

Handout materials for course

**‘Linguistic Areas and Language Contact (with a focus on Southeast Asia)’**

N. J. Enfield

(Nick.Enfield@mpi.nl)

**LSA 215, 27 July – 12 August, 2009**

Berkeley, CA

Mon/Wed 1.30-3.15pm

## Phonology

Some general statements about the phonological systems of MSEA languages:

- (a) **Consonants** Large number of initial segmental contrasts (tending to be heavily constrained in final position).
- (b) **Laryngeal features in lexical contrast:** Most languages utilize ‘suprasegmental’ laryngeal contrasts in their phonological systems (register, tone).
- (c) **Vowel systems** are among the largest in the world, from nine simple vowel contrasts, along with length contrasts, and a range of complex vowels (diphthongs or /VV/ sequences). High back unrounded vowel in many languages. Depending on how you count, some languages have over 50 vowel phonemes.
- (d) **Phonotactics** generally work on an /initial+rhyme/ basis, where final syllables are very restricted (i.e. widespread neutralization in syl-final position). There is a tension between analysis of phonotactic structure, on the one hand, and complex vowels and vowel length, on the other. Many languages have an initial unstressed ‘minor syllable’, in which vocalic distinctions are neutralized.

Eg. Khmu (Northern Mon-Khmer, Laos; Suwilai 1987)

		Labial	Alveolar	Palatal	Velar	Glottal
Stops	Voiceless unaspirated	p-	t-	c-	k-	ʔ-
	Voiceless aspirated	p <sup>h</sup> -	t <sup>h</sup> -	c <sup>h</sup> -	k <sup>h</sup> -	
	Voiced	b-	d-			
Fricatives	Voiceless		s-			h-
Nasals	Voiced	m-	n-	ɲ-	ŋ-	
Liquids	Trill		r-			
	Lateral		l-			
Glides	Voiced	w-			j-	

Khmu initials (Suwilai 1987:7) (notice: no *f*-)

		Trill	Lateral	Glide
Labial	Voiceless unaspirated stop	<b>pr-</b>	<b>pl-</b>	
	Voiceless aspirated stop	<b>p<sup>h</sup>r-</b>		
Alveolar	Voiceless unaspirated stop	<b>tr-</b>		
	Voiceless aspirated stop	<b>sr-</b>		
Palatal	Voiceless unaspirated stop	<b>cr-</b>		
	Voiceless aspirated stop	<b>c<sup>h</sup>r-</b>		
Velar	Voiceless unaspirated stop	<b>kr-</b>	<b>kl-</b>	<b>kw-</b>
	Voiceless aspirated stop	<b>k<sup>h</sup>r-</b>		<b>k<sup>h</sup>w-</b>

### Khmu clusters

		Labial	Alveolar	Palatal	Velar	Glottal
Stops	Voiceless unaspirated	<b>-p</b>	<b>-t</b>	<b>-c</b>	<b>-k</b>	<b>-ʔ</b>
	Voiceless aspirated	-	-	-	-	
	Voiced	-	-			
Fricatives	Voiceless		<b>-s [ç/x]</b>			<b>-h</b>
Nasals	Voiced	<b>-m</b>	<b>-n</b>	<b>-ɲ</b>	<b>-ŋ</b>	
Liquids	Trill		<b>-r</b>			
	Lateral		<b>-l</b>			
Glides	Voiced	<b>-w</b>			<b>-j</b>	

### Khmu finals

		unrounded		rounded
	front	central	back	
high	<b>i i:</b>		<b>ɯ ɯ:</b>	<b>u u:</b>
mid	<b>e e:</b>	<b>ə ə: ʌ:</b>		<b>o o:</b>
low	<b>ɛ ɛ:</b>	<b>a a:</b>		<b>ɔ ɔ:</b>

Diphthongs: **iə, uə, ɯə**

Note also two tones/register 'tense' and 'lax'.

### Khmu

Compare this with Lao - fewer initials, fewer initial clusters, fewer finals, tones.

		Labial	Alveolar	Palatal	Velar	Glottal
Stops	Voiceless unaspirated	<b>p-</b>	<b>t-</b>	<b>c-</b>	<b>k-</b>	<b>ʔ-</b>
	Voiceless aspirated	<b>p<sup>h</sup>-</b>	<b>t<sup>h</sup>-</b>		<b>k<sup>h</sup>-</b>	
	Voiced	<b>b-</b>	<b>d-</b>			
Fricatives	Voiceless	<b>f-</b>	<b>s-</b>			<b>h-</b>
Nasals	Voiced	<b>m-</b>	<b>n-</b>	<b>ɲ-</b>	<b>ŋ-</b>	
Liquid	Lateral		<b>l-</b>			
Glides	Voiced	<b>v-</b>			<b>j-</b>	

Lao initials. (Initial clusters in Lao are *kw-* and *k<sup>h</sup>w-*, in some varieties.)

		Labial	Alveolar	Palatal	Velar	Glottal
Stops	Voiceless unaspirated	<b>-p<sup>ʔ</sup></b>	<b>-t<sup>ʔ</sup></b>	-	<b>-k<sup>ʔ</sup></b>	<b>(-ʔ)</b>
	Voiceless aspirated	-	-		-	
	Voiced	-	-			
Fricatives	Voiceless		-			-
Nasals	Voiced	<b>-m</b>	<b>-n</b>	-	<b>-ŋ</b>	
Liquid	Lateral		-			
Glides	Voiced	<b>-w</b>			<b>-j</b>	

Lao finals somewhat fewer than Khmu

	front	unrounded central	back	rounded
high	<b>i i:</b>		<b>ɯ ɯ:</b>	<b>u u:</b>
mid	<b>e e:</b>	<b>ə ə:</b>		<b>o o:</b>
low	<b>ɛ ɛ:</b>	<b>a a:</b>		<b>ɔ ɔ:</b>

Diphthongs: *ia*, *ua*, *ɯa* (and in Northern varieties *au*).

Lao vowels.

Here is the vowel system of an Austronesian language heavily under the influence of Khmer, and Mon-Khmer generally – Modern Western Cham (Headley 1998:25ff; cf. Thurgood):

	front	central	back
high	<b>i i:</b>	<b>ɨ ɨ:</b>	<b>u u:</b>
mid	<b>e</b>	<b>ə ə:</b>	<b>o o:</b>
low	<b>ɛ ɛ:</b>	<b>a a:</b>	<b>ɔ ɔ:</b>

Diphthongs: **ia, ii, ea, au, ua, oa.**

Registers: **high, low.**

Modern Western Cham vowels.

Note much higher number of vowels than would normally be found in an Austronesian lg:

	front	central	back
high	<b>i</b>		<b>u</b>
mid	<b>e</b>	<b>ə</b>	<b>o</b>
low		<b>a</b>	

Bahasa Malaysia Vowels.

**Hmong-Mien;** One variety of ‘Miao’ spoken in China illustrates a very high number of initial consonant contrasts, which in combination with an eight-tone system is in turn related to the very low number of vowels and finals (from Ramsey 1987:282):

		Labial	Alveolar	Retroflex	Palatal	Velar	Uvular
Stops	Voiceless unasp'd	<b>p-</b>	<b>t-</b>		<b>c-</b>	<b>k-</b>	<b>q-</b>
	Voiceless aspirated	<b>p<sup>h</sup>-</b>	<b>t<sup>h</sup>-</b>		<b>c<sup>h</sup>-</b>	<b>k<sup>h</sup>-</b>	<b>q<sup>h</sup>-</b>
Prenasal Stops	Voiceless unasp'd	<b><sup>m</sup>p-</b>	<b><sup>n</sup>t-</b>		<b><sup>ɲ</sup>c-</b>	<b><sup>ŋ</sup>k-</b>	<b><sup>ɴ</sup>q-</b>
	Voiceless aspirated	<b><sup>m</sup>p<sup>h</sup>-</b>	<b><sup>n</sup>t<sup>h</sup>-</b>		<b><sup>ɲ</sup>c<sup>h</sup>-</b>	<b><sup>ŋ</sup>k<sup>h</sup>-</b>	<b><sup>ɴ</sup>q<sup>h</sup>-</b>
Affricates	Voiceless unasp'd		<b>ts-</b>	<b>tʂ-</b>	<b>tc-</b>		
	Voiceless aspirated		<b>ts<sup>h</sup>-</b>	<b>tʂ<sup>h</sup>-</b>	<b>tc<sup>h</sup>-</b>		
Prenasal Affrcts	Voiceless unasp'd		<b><sup>n</sup>ts-</b>	<b><sup>ɳ</sup>tʂ-</b>	<b><sup>ɲ</sup>tc-</b>		
	Voiceless aspirated		<b><sup>n</sup>ts<sup>h</sup>-</b>	<b><sup>ɳ</sup>tʂ<sup>h</sup>-</b>	<b><sup>ɲ</sup>tc<sup>h</sup>-</b>		
Fricatives	Voiceless	<b>f-</b>	<b>s-</b>	<b>ʂ-</b>	<b>ç-</b>		
	Voiced	<b>v-</b>		<b>ʐ-</b>	<b>ʒ-</b>		
Nasals	Voiceless	<b>ɱ-</b>	<b>ɳ-</b>		<b>ɲ-</b>		
	Voiced	<b>m-</b>	<b>n-</b>		<b>ɲ-</b>	<b>ŋ-</b>	
Liquid	Lateral, voiceless		<b>ɬ-</b>				
	Lateral, voiced		<b>l-</b>				
Glides	Voiced	<b>w-</b>					

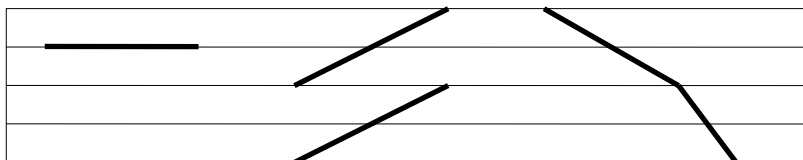
Initials of Dànáshān Miao, Guizhou (Ramsey 1987:282).

This system has only five simple vowels [**i, e, a, u, o**], and the only permissible consonant finals are **-n** and **-ŋ**, and these are not contrastive. Final combinations are **-en**, **-aŋ**, and **-oŋ**.

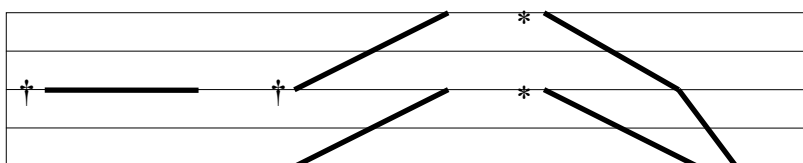
The system has eight tones. *falling* (*hi, mid, lo*); *level* (*hi, hi-mid, mid*), *rising*, (*low, mid*).

## Tones

Most non AA languages in MSEA have ‘tone’ – i.e., where distinct pitch shapes serve to consistently distinguish lexical items.



MSC tones (*mā* ‘mother’, *má* ‘hemp’, *mǎ* ‘horse’, *mà* ‘scold’).



Lao tones.

\* - also appears on [V<sub>long</sub>+stop] syllable

† - also appears on [V<sub>short</sub>+stop] syllable

[**saw**] 5 tones: *saw1* ‘rent’, *saw2* ‘stop’, *saw3* ‘post’, *saw4* ‘morning’, *saw5* ‘sad’

[**kha:**] 5 tones: *khaal* ‘galangal’, *khaa2* ‘stuck’, *khaa3* ‘leg’, *khaa4* ‘trade’, *khaa5* ‘kill’

[**het**] with two tones: *hêt1* ‘do’, *hêt2* ‘mushroom’

[**khop**] with two tones: *khop1* ‘meet’, *khop2* ‘bite’

[**sa:p**] with two tones: *saap4* ‘to know’, *saap5* ‘dank’

There are constraints on the occurrence of certain tones with certain syllable types, depending on (a) whether the major-initial consonant is a non-aspirated stop (b-, d-, p-, t-, c-, k-, ?-) or j- (versus any other consonant initial), (b) whether there is a final stop consonant, and (c) if so, whether the vowel is short or long. The constraints are as follows:

1. If C1 is a non-aspirated stop or j- and C2 is not a stop, Tones 2 and 5 do not occur (cf. *paa1* ‘forest’ versus *paa3* ‘fish’ versus *paa4* ‘elder aunt (Pa.eZ)’).
2. If C2 is a stop and the vowel is short, tones 3, 4, and 5 do not occur (cf. *khop1* ‘meet’ versus *khop2* ‘bite’); in addition, if C1 is a non-aspirated stop or j-, Tone 1 does not occur (cf. *kop2* ‘frog sp.’).
3. If C2 is a stop and the vowel is long, tones 1, 2, and 3 do not occur (cf. *saap4* ‘to know’ versus *saap5* ‘dank’); in addition, if C1 is a non-aspirated stop or j-, Tone 4 does not occur (cf. *daap5* ‘sword’).

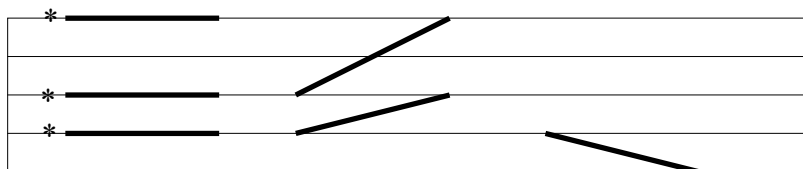
These constraints can be interpreted with reference to traditional, historical-comparative description of the tone system of Lao. Traditional analysis (Gedney 1989) makes reference to three parameters which determine the tone category of a word. First, the initial consonant of a

syllable will belong to one of three classes (as determined by the manner of articulation of the syllable in Proto Southwestern Tai): low (reconstructed as ‘voiceless, plus friction’), middle (reconstructed as ‘voiceless, minus friction’), or low (reconstructed as voiced). Second, a syllable will be either ‘live’ (i.e., with vocalic or sonorant final) or ‘dead’ (i.e., with stop final). Third, if ‘live’, it makes a difference whether the syllable is historically unmarked for tone (A), or historically marked as Tone 1 (B) or Tone 2 (C). Fourth, if ‘dead’ (D), it makes a difference whether a syllable has a long or a short vowel. This set of possibilities allows a historical-comparative map of the Lao tone system, as follows (for further information, see Gedney 1989):

Table 5. Modern Lao examples of reflexes of traditional Tai tones

OLD INITIAL	A	B	C	Dshort	Dlong
VOICELESS, +FRICTION	<i>khaa3</i> ‘leg’	<i>khaa1</i> ‘galangal’	<i>khaa5</i> ‘slave’	<i>khat2</i> ‘contrary’	<i>khaat5</i> ‘torn’
VOICELESS, –FRICTION	<i>kaa3</i> ‘crow’	<i>kaa1</i> ‘fish sp.’	<i>kaa4</i> ‘dare’	<i>kat2</i> ‘bite’	<i>kaat5</i> ‘cabbage’
VOICED	<i>khaa2</i> ‘stuck’	<i>khaa1</i> ‘cost’	<i>khaa4</i> ‘commerce’	<i>khat1</i> ‘select’	<i>khaat4</i> ‘hope’

### Cantonese tones:



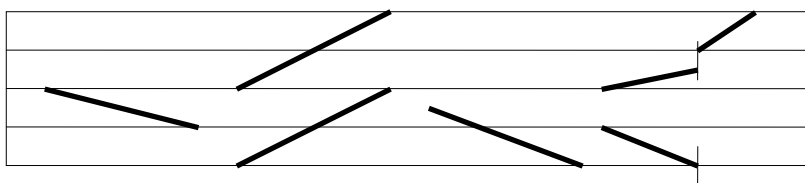
High-falling tone existed, but is being phased out now.

