From Atonal to Tonal in Utsat (A Chamic Language of Hainan)
Author(s): Graham Thurgood
Proceedings of the Eighteenth Annual Meeting of the Berkeley Linguistics Society: Special Session on The Typology of Tone Languages (1992)
pp. 145-156

From atonal to tonal in Utsat (a Chamic language of Hainan)

Graham Thurgood[1]
California State University, Fresno

0. Introduction. Utsat, a Chamic Austronesian of Hainan, has undergone striking typological changes, going from polysyllabic and atonal to monosyllabic and tonal. These Utsat developments are of broad interest for their historical parallels and for their typological insights. They provide a historical model for the Tai-Kadai languages, a family now monosyllabic and tonal but believed to have been originally disyllabic and atonal. Utsat also provides a model for another more controversial change, the large Austronesian component Sagart (1991) has reported for Chinese. Here, as with Tai-Kadai, a model for the transition from disyllabic and atonal to monosyllabic and tonal is needed to account for the changes in the original Austronesian forms. However, the focus of this paper is not the historical but the typological insights which Utsat provides into tonogenesis. In the literature on Southeast Asian tonogenesis, the studies typically report individual stages in the tonogenetic process rather than report the whole process. There are examples of atonal languages becoming registral, registral languages becoming tonal, tonal languages undergoing tonal splits, and so on. It is usually only through extrapolation from various parts of the whole that a complete picture of tonogenesis can be complied. However, with the Chamic Utsat language of Hainan the complete transition from atonal to tonal is still recoverable: the closely related insular Austronesian (=PAN) languages are atonal; on the mainland of Southeast Asia under the influence of Austroasiatic (=Mon-Khmer) languages, the Chamic languages have become increasingly monosyllabic and, in some cases, registral; and, on Hainan island under the influence of various tonal languages (Solnit 1982), the Chamic dialect of Utsat has become monosyllabic and fully tonal. In the case of Utsat, these dual developments are recent enough and transparent enough for us to be able to outline the full transition from disyllabic and atonal to monosyllabic and tonal.

This analysis builds upon prior comparative analyses of Utsat and Utsat tonogenesis by expanding the bases for comparison and by expanding the database. Benedict (1984) and Haudricourt (1984) compared selected Utsat etyma directly with Old Cham, each providing an insightful set of suggestions about the origins of Utsat tones. Zheng (1986) compared Utsat directly with Rade, another Chamic language, but here the focus was not on tonogenesis. Most recently Ni Dabai (1988, 1990ab) compared Utsat directly with Indonesian, providing additional insights into Utsat tonogenesis. Unlike the two prior analyses of Utsat tonogenesis, this paper uses all of the available Utsat data and compares Utsat directly with proto-Chamic. Both Benedict (1984) and Ni Dabai (1988, 1990ab) used only part of the database. Unlike the earlier analyses which compare Utsat directly with specific individual languages, this analysis compares Utsat with reconstructed proto-Chamic, although an eye has also been kept on Malay and Indonesian cognates for insights they might provide.

The Utsat database used for this study consists of virtually all the forms found in the sources available to me: Ouyang and Zheng (1983), Zheng (1986),

and Ni Dabai (1988, 1990ab). This amounted to some 500 forms. Of these, some 370 or so forms either can be related to proto-Chamic reconstructions or can be identified on one basis or another as loans from non-Austronesian languages.

Although describing the same language, the database represents two separate transcriptions: the transcription of Ouyang and Zheng (1983) and Zheng (1986), on the one hand, and the transcription of Ni Dabai, on the other. The two approaches agree in virtually all respects except in their choice of tone numbers to designate the tones:

Ouyang and Zheng	Ni Dabai	
55	55	high-level tone
53	42	high-falling tone
33	33	mid-level tone
35	24	low-rising tone
11	11	low-level tone

These notational differences are of limited significance, as both transcriptions show three level tones, one falling tone, and one rising tone.

Something that is necessary to mention but I suspect ultimately of no significance is the tonal designations that show up sporadically. Somewhat regularly Ouyang and Zheng have a 32 tone that only occurs with stopped rhymes and appears to be a variant of the 33 tone; in Ni Dabai, these same forms occur as one of his basic five tones but with a final glottal stop (indicating glottalization?). A 45 tone occurs twice in Ouyang and Zheng, but this makes sense as a variant of their 35 tone. Finally, a 21 tone occurs in both Ouyang and Zheng (several times) and in Ni Dabai (once), but I suspect this too is a variant of one of the five basic tones.

