile:	fortune teller	mikho:m	bowl
ʔahek	for a moment	miyo:y	to pull, to tug
ʔaphe:	(of place) other	nano:	to gore
ʔɔke:n	sea	ŋo:k	to bully
ʔɔte:t	cape	pho:ŋ	to bloom
		po:	to exceed, to be in excess of
		sɛsoy	beside
		tabo:t	kind of lobsters

e	o	
	tɔko:k	earthern jar
	tɔŋo:k	stump
	təŋok	to sit
	to:	very, more than
	toy	kind of game
	yəpon	Japan, Japanese
	yiʔo:	radio
	yiʔoy	winnowing basket
	yuloy	to ride on the back
	ʔaŋo:	to nod the head
	ʔayoy	sunshade, shadow
	ʔo:t	to give an answering call
	ʔubo:t	first time
	ʔugot	to threaten
	ʔutoŋ	benefit, tax

6 *Acehnese and the Aceh-Chamic Language Family*[1]

Paul Sidwell

1. Introduction

The starting point for this paper is the treatment of Acehnese as a Chamic language by Thurgood (1999) (henceforth 'Thurgood'). While many scholars (e.g. Niemann 1891, Cowan 1933, 1948, 1974, 1981, Shorto 1975, 1977, Collins 1969, 1975, Blust 1981, Durie 1990 and others) have noted that, although widely separated geographically (Aceh in northern Sumatra and Champa centred in Vietnam), Acehnese and Chamic form a genetic sub-grouping. Thurgood is explcit in treating Acehnese as a descendent of Proto-Chamic (PC), specifically as the first dialect to separate from a more or less united Chamic speech community, sometime late in the 1st millennium CE. However, scholarly views on the precise nature of the Aceh-Chamic relationship vary, with no clear consensus on the likely date of separation of the Aceh-Chamic speech community.

Thurgood's monograph length study has revealed the extent to which Chamic was relexified by borrowings, particularly from Mon-Khmer, from ancient through to modern times. Earlier studies, such as Headley (1976), had suggested that around 10% of the reconstructable Proto-Chamic vocabulary was borrowed from Mon-Khmer (MK), while Thurgood's work indicates that the real proportion is perhaps more than three times that, with around 40% of the Proto-Chamic basic lexicon replaced by borrowings of one source or another. Yet for many of these borrowings it is difficult to clearly identify a specific source, not withstanding their frequent co-occurrence in neighbouring Bahnaric languages. My comparative and distributional analyses indicate that the mass of lexicon shared between Chamic and Bahnaric (and to some extent Katuic), is almost entirely borrowed from Chamic into Bahnaric, which implies that they formed a language area at a somewhat later phase, rather than from the outset of Chamic settlement.

My hypothesis, presented in this paper, is that Chamic and to a lessor extent Acehnese, preserves a "substratumised" branch of Mon-Khmer[2] that is otherwise unattested and now extinct—presumably the result of a language shift. The substantial body of borrowed lexicon reconstructable to Proto-Chamic (according to Thurgood) is very difficult to etymologise, and it is clear that there is a very old stratum that has no source in any known languages. A much smaller proportion of this stratum is shared with

[1] There are many people who have assisted me with advice and support as I have researched the history of MK-AN language contact. In particular I would like to thank the Max Planck Institute (Leipzig) and the Australian Research Council for financial support, and the Australian National University for providing me with an office and some administrative and financial assistance, not to mention a supportive academic environment. I would also like to thank Anthony Grant, Graham Thurgood and Malcolm Ross for their comments on drafts of this paper.

[2] Please forgive the echos of the late Paul Benedit's (1976) imaginative hypothesis for explaining certain lexical aspects of his "Austro-Thai" hypothesis.

Sidwell, P. "Acehnese and the Aceh-Chamic language family". In Grant, A. and Sidwell, P. editors, *Chamic and Beyond: Studies in mainland Austronesian languages.*
PL-569:211-246. Pacific Linguistics, The Australian National University, 2005. DOI:*10.15144/PL-569.211*

Acehnese, so logically the separation of Aceh-Chamic occurred sometime during the substratumisation process. The pre-Acehnese must have moved away from the zone of language contact, in constrast to Dyan's (2001) that Aceh-Chamic orignated in Sumatra with the Proto-Chams moving on to Indo-China. Clearly Aceh-Chamic originated with initial settlement on the Indo-Chinese coastline, followed by the splintering off of the Acehnese.

Well after the separation of Acehnese there were other phases of significant MK influence upon Chamic, principally by Khmer, Mon and Vietnamese. Probably much of it was associated with historical events that led to the decline of Champa and the differentiation of Chamic into Coastal and Highland branches. The earliest and later contact phases must have been quite separate, as we find no identifiable traces of the oldest loan stratum exist elsewhere beyond mainland Chamic and the Mon-Khmer languages of the Annamite Range that came under strong Chamic influence.

We may speculate that some great historical event, perhaps a great political conquest, saw a foreign population absorbed completely into the nascent Champa, leaving no direct ancestor elsewhere in Indo-China. Alternatively the substratum may simply have been the language of the autochrones of the Indo-Chinese coastal plains that were first encountered, and then absorbed, by pre-Aceh-Chamic settlers. My favoured speculation is that we might connect the more obscure lexical stratum in Chamic with the mysterious kingdom of Funan, an ally of early Champa that was ovetaken by the pre-Angkorian Khmer Chenla (Zhenla) around the middle of the first millennium. I dare not pretend to have positively identified the "language of Funan"—presumably the name refers only to the political centre that ruled over an ethnically complex region—but one can claim at least to have identified a specific line of investigation.

Finally, from a programmatic perspective, I suggest that it is appropriate to build upon the solid foundation of Thurgood's data and analyses by drawing in more extensive sources, especially Mon-Khmer, to rework the reconstruction of the respective phonologies and lexicons of Proto-Aceh-Chamic and Proto-Chamic. A more extensive etymological compilation and stratification of the lexicon offers prospects for revealing the history underlying the remarkable contact-driven change which occurred in the Aceh-Chamic languages. It is also significant that, if as I suggest, the Acehnese have constituted an independent society for the better part of 2000 years, there will be historical implications for migration and settlement that other disciplines may be able to shed some light upon.

2. Malayo-Chamic

Thurgood approvingly cites Blust (1994) identifying a Malayo-Chamic (MC) subgrouping within Proto-Malayo-Polynesian (PMP), which split into Malayic and Chamic branches (see Fig. 1, below) sometime in the first Millenium BCE.

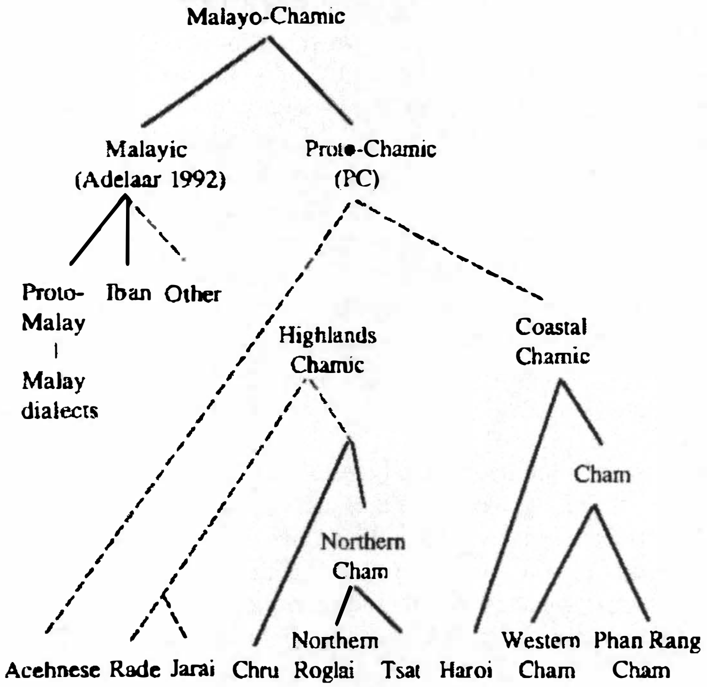


Figure 1: _Thurgood's Figure 6: the Malayo-Chamic Languages (p.36)_

Three principal sound changes that mark the formation of Proto-Malayo-Chamic (PMC) are discussed: 1) PMP *_R_ > PMC *_r_, 2) PMP *_w_- > PMC *_Ø_-, 3) PMP *_q_ > PMC *_h_:

1) PMP *_R_ > PMC *_r_, e.g.:
PMP *_Rusuk_ 'ribs', Malay _rusuk_, Aceh. _rusoʔ_, PC *_rusuk_
PMP *_daRaq_ 'blood', Malay _darah_, Aceh. _darah_, PC *_darah_

2) PMP *_q_ > PMC *_h_, e.g.:
PMP *_qataj_ 'liver', Malay _hati_, Aceh. _ʔate_, PC *_hataj_
PMP *_daqih_ 'forehead', Malay _dahi_, Aceh. _dhɔə_, PC *_ʔadhɛ̃j_
PMP *_baseq_ 'wet', Malay _basah_, Aceh. _basah_, PC *_basah_

3) PMP *_w_- > PMC *_Ø_-, e.g.:
PMP *_waRiH_ 'sun/day', Malay _hari_, Aceh. _ʔurɔə_, PC *_hurɛj._
PMP *_wakaR_ 'root', Malay _akar_, Aceh. _ʔukhɯə_, PC *_ʔughaar_
PMP *_wahiR_ 'water', Malay _air, ayer_, Aceh. _ʔiə_, PC *_ʔiar_

In the case of word initial PMP *_q_ the Acehense reflex is /ʔ/ which requires a sequence *_q_ > *_h_ > *_ʔ_. This initial glottal stop is not usually written in transcription, as it is predictable, a phonotactic artifact. This is also the occasional reflex in Malay, e.g. _abu_ 'ashes' < PMP *_qabu._

The loss of initial *_w_- is interesting as there appears to be a trace of it in the labial quality in the Aceh-Chamic minor-syllable[3] vowel, which shifted to /u/. At this point I caution the reader that I am approaching the topic of Austronesian historical phonology as an outsider, but it seems logical to me that the syllable *_wa_- must have been present at the PMC level, since a simple *_a_ would not have unconditionally shifted to [u] in Aceh-Chamic, any more than a secondary *_u_ would have unconditionally shifted back to [a] in Malayic. In the case of PMP *_wahiR_ 'water' an earlier regular loss of *_h_ resulted in a change of syllable structure that eliminated the minor-syllable, creating a diphthong, so there was no eligible vowel to labialise (note that Aceh-Chamic metathesised the resultant diphthong). Strikingly the 'sun/day' etymon shows special evidence of connection with Malayic—sharing the otherwise uniquely Malayic addition of an initial [h]. If it was a simple loan from Malay(ic) we would not expect the [u] vowel, so we are left to suggest some kind of contamination was caused by a knowledge of Malay(ic) among Aceh-Chamic speakers.

The above changes are not uniquely restricted to MC among MP: *_q_ > [h] also occurred in Balinese, Javanese, Sundanese and Batak, and the merger of *_R_ and *_r_ and the loss of *_w_ also occurred in Batak and Balinese. In these circumstances Blust's phonological arguments for MC also suggest that parallel changes elsewhere in MP were independent, and we may wonder why their occurrence in Malayic and Aceh-Chamic is not similarly coincidental, particularly in the light of the necessarily independent development of Aceh-Chamic *_ʔu_- < * _wa_-.

To the phonological data we can add the innovations among the numerals. Thurgood (p36-39) provides a detailed discussion of these, showing how Malayic and Aceh-Chamic replaced the PMP forms for 'seven', 'eight' and 'nine' with new words, the latter two based upon subtractive formulations. Thurgood concedes that the innovated 'eight' and 'nine' forms also occur in Maloh and Rejang, although Blust (1992) cautions that this "may be due to borrowing". One may also wonder whether the ancient Aceh-Chamic also acquired the new numeral forms by borrowing from Malayic.

My brief review of the Malayo-Chamic hypothesis leaves me with the strong impression that it does not demonstrate a very neat process of separation and branching such as we might like to see in a phylogentic model—instead it suggests a much messier (yet perhaps more realistic) dialect chain that saw prolonged contact and mutual influences, as sub-groupings emerged and population movements occurred. This is quite a normal thing in the real world, but we are still at a loss to understand the specific historical consequences this may have had for the place of Aceh-Chamic vis-à-vis Malayic, and the version of Malayo-Chamic I am relying upon in this paper. For now I do not wish to argue for any particular alternative to Blust's MC, as I am concerned with the Aceh-Chamic hypothesis in particular, but it is clear that the issue deserves further examination.

[3] The term 'minor-syllable' is used by Mon-Khmerists to designate the initial syllable within the typically MK phonological word pattern that maximally permits only iambic structures, with strong restrictions on which segments may occur in the initial syllable.

3. Aceh-Chamic

3.1 Phonological Innovations

We now turn to the issue of the relationship of Acehnese to Chamic. Restricting matters to the etymologically Austronesian material, Thurgood states that in Chamic and Acehnese the following changes occurred:

1) PMP **n-* > **l*;
2) PMP *-*r* > **Ø*;
3) PMP *-*i*, *-*u* > *-*ɛj*, *-*ɔw*, and later to [-ɔə, -ɛə] in Acehnese;
4) PMP stressed **a*, **e* (ə) > **aa*, **a*
5) Unstressed PMP initial syllables are reduced to clusters according to the same underlying patterning;
6) Imploded stops developed in some PMP etyma, reflected as /ʔ/ in Acehnese;

We will now discuss each of these in detail.

1) PMP **n-* > **l*. Two examples showing /l/ in Acehnese are adduced: PMP **h-in-ipi* 'to dream' > Malay *mimpi*, Aceh. *lumpɔə*, PC **lumpɛj*; PMP **nipis* 'thin' > Malay *nipis*, Aceh. *lipeh*, PC **lipih*. Blust (2000) challenges both of these comparisons. In the first it is not clear that etymological **n-* is the source of /l/, it is at least as likely the source of the nasal in the [mp] cluster, which case the /l/ is unexplained. The shift of **n-* > **l* in the 'thin' etymon is phonologically straightforward, although it may have been borrowed into Acehnese from Moklen/Moken (if not Chamic), which also shifted PMP **n-* > **l*, cf. MoklenLmp *lipih* 'thin (things)', MoklenKY *lipɯj* 'to dream'. Other apparent loans from Moklen/Moken are discussed below. An important counter example to this proposed sound change exists in the etymon for 'coconut': PMP **niuR* > Malay *nyiur*, Aceh. *bɔh ʔu*, PC **ləʔu*, where Acehnese and Chamic share the same loss of final and blocking of diphthongisation, but Acehnese has lost the initial lateral, rather than shifting it to /n/ (or potentially to [ɗ] if we accept the arguments concerning implosives, see below). There are at least two examples of this change which lack Acehnese forms: PC **lanah* 'pus' < PMP **nanaq*; PC **lasɛj* 'rice (cooked)' cf. Malay *nasi*. The limited comparisons we have seem to establish the general rule of PMP **n-* > **l* in Chamic, but we have only one reasonable example in Acehnese, and it is far from clear how it acquired the form, so it may be actually be a post-Aceh-Chamic change.

2) PMP *-*r* > **Ø*; this is a change that has occurred among other Mainland SEAsian languages, perhaps most importantly in Khmer (although other changes are also common, e.g.: /-r/ merged with /-n/ in Thai/Lao and with /-j/ in Vietnamese). In Aceh-Chamic the loss must have occurred after the diphthongisation of open syllable **u* had ceased to operate, i.e.: PMP **ikuR* 'tail' > Malay *ĕkor*, Aceh. *ʔiku*, PC **ʔiku*. Thurgood seems to be a little confused about the reconstruction of this final *-*r*, positing it in some proto-forms but not others, e.g. it is absent in his PC **ʔiku* 'tail', but it is present in his **ʔular* 'snake'. The change is common to both Acehnese and Chamic, so it properly belongs to the Proto-Aceh-Chamic level if it is not an independent change, although it must have occurred later, rather than earlier, in their unity.

3) PMP **-i, *-u* > **-ɛj, *-ɔw*, and later to [-ɔə, -ɛə] in Acehnese. E.g.: PMP **beli* 'buy' > Malay *bĕli*, Aceh. *blɔə*, PC **blɛj*; PMP **balu* 'widowed' > Malay *balu*, Aceh. *balɛə*, PC **balɔw*. Thurgood reconstructs the Acehnese /ɔə, ɛə/ deriving from PC **ɛj, *ɔw* (respectively) by dissimilation of vocalic onsets followed by neutralisation of final glides. This is a significant change that did not occur in Malayic, although it did occur in some other MP languages, in particular Moklen/Moken. Thurgood (p.58-59) takes pains to point out that the outcome of the diphthongisation in Moklen/Moken is different to Chamic, and therefore he considers it to be unrelated. However, Larish (1999:395-402) discusses the reconstruction of the diphthongisation in Moklen/Moken in considerable detail, arguing for precisely the same initial path of development as Thurgood posits for Chamic, namely a sequence: PMP **-i, *-u* > **-ɪi̯, *-ʊu̯* > **-ɛi̯, *-ɔu̯*, subsequently followed by dissimilations and mergers that ultimately yielded *-ɔj* ~ *-əj* and *-uj* in Moklen/Moken. The parallelism is remarkable, especially given the fact that Aceh-Chamic and Moklen/Moken do not subgroup genetically. What they have in common is their geographical location on the Asian Mainland, with the influence (to a greater or lesser extent) of Mon-Khmer languages (and others). Thus, while this kind of diphthongisation is otherwise rare or unknown in MP languages, it is common in MK, Cf. Khmer *dəj* 'hand' < **tii*.[4] Perhaps, given their apparent geographical separation, it was simply that under mainland influence the shift to fixed final stress set these processes on track, following parallel paths for reasons that are closed tied to universal phonetic processes. In that case Thurgood is correct to conclude that the diphthongisation in Moklen/Moken is genetically unrelated to that in Chamic, but the same argumentation works against the conclusion that Acehnese and Chamic must have derived these diphthongs together as one proto-language. The strongest evidence that they likely did is in the reflexes of words with final **ɯr* rhymes. As discussed above, the common loss of final **-r* must have occurred after the diphthongisation process had ceased to be productive, and therefore occurred before the separation of Aceh-Chamic, assuming that the loss was not itself also independent.

4) PMP **a*, **e* (ə) >* *aa*, **a* in Aceh-Chamic, with later diphthongisation of *aa to /ɯɪə/ in Acehnese closed syllables. E.g.: PMP **qudaŋ* 'shrimp' > Malay *hudang/udang*, Aceh. *ʔudɯəŋ*, PC **hudaaŋ*; PMP **halem* 'night' > Malay *malam*, Aceh. *malam*, PC **malam*. The same shift occurred in Moklem/Moken (Larish 1999), and the lowering of PMP **e* (ə) > /a/ was the normal result in most Malayic dialects (Adelaar 1992). Much ink has been spilled discussing the issue of the long /aa/ in Acehnese and Chamic. Writers such as Shorto (1975) and Cowan (1983) saw in it evidence of a much older, perhaps ProtoAN length distinction, an idea that has not survived closer examination. Clearly we are seeing an areal drift, again connected to some extent with the shift to final stress, and reinforced by contact with languages that already have length as an important component of their phonologies. It is apparent that the lengthening of PMP **a* > **aa* must have completed before PMP **e* (ə) > **a* to have prevented their merger. This clearly places these shifts before the separation of Aceh-Chamic, and we should probably treat them as a common inheritance in Aceh-Chamic.

[4] Note that this example of diphthongisation in Khmer is not related any devoicing of the initial consonant and is unrelated to the Middle Khmer register split.

5) Thurgood reconstructs PC word-initial consonant clusters of the types Cr/Cl/Ch, some of which are derived from reduction of initial syllables of AN disyllabic words, while others occur in borrowed vocabulary—Thurgood refers to them as "primary clusters". The former are attested as clusters in Acehnese and all Chamic languages, so their formation belongs to the earliest stage of the proto-language. Not all AN disyllables with medials /r,l,h/ reduced to clusters in this process: compare PMP *_beli_ 'buy' > Malay _bĕli_, Aceh. _blɔə_, PC *_blɛj_ with PMP *_balu_ 'widowed' > Malay _balu_, Aceh. _balɛə_, PC *_balɔw_. Thurgood does not offer an explanation of the distribution of reduced and non-reduced forms—although the presence of unstressed schwas in the first syllable of many of the relevant forms at the PMP level suggests a phonetic rule which is yet to be formulated. The point is that Acehnese and Chamic agree exceptionlessly in terms of the etyma that do and do not show the reduction to clusters. So although this kind of change is widespread in Mainland SE Asia, including within MP (including spoken Malay, not withstanding Malay authography[5]), the distribution across a specific restricted set of etyma strongly indicates an equivalent of a "Werner's Law" for Aceh-Chamic.

6) In at least two AN etyma imploded stops developed in Chamic, with /ʔ/ reflexes in Acehnese, e.g. PMP *_buhuk_ 'hair', Proto-Malayic *_buØ(uə)k_ (< PAN *_buSék_), Aceh. _ʔoʔ_, PC *_ɓuk_; PMP *_nahik_ 'climb'> (Proto-Malayic *_naØik_ ?) Malay _naik_, Aceh. _ʔeʔ_, PC *_ɗiʔ_, and rather speculatively PMP *_hideRaq_ 'lie down' > Aceh. _ʔeh_, PC *_ɗih_ (although Thurgood suggests MK origins). All three are rather problematic. Firstly, there are counter examples to the regularity of the 'hair' etymology in the reflexes of PMP *_bahu_ 'stench' > Malay _bau_, Aceh. _bɛə_, PC *_bɔw_, PMP *_bahut_ 'do' > Malay _buat_, Aceh. _buət_, PC *_buat_, indicating that AN medial *-_h_- is exceptionally, rather than regularly, reflected as *-_ʔ_- in Malayo-Chamic. Although the received view (since Lee 1966) is that PC *_ɓuk_ reflects a sporadic fusion of /b/ and /ʔ/, by implication it also requires the sporadic persistence of *-_ʔ_- in Malayo-Chamic.

Thurgood compares PC *_ɗiʔ_ 'climb' to Bahnar _dək_ 'go up' (citing Cabaton 1901, note that Banker et. al. 1979 gives the form as _dak_). One can also compare to Proto-Katuic (Sidwell 2005) *_dɨk_ 'lift up, raise', although these may not be helpful—the Katuic and Bahnaric suggest a prototype *_dak_, which simply does not correspond to the Chamic form. On the other hand there no problem deriving Acehnese _ʔeʔ_ from PMP *_nahik_ in the light of examples such as PMP *_niuR_ > Malay _nyor_, Aceh. _bɔh ʔu_, PC *_ləʔu_. The problem is how to account for the implosive initial in Chamic, and similarly the received view is a sporadic fusion of /n/ and /ʔ/.

