TONE CORRESPONDENCES AND TONOGENESIS IN VIETIC

Koichi Honda
School of Language Studies, ANU
Koichi.Honda@anu.edu.au

This paper[1] aims to examine tone correspondences in Vietic languages, namely Vietnamese, Ruc and Arem, and attest Haudricourt's tonogenesis hypothesis.

1. Haudricourt's Tonogenesis hypothesis of Vietnamese

Haudricourt's Tonogenesis hypothesis of Vietnamese (1954) is well known. It not only put an end to the controversy on what language family Vietnamese belongs to, but also proposed a model of the development of tones in a language. Even after fifty years since its publication, it is generally accepted and still very influential among linguists. The hypothesis is usually explained as follows.

(1)	1	2	3	4	1	2	3	4	1	2	3	4
	pa	=	pa	=	pa	=	pá	>	**p**a**h**	>	pà	>
	ba	=	ba	>	pà	>	bà		**b**a**h**	>	bà	>

	pá	=	pá	>	bá		**p**a**h**	>	pà	>	pả	> bả
	ba**X**	>	bá	>	pạ	>	bạ		**b**a**h**	>	bà	> pã > bã

In stage 1, Vietnamese started as a toneless language. In stage 2, three pitch contours were formed depending on the coda types: level pitch from null coda, rising pitch from stop coda including glottal stop, and falling pitch from voiceless fricative coda. In stage 3, they split into higher and lower series according to [+/- voice] feature of onsets, and the number of tones doubled causing onsets of lower series devoiced. In stage 4, voiceless onsets were voiced again, though this stage is not relevant to the number of tones. This model may be called a **consonantally-based model** of tonogenesis.

However, simple questions arise. Tone is a phonological category, and it is realized and perceived by phonetic pitch. This pitch is encoded acoustically by fundamental frequency (F0), and this F0 cannot exist without voicing. If I accept the consonantally-based model, such syllables as [ai] or [au] without consonants cannot have tones. And specifically, how can a change from [bá] to [pạ] between stage 2 and 3 explained? This change from voiced high rising pitch to voiceless low pitch does not seem to be plausible.

My question leads to a certain confidence that it is not consonants themselves, but interactions between various aspects of phonation that determines the pitch. This kind of idea is proposed by Thurgood (2001) in the name **laryngeally-based model** of tonogenesis.

[1] This paper is the main result from my Master of Linguistics sub-thesis at the ANU (2004). I'm indebted to my supervisor Dr. Paul Sidwell for his enormous efforts, and lots of comments on it by Dr. Mark Alves and Dr. Phil Rose. I'm also indebted to Dr. Graham Thurgood for a new concept specifically applied in this paper.

Paul Sidwell, ed. *SEALSXV: papers from the 15th meeting of the Southeast Asian Linguistics Society*. Canberra, Pacific Linguistic, 2005, pp.177-191.

2. Ruc and Arem – Vietic languages

I'm going to focus on the languages of Arem, Ruc and Vietnamese. The term "Vietic" is used to be known as "Viet-Muong". Some scholars such as Ferlus prefer to use the latter name.

Population of Ruc and Arem is extremely small. According to the census in 1985, it is *Arem* 76, *Rục* 125, *Maliêng* 715, *Mày* 715, *Sách* 625 (Phong et al. 1988). Another source shows *Chứt* 2400 (VNA 1996). The name *Chứt* is used by the government to represent these extremely minor ethnic people settled in the North-Central area of Vietnam.

The survey of Ruc language started in 1986 by Russian and Vietnamese linguists (Loi 1993). However, the report is not available at hand. Available materials at hand are the following three.

(2) **- Nguyễn Phú Phong, Trần Trí Dõi and M.Ferlus (1988) on Ruc**
- Nguyễn Văn Lợi (1993) on Ruc
- Michel Ferlus (1997, original version 1991) on Arem and Ruc

My study is based on these sources and not based on the field work of my own.

In the SEALS Conference, Mark Alves and Michel Ferlus have contributed papers on Ruc language in 1997, 1999 and 2001. These papers are also cited.

3. Lexicostatistic Data

The following lexicostatistic data show lexical distance between the languages of Vietic and Katuic.

(3) Compared Vietic and Katuic lexica (Samarina, I.V. 1989, cited in Loi 1993)

							Bru
						Taoih	63.0
					Pakoh	71.0	72.0
				Arem	40.0	38.0	41.0
			Ruc	66.5	37.5	33.0	37.5
		Poong	53.0	51.5	35.5	33.0	37.5
	Muong	50.5	50.0	56.5	27.0	26.5	31.5
Viet	74.0	51.5	53.0	45.0	26.0	24.0	29.5
←		Vietic			◄	Katuic	→

As is shown, Vietnamese shares nearly 50% common lexica with Ruc and Arem, while the rates are less than 30% with Katuic languages.

4. Tone Description

Tone description is crucially important but is a bit complicated because they are not shown by pitch alone but voice quality differences are involved in Vietic languages.

4-1. Standard Northern Vietnamese: 6 tones

Standard Northern Vietnamese is well known to have six tones, in Vietnamese name *ngang*, *sắc*, *hỏi*, *huyền*, *nặng*, *ngã*.

(4)

(plain)	(abrupt)	(contour)
1. *ngang* (1A)	2. *sắc* (1B)	3. *hỏi* (1C)
4. *huyền* (2A)	5. *nặng* (2B)	6. *ngã* (2C)

They are grouped into historical higher and lower series, and each two tones make three pairs: plain tones, abrupt tones and contour tones. Besides differences in pitch, *nặng* tone is accompanied by glottal stop, and *ngã* tone has creaky voice. Tone names in brackets, which I call **comparative tone numbers**, will be used to show tone correspondences more explicitly; the numbers (1 and 2) indicate higher and lower series, and the alphabets (A, B and C) indicate three contour types.

4.2 Ruc: 4 tones

Ruc is reported to have four tones.

