

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 modern 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 language change. If Rade is included and compared with the rather conservative 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.

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 they 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 Acehese, 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 Acehese (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 Acehese, 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 northern 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 Chamic 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 have 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 traditional 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 expressed 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 feature 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 /ʃa-/ has remained /ʃa-/. 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 Rang 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 preglottalised 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 Malayan 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 original 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 irregularities 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 innovations 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 'Austic' (= Austronesian plus Austro-Asiatic) and after we disregard the shared morphemic items which have been posited as existing within Austic, 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 reconstruction). 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 origins of some may simply be waiting to be recognised: for instance, the widespread form $\gamma\alpha\omega$ '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 languages 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-Malayo-Chamic 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-known 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 /b/ (/buk/, 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 /d/ in the Malayic stratum of Chamic languages, although there are 19 word-initial occurrences of both these sounds (meaning 19 occurrences for /b/ and 19 others for /d/) 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 /d/ 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. equivalents:	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. equivalents:	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 subgroup 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 numerals 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 /b d/ 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 burn; 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 language, 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 earlier³. 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' (*duduk*), '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 Javanese 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 undergone 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 *asj*, *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 *z̄ya: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	N Gen Person
NA*	<i>NA</i>	A N	<i>NA</i>	<i>NA</i>	<i>NA</i>
NNum	<i>N Num Cl</i>	Num Cl N	Num Cl N	<i>N Num Cl</i>	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	<i>V Neg (C)</i>	Neg V	Neg V/ <i>Neg C</i>	<i>V Neg(C)</i>	Neg V; <i>V Neg</i>
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?	<i>Some</i>	no	<i>some</i>	<i>some</i>	no
Suffixes?	<i>No</i>	some	<i>no</i>	<i>no</i>	<i>no</i>

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:	<i>Subj (TMA) Verb Obj</i>
Malay:	(tma) <i>Subj (TMA/Modal) (Prefix) Verb Obj</i>
Tagalog:	(TMA) (Voice) <i>Verb Subj (Object Marker) Obj</i>
Chrau:	<i>Subj (Preverb) (TMA) (Auxiliary) Verb Obj</i>
Khmer:	<i>Subj (TMA) Verb Obj</i>
Vietnamese:	<i>Subj (TMA) Verb Obj</i>

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' have 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 features 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 /ɲ-/ 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 superlatives 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 *frèv* 'new' (or the less allegro form *pihrèv*), 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 traditional 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	kēkuih
‘damp’	*basah	*basah	mēsah
‘salt’	*qasiRa	*sira	hra
‘thorn’	*duRi	*durey	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 *duduk* '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 Ronglai, 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 Rglai 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 Rglai from a form which is phonologically identical to the Rglai (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 Rglai, 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 Rglai 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 Rglai 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 monosyllabicity. But we should never forget that Tsat, like its sister language Rglai, 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 shared 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 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 donor 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 the typological 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 Zamboangueno 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 Zamboangeño.)

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 Acehese, 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 *mě-* 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 individuating 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 structures, 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 internal 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 borrowed 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 sesquisyllabic 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 evidence 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 is anomalous 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 admittedly 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 are 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 have 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 nonetheless.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 Ronglai 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 Rglai	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 acquire infixation from Malay, because Malay no longer had it to give. In this instance Acehnese had retained 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. *Язык Cham: ustnye govory vostochnogo dialekta*. St Petersburg: Institute for Oriental Studies.
- Aymonier, Etienne. 1889. 'Grammaire de la langue chame.' *Excursions et reconnaissances* XIV 31: 1-92.
- Aymonier, Etienne, and Antoine Cabaton. 1906. *Dictionnaire-cham-français*. Paris: Leroux. (Publications de l'Ecole française de l'Extrême-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 Pazez 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-East 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. *Zamboangueno 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-Hjelmsslev 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'Extrême-Orient LXIII.
- Larish, Michael D. 2005. 'Moken-Moklen.' *The Austronesian Languages of Asia and Madagascar*, edited by Alexander Adelaar and Nikolaus Himmelman, 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, Ilija. 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. Canberra: 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 *akey* '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 further 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 replaceive 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.