We do not have an obvious AN etymology for Aceh. _ʔeh_, PC *_ɗih_ 'lie down', although they could reflect a radical simplification of the trisyllabic PMP *_hideRaq_. I have yet to find a convincing mainland source—among MK languages Khasi _thiah_ 'lie down, sleep' potentially corresponds, but the geographical distance makes it a remote prospect, while Khmer _dek_, compared by Cowan, is phonologically too different (and probably ultimately related to Khasi _thiah_).

In addition to the above three sets with Acehnese reflexes, Thurgood reconstructs 12 PC words with initial *_ɓ_ and 10 with initial *_ɗ_ that lack Acehnese reflexes—all 22 are borrowings, which must have been acquired after the separation of Acehnese. So we have three words in which Chamic implosives correspond to Acehnese /ʔ/, but we don't know

[5] Drawn to my attention by David Gil in 2001 during a visit to the Max Planck Institute (Leipzig).

whether there was a shift of imploded stop to glottal stop in Acehnese, or a simple loss of initial syllable from a disyllabic PAC form.

On balance there are several phonological developments that solidly belong to a phase of Aceh-Chamic unity—the formation of Primary Clusters, the diphthongisation final **-i* and **-u* and the lass of final **-r* which followed the diphthongisations. To these phonological changes we can add the lexical innovations—borrowings—common to Acehnese and Chamic.

3.2 Lexical Innovations

In this section I discussus the data and results of two significant publications dealing with the sources of borrowings in Aceh-Chamic: Cowan (1948) and Thurgood (1999). Additionally I would have have liked to make use of Collins' (1975) PhD thesis on the sources of Acehnese vocabulary, but access to that work is restricted[6].

Cowan's 1948 paper made a fundamental contribution to discussion of the classifiation and history of Acehense half a century before Thurgood's recent synthesis appeared. Cowan discuses at length the position of Cham and Acehnese in respect of Austronesian, adducing many lexical comparisons with Malay. He groups Cham and Acehnese genetically on the basis of parallels in the phonology, morphology, lexicon and syntax, and interestingly contrasts them in respect of the use of pronouns and the "passive" voice (see Durie 1985 for a detailed analysis of Acehnese argument structure). Significantly for our present purposes, Cowan presents a list of 150 comparsions with mostly Mon and Khmer: of these I count 43 that can be confidently identified as MK loans into Acehnese, and perhaps another 60 into Aceh-Chamic, while the balance are put aside as either defective comparisons, misidentified Austronesian or other loans, imitative forms, or loans into MK languages from Chamic. A summary of Cowan's numbered examples thus excluded is at the end of Appendix 1. Of Cowan's MK loans into Aceh-Chamic, I count 17 sets not included in Thurgood's published data-set, which suggests that he did take full advantage of Cowan's contribution. This might seem a modest number at first, but in fact the total number of Thurgood's putative MK borrowings with an Acehenese reflex is modest—only some dozens—and is fact is given considerable attention in the following analysis.

Thurgood identifies some 277 Proto Chamic words of Mon-Khmer origin and another 179 of uncertain origin. One way or another we assume that the bulk of these are borrowings, although conceivably some are neologisms invented by Aceh-Chamic speakers. Dyen, in his 2001 review of Thurgood, expressing considerable scepticism about the Aceh-Chamic hypothesis. He pointed out that if Acehnese is descended from PC, it should preserve a substantial proportion of the borrowings reconstructable to PC, yet he counted only 44 Acehnese reflexes among the hundreds of PC items of MK origins. Reasoning further that those words also having Malay reflexes could well have diffused from Malay, only "twenty-eight entries, perhaps better reduced to twenty-six, then appear

[6] Durie (1975:3) reports Collins' conclusion that Acehenese "had contact with Old Mon, the Aslian languages of the Malay peninsula, and the languages of the Nicobar islands". In my own investigations so far I have found no particular lexical or structural features among the MK component in Acehnese that would identify an Aslian or Nicobaric source. I believe that this is consistent with the homeland of Aceh-Chamic being in Indo-China, and the reletively marginal importance of Aslian and Nicobaric in the trade networks of western Austronesia.

to constitute the basis of the hypothesis that Acehnese is a Chamic immigrant". In other words, only 10% of PC words of MK origin have Acehnese reflexes.

This is a very significant discrepancy. If Acehnese is a descendant of PC, it should reflect PC vocabulary pretty well as much as any Chamic language (subject to extraordinary social/historical factors). Furthermore, if Acehnese is the first branch of the Chamic family tree, the principal criteria for reconstructing a non-AN word to the PC level should be its attestation in at least Acehnese and one other Chamic language. Yet we have gross indications that Acehnese shares relatively few borrowings with the rest of Chamic, a fact that suggests that Acehnese separated before the bulk of borrowings into Chamic occurred.

Reviewing Dyen's count it seems that he did not consider the complete corpus of data presented by Thurgood—but ignored the words classified as of uncertain origin. I have made my own count combining both indices and the results are summarised as follows:

1. 16 borrowings also reflected in Malay
2. 7 words apparently borrowed separately into Acehnese and Chamic
3. 3 isoglosses with Moklen/Moken, origin and direction of borrowing uncertain
4. 28 AC borrowings of MK origins
5. 12 AC borrowings of unknown origins

1.)

Semantic	Aceh.	P-Chamic	Malay	Comment
'bean, pea'	rɯtɯəʔ	*rətaak	(Iban retak)	Cf. Khmer sandaek
'bitter'	phet	*phit	pahit	< Skt. pitta
'bowl, dish'	piŋan	*piŋan	pinggan	< Persian (> Bah.)
'branch, fork'	cabɯəŋ	*caɓaaŋ	cabang	>Aslian, Cf. Kh. ɟəŋrmaaŋ[7]
'broken, break'	picah	*picah	pĕcah	> Bah.
'buffalo'	kɯbɯə	*kabaw	kĕrbaw	> Bah.
'cotton'	gapɯəh	*kapaas	kapas	< Skt. karpaasa
'cow, ox'	lɯmɔ	*ləmɔ	lĕmbu	Cf. Khmu lmboʔ
'eggplant'	truəŋ	*trɔŋ	tĕrung	> Bah.
'form, image'	rupa	*rupa	rupa	< Skt. rupa
'g-grandchild'	cʌt	*cicɛ̃t	cicit	
'gold'	mɯh, mɯih	*ʔama(a)s	emas	> Khmer, Bah. < ?
'lizard, gecko'	cicaʔ	*cicaʔ	cicak	Cf. Mon həcɛk (imitative)
'net (casting)'	ɟɯə	*ɟaal	ɟala	< Skt. jāla
'pillow'	bantaj	*bantal	bantal	
'pineapple'	bɔh ʔanɯh	*manaas	nanas	< Portuguese

Group 1 is an etymologically heterogenous set of borrowings that fall mostly into two main types, Indic words that probably began to be diffused by traders even before the Common Era, but particularly from the middle first millennium (as Indic scripts and other

[7] 'forked stick'

cultural features were widely adopted), and MK words, some of which have clear etymologies, others identified on structural grounds that are inferred to be MK. A good example is Malay *kĕrbaw* 'buffalo'—close matches are found in Bahnaric and Katuic, but the Khmer reflex is *krəbɤj*, which shows phonological differences that eliminate it as the source. The other bovid term, reflected in Khmu *lmboʔ,* Bahnar *ləmɔɔ,* Vietnamese *bò*, is drived from PMK **[ ]bɔʔ* 'hump of ox' by Shorto (ms.) based on reflexes in Mon and Khmu. Speculatively the *kĕrbaw* word could have originated from the same root, assuming borrowing from a hypothetical MK language having lost the final glottal and added the small animal velar prefix (not uncommon changes in EMK).

Another interesting etymon is the 'gold' word. On the mainland it is restricted to languages historically in contact with Chamic, which suggests borrowing into MK, but that still leaves the question of its source in MP. An MK root **jaas* 'to shine' is reconstructable on the basis of widely distributed reflexes, and a hypothetical derivation via the *-m-* agentive infix in pre-Mon (cf. Old Mon /jimaas/) could have subsequently diffused with the very sought after trade item.

At this stage the main point I would like to make about these comparisons is that the borrowing of MK words into Malayic likely did not reflected a discrete historical process that might be localised in time or space. It is evident that the borrowings range from relatively recent Khmer, Mon and Vietic loans to very ancient times. Whatever the case Dyen is correct to set these aside from any discussion of Chamic sub-grouping.

Group 2 consists of words for which we have indications of independent borrowing of related or unrelated but similar forms:

2.)

Semantic	Aceh.	P-Chamic	Comment
'flesh, meat'	*siə, ʔasɔə*	**ʔusar*	Aceh. related to Malayic **isi*
'fly (v.)'	*phʌ/pʌ*	**pər*	PMK **par*. Anomalous aspiratred initial also found in Rade: *phiər* (Durie 1990)
'open (mouth)'	*hah*	**ʔaha*	PMK **haʔ, hah*, Ach. resembles B. & Viet.
'python'	*lhan, tlan*	**klan*	PMK **tlan* - Aceh. borrowed with apical initial; Chamic < form with velar initial
'strong, hard'	*kʌŋ*	**khaŋ*	Comp. Aceh. to Katuic**kəŋ,* Khmer *kɛ̀əŋ* (& Thai *khàŋ*) suggest **gaŋ*. Chamic < Vietnamese **khăŋ*
'wash'	*rhah*	**raw*	Cf. Viet. *rửa* (< **raah*), Katuic/Bahnaric **ʔəraaw*
'yawn'	*sɯmɯŋɯp*	**həʔaap*	PMK **sʔaap*, **sŋʔaap*, not all MK sub-groups have medial nasal

Group 2 items all show clear phonological indications that Acehnese and Chamic borrowed related forms from different MK sub-groups. This is quite understandable as lexical borrowing continued after separation, and therefore these forms are not relevant to the sub-grouping issue.

Group 3 is quite intriguing:

3.)

Semantic	Aceh.	P-Chamic	Proto-Moken/Moklen
'naked'	_lhon_	*_(ma)(sa) lun_	*_ɲɯlɔn._ No wider etymology apparent.
'urinate'	_ʔiəʔ_	*_maʔiãk_	*_niʔa̱ak_ >Pre-Moklen *_niʔiək_ < PMP *_[ ]iSeq_
'gecko'	_paʔɛɛ_	*_pak-kee_	*_tɔkɛ̱ɛʔ_, imitative word?

The phonological agreements in the first two sets above are excellent, and strongly suggest ancient contact involving Aceh-Chamic and Moklen/Moken—in particular the development of the diphthong in the 'urinate' etymon indicates Moklen/Moken as the source. Larish reconstructs the Moklen/Moken homeland as the Isthmus of Kra, with their marginalised to the islands off the western coast only later. This leaves the possibility of A-C and M-M contact somewhere on the Gulf of Thailand.

Group 4 items are the most numerous, all showing indications of being borrowed from MK:

4.)

Semantic	Aceh.	P-Chamic	MK comparisons
'arm'	_sapaj_	*_sapal_	Found in Asl., Kat., West-Bah.
'back'	_ruəŋ_	*_rɔŋ_	Katuic *_krɔŋ_ 'back' , Khmu _kndrɔɔŋ_ 'back'
'bail'	_sɯət_	*_sac_	PMK *_saac_ (all but Khmu, Asl., Nic.)
'bird'	_cicem_	*_cim_	PMK *_cim_ (all but Khmer)
'carry on sldr.'	_gulam_	*_gulam_	PMK *_klam_ or *_kləm_ (NMK & Aslian)
'chase'	_tijɯəp_	*_tijaap_	Khmu _ŋgjaap_, Ch. > Tampuon _tijạap_
'cheek, jaw'	_miəŋ_	*_miaŋ_	Khmu _miəng_ 'chew'. Vt. _miệng_, < PV *_mɛɛŋʔ_ 'mouth'
'chin, jaw'	_kɯəŋ_	*_kaaŋ_	PMK *_kaaŋ_ (Katuic, Bah., SNic., Vietic) (+ _kmaaŋ_ forms in Pearic, Vietic..)
'cover'	_gɔm_	*_gəm_	Khmer _kaem_ 'cover, encrust, decorate', PVietic *_kəmʔ_ 'to bury'
'crow'	_ʔaʔaʔ_	*_ʔaak_	PMK *_kʔaak_ (all but Khasi, Nic.) – Vietic reflexes typically _ʔaak_, e.g. Viet. _ác_, but such imitative words are problematic.
'cut off'	_kɔh_	*_kɔh_	PMK *_kɔh_ (Bah.,Kat.,Nic.,Asl.)
'dry'	_tho_	*_thu_	Temiar _təhool_, KhmuYuan _thúu_
'dumb'	_klɔ_	*_k-am-lɔ_	Khmer _kamlaw_ 'ignoramus'
'empty'	_sɔh_	*_sɔh_	Khm., Bah., (Katuic infixed forms only)
'escape'	_lhɯəh_	*_klaas_	> Bah., other MK suggests *_laas_ 'leave'
'forget'	_tuwʌ_	*_wər_	PMK *_wər_ 'go round' ? (all MK groups)
'hawk, kite'	_klɯəŋ_	*_klaaŋ_	PMK *_klaaŋ_ (all MK groups)
'house'	_sɯəŋ_	*_saaŋ_	Khmer _saaŋ_ 'to build' (also >Thai/Lao)

'lick'	liəh	*lijah	PMK *liət, also Khasi ɟliah
'mount. range'	cʌt	*cət	Khmer caot 'high, steep, sheer, abrupt'
'neck'	takuə	*takuaj	PMK *kuuj 'head' (Kat., Asl.)
'other, group'	gɔp	*gəp	PMK *gap, gəp 'friend, associate' (Khm., Bah., Viet.)
'peck (snake)'	cɔh	*cɔh	PMK *[ʔ]cɔh (EMK, Khmu, Asl.)
'pillar, post'	tamɛh	*tamɛh	Mon tmit 'post supporting veranda'
'river'	kruəŋ	*krɔɔŋ	PMK *ruŋ, *ruuŋ, *ruəŋ (all but Asl., Nic.)
'stand, stop'	dəŋ	*dʌŋ	Viet. đứng, or perhaps PMK *duŋ 'house'
'strike, pound'	pɔh, pɛh	*pɔh	PMK *pah, *puh, *puəh (NMK, Bah., Viet.)
'wrap'	sɔm	*səm	Old Khmer sum 'to wind, roll, wrap up'

To these we can add the Aceh-Chamic-MK comparisons from Cowan (1948) not used by Thurgood, yet which may be taken as highly indicative of MK borrowing.[8]

Aceh.	Cham	MK Comparisons
hu 'ablaze'	hu 'roast'	Kh. chur 'ignite', Bah. huur 'roast', Katu huar 'singe'
ɟa 'ancestor'	ɟa 'appelative'	OldMon ɲɟaʔ, OldKh. ɟi 'great-grandmother'
baʔ 'at, on'	pak 'at, towards'	OldMon bak 'up to, until'
luəŋ 'channel'	haluŋ 'pit, canal'	Khmer lùŋ 'dig hole', ʔɔnlùəŋ 'hole in stream-bed', Bah. səluŋ 'pit, ditch'
tɔm 'ever'	tom 'meet with, accomplish'	PMK *təm/*təəm/*tam 'begin', e.g. Mon tam /tɔm/ 'base, beginning' (widespread in MK)
ɲum 'flavour'	ɲəm, ɲam 'to taste'	Praok ɲɔm 'to taste', Bahnar ɲaam 'delicious'
wɛh 'go away'	weh 'to dodge'	Khmer veh / vèh/ 'to slip away, escape, dodge'
gət, gɛt 'good'	gɔt 'just'	Khmer gat / kɔt/ 'just, exact'
chen 'affection'	khin 'want, like'	Viet. xin 'beg', Palaung. sin 'desire', OldMon chān /chan / 'to pity/
khem 'laugh'	khim 'smile'	LitMon k'im / kʔim / 'to smile'
bɯŋəh 'morning'	paguh 'morning-light'	Mon peŋuh 'to awaken'
khem 'laugh'	khim 'smile'	LitMon k'im / kʔim / 'to smile'
weŋ 'to pedal'	wiŋ 'turn, whirl'	PMK *wiŋ &c. (with many variants) 'go round'
ʔuət 'polish, rub'	u̯ak 'rub'	Lawa ʔuət 'wipe', Khmu ʔɔɔt 'scrub body'
kuət 'scrape'(C.)	ku̯ac 'dig'	Khmer khvaac, Kensiw kəwɔɟ 'scratch up'
wɯə 'stable,pen'	wa(r) 'yard, stable'	Khmer val / viel/ 'plain, clearing, plaza', Mon wa / wɛ̀a/ 'open space, pasture'

[8] Note that Acehnese forms have been normalised to Daud & Durie (1999), Cham forms are from Cowan, MK comparisons have been corrected/augmented

dəm 'stay o.night'	*dəm* 'id.'	PMK **dəm*, e.g. Mon *døm* /*tɜ̀m*/ 'to lodge'
bʌt 'stretch'	*but* 'twisted'	Khmer *bot*/*pot*/ 'to curve, fold'
cɔʔ 'take, sieze'	*cɔk* 'id'	Khmu *cɔɔk* 'catch (e.g. pig)', *cok* 'take out (e.g. entrails)', WestBahnaric **cɔk* 'take'

Examining the above sets we note no convincing pattern of borrowing from a single dominant source—Khmer and Mon are well represented but this may simply reflect the reliance on those reference material. Some etyma are well distributed across the MK family with no particular phonological clues for their source in Aceh-Chamic (such as 'crow', 'fly', 'hawk'). There are several Khmer isoglosses (e.g. 'cover', 'dumb', 'gold', 'house', 'mountain range', 'wrap') although the lack of wider MK etymology is also suggestive of borrowing into Khmer. And there are several items where the closest MK comparisons are in Northern MK languages, and it is difficult to see how they could be the source of borrowings. It is also significant that there are so very few prospective Vietic or Katuic sources for these words, given Thurgood's suggestion that:

> ...the Acehnese were the most northerly of the Chamic groups, covering an area now populated by, among others, the modern Katuic speakers. (p.42)

This idea appears to be based on the overriding assumption that the break-up of Chamic was driven by one main historical process—the Vietnamese imperial drive southward. The model assumes that as the Acehnese were the first group to break away, they must have been the first to suffer Vietnamese pressure. Logically there are other possibilities to consider, such as a southern origin of Acehnese somewhere in the vicinity of the Mekong Delta/Funan. My problem is that no particular solution appears to be supported empirically by comparative linguistic data. Thurgood bases his claims upon supposed morphological and lexical arguments. The first of these is a comparison of the *tar-*, *t-*, *ta-* prefixes in Katuic with parallels in Austronesian which Thurgood (p240-241) asserts are "too close to be accounted for by mutual inheritance", and suggests that because some lexical borrowing from Chamic into Katuic is attested, the same is likely to explain the morphological parallels. A contra-opinion is offered by Diffloth (1994) who points out that the various MK affixes with parallels in An are actually widespread in MK. He concludes that:

> Ironically, it is the relative poverty of shared vocabulary between Austroasiatic and Austronesian, combined with evident agreement in morphology, that argues for a genetic, and against a contact relationship between the two families. (Diffloth (1994:312)

Thurgood writes (p.240-241):

> Other evidence of a contact with Chamic, particularly into Acehnese, and an apparent Austronesian morphological strata (sic.) in Katu (Reid 1994), which one would presume were due to Chamic influences.
>
> The obvious way to account for the Katuic strata found in Chamic is to assume that Chamic influence extended up along the coast into Katuic territory. Certainly, an examination of the appendix of forms makes it abundantly clear that there are a considerable number of MK forms, attested in the more northerly Katuic but not in the more southerly Bahnaric. Further, many of these are attested in Acehnese. Thus, the

> most likely scenario is to assume that the Acehnese are the descendents of the most northerly group of Chamic speakers.

Consistent with Diffloth above, Reid (1994) makes no claim of borrowed "Austronesian morphological strata in Katu". In his paper Reid compares the Austroasiatic prefixes _pa-_ and _ka-_, which "can be reconstructed with a causative function" with the Austronesian causatives *_pa-_ and *_ka-_, exemplifying the former with examples from Katu. The comparison is explicitly between two language families with consideration of the Austric hypothesis in mind, with much weight given to reflex the of *_pa-_ in Nicobarese.

Thurgood then refers to "Katuic strata found in Chamic", including a claim that that stratum is shared with Acehnese. No specific examples are presented for this claim, just the assertion that it is "abundantly clear" from perusing the appendix to the book. I strongly disagree that one could reach such a conclusion on that basis, since a careful examination of the appendix makes it clear that there are no examples where Katuic can be unambiguously identified as the source of an Aceh-Chamic word. Thurgood's comparisons of Acehnese with Katuic, with my commentary, follow:

PC *_ʔεh_ 'excrement', compares with both P-Katuic and P-Vietic *_ʔεh_; Acehnese _ʔεʔ_ matches neither as its final suggests *_ʔεk._

PC *_ʔaak_ 'crow', Acehnese _ʔaʔaʔ_, while Katuic suggests *_kaʔaak_, *_ʔaʔaak_, so do basically all MK languages, yet Acehnese fails to show the regular /ɯə/ reflex of /aa/, indicating a more recent imitative (re)formation.

PC *_ʔaha_, *_ha_ 'open mouth', Acehnese _hah_, most MK language share this clearly sound-symbolic formation, yet the Acehnese fail to agree in the final. Thurgood compares to Peiros' p-Katuic *_təha,_ *_ʔəhah_, but the back vowel does not match.

PC *_dəŋ_ 'stand; stop', Acehnese _dʌŋ_, compared to Peiros' p-Katuic *_ʔtəjɨŋ_, *_ʔəʔjɨŋ_, but there is no correspondence between the forms.

PC *_kaaŋ_ 'chin; jaw', Acehnese _kɯəŋ_, compared to Peiros' p-Katuic *_təʔbaaŋ_, but there is no correspondence between the forms.

PC *_kalaaŋ_ 'hawk; bird of prey', Acehnese _klɯəŋ_, compared to Peiros' p-Katuic *_kəlhaaŋ_, but the word is found throughout MK and is even in some Malayic languages, e.g. Malay _helang._

PC *_kapaas_ 'cotton', Acehnese _gapɯəh_, compared to Peiros' p-Katuic *_kəpaajh_, but the word is an Indic borrowing found throughout MK and Malayic languages, e.g. Malay _kapas._

PC *_klaas_ 'escape', Acehnese _lhɯəh_, compared to Thomas' p-Katuic *_-klah_, *_-lah_ but the distribution of the word suggests borrowing into Katuic and Bahnaric.

PC *_krɔɔŋ_ 'river', Acehnese _kruəŋ_, compared to Peiros' p-Katuic *_kərhuaŋ_, but other MK such as Vietic *_krɔɔŋ_ are more likely---even Thai has reflexes of this MK root.