Minimal set (Matthews and Yip 1994:21):

- High level: *yāu* ‘worry’
- High rising: *yáu* ‘paint’ (n.)
- Mid level: *yau* ‘thin’
- Low falling: *yàuh* ‘oil’
- Low rising: *yáuh* ‘have’
- Low level: *yauh* ‘again’

“It is important to recognize that pitch is frequently only one of the phonetic components of ‘tone’ as a phonological category. A phonological tone is in our area very frequently a complex of other features besides pitch—such as intensity, duration, voice quality, final glottal constriction and so on.” (Eugénie Henderson 1967)

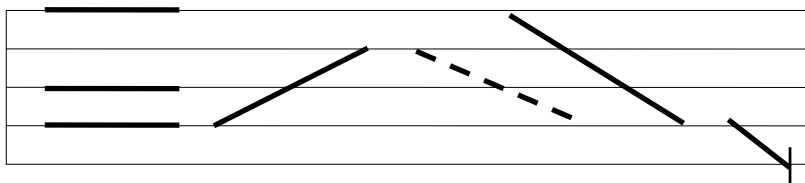
**Vietnamese:** From left to right, above before below – *ngang, sắc, hỏi, huyền, ngã, nặng*:



Vietnamese tones.

*ma* ‘ghost’    *má* ‘mother’    *mã* ‘horse’  
*mả* ‘tomb’    *mà* ‘which’    *mạ* ‘rice seedling’

Also in **White Hmong** there are phonation aspects to certain tones, here we see one tone breathy, and one tone checked (Ratliff 1992):



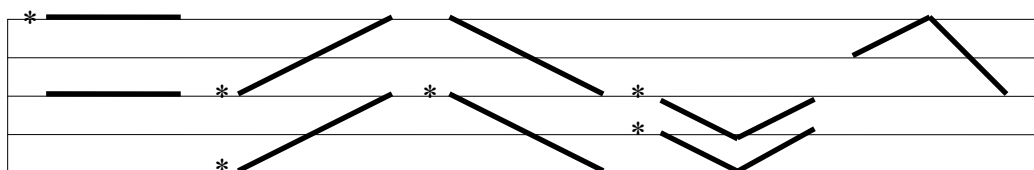
White Hmong tones.

Tones are indicated in the orthography by final letter, as follows:

level: *-o* [33] *-s* [22] *-b* [55]  
 rising: *-v* [24]  
 falling: *-j* [52] *-m* [21, checked] *-g* [42, breathy]

E.g. *koj* [ko<sup>52</sup>].  
*mus* [mu<sup>22</sup>].  
*kom* [ko<sup>21</sup>].  
*tau* [tao<sup>33</sup>].

Tones become largest in number in the Tai and Kam-Sui speaking areas towards the South of China. The ‘record number’ according to Ramsey (p244) is ‘15’ (9 on open syllables, 5 on checked) - Dong or Kam (corner of Hunan, Guizhou, and Guangxi: Long & Zheng 1998:31):



Dong (Kam) tones.

Minimal sets from Dong (Kam):

<i>sau</i> <sup>55</sup> ‘twist’	<i>jek</i> <sup>55</sup> ‘wet’
<i>sau</i> <sup>35</sup> ‘straw’	<i>jek</i> <sup>35</sup> ‘hard-working’
<i>sau</i> <sup>212</sup> ‘rear’ (v.)	<i>jek</i> <sup>21</sup> ‘pitiful’
<i>sau</i> <sup>323</sup> ‘steam’ (v.)	<i>jak</i> <sup>323</sup> ‘hungry’
<i>sau</i> <sup>13</sup> ‘grass carp’	<i>jak</i> <sup>13</sup> ‘fish pond’
<i>sau</i> <sup>31</sup> ‘husband’	<i>jak</i> <sup>31</sup> ‘rust’
<i>sau</i> <sup>53</sup> ‘soup’	
<i>sau</i> <sup>453</sup> ‘egret’	
<i>sau</i> <sup>33</sup> ‘create’	

**Tonogenesis** in historical linguistics in MSEA.

Vietnamese - competing features of high cognacy in basic vocab suggesting MK; phonology (and a lot of non-basic vocab) suggesting Sinitic. Acc. to Gage p497 there is a ‘great typological gulf’ (phonologically) between VN and other MK (especially, tone and monosyllabism). Maspero in 1912 (Maspéro, Henri, 1912. *Phonétique historique de la langue Annamite: les initiales. Bulletin de l’Ecole Française d’Extrême-Orient*, 12.1.) said that the conclusive point against a relationship was the tones - he said that VN tones could be derived from Proto-Tai, but he couldn’t see how they would arise in an MK lg (i.e. he said they were purely genetic phenomenon). At least Maspero raised the question of tones, which had been ignored - Haudricourt in 1954 published an article in French in which he argued that the tones *could* have arisen in an MK language (i.e. that they were not limited to genetic sources). (Haudricourt, André-Georges. 1954. De l’origine des tones en Vietnamien. *Journal Asiatique*, 242, 69-82.)

Norman on Sinitic - we know from Qieyun that Middle Chinese had four tonal categories: level, rising, departing, and entering. The latter had stop finals, the former three differed in tone contour. (Norman notes that historical correspondences are reconstructable, but not with respect to actual phonetic values.) Then, two registers emerged, when a voicing distinction was lost, and the difference in pitch compensated by becoming phonemic, and creating registers, splitting the tone system in two, and creating eight tones. (Still attested in some varieties - e.g. Yue languages.)

Initial class	Tone category			
	<i>píng</i> ‘level’	<i>shǎng</i> ‘rising’	<i>qù</i> ‘departing’	<i>rù</i> ‘entering’
<i>yīn</i> ‘upper’ (voiceless)	1. <i>yīnpíng</i> ‘upper level’	3. <i>yīnshǎng</i> ‘upper rising’	5. <i>yīnqù</i> ‘upper departing’	7. <i>yīnrù</i> ‘upper entering’
<i>yáng</i> ‘lower’ (voiced)	2. <i>yángpíng</i> ‘lower level’	4. <i>yángshǎng</i> ‘lower rising’	6. <i>yángqù</i> ‘lower departing’	8. <i>yánggrù</i> ‘lower entering’

Traditional Chinese tone categories (from Norman 1988:54)

Since then, most lgs have merged various of these categories (also note problems of the entering tones being hard to contrast, and misleading as to number of tones). (Any Sinitic language usually gets its tones mapped onto this plan.) **Vietnamese**: Haudricourt argued that VN could be plotted in the same way (Norman 1988:55):

<i>Tone category</i>				
Register	A	B	C	D
'Upper' (voiceless)	1. <i>ngang</i>	3. <i>sắc</i>	5. <i>hỏi</i>	7. <i>sắc</i>
'Lower' (voiced)	2. <i>huyền</i>	4. <i>nặng</i>	6. <i>ngã</i>	8. <i>nặng</i>

Traditional Vietnamese tone categories (from Norman 1988:54).

Category D emerged from words with original final stop. Words of category C were argued to correspond to words which historically ended in *-s* or *-h*; and words of category B were argued to derive from words with final glottal stop. (Norman.) Modern cognates show these. Then, a register split, due to voicing.

Step 1, four types of syllable in a toneless language, contrasting in final:

A	B	C	D
$C_{\pm\text{voiced}}-V-\emptyset/C_{+\text{voiced}}$	$C_{\pm\text{voiced}}-V-/\mathbf{}$	$C_{\pm\text{voiced}}-V-\mathbf{h/s}$	$C_{\pm\text{voiced}}-V-\mathbf{p/t/k}$

At this point, there are natural pitch distinctions associated with phonetic features of the finals (and of the  $\pm$ voicing of initials), but these are not contrastive.

Step 2, the finals are lost, but the lexical distinctions remain - the original non-contrastive pitch differences take up the load of lexical contrast:

A	B	C	D
$C_{\pm\text{voiced}}-V_{\text{tone A}}$	$C_{\pm\text{voiced}}-V_{\text{tone B}}$	$C_{\pm\text{voiced}}-V_{\text{tone C}}$	$C_{\pm\text{voiced}}-V-\mathbf{p/t/k}$

Precisely the same stage can be made more explicit by separating syllable types in terms of voicing, which again has associated regular non-contrastive pitch differences:

A	B	C	D
$C_{-\text{voiced}}-V_{\text{tone A}}$	$C_{-\text{voiced}}-V_{\text{tone B}}$	$C_{-\text{voiced}}-V_{\text{tone C}}$	$C_{-\text{voiced}}-V-\mathbf{p/t/k}$
$C_{+\text{voiced}}-V_{\text{tone A}}$	$C_{+\text{voiced}}-V_{\text{tone B}}$	$C_{+\text{voiced}}-V_{\text{tone C}}$	$C_{+\text{voiced}}-V-\mathbf{p/t/k}$

In step 3, these initial contrasts in voicing are lost (all initial voiced consonants become voiceless), and the regular pitch differences associated with erstwhile  $\pm$ voicing take on the load of maintaining lexical contrast:

A	B	C	D
$C-V_{\text{tone A1}}$	$C-V_{\text{tone B1}}$	$C-V_{\text{tone C1}}$	$C_{-\text{voiced}}-V-\mathbf{p/t/k}$
$C-V_{\text{tone A2}}$	$C-V_{\text{tone B2}}$	$C-V_{\text{tone C2}}$	$C_{+\text{voiced}}-V-\mathbf{p/t/k}$

This is borne out by comparative studies of Vietnamese. Norman (1988:55): ‘In comparing Vietnamese words with their Austroasiatic cognates, Haudricourt observed that Vietnamese words of category C (*hỏi-ngã*) frequently correspond to cognate words ending in *-h* or *-s* in other languages.’

Norman (1988:55)	Vietnamese	Mon	Mnong
‘seven’	bảy	tpah	poh
‘nose’	mũi	muh	mũh
‘root’	rễ	rɜh	ries

(Note Kri *pajh* ‘seven’, *muujh* ‘nose’, *leerh* ‘root’.)

Gage (1987:511)	Vietnamese	Khmer
‘scratch, draw’	kẻ	kɛs
‘toss (away)’	bỏ	poɰ
‘root’	rễ	rɜ
‘hole’	lỗ	luɰ

‘Likewise, in the case of category B (*sắc-nặng*), he noticed that Vietnamese words belonging to these tones were often related to words in Khmu and Riang (two related Mon-Khmer languages) which end in a glottal stop. (Table adapted from Norman 1988 and Gage 1987:512)

	Vietnamese	Khmu	Riang
‘leaf’	lá	hlaʔ	laʔ
‘rice’	gạo	rəŋkoʔ	koʔ
‘fish’	cá	kaʔ	-
‘dog’	chó	soʔ	soʔ
‘louse’	chí	-	siʔ
‘endure’/‘hurt’	chịu	cuʔ	-

(Note Kri *sulaaq* ‘leaf’, *qakaaq* ‘fish’, *cooq* ‘dog’, *ciiq* ‘louse’.)

Gage also notes the following for category A:

	Vietnamese	Khmu
‘bamboo’	tre	təlaa (Nth. Khmu)
‘thou’	mày	mee (2sg.m)

Norman, bottom of p56: Chinese may also have derived its original four tone system from finals (sonorant or zero final for level; glottal stop final for rising; final *-s* for ‘departing’ tone; and stop final for ‘entering’ tone; same as for Vn. cf. Norman 1988:57). Then later, value of *initial* became an issue (i.e. when loss of voicing created registers).

From Thurgood:

<b>Finals</b>	open, nasal (level)	stopped (rising)	Voiceless fricatives (falling)
<b>Initials</b>			
proto-voiceless (high pitch)	*pa > pa 'ngang'	*pak > pāk 'sắc'	*pas > pā 'hỏi'
proto-voiced (low pitch)	*ba > pà 'huyền'	*bak > pək 'nặng'	*bas > pã 'ngã'

It is now widely accepted that Vietnamese is a Mon-Khmer language, under Chinese influence (cf. quote in Ch2), minor syllables became lost. Gage, on VN in Mon-Khmer: 'The largest language is a maverick'. (Although there are some who still think that VN could be a Tai lg. cf. Gage p498.)

### Case of Kri (Vietic, Laos)

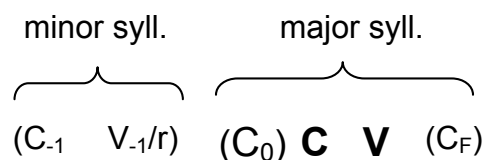


Figure 1. Segmental structure of the Kri word (non-segmental features not specified).

	labial	alveolar	retroflex	palatal	velar	laryngeal
voiced (implosive) stops	<b>ɓ-</b>	<b>ɗ-</b>		<b>f-</b>		
voiceless aspirated stops	<b>p<sup>h</sup>-</b>	<b>t<sup>h</sup>-</b>	<b>t<sup>ʂh</sup>-</b>		<b>k<sup>h</sup>-</b>	
voiceless unaspirated stops	<b>p-</b>	<b>t-</b>	<b>t<sup>ʂ</sup>-</b>	<b>c-</b>	<b>k-</b>	<b>ʔ-</b>
fricatives		<b>s-</b>			<b>ɣ-</b>	<b>h-</b>
nasal sonorants	<b>m-</b>	<b>n-</b>		<b>ɲ-</b>	<b>ŋ-</b>	
oral sonorants	<b>v-</b>	<b>l-</b>	<b>r/z<sup>ʰ</sup>-</b>	<b>j/j-</b>		

Table 1. 24 major-initial consonants.<sup>1</sup>

<sup>1</sup> Note: **r/z<sup>ʰ</sup>** and **j/j** represent freely varying allophones.

		bilabial	alveolar	palatal	(post) velar
stops	checked	<b>-p<sup>ʰ</sup></b>	<b>-t<sup>ʰ</sup></b>	<b>-c<sup>ʰ</sup></b>	<b>-k<sup>ʰ</sup></b>
	nasal	<b>-m</b>	<b>-n</b>	<b>-ɲ</b>	<b>-ŋ</b>
sonorants	oral	<b>-w</b>	<b>-l -r</b>	<b>-j</b>	<b>-V</b>

Table 2. Final segments (13), phonologically (V=vowel).

Contrasts in terminance (a three-way distinction for the final oral sonorants, and a two-way distinction for the final nasal sonorants):

- (1) (a) **t<sup>ʰ</sup>ɔːj** 'tail'                      **t<sup>ʰ</sup>ɔːj<sup>ʔ</sup>** 'bowl'                      **t<sup>ʰ</sup>ɔːj** 'follow'  
 (b) **z<sup>ʰ</sup>ɔː** 'turtle'                      **z<sup>ʰ</sup>ɔː<sup>ʔ</sup>** 'pig basket'                      **z<sup>ʰ</sup>ɔː** 'dry'  
 (c) **caməɪl** 'shiny'                      **ʔuməɪl<sup>ʔ</sup>** 'to hunt'                      **ɔɪl** 'to bounce'  
 (d) **kavər** 'stir'                      **kavər<sup>ʔ</sup>** 'embrace'                      **tər** 'to run out of workspace'  
 (e) **carɛ:w** 'green'                      **sarɛ:w<sup>ʔ</sup>** 'to raise/feed'                      **ɓlɛ:w** 'four-eyed turtle'  
 (f) **j<sup>ʰ</sup>ɔːm** 'sugar cane'                      **j<sup>ʰ</sup>ɔːm<sup>ʔ</sup>** 'to cry'  
 (g) **kə:n** 'oversize'                      **kə:n<sup>ʔ</sup>** 'to hunt by night'  
 (h) **ɓuːɲ** 'dust'                      **puːɲ<sup>ʔ</sup>** 'tree sp. (*licuala* sp.)'  
 (i) **cɔɲ** 'tree sp.'                      **cɔɲ<sup>ʔ</sup>** 'salty'

		bilabial	alveolar	palatal	(post) velar	
checked		<b>-p<sup>ʰ</sup></b>	<b>-t<sup>ʰ</sup></b>	<b>-c<sup>ʰ</sup></b>	<b>-k<sup>ʰ</sup></b>	← stops
		<b>-m<sup>ʔ</sup></b>	<b>-n<sup>ʔ</sup></b>	<b>-ɲ<sup>ʔ</sup></b>	<b>-ŋ<sup>ʔ</sup></b>	← nasals
		<b>-w<sup>ʔ</sup></b>	<b>-l<sup>ʔ</sup> -r<sup>ʔ</sup></b>	<b>-j<sup>ʔ</sup></b>	<b>-V<sup>ʔ</sup></b>	
voiced		<b>-m</b>	<b>-n</b>	<b>-ɲ</b>	<b>-ŋ</b>	← oral sonorants
		<b>-w</b>	<b>-l -r</b>	<b>-j</b>	<b>-V</b>	
voiceless		<b>-w̥</b>	<b>-l̥ -r̥</b>	<b>-j̥</b>	<b>-V̥</b>	

Table 3. Possible rime-endings (27 = 13 checked, 9 voiced, 5 voiceless).