The proto-Chamic reconstructions used for comparison are our own modifications of the reconstructions in Lee's 1966 dissertation. Although Lee's reconstructions are basically sound, much has been learned about Chamic languages in the last quarter of a century. Thus, our Chamic reconstructions ended up differing from Lee's in a number of ways: (a) The database is larger. In addition to the mainland sources found in Lee, our database includes additional mainland languages: Chru (Jrang et al. 1977), Haroi (Goschnick et al. 1976), Eastern Cham (Blood and Blood, 1977), and another description of Rade (J. A. Tharp and Y-Bham Buon-Ya (1980). And, of course, the database now includes Utsat on Hainan (Ouyang and Zheng 1983; Zheng 1986; Ni Dabai 1988, 1990ab), although this last addition has had minimal effects on the reconstruction of proto-Chamic itself. (b) Numerous loans from Austroasiatic, Indo-European, and occasionally elsewhere have been culled out (e.g., Headley 1976). (c) Our phonetic knowledge of Chamic has been expanded through valuable instrumental studies of the registral nature of Western Cham (Jerold Edmondson, and Kenneth Gregerson, to appear) and of the tonal system of Eastern Cham (Phú Văn Hẳn, Jerold Edmondson, and Kenneth Gregerson, to appear). (d) Insights have been gained from an increased understanding of the relationships of Chamic with the rest of the Austronesian languages (e.g. Blood 1962; Pittman 1959; Thomas 1963; Ni

Dabai1988, 1990ab; Benedict 1941, 1984) and from comparisons with PAN (various works of Blust and of others).[2] (e) Instrumental work has been done on Utsat itself, clarifying at least some of the problems inherent in the transcriptions used by Ouyang and Zheng and by Ni Dabai (Maddieson and Pang, 1991).[3]

1.0 Evolution of the Utsat tones. In at least some ways, the evolution of Utsat tones is strikingly straightforward. The first tonal split was between the voiceless finals and the voiced finals. The voiceless finals again split into syllables ending in *-h and the dead syllables (dead syllables end in voiceless final stops); the dead syllables then underwent one more tone split. The live syllables, that is, the remaining syllables, also split into two (live syllables end in nasal finals or vowel finals; for *-s see sec. 2.0 below). The live syllable forms containing an initial voiced stop or affricate produced a low-tone class; those syllables without a voiced stop produced a contrasting mid-tone class.

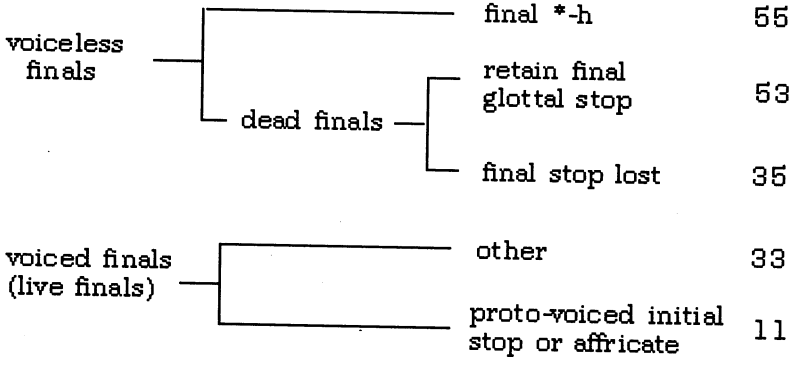

Chart 1.0: Utsat tonogenesis.

The remaining questions are mainly questions about individual etyma. Borrowings: In some cases, the non-Austronesian origin of particular forms has been suggested by Headley's work (1976) and other sources and confirmed by the existence of striking irregularities in proto-Chamic correspondence sets. Due to space limitations, words identified as borrowings are not dealt with in this shortened version of this paper. Some Austroasiatic and other borrowings were borrowed so early that they have perfectly regular reflexes within Chamic; as such forms behave regularly with respect to the Utsat developments, no effort have been made to cull these out. Anomalous forms: In any real data base, there are forms that occur with an unexpected initial, with an unanticipated vowel, or with an etymologically 'wrong' tone. This data base was no exception. However, the anomalous forms that occurred in this data base are just that—anomalous. None of these anomalies constitute problems for the analysis and thus most have, again for the lack of space, been simply left out of this shortened version of the paper.

1.1 From final *-h: (> tone 55). The regular Utsat reflex of Proto-Chamic final *-h is tone 55. This group contains 40 or so examples.

*kutŭt	tu?^{53}	tu?^{42}	kətù	kutŭk-f	ktŭt	'fart'
*?duai?	doi^{53}	---	ɗuaĭ	?duai?	ɗuĕ	'run; flee'
*hadip	---	thiu^{42}	hədiŭ	hadiu?	hdĭp	'live, alive'
*tiki?	---	ki?^{42}	təkì	tikī?	---	'few; little'

Tone 35. There are 31 examples of tone 35, a tone that correlates with the complete loss of final stops.