PC *_lɔɔk_ 'to peel', Acehnese _pluəʔ_, compared to Peiros' p-Katuic *_liεt_, *_luɔt_ but there is no correspondence.

PC *_picah_ 'broken; break', Acehnese _picah_, compared to Peiros' p-Katuic *_pəc[ə/a]h_, *_kəc[ə/a]h_ but the phonology and distribution suggest borrowing into Katuic and Bahnaric.

PC **pər* 'to fly', Acehnese *phʌ*, compared to Peiros' p-Katuic **par*, **paar*, although basically any MK language could be the source for Chamic, the Aceh. aspirated initial is not explained (some Pearic languages and Khasi did shift plain stops to aspirates but there is no convincing evidence of Pearic or Khasi influence).

PC **raw* 'wash', Acehnese *rhah*, compared to Peiros' p-Katuic **ʔəriaw* but the Acehnese form does not correspond.

PC **sapal* 'arm', Acehnese *sapai*, compared to Thomas' p-Katuic **qapaal* 'shoulder'. This etymon also found in Aslian (as 'upper arm') and Pearic (as 'palm (of hand)'). The problem is that the Chamic reflex has a short main vowel, and only Aslian shows a neat semantic and phonological match.

PC **sɔh* 'only; empty; free, leasure', Acehnese *sɔh*, compared to Peiros' p-Katuic **[s/c]ənhah* but Katuic all show infixed forms, unlike Bahnaric and Khmer.

PC **trɔŋ* 'eggplant', Acehnese *truəŋ*, compared to Peiros' p-Katuic **həŋgɨŋ*, **səkɨŋ* but there is no correspondence. The word is found in Malayic, e.g. Malay *terung*, which is probably more indicative of origin.

Of these 16 comparisons, few, if any, could be put forward as evidence of a Katuic stratum in Chamic, and certainly none demonstrate a Katuic stratum in Acehnese. Importantly several (such as 'wash', 'crow', 'excrement') show differences that suggest independent borrowing. As far as I can tell from the evidence I have assembled there is nothing to indicate a geographical location for Acehnese in relation to the present distribution of Chamic languages. For this reason my default hypothesis is that Acehnese separated from Chamicat a time before Chamic had developed any significant internal diversity.

The regularity of the phonological agreements between Acehnese and Chamic in their common borrowed vocabulary strongly indicates that most, if not all, these lexical items reflect a phase of Aceh-Chamic unity. Given that there is no standout source evident among known MK languages, two possibilities present themselves: a) proto-AC had contact with a range of MK languages from which it borrowed, or b) an unknown MK language that has not otherwise survived was in contact with proto-AC and contributed these borrowings—in the latter case the MK parallels adduced above are simply related MK reflexes rather than source forms.

Below I list the Aceh-Chamic borrowings without apparent wider etymologies (with borrowing into Bahnaric via Chamic indicated):

5.)

Semantic	Aceh.	P-Chamic	Comment
'arrive'	*troh*	**truh*	(> Bah.)
'descend, sink, collapse, destroy'	*lhʌh*	**gləh*	(> Bah.)
'dry weather; drought'	*khuəŋ*	**khɔɔŋ*	
'handle (of knife)'	*gʌ*	**gər*,	(> Bah.)
'many, much'	*lə*	**luu*,	(> Bah.)
'neg. imperative'	*bɛʔ*	**bɛʔ*	(> Bah.)
'peel'	*pluəʔ*	**lɔɔk*	(> Bah.)
'pick, pluck'	*pʌt, pɛt*	**pɛt*	(> Bah.)

'snail'	*ʔubo*	**ʔabaw*	
'straw (rice)'	*ɟɯmpuŋ*	**puuŋ*	
'that, there'	*sideh, hideh*	**dih*	(> Bah.)
'use'	*ŋuj*	**ʔaŋuj*	

Most of the above 12 items are also present in Bahnaric languages, although the lack of reflexes in West Bahnaric (see Sidwell & Jacq 2003) and in the rest of MK clearly indicates that what Thurgood took as straightforward MK > Chamic loans were actually borrowed from Chamic into Bahnaric, originating from an unknown source. Phonologically the words look like they are from MK—half are simple monosyllables while the rest have initial clusters or are sesquisyllabic, so our default hypothesis is that they come from some MK language or languages, the identity of which is unknown.

Can we link the group 4 and 5 etyma somehow without straining possibility too far, given that they are all at least reconstructable to PAC? I believe that it is worth speculating on this. First of all, it is a fact that each MK sub-group has a set of lexicon that is not shared with any other MK sub-group, since lexical innovation is a continuous process and an important aspect of the accretion of differences that drives linguistic diversification. Logically then, if an MK speaking community were absorbed by language shift into PC, a process that we strongly suspect did happen in ancient times, one of the consequences would be the borrowing of a set of words, some of which have a wider MK etymology, and some not, although the latter would none the less have the formal structural characteristics of MK lexicon.

This statement characterises not only the 42 AC borrowings discussed above, but also the bulk of the PC lexicon of borrowed or unknown origin reconstructed by Thurgood. Allowing for some errors and reassignments we have approximately 450 words in the PC lexicon that are borrowings or otherwise innovated, of which so far only 42 (or less than 10%) have been identified in Acehnese. It thus appears that Acehnese did not participate in a major phase of the lexical development of PC, presenting us with a significant problem of historical explanation.

4. Quantification of Etymological change and distance

Now that we have some rough indication that there is a significant difference in the absolute quantity of contact-induced change experienced by Acehnese and (the rest of) Chamic, I want to move forward to quantify this in a more representative fashion. My concern is that we don't know to what extent the PC lexicon reconstructed by Thurgood is representative of the real PC lexicon, and therefore the extent to which we can fairly compare and analyse the figures discussed above.

It is in the nature of proto-languages that they are constructs that, due to the availability of sources and various accidents of history, are necessarily incomplete or even skewed in terms of their representation of the lexicon. For example, it is commonly held that some areas of the lexicon are less stable than others, such as words representing more abstract meanings over the more concrete ones, and therefore concrete meanings will be potentially over-represented in a reconstructed lexicon. Now it is clearly beyond the scope of this paper to consider complete lexicons (whatever that might mean in practice), so I set about to devise a method that would go some way towards more fairly quantifying the proportions of lexical change in Acehnese and Chamic.

In the first place we acknowledge that Acehnese and Chamic are descended directly from Proto-Malayo-Chamic or something not very much removed from that. The Malayic sub-group of AN is already the subject of a comprehensive reconstruction (Adelaar 1992), so in the absence of PMC we might reasonably use it as a base line for quantifying the amount of lexical innovation in Acehnese and Chamic. Now I understand that there are a number of assumptions here that can be challenged, but I proceed on the basis that we are looking for a broadly indicative method, rather than a very precise tool, and one whose initial results can surely be improved by subsequent more detailed analysis. Accepting this programmatic rationale we move on to the details.

I take as my starting point the Malayic basic lexicon of 200 items as reconstructed by Adelaar (1992), using the diagnostic semantic list developed for MP languages by Hudson (1967). The 200 word list contains items from a range of semantic domains and word classes, and for our purposes I take it that for any MP language which we compare on the basis of this list its genetic classification will be evident, and the degree of lexical change from PAN, PMP or any other known starting point will be readily calculated. I copied the P-Malayic items into a spreadsheet and then added the etymologically equivalent PC and Acehnese reflexes. Where lexical replacements have occurred the new words are put in place. This is different to the strictly semantic approach of lexicostatistics which is necessarily blind to etymology in the initial compilation of the lists for comparison. I did this because I want to quantify the amount of lexical borrowing as opposed to the amount of semantic change within the lexicon.

Due to the incompleteness of the PC lexicon and Acehnese sources at my disposal the total list was reduced to 183 items.[9] The resultant list is presented as an appendix to this paper. The analysis of the list begins with counting the various common etymological retentions and innovations. Note that in some cases there is more than one form given in the sources for a given gloss, these are noted in the appendix, but in the counts below I have still treated these as single items. A summary of the results follows:

- 96 items (52.5%) where all three languages (Aceh., PC, PM) show direct inheritance of AN forms or Malayo-Chamic innovations
- 51 items (27.9%) Aceh innovations (discounting Malay borrowings)—of which 26 are shared with P-Chamic and 25 are unique to Aceh.
- 73 items (39.9%) Chamic innovations, including 26 shared with Aceh, and 47 unique to Chamic.

The above figures give a sense of proportion to the great extent of borrowing in PC in particular—approximately 40% of the basic lexicon replaced by mostly borrowed vocabulary. By contrast only just over a third (26/73), of those replacements in PC are also reflected in Acehnese.

Accepting the MC hypothesis, and Blust's estimate of MC separation around 2300 BP, plus Thurgood's estimate of a late 1st millennium break-up of PC, we would look to place the separation of Acehense somewhere in a 1000 or so year window from roughly 300 BCE forward. Taking the even bolder step of assuming a more or less stable rate of lexical replacement the above figures would place the separation of Acehnese in

[9] I considered supplementing with available items to bring it up to 200, but decided not to lest I further skew the results by my selections.

approximately the first century CE, shortly before the first historical references to Champa appear. Citing archaeological evidence, Thurgood (p.16) places the pre-proto-Chamic settlement of the Indo-Chinese coast at sometime before 600 BCE, which on my calculations would place the separation of Acehnese in the first or second century BCE.

This is only a broadly indicative calculation. Frankly I do not wish to make a claim for a stable rate of lexical replacement—since decades of experience with glottochronology have shown that the rate of change in language in respect of borrowings is quite unstable, given the possible social factors. None-the-less the fact that Acehnese demonstrably participated in only a minority of the contact driven lexical replacement that affected the rest of the Chamic strongly indicates that it separated at a much earlier than assumed by the Thurgood model. The stratum of common borrowings suggests that Acehnese split away during the early stages of a phase of assimilation of an unknown but presumably MK speaking population into the nascent Champa.

Thus one may take Thurgood's conclusion:

> The early arriving pre-Chamic peoples most likely landed south of Danang and thus probably encountered Bahnarics. Given the major restructuring of the arriving Austronesians language that took place, these pre-Chamic people must have become socially dominant, with this dominance leading many most probably Bahnaric speaking people to shift to Cham.
> [....] Probably sometime around the fall of Indrapura in the north, although it may have been as much as several centuries earlier or later, the Chamic speakers who were to become the Acehnese left the mainland on a joumey that would ultimately end in northern Sumatra. (p.251)

and reformulate it as follows:

> The early arriving pre-Chamic peoples most likely landed south of Danang and encountered a Mon-Khmer speaking population of undetermined classification. Given the major restructuring of the arriving Austronesians language that took place, these pre-Chamic people must have become socially dominant, with this dominance leading many or all of the Mon-Khmer speaking people to shift to Cham.
> [....] Sometime during this early phase of language shift, perhaps before the beginning of Common Era, the Chamic speakers who were to become the Acehnese left the mainland on a joumey that would ultimately end in northem Sumatra.

To what extent can we reconcile this with known history? Durie, discussing the founding of Champa in the second century CE, writes:

> From Chinese sources we know that there were several kingdoms during this period on the trade route to China around the Isthmus of Kra, the Malay peninsula, and the gulf of Thailand. One such was Funan, which was centred on the lower Mekong. Several kingdoms in the Isthmus of Kra were subject to it. It was overwhelmed by Khmers in the 6th century. We have no record of the language of Funan, but it could well have been a sister of early Chamic. During this period it would have quite likely for Funan traders to have been established in the Malay peninsula and even North Sumatra, which was in a strategic position for the trade with India. (Durie 1985:3)

So Durie suggests that Aceh may be a surviving fragment of Funan. Contra Thurgood, in that case the Acehnese were a southern branch of Aceh-Chamic that split off as Funan fell. The trouble I see with Durie's idea is that Funan fell to the Cambodians, and it is clear that the mysterious loan stratum found in Chamic and to a lessor extent Acehnese cannot be related directly to their language. I would like to suggest an alternative, in which the Funanese, or a segment of Funanese society, were speakers of an unrecognised branch of Mon-Khmer, and were absorbed into Champa as they lost their political and economic centre to Chenla/Ankor. Perhaps related events drove the Acehnese from the mainland, just as a thousand years later the Moklen/Moken were driven off the Isthmus of Kra by Thai expansion.

5. Conclusion

Thurgood's formulation of Acehnese as a "Chamic language" obscures an important distinction in the historical development of these languages. Alternatively I would suggest that we classify Acehnese as an "Aceh-Chamic" language, an offshoot of a stage intermediate between PMC and PC. The redrawn MC family tree, suggested by my analysis, is represented as follows:

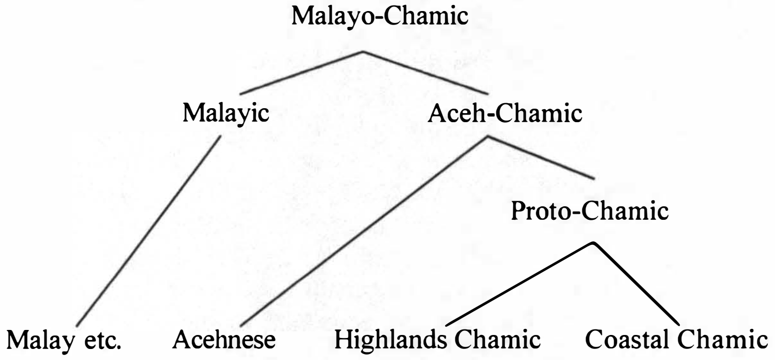


Figure 2: *Revised Malayo-Aceh-Chamic tree*

From a programmatic perspective the redrawing of the Stammbaum begs a major overhaul of the Acehnese and Chamic comparanda and their comparative-historical analysis. The resultant phonological and lexical reconstructions should be stratified into Aceh-Chamic and Proto-Chamic levels. Naturally one would seek to include in such a project:

- any new or otherwise un(der)utilised Chamic sources
- more extensive reference to Mon-Khmer sources, especially Khmer, Vietnamese and Mon, as well as more recent Mon-Khmer comparative reconstructions
- reconstruction of Proto-Acehnese based upon dialect comparison

I expect that the latter point may prove especially important, as Acehnese, although more affected by Malay, was protected by geography from much of the MK influence that has altered the face of Chamic.

References & Sources

Adelaar, K. Alexander. 1992. *Proto Malayic: The Reconstruction of its Phonology and Parts of its Lexicon and Morphology.* Pacific Linguistics Series C-119. Canberra, Australian National University.

Benedict, Paul K. 1976. Austo-Thai and Austroasiatic. In Jenner et al. (eds.) *Austroasiatic Studies, Part 1.* Oceanic Linguistics special publications No. 13. Honolulu, Unversity of Hawaii Press. pp 1-36.

Blust, Robert A. 1994. The Austronesian settlement of mainland Southest Asia. In *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society 1992.* Karen Adams & Thomas Hudak (eds.). Tempe Arizona, Arizona State University. pp:25-83.

----- 1981. The reconstruction of Proto-Malayic-Javanic: an appreciation. *Bijdragen tot de Taal-, Land- en Volkenkunde* 137.4.:456-469.

----- 2000. Review of Thurgood 1999. *Oceanic Linguistics* 39: 435-445.

Burnham, Eugene C. 1976. *The Place of Haroi in the Chamic Languages.* M.A. thesis, University of Texas at Arlington.

Collins, I. Vaughn. 1975. *The Austro-Asiatic substratum in Acehnese.* PhD. dissertation, University of California, Berkeley.

----- 1969. The position of Atjehnese among Southeast Asian languages. *Mon-Khmer Studies* III:48-60.

Collins, William. 1996. The Chams of Cambodia. In *Interdisciplinary Research on Ethnic Groups in Cambodia: Report on Research to United Nations Development Programme*, Center for Advanced Study, Phnom Penh, pp 15-107.

Cowan, H. K. J. 1933. Het Atjèsch metrum ‘sandja’’ in verband met een Tjamsch gedicht. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 90:149-55.

----- 1948. Aanteekeningen betreffende de verhouding van het Atjesch tot de Mon-Khmer-talen. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 104:429-514.

----- 1974. Evidence of long vowels in early Achehnese. *Oceanic Linguistics* XIII:187-211.

----- 1981. An outline of Achehnese phonology and morphology. *Bulletin of the School of Oriental and African Studies* LXIV 3:552-49.

----- 1983. The Achehnese diphthong ɯə and its possible implications for proto-Austronesian. *Acta Orientalia* 44:153-85.

Dempwolff, Otto. 1938. *Austronesisches Wörterverzeichnis.* Berlin, Dietrich Reimer.

Diffloth, Gérard. 1994. The lexical evidence for Austric, so far. *Oceanic Linguistics*, 33.2:309-21.

Durie, Mark. 1990. Proto-Chamic and Acehnese mid vowels: towards Proto-Aceh-Chamic. *Bulletin of the School of Oriental and African Studies,* Vol.53. London, University of London, pp 99-114.

Dyen, Isidore. 2001. Review of Thurgood 1999. *Anthropological Linguistics* 43.3:390-394.

Headley, Robert K. Jr. 1976. Some Sources of Chamic Vocabulary. In Jenner et al. (eds.) *Austroasiatic Studies: Part 1,* Oceanic Linguistics Special Publications. Honolulu, Unversity of Hawaii Press. pp 453-476.

Hudson, Alfred B. 1967. *The Barito dialects of Borneo: a classification based on comparative reconstruction and lexicostatistics*. Ithaca, New York, Department of Asian Studies, Cornell University.

Larish, Michael. 1999. *The Position of Moken and Moklen within the Austronesian Language Family*. PhD thesis, University of Hawaii.

Lee, E.W. 1966. *Proto-Chamic Phonologic Word and Vocabulary*. Unpublished Ph.D. dissertation, Indiana University.

Majumdar, R. C. 1985 (preprint of 1927 edition). *Champa: history and culture of an Indian colonial kingdom in the Far East 2nd-16th Century A.D.* New Delhi, Gyan Publishing House.

Niemann, G. K. 1891. Bijdrage tot de Kennis der Verhouding van het Tjam tot de Talen van Indonesië. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 40:27-44.

Reid, Lawrence, A. 1994. The morphological evidence for Austric. *Oceanic Linguistics*, 33.2:323-44.

Shorto, Harry. ms (no date). *Mon-Khmer Comparative Etymological Dictionary*. (being prepared for piublication with Pacific Linguistics)

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

Shorto, H.L. 1977. Proto-Austronesian *taqən: an anomaly removed. *Bulletin of the School of Oriental and African Studies* 40:128-129.

Sidwell, Paul. 2005. *The Katuic Languages: classification, reconstruction & comparative lexicon*, Munich, Lincom.

-----. 2002. Classification of the Bahnaric Languages: a comprehensive review. *Mon-Khmer Studies*, 32:1-24.

Sidwell, Paul & Pascale Jacq. 2003. *A Handbook of Comparative Bahnaric: Volume 1, West Bahanric*. Canberra, Pacific Linguistics 501.

Thomas, Dorothy M. 1967. *A Phonological Reconstruction of Proto-East-Katuic*. MA thesis, University of North Dakota.

Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: two thousand years of language contact and change*. Oceanic Linguistics special Publications No. 28. Honolulu, University of Hawaii Press.

Zorc, David. 1995. A glossary of Austronesian Reconstructions. In Darrell Tryen (ed.) *Comparative Austronesian Dictionary*, part 1, fascicle 2, pp.905-1197. Berlin & New York, Mouton de Gruyter.

Appendix 1: Summary of Acehnese words plausibly borrowed from MK sources, extracted from Thurgood (1999) and Cowan (1948). Note: 'PC' = Thurgood's reconstructions; 'C.' forms sourced from Cowan, 'C. No.' indicates Cowan's numbered comparison. MK comparisons cited are indicative only, and should not necessarily be interpreted and indentifying the particular donor language.

Aceh. *hu* 'ablaze'
Cham *hu* 'roast' (C.)
Khmer *chur* 'to ignite' (C.); Bah.
huur 'roast', Katu *huar* 'singe'
C. 64

Aceh. *ɟa* 'ancestor'
Cham *ɟa* 'appelative of poor people' (C.)
OldMon *'ja /ɲɟaʔ/*, OldKhmer *jī /ɟii/* 'great-grandmother'
C. 66

Aceh. *sapaj* 'arm'
PC * *sapal*
Reflexes in Aslian, Katuic & West-Bahnaric.

Aceh. *baʔ* 'at, on' preposition
Cham *pak* 'at, towards' (C.)
OldMon *bak* 'up to, until', *pâʔ* 'for, on, on behalf of' (C.)
C. 6

Aceh. *ruəŋ* 'back'
PC * *rɔŋ*
Katuic * *krɔŋ* 'back' , Khmu *kndrɔɔŋ* 'back'

Aceh. *sɯət* 'bail'
PC * *sac*
PMK * *saac*, widespread in MK.

Aceh. *tɤt* 'bake in fire, burn'
Khmer *tut/dot/* 'grill, roast; kindle, set fire to'(C.)
C. 140

Aceh. *sɯət* 'bale'
PC * *sac*
PMK * *saac* 'bale out' widespread in MK
C. 128

Aceh. *rɯtɯəʔ* 'bean, pea'
PC * *rətaak*
Khmer *sandaek,* Iban *retak*

Aceh. *cəgɛə* 'bear'
PC * *cagɔw*
EMK * *ɟkaw*, Asl. * *gaaw*
C. 18

Aceh. *cicem* 'bird'
PC * *cim*
PMK * *cim*, reflected in all brances but Khmer, note Nicobar has redup. initial.
C. 29

Aceh. *kap* 'bite'
(PC * *kɛʔ*)
PMK * *kap* 'bite' indicated by widespread reflexes
C. 74

Aceh. *blɛt* 'blink'
PC ?
Khmer *blet /plet/* 'appear and disappear like a flash'(C.)
C. 11

Aceh. *pot* 'blow (wind)'
PMK * *puut* 'blow' (NMK, Asl.)
C. 123

Aceh. *cabɯəŋ* 'branch, fork'
PC * *caɓaaŋ*
Malay *cabang* > Aslian, Cf. Kh. *ɟəŋrmaaŋ* 'forked stick'?

Aceh. *picah* 'broken, break'
PC * *picah*
Cf. Malay *pěcah.* Palatal stop indicates borrowing into Bahnaric also.

Aceh. *kɯbɯə* 'buffalo'
PC * *kabaw*
Aceh. = Kh. *krəbɤj*, while Chamic = Malay *kěrbaw*

Aceh. *gulam* 'carry on shldr'
PC * *gulam*
PMK * *klam* or * *kləm* on the basis of NMK & Aslian reflexes.

