

Introducing Kariì, a Vietic language of Laos¹

N. J. Enfield (MPI, Nijmegen)

G rard Diffloth (EFEO, Siem Reap)

Abstract

This article presents a first sketch of Kariì, a Vietic (Eastern Mon-Khmer) language spoken in upland Laos. This previously undescribed language is of special interest not only in so far as it adds to typology's sample of the world's languages, but also in so far as its complex phonological system adds to our understanding of the historical development of Vietic and Mon-Khmer, and more generally of the process of tonogenesis. Kariì syllables are defined not only in terms of segments and segmental slots, but also in terms of register ('heavy' versus 'light') and what we call 'terminance' (voiced, voiceless, checked). Register and terminance, both laryngeal features, are clearly associated with the development of lexical tone. The paper also contains a sketch of the main morphosyntactic features of the language.

Keywords: Language description, tonogenesis, Mon-Khmer, Vietic, mainland Southeast Asian languages, Laos.

In this article, we present new facts and analyses from recent field work on a previously undescribed language of mainland Southeast Asia. Kariì is an Austroasiatic language of the Vietic branch of Eastern Mon-Khmer. It is spoken by approximately 250 people living in the upper reaches of the N rong river (or *Nam Noy* in Lao), a tributary of the Nam Theun, and ultimately the Mekong River, in central Laos. The language has a number of typologically

¹ The data discussed here were collected in four field expeditions carried out by Enfield to Kariì-speaking villages of the upper N rong valley in 2004-6: first, a brief introductory visit in May 2004; second, five weeks in summer 2004; third, a month in upper N rong plus a further week with Kariì speakers in Nakai District centre, in summer of 2005; and fourth, a month in upper N rong in summer 2006. Both authors collaborated in analysis of audio-recorded data collected in 2004-5, and in addition, both authors spent a week working together with Kariì speakers in Nakai in August 2006. The phonological analysis presented in section 2 is joint work. Remaining descriptive sections (3-9) were written by Enfield with later input from Diffloth. We gratefully acknowledge the input of audiences at Siem Reap (2006), Nijmegen (2007), and Leipzig (2007). We thank the Max Planck Society and especially Steve Levinson for support of this work.

noteworthy features, including derivational infixes, a very crowded vowel system (18 long vowel contrasts), and a complex phonological template which combines segmental structure with cross-cutting rime-level distinctions of register and terminance. In many respects, this language differs from the common vision of a 'typical mainland Southeast Asian language' such as Thai or Vietnamese. Kariì is unlike those national languages in several respects, for example its lack of lexical tone, its complex phonological template, its derivational morphology, its possession of unusual sounds such as devoiced rhotic finals and retroflex affricates, and its high number of contrasts in syllable-final position. However, in the context of the types of language which are most numerous yet least described in the area, a language like Kariì may be more worthy of the title 'typical' in mainland Southeast Asia.

In terms of morphological type, Kariì tends toward an isolating and analytic profile (no inflectional morphology such as case-marking or verb-argument agreement). It has a modest inventory of (essentially non-productive) derivational morphology, including infixation (causative, verbalizing, and nominalizing) and reduplication. The phonological system features a large number of vowel distinctions, and has other properties of significant interest for research on the history of Vietic and other Mon-Khmer languages. These include the retention of syllable-level contrasts in register (heavy versus light) and terminance (checked versus voiced versus voiceless) which are understood to presage the kind of tonogenesis hypothesized for Vietic languages such as Vietnamese and Muong, as well as other languages of Southeast Asia such as Kmhmu and Chinese.

In the current state of knowledge of this language, the best worked-out component of the grammar is the phonology. We therefore concentrate in this article on the sound system in some detail before sketching some of the morphosyntactic features. But first, some information about the context of the language.

1 Kariì speakers and their setting

Kariì is spoken in the upper valley of the Ñrong river, in Khammouan Province, Laos. The area lies about 300km due East of the capital city, Vientiane, about 90km Northeast of the Lao city of Thakhek on the Mekong River, and about 60km Southwest of the Vietnamese coastal city of Ha Tinh on the South China Sea. The villages lie between 600m and 700m above sea level, just on the Western side of the Annamite Chain ridge. The Ñrong valley lies in the path of shortest distance anywhere in Laos from the Mekong to the South China sea.

For this reason, the area has long been a trade route, as documented in Vietnamese administrative archives since the early 17th century (Quy 2002).

Kariì territory is situated deep inside the area of highest biodiversity in all of Laos. The Northern Annamites Rain Forests area is one of the Global 200 ecoregions that are 'outstanding examples of biodiversity' in the world (ADB/UNEP 2004:72). It is also a site of significant ethnolinguistic diversity. Three main ethnolinguistic groups in the area are Katuic (e.g., speakers of Brou), Tai (e.g., speakers of Saek and Tai Mène), and Vietic (speakers of a range of highly endangered Vietic languages such as Thèmarou and Ahlao; see Ferlus 1996 on the languages and peoples of the Vietic group more generally). As Chamberlain (1997) describes it, the Vietic speakers are the earliest of current inhabitants, with speakers of Northern Tai languages such as Saek arriving some 300 years ago, and Brou speakers arriving (from lower areas along the Korat Plateau in present-day Thailand) over the course of the last century or so. Tai speakers came in search of flat irrigable land on which to grow wet rice crops. The group most closely related to the Kariì are the Phòngq, who live in several villages further downstream on the Ñrong. The Kariì and Phòngq languages are mutually intelligible – that is, they may be regarded as dialects of a single language. The Kariì also come into regular contact (as has likely been the case for some hundreds of years) with speakers of Vietnamese (mostly itinerant hiker-traders), and speakers of Lao, including local dialects spoken around the lowland villages far downriver from the Kariì, as well as the modern national language spoken in the extreme lowlands, along the Mekong River (Enfield 1999, in press).

The word *Kariì* is an autonym (as in *mleeng kariì* 'Kariì people', *meengq kariì* 'Kariì language', or *bruuq kariì* 'Kariì forest/territory'). The word also refers to other Vietic groups of the area. The Lao word *khaa* is widely considered to be pejorative with reference to upland minority people (the word also means 'slave'). Rivière's (1902) term for the Kariì is *Harème*, clearly from the word *Arem* used by Brou speakers to refer to the Kariì.² Another word which has been used for the Kariì is *Salang* or *Salaang*. Its origin is unknown, and its reference is uncertain (as it has been used with reference to other groups as well). The word *kariì* is otherwise a verb meaning 'to grow!', with a cognate of similar meaning in neighboring Brou.