*mañak	ña^{35}	ña^{24}	---	mañã?	mña̋k	'oil'
*rok	zə^{35}	zə^{35}	rŏ	---	rô̋k	'grass'
*cə?	tsə^{35}	tsə^{24}	chŏ	cə?	cŭ -i	'mountain'
*hua?	hua^{35}	---	huă	hua?	huă	'eat, to'
*?asap	sau^{35}	sau^{35}	asă	?asa?	asăp	'smoke'
*pa:?	pa^{35}	pa^{24}	pà	pā?	pă	'four'
*phĭ?	phi^{35}	---	phì	phī^{n}? ^{n}?	phĭ	'bitter'
*laŋi:?	ŋi^{35}	ŋi^{24}	-ləngì	laŋĭ^{n}?	---	'sky'
*?ana:k	na^{35}	na^{24}	anà	?anā^{n}?	anak	'child'
*kle?	ke^{35}	---	klĕ	tle?	tlĕ	'steal'
*?ako?	ko^{35}	---	akŏ	---	kŏ; akŏ	'head'
*le?	le^{35}	---	lĕ	le?	lĕ	'fall'
*saki?	ki^{35}	ki^{24}	-səkì	sakī?	---	'sick, painful'
*?bu:k	?bu^{35}	?bu^{24}	'bù	?bū?	ɓŭk	'hair, head'
*mvbŭ	---	phu^{24}	məbŭ	---	---	'drunk'
*situk	---	tu^{24}	stŭ -i	situ?	atŭk	'cook'
*bruk	---	zo^{24}	brŭ	biŭk	brŭ	'rotten'
*k_so?	so^{35}	---	kəlso ?o	sō?	ksŏ	'lungs'
*jhi:t	si^{35}	---	sì	chi?	jhĭt	'sew'
*chia:p	siau^{35}	---	siau'	---	siap	'wing'

1.3 Live finals: (i.e., non-stopped syllables). The live finals are those syllables ending in nasals or vowels. The basic division is between a low-toned reflex with a 11 pitch and a mid-toned reflex with a 33 pitch. The 11 tone is the conditioned tone with all live syllables containing a proto-voiced stop becoming 11. The remaining forms constitute the tonal residue class. All forms not containing a proto-voiced stop have the 33 tone, or what I view as one of its apparently-conditioned variants. Note: Proto-Chamic final *-as rhymes lost their *-s, thus behaving like live finals (section 2.0).

This analysis describes what happens to the vast majority of the live syllables. There remains, however, a thus-far puzzling residue of cases in which the Utsat reflexes have not the expected 33 tone but instead either a 32 tonal variant (or, much less frequently, a marginally-attested 21 tonal variant). In one subset of forms, however, the analysis is straightforward: the Proto-Austronesian (=PAN) final *-ay becomes *-a:i?^{32} in Utsat, with an epenthetic glottal stop. Analysis of this problem is complicated by the difficulties inherent in the identification of loans interacting with the limited size of the database.

Proto-Chamic	Utsat (Zheng)	Utsat (Ni)	Chru (Jrang)	Roglai (Lee)	Rade (Tharp)	gloss
*?buh	?bu^{55}	---	'buh	?buh	ɓuh	'see'
*?juh	?iu^{55}	---	iuh	?juh	djuh	'firewood'
*dilah	la^{55}	la^{55}	dəlah	dilah	êlah	'tongue'
*mumih	mi^{55}	mi^{55}	---	mumĩh	m'mih	'sweet'
*mumah	ma^{55}	ma^{55}	bəmah	mumãh	m'mah	'chew'
*tanah	na^{55}	na^{55}	tənah	tanãh	---	'earth, soil'
*lanah	na^{55}	lə^{11} na^{55}	---	lanãh	ênah	'pus'
*?abih	phi^{55} phi^{55}	---	abih	?abih	---	'all'
*buh	phu^{55}	---	buh	buh	buh	'wear'
*pəbah	-pha^{55}	---	ia pəbah	---	bah	'spittle'
*pluh	piu^{55}	---	spluh	pluh	pluh	'ten'
*pasah	sa^{55}	---	pəsah	pasah	msah	'wet; damp'
*bijeh	se^{55}	---	pəjeh -i	bijeh	mjeh	'seed'
*tijuh	su^{55}	---	təjuh	tijuh	kjuh	'seven'
*drah	sia^{55}	---	drah	diah	---	'fast'
*lupih	pi^{55}	pi^{55}	ləpih	lupih	êpih	'thin'
*wah	va^{55}	va^{55}	wah	uah	wah	'fish, to'
*mahirah	za^{55}	za^{55}	məriah	maiiah	hrah	'red'
*?urah	zua^{55}	---	arah	?uiah	areh	'bedbug'
*picah	tsa^{55}	tsa^{55}	pəchah	---	mčah	'broken'
*pitih	ti^{55}	ti^{55}	---	pitih	mtih	'white'
*labuh	phu^{55}	---	ləbuh	labuh	êbuh	'fall down'
*boh	pho^{55}	pho^{55}	boh 'egg'	boh	boh	'fruit, clf.'
*lagah	khe^{55}	---	ləgah	lagah	êgah	'tired'