Aceh. _drɔp_ 'catch, arrest'
Cowan notes Mon _rap /rɔp/_ 'to catch'; PMK *_rəp,_ *_rəp_ are indicated by widespread reflexes
C. 48

Aceh. _luəŋ_ 'channel'
Cham _haluŋ_ 'hole, pit, canal' (C.)
Cf. Khmer _lùŋ_ 'to dig hole', _ʔɔnlùəŋ_ 'hole in stream-bed'; Bahnar _səluŋ_ 'pit, ditch'
C. 107

Aceh. _tijɯəp_ 'chase, run aft.'
PC *_tijaap_
Cf. Khmu _ŋgjaap_; Tampuon _tijạap_ borrowed from Chamic.

Aceh. _let_ 'chase'
Mon _lemöt nâ_ 'to drive away' (with –m- infix?) (C.)
C. 97

Aceh. _miəŋ_ 'cheek, jaw'
PC *_miaŋ_
Cf. Khmu _miəng_ 'chew', Viet. _mi "ng_, < PV *_mɛɛŋʔ_ 'mouth'

Aceh. _kɯəŋ_ 'chin, jaw'
PC *_kaaŋ_
PMK *_kaaŋ_, reflexes in Katuic, Bahnaric, Nicobarese, Vietic, Pearic.

Aceh. _kruət_ 'citrus'
PC *_kruac_
PMK *_kruəc_ 'citrus'
C. 88

Aceh. _cah_ 'clear undergrowth'
Borrowed > Bahnaric , C. compares Khmer _ceh_ 'to cut with small blows'
C. 19

Aceh. _pɯdap_ 'cover, to'
PMK *_dəp_ (widespread etymon)
C. 40

Aceh. _khop_ 'cover; put face down'
PMK *_ckup_ 'cover'; PAn *_kubkub_ 'cover'
C. 80

Aceh. _gɔm_ 'cover'
PC *_gəm_
Cf. Khmer _kaem_ 'cover, encrust, decorate', PV *_kəmʔ_ 'to bury'

Aceh. _lɯmɔ_ 'cow, ox'
PC *_ləmɔ_
Cf. Khmu _lmboʔ_, Viet. _bò_, Malay _lĕmbu_; may be derived from MK *_[]bɔʔ_ 'hump of ox', cf. Mon _ba' /pòʔ/_ id.

Aceh. _ʔaʔaʔ_ 'crow'
PC *_ʔaak_
PMK *_kʔaak_ (all but Khasi, Nic.) – Vietic reflexes typically _ʔaak_, e.g. Vt. ác, but imitative! Aceh. reflex is irregular.

Aceh. _cɛh_ 'crush, pulverise'
Cham _cɛh_ 'hatch' (C.) ?
Khmer _ces_ 'to crush' (C.); C. also compares Bahnar _she,_ Cham _cɛh_ 'hatch' the connection to 'crush' is doubtful.
C. 22

Aceh. _kɔh_ 'cut off'
PC *_kɔh_
PMK *_kɔh_ (Bah.,Kat.,Nic.,Asl.)
C. 85

Aceh. _ɟluəh, gluəh_ 'deer (small kind)'
Khmer _chlus_ 'id.'
C. 73

Aceh. _kuəh_ 'dig'
PC *_kuah_ 'shave, scrape'
PMK *_kuəs_ 'scrape'
C. 90

Aceh. _ɟep_ 'drink'
Mon _jøp /cep/_ 'sip, taste'
C. 69

Aceh. _rɯəŋ_ 'dry, dry out'
Cf. Katuic: Ta'Oi _raaŋ_ 'drying rack
C. 124

Aceh. _tho_ 'dry'
PC *_thu_
Cf. Temiar _təhool_, KhmuYuan _thúu_
C. 137

Aceh. *ʔiteʔ ʔara* 'duck-wild'
PC **ʔada*
Khmer *dā /tïia/* < PMK **daʔ*, note doublets:
Srê *ʔara / ʔada*, Bahnar *həraa / tadaa*
C. 1

Aceh. *klɔ* 'dumb'
PC **k-am-lɔ*
Cf. Khmer *kɑmlaw* 'ignoramus'

Aceh. *ɟɯəp ɟɯəp* 'each, every'
OldMon *jāp / jap/* 'all, each, every'
C. 71

Aceh. *sɔh* 'empty'
PC **sɔh*
Khmer *suh / soh/*; Bah., Kat. may have borrowed via Chamic.
C. 131

Aceh. *lhɯəh* 'escape'
PC **klaas*
> Bah., other MK suggests **laas* 'leave'

Aceh. *tɔm* 'ever'
Cham *tom* 'meet with, accomplised'(C.)
PMK **təm/***təəm/***tam* 'begin' (all MK.); perhaps from Mon *tam / tɔm/*.
C. 139

Aceh. *ʔɛʔ* 'excrement'
PC **ʔɛh*
Borrowed separately, Ch. < K/V, Ach. < ***ʔɛk*
C. 51

Aceh. *toh* 'excrete'
PC **tɔh* 'remove clothing'
Cf. Khmer *tuh / doh/* 'remove clothing; to free, release'; > Bah.
C. 138

Aceh. *ba* 'father' (C.)
(PC **ʔama* < An.)
PMK **ʔbaaʔ*, cf. Khmer *baa*
C. 2

Aceh. *ɗit* 'few'
PC **dVt* 'small'
PMK **kɗït*, cf. Viet. *nít*, Khasi *khyndit*; > Bah. (T. incorrectly states "restricted to Highlands")
C. 45

Aceh. *gap* 'firm'
PMK **gap* 'fitting, sufficient' indicated by widespread reflexes
C. 53

Aceh. *ɲum* 'flavour'
Cham *ɲəm, ɲam* 'to taste'(C.)
Praok *ɲɔm* 'to taste', Bahnar *ɲaam* 'delicious', Khmer *ɲaaɛm* 'exclamation used mostly by children vaunting what they are eating or tasting'
C. 114

Aceh. *phʌ/pʌ* 'fly (v.)'
PC **pər*
PMK **par*. Anomalous aspiratred initial also found in Rade: *phiər* (Durie 1990)
C. 122

Aceh. *tuwʌ* (<*wʌ* 'stray, wander' C.) 'forget'
PC **wər*
PMK **wir* &c. 'turn' (all MK groups, with many varients)
C. 149

Aceh. *coh coh* 'frighten animals'
Cowan notes Mon *pecuh* 'to hound on, set on as a dog'
C. 33

Aceh. *kuət* 'gather up'
PC **kuac* 'gather, amass'
**kwaac* 'scrape up'

Aceh. *bit* 'genuine, real'
≠Cham *bjak* (C.)
Cowan notes Khmer *bit / pit/* 'correct, certain'
C. 10

Aceh. *wɛh* 'go away, leave'
Cham *weh* 'dodge' (C.)
Cowan notes Khmer *veh / vèh/* 'to slip away, escape, dodge'
C. 144

Aceh. *lop* 'go into, under'
Cf. Old Mon *lop / lop/* 'to enter': word is widespread in MK, but vowel varies considerably.
C. 104

Aceh. *ɟaʔ* 'go, walk'
PMK **jak* 'tread, set out' indicated by widespread reflexes
C. 67

Aceh. *mɯh, mɯih* 'gold'
PC **ʔama(a)s*
OldMon *jimās* 'gold' (<*jās* 'shine') > Kh. *maas* 'gold'

Aceh. *gət, gɛt* 'good'
Cham *gɔt* (C.) 'just'
Khmer *gat* /*kɔt*/ 'just, exact'(C.)
C. 55

Aceh. *rət* 'graze (on grass etc.)'
Mon *rat* /*rɔ̀t*/ 'to reap': word is widespread in MK, but vowel varies considerably.
C. 126

Aceh. *klɯəŋ* 'hawk, kite'
PC **klaaŋ*
PMK **klaaŋ* (all MK groups)
C. 84

Aceh. *gɯ-* 'he, she'
PMK **ge[e]ʔ* '3rd person pronoun' indicated by widespread reflexes
C. 56

Aceh. *sɯpət* 'hit with smth.'
Mon *sapot* 'stroke or rub with hand'(C.)
C. 129

Aceh. *sɯəŋ* 'house'
PC **saaŋ*
< Khmer *saaŋ* 'to build', also >Thai/Lao)

Aceh. *goh* 'hump'
PMK **guh* 'swell', e.g. Mon *kuh* 'to swell up', Kh, etc.
C. 60

Aceh. *chen, cen* 'in love, having strong desire'
Cham *khin* (C.)
Palaung *sin* 'desire', Viet *xin* 'beg' < PMK *siin ? (Cowan comparisons weak)
C. 26

Aceh. *panah* 'jackfruit'
Mon *panah* 'jackfruit' (C.)
C. 116

Aceh. *khem* 'laugh'
Cham *khim* 'smile'(C.)
LitMon *k̭im* 'smile'
C. 77

Aceh. *wiə* 'left side'
PC **ʔiə̄w*
< PMK **w[i]ʔ* 'left', with metathesis in Chamic?
C. 147

Aceh. *ɟawiə* 'left-handed'
OldMon **jwiʔ* 'left' < PMK **w[i]ʔ*
C. 68

Aceh. *buəŋ* 'morass'
Khmer *piŋ* /*bɤŋ*/ 'lake, pool'; > Stieng *bhəŋ* (C.), > Thai *bɯŋ*
C. 17

Aceh. *le* 'more, still more'(C.)
Riang-Lang *¯ləj* 'more, longer, else', Viet. *lại* 'again', Mon *lē* 'also', etc.
C. 94

Aceh. *bɯŋəh* 'morning'
Cham *paguh* 'morning light'(C.)
Mon *ŋuh* 'awake out of sleep' (C.)
C. 111

Aceh. *cʌt* 'mountain range'
PC **cət*
Cf. Khmer *caot* 'high, steep, sheer, abrupt'
C. 35

Aceh. *takuə* 'neck'
PC **takuaj*
Resembles PMK **kuuj* 'head' (Kat., Asl.), but doubtful. C. compared to a different etymon.
C. 135

Aceh. *kɯmuən* 'nephew'
PMK **kmun*, **kmuun*, **kmuən* 'nephew'
C. 92

Aceh. *coŋ* 'on top of'
Cowan notes Khmer *coŋ* 'end, tip'
C. 34

Aceh. *hah* ‘open (mouth)’
PC **ʔaha*
PMK **haʔ, hah*, Ach. resembles B. & Viet.
C. 61

Aceh. *gɔp* ‘other, group’
PC **gəp*
PMK **gap, gəp* ‘friend, associate’ (Khm., Bah., Viet.)

Aceh. *lap* ‘to paint’(C.)
Khmer /srlaap/ ‘to rub, anoint, smear, paint’
C. 101

Aceh. *cɔh* ‘peck (as snake)’
PC **cɔh*
PMK **[ʔ]cɔh* (EMK, Khmu, Asl.)
C. 32

Aceh. *weŋ* ‘pedal’(D&D), ‘turn around’ (C.)
Cham *wiŋ* ‘turn, whirl’ (C.)
PMK * *wiŋ* &c. ‘go round’ (all MK groups, with many varients)
C. 145

Aceh. *pət* ‘pick (fruit, flower)’
PC **pɛt*
MK forms suggest *pic, but connection is questionable.
C. 118

Aceh. *tamɛh* ‘pillar, post’
PC * *tamɛh*
Cf Mon *tmit* ‘post supporting veranda’- doubtful.

Aceh. *bantaj* ‘pillow’
PC * *bantal*
Cf. Malay *bantal*

Aceh. *cubet* ‘pinch’
(PC **kapit*?)
Cowan notes Khmer *cbec* ‘to pinch’
C. 38

Aceh. *bət* ‘pluck, uproot’
PC * *buc*
Khmer *boac* ‘to pull up’, Mon *bot* ‘unsheathe’; > Bah.,Stieng *buc*; also Malay *cabut*
C. 16

Aceh. *ʔuət* ‘polish, rub clean’
Cham *u̯ak* ‘rub’ (C.)
Lawa *ʔuət* ‘wipe’, Khmu *ʔɔɔt* ‘scrub body’
C. 142

Aceh. *bep* ‘pout like a monkey’ (C.)
Cf. Khmer *bep /pép/* ‘moue des lèvres, contracter les lèvres, grimacer’ (C.)
C. 9

Aceh. *lhan, tlan* ‘python’
PC **klan*
PMK **tlan* - Aceh. borrowed with apical initial (Kh.?); Chamic < form with velar initial (Bah./Mon?)
C. 102

Aceh. *wɔə* ‘return home’
PMK **wil* &c. ‘turn’ (all MK groups, with many varients)
C. 148

Aceh. *kruəŋ* ‘river’
PC **krɔɔŋ*
PMK **ruŋ*, **ruuŋ*, **ruəŋ*; low vowel reflexes in Bah. & Khmu’.
C. 87

Aceh. *kuət* ‘scrape/clear away’ (C.)
Cham *kwac* ‘dig’ (C.)
PMK **kwaac* ‘scratch up’, e.g. Khmer *khvaac*, Kensiw *kəwɔɟ*
C. 91

Aceh. *kɛh* ‘scratch’ (D&D ‘matches’)
Mon *keh* ‘write with stylus’ < PMK **kiəs* ‘scratch’
C. 75

Aceh. *ŋiəŋ* ‘see, look’
Aslian: Senoi, Blanya-Sakai *neŋ* ‘to see’ (C.)
C. 109

Aceh. *dɯə* ‘shallow’
PC **dɛl*
Mon *da* ‘shallow’(C.); PMK & Aslian reflexes show *[ɛ]*
C. 42

Aceh. *be* ‘size, amount’
Senoi *bē* ‘very’ (S&B); > Stieng
C. 7

Aceh. *caŋ* 'slash, strike, slice, chop'
Mon *caŋ* 'prick, pierce' (C.), also >
Stieng.Cf. Malay *cincang*
C. 19

Aceh. *cut* 'small', *bacut* 'a little'
Aslian: Senoi *maʔcut*, Sakai *macut*
'small'(C.)
C. 39, 4

Aceh. *chuəŋ* 'smelling of urine'
Cowan notes Khasi *jung* 'urine'
C. 28

Aceh. *lɯəŋ* 'spread out'
PC **laaŋ*
PMK **laaɲ* 'spread out'
C. 99

Aceh. *wɯə* 'stable, pen'
Cham *wa, war* 'yard (buffalo), stable' (C.)
Khmer *val/viel/* 'plain, field, clearing, courtyard, plaza, threshing floor'; Mon *wa* */wɛ̀a/* 'open space, pasture'
C. 148

Aceh. *dʌŋ* 'stand, stop'
PC **dəŋ*
Viet. *đứng* (doubtful); Cowan notes Mon *demɔŋ* 'remain, dwell' (with infix)
C. 47

Aceh. *dəm* 'stay overnight'
Cham *dəm* (C.)
Mon *døm/tɜ̀m/* 'to lodge'(C.); PMK **dəm* is indicated by widespread reflexes
C. 46

Aceh. *cɯɪt* 'stinging pain'
Khmer *cɔt* 'sour', Stieng *cət* 'astringent'(C.)
C. 24

Aceh. *cɯŋɛh* 'stink, unpleasant smell'
Khmer *chʔɛh,* Mon *həʔeh*, Stieng *ciʔih* 'to stink'(C.)
C. 23

Aceh. *gɔp* 'stranger, other'
PMK **gəp, *gap* 'friend, to associate'; C. notes Aslian forms with semantic match
C. 59

Aceh. *bʌt* 'stretch'
Cham *but* 'twisted' (C.)
Khmer *bot/pot/* 'to curve, fold'; also >
Stieng
C. 15

Aceh. *pɔh, pɛh* 'strike, beat'
PC **pɔh*
Khmer *pah* 'hit', *poh* 'hammer', *puh* 'hit with stick', Mon *peh* 'kick (of horse)', *kəpɔh* 'hit with hand'
C. 117

Aceh. *pɔh, pɛh* 'strike, pound'
PC **pɔh*
PMK **pah, *puh, *puəh*, NMK, Bahnaric, Vietic.

Aceh. *kʌŋ* 'strong, hard'
PC **khaŋ*
Katuic**kəŋ,* Khmer *kɛ̀əŋ* < **gaŋ* ? Chamic < Viet. **khăŋ* ?
C. 86

Aceh. *ŋɔp* 'submerged'
Khmer *ŋup* 'incline, drop', Khasi *ŋop* 'subside' (C.)
C. 110

Aceh. *ba* 'take, carry'
PC **ba*
OldKhmer *va*, Temiar *baʔ* 'carry on back'
C. 3

Aceh. *cɔʔ* 'take, seize'
Cham *cɔk* (C.)
WestBahnaric **cɔk* 'take'; Khmu *cɔɔk* 'catch (e.g. pig)', *cok* 'take out (e.g. entrails)' although other MK suggest **jɔ(ɔ)k*, e.g. Khmer *jɔ̀ɔk* 'take'.
C. 31

Aceh. *criəʔ* 'tear, rip'
Khmer *criək* 'to split'(C.)
C. 36

Aceh. *sideh* 'that, there'
PC **dih*
Mon *ɗeh* 'he or she (disrespectful)' (C.)
C. 41

Aceh. _bʌh_ ‘throw away’
Khmer _poh_ /_bɔh_/ ‘to throw’
C. 14

Aceh. _wɛt_ ‘turn’
PMK * _wac_ ‘twist’, e.g. Bah. _wɛc_ ‘twist’,
Mon _wòt_ ‘wring out’ etc.
C. 146

Aceh. **_plòïh_** ‘unroll’ (C.)
Mon **_plöh_** ‘untwist’(C.)
C. 121

Aceh. _that_ ‘very’
Mon _that_ /_thɔ̀t_/ ‘well, healthy, strong’,
Khmer _hat_ ‘to exert’, _that_ ‘large, obese’
(C.)
C. 136

Aceh. _sɯʔuəm_ ‘warm’
Khmer _sʔɔm_ ‘to heat, warm’(C. compares a different Kh. root)
C. 130

Aceh. _rhah_ ‘wash’
PC *_raw_
Aceh. cf. Viet. _rửa_ (< *_raah_ ?), Chamic cf.
Bah., Kat. *_ʔaraaw_
C. 133

Aceh. _sɔm_ ‘wrap’
PC *_səm_
Old Khmer _sum_ ‘to wind, roll, wrap up’

Aceh. _lɯən_ ‘yard’
Khmer /diilaan/, /lan/ ‘flat open area, square,
yard’
C. 98

Aceh. _sɯmɯŋɯp_ ‘yawn’
PC *_həʔaap_
PMK *_sʔaap_, *_sŋʔaap_, not all MK subgroups have medial nasal

Summary of rejected comparisons from Cowan (1948):

Phonological correspondence(s) defective: 5, 12, 13, 21, 25, 26, 30, 37, 43, 54, 57, 62, 70, 89, 93, 96, 100, 105, 106, 108, 119, 125, 127, 134, 135, 143
Semantic comparison unconvincing: 103, 113
An. or Malay: 8, 27, 65, 132
Indic: 120
Expressive/sound symbolic: 63, 82, 83, 115
No resemblant forms found beside obvious loans into Bahnaric: 49, 50, 58, , 76, 78, 79, 95, 112, 141

Appendix 2: Basic vocabulary of Acehnese, Proto-Chamic, Proto-Malayic, 183 items.

Sematic	Acehnese	P-Chamic	P-Malayic	Commentary
above/on top	_ʔatɯəh_	*_ʔataas_	*_atas_	All < PAN *_Caʔas_
ashamed	_malɛə_	*_malɔw_	*_malu_	All < Malayo-Chamic etymon
ashes	_abɛə_	*_habɔw_	*_habu_	All < PAN *_qabúH_
at	_di_	*_di_	*_di_	All < PMP *_di_. although the failure to diphthongise in Aceh.-Chamic is odd.
back (anat.)	_ruəŋ_	*_rɔŋ_	*_bA lakaŋ_	Aceh-Chamic replaced by MK, Cf. Bahnar _rɔŋ_, Khmu _kndrɔɔŋ_. Note: Bahnaric may have back-borrowed from Chamic. the original MK form retained in West Bahnaric *_krɔŋ_ 'back of knife blade'
bad	_ɟɯhɯt_	*_ɟəhaat_	*_ɟahət_	All < PMP *_zaqát_
belly/guts	_pruət_	*_pruac_	*_pərut_	Metathesis in Aceh-Chamic
below	_baroh_	*_ʔala_	*_babah_	Aceh corresponds to Iban _baruh_ and Maningkabau _barɯʾh_; Chamc obscure
big	_raja, rajəʔ_	*_raja_	*_raja_	All < PAN *_Raja_
bird	_cicem_	*_cim_	*_buruŋ_	Aceh-Chamic borrowed < MK *_cim_
bite	_kap_	*_kɛʔ_	*_gigit_	Aceh < MK *kap; Chamic form obscure
black	_ʔitam_	*_hitam_	*_hitəm_	All < PAN *_qitém_
blood	_darah_	*_darah_	*_darah_	All < PAN *_dáRaq_
blow	_jop_	*_ʔajup_	*_t/iup_	All < PAN *_Siúp_
bone	_tulɯəŋ_	*_tulaŋ_	*_tulaŋ_	All < PHF *_CuqelaN_
branch	_dhɯən_	*_dhaan_	*_dahan_	All < PMP *_daqan_
breast	_tɛʔ, dɛʔ_	*_tasɔw_	*_susu(ʔ)_	Aceh. < Malay _tetek_; Chamic shares initial stop with Iban _tusu_
breathe	_naphãh_	*_ɲawa_	*_ɲa wa_	Aceh. < Malay _napas_ < Arabic; Malayo-Chamic < PMP *_ɲá wa_
burn	_tət_	*_ɓəŋ_	*_bakar_	All three apparently innovated; Cf. OKhmer _tut_ (_dɔt_) 'brûler'
buy	_blɔə_	*_blɛj_	*_bəli_	All < PAN *_bělí_
chew	_mamʌh_	*_mamah_	*_mamah_	All < PMP *_mamáq_
child	_ʔanɯʔ_	*_ʔanaak_	*_anak_	All < PAN *_aNak_, widely borrowed (via Malay?) in SEAsia
choose	_pileh_	*_ruah_	*_pilih_	Aceh. & Malayic < PAN *_píliq_, Chamic borrowed from MK, Cf. Khmer _rɤh_, Stieng _rɔɔjh_, although the Chamic vocalism is not explained
claw/nail	_gukɛə_	*_kukɔw_	_kuku_^{Malay}	All < PAN *_kuS+ kuS_
climb	_ʔeʔ_	*_ɗiʔ_	*_naik_	All < PMP *_nahik_
cloud	_awan_	*_hual_	*_a(bw)an_	Aceh. borrowed Malay _awan_, Chamic obscure
cold	_siɟuək, lɯpiə_	*_laʔən_	*_diŋin_	Aceh. borrowed Malay _sejuk_, other Malayic < PMP *_diŋ+ diŋ_; Chamic