(a) *Distribution of terminance distinctions across the three classes of final segment.*

	Final stops	Final nasals	Final oral sonorants
Realized with Checked Termination	✓	✓	✓
Realized with Voiced Termination	✗	✓	✓
Realized with Voiceless Termination	✗	✗	✓

### Vowels

	Front		Back
High	ɨ:	ɨ:	u:
	ɛ̃:	ɛ̃:	ɔ̃:
	ɨ:	ɨ:	u:
	ɛ̃:		ɔ̃:
	ɛ̃:	ɛ̃:	ɔ̃:
	ɛ̃:		ɔ̃:
Low		a: ɛ̃a:	ɔ̃:

Table 4. Long vowels, phonetically (18).

### (2) *Front vowels*

<b>pi:k</b>	'a tadpole'
<b>pɛ̃:t</b>	'to plant'
<b>vi:k</b>	'(to) work'
<b>savɛ̃:k</b>	'to comb back'
<b>pɛ̃:ŋ</b>	'to be dear'
<b>cavɛ̃:k</b>	'peel sth. apart'
<b>cavɛ̃:k</b>	'fish sp.'

### (3) *Back vowels*

<b>t̃lu:</b>	'bamboo sp.'
<b>c̃lɔ̃:</b>	'buffalo'
<b>kalu:</b>	'fruit sp.'
<b>klɔ̃:</b>	'star'
<b>kl̃ɔ̃:</b>	'fish sp.'
<b>culq:</b>	'palm sp. ( <i>licuala</i> sp.)'

(4) *Central vowels*

<b>c<sup>i</sup>cɿ:</b>	'now'
<b>h<sup>o</sup>ɿ:</b>	'broken'
<b>mlɿ:</b>	'afraid'
<b>murə:</b>	'disgusted'
<b>tu<sup>o</sup>ɿ:</b>	'weird'
<b>ka<sup>o</sup>ɿ:</b>	'bamboo sp.'

We can group these 18 vowels into 9 pairs, where the two members of each pair contrast in terms of phonation:

	Front		Back
High	<i>i</i> : <i>ɿ</i> :	<i>i</i> : <i>ɿ</i> :	<i>u</i> : <i>ɿ</i> :
	<i>ɿ</i> : <i>e</i> :	<i>ɿ</i> : <i>e</i> :	<i>ɿ</i> : <i>ɿ</i> :
	<i>ɿ</i> : <i>e</i> :	<i>ɿ</i> : <i>e</i> :	<i>ɿ</i> : <i>ɿ</i> :
Low	<i>a</i> : <i>ɿ</i> :	<i>a</i> : <i>ɿ</i> :	<i>a</i> : <i>ɿ</i> :

Table 1. Long vowels, showing vowels paired by register.

(1) *Some minimal pairs contrasting in register*

<b>c<sup>i</sup>:n<sup>?</sup></b> 'nine'	<b>c<sup>o</sup>ɿ:n<sup>?</sup></b> 'cooked'
<b>kar<sup>o</sup>ɿ:ŋ</b> 'sand'	<b>kar<sup>o</sup>ɿ:ŋ</b> 'sunshine'
<b>mat<sup>o</sup>ɿ:m</b> 'son-in-law'	<b>kata:m</b> 'crab'
<b>km<sup>o</sup>ɿ:<sup>?</sup></b> 'thumb'	<b>km<sup>o</sup>ɿ:<sup>?</sup></b> 'female (chicken)'
<b>z<sup>o</sup>ɿ:<sup>?</sup></b> 'clear'	<b>z<sup>o</sup>ɿ:<sup>?</sup></b> 'pig basket'
<b>z<sup>o</sup>u:<sup>?</sup></b> 'fence'	<b>z<sup>o</sup>u:<sup>?</sup></b> 'to know'

Emerging from the array in Table 1 is an orderly pattern of nine heavy-light pairs, giving 18 vowels in all. The pattern is recognizable as the 9-position system widely found in languages of the mainland Southeast Asia area (cf. Enfield 2005:182-4):

		Front	Central	Back
High	<i>Heavy</i>	<b>ɪ:</b>	<b>ɨ:</b>	<b>ʉ:</b>
	<i>Light</i>	<b>°ɪ:</b>	<b>°ɨ:</b>	<b>°ʉ:</b>
Mid	<i>Heavy</i>	<b>ɪ:</b>	<b>ɨ:</b>	<b>ʉ:</b>
	<i>Light</i>	<b>°e:</b>	<b>°e:</b>	<b>°ɔ:</b>
Low	<i>Heavy</i>	<b>ɛ:</b>	<b>°a:</b>	<b>°ɔ:</b>
	<i>Light</i>	<b>ɛ:</b>	<b>a:</b>	<b>ɔ:</b>

Table 6. Long vowels, laid out in a standardized 9-place system.

		Front	Central	Back
	<i>Heavy</i>	<b>ɪ</b>	<b>ɨ</b>	<b>ʉ</b>
Non-Low		<b>°ɪ</b>		<b>ɔ</b>
	<i>Light</i>		<b>°e</b>	
Low	<i>Heavy</i>	<b>ɛ.</b>	<b>°a</b>	<b>°ɔ</b>
	<i>Light</i>	<b>ɛ</b>	<b>a</b>	<b>ɔ</b>

Table 7. Short vowels.

Front	Central	Back
<b>ia</b>	<b>ia</b>	<b>ua</b>

Table 8. Diphthongs.

The minor syllable vowels constitute a radically reduced system, with the following properties:

- only three vowels occur: **-a-**, **-i-**, **-u-** (along with **-r-**)
- there is no vowel length distinction (always short/unstressed)
- there is no independent register distinction

While all words have major syllables, minor syllables are relatively infrequent. Around a third of all words have a minor syllable (998/2778=36%). Of these, many do not have a vowel, but an **-r-** (203/998=20%). Of those minor syllables which do have a vowel, most have **-a-** (562/795=71%), while the other two are significantly less frequent (**-u-** 177/795=22%, **-i-** 56/795=7%).

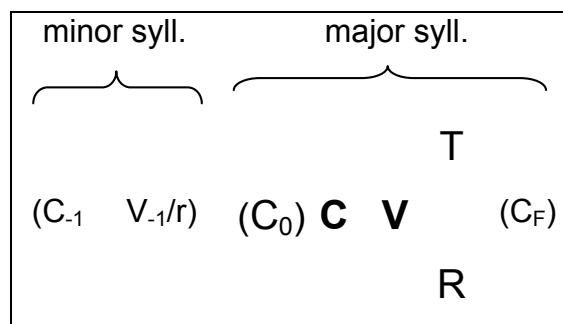
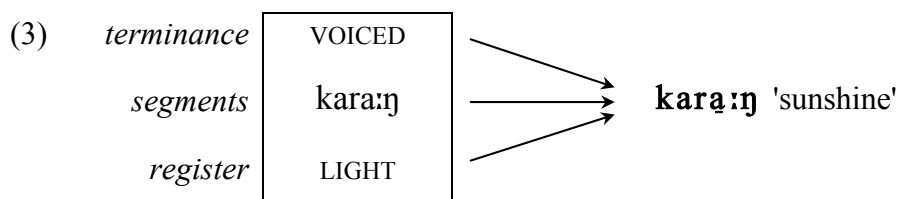
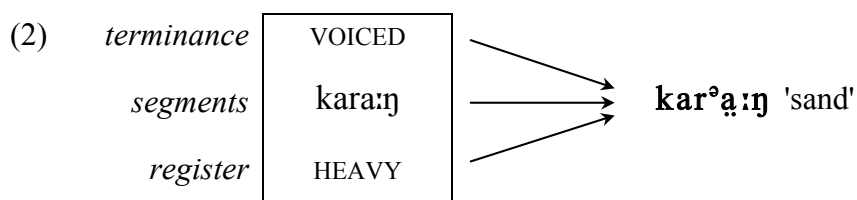
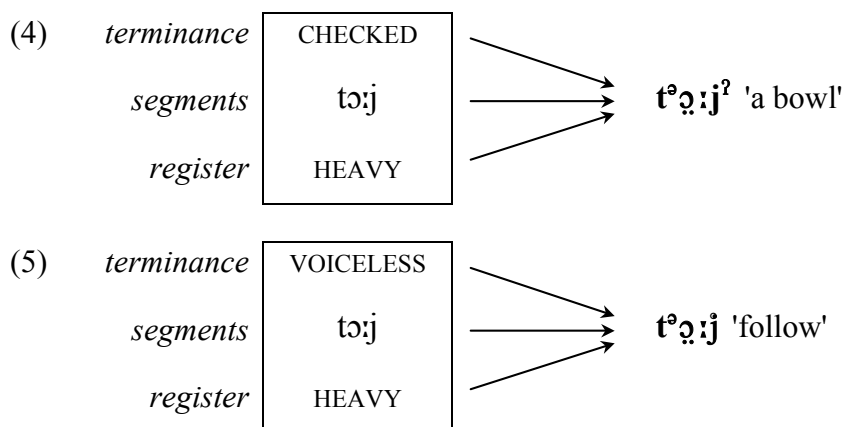


Figure 2. Word template (C=consonant, V=vowel, R=register, T=terminance, F=final)

The surface realization of each word is a product of (a) a string of segments, (b) a terminance value, (c) a register value. Here are a few examples, making this explicit:





The intersection of register (with two values) and terminance (with three values) yields six distinct rime types in Kri:

		register	
		<i>heavy</i>	<i>light</i>
terminance	<i>checked</i>	HEAVY-CHECKED	LIGHT-CHECKED
	<i>voiced</i>	HEAVY-VOICED	LIGHT-VOICED
	<i>voiceless</i>	HEAVY-VOICELESS	LIGHT-VOICELESS

Figure 3: The six Kri rime types, defined by intersections of terminance and register distinctions.

		old initials	
		<i>voiced</i>	<i>voiceless</i>
old finals	<i>stopped</i>	*bak > pək 'nang'	*pak > pək 'sắc'
	<i>open</i>	*ba > pà 'huyền'	*pa > pa 'ngang'
	<i>voiceless fricative</i>	*bas > pã 'ngã'	*pas > pã 'hỏi'

Figure 4: Vietnamese tone categories.

Register can lead to multiplication of vowel distinctions instead of emergence of tones (cf. the Kri vowel mutation, that could send it in either direction) Khmer has a large system of vowels (based on Headley 1998:24 – for lexical contrasts, cf. Huffman 1970):

i	i:	ɨ	ɨ:	u	u:			
e	ɛ:	ə	ə:	o	ɔ:			
	ɛ̄:				ɔ̄:			
	ɛ:			ɔ	ɔ:			
a	a:			ɑ	ɑ:			
ie	ae	ao	aə	ea	iə	uə	oa	wə

Since former stages of the language (Old Khmer, Headley 1998:23), many of the diphthongs have appeared, and certain vowels. Here are some vowel changes which have taken place in Khmer, ultimately deriving from an erstwhile register system (from Headley 1998:24):

Original vowel nucleus	High register <i>(ex-voiceless initial, ex-clear)</i>	Low register <i>(ex-voiced initial, ex-breathy)</i>
a'	u/wə	ɑ
a	ɔ:	ɑ:
ā'	ea/oa	a
ā	ie	a:
i	ɨ	ə
ī	i:	əy
ɨ	ɨ	ə
ī	ɨ:	əi
u	u	o
ū	u:	ɔ:
œ	ə:	aə
e	ɛ:	ɛ:
ē	ɛ:	ae
o	ɔ:	ao

### A continuum of systems:

tones & few Vs --- register & no tones, more Vs --- no register/tones, even more Vs

### Discussion points:

1. Is 'tonality' an areal phenomenon?
2. What are the virtues of going for more vowels in a register system? (If Kri has 30 vowels, does Lao have 90+?; i.e 9x2x5)

## Typology -

MSEA languages are at the extreme of what Sapir called *isolating* (in terms of ‘technique’, where the number of morphemes per word is approaching one) and *analytic* (in terms of style of ‘synthesis’).

Those languages that always identify the word with the radical element would be set off as an ‘**isolating**’ group against such as either affix modifying elements (affixing languages) or possess the power to change the significance of the radical elements by internal changes (reduplication; vocalic and consonantal change; changes in quantity, stress, and pitch)... An **analytic** language is one that either does not combine concepts into single words at all (Chinese) or does so economically (English, French). In an analytic language the sentence is always of prime importance, the word is of minor interest. (p.128)

### *Two distinct clines*

1. morpheme-to-word ratio: analytic-to-synthetic
2. ‘fusion’ of morphemes in word: agglutinating-to-fusional

In Sapir’s famous characterization of morphological type, he used Vietnamese (what he termed Annamese) as the paradigm case of a language in which words are not morphologically complex. Many people still think this. It’s probably true that these languages are as far as you get on that scale, but they do not approach the logical extreme – i.e., all words are one morpheme. Some people realize there is significant morphological complexity, along the same lines as Chinese (i.e., compounding), but there is a domain of morphological complexity that is unappreciated in MSEA. This is what we may call elaborative reduplication. It is found in Tai languages like Lao, as well as Austroasiatic languages like Vietnamese and Semelai. (Another type of morphological complexity is the derivation morphology found in languages like Khmer – though unproductive – and Semelai and Kmhmu – fairly productive – and Kri – somewhere in between.) The real generalization that can be made is that there is no ‘obligatory inflectional’ morphology, specifically no case marking, no person marking, and no tense marking.

An important point to make is that – pace Bisang among many others – it is not the case that MSEA languages demand strict constituent order. (The presumption that they do is, I think, based on the presumption that it *must be*, since there is no head-marking or dependent-marking to rely on.) But a look at data as it actually occurs tells a different story. The problem is that descriptions do not give us this kind of data. (See Enfield 2008 on ‘case relations’ in Lao). From Enfield 2007: A widely presumed answer is that hearers of languages like Lao are forced to rely on strict constituent order to maintain informational coherence in predicate-argument relations. This claim is, however, weak at best, since extensive ellipsis and movement create widespread surface ambiguity, and without compromising communication. Accordingly, for Mandarin, Li and Thompson (1981:26) state that ‘no basic word order can be established’. Similarly, on Riau Indonesian, Gil (2005b) shows that there are ‘no distinctions between major syntactic categories’. Gil argues that observed word order patterns in the language are epiphenomenal (cf. also LaPolla 1993). (See texts.)

## Derivational morphology

Thompson p150ff, on ‘specializing derivatives’ in Vietnamese, which employ what he calls ‘chameleon affixes’. There are subtypes of *chameleon affix*, which ‘have some kind of consistent phonetic resemblance to their bases’.