² Ethnologue.com has no entry for Kariì, but it may correspond to the language listed as *Arem* (although the information supplied there differs in some ways from that given here).

Kariì tradition is to periodically move around, within a delineated 'home range' territory in the upper reaches of the N̄rong, practicing shifting cultivation, mostly on flat land alongside riverbanks. Relocation is frequent. A house will be abandoned by its builder/owner within about five years from construction, for example due to death in the family. Traditionally, upon death of a family member (small children are an exception), the house is dismantled and rebuilt on another location, anywhere from a few meters away to several kilometers away in another village location. The floor of the house is always discarded after a death in the household. (See Enfield 2007a for an ethnographic description of the Kariì house.)

Material culture is unelaborate. There is little in the way of decorative art, no weaving of cloth or other production of clothing, no decorative carving apart from very simple fashioning of window frames, for example. Unlike many Bahnaric and Katuic speaking groups further South, there is no special decoration of the house. Basketry is the one relatively elaborated area of material culture. The techniques used are widespread across mainland Southeast Asia. A well-sharpened machete is the main tool. An initial phase of preparing weave-strips (*laac*) precedes the hand weaving of a large range of practical artifacts, including baskets of many shapes and sizes, and several types of fish trap (although of limited range compared to the repertoire of the lowland Lao). Basketry is men's work.

Kariì adults are multilingual, speaking Kariì, Vietnamese, Saek, Brou, and Lao. The men have some advantage over the women in this respect, since they travel more widely and more often. The men go to market semi-regularly, and many of them spend a period in the army, traveling to other provinces of Laos. Kariì villagers have extensive contact with other ethnolinguistic groups, most proximally Saek and Vietnamese, but also including other Vietic groups, especially the Phòngq who live a short way downstream.

Descent is patrilineal, as reckoned through house spirits. In practice, a family is typically first established under the bride's house spirits, the young couple later moving to the groom's father's house when he can afford the bride price (which includes 2 pigs, one very big, one very small). Kinship is utterly central to the organization of life and of people's relationship practices, particularly as manifest in linguistic practice. Religion is animist, spirits are often consulted. Women are required to 'remain on the ground' during menstruation – that is, they are not to go up into their houses. Instead, they sleep in small raised huts built for this

purpose. The Kariì have traditionally upheld widespread restrictions on what kinds of food may be eaten, a very unusual trait for people of this area: they avoid eating domesticated red meat, large animals of any kind, and many kinds of smaller non-plant food sources such as snakes, insects and birds. However, this, like many traditional aspects of Kariì life, is rapidly changing.

2 Sound system of Kariì

Some notable features of the Kariì phonological system are:

- **Word structure:** Any Kariì word has one stressed syllable. If a word has two syllables, the first will be unstressed, and will be phonotactically constrained in other ways as well.
- **Vowels:** The vowel system features a large number of contrasts: 18 long vowels, 12 short vowels, and 3 diphthongs.
- **Register:** The high number of vowels is brought about by a thoroughgoing contrast between heavy and light registers for each of nine points in the vowel space.
- **Initial vs. final consonants:** In the inventory of consonants, there are many more initials than finals. Initial stops show a three-way contrast in voice onset time, including a series of implosives. Word-final stops show no contrast in voicing.
- **Terminance:** In word-final position, there is a systematic contrast between three types of terminance (see below for explication of this term) – voiced, voiceless, and checked. The full three-way contrast is realized for oral sonorant finals (e.g. [-l] vs. [-l̥] vs. [-l̚]), whereas only a two-way distinction is realized for nasal finals (voiced versus checked, e.g., [-n] vs. [-n̚]), and final stops show no contrast, being always checked (e.g., [t̚]).
- **Register and terminance cross-cut,** yielding six possible syllable types, structurally and historically comparable to tones: heavy checked, heavy voiced, heavy voiceless, light checked, light voiced, light voiceless. Such a division of syllable types is understood to have set the stage for tonogenesis in Vietic languages such as Vietnamese and Muong (cf. comparable processes in other tone languages such as varieties of Chinese).

2.1 Consonants

As many have done in research on sounds systems of Southeast Asian languages, we treat initial and final consonants as belonging to distinct systems.³ First, it is useful to note the segmental structure of the Kariì word:

$$(c_{-1} \ v_{-1}/r) \ (C_0) \ \mathbf{C} \ \mathbf{V} \ (C_F)$$

Figure 1. Segmental structure of the Kariì word (non-segmental features not specified here; see Figure 2, below, for full details, including terminance and register).

Within this structure, **C** is defined as the segmental onset of the stressed or major syllable. **C** may or may not be initial to the word, depending on whether the major syllable is preceded by a minor (unstressed) syllable. The term we use for **C** is *major-initial consonant*. This does not mean that **C** is necessarily WORD-initial, although it may be. The term we use for **C_F** in Figure 1 is *final consonant*. The term we use for **C₋₁** in Figure 1 is *minor-initial consonant*. When **C₀** is realized, the major syllable onset is a **CC** cluster.⁴ Where necessary, we use the term *Pre-C consonant* to refer to any consonant which comes before **C** (i.e., either **C₀** or **C₋₁**). We use the term *rime* to refer to the stressed vocalic element along with the following consonant (i.e., the word final consonant) if there is one. We use the term *onset* to refer to the material which precedes the rime.

2.1.1 Major-initial consonants

Table 1 lists the full inventory of major-initial consonants:

³ Following work on other Southeast Asian languages, we regard a distinction between onset and rime as the main distinction in the Kariì syllable. Sino-Tibetan, Tai, and Hmong-Mien linguistics have done this traditionally.

⁴ These are not always true clusters in phonetic terms, due to the insertion of a predictable epenthetic vowel in many **CC** sequences (e.g., /kt-/ [kʰt-] **CC**- vs. /kat-/ [kat-] **CvC**-). See section on Phonotactics, below.

	bilabial	alveolar	retroflex	palatal	velar	laryngeal
voiced (implosive) stops	ɓ	ɗ		ɟ		
voiceless aspirated stops	p^h	t^h	t^{sh}		k^h	
voiceless unaspirated stops	p	t	t^s	c	k	ʔ
fricatives		s			ɣ	h
nasal sonorants	m	n		ɲ	ŋ	
oral sonorants	ʋ	l	r/zɿ	j		

Table 1. The 24 major-initial consonants.