				obscure
come/arrive	troh	*truh	*datəŋ	Aceh-Chamic etymon is shared with North+Central Bahnaric, source unknown.
cook	tagɯn	*tanak	*tanak	all < PHF *taNek , assuming that Aceh. shows metathesis
count	bilɯəŋ	*jaap	*hituŋ	Aceh. < PHF *bílaŋ; Chamic < PHF *Hiáp; Malayic < PAN *qi-(n)tuŋ
cry/weep	kliʔ, mɔə	*cɔk	*taŋis	Malayic < PAN *Cáŋis, Aceh. & Chamic forms obscure
cut/hack	tɛktɛk	*tarah	*tətək, *taRas	Aceh. & Malayic < PAN *tek+ tek, Chamic & Malayic < PAN * taRáq
day/sun	ʔurɔə	*hurɛj	*hari	All < PAN * waRiH
die	mate	*mataj	*mati	All < PAN *maCéj
dig	kuəh	*kalɛj	*kali	Chamic & Malay < PAN *kálih, Aceh. appears to have borrowed from MK, Cf. Bahnar kwajh 'dig up, scratch around for'
dirty	kutɔ, tɨbəh, mɨluteŋ	*chəp, *grit	*kamah/ *kumuh	Aceh. kutɔ from Malay kotor, but other forms obscure.
dog	ʔasɛə	*ʔasɔw	*asuʔ	All < PAN *asu, with semantic shift > 'canine' in Malay
dream	lumpɔə	*lumpɛj	*m/impi/ *impi	All < PMP *nipi, note the Aceh-Chamic shift *n- > *l-
drink (water)	minom	*minum	*inum	All < PMP *inúm
dry	kraŋ, tho	*raŋ, *thu	*kəriŋ	All < MP doublet *kaRaŋ/*kaRiŋ, plus Aceh-Chamic has innovated *thu - origin obscure
dull/blunt	tumpoj	*ʔabual	*tumpul	Aceh. & Malayic < PAN *dump+ pel , Chamic obscure
dust	dhoj, ʔabɛə	*dhual/r	*dəbu	Aceh. + Malayic < PMP *debu; but *dhual/r (more probably *dhul) is obscure
ear	gɯliɲuəŋ	*təliŋa	*ƭ4 liŋa(ʔ)	All < PHF *taŋíla
earth/soil	tanɔh	*tanah	*tanah	All < PMP *tanaq or *taneq
eat	makɯən	*ɓəŋ	*ma/kan	Aceh. & Malayic < PAN *kán, Chamic obscure
egg	bɔh	*bɔh	*təlur	Aceh-Chamic replaced PAN *tĕlúR 'egg' - Thurgood suggests *bɔh < PAN *buáq 'fruit', although the vocalism is problematic
eye	mata	*mata	*mata	All < PAN *maCá
fall down	rhət	*labuh	*labuh	Chamic & Malayic < PMP *ka-nabúq, Aceh. obscure
far/distant	ɟɯʔoh	*ɗɔh	*ɟauh	Aceh. & Malayic < PMP *Zaúq, Chamic obscure

fat, grease	gapah	*ləmaʔ	*ləmək	Chamic + Malayic < PMP *lemak; Aceh. obscure
father	ʔajah, jah, ʔa bu, du, abi	*ʔama	*apa(ʔ)	Aceh. forms all secondary; Chamic < PAN *ama, Malayic < PHN *bapaʔ
fear, afraid	takot	*huac	*takut	Aceh. + Malay(ic) < *PAN *tákut, Chamic obscure
feather	bulɛə	*bulɔw	*bulu	All < PMP *búlu
fire	ʔapuj	*ʔapuj	*api	All < PAN *Sapúj
fish (n.)	ʔɯŋkot	*ʔikaan	*ikan	Chamic & Malayic < PAN *Si-káʔen; Aceh. obscure
flow	ʔile	*ɗuac	*alir	Aceh. & Malayic < PMP *a+ liR, although Aceh. may have borrowed Minangkabau ilʔ ; Chamic obscure
flower	buŋɔŋ	*buŋa	*buŋa(ʔ)	All < PMP *búŋah
fly (v.)	phʌ, pʌ	*pər	*tAr(ə)baŋ	Aceh-Chamic has borrowed < MK. Cf. PMK *par
foot/leg	gaki	*kakaj	*kaki	Aceh. has borrowed directly from Malay(ic).
forest	ʔutɯən	*hutaan	*hutan	All < PMP *qutan
four	pɯət	*paat	*əmpat	All < PAN *Sĕ(m)pát
full (sated)	pɯnɔh, trɔə	*trɛj	penuh^{Malay}	Aceh. & Malay < PMP *pĕnúq, + Aceh-Chamic innovated
give	bri, ɟok	*brɛj	*bəriʔ	Chamic & Malayic < PAN *bĕRáj, Aceh. has borrowed Malay beri & an MK form, Cf. Khmer jɔ̀ɔk 'take'
good	gʌt, gɛt	*biaʔ, gɔt^{Cham}	*baik	Aceh. + Cham < Khmer gɔt /kɔt/
grass	nalɯəŋ	*rək	*rumput	All show independent innovation
green	ʔijo	*hiɟaw	*hiɟaw	All < Malayo-Chamic etymon
grow	timoh	*tamuh	*t/um/buh	All < PAN *Cú(m) buq
hair (of head)	ʔok	*ɓuk	*buø(uə)k	All < PAN *buSék
hand	ɟarɔə	*taŋaan	*taŋan	Cf. Malay jari 'finger'. Acehnese shares with Iban the semantic shift 'finger' > 'hand', using the compound ʔanɨk ɟarɔə 'child hand' for 'finger'. Chamic *caɗiaŋ 'finger' borrowed from unknown source.
he/she	ɟih	*ɲu	*ia	Chamic correspondes to Minangkabau iɲo; Malayic < PAN *siá; Aceh. shows a variety of forms
head	ʔulɛə	*ʔakɔʔ	kepala^{Malay}	Aceh. regularly < PMP *qúluH; Malay < Indic; Chamic < MK, Cf. Mon kɔʔ 'neck'
hear	dɯŋʌ, lɯŋʌ; simaʔ 'listen attentively'	*həməʔ	*dəŋər	Aceh. + Malayic < PMP *dʒĕ+ ŋéR, although Cf. PMK *[t$_1$]ŋər, e.g. Viet. nghe 'to hear', RiangLawa ¯təkŋar 'to listen'; The Aceh-Chamic simaʔ/*həməʔ etymon is obscure.

heavy	_ghən, brat_	*_traap_	*_bərat_	Aceh. * Malayic < PMP *_beRʔat_, other Aceh. and Chamic obscure
hit/slap	_tampa_	*_pah_	_tampar_^{Malay}	Aceh. < Malay; Chamic < MK, Cf. Khmer _pah_ ‘hit’
strike/beat	_pɔh, pɛh_	*_pɔh_	*_pukul,_ *_paluʔ_	Aceh. & Chamic < MK, Cf. Khmer _pah_ ‘hit’, _poh_ ‘hammer’, _puh_ ‘hit with stick’, Mon _peh_ ‘kick (of horse)’, _kəpɔh_ ‘hit with hand’
hold	_rɯgam, mat_	*_ʔɟaaʔ,_ *_ʔapan_	*_pəgaŋ_	Aceh. < PAN *_gem_ + _gem_, Chamic & Malayic obscure although MK forms such as OldMon _bgan_ ‘to yoke, take hold of’ are suggestive
hom	_luŋkɛə_	*_tuki_	_tanduk_^{Malay}	Ache-Chamic has borrowed from MK, the etymon is found in Bahnaric & Katuic, Cf. Bahnar _ʔəkɛɛ_
house	_sɯəŋ_	*_saaŋ_	*_rumah_	Malayic < PAN *_Rumaq_, Aceh. & Chamic borrowed , Cf. Thai/Lao _saaŋ_ ‘granary, warehouse’
I	_kɛə_	*_kɔw_	*_aku_	Aceh-Chamic < PAN *_ku_, Malayic < PAN *_akú_
inside	_dalam_	*_dalam_	*_(d-)aləm_	All < PAN *d_2_á_+ _lem_
knee	_tɯʔot, tuʔot_	*_tuʔut_	*_tuʔ(uə)t_	All < PHF *_túSud_
know (things)	_thɛə_	*_thɔw_	*_tahu_	All < PMP *_taqú_
lake	_danɔ_	*_danaw_	*_danaw_	All < PAN *_dánaw_
laugh	_khem_	*_klaw_	*_tawaʔ_	Malayic < PAN *_Cáwa_, Aceh. & Chamic forms obscure.
leaf	_ʔon_	*_sula_	*_daun_	Aceh. & Malayic < PMP *d_2_ahun_, Chamic < MK, Cf. PMK *_slaʔ_
left side	_wiə_	*_ʔiə̃w_	*_kA-iri/* kibaʔ_	Malayic < PAN *_ka-wiRi_; Aceh-Chamic < MK Cf. Khmu _trweʔ_, Jenai _wĩʔ_, Mon _c’wei_ (with metathesis in Chamic and > Bahnaric).
lightning	_kilat_	*_kataal_	*_kilat_	Aceh. & Malayic < PHF *_kilát_, Chamic is obscure, but could be derived by metathesis
live	_ʔudep_	*_hudip_	*_hudip_	All contine PAN *_qúd_$_2$_ip_
liver	_ʔate_	*_hataj_	*_hati_	All < PAN *_qaCéj_
louse	_gutɛə_	*_kutɔw_	*_kutu_	All < PAN *_kúCuH_
man/male	_lakɔə_	*_ʔakɛj_	*_laki_	All < PMP *_láki_
many	_lə_	*_lu_	_banyak_^{Malay}	Aceh-Chamic obscure
meat/flesh	_ʔasɔə_	*_ʔasɛj_	*_isiʔ_	All contine PAN *_Sesi_ (Malayic also innovated *_dagiŋ_)
moon	_bulɯən_	*_bulaan_	*_bulan_	All < PAN *_bulaN_
mosquito	_ɟamɔʔ, ɲamɔʔ_	*_ɲamuk_	*_ɲamuk_	All < PMP *_ɲamúk_
mother	_maʔ, ma_	*_mɛʔ_	*_(ə) ma(ʔ)_	Aceh. corresponds to Malayic, Chamic resemble numerous MK forms

				suggesting PMK *mee?
mountain	gunɔŋ, cɔt/cʌt	*cət	gunung^{Malay}	Cf. Khmer cót 'escapé'; Aceh. gunɔŋ < Malay
mouth	babah	*babah	*mulut	Aceh-Chamic < PMP *baqbaq
name	nan	*ʔanan	(Malay nama < Skt.)	Aceh-Chamic etymon obscure, borrowed into Bahnaric, Cf. Bahnar ʔənan
narrow	ʔubit/ʔubɯt	*ganiat	*səmpit	Aceh. and Malayic may reflect independent varients of PMP *kapit; Chamic obscure
near	tɔə, rap	*ɟɛʔ	*dəkək	All show independent developments
neck	takuə	*takuaj	*lihər	Aceh-Chamic resembles PMK *kuuj 'head'
needle	ɟarom	*ɟarum	*ɟarum	All < PAN *ZáRum
new	baro	*bahrɔw	*baharuʔ	All < PAN *baq(e)RuH
night	malam	*malam	*ma-lə(hø)m	All < Malayo-Chamic etymon
nose	ʔidoŋ	*ʔiduŋ	*hiduŋ	All < PAN *ɨ+júŋ
not	hʔan, tan	*ɓuh...ʔɔh	*-daʔ	All show independent developments
old (person)	tuha	*klap	*tuha(ʔ)	Aceh. & Malayic < PAN *tuqáS; Chamic obscur
one	sa	*sa	*əsaʔ	All < PAN *sa
open/uncover	pɯhah	*pəh	*bukaʔ	Chamic < MK, Cf. Bahnar pɔh, Palaung pɯh, Aceh. Cf. Viet. ha?; Malayic < PMP *bukáʔ
other	bukʌn	*bukən	*bukən	All < Malayo-Chamic etymon
person/human	ʔurɯəŋ	*uraaŋ	*uraŋ	All < Malayo-Chamic etymon
rain	ʔuɟɯən	*huɟaan	*huɟan	All < PAN *quZáN
rat	tikoh	*tikus	*tikus	This Malayo-Chamic etymon resembles MK words for 'porcupine', e.g. PWaic *ŋkos, PSemai *kuus; also borrowed into Moken as koh 'porcupine'
red	mirah	*mahirah	*(ma-)irah	All < PMP *ma+ iRaq
right side	ʔunɯn	*hanuãʔ	*k/anan	Aceh. corresponds to Malayic. Chamic is obscure, but is perhaps an infixed reflex of the same etymon as Minangkabau suoʔ 'right side'
road/path	ɟalan	*ɟalaan	*jalan	Aceh. < Malay(ic) (otherwise ɟalɨən expected)
root	ʔukhɯə	*ʔughaar	*akar	Aceh-Chamic < PMP * wakaR (note influence of * w on minor-syllable vocalism), Malayic < PMP *akaR
rope/string	talɔə	*talɛj	*tali	All < PAN *Calís
rotten	broʔ	*bruʔ	*busuk	Aceh-Chamic < PAN * buRúk, Malayic < PMP *busuk
salt	sira	*sira	*sira, garam^{Malay}	All < PAN *qasiRa, plus some replacement with garam in Malay and others

sand	*ʔanɔə*	**cuah*	**pasir*	Aceh. < *PAN **qĕnaj*; Chamic & Malayic independently innovated
say/speak	*mɯtuto*	**lac*	**tutur*	Aceh. corresponds to Malayic; Chamic etymon obscure
scratch (itch)	*krut*	**kabac*	**garut*, **garuk*	Aceh. & Malayic < PMP **ka+Rud*; Highlands Chamic borrowed from Bahnaric. Cf. Bahnar *kəbajʔ*, infixed PMK **kaac*
sea/ocean	*laot*	**tasiʔ*	**tasik*	Malayo-Chamic < PMP **tasik*, Aceh. borrowed Malay *laut*
see	*kalʌn*, *ɲieŋ*, *ʔɯ*	**ɓuh*	**lihat*	Aceh. forms obscure; Chamic > Bahnar *ɓoh*. Cf. also OldMon /təmɓah/ 'to appear'
sew	*cɔp*	**ɟahit*	**ɟahit*	Chamic & Malayic < PMP **záqit*. Aceh. obscure
sharp	*taɟam*	**haluaʔ*	**taɟəm*	Aceh. & Malayic < PMP **tazím*, Chamic obscure
shoot (arrow)	*panah 'arrow'*	**panah*	**panah*	All < PAN **panaq*
shoulder	*baho*	**bara*	*(PAN **qabáRaH)*	Aceh. < Malay *bahu*
sick	*saket*	**sakit*	**sakit*	All < PMP **sakít*
sit	*duəʔ*	**dɔɔk*	**duduk*	All < PMP **d_2uk+d_2uk*, note: Aceh. resembles Minangkabau *duduəʔ*. Chamic vowel quality is not explained
skin	*kulet*	**kulit*	**kulit*	All < PAN **kúliC*
sky	*laŋɛt*	**laŋit*	**laŋit*	All < PAN **láŋit*
sleep/lie down	*ʔeh*	**ɗih*	**tidur*	Aceh-Chamic < PMP **hideRáq* 'lie down'; Malayic < PAN **tid_2ur* 'to sleep'
small	*ʔubɯt*, *ʔubit*, *cut*	**ɗVt*	**kəcil*, **kətik*	Aceh. & Chamic forms obscure
smoke	*ʔasap*	**asap*	**asəp*	All < Malayo-Chamic the etymon
snake	*ʔulɯə*	**ʔular*	**ulər*	All < PAN **ulaR*
sniff, smell	*com*	**cum*	*cium*^{Malay}	Malayo-Chamic etymon of obscure origin, also borrowed into North & Central Bahnaric
spider	*rambidɯən*	**waj*	**lawaʔ*, **laba(ʔ)*	Aceh. appears to correspond, at least partially, to Iban *əmpəlawaʔ*. Highlands Chamic has borrowed a word meaning 'turn' (> 'spin (web)' Cf. Bahnar *waaj* 'roll up, turn'
spit	*ludah*, *rudah*	**kacua*, **kacuh*	**ludah*	Aceh. borrowed < Malay: Malayic < PMP **luZáq*. Chamic < MK. Cf. Khmu *kɟuh*. Bahnar *ksɔh*
split (v.t.)	*plah*	**blah*	**bəlah*	All < PAN **bĕ+láq*
squeeze	*ɟɯpat/ɟupat*, *prah*	**kapit*, **cupa/et*	**pərəs*, **pərah*	Aceh. and Malayic < PMP **peRáq*, while Aceh-Chamic has borrowed a prefixed from of PMK **pat*
stab	*tɔp*	**kləp*	**tikəm*,	Aceh. and Chamic have independently

			*tusuk	borrowed from MK while Malayic < AN etyma
stand/stay	dʌŋ	*dəŋ	*diri	Malayic < PMP *d_2iRi; Aceh-Chamic resembles Viet. đứng, 'be standing, to set' but initial voicing is problematic, an alternative comparison is PMK *duŋ 'house'
stand up	bɯdəh	*taguuʔ	bangun^{Malay}	?
star	bintaŋ	*bituʔ	*bintaŋ	Chamic < PAN *bi-(n)túqen, while Aceh. has borrowed the Malayic varient with final velar nasal
steal	puplɯŋ, cuə	*klɛʔ	*maliŋ	Aceh. puplɯŋ relates to Malayic, but cuə is obscure, as is Chamic *klɛʔ
stick (wood)	kajɛə 'wood'	*kajɔw 'tre e, wood'	*kajuʔ	All < PAN *kájuH
stone	batɛə	*batɔw	*batu	All < PAN *batú
suck, sip	hirop, pĩəp	*sarip, *mam	*hiRup^{PMP}, *hi(ŋ)səp	Aceh. pĩəp plausibly < Malayic *hi(ŋ)səp, Chamic *mam is clearly a nursery word
swell (abscess)	barah	*barah	barah^{Malay}	All < PMP *baReq
swim	laŋuə	*luaj	*(mb)A-rənaŋ	Aceh. < PHF *laŋúj, Chamic is replaced by MK
tail	ʔiku	*ʔiku	*ikur	All < PAN *íkuR
that (far)	ɲan, nan	*ʔanan	*(ı)na(n), *(a)na(ʔ)	All < PAN *i-náʔ
thick	tɯbaj	*kapaal	*təbəl	Chamic < PMP *kapal; Aceh. & Malayic appear to reflect MK loan, Cf. PMK *[t]ɓəl
think	pike	*saniŋ	——	Aceh. < Malay pikir < Arabic; Chamic is obscure
this (near)	nɔə	*ʔiniʔ, *inɛj	*(ı)ni(ʔ)	All < PAN *i-ní
three	lhɛə	*klɔw	*təlu	All < PAN *tĕlú
thunder	gɯlantɯə	*grəm	*guntur	Aceh. corresponds to Malayic, plus -l- infix which MK languages use to indicate repeated action; Chamic < MK, Cf. PMK *grəm[ʔ]
tie/fasten	ʔikat	*ʔikat	*ʔikət	All < PMP *hi+ket
tongue	dilah, lidah	*dilah	*dilah	All < PHF *$d_3$ílaq 'lick', Aceh. also shares metathesised reflex with Malay
tooth	gigɔə	*gigɛj	*gigi	All < the Malayo-Chamic etymon
true	bəna	*biaʔ	*bənər	Aceh. & Malayic < PMP *bener, while Chamic has merged with *biaʔ 'good'
turn over	baleʔ	*blək	*biluk	Aceh-Chamic < PAN *balík 'turn around'
two	duwa	*dua	*dua(ʔ)	All < PAN *d_3uSá

vomit	_muntah_	_*patah_	_*m/u(n)tah_	Aceh. < Malay; Aceh-Chamic < PAN _*utaq+ m_
walk/go	_gaki_	_*labaat,_ _*naw_	_*((mb)Ar)ɟalaŋ_	Aceh. borrowed Malay _kaki_
warm	_sɯʔuəm_	--	_*panas_	Aceh. < MK, Cf. Khmer _sʔɔm_ 'warm'; Malayic < PMP _*panas_
water	_ʔiə_	_*ʔiar_	_*air_	All < PMP _*wáhiR_
we (excl.)	_kamɔə_	_*kamɛj_	_*kami_	All < PAN _*kamí_
wet	_basah_	_*basah_	_*basah_	All < PMP _*basáq_
what?	_pɯə, puə_	_*hagɛt_	_*apa_	Aceh. & Malayic < PMP _*apa_, Chamic obscure
white	_puteh_	_*putih_	_*putih_	All < PAN _*putíq_
who?	_sɔə_	_*sɛj_	_*sai, *si-apa_	All < PMP _*i-sai_
wind	_ʔaŋɛn_	_*ʔaŋin_	_*ʔaŋin_	All < PMP _*háŋin_
wing	_sajɯəp_	_*sajaap_	_*sajap_	All < PHN _*sajap_
woman/female	_binɔə_	_*kumɛj_	_*bini_	Aceh. & Malayic < PMP _*ba-b(in)áHi_, Chamic obscure
work, do	_buət_	_*buat,_ _*bruãʔ_	_*buat_	All < PAN _*buhat_; Chamic _*bruãʔ_ borrowed into some Katuic & Bahnaric langs., but origin obscure, possibly secondary from _*buat_
worm	_ʔulat_	_*hulat_	_*hulət_	All < PAN _*qúleɟ_
yawn	_sɯmɯŋɯp_	_*həʔaap_	_*uap_	Aceh-Chamic < MK, Cf. Khmer _sŋaap_, Bahnar _kəʔaap_; Malayic < PAN _*Suab_
year	_thon_	_*thun_	_*tahun_	All < PMP _*taqún_
yellow	_kunɛŋ, kuɲɛt_ 'tumeric'	_*kuɲit_	_*kunit,_ _kuning_^{Malay}	Malayic forms indicate _*kuniŋ_ yet Adelaar reconstructs _*kunit_ from PMP _*kuniɟ_. Both are found in Aceh.
you (pl.)	_kah_	_*hã_	_*kamu(?)_	Malayic < PAN _*kamú_, Chamic < MK(?), Aceh. obscure
you (sg.)	_gata, kah_	_*ih_	_*kau_	Malayic < PAN _*i-kaSú_, Chamic/Aceh.?