(5)
1. High level pitch
2. Low level pitch (with breathy voice) or falling pitch (with breathy voice)
3. Rising pitch (with optional glottal stop)
4. Concave pitch or falling pitch (with optional glottal stop)

According to Loi (1993), tone 1 has high level pitch, and tone 2 has two variations in free variation, accompanied by breathy phonation. Tone 3, rising pitch can accompany glottal stop optionally. Tone 4 is reported to have two free variations, concave pitch and low falling pitch, and the falling pitch is accompanied by glottal stop optionally. Phong et al. (1988) describes six tones for Ruc, which counts tones 3 and 4 with and without glottal stop as an individual one.

4.3 Arem: Toneless

Arem is reported to be toneless, but to have breathy phonation, and post-vocalic laryngeal constriction (Ferlus 1997, 2001).

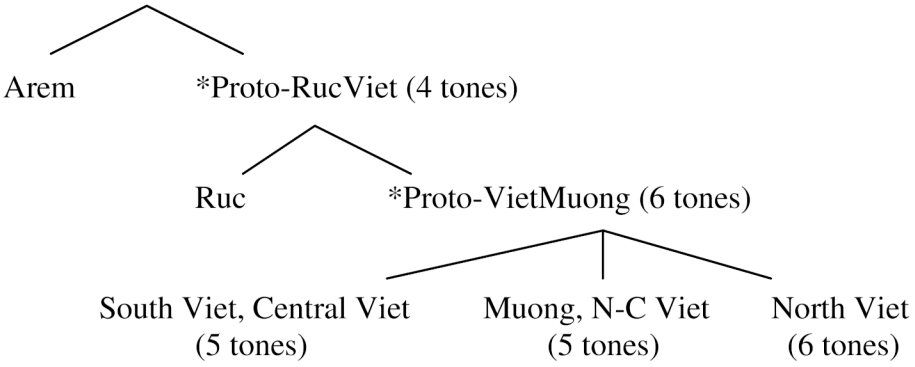
4.4 Research Questions

Here are some questions. According to Haudricourt, the tones developed from 0 > 3 > 6. How are they different from Ruc's 4-tone system? What are the tone correspondences among Arem, Ruc and Viet?

5. Tone Developments – A tentative tree diagram

I propose a tentative tree diagram of Vietic from a viewpoint of the development of tones.

(6) *Proto-Vietic (toneless, contrastive phonation types)



I assume the tones developed from 0 > 4 > 6, in the same way as Arem, Ruc and Viet. 5-tone systems of Southern, Central, North-Central varieties of Vietnamese as well as Muong are assumed to be the result of merger of two tones of the Proto-VietMuong, though categorisation of the two tones are not the same.

6. Word list

Word list is attached as an Appendix to this paper. I have found 109 cognate sets from the three materials and made them in a comparative list. On top from the left, after numbers and English glossary, there are transcriptions of Arem in Ferlus (1997), Ruc in Ferlus (1997), Ruc in Loi (1993) and Ruc in Phong et al. (1988) as in the original source. Next three columns are Vietnamese orthography and comparative tone numbers. In the last column, Ferlus' 1997 reconstruction is shown as a reference. On left from the top, the 109 lexica are categorised into five groups by their coda types: (1) null coda, (2) fricative coda, (3) liquid coda, (4) stop codas, and (5) nasal codas. The colour further categorises the lexica within the coda types: coloured cells show basic regular correspondences, and colourless cells show irregular correspondences. From next section, I will demonstrate several typical examples of regular and irregular correspondences.

7. Regular Tone Correspondences (79/109 cases)

I will explain the regular tone correspondences according to the six tones of Standard Northern Vietnamese. This regular correspondence applies to 79 cases out of 109 lexica. The number after Ruc and Viet IPA transcription refers to its tone.

7.1 Ngang tone (1A)

(7)	No.	Gloss.	Arem	Ruc	Viet	*P-V(Ferlus)
	1	'three'	pæː	paː 1	baː 1A	paː
	47	'fly (verb)'	pal	pəl 1	băj 1A	pər
	78	'cooked rice'	kəːm	kəːm 1	kəːm 1A	kəːm

Vietnamese tone 1A corresponds to Ruc tone 1 in null coda, liquid coda and nasal coda syllables. Liquid coda in Arem and Ruc corresponds to the glide coda [j] in Vietnamese.

7.2 Sắc tone (1B) with null coda and sonorant codas

(8)	No.	Gloss.	Arem	Ruc	Viet	*P-V(Ferlus)
	20	'fish'	akæːʔ	akaː 3	kaː 1B	ʔa-kaːʔ
	55	'roof'	(n.a.)	baːl 3	măj 1B	ɓaːlʔ
	92	'four'	puənʔ	poːn 3	bon 1B	poːnʔ

Vietnamese tone 1B corresponds to Ruc tone 3 in the above mentioned three coda types, and final *glottal stop* in Arem. Ferlus (2001) indicates that this syllable-end *glottal stop* of Arem is used to represent "the constricted rimes in voiced ending".

7.3 Huyền tone (2A)

(9)	No.	Gloss.	Arem	Ruc	Viet	*P-V(Ferlus)
	10	'fly (insect)	uruə̀j	mərɔːj 2	ruăj 2A	m-rɔːj
	53	'long'	(n.a.)	jəal 2	zaːj 2A	jaːr

Vietnamese tone 2A corresponds to Ruc tone 2, and Arem transcription *grave accent* above vowels in most cases. According to Ferlus 2001, this diacritic indicates "breathy voice".

7.4 Nặng tone (2B) with null coda and sonorant codas

(10)	No.	Gloss.	Arem	Ruc	Viet	*P-V(Ferlus)
	57	'wake up'	jɨ̀lʔ	jɨl 4	zăj 2B	jɨrʔ
	108	'heavy'	naŋʔ (?)	naŋ 4	năŋ 2B	naŋʔ

Vietnamese tone 2B corresponds to Ruc tone 4, which corresponds to Arem combination of 'breathy voice' and 'the constricted rimes in voiced ending' in most cases.