**Perfect** chameleon affixes ‘have exactly the same phonetic shape as the bases with which they occur’ (p139).

**Tonal** chameleon affixes ‘differ from their bases only in tone’ (p139).

**Riming** chameleon affixes retain the rime of the base (p140).

**Alliterative** chameleon affixes retain the initial of the base (p140).

Thompson goes in great detail through patterns and regularities of formation of these ‘affixes’. In section 7.4, from p150 onwards, he goes through an amazing variety of types of reduplication.

Note that when tones differ from one syllable to the next in a derivative, the relationship is mostly regular, and can be expressed in terms of historical tone relationships. As Thompson explains, there are ‘two registers and three types’ (cf. the Kri case). ‘In the most common patterns the tone of the affix is in the same register as that of its base’. (p140).

	Even	Short	Long
<b>Register</b>			
First	<i>ngang</i>	<i>sắc</i>	<i>hỏi</i>
Second	<i>huyền</i>	<i>nặng</i>	<i>ngã</i>

### Tone relationships in derivatives

**Specializing derivatives.** These are productive, all involving ‘perfect chameleon affixes’, four types (151-153).

1. Distributives: based on nouns.

(a) *nhà-nhà* ‘all the houses’ [*nhà* = ‘house’]

2. Iteratives: based on verbs.

(a) *làm-làm* ‘continually working, unfinished’ [*làm* = ‘work’]

3. Attenuatives: based on stative verbs.

(a) *đẹp-đẹp* ‘somewhat pretty’ [*đẹp* = ‘pretty’]

4. Intensives: formed from range of bases.

(a) *mau-mau* ‘very rapidly’ [*mau* = ‘rapidly’]

(b) *vụn-vụn* ‘in very small pieces’ [*vụn* = ‘in small pieces’]

(c) *đời-đời* ‘eternity’ [*đời* = ‘life, period of existence’]

**Emphatics** are described from p154. These are formed ‘with extremely diverse affixes’. (Compare this way of talking about the forms – as ‘affixes’ – with the kind of approach taken by Kruspe in her description of Semelai; her approach is to give ‘formulae’ for constructing the appropriate output; cf. also the Enfield grammar of Lao.)

1. Emphatics with tonal affixes (p156).

- (a) *bư-bư* ‘be very big’ [*bư* = ‘be big’]
- (b) *xốp-xốp* ‘be very spongy’ [*xốp* = ‘be spongy’]

2. Emphatics with vocalic affixes (p157); mostly ‘suffixing’ (i.e., infixing in the second syllable).

- (a) *hóc-hác* ‘be emaciated, gaunt’ [*hóc* = ‘hole, hollow’]
- (b) *rỗi-rãi* ‘have leisure time’ [*rỗi* = ‘be unoccupied’]
- (c) *vùng-vằng* ‘be shaking with anger’ [*vùng* = ‘shake’]
- (d) *múp-múp* ‘be fat, plump’ [*múp* = ‘be fat’]

3. Emphatics with riming prefixes (p158); most common are *b-* and *l-*.

- (a) *bàu-nhàu* ‘be very wrinkled’ [*nhàu* = ‘be wrinkled’]
- (b) *bối-rối* ‘be troubled, perplexed, upset’ [*rối* = ‘be confused’]
- (c) *lò-mò* ‘to grope feverishly’ [*mò* = ‘grobe for’]
- (d) *tò-mò* ‘to be curious, want to know all’ [*mò* = ‘grobe for’]

4. Emphatics with riming suffixes (p158-9).

- (a) *tham-lam* ‘be greedy, covetous’ [*tham* = ‘be greedy’]
- (b) *tạp-nhạp* ‘be mixed’ [*tạp* = ‘be mixed, miscellaneous’]
- (c) *cười-nhười* ‘to tease, joke’ [*cười* = ‘to laugh’]

5. Emphatics with alliterative prefixes (p159-60).

- (a) *rắc-rối* ‘be complicated, intricate’ [*rối* = ‘be be mixed up, tangled’]
- (b) *lại-láng* ‘overflow’ [*láng* = ‘abound’]
- (c) *đắn-đo* ‘weigh both sides of a question’ [*đo* = ‘to measure’]

6. Emphatics with alliterative suffixes (p160ff). These are ‘by far the most numerous and diverse’.

- (a) *nôm-na* ‘popular language’ [*nôm* = ‘demotic script’]
- (b) *xấu-xa* ‘be wicked’ [*xấu* = ‘be bad, ugly’] (over 20 exx with *-a*)
- (c) *đất-đai* ‘territory’ [*đất* = ‘land’]
- (d) *rầm-rập* ‘be noisy’ [*rầm* = ‘make a heavy noise’]

Thompson provides 8 pages of examples (over 300 examples), exemplifying dozens of different ‘suffixes’. These run the gamut of ‘productive’ (e.g., *-a* with over 20 examples given) to ad hoc (many cases have only one or two forms, or where the ‘source’ morpheme is marked as having ‘severely limited distribution’). He notes some ‘problems’ from p168, such as the occurrence of unanalyzable cases (i.e., where the two-syllable structure is itself a ‘single

morpheme’), the occurrence of compounds (where both syllables have independent occurrence with similar/related meanings – e.g., *béo-bở* ‘be advantageous’; *béo* ‘be fat’, *bở* ‘be profitable’), and so-called phrases (i.e. where the two elements are from different parts of speech in a construction). The favoured account from among various possibilities is this (p178): ‘it may be argued that at an earlier stage of the language there was an elaborate morphological system, involving many types of chameleon affixes with more or less consistent meanings. These may or may not have been related historically to a variety of onomatopoeic and other single-morpheme forms of reduplicative type.’

Compare this system (or its *description*) to what is described for Lao (Enfield) and Semelai (Kruspe) to show (1) how similar they are (though of course differing in details), and (2) how different their descriptions can be.

### Expressive reduplication in Lao.

#### Original Form

*cɔk5* ‘drinking glass’  
*khuq1* ‘bucket’  
*pum4* ‘book’  
*toq2* ‘table’  
*kɔng4* ‘camera’  
*pɔɔ3* ‘hessian’  
*toong3-tooj3* ‘expressive phrase’  
*kua3* ‘salt’  
*hua4* ‘fence’  
*cia4* ‘paper’  
*tip2* ‘rattan rice container’  
*piik5* ‘wing’  
*pet2* ‘duck’  
*pheet4* ‘sex’  
*sek1* ‘cheque’  
*dɛk5* ‘eat (vulgar)’  
*kiip5* ‘Kip’ (Lao currency)  
*nam4* ‘water’  
*nangsuuu3* ‘books’  
  
*hoong2-luaj1* ‘timber mill’  
*thuaj4* ‘bowl’  
*nam4-saa2* ‘tea’  
*nangsuuu3 kaj1-khia1* ‘bad handwriting’

#### Reduplicated Form

*cɔk5 cɛk5*  
*khuq1 khiq1*  
*pum4 pəm4*  
*toq2 teq2*  
*kɔng4 kɛng4*  
*pɔɔ3 pɛɛ3*  
*toong3-tooj3 toong3-təaj3*  
*kua3 kia3*  
*hua4 hia4*  
*cia4 cəq2*  
*tip2 təp2*  
*piik5 pəək5*  
*pet2 pət2*  
*pheet4 phəət4*  
*sek1 sək1*  
*dɛk5 dək5*  
*kiip5 kəəp5* or *kiip5 kaap5*  
*nam4 naj2* or *nam4 thaal*  
*nangsian5 nangsuu3* or  
*nangsɔk5 nangùù3*  
*hoong2-lɔk5 hoong2-luaj1*  
*thɔk5 thuaj5*  
*nam4 sɔk5 nam4 saa2*  
*nangsuuu3 kaj1-khɔk5 kaj1-*  
*khia1*

## Ideophones/expressives

Most Southeast Asian languages have a class of ideophones or ‘expressives’ (Diffloth’s term), which are verb-like, or adverbial, with their own syntactic behaviour. For example, they are often very limited in the kinds of TAM marking (if any) they can take. Expressives often have very specific semantic value, associated with modification of very particular property concepts (e.g., colours, states) or events. Tai and Mon-Khmer languages have the most developed classes of expressives. (Note that this is one of many areas in which MSEA and West African languages pattern alike). Expressives are often simply separate lexical items (usually two syllables, with some ‘rhyming’ pattern), which function as adverbials, often with highly evocative meanings:

- (a) *dam3 pii4-pii4*  
black expr  
‘totally black’
- (b) *laan4 qùù4-khùù4*      (c) *laan4 khim1-mim1*  
bald expr Lao      bald expr  
‘completely bald’      ‘lightly balding’ (‘with a frog by the pond’)
- (d) *laan4 sameng4-kheng4*      (e) *laan4 samook5-khook5*  
bald expr      bald expr  
‘bald across the whole dome’      ‘bald, with fuzz all over,  
(‘swidden in the jungle’)      and a protruding forehead and back of head’

‘standing’

- (a) *tòk2-pòk2*  
‘of old lady, “skinnily” standing, angular.’
- (b) *sêê1-lêê1*  
‘of someone standing, stooped over “all alone, no-one talking to them”.’

‘lying’

- (a) *khòò5-lòò5*      (b) *qêêk1-lêêk4*  
‘like a child asleep’      ‘lazily, not doing anything’

‘hair’

- (a) *keng4-ceng4*  
‘of messed up hair, “uncombed, not flat, going this way and that” ’
- (b) *phùk1-vùk1*      (c) *khuk1-ñuk1*  
‘of hair bent from being slept on’      ‘of curly, flowing, hair’

‘white’

- (a) *pen4-ven4*      (b) *còn1-phòn1*  
‘of face or body, pale,      ‘of completely white hair’  
“like too much powder applied” ’

‘hole/perforation’

- (a) *cingl-pingl*  
‘tiny, cannot put finger in’  
(b) *còngl-pòngl*  
‘of a lettuce, with holes in it from snails (can put a finger in)’  
(c) *công-pôôngl*      (d) *cangl-paangl*  
‘can put hand in’      ‘can fit head through’

other

- (d) *tuj4 quj-lujl*  
‘fat, tubby, of a baby’  
(e) *lèèn1 thi-lii1*  
‘running, of a 1-yo girl, to go swimming, “not interested in anything but getting to the destination”’

Common ‘default’ syntactic frame: het1 EXP juu1

Aslian languages (Kruspe 2004:396)

*Semelai*

- c<sup>h</sup>ɨjis*      ‘the smell of ammonia, urine’  
*crãlãp*      ‘the sound of something entering the undergrowth’

*Semai*

- cnayur*      ‘the appearance of bushiness in several places’  
*blbʔəl*      ‘the feeling of painful embarrassment’

*Temiar*

- kəraləg ləg*      ‘the sound of heavy footsteps’  
*gengerlət*      ‘spindly-ness’

Chrau (Thomas 1971 grammar, p155ff)

- hũl hũl*      -      ‘sitting quietly’  
*khũch khũch*      -      ‘many fish’  
*khyong khyong*      -      ‘walking stiffly’  
*klə klə*      -      ‘darkly, unclear’  
*lăq lăq*      -      ‘sitting still, sick’
- caprăh caprăng*      -      ‘scattered’  
*ravênh rawai*      -      ‘dizzy’  
*mbăq mban*      -      ‘unskillfully’  
*mlăq mlăq*      -      ‘dirty’

‘The judicious use of reduplication is a mark of an expert story teller.’ (p155)

**Four-syllable expressions**

- (a) *khaw5-paa3-qaahaan3*  
Lao rice-fish-food  
'foodstuffs'
- (b) *kêp2-phak2-hak2-nò01*  
Lao collect-vegetables-snap.off-shoots  
'Gather edible vegetables and shoots.'
- (c) *khiaw3-sot2-ngot1-ngaam2*  
Lao green-fresh-EUPH-beautiful  
'luscious, fresh and green' (as the Lao countryside in the rainy season)

Here are some examples from Dong (Long and Zheng 1998:162-3):

- (a) *ʔen<sup>53</sup> ta<sup>55</sup> ʔen<sup>53</sup> neŋ<sup>55</sup>*  
Dg colourful eye colourful nose  
'face is very dirty'
- (b) *weŋ<sup>212</sup> kau<sup>323</sup> weŋ<sup>212</sup> kha<sup>35</sup>*  
Dg horizontal head horizontal ear  
'do not listen to reason'

Books of these are popular in China ('cheng-yu'). Liem has listed over 800 of these in Vietnamese:

'[V-OBJ]-[V-OBJ]': Liem (1970:9) -

- (a) *cấm chợ ngăn sông*  
forbid market block river  
'to be isolationist'
- (b) *buôn mây bán gió*  
buy cloud sell wind  
'engage in risky commerce'
- (c) *cày mây cuốc gió*  
plough cloud hoe wind  
'have a free life in the open air'

'[N-ADJ]-[N-ADJ]': Liem (1970:21):

- (a) *ma thiêng nước độc*  
ghost powerful water poison  
'of the highlands'
- (b) *mặt to tai lớn*  
face big ear long  
'with power, influence'
- (c) *minh gầy mặt võ*  
body thin face exhausted  
'emaciated'

### Sentence-final particles.

The mainland Southeast Asian languages have a well developed class of sentence-final particles, such as these examples (Clark 1989:182):

- (a) *Kuv tsis hnov koj hais os*  
 Hm I not hear you say pcl(emph)  
 'I didn't hear what you said!' (Clark 1989:182)
- (b) *Koj ua dabtsi lawm os ob peb hnub no na*  
 Hm you do what pfv pcl(emph) two three day this pcl(eh?)  
 'What have you been doing these few days?'
- (c) *Ngày mai anh có đi không? đi chứ.*  
 Vn tomorrow bro. have go pcl(q) go pcl(sure)  
 'Are you going tomorrow? Of course I'm going.'

Cantonese has a developed system of particles (*two nice studies Kwok 1984 Luke 1990*, and good discussion in Matthews and Yip 1994:338-356).

- (d) *husb: léih mhóu joi máaih yéh la*  
 Ca you don't again buy stuff pcl  
 'Don't you go buying more stuff.'
- wife: ngóh tái-háh jē ma*  
 I look-del pcl(playing:down) pcl(of course)  
 'I'm just going to take a look.' (Matthews and Yip 1994:344)
- (e) *làahmjái tek-bō tühngmáaih léuihjái tek-bō*  
 Ca boy kick-ball and girl kick-ball  
*móuh māt fānbiht ge je bo*  
 not:have any difference pcl(assert) pcl pcl(invite agrmnt)  
 'Boys play football and girls play football - what's the difference?'  
 (Matthews and Yip 1994:345)

Kwok (1984) gives 30 basic forms, and 74 possible combinations of two or three particles (much lower than the theoretically possible set of combinations of her thirty basic particles.)

Matthews and Yip (1994:338):

- functions:
  - a. indicating speech-act types (q, assertion, request)
  - b. evidentiality
  - c. affective and emotional colouring
- they are very common and important - colloquial
- often correspond to intonation in English and other lgs (sometimes said to compensate for lack of freedom to intone, due to lexical tone)
- problems deciding on whether they belong in semantics, pragmatics, or discourse
- sometimes non-arbitrary relationship to tone (some have claimed that there are 'tonal particles' (has been claimed for Mandarin which has only about 7 particles, which can't stack); many of the Cantonese particles have 'tonal variants'; 'Typically, the high-tone variants are more tentative, the low-tone ones assertive and the mid-tone ones neutral'.

This is an area in need of attention, particularly in the domain of meaning.

### Topic-prominence

- (h) *Sakana-wa tai-ga oisii*  
 Jpn fish-TOPIC red snapper-SUBJ delicious  
 ‘(As for) Fish, red snapper is delicious.’
- (i) *Pihengki-nin 747-ka khi-ta*  
 Krn aeroplane-TOPIC 747-SUBJ big  
 ‘(As for) aeroplanes, the 747 is big.’
- (j) *dà-xiàng bí-zi hěn cháng*  
 Man elephant nose very long  
 ‘Elephants, (their) noses are very long.’
- (k) *nèi-kè shù yè-zi dà*  
 Man that tree leaf big  
 ‘That tree, (its) leaves are big.’
- (l) *làng1 nii4 caw4 bò1 tòng4 pên3-huang1*  
 Lao matter this you don’t have-to worry  
 ‘This matter, you don’t have to worry (about it).’
- (m) *saang4 dang3 ñaaw2 laaj3*  
 Lao elephant nose long very  
 ‘Elephants, (their) noses are very long.’