We now supply some illustrative lexical contrasts.⁵

(1) *Bilabial initial stops of different manner of articulation*

Impl. *V'less* *Asp'd*

ɓa: 'flank' **pa:** 'three' **p^ha:** 'to cool'

(2) *Alveolar initial stops of different manner of articulation*

Impl. *V'less* *Asp'd*

ɗam 'five' **taj** 'French' **t^haj** 'to change'

(3) *Retroflex initial stops of different manner of articulation*

V'less *Asp'd*

t^{sə}ɹ: 'a swidden' **t^{shə}ɹ:n** 'a rack'

Note also that the retroflex initials contrast with phonetically similar clusters:

⁵ Throughout this article, for convenience of presentation, we give simplified glosses of the meaning of example words. Often, the lexical meanings are more specific than we supply here. The authors may be contacted for more information.

(4) *Retroflex initial stops and phonetically similar clusters*

<i>V'less</i>	<i>Asp'd</i>	<i>tr-</i>	<i>tar-</i>	<i>trh-</i>
t̪ʰa:n 'bird sp.'	t̪ʰa:j 'upset'	tra:kʰ 'lead'	ta:ra:kʰ 'wake'	trh^hɛ:ŋtʰ 'to suck'

(5) *Palatal initials of different manner of articulation*

<i>Impl.</i>	<i>V'less</i>	<i>Glide</i>	<i>Nasal</i>
f^ha:ŋʰ 'to step over'	c^ha:wʰ 'a strip'	j^ha:jʰ 'crazy'	ɲ^ha:ŋ 'tree sp.'
fa: 'to release'	ca: '1PL.EXCL'	ja: 'FeZ'	ɲa:r '1DU.EXCL'

(6) *Velar initial stops of different manner of articulation*

<i>V'less</i>	<i>Asp'd</i>
ka:ŋ 'to scrape'	k^ha:ŋ 'to escape'

(7) *Rhotic initials*

<i>l-</i>	<i>r-</i>
la:cʰ 'oversized'	z^ha:cʰ 'belly, guts'

As Table 1 shows, the system of major-initial consonants is not perfectly symmetrical. There are gaps. In the stops, we do not see full symmetry in distinctions in voicing for all places of articulation. First, there is no voiced stop in velar position to complement **k** and **k^h**. A voiced velar stop is occasionally found in the phonology of related languages (e.g. Katu; Costello 1993:vii), but these contrasts have disappeared in much of the MSEA area, apparently due to a historical process of devoicing (Haudricourt 1946). Languages like Brou, Thai, Lao, and Khmer do not have voiced velar stops in their phonological systems. These languages only have voiced initials in positions further forward, showing an analogous gap.

Of typological note in the initial consonant series is the implosive articulation of the voiced stops. While typologically unusual across languages of the world, this is not unusual for an Eastern Mon-Khmer (EMK) language. It represents the proto-EMK state of affairs. Where languages of this area have implosive stops, these are often a historically recent development, and are often even fewer in number, normally only including **b** and **d**, as in some modern

varieties of Khmer. If a language has more than **ɓ** and **ɗ** in an implosive series, the next addition will be palatal **ɟ**, as found in Kariì. It is more rare to find an implosive in velar position **ɠ** – this occurs, for example, in Nya Hoen (e.g., **ɠɔk** 'to cough'; Diffloth field notes; cf. Sidwell and Jacq 2003:137, and Davis 1968, cited there). In Vietnamese, there is a phonological category of voiced velar stop (Thompson 1987:28), but it is phonetically not a stop, rather it is (in most contexts at least) a fricative **ɣ**.

The retroflex stops are affricated, and are anomalous in not having voiced, nasal, or oral sonorant counterparts, and in not having any counterparts in final position. It is not possible to regard these as, say, clusters of t and r, since there is a clear contrast between [**tr-**] and [**tʂ-**] (which both contrast, in addition, with [**tar-**]; cf. [**tra:kʔ**] 'lead (a type of metal)' vs. [**tʂa:n**] 'a rack' vs. [**tarakʔ**] 'to wake in fright'). Additionally, only the unaspirated retroflex initial occurs in minor-initial position, and in that position it is not contrastive. It occurs there in free variation with [**r-**] (e.g., the word which means 'new' may be pronounced variously as [**rʰmɑː**], [**tʂʰmɑː**], and [**tʂʰmɑː**]). It is unusual for a mainland Southeast Asian language to have retroflexes (but cf. some varieties of Vietnamese for a retroflex initial in words like *trên* 'above', and Beijing Mandarin for retroflex fricatives in words like *sī* 'ten').

Unusually, there is a fricative/continuant **ɣ** at the velar place of articulation. This segment occurs in two other languages in close contact with Kariì: Vietnamese and Saek (Thompson 1987, Gedney 1993:xxv). However, it is marginal in Kariì. The only clear case we have is the borrowing [**ɣɛːʔ**] 'scabies' (cf. Vietnamese *ghẻ*). There is no historical background for such a velar fricative segment in Mon-Khmer (i.e., there is no reconstruction of a velar fricative at any early stage of Mon-Khmer). Like in Vietnamese, it would seem suggestive of a counterpart to the voiced/implosive **g**.

Aspirated stops plays a minor role in the system. The lexical type frequency⁶ of aspirated stops is very much lower (less than one to seven overall) than the unaspirated counterparts. In a list of 2778 lexical items, we have:⁷

(8) *Lexical type frequency of unaspirated versus aspirated stop initials*

p vs. p^h	=	208 vs. 26
t vs. t^h	=	229 vs. 35
k vs. k^h	=	246 vs. 43

Furthermore, most words with aspirated stops are borrowings (Saek, Vietnamese, or Lao):

(9) *Borrowings with aspiration*

- (a) **p^hɑ:n^ʔ** 'to dance' (Lao *fòòn4*)
- (b) **t^{shə}ɑ:p^ʔ** 'a meal' (Saek *thraap⁵* 'a meal')
- (c) **t^hɛ:k^ʔ** 'to measure' (Lao *thèèk4* 'to measure')
- (d) **k^{hə}ɑ:k^ʔ** 'great' (Lao *khakl* 'great')

There is no aspirated palatal [**c^h**] to complete a series with [**p^h, t^h, k^h**], by analogy with [**p, t, c, k**]. This gap is also attested in both Lao and Vietnamese, two major contact languages for Kariì speakers. Some languages elsewhere in the area do feature an aspirated palatal stop (e.g. Thai and Khmer).