7.5 Hỏi tone (1C) and ngã tone (2C)

(11)	No.	Gloss.	Arem	Ruc	Viet	*P-V(Ferlus)
	38	'grass'	(n.a.)	kɔh 1	kɔː 1C	kɔh
	44	'nose'	mùh	muːr^{h} 2	muːj 2C	muːs

Next sets show Vietnamese tones 1C and 2C, which correspond to Ruc tones 1 and 2 respectively whose rhymes have voiceless fricative [h], and which correspond to Arem voiceless fricative coda, too. In Arem and Ruc, final –h is still segmental and not tonal yet. In this point, the evidence is different from Haudricourt's hypothesis.

7.6 Sắc tone (1B) and nặng tone (2B) with stop codas

(12)	No.	Gloss.	Arem	Ruc	Viet	*P-V (Ferlus)
	63	'sing'	ahæːt	haːt 1 or 3	haːt 1B	haːt
	71	'wear'	mɛ̀k	măak 2 or 4	măk 2B	mak

Next sets show Vietnamese tones 1B and 2B with stop codas, which correspond to Ruc tone 1 and 2, or 3 and 4 respectively. What is interesting is the difference between Ferlus and Loi. While Ferlus describes 1 and 2, high level and low level, Loi describes 3 and 4, high rising and low falling. Since transcriptions of the two authors are coherent, I understand it is due to the perception difference.

7.7 Proposed Tonogenesis model

To summarise the above regular tone correspondences, proposed Tonogenesis model is shown as below.

(13)	*Proto-Vietic		Arem	Ruc	Viet	
	[modal]	a [-constricted]	a	a 1A	a 1A	*ngang*
			al	al 1A	aj 1A	
			an	an 1A	an 1A	
		a' [+constricted]	a'	a 1B	a 1B	*sắc 1*
			al'	al 1B	aj 1B	
			an'	an 1B	an 1B	
			at	at 1A (or 1B)	at 1B	*sắc 2*
			ah	ah 1A	a 1C	*hỏi*
	[breathy]	a̤ [-constricted]	a̤	a 2A	a 2A	*huyền*
			a̤l	al 2A	aj 2A	
			a̤n	an 2A	an 2A	
		a̤' [+constricted]	a̤'	a 2B	a 2B	*nặng 1*
			a̤l'	al 2B	aj 2B	
			a̤n'	an 2B	an 2B	
			a̤t	at 2A (or 2B)	at 2B	*nặng 2*
			a̤h	ah 2A	a 2C	*ngã*

Several remarks should be noted here. Comparative tone numbers are applied to Ruc, where Ruc tones 1, 2, 3 and 4 are shown as 1A, 2A, 1B and 2B respectively. The letter "a" represents all the vowels, the letter "n" represents nasal codas in /m, n, ŋ/, the letter "t" represents stop codas /p, t, k/. The *apostrophe* shows post-vocalic laryngeal constriction.

As we see above, Proto-Vietic is assumed to have contrasts in voice quality; **modal vs. breathy**, as Arem. Another laryngeal feature is the post-vocalic laryngeal constriction. This glottal constriction is what Diffloth (1989) called 'creaky voice'. These two laryngeal configurations seem to be closely related to the origin of tones.

To express in binary features, breathiness and laryngeal constriction may be **marked features**. Following figures may be an evidence for this.

(14) 1A (*ngang*) 1136 1B (*sắc 1*) 891
2A (*huyền*) 951 2B (*nặng 1*) 636

These figures show number of syllables by tones, which I counted from a modern Vietnamese-Chinese dictionary (Honda 2004), which lists 5795 syllables as a whole. According to this, 1B and 2B are less than 1A and 2A respectively due to the marked laryngeal constriction, and 2A and 2B are less than 1A and 1B respectively due to the marked breathiness.

The second important point is the presence of fricative coda /–h/ in Arem and Ruc, which is assigned to tone 1A in Ruc. I see chronological order of the emergence of tone 1C was much later than 1B.

The reason for my assumption is not only due to the presence of /–h/ in Ruc. In modern Vietnamese, we can easily find two more syllable types that are not shown in (13); that is nasal coda with tones 1C and 2C. There is no clarification why there are no Mon-Khmer cognates in this category (Gage 1985). They are assumed to be added by the influence of another language. In other words, emergence of contour tones might be triggered by the language contact, and that time the coda /–h/ worked as medium to give slots to contour tones. Arem and Ruc were exempted from this phonological change.

8. Irregular Tone Correspondences (31/109 cases)

Next, I will explain irregular correspondences. I posit three factors for the irregular correspondences, by which 17 cases out of 31 can be solved. The reasons for the rest 14 cases are unknown at present.

8.1 Spreading of voicing from minor-syllable to main-syllable (7 cases)

The first factor is spreading of voicing from minor-syllable to main-syllable. Before showing the examples, we should bear in mind the word structure of Arem and Ruc.

(15)			
	Arem	CV+CCVC (ʔ)	sesqui-syllable
	Ruc	CV+CCVC / T	sesqui-syllable
	Middle Vietnamese (17C)	CCVC / T	mono-syllable
	Modern Vietnamese	CVC / T	mono-syllable

Word structure of Arem and Ruc are sesqui-syllable, which consists of minor-syllable and main-syllable. Examples of up-shifting and down-shifting are shown as below.

(16) Up-shift (#5, 6, 11, 104)

No.	Gloss.	Arem	Ruc	Viet
5	'deep'	cirùː	ʈuː 2A	sə̆w 1A
104	'thunder'	kərɨ̀mʔ	k^hrɨm 2B	sə̆m 1B

Above examples show tones are up-shifted due to the [-voice] feature of the minor-syllables. Although lower tones (2A, 2B) are expected, they are higher tones (1A, 1B) in Viet.