In a verb-final language (cf. Li & Thompson 1976:472, Fuller 1985):

- (a) *làthyu nya ánà khù-ǎ*  
 Lisu people TOP dog bite-DECL  
 i. ‘People, dogs bite (them).’  
 ii. ‘People, (they) bite dogs.’
- (b) *ánà nya làthyu khù-ǎ*  
 Lisu dog TOP people bite-DECL  
 i. ‘Dogs, people bite (them).’  
 ii. ‘Dogs, (they) bite people.’

### Argument structure

Consider these examples from Khmu (Suwilai 1987:60-61). Suwilai writes that A-V-O is ‘the normal order of transitive clauses’. She adds, however, that ‘for emphasis it is common for the Object to occur clause initial’, resulting in O-A-V order being ‘also common’:

- (a) *rpà:ng nò: tám*  
 Km gong they beat  
 ‘The gong, they beat (it).’
- (b) *kó:n màq nò:k*  
 Km child mother beat  
 ‘The child was beaten by the mother.’

She adds, that the ‘reversed’ order — O-V-A — is used ‘for special emphasis’. This is an area for further research – i.e., what does ‘special emphasis’ really mean. Corpus studies will help.

(c) *sqð:ng péng qòq*

Km tree cut.down 1

‘I cut down a tree.’

(d) *réq ìøm qòq*

Km farm clear 1

‘I clear the land for crops.’

Thompson (1987:217) on Vietnamese:

‘An important way in which Vietnamese verbs differ from English verbs is that they do not in themselves imply a clear notion of “voice” in the grammatical sense.’

(1) (a) *Làm việc này rồi.*

do work this already

‘[We’ve] already done this.’

(b) *Việc này làm rồi.*

work this do already

‘This has already been done.’

(2) (a) *Ông ấy cất nhà ở gần trường-học.*

sir that build house be.at near school

‘He’s building a house near the school.’

(b) *Hai cái nhà kia cất gần nhau cùng trên một miếng đất.*

two clf house that build near rep tgthr on one clf lot

‘Those two houses are built close together on the same lot.’

(3) (a) *Anh ấy dùng một chữ ít lễ-phép.*

fellow that use one word little polite

‘He used a word that was not very polite.’

(b) *Chữ này không dùng một-mình.*

word this neg use one-self

‘This word is not used by itself.’

### ‘Passives’:

(a) *Sohput khan hlũn*

WCh friend tell 1sg.loresp

‘A friend told me.’ (Baumgartner 1998:5)

(b) *Hlũn djauk sohput khan*

WCh 1sg.loresp ‘passive’ friend tell

‘I was told by a friend.’ (Baumgartner 1998:5)

*Djauk* (before or after the sentential subject) also means ‘must’:

(c) *Yah rôk nao ðôh drăp kau, hư djauk ta bray ni*

WCh if dig away see thing 1.hiresp 2.loresp must only give this

*ka kau wok*

to 1.hiresp back

‘If you go dig it up and find my things, you have to give it back to me.’ (1998:13)

- (d) *Rani djauk hr bray pagöl tanuh.ea ni mai ka*  
 WCh now must 2.loresp give hand.over territory this toward to  
*kau wok*  
 1.hiresp back  
 ‘Now you must hand over this territory back to me.’

In Khmer, Lao, and Thai (and other languages), the ‘passive’ marker is derived from a verb to ‘strike’, and this verb (or other ‘strike’ verbs) are also used to mark these ‘passive’ constructions. (‘Acquire’ verbs may also be used.)

On Khmer, from Huffman (1970:302):

- (1) *trəw haəy*  
 correct pfv  
 ‘That’s right/correct.’
- (2) *comlaəy look m̄in trəw tee*  
 answer 2hon neg correct neg  
 ‘Your answer isn’t correct.’
- (3) *kñom trəw t̄iw psaa thngay-nih*  
 1 must go market day-this  
 ‘I have to go to the market today.’
- (4) *koun trəw sliəq-peəḍq qaoy sqaat-baat*  
 child must dress give/cause clean/tidy  
 ‘You must dress neatly, child.’
- (5) *wiə crəluəh mək trəw cəngkuəng wĩñ*  
 3 slip come contact knee return  
 ‘It slipped and (came and) hit my knee.’
- (6) *p̄ii-msəl-məñ kñom trəw ckae kham*  
 yesterday 1 contact/undergo/suffer dog bit  
 ‘Yesterday I was bitten by a dog.’ (I was subjected to a dog biting.)

Cf. Hmong (fieldnotes):

- (1) *Kuv tua raug phiaj*  
 1 shoot strike target  
 ‘I hit the target.’
- (2) *Nws raug mus MisKuj, (tabsis mus t̄sis tau)*  
 3 strike go M. but go neg can  
 ‘S/he got to go to America.’
- (3) *Nws raug tub ceevwm ntes.*  
 3 strike clf police catch  
 ‘S/he got caught by the police.’

The functions of *passive* are achieved often by other means: ellipsis, freedom of pragmatically-determined argument movement, and versatility in verbs.

## Verb Serialization

Data from Halǎng:

Regarding the next three Halǎng examples, Cooper says ‘Verbs in this relationship share the same subject and object’ (Cooper 1966:33):

- (a) *Ger bǔh oak cha ka*  
Hal he roast clean eat fish  
‘He roasted, cleaned, and ate the fish.’ (Cooper 1966:33)
- (b) *Pataw ruup bǔh 'maang pahaang giaw' àih*  
Hal rich-man capturepound beat roast cook-on-a-stick you  
‘The rich man will capture, pound, beat, roast, and cook you.’  
(Cooper 1966:33)
- (c) *Kataam yok cha ger*  
Hal crab get eat her  
‘The crab caught and ate her.’ (Cooper 1966:33)

Purposive/directional:

- (d) *Hèey chiw hùum*  
Hal we go bathe  
‘We go and bathe.’ (Cooper 1966:33)
- (e) *Aw lah chong pang ih*  
Hal I come eat with you  
‘I come and eat with you.’ (Cooper 1966:33)
- (f) *'ni chiw dak cha*  
Hal you go ascend eat  
‘You go up to eat.’ (Cooper 1966:33)

And here is some data from Dong, a distributive string, and a purposive (transcription adjusted for convenience; example from Long and Zheng 1998:161):

- (a) *qat<sup>323</sup> nang<sup>13</sup> sang<sup>31</sup> pa<sup>55</sup>*  
Dg cut grass raise fish  
‘cut grass, raise fish’
- (b) *lau<sup>323</sup> jan<sup>212</sup> paī<sup>55</sup> səm<sup>33</sup> na<sup>212</sup>*  
Dg enter house go look.for you  
‘went into the house to look for you’

## Disposal constructions

Wangmo Bouyei (Zhou 2000:444):

- (1) *muŋ<sup>2</sup> ?au<sup>1</sup> va:n<sup>1</sup> ?jau<sup>4</sup> ?di<sup>1</sup>*  
WB you take ax hide good  
‘You hide the ax well.’

## Zhenfeng Bouyei

- (2) *ku<sup>1</sup> pa<sup>4</sup> pa<sup>3</sup>tu<sup>3</sup> yu<sup>7</sup> ?di<sup>1</sup> leu<sup>4</sup>*  
 ZB I - door close well pcl  
 'I have closed the door already.'

These are the same constructions required for three-place predicates:

- (3) *mu<sup>2</sup> ?au<sup>1</sup> θa:n<sup>1</sup> ɬo<sup>5</sup> la<sup>3</sup>, ?au<sup>1</sup> mian<sup>6</sup> ɬo<sup>5</sup> ku<sup>2</sup>*  
 WB you take husked.rice put below, take flour put above  
 'Please put the husked rice underneath, and the flour on the top.' (Zhou 2000:447)
- (4) *te<sup>1</sup>fi<sup>6</sup> ðiu<sup>3</sup> tsuo<sup>2</sup>ŋi<sup>2</sup> tɕiao<sup>1</sup> hau<sup>32</sup> ku<sup>2</sup>*  
 ZB s/he not carry homework hand.in give me  
 'S/he hasn't given her homework to me.' (Zhou 2000:445)

&Hmong:

- (a) *Kuv yuav muab qho no povtseg*  
 I irr take thing this dispose:of  
 'I will dispose of this thing.'

**Polar questions**

- (6) (a) *kin<sup>3</sup> qaahaan<sup>3</sup> laaw<sup>2</sup> pen<sup>3</sup> bò<sup>3</sup>*  
 Lao eat food Lao can q  
 'Can (you) eat Lao food?'  
 (b) *rwep còmnuen krwep baaj baan têê*  
 Kh count amount grain rice can q  
 'Can (you) count the amount of rice grains?'  
 (c) *Ge wak bia bwan rw*  
 Km 3.m.sg drink beer can q  
 'Can he drink beer?'  
 (d) *Anh nă thit chó được không*  
 Vn 2 eat flesh dog can q  
 'Can you eat dog meat?'
- (7) (a) *Nws tshuab qeej keej*  
 Hm 3 blow pipe good.at  
 'S/he is good at playing the qeej.'  
 (b) *Nws tshuab qeej puas keej?*  
 3 blow pipe q good.at  
 'Is s/he good at playing the qeej?'
- (8) *Ông có biết nói tiếng tây không?*  
 Vn sir have know speak language West neg  
 'Do you (know how to) speak French [i.e. 'Western Language']?'

- (9) *tā qù*  
 MSC 3 go  
 ‘S/he goes/will go.’
- (10) *tā qù-bu-qù*  
 MSC 3 go-neg-go  
 ‘Is s/he going or not?’

Here are some p-not-p questions from Dong (under influence of Sinitic) from Long and Zheng 1998:177, transcription altered for convenience):

- (a) *tui<sup>55</sup> nai<sup>33</sup> khwan<sup>35</sup> kwe<sup>212</sup> khwan<sup>35</sup>*  
 Dg plum this sweet neg sweet  
 ‘Is this plum sweet or not?’
- (b) *na<sup>212</sup> wo<sup>31</sup> kwe<sup>212</sup> wo<sup>31</sup> wa<sup>33</sup> kəm<sup>55</sup>*  
 Dg 2sg know neg know speak Dong  
 ‘Do you speak Dong or not?’
- (c) *na<sup>212</sup> wo<sup>31</sup> wa<sup>33</sup> kəm<sup>55</sup> kwe<sup>212</sup> wo<sup>31</sup>*  
 Dg 2sg know speak Dong neg know  
 ‘Do you speak Dong or not?’

Here, an MSC example illustrates the ‘V-neg-V O’ pattern:

- (11) *wǒ mài wǒ de zì-xíng-chē*  
 MSC 1 sell 1 poss bicycle  
 ‘I sold/am selling my bicycle.’
- (12) *nǐ mài-bu-mài nǐ de zì-xíng-chē*  
 MSC 2 sell-neg-sell 2 poss bicycle  
 ‘Are you selling your bicycle (or not)?’

### Numeral classifier constructions

In MSEA languages, a common trait of members of the noun class is their participation in classifier constructions:

- (a) *hkwei hnă kauñ*  
 Bu dog two animal  
 ‘two dogs’ (Okell 1969:209)
- (b) *lot sohng khan*  
 Thai car two handle  
 ‘two cars’
- (c) *neaknipon pii (ru:p)*  
 Kh writer two form  
 ‘two writers’ (Capell 1979:13)
- (d) *yim \*(tus) menyuam*  
 Hm eight clf child  
 ‘eight children’ (Clark 1989:183)
- (e) *léuhng jek gáu*  
 Ca two clf dog  
 ‘two dogs’ (Matthews and Yip 1994:93)

A typical set is from Khmu (Suwilai 1987:34, slightly adapted transcription):

- CLF Used with reference to
1. kló:ng solid, round objects (e.g. fruit, vegetables, sun, moon, house)
  2. trl`əm long, flat objects (e.g. leaves, piece of meat, hammer, comb, bag)
  3. tó: animals
  4. p<sup>h</sup>é:n piece of cloth, shirt, trousers
  5. sé:n flexible, long, narrow objects (e.g. rope)
  6. kòn human beings
  7. smtúm bunch of fruit or ‘other things’
  8. plàh one of a pair of human organs
  9. cùm one of a pair of things such as human organs
  10. krlòh mouthful of rice
  11. bò:n piece of land
  12. kmlòq piece of meat
  13. lém oblong object (e.g. piece of wood, finger, drinking water container)

	<b>Numeral classifiers</b>	<b>Modifier classifiers</b>			<b>Class terms</b>	<b>Kin prefixes</b>
		<i>a. unitizing</i>	<i>b. demonstrative</i>	<i>c. modified</i>		
<b>Functional context</b>	counting	unitizing	situational or discourse deixis, anaphora	unitizing, hosting, modification	lexically derivational	social deixis for personal reference
<b>Grammatical context</b>	NUM __	__ NOM	__ DEM	_ ADJ/ REL.CLS	__ MOD	__ NAME
<b>Phonological status</b>	stressed	u n s t r e s s e d , d e p e n d e n t				
<b>Semantic properties</b>	fine semantic contrast, encoding shape/form, function; variable assignment; object of speaker awareness	only refer to shape/form of unit	semantic specificity unnecessary in these functions; distinctions often neutralized (using just <i>too</i> <sup>3</sup> ‘body’ and <i>qan</i> <sup>3</sup> ‘small thing’), and use of more specific terms in these contexts is pragmatically marked		refer to taxonomic ‘essence’	refer to ‘higher’ kinship relation, mark expression of respect
<b>Approximate size of set</b>	100+	20+	2–100+		60+	10+

Table 1. Summary of systems of nominal classification in Lao.

## Pronominal elaboration

Khmer forms from Huffman (1970:356-7):

Person	Form	Comments
1	<i>kñom</i> <i>qañ</i> <i>yeeng</i> <i>knie</i> <i>kñom-prjèh-baat</i> <i>kñom-kenaa</i> <i>qaatmaa</i>	General, polite; literal meaning ‘slave’. Between intimate friends, superior to inferior, or adult to child; otherwise insulting. Singular between equals in some dialects; otherwise plural. Singular, among intimates. Inferior to superior, respectful or extremely formal. Layman to priest; inferior to superior of exalted rank; commoner to king. Priest to layman.
2	<i>look</i> <i>look-srey</i> <i>njèq</i> <i>njèq-srey</i> <i>qaeng</i> <i>qaa-qaeng</i> <i>prjèh-dac-prjèh-kun</i>	Masculine, formal, polite; inferior to superior; between equals of relatively high status. Feminine, formal, polite; inferior to superior; between married women of relatively high status. Superior to inferior; older to younger; between equals of relatively low status; to one’s mother. Feminine, polite; superior to inferior; between equals. Reflexive connotation; between intimates; superior to inferior; otherwise insulting; usually paired reciprocally with <i>qañ</i> ‘I’. Derogatory, good-natured insult between friends; adult to child. Inferior to superior of exalted rank; layman to priest.
3	<i>kwet</i> <i>kêê</i> <i>vie</i>	Respectful; younger of older; inferior of superior. Informal; other; indefinite (generic). Superior of inferior; adult of child; otherwise insulting. (‘It’.)