⁶ By *lexical type frequency* of a phonological element, we mean its frequency of occurrence across entries in the lexicon, independent of its *discourse token frequency*, i.e., how often it occurs in discourse. (The distinction can make a world of difference – witness English /ð/, with very low lexical type frequency, but very high discourse token frequency, thanks to the extremely common occurrence of a small set of words including *the, this, that, there, then.*)

⁷ We do not list the retroflex initials here, since they do not pattern as a set with the other initial stops – the two retroflex initials have very low type frequency, occurring only about 10 times each in our lexicon of 2778 words.

2.1.2 Final consonants

In mainland Southeast Asian languages, initial consonants and final consonants typically belong to distinct systems. The finals always form a smaller set, showing absence or neutralization of contrasts which are found in initial position. Extreme cases can be observed in languages of the Hmong-Mien family (see Ramsey 1987:282 on Miao). A more typical example is Thai. In Thai, initial stops show a three-way contrast in voice onset time (e.g., /b-/ , /p-/ , /p^h-/), while final stops have a single value, namely checked or unreleased (e.g., /-p̚/). The same thing is true for stops in Kariì, as shown above (Table 1 and following examples). In this sense, Kariì is typical. However, going beyond the stops, the Kariì system shows an unusually high number of contrasts in final position for a mainland Southeast Asian language. The number of contrastive oppositions in manner for sonorant finals is greater than that for initials. In addition, the NATURE of the contrasts found there (devoicing, post-glottalization) is areally unusual. The system furnishes systematic contrasts between voiced, post-glottalized, and devoiced finals. Consider the following examples:

- (10) (a) **t̚ʰɔ̃j** 'tail' **t̚ʰɔ̃j̥** 'bowl' **t̚ʰɔ̃j̥** 'follow'
- (b) **z̥ʰɑ̃:** 'turtle' **z̥ʰɑ̃:** 'pig basket' **z̥ʰɑ̃:** 'dry'
- (c) **cam̥ɑ̃:l** 'shiny' **?um̥ɑ̃:l̥** 'to hunt' **ɗ̥ɑ̃:l̥** 'to bounce'
- (d) **kav̥ər** 'stir' **kav̥ər̥** 'embrace' **tar̥** 'to run out of workspace'
- (e) **car̥ɛ̃:w** 'green' **sar̥ɛ̃:w̥** 'to raise/feed' **ɓ̥lɛ̃:w̥** 'four-eyed turtle'
- (f) **ɟ̥ʰɑ̃:m** 'sugar cane' **ɟ̥ʰɑ̃:m̥** 'to cry'
- (g) **kɑ̃:n** 'oversize' **kɑ̃:n̥** 'to hunt by night'
- (h) **ɓ̥ũj̥** 'dust' **p̥ũj̥** 'tree sp. (*licuala* sp.)'
- (i) **cɑ̃ŋ** 'tree sp.' **cɑ̃ŋ̥** 'salty'

We do not analyze these syllable-final contrasts as involving extra segmental material (e.g., in an extended set of final consonants, or in a second final consonant slot in the phonotactic template). Instead, we treat the distinctions in (10) as being the product of non-segmental,

rime-level features. Table 2 lists the full inventory of final consonants (over which we shall later see the distinctions in terminance are laid):

		bilabial	alveolar	palatal	(post) velar
stops	checked	p'	t'	c'	k'
	nasal	m	n	ɲ	ŋ
sonorants	oral	w	l r	j	v

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

We first establish that the final segments listed in Table 2 are lexically contrastive:

(11) *Stop finals at different place of articulation*

<i>Lab</i>	<i>Alv</i>	<i>Pal</i>	<i>Velar</i>
kɑ:p' 'to wash'	ʙɑ:t' 'to rub'	ʙɑ:c' 'mushroom'	ʙɑ:k' 'to scoop'

(12) *Nasal finals at different place of articulation*

<i>Lab</i>	<i>Alv</i>	<i>Pal</i>	<i>Velar</i>
ɔ̃:m 'to perch'	ɔ̃:n 'bent'	lɔ̃:ɲ 'sharp'	tɔ̃:ŋ 'to tap'

(13) *Oral sonorant finals at different place of articulation*

<i>Lab</i>	<i>Alv</i>	<i>Lat</i>	<i>Pal</i>
sɑ:w 'ascend'	hɑ:r 'two'	masɑ:l 'tuber sp.'	sɑ:j 'ear'

(14) *Bilabial finals of different manner of articulation*

<i>Stop</i>	<i>Nasal</i>	<i>Glide</i>
katɑ:p' 'basket type'	katɑ:m 'crab'	kɑ:w 'turtle sp.'

(15) *Alveolar finals of different manner of articulation*

<i>Stop</i>	<i>Nasal</i>	<i>Lateral</i>	<i>Rhotic</i>
kɑ:t 'to bite'	kɑ:n 'oversize'	z^hpɑ:l 'butt, end'	hɑ:r 'two'

(16) *Palatal finals of different manner of articulation*

<i>Stop</i>	<i>Nasal</i>	<i>Glide</i>	<i>Post-glottalized</i>	<i>De-voiced</i>
ɸɑ:c 'mushroom'	ɸɑ:ŋ 'dust'	ɸɑ:j 'scold'	ɸɑ:j^ʔ 'salt'	pɑ:j̥ 'muntjac'

Note with regard to the palatal finals that there is a contrast between the phonetically similar checked palatal stop final **-c** and post-glottalized palatal glide **-j^ʔ**:

(17)

<i>Stop</i>	<i>Post-glottalized</i>
ɸɑ:c 'mushroom'	ɸɑ:j^ʔ 'salt'
hɑ:c 'slippery'	hɑ:j^ʔ 'to strip'
ʔɑ:c 'go'	kɑ:j^ʔ 'head hair'

(18) *Velar finals of different manner of articulation*

<i>Stop</i>	<i>Nasal</i>
ɸɑ:k 'water'	ɸɑ:ŋ 'tree sp. (<i>Murraya paniculata</i>)'

Comparing Tables 1 and 2, we see that the systems of initial and final consonants are strikingly non-equivalent. Looking first at the stops, there are thirteen distinct units in initial position (Table 1) versus four in final position (Table 2). In final position, stops are always checked (as [-p^ʔ, -t^ʔ, -c^ʔ, -k^ʔ]) and there is no realization of lexical contrast in aspiration, voicing, or implosive articulation.⁸ Accordingly, none of the gaps in the initial stop series (e.g. implosive velar stop, aspirated palatal stop) are evidenced in final position. That is, there

⁸ In utterance-final position, final stops are often pronounced with a kind of delayed voiceless aspiration (e.g., [siit^h] as a variant of [siit^ʔ] for 'animal'). This is not lexically contrastive, but is some kind of practice whose meaning is not yet clear.

are no contrasts in manner of articulation for final stops. However, beyond the stops, we do find contrasts in manner of articulation. This arises through a rime-level phenomenon which we refer to as syllable terminance, to which we now turn.