(17) Down-shift (#7, 36, 109)

No.	Gloss.	Arem	Ruc	Viet
7	'chicken'	lakæː	rəkaːʔtəkaː 1A	ɣaː 2A
36	'husked rice'	ŋkɔːʔ	rəkɔːʔtəkɔː 1B	ɣaːw 2B

Above examples show tones are down-shifted due to the [+voice] feature of the minor-syllables. Although higher tones (1A, 1B) are expected, they are lower tones (2A, 2B) in Viet.

8.2 Misperception of phonation type difference with voicing of onsets and tones (4 cases)
The second factor is misperception of phonation type difference with voicing of onsets, which affected tones.

(18) Up-shift (#12, 13, 87, 88)

No.	Gloss.	Arem	Ruc	Viet	*P-V (Ferlus)
12	'go'	tì̤	tiː 2A	diː 1A	diː ʔ tiː
87	'bone'	sì̤əŋ	saːŋ 2A	sɯəŋ 1A	dʒaːŋ ʔ tʃaːŋ

In both cases, although lower tone (2A) is expected, they are higher tone (1A). Ferlus' reconstruction also suggests two options: voiced and voiceless onsets. I posit **breathy voice** because I have a similar experience in transcribing East Javanese. Examples below show sub-minimal pair.

(19) East Javanese

Gloss.	*Mistranscription	Corrected transcription
'four'	*paˑpatˀ ˧˥	paːpatˀ ˧˥
'father'	*p^haˑpaʔ ˧˥	pa̤ːpaʔ ˧˨

At first, I transcribed the first syllables as contrastive [-aspiration] and [+aspiration] on the initial consonant. The informant insisted they are 'voiceless' and 'voiced'. However, both initial consonants are apparently voiceless. After repetitions of reproduction, I found the difference is rather in phonation type on the following vowels: the contrast between [modal voice] and [breathy voice], or *modal voice* and *slack voice* (Ladefoged and Maddieson 1996: 63). The latter is also accompanied by a slightly low pitch. I suppose similar sounds existed and similar misperception took place in the history of Vietnamese.

8.3 Semantic change and innovations (6 cases)
The third factor is semantic change and innovations. They are shown in pairs. One example is as below.

(20) Semantic change and innovations (#73/74, 99/100, 105/106)

No.	Gloss.	Arem	Ruc	Viet	*P-V (Ferlus)
73	'eye'	mɛ̀t	mə̆at 1B/2B	mă̆t 1B	mat
74	'face'	(n.a.)	mə̆at 1B/2B	mă̆t 2B	mat

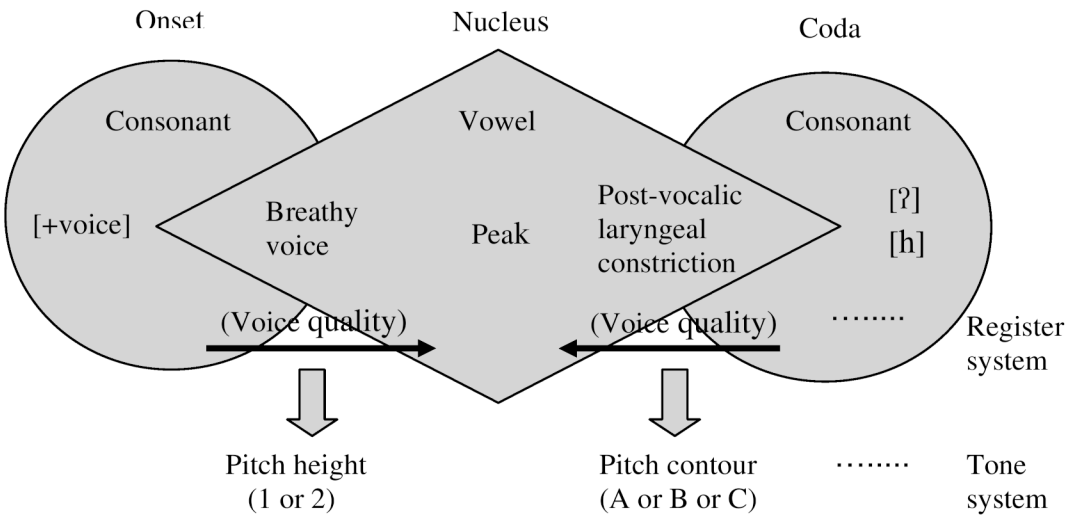
Although Vietnamese 'eye' and 'face' is phonologically different, it seems they have been derived from one etymon.

9. Summary of comparison – Haudricourt's model vs. Vietic evidence
Above analysis is summarised as follows:

- Phonation type difference (modal vs. breathy) is consistently reflected within Vietic.
- Chronological order of 2-way tone split by initial consonants to higher and lower series is not placed in the third stage of the development of tones, but the distinction between modal vs. breathy voice and the accompanying pitch distinction, had supposedly existed since the early stage of the proto language.
- Compared glottal stop [-ʔ] with fricative [-h] in syllable-end position, glottal stop was incorporated to tone system much earlier, while glottal fricative was incorporated to tone system later, as Arem and Ruc evidence shows.
- Therefore, the tone development was not 0 > 3 > 6 as Haudricourt, but 0 > 4 > 6, or even 0 > 2 > 4 > 6 may be possible, though evidence of 2-tone system is not found yet.
- Three main factors for irregular correspondences are posited: (1) spreading of voicing from minor-syllable to main-syllable, (2) misperception of phonation type difference with voicing of onsets and tones, and (3) semantic change and innovations.

10. Toward a laryngeally-based model of Tonogenesis
Lastly, I present a chart of the laryngeally-based model of tonogenesis below.