Sidwell, P. "Acehnese and the Aceh-Chamic language family". In Grant, A. and Sidwell, P. editors, _Chamic and Beyond: Studies in mainland Austronesian languages._
PL-569:211-246. Pacific Linguistics, The Australian National University, 2005. DOI:_10.15144/PL-569.211_

7 *The tones from Proto-Chamic to Tsat [Hainan Cham]: insights from Zheng 1997 and summer 2004 fieldwork*[1]

Graham Thurgood and Ela Thurgood

The Language, the People, The History

The Cham of Hainan, termed Huihui (that is, Muslim) by the Chinese, mainly live just within the Yanglan township within the Sanya municipality on Hainan, in the villages of Huihui and Huixi. These people call themselves u^{33}tsa:nʔ^{32}, an autonym composed of U 'people' + Tsat 'Cham'(< Proto-Chamic [PC] *cam 'Cham'), and their language tsa:nʔ^{32} 'Tsat' (< PC *cam 'Cham').[2]

Historical reconstruction of Chamic (Thurgood 1999:224-227) makes it clear that the Tsat represent an offshoot of the Northern Cham of the Champa Kingdom. The first mention of the Kingdom of Champa, according to Coedès (1968:42), around 190 to 193 AD, but the first reference to what was the Tsat was undoubtedly in the Chinese dynastic records, which seem to have referred to the Tsat twice, once in 986 and once in 988. These dates follow not too long after the fall of the Vietnamese, in their 'Push to the South' sacked the capital in Indrapura in 982. This fall of the capital in 982 accounts for the refugees mentioned in the Chinese dynastic records of 986 (*History of the Song Dynasty* (960-1279), which records in 986 the arrival of some Cham in Hainan from Zhancheng [=Cham City] (Zheng 1986:37)). Another group is recorded in 988 in Guangzhou (Canton). The capital then moved to the south, but in 1491, the southern capital at Vijaya also fell resulting in another influx of refugees in Hainan.

The first modern account and the first account containing linguistic data, however, is H. Stübel's short note on the language found in his 1937 work entitled *Die Li-Stämme der Insel Hainan: ein Beitrag zur Volkskunde Süd-Chinas, unter Mitwirkung von P. Meriggi,* which, despite showing no indication of the tones and being limited to a relatively small number of forms provided the basis for Benedict's 1941 identification of the language as Chamic. Then, beginning with work by Zheng and Ouyang in Ya county in 1956, who ran across the language in the course of their monumental work on the Li languages (1983b), there has been sporadic modern work on the language. Initially busy

[1] This paper was originally given at the SEALS XV (Southeast Asian Linguistics Society) conference, in Canberra 20-22, April 2005. It has been significantly improved by the comments of participants, in particular, Phil Rose, Justin Watkins, Harold Koch, Paul Sidwell, Koichi Honda, Pittayawat Pittayapom, and Charatdao Intratat.

[2] As Goschnick (1977:106) notes, other Chamic subgroups have also used 'Cham' in their name: the Cham Raglai (the Roglai; from ra 'people' + glai 'forest'), the Cham Jarai (the Jarai), the Cham Kur (Cham + kŭr 'Khmer', the Western Cham of Cambodia and Southern Vietnam), and the Cham Ro > Chru (from Cham + rɔ 'remnant').

Thurgood, G. and Thurgood, E. "The Tones from Proto-Chamic to Tsat [Hainan Cham]: Insights from Zheng 1997 and from Summer 2004 fieldwork". In Grant, A. and Sidwell, P. editors, *Chamic and Beyond: Studies in mainland Austronesian languages.*
PL-569:247-271. Pacific Linguistics, The Australian National University, 2005. DOI:*10.15144/PL-569.247*

with other work, they collected a small amount of data and left its analysis and further investigation for a later date. The couple returned in the spring of 1981 to the Yanglan Township of Ya County to carry out a detailed investigation of the language, collecting some 3000 words and some 300 sentences. The work by Zheng and Ouyang attracted much attention from scholars both inside and outside China, particularly because, despite being a Chamic language (that is, in the Malayo-Chamic subgroup of Austronesian, see Figure 1) it was fully tonal.

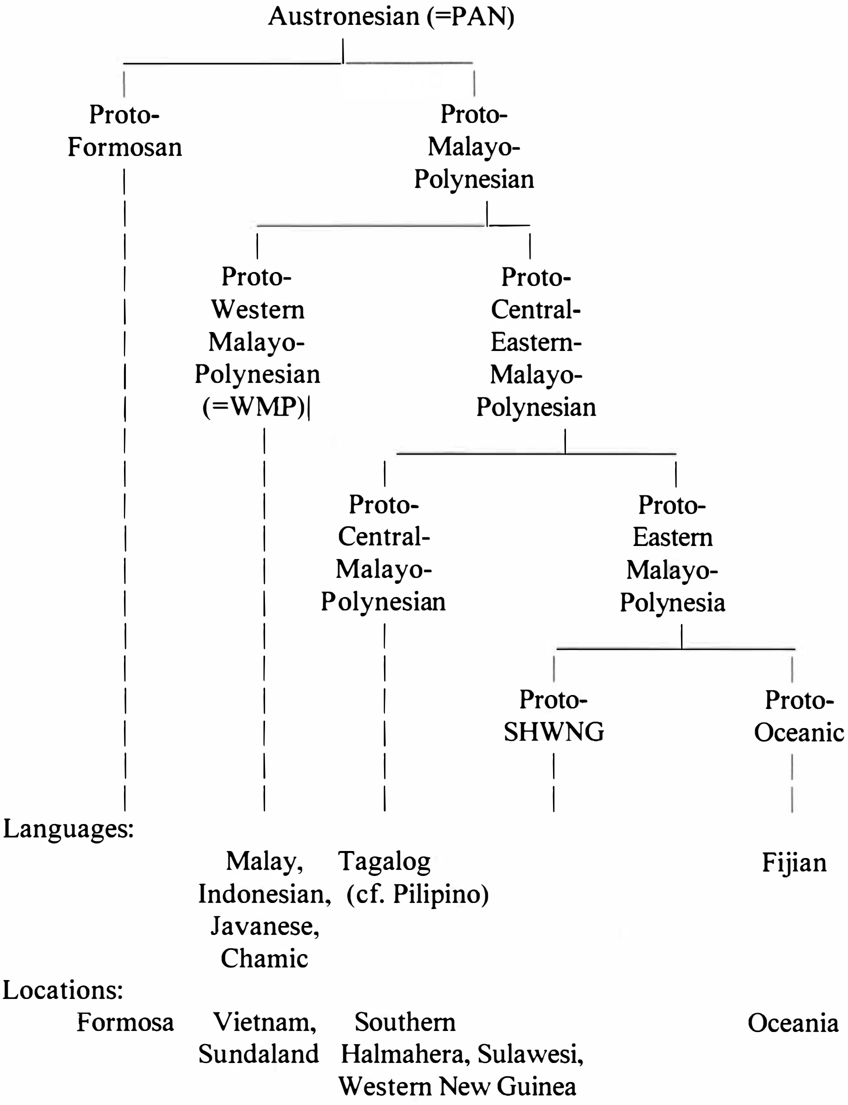

Austronesian (=PAN)
Proto-Formosan
Proto-Malayo-Polynesian
Proto-Western Malayo-Polynesian (=WMP)|
Proto-Central-Eastern-Malayo-Polynesian
Proto-Central-Malayo-Polynesian
Proto-Eastern Malayo-Polynesia
Proto-SHWNG
Proto-Oceanic
Languages:
Malay, Indonesian, Javanese, Chamic
Tagalog (cf. Pilipino)
Fijian
Locations:
Formosa
Vietnam, Sundaland
Southern Halmahera, Sulawesi, Western New Guinea
Oceania

Figure 1: *Proto-Austronesian family tree (Blust c. 1980)*

Although there was initially some debate over the status of Tsat within China, foreign scholars recognized it immediately as Chamic. What made it of particular interest in general was its complete typological restructuring under the influence of the languages of Hainan. Several scholars wrote about the syntactic restructuring (Ni Dabai 1988ab, 1990ab; Thurgood and Li, forthcoming, 2003), but the major focus was on the development of a full tonal system from a completely atonal source. This development caught the interest of a number of scholars (Benedict 1984, Haudricourt 1984, Maddieson and Pang 1993, Thurgood 1992, 1993, 1996, 1999). The work on Tsat tonogenesis was done exclusively on the basis of several early articles by Ouyang and Zheng (1983a), Zheng (1986, 1997), and Ni (1988ab, 1990ab). With the publication of Zheng's 1997 grammar, however, the data base has been expanded significantly; that expanded database coupled with the results of our fieldwork in the summer of 2004 makes it possible to paint a more detailed, richer picture of Tsat tonogenesis.

However, first, a few comments will be made on one of the important precursors to tonogenesis, the development of Tsat monosyllables from Malayo-Chamic disyllables. Contact with the iambic Austroasiatic patterned languages along the coast of Vietnam caused a switch in stress from penultimate to iambic, ultimately resulting in a reduction first to iambic and then to monosyllabic Chamic forms, an explanation that looks likely even for the pattern of Table 1, for which the disyllabic form can no longer be recovered from the data of the modern languages.

Table 1 shows the reduction of pre-Proto-Chamic forms with a medial *-h- to monosyllabic after the loss of the first syllable vowel. PAn is Proto-Austronesian, while Written Cham is the oldest written records of Cham; Malay is included simply for comparison.

PAn	Malay	PChamic	Wr. Cham	Tsat	gloss
*taqun	tahun	*thŭn	thun	t^{h}un^{33}	'year'
*puqun	pohon	*phŭn	phun	p^{h}un^{33}	'plant'
*paqit	pahit	*phiəʔ	---	p^{h}iʔ^{24}	'bitter'
*paqat	pahat	*phaːʔ	phaʔ	p^{h}aʔ^{24}	'chisel'
*paqa	paha	*pha	phā	p^{h}a^{33}	'leg, thigh'
*daqiS	dahi	*dhəi	dhei	t^{h}ay^{33}	'forehead'

Table 1. *Monosyllables from disyllables with medial *-h-.*

In a parallel way, Table 2 shows the reduction of pre-Proto-Chamic forms with a medial *-l- or *-r- to monosyllabic after the loss of the first syllable vowel. Note that the medial *-l- or *-r- is retained in Tsat as an –i- glide.

PAN	Malay	PChamic	Wr. Cham	Tsat	gloss
*baqeRu	baharu	*bahrəu	barăhau	p^{h}i^{11}	'new'
*qabaRa	---	*bara	bara	p^{h}ia^{11}	'shoulder'
*bulan	bulan	*bilaːn	bulan	p^{h}ian^{11}	'moon'
*bulu	bulu	*biləu	bulău	p^{h}iɤ^{11}	'body hair'

Table 2. *Monosyllables from disyllables with medial liquids.*

In the remaining cases, as Table 3 shows, the initial syllable is simply lost without any trace in Tsat. A glance at the Chru and Rade columns reveals the path of development: first the vowel of the initial syllable was reduced and then the whole syllable was lost.

PAN	PChamic	Wr. Cham	Chru	Rade	Tsat	gloss
*baseq	*basah	basah	pəsah	msah	saʔ^{43}	‘wet; damp’
*qubi	*hubəy	hubei	həbəi	hbei	p^{h}ay^{11}	‘taro; yam’
*quzan	*huja:n	hujan	həjăn	hjan	sa:n^{11}	‘rain’
*qumah	*huma	humā	ləma	hma	ma^{33}	‘dry field’
*lapaR	*lapa	lapa	ləpa	epa	pa^{33}	‘hungry’
*lima	*lima	limə	ləma	ema	ma^{33}	‘five’
*m-uda	*muda	medā	məda	mda	t^{h}a^{11}	‘young; unripe’
*mamaq	*mumăh	memeh	bəmah	mmah	maʔ^{43}	‘chew’
*pajay	*paday	padai	pədai	mdie	t^{h}a:yʔ^{24}	‘rice (paddy)’
*panaq	*panah	paneh	pənah	mnah	naʔ^{43}	‘(shoot) bow’
*taliS	*taləy	talei	tələi	klei	lay^{33}	‘rope; string’
*tangan	*taŋa:n	taŋin	təngăn	kngan	ŋa:n^{33}	‘hand’

Table 3. *Monosyllables with no Tsat evidence of an original initial syllable.*

This reduction of disyllables to monosyllables did not complete the restructuring of the Tsat word along the lines of the languages of Hainan—most of the finals were to be lost, but the transition to monosyllabic was complete.

The Modern Tones

The notational system used for tones in this work represents an adaptation of the tonal system found in Zheng (1997:24-25), which is itself an adaptation of Ouyang and Zheng (1983a);[3] however, it differs only in minor details, the most obvious one being that all glottal stops are overtly marked. Nonetheless, the Zheng 1997 analysis is consistent with the instrumental data obtained during our fieldwork on Hainan in the summer of 2004.[4]

Procedure

The acoustic analysis of Tsat is based on recordings during fieldwork in the summer of 2004 in Hainan consisting of words produced in citation form by six speakers (three

[3] Tones are indicated using Chao tone numbers (1930), a five point scale with 1 the lowest and 5 the highest; the first number indicates the starting point, the second indicates the ending point. Thus, a high level tone would be 55, starting high and remaining high, while a 24 tone would indicate starting slightly below the mid point and rising slightly above the mid point.

[4] Our fieldwork in the summer of 2004 was supported in part by National Science Foundation Grant NSF #606232 Endangered Languages in China. During this time Professor Sun Hongkai arranged for us to be accompanied by two distinguished Chinese scholars, Professor Ouyang Jueya and Professor Jiang Di.

female subjects (F1, F2, F3) and three male subjects (M1, M2, M3)), with ages form the early 20s to the mid 60s. All were fluent in Hainan Cham [Tsat]; all also knew Hainanese, the Min dialect of Hainan, and all knew Mandarin. Each speaker repeated each word three times. The data was recorded in a quiet room on a laptop computer using the SoundEdit software and a head mounted Telex H-831 mic. The analyses were performed using Macquirer software. The recordings were digitized at a sampling rate of 11, 025 Hz.

Prior analyses

All analyses of Tsat treat the language as essentially having five phonemic tones. Zheng (1997), for instance, posits five basic phonemic tones, plus some allophonic variation conditioned by the existence of a final stop—usually a glottal stop, see Table 4.[5] Likewise, historical analyses of the origins of Tsat tones based on Ouyang and Zheng (1983a)—Haudricourt (1984), Benedict (1984), Maddieson and Pang (1993), Thurgood (1992, 1993, 1996, 1999)—treat also Tsat as having five etymological tones: three level tones and two contour tones in checked syllables, but none of these studies contains a fully adequate treatment of the allophonic variants. However, none of the works had access to Zheng (1997); Zheng supplemented with instrumental analyses of our own fieldwork allows a much more complete picture of the Tsat tones and allotones, both synchronically and diachronically.

	/11/	/24/	/33/	/43/	/55/
'live' finals	[11~21] sa^{11} 'tea'		[33] sa^{33} 'one'		[55~45] sa^{55} 'wet'
glottal finals	[21] ta:nʔ^{21} 'intestine'	[24] saʔ^{24} 'cook'	[32] sa:yʔ^{32} 'lay bricks'	[43] saʔ^{43} 'ladder'	
stop finals (borrowings)		[24] tsat^{24} 'narrow'		[43] tsat^{43} 'photo'	

Table 4: *The five phonemic tones (< Zheng (1997), modified)*

A handful of additional forms exists with patterns that fall outside of Table 4. Without exception, these forms represent loanwords not fully assimilated into the older Tsat phonological patterns; in fact, in many such cases it is hard to distinguish code-switching from borrowing. In some instances, these historically-interesting aberrancies give clues about historical contact patterns and when this is so, it is usually commented on. However, before historical inferences are drawn about inheritance or contact on the bases of forms that pattern irregularly, a fuller description of the regularly-patterning forms is in

[5] Organizationally, this table differs from the corresponding table in Zheng in the fact that the second row here consists of syllables with final glottal stops and, notationally, in that the final glottal stops are written as such.

order.

The tone system of Tsat is one in which the diachronic origins are still reflected in the modern distribution of the phonetic, and, thus, phonemic tones: to use Thai terms, the 'live' syllables, that is, open syllables and syllables ending in nasals have one set of tones, syllables, and dead syllables, that is, checked syllables, have another set of tones. The live syllables co-occur with the so-called level tones, tones 11, 33, and 55; the contour tones co-occur with stopped syllables.

The 'level' tones. Tsat has three level tones: 11, 33, and 55. Figures 1 and 2, show the level tones for the two youngest speakers.

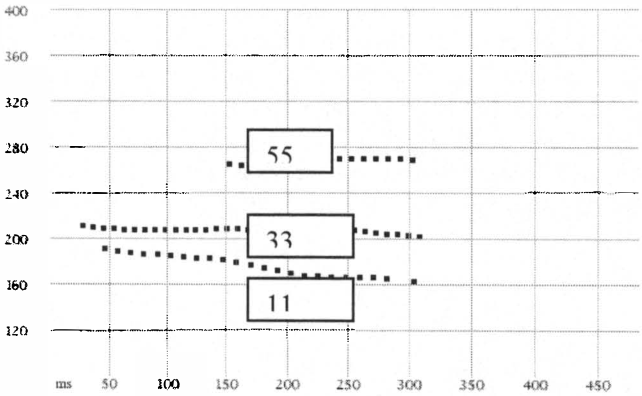


Figure 1: *The three level tones (speaker F1)*

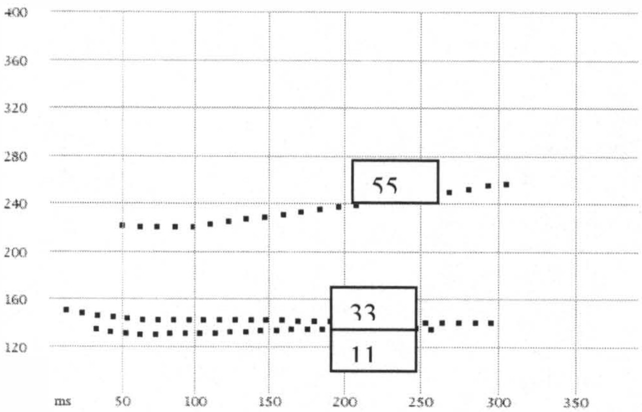


Figure 2: *The three level tones (speaker M1)*

These two speakers have different fundamental frequency (F0) ranges, as measured from the highest point of the 55 tone to the lowest point of the 11 tone. For the female speaker (F1), the highest point is ca. 320 Hz and the lowest point is ca. 168; for the male speaker

(M1), the highest point is ca. 255 Hz and the lowest point is ca. 137 Hz. Thus, the pitch range for the female speaker is 152 Hz, while for the male speaker it is 118 Hz. It is within these ranges that both the level tones and the contour tones are found.

Tone 55.

The pitch value of the 55 tone is strikingly high, separating it clearly from all the other tones (see Figures 1 and 2). In Figure 1, the 55 tone is essentially level, while in Figure 2 the 55 tone is rising, in this instance by 40 Hz. The distinctiveness of this extra-high tone, which at times has a falsetto quality to it, is immediately obvious, and has been commented on by all observers.[6] Ni Dabai (1988a) labeled it 55, but explicitly notes that it can be either 55 or 45, a characterization that matches Maddieson's instrumental work in Maddieson and Pang (1993:80), in which their figure shows it as initially rising.[7]

A historical note is in order: With the exception of one word in the data, the word zoʔ^{55} 'powder' listed in Ni (1988a:19), the 55 tone is restricted to non-stopped syllables.

Tone 33.

The pitch of the 33 tone is described as mid level although in relative terms it is sometimes a little lower in comparison with the extra-high pitch of the 55 tone. Ouyang and Zheng suggest that phonetically it might be described as 22, but leave it as 33 for reasons having to do as much with notation as phonetics; similarly, Maddieson and Pang (1993:79) note that the onset of the 33 tone is quite close to the onset of the 11 tone. As Figures 1 and 2 show the 33 is a level tone, but at times towards the end it drifts downward. When it drifts down, it would not be readily distinguishable from the checked 32 variant except by its final glottal stop.

The 32 pattern is the checked variant of the 33 tone. It is worth noting that all the syllable finals with a 32 (or 21) tone have a glottal catch when they end with a final *-n* or *-ŋ*. See the discussion of preploded final nasals below. The 32 tone variant will be discussed further with the contour tones.

Tone 11.

All authors note that the low level tone frequently is phonetically more a 21 than a 11 (see Figure 1). Maddieson's figure shows this tone as 21 and Ni's notes have it as 21.[8] Similarly, Ouyang and Zheng (1983a:32) note, even when this tone does not end in a glottal stop, the pitch is often 21. Our own recordings show low tone items with a final glottal stop as consistently falling, while there is variation between level and falling in those without a final glottal stop. When this non-checked low tone drifts down toward the end, it is largely distinguishable from the checked phonetic variant of the 11 tone, labeled

[6] Rose (1997:19) comments similarly on a super-high tone in Pakphanang Thai, noting that it is sometimes falsetto, convex in shape, and very salient. The Tsat 55 differs in that it is only sometimes convex.

[7] It is not clear to us whether the initially rising onset has any particular significance. Tones labeled 55 elsewhere also sometimes are actually more of a rising tone. For instance, in instrumental work we did on Jiamao, a Tai-Kadai (Kra-Dai) language of Hainan, the so-called 55 tone turned out upon instrumental analysis to be 45 or even 35. Whether this represents a change, notational convention, or phonetic variation is unclear.

[8] Although Ni notes the fall occurs with non-checked tone 11 forms, he observes that the fall is particularly noticeable when the form ends in a glottal stop.

21 by Zheng, only by the presence of a final glottal stop. The checked 21 variant will be further discussed with the contour tones.

The contour tones.

The remaining tones have contours: three falling: 43, 32 (the checked allophone of 33), and 21 (the checked allophone of 11); and, one rising, 24. Maddieson and Pang (1993) note quite correctly that the contour tones are always associated with checked syllables. By checked syllables, Maddieson and Pang mean Zheng's rising 24 and her falling 43, 32, and 21. Although it is not obvious from the transcriptional conventions used, Zheng and Ouyang were fully aware that these tones had a glottal final. The Zheng/Ouyang transcriptional system systematically distinguishes glottal final syllables—the labels 24, 43, 32, and 21 indicate the existence of both a contour and a final stop: if the segment does not already end in a *-p*, *-t*, or *-k*, the existence of a glottal stop is implicit. During our fieldwork together Ouyang mentioned more than once that it was unnecessary to mark final glottal stops, as they were implicit in the tone numbers.

It is important in terms of tonogenesis to note that the rising tone and the three falling tones differ in more than just pitch contour. In fact, other cues may be more salient for distinguishing the rising from the falling tones than simple pitch differences. The rising tone, which rises somewhat slowly, occurs with phonetically long vowels; the three falling contour tones, which fall somewhat abruptly, occur with phonetically shorter vowels. The patterns suggest that at the point where the contour pitch developed, the finals in question (*-ʔp, *-ʔt, *-ʔk; *-ʔn, *-ŋʔ; *-yʔ, *-wʔ) had co-articulated finals, perhaps accounting for why the effect of the glottal final of the rising tone differs from the effect of the glottal final of these falling tones. It has been noted in the tonogenesis literature that final glottal stops are sometimes related to rising contours and sometimes to falling (cf. Thurgood 2002).