1.2 From dead finals: (i.e., stopped syllables). The regular reflexes of the dead syllables (that is, those syllables ending in *-p, *-t, *-k, *-ʔ) are tones 53 and 35.

Tone 53. The 53 tone occurs when a stopped final has a glottal stop reflex in Utsat, or in the case of variant forms, if one of the variants retains a glottal stop as a reflex. Several forms no longer have an attested variant with a glottal stop, however, on the basis of comparative evidence it appears that this lack is due either to the limited size of my Utsat database or the recent loss of the glottal stop. Nine forms fit this pattern.

*bruãʔ	phua^{53}	---	bruă	buãʔ	bruă	'work'
*do:k	thoʔ^{53}	thoʔ^{42}	dò	dõʔ	dôk	'live; sit'
*luba:t	pha^{53}	---	---	lubāʔ	êbat	'walk, go'
*pituk	tuʔ^{53}	tuʔ^{42}	pətŭ	pituʔ	mtŭk	'cough'

Tone 33: (a) with regular final reflexes.

*lapa	pa^{33}	pa^{33}	ləpa	lapa	êpa	'hungry'
*lupəi	pai^{33}	pai^{33}	ləpəi	lupəi	êpei	'dream'
*mata	ta^{33}	tiŋ^{33} ta^{33}	məta	---	---	'eye'
*ʔiku	---	ku^{33}	aku	ʔiku	aku	'tail'
*takuai	kuai^{33}	---	təkuai	takuai	k'kuê	'neck'
*ʔapui	pui^{33}	pui^{33}	apui	ʔapui	apui	'fire'
*ʔadhə̃i	thai^{33}	thai^{33}	thəi	thẽi	dhei	'forehead'
*thun	thun^{33}	---	thun	thut	thŭn	'year'
*thu	thu^{33}	---	thu	---	thu	'dry'
*pha	-pha^{33}	pha^{33}	pha	pha	pha	'leg, thigh'
*ʔasəu	a^{11} sau^{33}	a^{11} sau^{33}	asəu	ʔasəu	asâo	'dog'
*sa:ŋ	sa:ŋ^{33}	---	sàng	sāk	sang	'house'
*sa	sa^{33}	sa^{33}	sa	sa	sa	'one'
*hia	hia^{33}	---	hia	hia	hia	'cry; weep'
*hã	ha^{33}	ha^{33}	hã	hã	ih	'you; thou'
*kəu	kau^{33}	kau^{33}	kəu, kəmi	---	kâo	'I'
*ʔikan	ka:n^{33}	---	akàn	ʔikāt	akan	'fish'
*ʔia	ʔia^{33}	ʔia^{33}	ia	ʔia	êa	'water'
*ʔiəu	ʔiə^{33}	---	---	ʔiəu	iêô	'call; cry'
*ʔbəu	---	ʔbə^{33}	'bəu	ʔbəu	ɓâo	'stench'
*ʔdau	dau^{33}	---	ɗau	---	---	'all'
*kami	mi^{33}	---	---	kamin	hmei -i	'we (ex.)'
*taŋa:n	ŋa:n^{33}	ŋa:n^{33}	təngàn	təŋan	kngan	'hand'
*tariŋa	ŋa^{33}	lə^{11} ŋa^{33}	tənia	iiŋiã	knga	'ear'
*ʔuni	ni^{33}	ni^{33}	ni	ʔunĩ	anei	'this'
*ʔaŋin	ŋin^{33}	ŋin^{33}	angin	ʔaŋin	angĭn	'wind, the'
*trun	tsun^{33}	---	trun	tiut	trŭn	'descend'
*paləi	piai^{33}	piai^{33}	pləi -v	pələi	----	'village'
*pila	pia^{33}	pia^{33}	pəla	pila	pla	'plant, to'
*haləu	lau^{33}	---	hələu	haləu	hlâo	'pestle'
*talәi	lai^{33}	lai^{33}	tələi	talәi	klei	'rope; string'
*tulak	---	la^{33}	tələ̆	---	klŏ, klŭ	'push, to'
*ʔula	a^{11} la^{33}	-la^{33}	ala	ʔula	ala	'snake'
*kləu	kiə^{33}	kiə^{33}	kləu	tləu	tlâo	'three'
*klau	kiau^{33}	---	klau	tlau	tlao	'laugh'
*kra	kia^{33}	---	kra	kia	kra	'monkey'
*kayəu	zau^{33}	---	kəyəu	---	kəyâo	'tree; wood'
*riya	za^{33}	za^{33}	lia	iiia	êya	'ginger'
*ruəi	zoi^{33}	---	---	iuəi	rui	'crawl, to'
*hurəi	zai^{33}	zai^{33}	hərəi	huiəi	hruê	'day; sun'
*mua	mua^{33}	---	mua	muã	muôr	'termite'