(21)



As it shows, the acoustic energy to create pitch difference was not directly supplied from consonants. It is the phonation, or voice quality, originally derived from initial and final consonants, which is the overlapping portion between consonants and vowel, that worked as medium or reservoir of acoustic energy. In the case of Vietic, this voice quality is both breathy voice and post-vocalic laryngeal constriction. Each of them becomes responsible for pitch height and pitch contour respectively, when the pitch element becomes more salient than voice quality, that is when the system changes from register system to tone system.

References

Alves, Mark (1997) "Ruc and other minor Vietic languages: linguistic strands between Vietnamese and the rest of the Mon-Khmer language family", papers from SEALS 7th conference.

Alves, Mark (1999) "What's so Chinese about Vietnamese?", papers from SEALS 9th conference.

Diffloth, Gérald (1989) "Proto-Austroasiatic creaky voice", *Mon-Khmer Studies* 15, 139-154.

Ferlus, Michel (1997 modified version of 1991) "Vocalisme du Proto Viet-Muong", papers from 24th International conference on Sino-Tibetan languages.

Ferlus, Michel (2001) "The origin of tones in Viet-Muong", papers from SEALS 11th conference.

Gage, William W. (1985) "Glottal stops and Vietnamese tonogenesis", *For Gordon H. Fairbanks ed. by V.Z. Acson & R.L. Leed, 21-36 (Oceanic Linguistics Special Publication, 20)*, University of Hawaii Press, Honolulu.

Haudricourt, André-George (1954) "De l'origine des tons en vietnamien", in *Journal Asiatique* 242, 69-82.

Honda, Koichi (2004) *Phonological history of Vietnamese – with special focus on the phonological system of Middle Vietnamese*, Master of Linguistics sub-thesis, The Australian National University, Canberra.

Ladefoged, Peter and Ian Maddieson (1996) *The sounds of the World's Languages*, Blackwell, Oxford.

Nguyễn Phú Phong, Trần Trí Dõi and M. Ferlus (1988) *Lexique Vietnamien – Rục – Francais*, Unversiti de Paris VII, Sudestasie.

Nguyễn Văn Lợi (1993) *Tiếng Rục [The Ruc language]*, Nhà Xuất bản Khoa học Xã hội.

Pham, Andrea Hoa (2003) *Vietnamese Tone – A New Analysis*, Routledge, New York.

Thurgood, Graham (2002) "Vietnamese and tonogenesis – Revising the model and the analysis", *Diachronica* vol.19, no.2, 333-363.

Vietnam News Agency (1996) *Vietnam: Image of the community of 54 ethnic groups*, The ethnic cultures publishing house, Hanoi.

	Item	**Arem** Ferlus 1997	**Ruc** Ferlus 1997	**Ruc** Lợi 1993	**Ruc** Phong et al. 1988	***Mod VN*** *orthography*	Tone	Exp. Tone	***P-VM** Ferlus 19
nal -Ø, -w, -j, -Ø?, -w?, -j?]									
	three	pæː	paː^{1}	pa^{1}	pa	*ba*	1A	1A	paː
	drunk	pərɪː	p^{h}riː^{1}	pri^{1}	pri pə̆do	*say*	1A	1A	priː > p^{h}riː / k
	arm, hand	t^{h}ɪː	siː^{1}	si^{1}	si	*tay*	1A	1A	siː
	ear	t^{h}aːj	saːj^{1}	saj^{1}	saj	*tai*	1A	1A	saːj
	deep	cirùː	ʈuː^{2}	chơru^{1}, djơru^{1}, tru^{1}, tu^{1}	rpï / ʒŭpï kàsă̆ŋ	*sâu*	1A	2A	cruː > k^{h}ruː
	buffalo, carabao	(n.a.)	cəluː^{2}	tlu^{1}, klu^{1}	klu	*trâu*	1A	2A	cluː > kluː / tl
	chicken (gen.)	lakæː	təkaː^{1}	rơka^{1}, tơka^{1}	rə̆ka	*gà*	2A	1A	rkaː
	betel	ulɛ̀w	(pluː^{2}) S	plu^{2}	plù	*trầu, giầu*	2A	2A	bluː > bluː / tl
	sky, heaven, God	tləːj^{1} (?)	pləj^{2}	ploj^{2}	pləj	*trời, giời*	2A	2A	b-ləːj > bləːj /
	fly (insect)	urùəj	pərɔːj^{1}	mơrôoj^{2}, pơrôoj^{2}	mŭrɔj	*ruồi*	2A	2A	m-rɔːj
	rain	mìa	kuməa^{2}	kumơa^{2}	kŭmʌa	*mưa*	1A	2A	k-maː > k^{h}ma
	go	t̂ɨː	tiː^{2}	ti^{2}	ti	*đi*	1A	2A	diː / tiː
	tail	tùəj	(tuej^{2})	tuoj^{2}	tuòj	*đuôi*	1A	2A	dɔːj / tɔːj
	lip	(n.a.)	cəboːj^{3}	chưbôj^{3}	cŭbó?j	*môi*	1A	1B	c-ɓuːj? / c-ɓuː
	bundle	pɑːʔ	pɔː^{3}	(n.a.)	(n.a.)	*bó*	1B	1B	pɔːʔ
	blood	mawʔ	təmuː^{3}	[asam^{3}]	[ă̆saʔ́m]	*máu*	1B	1B	t-muːʔ > t^{h}mu
	clothes	ʔɛːwʔ	ʔaːw^{3}	ʔaw^{3}	ʔáw	*áo*	1B	1B	ʔaːwʔ
	dog	acɑːʔ	acɔː^{3}	acho^{3}	acɔ́ʔ	*chó*	1B	1B	ʔa-cɔːʔ
	female	keːʔ	pəkiː^{3}	puki^{3}	pə̆kíʔ	*gái*	1B	1B	keːʔ
	fish	akæːʔ	akaː^{3}	aka^{3}	ă̆kaʔ́	*cá*	1B	1B	ʔa-kaːʔ
	have	kɑʔ	kɔː^{3}	ko^{3}	kɔ́	*có*	1B	1B	kɔːʔ
	leaf	ulæːʔ	ulaː^{3}	ula^{3}, hla^{3}	ŭlaʔ́	*lá*	1B	1B	s-laːʔ
	louse, head	cɪːʔ	ciː^{3}	chi^{3}	cí	*chấy, chí*	1B	1B	ciːʔ
	paper	kacajʔ	kəcaj^{3}	kưchạj^{3}	kə̆cắj	*giấy*	1B	1B	k-cajʔ
	remember, think about	ɲʌːʔ	kəɲəː^{3}	kưnhô^{3}	kă̆ɲɔ́	*nhớ*	1B	1B	k-ɲəːʔ > k^{h}ɲəː
	salt	bɑːjʔ	bɔːj^{3}	boj^{3}	bɔ́j	*muối*	1B	1B	ɓɔːjʔ
	six	pərawʔ	ʃaw^{3}	prạw^{3}, psaw^{3}, phrạw^{3}, saw^{3}	ʃráw	*sáu*	1B	1B	p-ruːʔ > p^{h}ruː