This sort of pattern is found in Thai and Lao. Vietnamese and surrounding languages also show full use of kinship terms as pronouns. Here are Kri pronouns (not including kin terms):

	SG		DU		PL		
	BARE	POLITE					
		F	M	INCL	EXCL	INCL	EXCL
1	<i>teeq/pàànq</i>	<i>koon</i>		<i>saa</i>	<i>ñaar</i>	<i>cawq</i>	<i>caa</i>
2	<i>cak</i>	<i>mii</i>		<i>maar</i>		<i>prii</i>	
3	<i>hanq</i>	<i>mooq</i>	<i>qôông</i>	<i>qaar</i>		<i>paa</i>	

The pronouns can be used in inclusory constructions, of the type PRONOUN+[TITLE+NAME], for referring to groups of people identified with a sub-member who is mentioned by name:

- (13) *paa vòòk nôn*  
3pl grandpa N  
‘they of whom Grandpa Non is one’
- (14) *qaar qaj kham*  
3du non.resp.pref K  
‘those two of which Kham is one’
- (15) *ñaar saaw knin*  
1du.excl aunt K  
‘we two (excl) of which the other is Aunt Knin.’

Kin category of addressee	1st person pronoun	2nd person pr	Can use name?
1. <i>cià-maangq</i> 'kin of same descent group' (marriage not allowed)	<i>teeq</i> 1SG	<i>cak</i> 2SG	Yes
2. <i>mree</i> "'higher" <i>cià-maangq</i> of your spouse'	<i>ñaar</i> 1DU.EX	kin term	No
3. <i>matààm/qujùù</i> 'so. married to your lower <i>cià-maangq</i> (e.g. younger sib. or child)'	<i>pààng</i> 1SG	<i>maar</i> 2DU	No
4. <i>kmoon</i> "'lower" <i>cià-maangq</i> of your spouse'	<i>koon</i> 1SG	<i>mii</i> 2SG	Yes
5. <i>tààm/sawq</i> 'someone married to your elder sibling'	<i>teeq</i> 1SG	kin term	No
6. <i>sdoong</i> ' <i>mree</i> of your own child' (i.e. someone whose child you are <i>mree</i> to) [these pronouns used reciprocally]	<i>khoojq</i> 1SG	<i>nôôq</i> 2SG	No

Some kin categories, consequential for person reference. The distinction between 'higher' and 'lower' is grounded in sibling relations; those related to you via elder siblings (e.g. your father's older brother) are 'higher', and *mutatis mutandis* for younger siblings. Note: For higher (i.e., older) *cià-maangq* you can't use *cak* as 2sg pronoun.

Here is a list of basic Kri kin terms:<sup>2</sup>

F	-	<i>pòòq</i>
M	-	<i>mêêq</i>
eB	-	<i>maangq</i>
eZ	-	<i>cìiq</i>
yG	-	<i>qeem</i>
FyZ	-	<i>qoo</i>
FyB/HyB	-	<i>pòòq</i>
MyB/WyB	-	<i>kùùq</i>
MyZ	-	<i>murq</i>
FeZ/HeZ	-	<i>jaa</i>
FeB	-	<i>puu</i>
MeZ/WeZ	-	<i>naaj</i>
MeB/WeB	-	<i>taa</i>
PP	-	<i>vòòk</i>
PPP	-	<i>kôôq</i>
CC, etc.	-	<i>cuuq</i>

<sup>2</sup> Abbreviations are: F=father, M=mother, B=brother, Z=sister, G=sibling, H=husband, W=wife, C=child, y=younger, e=elder.

## Word Order Typology

Mon-Khmer and Tai are generally left-headed, while Chinese, Tibeto-Burman, and Hmong-Mien exhibit a mix of left- and right-headed structures. Languages to the North and West are generally right-headed (Japanese, Mongolian, Uralic, Indo-Iranian, Dravidian).

### Features of WO Typology in Southeast Asian Languages

#### 1. Relative position of verb and object

##### 1.1. V-O

- (a) *dèng3 kin3 khaw5 lèw4*  
Lao D. eat rice finish  
'Deng has eaten.'
- (b) *mae: hên khon nan*  
Thai mother see person that  
'Mother saw that person.'
- (c) *kñom sɔm'aat ptéah*  
Kh I clean house  
'I clean the house.' (Ehrman 1972:29)
- (d) *Tus dev tom tus npua*  
Hm clf dog bite clf pig  
'The dog bit the pig.' (MC 1985:178)
- (e) *kéuih ngoi ngóh*  
Ca 3p love I  
'S/he loves me.' (Matthews & Yip 1994:68)
- (f) *mẹ yu con*  
Vn mother love child  
'I love you.' (said by mother to child)
- (g) *təngaj trənau kau jɔ' təlaat*  
Ng tomorrow I go market  
'Tomorrow I will go to the market.' (Field notes)
- (h) *tse<sup>2</sup> mɔ<sup>6</sup> chi<sup>1</sup> ffa<sup>1</sup>*  
Mu sister he comb hair  
'His sister was combing her hair.' (Wang & Zheng 1993:79)

##### 1.2. O-V:

- (a) *yô yè te chê ve*  
Lahu 3sg house make prog  
'He is building a house.' (Matisoff 1991:404)
- (b) *thu sa yei-pa-te*  
Bu he letter write-polite-V.S.  
'He wrote a letter.'

## 2. Relative position of ‘adjective’ and noun

### 2.1. N-A

- (a) *khon2 suung3*  
Lao person be.tall  
‘tall person’
- (b) *mənuh l'ɔɔ*  
Kh person beautiful  
‘beautiful person’ (Capell 1979:5)
- (c) *rapyā kariār*  
Pal girl beautiful  
‘the beautiful girl’ (Milne 1921:38)
- (d) *hœ' hnòk*  
Mon house large  
‘a large house’ (Bauer 1982:333)
- (e) *nhà nhỏ*  
Vn house small  
‘a small house’ (Thompson 1965:221)
- (f) *tsən<sup>1</sup> ga:i<sup>3</sup>*  
Mu hair long  
‘long hair’ (Wang & Zheng 1993:82)
- (g) *ye è*  
Bu water cold  
‘cold water’ (Thein-Tun 1995:173) (where ‘cold’ is a stative verb)

### 2.2. A-N

- (a) *dākyi ge gáu-jái*  
Ca cute LP puppy  
‘cute puppy’ (Matthews and Yip 1994:88)
- (b) *səp<sup>8</sup> at<sup>7</sup> hən<sup>3</sup> lo<sup>4</sup> kɔ ku<sup>3</sup>*  
Mu ten clf very big pcl drum  
‘ten very big drums’ (Wang & Zheng 1993:87)
- (c) *tai'-eiñ*  
Bu brick-house  
‘brick house’ (Okell 1969:68) (nominal attribution)

### 2.3. Some apparent exceptions.

- (a) *θau' ye*  
Bu drink water  
‘drinking water’ (Thein-Tun 1995:173) (‘drink’ is a ‘factive verb’)
- (b) *khaaw phat*  
Thai rice fry  
‘fried rice’
- (c) *khua5 khaw5*  
Lao fry rice  
‘fried rice’ (‘rice fry’)

Cf. case of Lao verbs *khua5*, *laap4*, *som5*, *mok1*, *ping4*, etc. These are transitive verbs, and also work as nouns (i.e. fit into the frame *I'm going to eat \_ tonight*). And note that *khaw5 khua5* is something else in Lao.

Cf. Matthews and Yip (1994:49) on Cantonese:

‘Some compound nouns like food and animal terms have their head nouns in the initial position (see Y-S. Cheung 1969):

- |     |                  |     |                       |
|-----|------------------|-----|-----------------------|
| (a) | <i>yùh-sāang</i> | (b) | <i>gyū-yuhk-gōn</i>   |
| Ca  | fish-raw         |     | pig-meat-dry          |
|     | ‘raw fish’       |     | ‘pork jerky’          |
| (c) | <i>choi-gōn</i>  | (d) | <i>ngàuh-yuhk-gōn</i> |
| Ca  | vegetable-dry    |     | cow-meat-dry          |
|     | ‘dry vegetable’  |     | ‘beef jerky’          |
| (a) | <i>gāi-gūng</i>  | (b) | <i>gāi-lá</i>         |
| Ca  | chicken-male     |     | chicken-female        |
|     | ‘rooster’        |     | ‘chicken’             |
| (c) | <i>gyū-gūng</i>  | (d) | <i>gyū-lá</i>         |
| Ca  | pig-male         |     | pig-female            |
|     | ‘male pig’       |     | ‘female pig’          |

(Cf. MSC *gōng-jī*)

### 3. Relative position of Noun and Relative Clause

#### 3.1. N-Rel

- |     |  |
|-----|--|
| (a) | <i>qan3 khòj5 hên3</i>                                     |
| Lao | clf I see  |
|     | ‘the thing I saw’  |
| (b) | <i>sach (mà) anh mua hôm-quyên ở đâu</i>                   |
| Vn  | book ALP you buy yesterday be.at where                     |
|     | ‘Where is the book you bought yesterday?’ (Capell 1979:10) |
| (c) | <i>siəwphəw dæl look msəl-məñ teñ nəu 'εε-naa</i>          |
| Kh  | book which you yesterday buy be.at where                   |
|     | ‘Where is the book you bought yesterday?’ (Capell 1979:10) |
| (d) | <i>ib tug dev tom npua</i>                                 |
| Hm  | one clf dog bite pig                                       |
|     | ‘a dog who bites pigs’ (Clark 1985:183)                    |
| (e) | <i>ngwa<sup>l</sup> lik<sup>8</sup> na:n<sup>4</sup></i>   |
| Mu  | dog chase deer   |
|     | ‘a (deer-)hunting dog’ (Wang & Zheng 1993:29)              |

## 3.2. Rel-N

- (a) *ngóh chéng ge gūngyàhn*  
 Ca I hire LP maid  
 ‘the maid I hire’ (Matthews & Yip 1994:88)
- (b) *mə-`hma-te’ `sheshəya*  
 Bu not-mistake-makes doctor  
 ‘a doctor who makes no mistakes’ (Capell 1979:11)
- (c) *ṅa:u<sup>3</sup> twa<sup>3</sup> jəu<sup>6</sup> fong<sup>1</sup> jəu<sup>6</sup> kyang<sup>1</sup> kə ga:n<sup>2</sup> ngwa<sup>4</sup>*  
 Mu one clf both tall and bright pcl house tile  
 ‘a tiled house which is both tall and bright’ (Wang & Zheng 1993:87)

3.3. Note the mixed data from Mulao, a Kadai language spoken in Guangxi. More examples:

- (a) *a:n<sup>1</sup> ma<sup>4</sup> / ma<sup>4</sup> a:n<sup>1</sup>*  
 Mu saddle horse horse saddle  
 ‘saddle’ (Wang & Zheng 1993:32)  
 (cf. Mandarin *mǎ`ān*, Lao *qaan1 maa4*)
- (b) *fɪ<sup>1</sup> tshja<sup>1</sup>*  
 Mu fire vehicle  
 ‘train’ (Wang & Zheng 1993:31)  
 (cf. Mandarin *hǔo chē* [fire vehicle], Lao *lot1 faj2* [vehicle fire])
- (c) *fā<sup>5</sup> tɔŋ<sup>1</sup> jep<sup>8</sup>*  
 Mu leaf dumpling leaf  
 ‘reed leaves for making pyramid-shaped glutinous-rice dumpling’  
 (Wang & Zheng 1993:31)  
 (cf. Mandarin *tuan ye`?*)
- (d) *pua:n<sup>6</sup> tshə<sup>5</sup> sjen<sup>4</sup>*  
 Mu thread machine thread  
 ‘thread used for sewing machine’ (Wang & Zheng 1993:31)  
 (cf. Mandarin *jī xiàn*)

## 4. Relative position of Possessor (R) and Possessed (D)

## 4.1. D-R

- (a) *hùan2 (khòòng3) khòòj5*  
 Lao house of I  
 ‘my house’
- (b) *ptěah (rəbɔh) kñom*  
 Kh house of I  
 ‘my house’
- (c) *nhà (cua) tòi*  
 Vn house of I  
 ‘my house’
- (d) *pu<sup>4</sup> niu<sup>2</sup>*  
 Mu father I  
 ‘my father’ (Wang & Zheng 1993:59)

## 4.2. R-D

- (a) *koj (lub) tsev*  
Hm you clf house  
'your house' (Clark 1985:185)
- (b) *ngóh sailóu*  
Ca I YBro  
'my younger brother' (Matthews & Yip 1994:107)
- (c) *hohksāang ge gājéung*  
Ca studentposs parents  
'the student's parents' (Matthews & Yip 1994:107)
- (d) *ñá<sup>2</sup> təm<sup>1</sup> tau<sup>2</sup>*  
Mu you chest  
'your chest' (Wang & Zheng 1993:62)
- (e) *uù tháñ-yé ǎmyiñ*  
Bu U-poss view  
'U Thant's view' (Okell 1969:169)

Cf. Clark and Prasithratsint (1985:66-71) on Hmong possession being somewhat different to the other languages, in its use of classifiers to indicate alienability. Sometimes the syntactic frame can disambiguate polysemous nominals, such as for example *txĩ* 'husband'/'father':

- (a) *kũ txĩ*  
Hm 1 -  
'my father'
- (b) *kũ tù txĩ*  
Hm 1 clf -  
'my husband'

## 5. Relative order of Adposition and Noun

## 5.1. Preposition-N

- (a) *man<sup>2</sup> kin<sup>3</sup> khaw<sup>5</sup> juu<sup>1</sup> theng<sup>2</sup> phuu<sup>2</sup>*  
Lao 3p eat rice be.at top mountain  
'S/he eats rice on top of the mountain.'
- (b) *Nws cog zaub ntsuab rau hauv vaj*  
Hm 3p plant vegetable green to/at inside garden  
'She planted green leafy vegetables in the garden' (Clark 1989:188)
- (c) *kñom caq mcul knong sac*  
Kh I insert needle inside flesh  
'I stuck a needle in her skin.' (Clark 1989:188)
- (d) *Chị ngồi trên ghế*  
Viet sister sit top chair  
'You sit on the chair.' (Clark 1989:189)
- (e) *suĩ<sup>53</sup> nau<sup>33</sup> te<sup>23</sup> mət<sup>31</sup> ta<sup>33</sup>*  
Dg sit at below tree that  
'Sit underneath that tree.' (Long and Zheng 1998:141, adjusted)

## 5.2. N-Postposition

- (a) *léih deui hàaih hái jēung tói hahbihn*  
 Ca you pair shoes be.at clf table underneath  
 ‘Your shoes are under the table.’ (Matthews & Yip 1994:117)
- (b) *kéuih hái jáudim chēutbihn dāng ngóh*  
 Ca 3p be.at hotel outside wait I  
 ‘S/he is waiting for me outside the hotel.’ (Matthews & Yip 1994:118)  
 (Chinese combines ‘coverbs’ with ‘localizers’)
- (c) *chò kà’*  
 Lahu here loc  
 ‘here, hither, hence.’ (Matisoff 1973:164)

## 6. Relative Order of Adjective and Standard of Comparison

## 6.1. Adj:SComp

- (a) *suung<sup>3</sup>kual caw<sup>4</sup>*  
 Lao tall surpass you  
 ‘taller than you’
- (b) *fong<sup>1</sup> ta<sup>6</sup> ña<sup>2</sup>*  
 Mu tall over you  
 ‘taller than you’ (Wang & Zheng 1993:56)
- (c) *cāu hơn anh*  
 Vn tall more.than you  
 ‘taller than you’
- (d) *A-wàhn leng gwo kéuih mùihmúi*  
 Ca W pretty than 3p sister  
 ‘Wan is prettier than her sister.’
- (e) *niəng nih l'ɔɔ ciəng niəng nuh*  
 Kh girl this pretty surpass girl that  
 ‘This girl is prettier than that girl.’

(The comparative marker is often derived from a verb meaning ‘surpass’ or ‘exceed’ - Ansaldo 1999; cf. some African languages.)