*miau	miau^{33}	---	miau	miãu	miêô	'cat'
*la	la^{33}	---	---	la	la	'spleen'
*lu:n	luan^{33}	luan^{33}	luən	luət	lun	'swallow, to'

(b) final nasal > Utsat stop (unexplained). The tonal reflex in these forms is 33, as expected. The problem is the denasalization of the final. In Utsat, the denasalization does not occur with tone 11 forms, only the 33 forms. Further, even here only some of the final proto-nasals have denasalized; many have not. The question is why just these forms?

*tia:n	tet^{33}	---	tiàn	---	tian	'stomach'
*proŋ	pyok^{33}	---	prong	piok	prŏng	'big'
*troŋ	tsyok^{33}	---	trong	tiok	trong	'eggplant'
*dhoŋ	thok^{33}	---	thong	thok	dhŏng	'knife'
*ʔuraŋ	za:k^{33}	za:n?^{33}	aràng	ʔuiāk	arăng	'person'
*ʔariaŋ	liak^{33}-i	---	rə-iang	ʔaiak	ariêng	'crab'

Tone 32 (PAN *-ay > -a:?^{32} (and so on)). Tone 32 is variant of tone 33 that occurs with secondarily-derived final stops. Regular addition of a final glottal stop occurs with PAN *-ay syllables. And, although there is considerable unexplained and unexpected tonal variation, the tone assignment essentially matches that found with Utsat live syllables: 33 toned forms become 32 tones with the addition of the glottal final, while 11 toned forms remain 11 (only one example!).

*takay	ka:i?^{32}	kai^{11}	təkai	takai	---	'foot'
*məray	za:i?^{32}	za:i?^{33}	mərai	---	---	'come'
*haway	va:i?^{32}	va:i?^{11}	həwai	hauai	hwiê	'rattan'
*matay	ta:i?^{42}-t	tai?^{42}-t	mətai	---	mdjiê	'die'
*hatay	ta:i?^{32}	tai^{11}	hətai	hatai	atiê -i	'liver'
*paday	tha:i?^{32}; tha:i?$^{21/11}$	thai?^{11}; thai^{11}	pədai	padai	mdiê	'paddy'
*glay	khiai?^{11}	---	glai	---	dliê	'jungle'

*naw na:u?^{32} na:u?^{33} nau nãu nao 'go; leave'
The final glottal stop with this form may reflect a PAN final *-aw, or it may simply be irregular. Without more data, it is not possible to tell.

Tone 11: (from proto-voiced obstruents). Tone 11 forms all show a proto-voiced stop (or affricate) either in the pre-syllable or in the main syllable. However, it cannot simply be voicing that causes these forms to have the low 11 tone, since voiced resonants are in the 33 tone class, not in the 11 tone class. Further, in some Utsat forms that descend from disyllabic forms, the main syllable originally had and still has a voiceless onset; the crucial voiced stops were in the now-lost pre-syllable. Thus, the voiced stop onset of the pre-syllable managed to affect the tone class of the main syllable, but without voicing the main syllables'

voiceless initial. This strongly suggests that voiced stops produced a breathy phonation that lead to the 11 tonal reflex. When the voiced consonant occurred as the first consonant of the pre-syllable, the breathiness spread to the main syllable before the pre-syllable was lost. Whatever the explanation, it is clear that tone 11 is associated with proto-voiced stops and affricates.