	Item	**Arem** Ferlus 1997	**Ruc** Ferlus 1997	**Ruc** Lợi 1993	**Ruc** Phong et al. 1988	***Mod VN*** *orthography*	Tone	Exp. Tone	***P-VM** Ferlus 1997
	smoke	ahɑ:jʔ	kəhɔ:j^{3}	khahoj^{3}	kăhɔ́j	*khói*	1B	1B	k-hɔ:jʔ > klɔ:jʔ / tlɔ:jʔ
	wind (blowing)	kajɑ:ʔ	kəjɔ:^{3}	kajo^{3}	kăjɔ́ʔ	*gió*	1B	1B	k-jɔ:ʔ
	bee, wild honeybee	kwi:	kwi:^{3}	kwi^{2}	(n.a.)	*khoái*	1B	1B	kwe:ʔ
	rice (plant and grain)	alɑ:ʔ	alɔ:^{3}	alo^{3}	ălɔ́	*lúa*	1B	1B	ʔa-lɔ:ʔ
	fruit	ule:ʔ	pəli:^{3}	puli^{3}	pəlí	*trái*	1B	1B	p-le:ʔ > ple:ʔ / tle:ʔ
	stone	atæ:ʔ	lata:^{3}	lata^{3}, ta^{3}	taʹʔ	*đá*	1B	1B	l-ta:ʔ
	forest	brawʔ	bru:^{3}	bru^{3}	brú	*[rừng] rú*	1B	1B	m-ru:ʔ
	mother	(n.a.)	miɛ^{4}	mêe^{4}	meɛ.	*mẹ*	2B	2B	me:ʔ
	rice (husked)	ŋkɔ:ʔ	təko:^{3}	rơkô^{3}	rə̆kó	*gạo*	2B	1B	r-ko:ʔ
hal -h, -s]									
	break	pɛh	peh^{1}	(n.a.)	pɛh	*bẻ, bể*	1C	1C	pɛh, peh
	grass	(n.a.)	kɔh^{1}	kọh^{1}	kɔh	*cỏ*	1C	1C	kɔh
	red	(n.a.)	(tɔ:^{6})N, tɔh^{1}	tɤkơ^{3}	tọʔ	*đỏ*	1C	1C	tɔh
	cloth	kupɐ:lʔ (?)	(kupa:l^{6}) N	kupal^{3}	kupạʔl	*vải*	1C	1C	kpa:s
	vomit	abah	bah^{1}	bah^{1}	(n.a.)	*mửa*	1C	1C	ɓah
	breathe	aŋòh	təŋəh^{1}	tangơh^{1}	tă̆ŋəh	*thở, ngửi*	1C	1C	tŋəs > t^{h}ŋəw
	nest	nʊh	suh^{1}	ʔô^{4}	ʔo [to]	*ổ, tổ*	1C	1C	suh / ʔuh
	seven	pʌh	(paj^{6}) N	paj^{4}	pặʔj	*bảy, bẩy*	1C	1C	pəs
	nose	mùh	mu:r^{h1}	mujh^{2}, murh^{2}	muʌh, mulʃ	*mũi*	2C	2C	mu:s
	tongue	lì̀əh	ləar^{h1}	lơarh^{2}, lơajh^{2}	ləarh, lʌarl	*lưỡi*	2C	2C	la:s
hal -l, -r]									
	fly (verb)	pal	pəl^{1}	pơr^{1}	pəl	*bay*	1A	1A	pər
	lime (substance)	apʊ:l	kəpu:l^{1}	kapur^{1}, kapul^{1}	kă̆pul	*vôi*	1A	1A	kpu:r
	two	hæ:l	ha:l^{1}	hal^{1}	hal	*hai*	1A	1A	ha:r
	cloud	(n.a.)	məl^{1}	mọl^{1}	mă̆l	*mây*	1A	1A	-məl
	tree	(n.a.)	(kəj^{1})	kơâj^{2}	kəʌj	*cây*	1A	1A	kəl
	return	(n.a.)	(vi:^{2})	[Bi^{2}]	(n.a.)	*về*	2A	2A	ve:l
	long	(n.a.)	jəal^{2}	jar^{1}, jal^{1}	ʔjal	*dài*	2A	2A	ja:r