## 6.2. SComp:Adj

- (a) *A-wàhn béi kéuih mùihmúi leng*  
 Ca W than 3p sister pretty  
 ‘Wan is prettier than her sister.’ (formal: Matthews & Yip 1993:167)
- (b) *xiǎo-Wǎng bǐ tā měimei piàoliang*  
 MSC little-W. than 3p sister pretty  
 ‘Wang is prettier than her sister.’
- (c) *mau<sup>33</sup> pi<sup>323</sup> na<sup>212</sup> phang<sup>35</sup>*  
 Dg 3sg compare 2sg tall  
 ‘He is taller than you.’ (Long and Zheng 1998:141, adjusted)
- (d) *ngà = a-ké N V*  
 Lahu I=more:than n v  
 ‘n is more v than me’ (Matisoff 1973:131)

(e) *mì-kṭ̄ chi N a-cí dà' ve*  
Lahu chair this n more good pcl  
'This chair is better than *n*.' (Matisoff 1973:131)

**Kri text examples**

(050801a)

03.41

A *nèèw teeq hòòm ruuq mee kuu qan ni*  
 thing 1SG not yet know make PURP eat TPC  
 ‘The things I don’t know how to make to eat.’

A *liiw siiw kii qa-*  
 plant sp. and HES  
 ‘*Li iw si iw* and um-’

03.46

A *liiw siiw kudu u*  
 plant sp. plant sp.  
 ‘*Li iw si iw, ku du u.*’

03.47

B *poo-haj tuqèè qa tangq qiin tà*  
 enough-nice INDEF HES bitter say PCL  
 ‘It’s just something (like) um bitter, so to say.’

A *ñaraaq, hòòm mee, teeq naaq*  
 plant sp. not yet make 1SG TPC.EXT  
 ‘(and) *ñaraaq*, (I) haven’t made (them), I.’

03.49

A *mee kuu qan*  
 make PURP eat  
 ‘Made them to eat.’

A *qor paa nèèw*  
 still three thing  
 ‘Still three things (I haven’t made).’

03.54

C *teeq teeq laa qan*  
 1SG 1SG NEG eat  
 ‘I, I didn’t eat (those).’

03.55

A *ñaraaq teeq màk kwaa qaboo teeq*  
 plant sp. 1SG like much PCL 1SG  
 ‘*Ñaraaq* I really like, I (do).’

03.57

C *teeq hoo (boo mìrùq sùùj)*  
 1SG TPC.LNK PCL MyZ S.  
 ‘I (Aunty Sùùj).’

03.58

A *teeq dêêh ruuq mee kuu teeq tahôôt toom nan*  
 1SG NEG know make PURP 1SG steam put cooked.rice  
 ‘I don’t know how to make them so as to steam them with rice.’

03.59

B *màk qa paj mee mêêng paj baah ckhanq*  
 like HES PRED make careless PRED vomit really  
 ‘They would tend to prepare it carelessly and then they’d really vomit.’

04.02

C *teeq ñaraaq teeq laa ngààj qan*  
 1SG plant sp. 1SG NEG NOT.YET eat  
 ‘Me, ñaraaq, I haven’t eaten (it) yet.’

C *teeq hoo*  
 1SG TPC.LNK  
 ‘I (haven’t).’

B *qaaj*  
 INTJ  
 ‘Oh.’

04.04

A *coo qan lèèwq qeet, cak dêêh ñooq*  
 really eat PFV PCL 2SG NEG remember  
 ‘(You) really have eaten (it), you don’t remember.’

04.05

A *cnam kooq taa ckaa naaq tôôq*  
 year live LOC C. TPC.EXT PCL  
 ‘The year (we) lived at Ckaa there.’

A *ma dêêh kooq quqêê ma dêêh qan*  
 PRED NEG there.is INDEF.HUM PRED NEG eat  
 ‘There was nobody who didn’t eat it.’

04.08

A *pòòq pr- qoo naaq naaq kii qaa qôông pèèng*  
 father pr- FyZ TPC.EXT TPC.EXT and HES MR. P.  
 ‘The father of- you there, and um Mr. Pèèng.’

04.11

A *kii maangq paa thàj kalq hanq dikêêq coon qaj*  
 and eB 3PL people cut.down 3SG woman give title-  
 ‘and (your) brother, those who were cutting the tree down, the women, (those guys) had miss-.’

A *côôm qaj nêêt koon*

C. title N. 1SG  
‘Côôm and Miss Nêêt of mine.’

A *qa pòòq mi naaq qiin ñùàk qa caa*

father 2SG TPC.EXT say tease NP 1pl.excl  
‘Your father there was teasing us.’

04.17

A *cuuq prmaajq qa nù pàr bulii seeh ma piin pulêêq naaq*

group maiden NP TPC roll cigarette TPC.FAR PRED give fruit TPC.EXT  
‘(You) girls here, roll cigarettes (for us) and (we) will give the fruit (to you).’

A *dêêh ci dêêh piin*

NEG PRED NEG give  
‘Otherwise (we) won’t give (you the fruit).’

04.21

A *koon ci koon laa haan koon ma pàr*

1SG PRED 1SG NEG dare 1SG PRED roll  
‘I didn’t dare to roll (cigarettes).’

A *ma coon, mlòò qôông màk*

PRED give afraid Mr. LIKE  
‘to give (them), (I was) afraid he would like (me).’

04.25

A *khàm muu-muuq naaq, khàm qor mnee*

because naïve TPC.EXT because still child  
‘Because (I was) naïve, because (I was) still a child.’

04.29

A *pù- pù mòòr pù pàr pàr qôông qiin*

like like INDEF like roll roll Mr. say  
‘Like- How was it, it was like ‘Roll, roll!’ he said.’

A *bangkhap qaj côôm qaj nêêt pàr coon qôông qiin*

force title C. title N. roll give Mr. say  
‘(He) forced Côôm and Nêêt, “Roll (cigarettes) for (me)!” he said.’

04.33

A *qòòr qòòr, coo engalq bulii prmaajq, qiin*

INTJ INTJ really delicious cigarette maiden say  
‘Uh-huh, uh-huh, really delicious, cigarettes of maidens, (he) said.’

A *qôông pèèng qa naaq toom*

Mr. P. NP TPC.EXT put  
‘Mr. P. there said.’

04.37

A *phaangq qor lêêq moo van qa naaq*  
 time still take Mrs. V. NP TPC.EXT  
 ‘(It was) the time that (he) was still married to Mrs. Van then.’

04.39

A *teeq mlòrò teeq*  
 1SG afraid 1SG  
 ‘I was afraid (I was).’

**LNEKVDP13Jul04**

TC 00:17:21.010 - 00:17:22.160

A *qaaj, kunààq-kunààq*  
 INTJ long.ago-RDP  
 ‘Oh, long ago.’

TC 00:17:22.160 - 00:17:23.320

A *cààw meengq qunêêq kunààq*  
 speak language Vietnam long.ago  
 ‘Speaking Vietnamese, long ago.’

TC 00:17:23.320 - 00:17:24.370u

A *còrò teeq ruuq tuqèè*  
 NEG 1SG know INDEF  
 ‘I knew nothing.’

(0.62)

B &lt;laugh&gt;

(0.53)

TC 00:17:26.136 - 00:17:29.346e

A *qos̄c paañq kùr̄rq quu seeh ch̄ợ Jaa seeh*  
 go sell pig LOC DEM.ACR market J. DEM.ACR  
 ‘Going to sell pigs over there, at Jaa market over there.’

(5.2)

TC 00:17:34.546 - 00:17:36.536a

A *hanq haañ teeq, teeq laa ruuq tuqèè teeq ma toop*  
 3SG ask 1SG 1SG NEG know INDEF 1SG PRED respond  
 ‘They’d ask me, I’d know nothing, to respond.’

(3.45)

TC 00:17:39.986 - 00:17:41.146

A *người Lào không biết gì*  
 person Lao NEG know INDEF  
 ‘A Laotian, he knows nothing.’

(1.04)

TC 00:17:42.696 - 00:17:43.866

A *tùngq ko- kooq quu*  
 stand be- be.at LOC  
 ‘Standing there at-.’

(0.32)

TC 00:17:44.186 - 00:17:45.526

A *kòòh ngooq naboo*  
 door gateway PCL  
 ‘the gateway.’

(0.35)

TC 00:17:45.876 - 00:17:48.076

A *kòòh ngooq talaat ma mee ròòq ma mee coot tipeeh*  
 door gateway market PRED make fence PRED make meet PCL  
 ‘the gateway to the market, they made a fence which met up there.’

(1.33)

TC 00:17:49.406 - 00:17:50.216

A *chø Jaa qaboo*  
 market J PCL  
 ‘(at) Jaa market.’

TC 00:17:50.216 - 00:17:51.736

A *tèèk tuqèè, thôm-thôm pii qôôq*  
 play INDEF ONOM like bird  
 ‘No game, *thôm-thôm*, like birds.’

(0.34)

TC 00:17:52.076 - 00:17:53.196

A *pii kòòt tuk-taak nù*  
 like frog sp. TPC.PROX  
 ‘like *tuk-taak* frogs’

(1.01)

TC 00:17:54.206 - 00:17:55.026

B *mleeng naaq nòòq*  
 person TPC.EXT PCL  
 ‘The people, right?’

TC 00:17:54.996 - 00:17:55.736

A *mleeng naaq*  
 person TPC.EXT  
 ‘The people.’

(2.03)

TC 00:17:57.775 - 00:17:59.435

A *loon quu naaq ci ()*  
 enter LOC DEM.EXT PRED  
 ‘Entering there,’

TC 00:17:59.435 - 00:18:00.845

A *teeq sòòk qa nèw teeq mon turr*  
 1SG seek NP thing 1SG want acquire  
 ‘I looked for things I wanted’

TC 00:18:00.845 - 00:18:02.315

A *teeq dèèh ruuq lêq tuqèè*  
 1SG NEG know take INDEF  
 ‘I didn’t know what to take.’

(1.4)

TC 00:18:03.715 - 00:18:04.955

A *prak naaq, haar-*  
 money TPC.EXT two  
 ‘As for money (I had) two-’

TC 00:18:04.955 - 00:18:05.955

A *haar mòòj saamq phàn*  
 two ten eight thousand  
 ‘twenty eight thousand’

(0.37)

TC 00:18:06.330 - 00:18:07.820

A *qiin phàn ckhanq kunàq-kunàq*  
 say thousand really long.ago-RDP  
 ‘It was really referred to in thousands, back then.’

(3.75)

TC 00:18:11.570 - 00:18:13.730

A *braaq naaq, môôc lààjh naaq, haar phàn saamq*  
 machete TPC.EXT one tongue TPC.EXT two thousand eight  
 ‘As for machetes, for one, two thousand eight (hundred).’

(4.16)

TC 00:18:17.890 - 00:18:19.190

A *paj jòònq hanq ckhanq*  
 COP.COMP good 3SG really  
 ‘That would be a really good one.’

(7.68)

TC 00:18:26.870 - 00:18:29.530

A *ci paa caak paa () thàj mleeng ruuq naaq*  
 PRED 3PL buy 3PL people people know TPC.EXT  
 ‘They’d buy (things) and trade (things), those who knew.’

*paa caak qa-*  
 3PL buy HES  
 ‘They’d buy-’

(1.03)

TC 00:18:30.564 - 00:18:31.824

A *() bún ma qiin qa naaq*  
 b. PRED say NP TPC.EXT  
 ‘bún, it’s called’

(1.88)

TC 00:18:33.704 - 00:18:34.564

B *caak tuqèè qiin citii*  
 buy INDEF say PCL  
 ‘What would they buy?’

TC 00:18:34.524 - 00:18:35.124

A *bún*

b.

'*bún*'

(0.79)

TC 00:18:35.914 - 00:18:37.094

A *cicùr qiin khaw5 pun4 ckhanq*

now say rice P. really

'Now it's called *khaw5 pun4*.'

TC 00:18:37.094 - 00:18:37.974

A *manêêt naaq*

probably TPC.EXT

'It seems.'

(2.55)

TC 00:18:41.283 - 00:18:42.213

A *paa caak ka-*

3PL buy

'They'd buy-'

(0.17)

TC 00:18:42.383 - 00:18:43.103

A *sùt kùurq*

flesh pig

'Pork'

(0.68)

TC 00:18:43.783 - 00:18:44.953

A *caak kaah see*

buy all tea

'Along with tea'

(2.89)

TC 00:18:47.843 - 00:18:48.353

A *qan*

eat

'to eat'

(0.84)

TC 00:18:49.193 - 00:18:51.653

A *qooc bruuq qunêêq qan pii nii ckhanq qaboo teeq loo*

go forest Viet eat like DEM.PROX really PCL 1SG PCL

'Going to Vietnam, that's how I would eat, I.'

(0.04)

B &lt;laugh&gt;

TC 00:18:51.693 - 00:18:52.303

(6.05)

TC 00:18:58.361 - 00:18:59.601

A *qajààk kùùrq qʃʊc boo*  
 carry pig go PCL  
 ‘Carrying pigs there.’

TC 00:18:59.601 - 00:19:00.591

A *tèèk tuqèè*  
 play INDEF  
 ‘No game.’

(1.42)

TC 00:19:02.011 - 00:19:02.661

A *lam*  
 lie.down  
 ‘Sleeping’

TC 00:19:02.661 - 00:19:04.760

A *taa jaang naaq,*  
 LOC J. DEM.EXT  
 ‘In Yang there,’

*lôôh dêêwq jaang paj têêngq seeh paj-*  
 exit from J. PRED reach DEM.ACR PRED  
 ‘Leaving from Yang, to reach there it would be’

*tngòòp*  
 night  
 ‘Evening time.’

TABLE 1. Some characteristic areal features

A (for 'all')—the feature applies across the whole continent, not necessarily in every language, but in some languages from every region. In some cases a figure is given (e.g. c.80%) indicating the approximate proportion of languages the feature is found in.

R (for 'region')—the feature is found in one (or a few) continuous regions and is an areal feature for those regions.

D (for 'discontinuous')—the feature is found in a few languages or small groups of languages, spotted across the continent, rather than making up a solid geographical block.

## Phonology

- 1 A, c.98% A nasal corresponding to every stop.
- 2 A, 100% At least four places of articulation for stops and nasals (best specified in terms of active articulator): labial, dorsal, apical, and laminal.
- 3 R A contrast between apico-alveolar and apico-postalveolar (retroflex) stops and nasals (and laterals)—everywhere except in a strip down the east coast and some languages in groups NH, NBl, and X.
- 4 R A contrast between lamino-dental and lamino-palatal stops and nasals (and laterals)—in two large areas, one on the west coast and the other comprising an east-central block, plus a couple of small areas elsewhere (see the maps in Dixon (forthcoming)).
- 5 A, c.98% Two semivowels, *w* and *y*. (Note that two contiguous languages in groups NF and NG, and a geographically separate language in WH, each have three semivowels: dorsal-labial *w*, lamino-palatal *y*, and lamino-dental *yh*.)
- 6 A, c.85% Two rhotics (grooved-tongue sounds), one articulated further forward in the mouth (generally an apico-alveolar trill or tap) and one further back in the mouth (apico-postalveolar or semi-retroflex, generally a continuant, sometimes a trill).
- 7 D Three rhotics, in c.7 distinct geographical regions.
- 8 A, c.75% A single series of stops; and no fricatives.
- 9 D Contrastive series of obstruents (either fortis and lenis stops, or stop and fricative, or two series of stops plus fricatives), found in c.60 languages in c.16 distinct geographical regions.
- 10 A, c.80% Basic syllable structure CV(C), all words beginning in a single consonant (not with a vowel or a consonant cluster).
- 11 D Initial dropping (of C or CV from the beginning of a word) has taken place in about ten distinct geographical regions (sometimes in just one dialect of a language, but at other times over all the languages in a small diffusion area). This leads to atypical word structures, e.g. CCV(C) or V(C)CV(C) and to new paradigmatic distinctions, e.g. stop contrasts, fricative phonemes, additional vowel phonemes.
- 12 A, c.67% System of three vowels, *i*, *a*, and *u*. [All examples of different-size systems are the result of phonological change—just two vowels for some dialects in group WL (also suggested for NBd3, Anindhilyagwa) and more than three vowels in some languages of groups A, B, D, E, NB, ND, NG–NJ, NL (plus a few odd languages in J, M, T, U, WF, WM, NA, NE).]