2.1.3 Syllable terminance

Every rime (and therefore every major syllable – see phonotactics below) must have a terminance value, where there are three possibilities: voiceless, voiced, and checked. Not all classes of final consonant show all terminance values:

(19)

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

Voiceless terminance. Oral sonorant finals /-**V**, -**j**, -**r**, -**l**, -**w**/ may occur with voiceless terminance, where offset of the final element is devoiced, as [-**V**^ʰ, -**j**^ʰ, -**r**^ʰ, -**l**^ʰ, -**w**^ʰ]. This contrasts with voiced terminance [-**V**, -**j**, -**r**, -**l**, -**w**] and checked terminance [-**V**^ʔ, -**j**^ʔ, -**r**^ʔ, -**l**^ʔ, -**w**^ʔ].

Since terminance distinctions of this kind are decidedly rare in mainland Southeast Asia, we want to provide as extensive evidence as we can to characterize these contrasts. First, it is important to note asymmetries in frequency in our set of 2778 lexical entries:

Final segment	Total type occurrences	w/ Voiced terminance	w/ Checked terminance	w/ Voiceless terminance
- V	616	253 (41%)	184 (30%)	179 (29%)
- j	202	102 (50%)	63 (31%)	37 (19%)
- r	189	98 (52%)	66 (35%)	25 (13%)
- l	135	74 (55%)	58 (43%)	3 (2%)
- w	94	55 (59%)	38 (40%)	1 (1%)

Table 3. Type frequencies of final oral sonorants and their occurrence with different terminance distinctions (as absolute number and as percentage of all type occurrences of that final segment with the relevant terminance value).

We now list examples which establish the relevant contrasts, beginning with the lowest frequency case, *-w*, for which we have a single example to date of voiceless terminance:

(20) *Terminance distinctions in -w*

<i>Voiced</i>	<i>Voiceless</i>	<i>Checked</i>
carɛ:w 'green'	ble:w 'four-eyed turtle'	sarɛ:w[?] 'to raise; to feed'

(21) *Terminance distinctions in -l*

<i>Voiced</i>	<i>Voiceless</i>	<i>Checked</i>
dəl 'to shut off'	dəl 'to bounce'	val[?] 'bent'
kəl 'fish sp.'	pəl 'a snare'	bəl[?] 'how many'
t^oɔ:l 'to cook in tube'	p^oɔ:l 'tree sp.'	cam^oɔ:l[?] 'dibble stick'

(22) *Terminance distinctions in -V*

<i>Voiced</i>	<i>Voiceless</i>	<i>Checked</i>
z^ɪɑ: 'turtle'	z^ɪɑ: 'to be dry'	z^ɪɑ:[?] 'pig basket'
kɑ: 'chicken'	kɑ: 'all'	?akɑ:[?] 'fish'
kɯ: 'a shoot'	tukɯ: 'leaf'	cakɯ:[?] 'a bear'

(23) *Terminance distinctions in -r*

<i>Voiced</i>	<i>Voiceless</i>	<i>Checked</i>
ɕir 'strips'	ɕir 'to shake'	ɕir[?] 'wake up'
kinɛ:r 'spur'	lɛ:r 'root'	ŋɛ:r[?] 'yellow'
svɛ:r 'chipped'	vɛ:r 'to separate'	vɛ:r[?] 'to pivot'
kumi:r 'ant sp.'	?ipi:r 'tree sp.'	bi:r[?] 'to wipe'
pc^əar 'slat'	j^əar 'far'	c^əar[?] 'to flow'

(24) *Terminance distinctions in -j*

<i>Voiced</i>		<i>Voiceless</i>		<i>Checked</i>	
taua:j	'happy'	ɬava:j̥	'tuber sp.'	rua:j̥ʔ	'a spirit'
la:j	'patterned'	cala:j̥	'sit at fire'	la:j̥ʔ	'to succeed'
t̚o:j̥	'a tail'	t̚o:j̥	'to chase'	t̚o:j̥ʔ	'a bowl'
k^ha:j	'tree sp.'	k^ha:j̥	'to escape'	k^ha:j̥ʔ	'1sg'
kut̚o:j̥	'dog's vagina'	kut̚o:j̥	'tuber sp.'	krt̚o:j̥ʔ	'bird sp.'
pro:j̥	'forget'	turu:j̥	'boiling'	ru:j̥ʔ	'fall off'

Note that the devoiced final oral sonorants **-r̥** and **-j̥** are almost in complementary distribution, with **-r̥** occurring after front vowels, and **-j̥** occurring after back vowels. The two do, however, contrast after **-a-**, as the following examples show:⁹

⁹ The explanation is historical, with devoiced **-r̥** and **-j̥** corresponding to a word final ***-s** (now absent in Kariì). Words with final **-r̥** and **-j̥** correspond to words with final **-s** in related languages: for **-r̥** versus **-r**, compare cognates Kariì **le:r̥** 'root' and Tariang **rias** 'root' (Theraphan 2001:485) versus Kariì **sapi:r̥** 'pumpkin' and Nyah Kur **cmpiir** 'pumpkin' (Diffloth 1982:92); for **-j̥** versus **-j**, compare cognates Kariì **ma:j̥** 'gnat' and Stieng **mo:s** 'mosquito' (Haupers 1979) versus Kariì **tm̩a:j̥** 'guest' and Kmhmu **tm̩ɔj** 'stranger' (Swantesson et al 1994:215). The idea that devoiced **-r̥** and **-j̥** correspond to a word final ***-s** is further supported by type frequency: TAKEN TOGETHER, **-r̥** and **-j̥** show a similar frequency to other SINGLE palatal finals (e.g., **-c** or **-ɲ**). If we identify devoiced vowel finals **-V̥** with erstwhile **/h/** in final consonant position, a possible scenario is that after a development of **-s** into **-r̥** and **-j̥** in complementary environments, these became aligned with **-V̥** in a paradigmatic series of devoiced terminant sonorants. Subsequent innovation of parallel devoiced versions of **-w** and **-l** (**-w̥** and **-l̥**), might have then filled the gaps, bringing all oral sonorants into line as a set (through Sapir's 'drift', to avoid the 'psychological shakiness' of having empty cells in the system; Sapir 1921:154-8).