Maddieson and Pang suggest on the basis of the limited data then available to them that there is a single rising tone and a single falling tone, both of which occur only in checked syllables. Essentially, this position is borne out by our study, with two qualifications: First, the claim that there is only one falling tone must be interpreted as phonemic rather than phonetic as there are three phonetically-distinct falling pitch patterns-—43, 32, and 21; these can, however, be easily phonemicized, leaving only one phonemically distinct falling tone. The six speakers we recorded all had three distinct phonetic patterns—although not for all words and not at all times, but no minimal pairs exist and the pitch patterns are only a part of a cluster of features that distinguish the various pitch patterns involved. Second, at times, all of the non-checked so-called level tones show some contour; more specifically, the so-called 55 tone is often phonetically a 45, something obvious in both Maddieson's figure and our own instrumental examination, the so-called 11 tone is often actually 21, something again in Maddieson's figure and our instrumental examination, and even the 33 tone at times noticeably drifts downward.

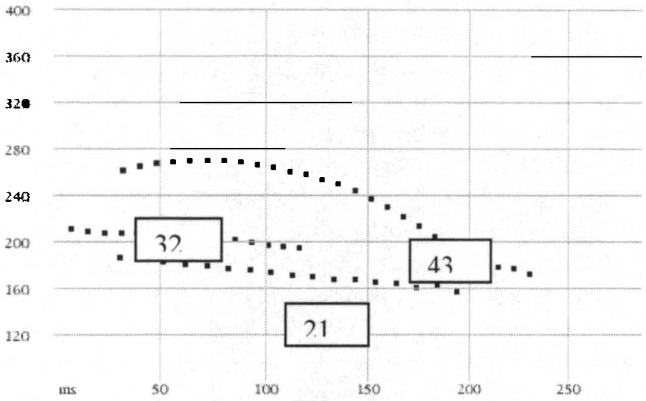


Figure 3: *Falling tones (speaker Fl)*

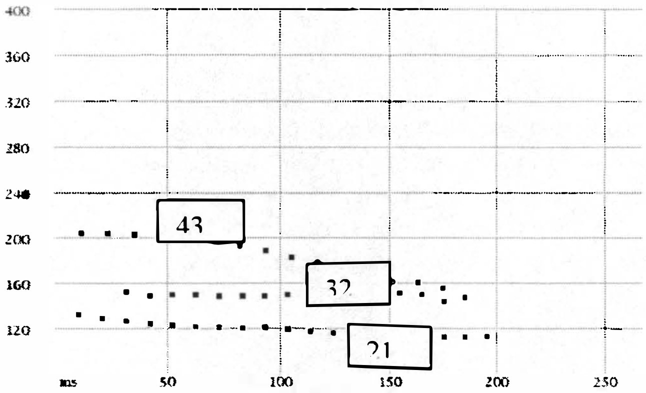


Figure 4: *Falling tones (speaker Ml)*

Tone 43.

Tone 43 is high-falling, ending in a glottal stop. Maddieson and Pang (1993) described it as 42, Ni (1988ab, 1990ab) has it as 42, and Zheng (1997) has it as 43. In our data, it varies between 42 and 43 (see Figures 3 and 4). For example, in Figure 3 tone 43 looks like 42; this particular token has a rise in frequency at the beginning of the vowel followed first by a gradual fall, and then by a relatively steep fall of 60 Hz—an overall fall of roughly 100 Hz. In Figure 4, we have 43, with a level plateau through approximately the first half of the vowel followed by a gradual decline through the second half of the vowel—an overall fall of a little over 40 Hz.

Two historical notes are in order. Forms in 43 inherited from Proto-Chamic [PC] had a PC final stop, and all 43 forms ending in *-t* or *-k* are borrowings.

Tones 32 and 21

Tones 32 and 21 are squashed into the narrow range occupied by level tones 33 and 11. For the female speaker (F1), the two falling tones are between ca. 210 Hz and 165 Hz (Figure 3); and, for the male speaker (M1), the two falling tones are between ca 150 Hz and 130 Hz (Figure 4). As a result, tones 33 and 32 are often in the same pitch range, and tones 11 and 21 are often in the same pitch range. For M1, neither the pitch shape nor the fundamental frequency differentiates between tones 33 and 32; for F1 neither the pitch shape nor the fundamental frequency differentiates between tones 11 and 21. Clearly, there are other acoustic cues that differentiate between tones 33 and 32 and between 21 and 11. These cues are provided by the final glottal stop.

The final glottal stops in tones 32 and 21 are secondarily derived either from syllables with PC final nasals or from the PC diphthongs *-ay or *-aw. Although the 32 and the 21 tones consistently show a falling contour, this by itself would not always distinguish them from 33 and 11, respectively, as at times these also manifest a falling pitch pattern. Our recordings and measurements show the items without a final glottal stop varying between a level and a falling pitch pattern, while those with a final glottal stop are consistently falling and accompanied by creakiness (discussed below).

A historical note: While at times all six of our speakers keep these three falling pitch patterns distinct, on occasion some of them do not; further, the historical evidence indicates that at least for some speakers in some words the 32 and the 21 have begun to merge. This is an area for further investigation.

Tone 24.

Tone 24 is the mid-rising tone, ending in a glottal stop. As Maddieson notes in Maddieson and Pang (1993:80), the sources indicate that all rising tones occur in checked syllables, which Maddieson treats as a 24 tone. In the most recent work, Zheng (1997:24) also has a single 24 tone found in checked syllables. While it is not particularly clear from their notational system, it is clear from conversations with both Zheng and Ouyang, that they too regard the 24 as ending in a glottal stop. For examples, see Figures 5 and 6.

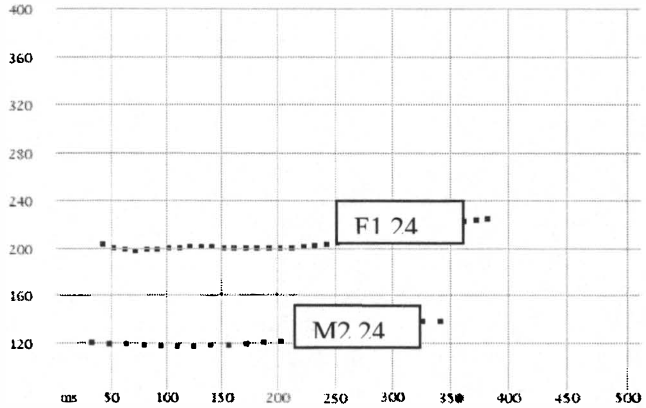


Figure 5: *The 24 rising tone (two tokens; speakers F1 and M2)*

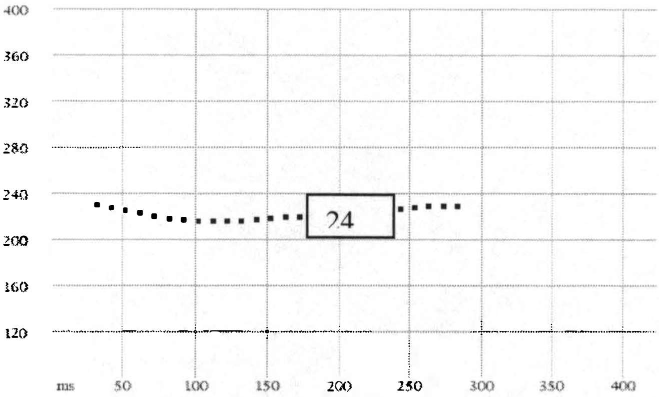


Figure 6: *The 24 rising tone (speaker F2)*

In our data, tone 24 either displays a level plateau through approximately the first half of the vowel followed by a rise in frequency (Figure 5) or it falls for the first third of its duration, flattens out for the second third of its duration, and goes back up for the last third of its duration (Figure 6). The tone itself is easily distinguished, although at times it patterns enough like the unchecked 33 tone that the glottal final is needed to distinguish it.

On inherited PC forms, the final glottal stop is the remnant of old final stops. All 24 forms ending in *-t* or *-k* are borrowings.

Creaky voice

All contour tones correlate with a creaky voice quality, which in Tsat is always associated with a final glottal stop. The non-modal phonation types, including creaky voice, are described by their unique spectral properties, including periodicity; overall acoustic intensity; level of intensity of the first harmonic (H1), level of intensity of the second harmonic (H2), and level of intensity of the highest harmonic both in the first formant (F1) and in the second formant (F2). In this study, the voice quality effects of a glottal stop on the preceding vowel were analyzed on the FFT power spectra with superimposed LPC spectra using a bandwidth of 43 Hz and a window with 256 points. We concentrated on the creaky phonation of the /a/ vowel, because intensity differences shown by FFT spectra are best shown for low vowels, for which the frequency location of F1 is far enough from H1 not to influence the amplitude of H1 (Jessen and Roux 2002; Ladefoged 2003). The voice quality effects induced by a glottal stop were analyzed on the second FFT spectrum computed in the middle of the vowel. For each vowel token, a number of amplitudes were measured from the spectra: the amplitude of the first harmonic (H1) of the fundamental frequency, the amplitude of the second harmonic (H2), and the amplitudes of the first formant (F1) and of the second formant (F2).

Table 5 gives the list of words in which the /a/ vowel was analyzed. The transcription is a modification of the notational system of Zheng (1997).

[a̰]			[a]		
pa:ʔ^{24}	(PC *pa:t)	‘four’	pa^{33}	(PC *kapa:s)	‘cotton’
p^{h}a:ʔ^{24}	(PC *p^{h}a:t)	‘chisel’	p^{h}a^{11}		‘bland’
ta:ʔ^{24}	(PC *rəta:k)	‘beans, peas’	ta^{33}	(PC *ʔata:s)	‘far’
taʔ^{43}		‘pillow’	ta^{11}	(PC *buta)	‘blind’
t^{h}aʔ^{43}		‘white gourd’	t^{h}a^{33}		‘speech’

Table 5: */a:ʔ/ and /a/*

In terms of spectral characteristics the biggest difference between creaky voice and modal voice lies in the comparison of the intensity values of higher frequencies to the intensity values of lower frequencies; this is the case in Tsat. For the female speakers (Figure 7), when we compare H1 with the highest harmonic in the first *formant*, for creaky voice, the highest harmonic in the first formant is 11.8 dB above H1 (H1-F1 = -11.8 dB), while for modal voice F1 is only 2.8 dB above H1 (H1-F1 = -2.8 dB). The difference between creaky and modal is even more marked when we compare H1 with the highest harmonic in the second formant. For creaky voice, F2 is 7.7 dB above H1 (H1-F2 = -7.7 dB). In contrast, for modal voice, F2 is 3.2 dB below H1 (H1-F2 = 3.2 dB). For the male speakers (Figure 8), for creaky voice, when we compare H1 with the highest harmonic in the first formant, the highest harmonic in the first formant is 11.6 dB above H1 (H1-F1 = -11.6 dB), while for modal voice F1 is 6.3 above H1 (H1-F2 = -6.3 dB). Similarly, for creaky voice, when we compare H1 with the highest harmonic in F2, F2 is 9.3 dB above H1 (H1-F2 = -9.3 dB), while for modal voice, F2 is 6.2 dB above H1 (H1-F2 = -6.2 dB).

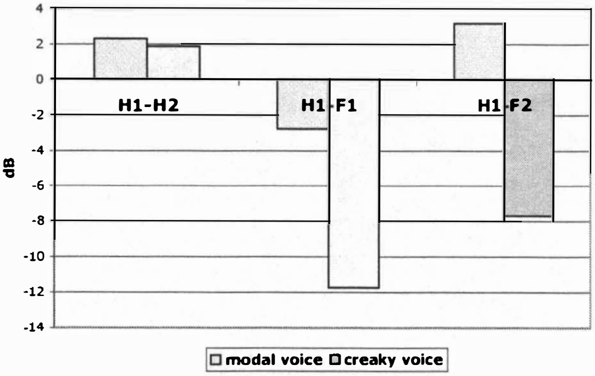


Figure 7: *Differences in H1-H2, H1-F1, and H1-F2 for female speakers*

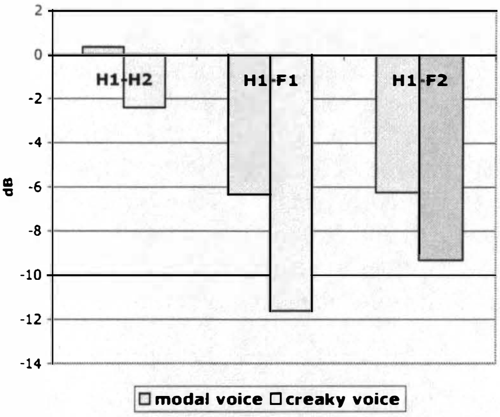


Figure 8: *Differences in H1-H2, H1-F1, and H1-F2 for male speakers*

During our recording sessions we often noticed, not only that the /a:yʔ/ was creaky, but also that /a:yʔ/ was sometimes pronounced as [a:ʔ]. To examine both features, we compared /a:y/ words with a final glottal stop to /ay/ words without a following glottal stop. For the words used, see Table 6.

pa:yʔ^{32} (PC *tapay) ‘wine, liquor’	pay^{33} (PC *tampɛy) ‘to winnow’
ta:yʔ^{32} (PC *hatay) ‘liver’	pay^{33} (PC *lumpɛy) ‘to dream’
ta:yʔ^{32} (PC *matay) ‘to die’	t^{h}ay^{11} (PC *ʔadɛy) ‘younger sibling’
t^{h}a:yʔ^{32} (PC *paday) ‘paddy’	t^{h}ay^{33} (PC *ʔadhɛ̃y) ‘forehead’

Table 6: */a:yʔ/ versus /ay/*

Figure 9 presents the distribution of [a:yʔ] and [a:ʔ] in 72 forms produced by the six subjects (4 tokens of /a:yʔ/ repeated 3 times by 6 speakers = 72 tokens of /a:yʔ/). The two older male speakers (M3 and M2) pronounce the /a:yʔ/ diphthong variably as either [a:yʔ] or [a:ʔ] (with slight preference for [a:yʔ] for M2). The remaining four speakers favor [a:ʔ] over [a:yʔ], with three of the four speakers (M1, F3, and F2) only producing [a:yʔ] once, interestingly all in the same word [t^{h}a:ʔ^{32}] ‘paddy’. The youngest subject (F1) did not produce [a:yʔ] at all.

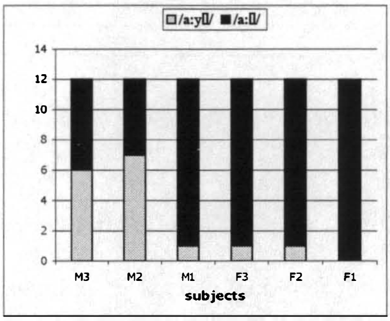


Figure 9: *Frequency of [a:yʔ] versus [a:ʔ] among the six Tsat speakers*

The characteristics of the Tsat creaky /a:yʔ/ can be seen in the spectrogram and waveform of /t^{h}a:yʔ^{32}/ 'paddy' in Figure 10. For comparison, the spectrogram of modal /ay/ in /t^{h}ay^{33}/ 'forehead' is also given. The spectrogram of /t^{h}a:yʔ^{32}/ 'paddy' shows the sequence of diphthong + glottal stop realized as a modal vowel (between 150 and 209 milliseconds) which then becomes creaky (between 209 and 240 milliseconds) before turning into a glide; thus phonetically [t^{h}aa̰yʔ]. The waveform corresponding to the portion of the spectrogram between 270 and 320 milliseconds encompasses both modal and creaky phases. The phonetic transcriptions carefully positioned above the spectrograms indicate the approximate location of different acoustic events.

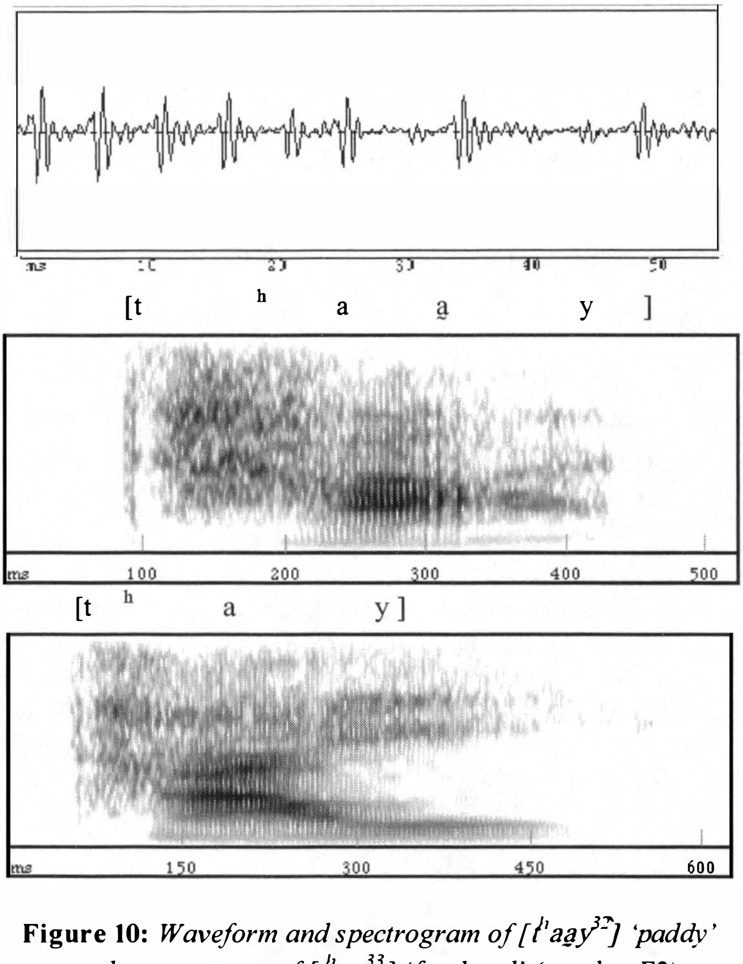


Figure 10: *Waveform and spectrogram of [t^{h}aa̰y^{32}] 'paddy' and spectrogram of [t^{h}ay^{33}] 'forehead' (speaker F2)*

The pitch periods of the creaky phase are irregular in terms of their duration and considerably longer than those of the modal phase. They are also relatively infrequent compared to the pitch periods of the modal phase. The increased length of the pitch periods indicates the lowered fundamental frequency values of the creaky [a̰] vowel.

A different phonetic realization of /a:yʔ/ is illustrated in Figure 11. In this case, /a:yʔ/ is pronounced as [aʔa̰]. The waveform and the corresponding spectrogram encompass first a modal phase (between 20 and approximately 125 ms), followed by a glottal stop, and then a creaky phase (between 150 and approximately 250 ms).

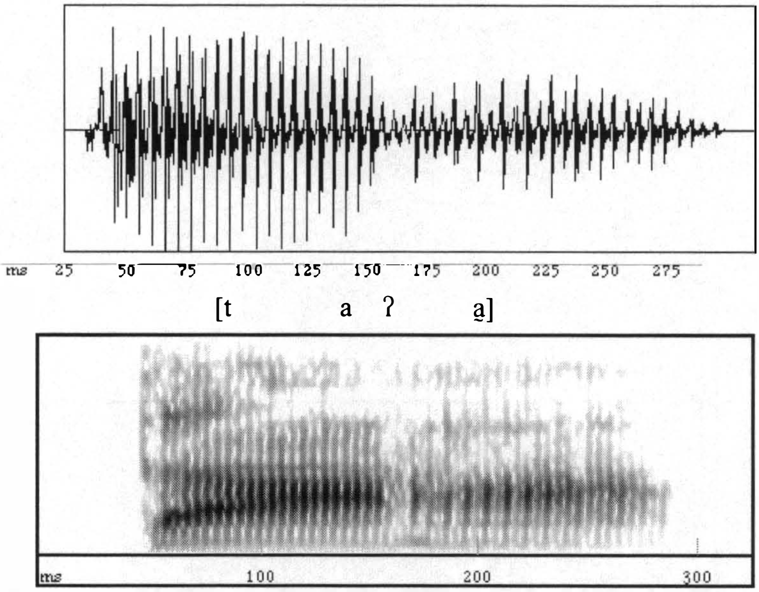
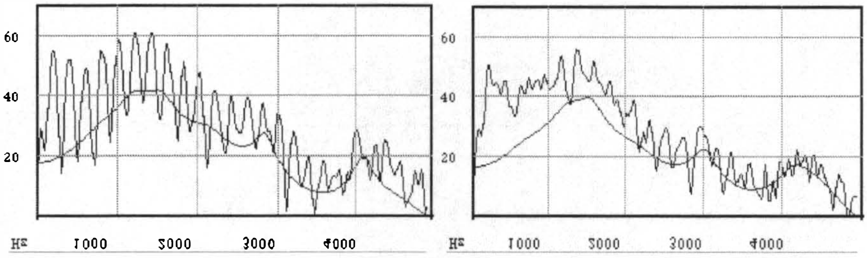


Figure 11: *Waveform and spectrogram of [taʔa̰ʔ^{32}] 'liver' (speaker F1)*

The FFT spectra with the superimposed LPC spectra of the modal phase and the creaky phase of /a/ are given in Figures 12-13. A shallower spectral tilt and a lower first harmonic (H1) clearly differentiate the creaky phase from the modal phase.



Figures 12-13: *FFT spectra with LPC spectra of modal and creaky /a/ in [taʔa̰] 'liver' (speaker F1)*

Creakiness in /a:wʔ/ was analyzed on the basis of the words given in Table 7. Table 7 also gives the words with /aw/ without a following glottal stop, the spectrograms of which were compared with the spectrograms of /a:wʔ/.

/a:wʔ/		/aw/	
ta:wʔ^{21} (PC *pataw)	'master, lord'	taw^{11} (PC *katɔw)	'louse'
ta:wʔ^{21} (borrowing)	'10 liters, clf.'	taw^{11} (PC *kukɔw)	'fingernail'
tʰa:wʔ^{21}	'hide something'		
tʰa:wʔ^{21}	'avoid (rain)'		

Table 7: */a:wʔ/ versus /aw/*

In Figure 14, /a:wʔ/ pronounced as [a:w̰ʔ] is illustrated and contrasted with modal /aw/ pronounced as [aw]. In the spectrogram of /tʰa:w̰ʔ^{21}/ 'hide something', the pitch periods of the glide have a greater distance between the vertical striations than the pitch periods of the glide in /taw^{11}/ 'louse'.