	Item	**Arem** Ferlus 1997	**Ruc** Ferlus 1997	**Ruc** Lợi 1993	**Ruc** Phong et al. 1988	***Mod VN*** *orthography*	Tone	Exp. Tone	***P-VM** Ferlus 1997
	plow, plough	(n.a.)	kə̆al^{2}	kơâl^{2}	kʌ̀l	*cày*	2A	2A	gal
	roof	(n.a.)	baːl^{3}	bal^{3}	bál	*mái*	1B	1B	ɓaːlʔ
	egg	[uluːlʔ]	[tulul^{3}]	[talur^{3}, tulul^{3}]	təlúl	*trứng*	1B	1B	tluːrʔ
	wake, get up, rise	ɟɨlʔ	jɨl^{4}	(n.a.)	ɲəʌ́p ʒɨrɨ	*dậy*	2B	2B	jɨrʔ
	rise, raise	ajalʔ	ajəl^{3}	(n.a.)	(n.a.)	*dấy*	1B	1B	ʔa-jərʔ
	run	(n.a.)	(n.a.)	tangaj^{3}	tə̆ŋə́j	*chạy*	2B	1B	ɟalʔ
	light (weight)	(n.a.)	ɲeːl^{3}	nhel^{3}	ə̆ɲɛ́l	*nhẹ*	2B	1B	ɲeːlʔ
	dust	(n.a.)	(n.a.)	[puj^{4}]	[kə̆́ʔjíh, kăjih]	*bụi*	2B	2B	buːlʔ
nal -p, -t, -c, -k]									
	lightning	cʌːp	cəːp^{1}	[alar^{1}]	cə́p	*chớp*	1B	1B	cəːp
	sing	ahæːt	haːt^{1}	hat^{3}	hát	*hát*	1B	1B	haːt
	iron	(n.a.)	k^{h}lat^{1}	khlạt^{3}	klắt	*sắt*	1B	1B	krac > k^{h}rac
	cool	(n.a.)	rəmaːc^{1}	(n.a.)	mát	*mát*	1B	1B	tmaːc
	cut	kac (?)	kac^{1}	[pơâm^{2}]	kắc	*cắt, chặt*	1B	1B	kac
	sand	təkaːc	təkaːc^{1}	tɯkach^{3}	tə̆kắc	*cát*	1B	1B	tkaːc
	hair (head)	ut^{h}ʊk	usuk^{1}	usuk^{3}	usúk	*tóc*	1B	1B	suk
	stump, base of tree	təkoːk	təkoːk^{1}	[kɯl kơâj]	tə̆kòk kəʌj	*gốc*	1B	1B	tkoːk
	water	dæːk	daːk^{1}	dak^{3}	dák	*nước*	1B	1B	ɗaːk
	one	mùːc	mʊːc^{2}	môc^{4}	moiʔc	*một*	2B	2B	moːc
	wear	(mə̀k)	mə̆ak^{2}	mơâk^{4}	məắk ʔáw	*mặc*	2B	2B	mak
	eye	mə̀t	mə̆at^{2}	mơât^{4}	(n.a.)	*mắt*	1B	2B	mat
	face	(n.a.)	(n.a.)	mơât^{4}	məʌ́t	*mặt*	2B	2B	mat
nal -m, -n, -ɲ, -ŋ, -mʔ, -nʔ, -ɲʔ, -ŋʔ]									
	bird	iciːm	ciːm^{1}	ichim^{1}	ĭcim	*chim*	1A	1A	ciːm
	five	dam	dam^{1}	dạm^{1}, ʔdạm^{1}	dăm	*năm*	1A	1A	ɗam
	hundred	tlam	klam^{1}	klạm^{1}	klăm	*trăm*	1A	1A	klam > klam / tlam
	rice (cooked)	kəːm	kəːm^{1}	kơm^{1}, (chaw^{3})	kəm	*cơm*	1A	1A	kəːm
	year	(n.a.)	nam^{1}	nạm^{1}	năm	*năm*	1A	1A	cnəm
	child	kɑːn	kɔːn^{1}	kon^{1}, kon^{4}	kɔn	*con*	1A	1A	kɔːn

Item	**Arem** Ferlus 1997	**Ruc** Ferlus 1997	**Ruc** Lợi 1993	**Ruc** Phong et al. 1988	***Mod VN*** *orthography*	Tone	Exp. Tone	***P-VM** Ferlus 1997
eat	ʔan	ʔan[1]	ʔan	ʔăn	*ăn*	1A	1A	ʔan
flower	(n.a.)	poːŋ[1]	pông[1] pơar[1]	pɨarl, pɨal	*bông*	1A	1A	poːŋ
in	tlɑːŋ	klɔːŋ[1]	klọng[1]	klɔŋ	*trong*	1A	1A	klɔːŋ > klɔːŋ / tlɔːŋ
on, on toop	ləːŋ	liːŋ[1]	aling[1], ling[1]	ălɨŋ, leɲ	*trên, lên*	1A	1A	kleːŋ > kleːŋ / tleːŋ
tooth	at^{h}aŋ	kəsaŋ[1]	kasạng[1]	kàsăŋ	*răng*	1A	1A	ksaŋ
ask, question	(n.a.)	haːɲ[1]	hanh[1]	(n.a.)	*[hỏi] han*	1A	1A	haːɲ
bone	sı̀əŋ	saːŋ[2]	sang[1]	sàŋ	*xương*	1A	2A	dʒaːŋ / tʃɨəŋ
foot, leg	cı̀ɲ	ciːŋ[2]	ching[2]	ciŋ	*chân*	1A	2A	ɟiːŋ / ciɲ
bed	(n.a.)	kəciːŋ[2]	kưchơng[1]	ciə̀ŋ	*giường*	2A	2A	kɟəːŋ
near	(n.a.)	təkiɲ[1]	chưkinh[2]	ckɨ̀ɲ, ckɨ̌ɲ	*gần*	2A	2A	tkiɲ
eight	t^{h}æːmʔ	(t^{h}aːm[3]) N	tham[3], tơham[3]	thám	*tám*	1B	1B	saːmʔ
four	puənʔ	poːn[3]	pôn[3]	pón	*bốn*	1B	1B	poːnʔ
nine	ciːnʔ	ciːn[3]	chin[3]	cín	*chín*	1B	1B	ciːnʔ
ripe, cooked	ciːnʔ	ciːn[3]	chin[3]	cín	*chín*	1B	1B	ciːnʔ
shoot	piɲʔ	piɲ[3]	pịnh[3]	píɲ	*bắn*	1B	1B	pəɲʔ
snake	ut^{h}iɲʔ	pəsiɲ[3]	pusịnh[3]	pəsíɲ	*rắn*	1B	1B	psəɲʔ
palm (of hand)	adæːŋʔ	kədaːŋ[3]	(n.a.)	(n.a.)	*náng*	1B	1B	kɗaːŋʔ
month	t^{h}æːŋʔ	t^{h}aːŋ[3]	thang[3]	tháŋ	*tháng*	1B	1B	kraːŋʔ
wing	(n.a.)	keːŋ[3]	kêng[3]	kéŋ	*cánh*	1B	1B	kɛːŋʔ
branch	kɑːŋʔ (V!)	təkɛːŋ[3]	takeng[3]	tăkéŋ	*cành*	2A	1B	tkɛːŋʔ / gɛːŋ
bitter	taŋ[3]	tạng[3]	ətắʔŋ	*đắng*	1B	1B	taŋʔ
live, be alive	tloŋʔ	kloːŋ[3]	tlung[4], klung[4]	klúŋ, kluóŋ	*sống*	1B	1B	k-loːŋʔ / kroːŋʔ
salty	mènʔ	màn[4]	mạn[4]	mặn	*mặn*	2B	2B	
thunder	kəɤ̀mʔ	k^{h}ɤ̆m[4]	krưm[4]	krɨʔm	*sấm*	1B	2B	krɨmʔ > k^{h}ɤ̆mʔ
stand	tɨŋʔ (?)	tɨŋ[6] (?)	tựng[4]	tɨ́ʔŋ	*đứng*	1B	2B	tɨŋʔ
build	pətɨŋʔ	pətɨŋ[6]	(n.a.)	pătɨ̀ŋ	*dựng*	2B	2B	p-tɨŋʔ
borrow	maːɲʔ (?)	məaɲ[4]	mơan[4]	mɨạʔn	*mượn*	2B	2B	maːɲʔ
heavy	naŋʔ (?)	naŋ[6] (?)	nạng[4], ʔnạng[4]	nắʔŋ	*nặng*	2B	2B	naŋʔ
louse (body)	^{n}rɨɲʔ, ^{n}tʃɨɲʔ	briɲ[3]	brợnh[3]	bríɲ	*rận*	2B	1B	mrɨɲʔ > p^{h}rɨɲʔ