*dua	thua^{11}	thua^{11}	dua	dua	dua	'two'
*dəŋ	thaŋ^{11}	---	dəng	dək	dŏng	'stand; stop'
*glay	khiai?^{11}	---	glai	---	dliê	'jungle'
*habəu	phə^{11}	phə^{11}	həbəu	habəu	hbâo	'ashes'
*təbəu -v	phə^{11}	phə^{11}	tərbəu	tubəu	kbâo	'sugarcane'
*hubəi	phai^{11}	phei^{11}-f	həbəi	habuəi	hbei	'yam'
*babui	---	phui^{11}	pəbui -i	babui	---	'pig, wild'
*?ada	tha^{11}	---	---	---	---	'have; be'
*?ada	na^{11} tha^{11}	---	ada	---	---	'duck'
*dada	tha^{11}	tha^{11}	---	dada	ɗahda-i	'chest'
*?iduŋ	thuŋ^{11}	thuŋ^{11}	adung	?idŭk	adŭŋ	'nose'
*sidɔm	-than^{11}	-than^{11}	adəm	sudəp	hdăm	'ant'
*padam	---	than^{11}	---	padap	---	'extinguish'
*digəi	khai^{11}	-khai^{33}-t	təgəi	digəi	êgei	'tooth'
*hujaːn	saːn^{11}	saːn^{11}	hə jàn	hujāt	hjan	'rain'
*drəi	se^{11}	---	---	diəi	drei	'body; self'
*bara	phia^{11}	phia^{11}	bra	baia	mra	'shoulder'
*braːs	phia^{11}	phia^{11}	bràh	bia	braih	'paddy'
*bahrəu	phiə^{11}	phiə^{11}	bərhəu	bahiəu	mrâo	'new'
*bilaːn	phian^{11}	-phiaŋ^{11}	blàn -v	bilāt	mlan	'moon'
*biləu	phiə^{11}	phiə^{11}	bləu -v	biləu	mlâo	'body hair'
*jrãu	siau^{11}	---	jràːũ	jiãu	drao	'medicine'
*jalaːn	laːn^{11}	laːn^{11}	jəlàn	jalāt	êlan	'road; path'
*buŋa	ŋa^{11}	---	bən ga	---	mn ga	'flower'
*bitəi	tai^{11}	---	pətəi	pitəi	mtei	'banana'
*batəu	tau^{11}	tau^{11}	pətəu	patəu	tâo	'stone'
*bato	to^{11}	to^{11}	pəto -f	pato	mtô	'teach'
*bisəi	sai^{11}	sai^{11}	pəsəi	pisəi	msei	'iron'
*dupa	pa^{11}	---	təpa	tupa	êpa	'armspan'
*dihləu	lau^{11}	---	dərhəi	dihləu	êlâo	'first (go); formerly'

2.0 Proto-Chamic *-h versus *-s and subgrouping. In most cases, the distinction between the PAN finals *-s and *-h has been lost in Chamic with the two finals merging into proto-Chamic *-h. However, although the *-s is nowhere retained as an -s, evidence of its earlier presence is still found after the

Chamic vowels *u and *a. As a consequence, there is a distinction in Chamic between the vocalic reflexes of *-uh and *-us and the vocalic reflexes of *-ah and *-as.

The final *-s of the final *-us merged totally with proto-Chamic *-h, but not before causing the *-u- to diphthongize in Jorai and Rade. Thus, in these two languages, the old *-us results in a diphthongal -uih reflex, while the old *-uh results in a monophthongal -uh reflex. In a partially but not totally parallel way, the final *-s of the final *-as merged totally with proto-Chamic *-h, but not before causing the *-a- to diphthongize in Jorai and Rade. Thus, in these two languages, the old *-as results in a diphthongal -aih reflex, while the old *-ah results in a monophthongal -ah reflex.

gloss	proto-Chamic	Jorai (Lee)	Rade (Jrang)	Roglai (Lee)	Utsat (Zheng)
	*-as	**-aih**	**-aih**	**-a**	**-a$^{33/11}$**
'husked rice'	*braːs	braih	braih	bia	phia^{11}
'escape'	*klaːs	klaih	tlaih	tla	---
'escape, c.'	*paklaːs	pəklaih	mtlaih	patla	---
'fish scales'	*kakăs	---	k'kaih	kaka	ka^{33}
'far'	*ʔatas	ʔataih	taih	ʔata	ta^{33}
'scratch'	*pras	---	praih	---	---
	*-ah	**-ah**	**-ah**	**-ah**	**-a^{55}**
'earth, soil'	*tanah	tənah	---	tanãh	na^{55}
'shoot bow'	*panah	pənah	mnah	panãh	na^{55}
'pus'	*lanah	rənah	ênah	lanãh	na^{55}
'middle'	*krãh	krah	krah	kiãh	kia^{55}
'chew'	*mumãh	məmah	m'mah	mumãh	ma^{55}
'mouth'	*mubah	məbah	mbah	mubah -i	pha^{55}
'leech'	*ritah	---	êtah	iitah	a^{11} ta^{55}
'blood'	*darah	drah	êrah	daiah	sia^{55}
'wet; damp'	*pasah	pəsah	msah	pasah	sa^{55}
'bedbug'	*ʔurah	ʔarăh	areh	ʔuiah	zua^{55}