(25) *Contrast between -r̥ and -j̥*

-r̥		-j̥	
tar̥	'to reach the end'	tḁj̥	'to tap'
j̥ar̥	'far'	c̥ḁj̥	'sneeze'
?var̥	'to jiggle'	kap̥aj̥	'cloth'
mar̥	'to speed up'	pḁj̥	'seven'

Checked terminance. A syllable with checked terminance displays complete obstruction of airflow (without immediate release) at the end of the word. Rimes which end in checked stops (i.e., /-p̥, -t̥, -c̥, -k̥/) always have checked terminance. When, however, the final segment is a sonorant /-m, -n, -ŋ, -V, -j, -r, -l, -w/, the Karĩ system allows a distinction between checked and voiced terminance values (in addition to the voiceless terminance just outlined for /-V, -j, -r, -l, -w/). With voiced terminance, a final continuant phoneme is simply realized as a continuant final (as in [p̥ḁ:] 'cow', [j̥ḁ:m] 'sugar cane', [b̥uaj̥] 'tea scoop'), while with checked terminance, a continuant final segment is post-glottalized (as in [p̥ḁ:] 'father's younger brother', [j̥ḁ:m̥] 'cry', [b̥uaj̥] 'time, occasion'). Phonetically, for sonorant finals, checked terminance simply means the presence of a glottal stop. However, to reiterate a point made above, while in phonetic terms these syllables do have a glottal stop final, this glottal stop does not occupy a segmental 'slot', but represents a rime-level laryngeal distinction in terminance.

We now supply some examples showing contrasts involving final checked terminance. We have already supplied a good number of examples of checked terminance in presenting the three-way terminance distinctions for oral sonorants, above. But since -w showed so few cases, we add further examples here focusing on the voiced/checked distinction for -w:

(26)

<i>Voiced</i>		<i>Checked</i>	
tɛ:w	'to wear'	tɛ:wʔ	'tree sp.'
cʰa:w	'speak'	cʰa:wʔ	'wood strips'
kɔ:w	'rich'	kɔ:wʔ	'1pl INCL'

Since nasals do not show voiceless terminance, there is only a two-way terminance distinction in final nasals. Here are examples of the voiced versus checked terminance contrast for final nasals:

(27)

<i>Voiced</i>		<i>Checked</i>	
jʰa:m	'sugar cane'	jʰa:mʔ	'cry'
kata:m	'a crab'	sa:mʔ	'eight'
dɔ:m	'to perch'	lɔ:mʔ	'to steal'

(28)

<i>Voiced</i>		<i>Checked</i>	
kɓa:n	'washbowl'	ɓa:nʔ	'village'
palu:n	'earthworm'	lu:nʔ	'to cross under'
nɔ:n	'cooked rice'	tɔ:nʔ	'to chop'

(29)

<i>Voiced</i>		<i>Checked</i>	
ɓu:n	'ashes, dust'	pu:nʔ	'tree sp.'
lɔ:n	'be pointed'	kɔ:nʔ	'shoots'
tɔ:n	'to weave'	pɔ:nʔ	'to sell'

(30)

<i>Voiced</i>		<i>Checked</i>	
pa:ŋ	'to signpost a swidden'	ma:ŋʔ	'elder brother'
cal^o:ŋ	'tree sp.'	kal^o:ŋʔ	'a loop'
caŋ	'tree sp.'	caŋʔ	'salty'

Note that there is also a type frequency asymmetry for the nasal terminance types:

Final segment	Total type occurrences	w/ Voiced terminance	w/ Checked terminance
-m	165	99 (60%)	66 (40%)
-n	191	109 (57%)	82 (43%)
-ɲ	67	29 (43%)	38 (57%)
-ŋ	440	276 (63%)	164 (37%)

Table 4. Type frequencies of final nasals and their occurrence with different terminance distinctions (as absolute number and as percentage of all type occurrences of that final segment with the relevant terminance value).

The relative frequencies of voiced versus checked terminance for final nasals look fairly consistent (in the 60/40 range), with the exception of **-ɲ**, which shows the reverse pattern.

Post-glottalization occurs in nearly half of all nasal-final words, showing that it is not a marginal feature, but a robust phonological feature of the Kariì lexicon. Further work is needed in order to understand how it is that the feature of lexically specified post-glottalization could have originally developed. A conceivable hypothesis might be that post-glottalization has its source in borrowings from languages like Saek or Lao varieties, since glottalization is often a predictable feature of specific lexical tones in those languages.

However, such an account offers little promise, due to the fact that post-glottalized **-ɲʔ** is common in Kariì (indeed more common than the non-glottalized version): the Tai languages do not have palatal finals, and could therefore not have contributed to the Kariì pattern here.

Summary of terminance. We now summarize the discussion of terminance distinctions in Kariï. There are three terminance values: checked, voiced, and voiceless. Not all finals (Table 2) can occur with all three terminance values. Final stops show no terminance contrast (they are always checked); final nasals show a two-way contrast between checked and voiced terminance; and final oral sonorants show a three-way contrast between checked, voiced, and voiceless terminance. This gives the following full set of rime-endings:

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

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

A consequence of this terminance analysis is that many of the large number of rime-endings listed in Table 5 are not to be considered as composed of distinct segmental units. Specifically, the two rows of nasals and the three rows of oral sonorants each reduce to one, in segmental terms. The contrasts arise from a combination of a particular final consonant with a particular rime-level terminance value. This analysis is partly motivated by the relative cost of the alternatives. One alternative would be to add a set of nine new final segmental units in a post-glottalized series (with no equivalents in the system of initials),¹⁰ plus a set of five new phonemic segmental units in a voiceless series (with no equivalents in the system of initials), as listed in Table 5. Another alternative would be to treat the final glottal stop and **h** as distinct segments in the system of finals, thereby allowing final consonant clusters. However, there would then be only two permissible segments in the second position of such a cluster, namely [-^ʔ] and [-**h**], making it a somewhat unbalanced system. Furthermore, it is no

¹⁰ This is a solution currently adopted in Vietic historical reconstruction (Ferlus 1998).

coincidence that these are both laryngeal, a feature strongly associated with suprasegmentals (cf. Diffloth 1974). To treat these as final clusters would be a major departure from the otherwise required phonotactic template (see below). By treating terminance as a non-segmental phenomenon, our analysis avoids these problems.