narks:	(?)	: borrowing, noted by Ferlus	[ ]	: suspicious cognates, noted by Honda
	( ) S	: Sach (alternative), noted by Ferlus	( ) N	: Nguôn (alternative), noted by Ferlus

for Appendix:

Michel (1991, 1997 modified version) "Vocalisme du Proto Viet-Muong", Paper to be circulated at the 24th International Conference on Sino-Tibetan Languages and linguistics

2001) "The Origin of Tones in Viet-Muong", Paper circulated at the Southeast Asian Linguistic Society 11th Conference

Văn Lợi (1993) Tiếng Rục [The Ruc Language], Nxb Khoa Học Xã Hội , Hanoi, 166 pp.

Phú Phong - Trần Trí Dõi - M. Ferlus (1988) Lexique Vietnamien - Rục - Francais, Universite de Paris VII, Sudestasie, 93 p

PROTO-KATUIC PHONOLOGY AND THE SUB-GROUPING OF MON-KHMER LANGUAGES[1]

Paul Sidwell
Australian National University
paul.sidwell@anu.edu.au

Summary
The Katuic languages are a branch of the Mon-Khmer (MK) family spoken by more than one million people living in Thailand, Cambodia, Laos and Vietnam. While the majority of Katuic speakers live in eastern Thailand and Cambodia, the greatest diversity of Katuic languages lies in the Salavan and Sekong provinces of Laos and adjacent border areas of Vietnam, part of a complex patchwork of small ethnic communities. From a comparative-historical point of view Katuic has particular importance, as between them the languages appear to have conserved some very ancient phonological and lexical features. At the same time some Katuic languages have been remarkably innovative and developed some of the richest vowel systems in the world. The recent advances in the reconstruction of Proto Katuic (Sidwell 2005) potentially allow us to investigate the sub-grouping of Katuic within Mon-Khmer on the basis of comparative phonology. However, the results are somewhat ambiguous, and do not support any special sub-grouping of Katuic within Mon-Khmer.

Classification of Katuic with the Mon-Khmer family

During the first major phase of comparative-historical work on the MK languages, which lasted into the 1960s (effectively beginning with the work of Schmidt (1901, 1904, 1905 etc.) until Pinnow (1959) and Shafer (1952, 1965)), there was no coherent account of the real extent and internal structure of the Mon-Khmer family.

Thomas and Headley (1970) established a new paradigm when they successfully applied lexicostatistics to the emerging body of new field data, distinguishing nine branches: *Pearic, Khmer, Bahnaric, Katuic, Khmuic, Monic, Palaungic, Khasi* and *Viet-Mɯ̀ɔng*. Adding *Aslian* and *Nicobarese* (not examined by Thomas and Headley although already long recognised as MK), Diffloth's (1974) expanded listing became the received classification[2]. The Munda languages of India are also generally recognised as related to MK, although opinion is divided over how close that relationship is. All together they are recognised as forming the Austroasiatic phylum, but in this paper I am only concerned with analysis up to the MK level.

[1] I would like to thanks Mark Alves for valuable advice and comment on this paper, in addition to his ongoing cooperation and encouragement in respect of my broader Mon-Khmer research activities. This paper was also made possible by assistance from the Department of Linguistics of the Max Planck Institute (Leipzig) in the form of stipend support during 2005.

[2] Since the 1980s some minor languages of China have come to light that may or may not constitute a new branch.

Paul Sidwell, ed. *SEALSXV: papers from the 15th meeting of the Southeast Asian Linguistics Society*.
Canberra, Pacific Linguistic, 2005, pp.193-204.