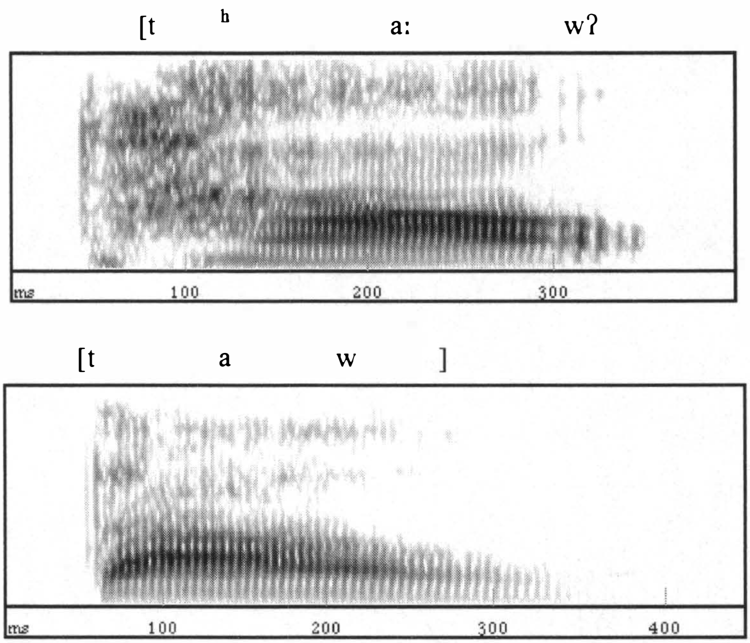


Figure 14: *Spectrograms of [tʰa:w̰ʔ^{21}] 'hide' and [taw^{11}] 'louse' (speaker F1)*

In Figure 15, /a:wʔ/ pronounced as [aa̰ʔ] is illustrated. The waveform and the corresponding spectrogram of /ta:wʔ^{21}/ 'master, lord' show a modal phase of /a/ followed by a creaky phase that culminates in the final glottal closure.

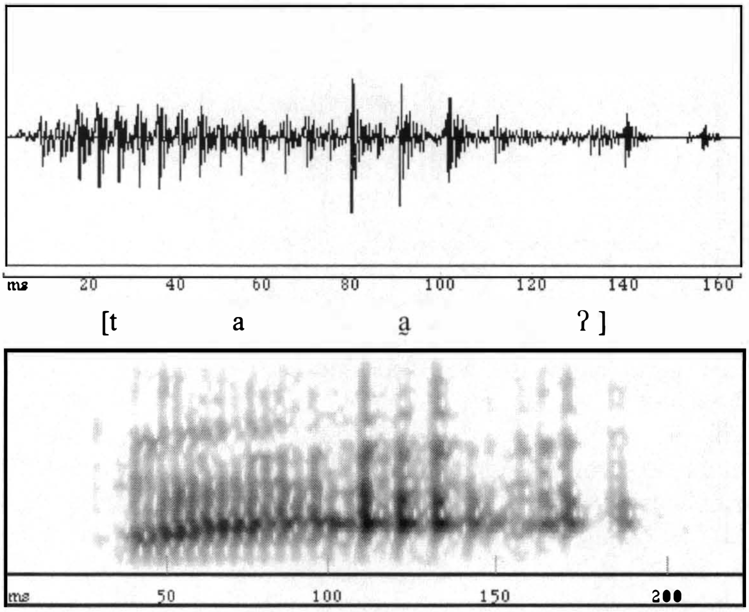


Figure 15: *Waveform and spectrogram of /ta̰:wʔ^{21}/ 'master, lord' (speaker F3)*

The /a:wʔ/ can also be pronounced as [a̰:ʔ] with creakiness spreading over the whole vowel. Figure 16 presents the waveform of the creaky vowel [a̰] in /ta:wʔ^{21}/ 'master, lord' taken over a 58 millisecond interval centered around the middle of the vowel. As comparison, Figure 17 presents the modal [a] vowel in /ta^{33}/ 'far' also taken over a 58 millisecond interval centered around the middle of the vowel. Creaky voice in Figure 16 is readily differentiated from the modal voice in Figure 17 by its irregularly spaced glottal pulses and reduced acoustic intensity relative to modal voice.

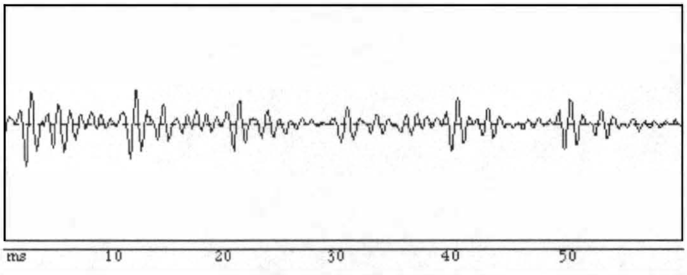


Figure 16: *Waveform of /a/ in /ta:wʔ^{21}/ 'master, lord' (speaker M3)*

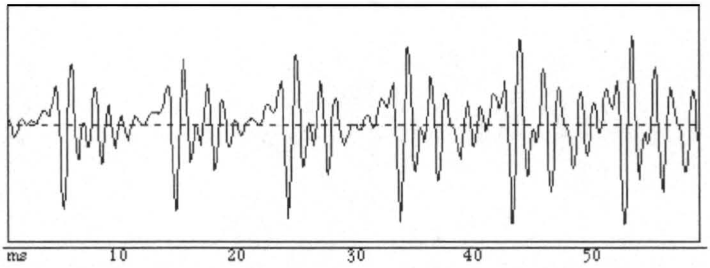


Figure 17: *Waveform of /a/ in /ta^{33}/ 'far' (speaker M3)*

Figures (18-19) show the FFT spectra with the superimposed LPC spectra, measured in the middle of /a/ in /ta:wʔ^{21}/ 'master, lord' and /ta^{33}/ 'far'. The creaky vowel [a] is characterized by a bigger increase in intensity as one moves from H1 to H2, to F1, and to F2 for the creaky vowel (Figure 18). In contrast, the modal voice shows a relatively small increase in intensity moving from H1 to H2 and a drop in intensity moving from F1 to F2. Also, the first harmonic is of lower frequency for the creaky [a̰] than for the modal [a].

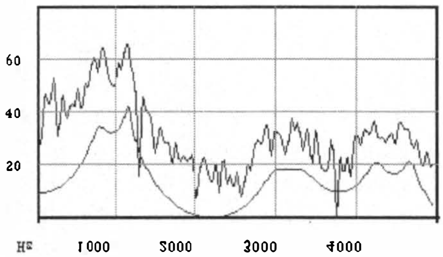


Figure 18: *Creaky [a̰] (speaker M3)*

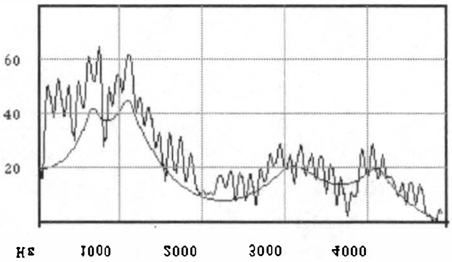


Figure 19: *Modal [a] (speaker M3)*

In Figures 20-21, for clarity [a̰] as the phonetic realization of /a:ʔ/ is referred to as creaky voice1; and, [a̰] as the phonetic realization of /a:wʔ/ is referred to as creaky voice2. The figures show that in terms of spectral characteristics, /a:wʔ/ pronounced as [a̰] patterns with the creaky [a̰] earlier discussed.

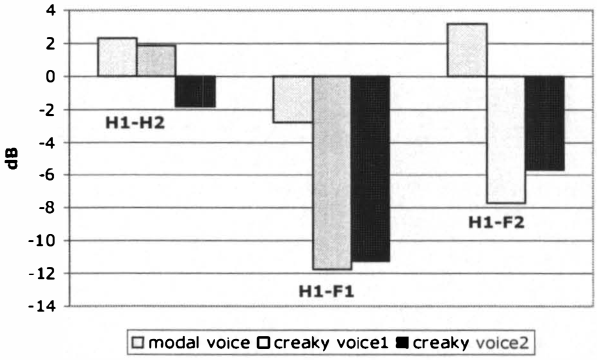


Figure 20: *Differences in H1-H2, H1-F1, and H1-F2 amplitude for female speakers*

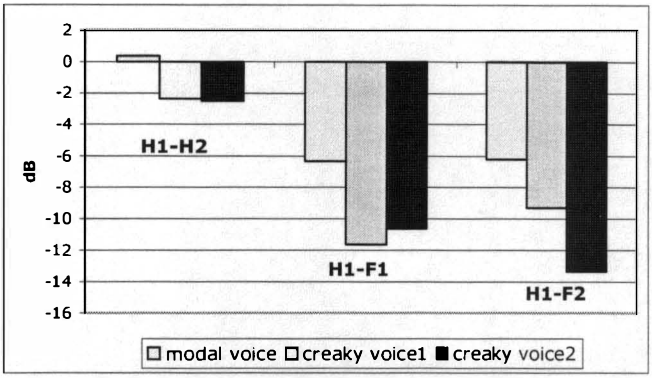


Figure 21: *Differences in H1-H2, H1-F1, and H1-F2 amplitude for male speakers*

Summary of *-a:y? and *-a:w?.

PC *-ay and *-aw developed glottal stops within the history of Tsat. The modern Tsat reflexes are -a:y? and -a:w?, respectively, but both exhibit considerable interspeaker and cross-speaker variation. In all cases, the glottal stop is consistently associated with creaky voice quality. This creaky voice quality begins at the back: sometimes only the final glide is creaky, sometimes the creaky voice ranges forward enough so that the second half of the long [a:] is creaky, and sometimes the whole vowel plus the glide is creaky. Also, in some cases, the final glide is voiceless. And, in some cases, presumably masked by the creaky voice and the voicelessness, the final glide has disappeared completely, leaving only the long monophthong with creaky voice. In summary, the two PC diphthongs *-ay and *-aw developed final glottal stops; in some cases the final glides are manifested as voiceless; and, in other manifestations, the glides have dropped, leaving only a long, monophthong behind. Thus, PC *-ay and *-aw have developed into Tsat -a:y? and -a:w? and this Tsat pair is in the process of developing into the long monophthong -a:?.

Tonogenesis

The history of Tsat tones correlates directly with the interaction of finals, initials, and phonation types. Of these, only the finals are still represented as such in Tsat; the initials and finals, however, are present in the reconstruction of PC, while the phonation types are clear from an examination of the other Chamic dialects.

The three so-called level tones have straightforward origins (Table 8). Items ending in a PC *-h have a 55 tone. Open syllables or syllables ending in a simple nasal, not a preploded final nasal, split into tones 11 and 33. The condition for the split is straightforward: syllables with a PC voiced obstruent initial developed breathy voice, which was accompanied by lower pitch; if in disyllabic items the initial of the first syllable was voiced, this led to breathy voice—not retained in Tsat, then the breathy voice spread to the next syllable, and then the breathy voice produced the low pitch, that is, the 11 tone.

final *-h:	[55~45]
*bah >	p^{h}a^{55}
*pah >	pa^{55}
other initials:	[33]
*ma >	ma^{33}
*pa >	pa^{33}
voiced obstruent initials:	[11~21]
*dapa (spreading)	pa^{11}
*ba (same syllable)	p^{h}a^{11}

Table 8: *Pitch patterns of PC *-h, and of open and nasal-final syllables*

The PC final stops *-p, *-t, *-k, and *-ʔ are manifested in Tsat as tones 43 and 24, with the 43 tone emerging if the syllable initial was a voiced obstruent and the 24 tone emerging otherwise, depending upon whether the initial was a voiced obstruent or not; again, if in disyllabic items the initial of the first syllable was voiced, this led to breathy voice, the breathy voice spread to the next syllable, and the breathy voice produced the low pitch, that is, the 43 tone. (see Table 9 below).

other initials:	[24]
*mak >	maʔ^{24}
*pak >	paʔ^{24}
voiced obstruent initials:	[43]
*bak > (same syllable)	p^{h}aʔ^{43}
*dapak > (spreading)	paʔ^{43}
*dapay > (spreading from first-syllable)	p^{h}a:yʔ^{21} (> p^{h}a:yʔ^{32})

Table 9: *Pitch patterns of PC final stops*

Preploded final nasals.

Another source of final glottal stops is from syllable-final preploded nasals, an earlier feature of Tsat subgroup that was almost gone by the time Ouyang and Zheng (1983a) began working on the language in the 1980s. However, Ouyang and Zheng still managed to record a handful of forms (see Thurgood 1999:165):

tatn^{33} la:n^{11}	'section'	---
tsiakŋ^{33} lai^{11}	'where'	---
t^{h}okŋ^{33}	'knife'	PC ^{x}*dhɔŋ
t^{h}atn^{33}	'extinguish'	PC *padam

Table 10: *The attested forms with preploded final nasals*

The PC nasal finals *-am, *-an, *-aŋ, *-ɔŋ and the resonant finals *-al, and *-ar initially developed into preploded finals and then into final glottal stops. The *-al and *-ar, of course, first went to *-an before becoming preploded final nasals. The development of preploded final nasals dates back to the P-Roglai-Tsat subgroup of Chamic (for discussion of preploded final nasals see Thurgood 1999:164-177), but the subsequent developments described here are unique to Tsat.

The data on Tsat and on Northern Roglai indicates that all final nasals first developed into preploded final stops followed by homorganic nasals and then in most contexts these highly-marked codas were simplified. In fact, except after short̤ ɔ and the short -a- in *-am, *-an, *-aŋ, *-al > *-an, and *-ar > *-an the complex coda was simplified (Thurgood 1999:164-177), leaving *-kŋ and *-tn (the final -m merged with final -n). The *-k- and *-t- then became glottal stops, hence the modern forms. The forms with these glottal finals have developed into 21 and 32 tones, with the split again depending on the presence or absence of an earlier voiced obstruent initial, respectively.

	Proto-Chamic	Tsat	gloss
other initials:	*cam	tsa:nʔ^{32}	'Tsat'
	*masam	sa:nʔ^{32}	'sour; vinegar'
	*klam	kianʔ^{32}	'dark; afternoon'
voiced obstruents:	*dar	t^{h}a:nʔ^{21}	'encircle'
	*padam	t^{h}a:nʔ^{21}	'extinguish'
	*hadaŋ	t^{h}a:ŋʔ^{32}	'charcoal'
voiced obstruent initial spreading	*gunam	na:nʔ^{21}	'cloud'
	*dalam	la:nʔ^{21}	'deep; inside'
	*gatal	ta:nʔ^{21}	'itchy'

Table 11: *Tones from PC forms with glottalized stops with final nasals, *-l, and *-r*

A third internal source was the development of final glottal stops from the epenthesis of the PC diphthongs *-ay and *-aw.

other obstruent initials	[32]
*pay >	pa:yʔ^{32}
voiced obstruent initials	[21]
*bay >	p^{h}a:yʔ^{21} (> p^{h}a:yʔ^{32})

Table 12: _Pitch patterns from PC forms *-ay > -a:yʔ and *-aw > -a:wʔ_

Table 13 provides examples of these developments, along with the PC reconstructions. Some of these forms pre-date PC, but those marked with #* only date back as far as PC.

	Proto-Chamic	Tsat	gloss
other initials	*maray	za:yʔ^{32}	'come'
	*matay	ta:yʔ^{32}	'die'
	*kakay	ka:yʔ^{32}	'foot; leg'
	*tapay	pa:yʔ^{32}	'rice wine'
	*hatay	ta:yʔ^{32}	'liver; heart'
	*haway	va:yʔ^{32}	'rattan'
	*naw	na:wʔ^{32}	'go; walk'
	#*paʔdaw	(kia^{33})ʔda:wʔ^{32}	'warm, hot'
voiced obstruents:	#*gay	k^{h}a:yʔ^{21}	'walking stick'
	*paday	t^{h}a:yʔ^{21}(>^{32})	'rice (paddy)'
	*glay	t^{h}a:yʔ^{21}(>^{32})	'forest; wild'
voiced obstruent initial spreading	*gatal	ta:nʔ^{21}	'itchy'

Table 13: _Examples of tones with glottal stops from PC *-ay and *-aw_

The fourth source of glottal finals is borrowings. That study is in progress; while it is easy to spot many of the borrowings, it is often quite difficult trying to determine where they are borrowed from and when. Presentation and evaluation of these forms will have to be left to another time.

Conclusions

The addition of the data from Zheng (1997) and from the summer 2004 fieldwork gives us a much clearer, much richer picture of the Tsat reflexes of Proto-Chamic and thus of Tsat tonogenesis. The three distinguishable falling tones are of particular interest. The Tsat reflexes of proto-Chamic, although not presented here, require no significant adjustments of Proto-Chamic and are straightforward, making the segmental origins of the various phonemic and subphonemic tones non-problematic. In fact, the relationships, although richer and more detailed, remain remarkably clear.

Several other points of interest emerged. First, it is speculated that the association found in the literature between final glottal stops and falling, rather than rising, tones might be, upon closer inspection an association between final glottal stops co-articulated with an oral closure of some kind, rather than simply a glottal stop. This would certainly account for the Tsat data, and, if co-articulation of oral final stops with glottal stops is as widespread in Southeast Asia as it now appears to be, it would account for many of the reported instances of glottal stops associated with falling contours.

Second, the variation in the Tsat data found in the reflexes of PC *-ay and *-aw show a path from diphthong to diphthong with a final glottal stop, to a monophthong with a final glottal stop. Here, the data are rich enough to posit a plausible path of change for the developments to have followed.

Finally, there is the extra high 55 tone, which is of wider interest largely because of its apparent rarity.

References

Benedict, Paul K. 1941. A Cham colony on the island of Hainan. *Harvard Journal of Asiatic Studies* 4:129-34.

-----. 1984. Austro-Tai parallel: A tonal Cham colony on Hainan. *Computational Analyses of Asian & African Languages* 22:83-86.

Burling, Robbins. 1966. The addition of final stops in the history of Maru. *Language* 42.3:581-586.

Chao, Yuen Ren. 1930. A system of tone letters. *Le Maître Phonetique.* Troisème série. 30.24-27.

Coedès, Georges. 1968. *The Indianized states of Southeast Asia.* Ed. Walter F. Vella. Trans. Susan Brown Cowing. Kuala Lumpur: U. of Malaya Press.

Goschnick, Hella. 1977. Haroi clauses. *Papers in South East Asian Linguistics No. 4: Chamic Studies.* Edited by David Thomas, Ernest W. Lee, and Nguyen Dang Liem. Pacific Linguistics Series A, No. 48:105-124.

Haudricourt, André-G. 1954. De l'origine des tons viêtnamien. *Journal Asiatique* 242:69-82.

-----. 1984. Tones of some languages in Hainan. *Minzu Yuwen* 4:17-25. Also published in *Bulletin de la Société de Linguistique de Paris* as "La tonologie des langues de Hainan." 79.1:385-394.

Jessen, Michael and Justus C. Roux. 2002. Voice quality differences associated with stops and clicks in Xhosa. *Journal of Phonetics* 30, 1-52.

Ladefoged, Peter. 2003. *Phonetic Data Analysis. An Introduction to Fieldwork and Instrumental Techniques.* Blackwell Publishing Ltd.

Maddieson, Ian and Keng-Fong Pang. 1993. Tone in Utsat. *Tonality in Austronesian Languages.* Edited by Jerry Edmondson and Ken Gregerson. Oceanic Linguistics Special Publication No. 24. Honolulu, Hawaii: University of Hawaii Press. pp. 75-89.

Ni, Dabai. 1988a. The genealogical affiliation of the language of the Hui people in Sanya Hainan. *Minzu Yuwen* 2:18-25.

-----. 1988b. The Kam-Tai of China and the Hui of Hainan. *Bulletin of the Central Institute of Minorities* 3:54-65.

-----. 1990a. The origins of the tones of the Kam-Tai languages. ms.

-----. 1990b. The Sanya (= Utsat) language of Hainan island: a living specimen of a linguistic typological shift. ms.

Ouyang, Jueya and Zheng Yiqing. 1980. *A brief description of Li (Hainan).* Chinese Minority People's Language, Basic Description Series. Beijing.

-----. 1983a. The Huihui speech (Tsat) of the Hui nationality in Yaxian, Hainan. *Minzu Yuwen* 1:30-40.

-----. 1983b. Survey of the Li (=Hlai) languages. Beijing.

Rose, Phil. 1997. A seven-tone dialect in Southern Thai with Super-High: Pakphanang tonal acoustics and physiological inferences. In *Southeast Asian Linguistics Studies in Honour of Vichin Panupong*, edited by Arthur S. Abramson. Chulalongkorn University Press. Pp. 191-208.

Stübel, Hans. 1937. *Die Li-Stämme der Insel Hainan: ein Beitrag zur Volkskunde Süd-Chinas, unter Mitwirkung von P. Meriggi.* Bern: Klinkhardt and Biermanm Verlag. (Reproduced 1976, Taipei, Orient Cultural Services, 2 volumes. (*Asian folklore and social life monographs 83*).

Thurgood, Graham. 1992. From atonal to tonal in Utsat (a Chamic language of Hainan). *Proceedings of the Eighteenth Annual Meeting of the Berkeley Linguistics Society. February 14-17, 1992.* Special Session on the Typology of Tone Languages. Edited by Laura A. Buszard-Welcher, Jonathan Evans, David Peterson, Lionel Wee, and William Weigel. Pp. 145-156.

-----. 1993. Phan Rang Cham and Utsat: tonogenetic themes and variants. *Tonality in Austronesian Languages.* Edited by Jerry Edmondson and Ken Gregerson. Oceanic Linguistics Special Publication No. 24. Honolulu, Hawaii: University of Hawaii Press. Pp. 91-106.

-----. 1996. Language contact and the directionality of internal 'drift': the development of tones and registers in Chamic. *Language* 71.1:1-31.

-----. 1999. *From Ancient Cham to modern dialects: Two thousand years of language contact and change. With an appendix of Chamic reconstructions and loanwords.* Oceanic Linguistics Special Publications, no. 28. June 1999, 6 x 9. ISBN: 0-8248-2131-9. Honolulu: University of Hawai'i Press. 403pp.

-----. 2002. Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica* XIX.2:333-363.

Thurgood, Graham and Fengxiang Li. forthcoming. From Malayic to Sinitic: The Restructuring of Tsat under Intense Contact. *Proceedings of the Southeast Asian Linguistics Society XII*, University of Arizona, Tempe, Arizona. 13 pp.

-----. 2003. Contact induced variation and syntactic change in the Tsat of Hainan. In Language variation: *Papers on variation and change in the Sinophere and the Indosphere in honour of James A. Matisoff.* Edited by David Bradley, Randy LaPolla, Boyd Michailovsky, and Graham Thurgood. Pacific Linguistics. Research School of Pacific and Asian Studies., Pp. 285-200.

Zheng, Yiqing. 1986. A further discussion of the position of Huihui speech and its genetic relationship. *Minzu Yuwen* 6:37-43.

-----. 1997. *Huihui Yu Yanjiu* [A Study of Tsat]. Shanghai Yuandong Chuban She [Shanghai Far East Publisher]

Zhangmianyu Yuyin he Cihui (= ZMYYC) Tibeto-Burman phonology and lexicon. 1991. China Social Sciences Press.