2.2 Vowels

As in many other Mon-Khmer languages of MSEA, Kariì words may be monosyllabic or disyllabic. When a word is disyllabic, the first of the two syllables will be heavily restricted. We use the term *minor syllable* to refer to the first part of words which have an iambic prosodic pattern (e.g., [ka-] in [ka'ta:m] ‘crab’, [pu-] in [pu'le:ʔ] ‘fruit’). Some key features of the minor syllable are: reduced stress, lack of vowel length distinction, a much reduced system of vowel quality contrasts, and absence of independent register distinction (see below). The *major syllable* (i.e., the rest of the word) receives primary word-level stress, and shows the full range of contrasts in the vowel system (quality, length, register), just as a monosyllabic word does.

To describe the vowel system of Kariì, it is necessary to begin with the site of its greatest elaboration, the major syllable. As we shall later see, the minor syllable vowels constitute a radically reduced system.

2.2.1 Major syllable vowel nuclei

There is a fundamental distinction in Kariì major syllables between long and short vowels. The system of long vowels shows a greater number of quality distinctions than the system of short vowels. Long vowels also show greater type frequency: In words with types of syllable where vowel length contrast is possible, 67% of these words have long vowels.¹¹ In order to illustrate the full set of contrasts in the vocalic system, we begin then with the long vowels.

We observe 18 long vowel distinctions in Kariì:

¹¹ With stop final, short=224, long=400; with nasal, rhotic, glide finals (not -h or -V), short=444, long=980; i.e., 668 short versus 1380 instances of long vowels where contrast is possible.

	Front		Back
High	i:	ɨ:	u:
	^ei:	^aɨ:	^ou:
	ɨ:	ɥ:	ɯ:
	^ee:		^oɯ:
	ɛ:	ə:	^oɔ:
Low	ɛ:		ɑ:
		a: ^aa:	

Table 6. Long vowels, phonetically.

Here are some sets of illustrative (near) minimal contrasts between these long vowels:

(31) *Front vowels*

- pi:k⁷** 'a tadpole'
p^ei:t⁷ 'to plant'
vɨ:k⁷ '(to) work'
sav^ee:k⁷ 'to comb back'
pɛ:ŋ 'to be dear'
cavɛ:k⁷ 'peel sth. apart'
cavɑ:k⁷ 'fish sp.'

(32) *Back vowels*

- t^alu:** 'bamboo sp.'
cⁱl^ou: 'buffalo'
kalu: 'fruit sp.'
klɔ: 'star'
kl^oɔ: 'fish sp.'
culq: 'palm sp. (*licuala* sp.)'

(33) *Central vowels*

c^hce:	'now'
h^oɛ:	'broken'
mlɛ:	'afraid'
murə:	'disgusted'
tul^oa:	'weird'
kalə:	'bamboo sp.'

In comparison to other languages of the world, the set of contrasts represented here shows unusually fine phonetic distinctions. There is evidence, however, for a simpler underlying system. We can group these 18 vowels into 9 pairs, where the two members of each pair contrast in terms of phonation:

	Front		Back
High			
Low			

Table 7. Long vowels, showing vowels paired by register.

For each pair, as grouped by dotted ellipses in Table 7, one member has a relatively breathy or *heavy* phonation type, the other has a relatively tight and clear or *light* phonation type.¹²

¹² Native speakers have a folk category of ‘heavy’ versus ‘light’ in reference to the language’s sound system, but this is not consistently applied to the phenomenon of phonological register. There are three native applications of the term: (1) to the lexically specified phonological register distinction (heavy [cɪ:^h] ‘elder sister’ versus tight [c^hɪ:^t] ‘louse’); (2) lexically specified terminance distinction, especially where this determines the presence versus absence of a final glottal stop in a syllable (checked terminance [ʔakə:^h] ‘fish’ versus voiced terminance

This type of phonation-based distinction is common in Mon-Khmer languages. In phonological terms, we refer to this distinction as *register* (Henderson 1952, Gregerson 1976, inter alia). In addition to the basic difference in phonation type, there are noticeable vowel quality differences between heavy and light counterpart vowels, as shown in Table 7, above.

(34) *Some minimal pairs contrasting in register*

c̣i:ṇ 'nine'	c^he:ṇ 'cooked'
kar^ha:ŋ 'sand'	kar^ha:ŋ 'sunshine'
mat^ha:m 'son-in-law'	kat^ha:m 'crab'
kmɛ:̣ 'thumb'	km^he:̣ 'female (chicken)'
ẓ^hɔ̣: 'clear'	ẓ^hɔ̣: 'pig basket'
ẓ^hu: 'fence'	ẓ^hu: 'to know'

Emerging from the array in Table 7 is an orderly pattern of nine heavy-light pairs, giving 18 vowels in all. The pattern is easily 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>	ɨ̣	ɨ̣	ụ
	<i>Light</i>	^hɨ̣	^hɨ̣	^hụ
Mid	<i>Heavy</i>	ɪ̣	ɪ̣	ʊ̣
	<i>Light</i>	^hɪ̣	ə̣	ɔ̣
Low	<i>Heavy</i>	ɛ̣	^hạ	^hɔ̣
	<i>Light</i>	ɛ̣	ạ	ɔ̣

Table 8. Long vowels, laid out in a standardized 9-place system.

[**kạ**] 'chicken'); (3) dialect differences in pronunciation of the same word (any two cognates spoken in Kariì vs. the related dialect Phòngq).