The parallelism between the reflexes of *-us and *-as is broken in the Roglai and, as we shall see, in the Utsat reflexes. In Roglai, the final *-s in the rhyme *-as simply disappears, but **only** in the rhyme *-as. In other words, in Roglai, the Chamic final *-as simply lost its final *-s and became -a, an open syllable. In Utsat, the tonal reflexes indicate that it is also just those Utsat forms reconstructed with *-as that behave tonally as if *-as had lost its final *-s and become an open syllable. This loss of the final *-s in *-as rhymes, attested in Roglai, accounts for the tonal behavior of *-as finals in Utsat. Note: This shared, idiosyncratic innovation in the treatment of *-as finals at least suggests that it might

have been Roglai speakers who originally migrated to Hainan and became the Utsat.

3.0 Disyllabic to monosyllabic. The Utsat monosyllabic forms resulted from the collapse of what were in most cases disyllabic Austronesian forms. Under the influence of the mainland Austroasiatic forms, stress came to rest on the final syllable, leaving the pre-syllable unstressed and subsequently, this unstressed vowel dropped. In most cases, the whole pre-syllable was eventually completely lost without any trace. However, sometimes a voiced obstruent in a pre-syllable left a tonal reflex before it dropped (see the discussion above of tone 11). In another subset of forms, the loss of the unstressed vowel in the pre-syllable resulted in a cluster consisting of a stop followed by -r- or -l-; in these cases, the -r- or -l- became Utsat -i-, allowing the initial stop of the pre-syllable to be reinterpreted as the onset in the newly-formed monosyllable (see examples above).

4.0 Summary. Although numerous details still remain to be worked out, the outlines of Utsat tonogenesis are clear. It is largely based on rhymes and finals, not initials. The initial division was into two groups on the basis of whether the proto-rhyme ended with a voiced or a voiceless final. Each of these then underwent further splits. The voiceless finals split into two depending upon whether the syllable ended in *-h or ended with a final stop (=dead syllables). Those proto-rhymes ending in *-h invariably went to Utsat tone 55; the dead syllables again split into two, with those rhymes still retaining a final glottal stop becoming tone 53, while those rhymes having completely lost their final stop became tone 35. The voiced finals (the live syllables) also split into two basic variants. Those syllables with a proto-voiced stop or affricate became tone 11. Most of the remaining live syllables became 33. The exceptions involve secondarily-derived final stops: PAN *-ay develops an epenthetic glottal stop in Utsat, with tone 33 forms becoming tone 32. Only this last tonal development can be interpreted as initial-based, and even here it may be more the phonation type than the initial itself (see earlier discussion).

Beyond their interest as a model of tonogenesis, the Utsat developments have historical implications for the history of the Tai-Kadai languages and, quite possibly, for understanding the apparently Austronesian component in Chinese.

Endnotes

[1]I shall be astonished if all my errors should prove minor and grateful to readers for their corrections. I wish to thank Bob Blust for a marvelous seminar he gave on Austronesian, for his feedback on this paper, and for his generous help finding the materials I needed for this paper; unless he wishes, he need take no responsibility for the positions in this paper. I also wish to thank Joel Nevis, Jerry Edmondson, Ian Maddieson, Gérard Diffloth, Jim Matisoff, Paul Benedict, Mark Durie, and Eric Oey for their useful feedback.

Symbols used: forms prefaced by a single asterisk (*) are proto-forms, forms followed by -i have an irregular initial, by -f have an irregular final, by -v have an irregular vowel, and by -t have an irregular tone. As the historical phonology is better understood, at least some of these apparent irregularities should disappear, while others will remain puzzles.

The data used in this paper has been normalized in only two ways: where the original sources used ɔ' for a ə, a ə has been substituted, and where the original sources used a ɯ' for a -i̱, a -i̱ has been substituted.

[2]The treatment of nasalized vowels in Chamic remains unsolved, but is happily irrelevant to the concerns of this paper.

[3]I first became aware of the Maddieson and Pang paper only after I sent Maddieson an earlier draft of this paper. Thus, its insights are not fully incorporated. Maddieson and I intend to collaborate on a joint effort on Utsat in the near future.

Bibliography

Benedict, Paul K. 1941. A Cham colony on the island of Hainan. *Harvard Journal of Asiatic Studies* 4:129-34.

Benedict, Paul K. 1984. "Austro-Tai parallel: A tonal Cham colony on Hainan." *Computational Analyses of Asian & African Languages* 22:83-86.

Blood, Doris W. 1962. Reflexes of Proto-Malayo-Polynesian in Cham. *Anthropological Linguistics* 4.9:11-20.