Nouns

- 13 A, 90%+ Nouns take derivational suffixes between root and final inflection: typically, genitive, comitative, privative, dual, plural (may all be followed by case inflection).
- 14 A, c90% Nouns take final case inflections to distinguish core functions; generally ergative case for A, absolutive case (with zero realization) for S and O functions.
- 15 R Two groups of languages (in WH and NA) have shifted to a case system with nominative for S and A and accusative for O.
- 16 R Many languages with well-developed head-marking have lost case-marking of NPs in core functions—in WJ, NB, ND, NG, NI, NK, NL.
- 17 A, c.98% Nouns have case inflections for dative, purposive, and instrumental functions and also for local functions including locative, allative, and ablative.
- 18 A, c.85% Instrumental function marked by the same suffix as ergative (where there is an ergative suffix).
- 19 D If in a non-prefixing language instrumental is not the same as ergative, then it coincides with locative—in eight languages, scattered across the continent.

Pronouns

- 20 A, c.85% Singular/dual/plural number system in pronouns.
- 21 R Minimal/unit-augmented/augmented pronoun system, where the minimal terms are 1st person, 2nd person, and 1st-plus-2nd person ('me and you') plus, in some languages, 3rd person; unit-augmented involves one participant added to the minimal set and augmented more than one—in three areas: (i) NE in the north-west; (ii) B in the north-east; and (iii) some languages from WJ, NB, NH, NI and NL in the central north.
- 22 D Inclusive/exclusive distinction for 1st person dual and/or plural; found in about two-thirds of the languages with a singular/dual/plural system
- 23 A, c.75% Free and bound non-singular pronouns show a nominative (S and A)/accusative (O) case system.
- 24 D Singular free pronouns have different forms for each of the three core functions, S, A, and O; this is an archaic feature found in about eight geographical enclaves across the continent.
- 25 D Free pronouns follow an absolutive/ergative system, like nouns (all languages with this feature have bound pronouns, with nominative/accusative forms)—in a geographical block encompassing WI, WJ, some dialects of WD; plus scattered languages elsewhere (including P, W, WA/WB).

Verbs

- 26 A, c.98% Final suffixal inflection, indicating tense and/or aspect plus imperative mood.
- 27 A, c.80% Derivational suffixes between root and final inflection, generally including valency-decreasing derivation(s) (covering reciprocal and usually also reflexive).

- 28 D Some languages have developed either a reflexive/reciprocal pronoun, in place of verbal derivations; or just a reflexive pronoun, retaining a verbal suffix for reciprocal (but never the reverse, with reciprocal pronoun but reflexive derivational suffix to the verb)—these pronouns are all individual developments, in languages spotted around the continent.
- 29 R Prefixes to verbs (and sometimes also to nouns), always including bound pronominal prefix for S and A functions, and generally also one for O function; in most cases there is also a TAM prefix, often fused with the pronominal prefixes—in WMa; NB–NL.

## FORMS

## Lexemes

- 30 A *mayi* 'vegetable food'—in 17 of the 38 groups A–Y, WA–WM and in 6 of the 12 groups NA–NL.
- 31 A *dhalanj* 'tongue'—in 29 of the 38 groups A–Y, WA–WM and in 6 of the 12 groups NA–NL.
- 32 A *dirra*, *lirra*, *rirra* or *yirra* 'tooth' (sometimes extended to 'mouth')—in 23 of the 38 groups A–Y, WA–WM, and in NB and NH.
- 33 A *bu(-m)* 'hit'—in 28 of the 38 groups A–Y, WA–WM and in 7 of the 12 groups NA–NL.
- 34 A *na(-ŋ)* 'see'—in 4 of the 12 groups NA–NL and in W and X; and *nha(-ŋ)*—in 29 of the 36 groups A–V, Y, WA–WM.

## Case allomorphs

- 35 A Ergative allomorph related to *\*-lu*, on demonstratives, interrogatives, proper nouns, kin terms, generic nouns and pronouns; allomorph related to *\*-dhu*, on other nouns—one or both of these is found in about 80% of languages with ergative case marking (see Sands 1996).
- 36 R ergative allomorph *-ŋgu* (developed from *\*-dhu*, probably by different routes in different regions)—in (i) c.25 languages in WD, WE, WG–WM; (ii) in c.30 languages in B–K plus the adjoining Nd and W; (iii) in Mg1, Gumbaynggirr. (See the Appendix.)

## Pronoun forms

- 37 A 2sg *ŋinj-* in about half the languages in NA–NL; 2sg based on *\*ŋin* in c.95% of the languages in A–Y, WA–WM.
- 38 A 2n-sg *nu-* in c.70% of the languages in NA–NL (and in X); 2n-sg *nhu-* in c.60% of the languages in A–W, Y, WA–WM.
- R 1du(inc) *ŋali* in c.80% of the languages in A–Y, WA–WM, but in none from NA–NL. (See the Appendix.)

This is just a selection of the recurrent features in the Australian linguistic area, or in sub-areas within it. Others that could be added to the list include: lateral consonants, all words in a language ending in a vowel or all ending in a consonant, vowel length, glottal stop as a syllable prosody, stress placement, aversive case marking ('for fear of') on nouns, classifiers and noun classes, kin-determined pronouns, transitivity classes of verbs, nominal incorporation in verb stems, associated motion derivational suffixes to verbs ('done while going/coming', etc.), number and person neutralization in bound pronouns, interrogatives/indefinites, deictics, verbless and copula constructions, marking of negation, switch-reference marking, types of subordinate clauses. Under recurrent forms we could add nominal/verbal purposive suffix *-gu*, reflexive/reciprocal verbal suffix *\*-dharri-*, imperative *-ga*, several more personal pronoun forms and about 130 lexemes (over 60 verbs, about 60 nouns, and about 6 adjectives).

Table 24. Areal properties in the CB area (figures in the first column refer to the sections in the paper where the phenomenon is discussed)

Phenomenon	Languages primarily involved	Possible source(s)	Typological status	
			Globally	Europe
Pluralia tantum (4)	Slavic, Baltic, Finnic, Latvian Romani, Icelandic	Extension of Indo-European properties in Baltic and Slavic and their influence on the other languages of the region (each with its own sphere of influence)	Relatively unusual <sup>a</sup>	Relatively unusual <sup>a</sup>
Initial stress (5.1)	Finnic, Latvian, Germanic, (Northern Russian dialects)	Probably Finnic for Latvian; Uralic source for Germanic highly disputable	Frequent	A relatively high concentration in the CB area
Polytonicity: a. Tones on long syllables (5.2.1)	Baltic	Relict of a more wide-spread phenomenon in Indo-European dialects (also Slavic, Old Greek)	Probably relatively unusual. <sup>a</sup> Unclear whether the three phenomena are related to each other.	
b. Overlength (5.2.2)	Estonian, Livonian, Low Latvian, some Low German dialects?	Common innovation as a consequence of initial stress and reduction of non-initial syllables		
c. Phonologization of secondary stress (5.2.3)	Swedish, Norwegian, Danish, Scottish Gaelic	Innovation		
Case alternation for marking Total vs. Partial objects/subjects (6.3, 6.5)	Finnic, Baltic, Eastern and Western Slavic	In embryo attested in Indo-European; common innovation in the CB area with several layers of influence (Baltic -> Finnic -> Russian)	Probably unusual, but not unique (cf. Basque) <sup>a</sup>	
Nominative object (6.4, 6.5)	Finnic, Baltic, Northern Russian	Various hypotheses; in Baltic and Northern Russian; probably a combination of inherited Indo-European models reinforced by contacts with Finnic	Probably unusual, but not unique (cf. Nenets, Kamassian, Southern Paiute, Yindjibarndi) <sup>a</sup>	

Case alternation in nonverbal predication (7.1)	Finnic, Sami, Mordvin, Komi, Baltic, Eastern Slavic, Polish	Various hypotheses	Fairly infrequent, but far from unique. Occurs mainly at the fringe of Indo-European and is most probably a non-Indo-European characteristics	
Predicative possession not based on 'have'-verbs (7.2)	Finnic (and most of Uralic), Latvian, Eastern Slavic	For Latvian and Eastern Slavic, most probably, a combination of the inherited Indo-European model reinforced by contacts with Finnic	Very frequent, in particular in Eurasia.	Outside of the SAE area, where 'have'-verbs are a common innovation
Syncretism of instrumental/commitative (7.3)	Estonian, Livonian, Sami, Latvian, Germanic	Germanic influence (in Latvian, probably combined with language-internal factors)	In 25% of languages	A high concentration in the SAE area (a common innovation?)
Comparatives: "Separative" and "Goal"	Finnic, Slavic, Baltic	Retention of old Indo-European and Uralic models; various local influences (Baltic -> Eastern Slavic dialects, etc.)	"Separative" and "Goal" comparatives frequent, particularly in Eurasia; "Particle" comparatives relatively infrequent	The SAE area shows the cross-linguistically highest concentration of "Particle" comparatives
Comparatives involving particles (7.4)	Finnic, Slavic, Baltic, Germanic			
Zero-subject constructions: non-referential indefinite zero subjects;	Finnic, Latvian, Western Slavic	Finnic influence on Latvian; an innovation in Western Slavic	No data available	SAE allows no zero-subjects, but resorts to generic pronouns like <i>man/on</i>
generic zero subjects (7.5.1)	Baltic, Slavic; Veps, Votic	Slavic influence on Veps and Votic		
reflexive postfixes as markers of valence recession (7.5.3)	Northern Germanic, Baltic, Eastern Slavic	Various hypotheses	Fairly unusual	Only in the CB languages
adjective agreement (8.1)	Germanic, Baltic, Slavic, Finnic	Indo-European influence on Finnic	Fairly usual, but areally and genetically unevenly distributed	
alternation between case-government and agreement within numerical constructions (8.3)	Baltic, Slavic, Finnic, Sami	Expansion of the Indo-European model in Baltic and Slavic. Indo-European (most probably Baltic) influence on Finnic and Sami.	Very unusual, probably not unique (cf. Arabic) <sup>a</sup>	
flexible SVO (9.1)	Baltic, Slavic, Finnic, Komi, Northern and Eastern Sami	Word order flexibility is, probably, a retention of older Indo-European and Uralic properties in transition from SOV to SVO	Rare — 4%	21%, with a high concentration in the CB (and North-eastern European) area

SVO/GN (9.2)	Baltic, Finnic, Komi, Mordvin, Continental Scandinavian	Retention of the older Eurasian order within NPs in Uralic; probably a combination of inherited structures and Finnic influence in Baltic. Independent innovation in Continental Scandinavian	Frequent	Only in the CB (and North-eastern European) area
Mixed adpositional systems (9.3)	Finnic, Latvian	Probably a combination of internally motivated factors and Finnic influence for Latvian. Unclear for Finnic	Infrequent — 7.7%	Infrequent
<i>Yes/no</i> -questions: Particle-initial <i>yes/no</i> -questions	Baltic, Estonian, Livonian, Southern Sami, Yiddish, Polish, Ukrainian, Belarusian, Russian dialects; Swedish dialects in Finland	Particle-initial questions: common innovation of Baltic and some of the Slavic; Baltic influence on Estonian and Livonian; Slavic influence on Yiddish; indirect Finnish influence on Swedish. Various local lexical connections.	Particle-initial questions — 14%;	Particle-initial questions mainly on the periphery; SAE area shows the cross-linguistically highest concentration of languages with verb-fronting in questions
Verb-fronting (9.4)	Germanic, Estonian, Finnish, Russian	A shared SAE innovation	verb fronting — 8.9%	
Evidential mood (10.2)	Baltic, Southern Finnic (Estonian, Livonian)	A shared innovation, probably starting from Finnic. However, also some Indo-European preconditions.	Difficult to estimate to what extent the Baltic/Finnic phenomena have parallels elsewhere <sup>a</sup>	Certain parallels in the Balkans

<sup>a</sup> no large scale cross-linguistic comparisons available

Table 2. The diagnostic features of the Caucasian Sprachbund

Phonology	Morphology	Syntax	Lexical semantics	Lexicon
rich consonantism	agglutination polysynthesism predominance of prefixal conjugation	identical word order (SOV, Attr-N)	stative vs. dynamic verbs	common cultural terms not found outside the Caucasus
ternary contrast of stops and affricates		ergative construction	inversive verbs	common phraseology specific to the area
glottalization		inversive construction	ambitransitive (labile) verbs	common semantic patterns
rich sibilant systems rich postvelar (uvular, pharyngeal and laryngeal) systems	predominance of postpositional constructions masdar (verbal noun) morphological marking of causative	the possessor constituent precedes the possessed one	suppletive verbs for singular and plural arguments	
similarly built harmonic clusters				
presence of schwa	category of evidentiality			
lack of phonemic diphthongs	category of potential			
lack of vocalic clusters ablaut	attachment of coordination markers to each conjunct directional and orientational preverbs group inflection a three-grade deictic distinction vigesimal numeral system			

isolated languages is high. The region probably represents a “residual zone”, in Nichols’ (1992) terms, or, in other words, a region in which linguistic diversity has been accumulating as languages were pushed aside by those of politically and economically powerful populations in adjacent linguistically homogenous “spread zones”. When the distribution of language families and isolates is shown in different colors on a map (e.g., Nimuendajú 1981, Queixalos and Renault-Lescure 2000), the Guaporé-Mamoré region comes across as a focal region of prehistoric population movements. However, from the available linguistic material many similarities between the languages come to the fore. These similarities cross at least ten genetic linguistic borders, the borders of the three culture areas Moxos, Guaporé and Tapajós-Madeira and the geographical border of the Guaporé River itself. They include features such as minimal shared vocabulary, complex verbal morphology, evidentials, directionals, inclusive/exclusive distinction (cf. Crevels and Muysken 2005), and, surprisingly, lack of classifiers. The similarities may be phonological, grammatical, morphological or, sporadically, lexical in nature. The phonological similarities include, for example, nasal harmony. Grammatical similarities concern, for example:

- a high incidence of prefixes
- evidentials
- directionals
- verbal number
- lack of nominal number
- lack of classifiers
- inclusive/exclusive distinction

At the morphological level we find similar forms for grammatical categories. Note that at this point we have included forms from the Bolivian highland languages Uru (URU-CHIPAYAN) and Aymara (AYMARAN) as well:

- locative case KAN *-ni*; KWA *-na*; AIK *-ne*; AYM *-na*
- applicative derivation KAN *-ta-* or *-to-*; KWA *-ta-*; LAK *-ka*; KAR *-ta-*; MOS *-tya-* or *-te-*
- emphatic marker KWA/MEK *-te(te)*; KAN *-kete-*; GAV *tere*; SIR *te*
- semantically empty noun formative root *e-/i-* in KWA/KAN/LAK/CAV/ESE
- specific classifiers such as KWA/KAN/AIK *-mū* ‘liquid’, KWA/KAN/AIK/ARI/LAK *-nū* ‘powder’, ‘porridge’, KWA/AIK/NAM/SAB *-su* ‘bone’.

Some of these categories may rather characterize subareas than the entire region. Similar forms are also found at the lexical level.

- banana: AKU/TUP/MEK/KWA *apara*; AIK *dipara*; ITO *upatsū*; MOV *pere*; MOS *chipeʔre*
- arrow: AKU *mambi*; ARI *mbu*; DJE *kubi*; KAN *mapi*; KWA *mābi*; MEK *mampi*