Blust, R. 1980. Austronesian etymologies. *Oceanic Linguistics* 19.1-2:1-189.

Blust, R. 1980. More on the origins of glottalic consonants. *Lingua* 52:125-156.

Blust, R. 1981. The reconstruction of Proto-Malayo-Javanic: an appreciation. *Bijdragen tot de taal-, land-, en volkenkunde* 137.4:456-469.

Blust, R. 1983-4. Austronesian etymologies—II. *Oceanic Linguistics* XXII-XXIII.1 & 2: 29-150.

Blust, R. 1986. Austronesian etymologies—III. *Oceanic Linguistics* XXV.1 & 2: 1-123.

Blust, R. 1989. Austronesian etymologies-IV. *Oceanic Linguistics* XXVIII.2:111-180.

Dyen, Isidore. 1971. The Chamic languages. In Thomas A. Sebeok (ed.), *Current Trends in Linguistics*. Mouton: The Hague: Volume 8.1:200-210.

Edmondson, Jerold A. and Kenneth J. Gregerson. To appear. Western Cham as a register language.

Goschnick, H. and Alice Tegenfeldt-Mundhenk. 1976. Haroi dictionary. SIL. Huntington Beach, Calif. 325pp. ms.

Haudricourt, A.-G. 1984. Tones of some languages in Hainan. *Minzu Yuwen* 4:17-26.

Headley, Robert K. 1976. Some sources of Chamic vocabulary. *Austroasiatic Studies, Part I.* Oceanic Linguistics Special Publication No. 13:453-476.

Jrang, Ja Kuang, Ja Wi, Ja Dai, Ja Ngai, and Eugene Fuller. 1977. *Chru Vocabulary*. Summer Institute of Linguistics. Manila.

Lee, Ernest Wilson. 1966. Proto-Chamic phonologic word and vocabulary. Ph.D. Indiana University. 67-3690.

Lee, Ernest Wilson. 1974. Southeast Asian areal features in Austronesian strata of the Chamic languages. *Oceanic Linguistics* XIII.1-2:643-668.

Lee, Ernest Wilson. 1977. Devoicing, aspiration, and vowel split in Haroi: evidence for register (contrastive tongue-root position). *Papers in South East Asian Linguistics No. 4: Chamic Studies.* Pacific Linguistics Series A, No. 48:87-104.

Maddieson, Ian and Keng-Fong Pang. 1991. Tone in Utsat. Sixth International Conference on Austronesian Linguistics. Honolulu.

Ni Dabai. 1988. The genealogical affiliation of the language of the Hui people in Sanya Hainan. *Minzu Yuwen* 2:18-25.

Ni Dabai. 1990a. The origins of the tones of the Kam-Tai languages. ms.

Ni Dabai. 1990b. The Sanya (=Utsat) language of Hainan island: A living specimen of a linguistic typological shift. ms.

Oey, Eric M. n.d. The linguistic position of the Chamic languages. [preliminary draft]. m.s.

Ouyang Yueya and Zheng Yiqing. 1983. The Huihui speech (=Utsat) of the Hui nationality in Yaxian Hainan. *Minzu Yuwen* 1:30-40.

Phú Văn Hẳn, Jerold Edmondson, and Kenneth Gregerson. To appear. Eastern Cham as a tone language.

Pittman, Richard S. 1959. Jarai as a member of the Malayo-Polynesian family of languages. *Asian Culture* 1.4:59-67.

Sagart, Laurent. 1991. Chinese and Austronesian: Evidence for a genetic relationship. Sixth International Conference on Austronesian Linguistics. Honolulu.

Shorto, H. L. 1975. Achinese and Mainland Austronesian. *Bulletin of the School of Oriental and African Studies* 38:81-102. University of London.

Solnit, David. 1982. Linguistic contact in Ancient South China: the case of Hainan Chinese, Be, and Vietnamese. *BLS* 8:219-230.

Tharp, J. A. and Y-Bham Buoy-Ya. 1980. *A Rhade-English dictionary with English-Rhade finderlist.* Pacific Linguistics. Series C - No. 58.

Thomas, David D., Ernest Wilson Lee, and Dang Liem Nguyen. 1977. *Papers in South East Asian Linguistics. No. 4. Chamic Studies.* Canberra: Pacific Linguistics. Series A, No. 48.

Thomas, Dorothy M. 1963. Proto-Malayo-Polynesian reflexes in Rade, Jarai, and Chru. *Studies in Linguistics* 17:59-75.

Thurgood, Graham and Elzbieta Thurgood. In preparation. Proto-Chamic.

Zheng Yiqing. 1986. A further discussion of the position of Huihui speech and its genetic relationship. *Minzu Yuwen* 6:37-43.