The set of short vowels is significantly reduced in comparison to the long vowels, in having fewer members, lower type frequency, and more restricted phonotactics (see below):

		Front	Central	Back
	<i>Heavy</i>	ɪ̥	ɨ̥	ʊ̥
Non-Low		ɛ̥		ɔ̥
	<i>Light</i>		ə̥	
Low	<i>Heavy</i>	ɛ̣	ʌ̣	ɔ̣
	<i>Light</i>	ɛ̣	ʌ̣	ɔ̣

Table 9. Short vowels.

Of note here is that the two high central short vowels **ɨ̥** and **ə̥** are spaced further apart than their front and back counterparts. For the front and back vowels, comparing the short and long vowel systems, it is as if the mid vowel pairs are absent. The low central pair **ʌ̣** and **ʌ̣** is also present as in the long vowel system. For the mid and high central vowels, however, while the heavy high central short vowel **ɨ̥** is more or less of the same height as its counterpart in the long vowel system, the light high central short vowel **ə̥** is more similar in quality not to its counterpart the light high central long vowel, but to the light MID central short vowel, i.e. schwa. The behavior of this mid non-low short vowel pair prevents us from making the general statement that in the short vowel system the mid vowels are simply absent.

An advantage of taking the register distinction as primary is that we can then make some generalizations about the phonetics of vowel quality. For example, any heavy register vowel will be higher (or partially higher) in the vowel space than its light counterpart. One consequence of this is that the maximally high vowels [**ɪ̥ ɨ̥ ʊ̥**] are all heavy while the maximally low vowels [**ɛ̣ ʌ̣ ɔ̣**] are all light. These maximally high and maximally low

vowels are all steady state vowels, while their counterparts – high light [ᵉᵏ ʰᵏ ʰᵒᵏ] and low heavy [ᵉᵏ ʰᵏ ʰᵒᵏ] – all of which are closer to mid in height, mostly have noticeably distinct onsets ([ᵉᵏ] being the only exception). These onsets are even closer to mid in height.

Had we focused on vowel quality rather than phonation as a criterion for pairing, the result would have failed to capture the system concisely. Looking at Table 6, one might have thought that [ᵉᵏ ᵏ], [ʰᵏ ᵏ], and [ʰᵒᵏ ᵒᵏ] were relevant pairs, each at a similar height and each differing in phonation. This, however, would have left a number of vowels unpaired, and would not have cleanly captured the set as a whole. In addition, there is evidence from loanwords that these pairings are correct. Loans from Lao will have one of the two registers in Kariì (with even distribution: out of 188 Lao loans, 57% have light register). Depending on the register, the phonetic quality of the vowel in a Kariì word can be significantly different from that of the Lao source word. For example, the vowel in Kariì [ᵉᵏwᵏ] (a perfective marker) is phonetically much higher than the vowel in its Lao source [ᵉw⁵¹]; in fact, so much higher that it sounds closer to the next vowel up in the Lao vowel space: /e/ as in [ᵉw³⁵] 'to fight'. Similarly, the Kariì word [ᵉᵏᵏ] 'to like' is borrowed from Lao [ᵉᵏ⁴⁴] 'to like', with a significantly higher vowel, near to schwa. The vowel of the source word in Lao is phonetically more similar to that of another Kariì word, the light register [ᵉᵏ] 'to wear'. By contrast, Kariì ᵏᵏ 'sneaky', borrowed from Lao ᵏᵏ⁴⁴ 'to steal, to do secretly' has LIGHT register, and thus ends up with a vowel which is identical with that of the Lao source. This provides some support to the idea that each of the pairs shown in Table 7 correspond to a single underlying point in the vowel space.

There are three diphthongs, all of which seem to be found in borrowings. These are low in lexical type frequency (all together, 42 out of 2778 words have one of these diphthongs; less than 2 percent):

Front	Central	Back
ia	ia	ua

Table 10. Diphthongs.

There is no evidence for a contrastive register distinction in the diphthongs. They are reliably pronounced with heavy register (perhaps unexpected given that heavy is the marked register – see below). Attempted elicitation of register distinction did not yield recognition of a heavy-light contrast by speakers. This contrasts with the complete ease with which speakers recognize and produce register distinctions for regular vowels. Note that there are a number of cases which look like diphthongs, but which we treat as sequences of vowel plus glide (*aj*, *aw*, etc.), analyzing these as VC strings.

2.2.2 On the lack of relation between register and initial consonant class in Kariì

There is a type frequency asymmetry in the heavy versus light register distinction in Kariì. About two-thirds of words in our lexicon are light; 1892 vs. 887 (that is, 68% of words have light register). In related languages, register distinctions often show radically different distributions depending on classes of initial consonant, especially regarding implosives and glottal stops.¹³ This does not appear to be the case in Kariì, with glottal stops and implosive initial consonants showing a similar relation to register as other types of initial consonant: for example, in the set of words with voiceless stop initial **t-**, 68% (154 vs. 72) have light register; in the set of words with implosive stop initial **ɗ-**, 76% (122 vs. 38) have light register; in the set of words with glottal stop initial **ʔ-**, 73% (60 vs. 22) have light register. These figures are in line with the overall distribution of register in the language (68% of all words have light register).

2.2.3 Minor syllable vowel nuclei

¹³ For example, in modern Mon, words with a voiceless stop initial such as **t-** may have either clear or breathy register, with roughly equal type frequency, while words with an implosive stop initial such as **ɗ-** show almost exclusively clear register. This has a straightforward historical explanation: ***ɗ-** conditioned clear register, whereas modern **t-** has two historical sources, ***ɗ-** (which conditioned breathy register) and ***t-** (which conditioned clear register). Similarly, Mon words which begin with glottal stop will be almost exclusively clear in register.

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

- only three vowels occur: **-a-**, **-i-**, **-u-**
- 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 cases). 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%).

2.3 Phonotactics

2.3.1 Phonotactic template for the word

Here is a template for the possible types of Kariì word, in terms of segmental and suprasegmental components:

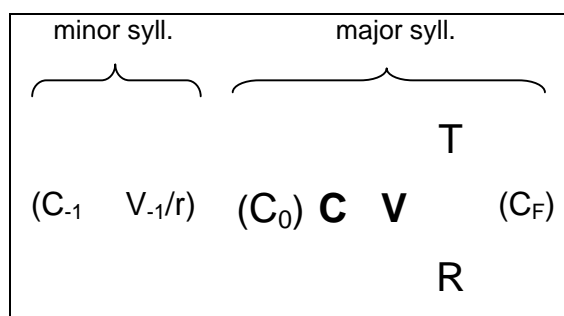
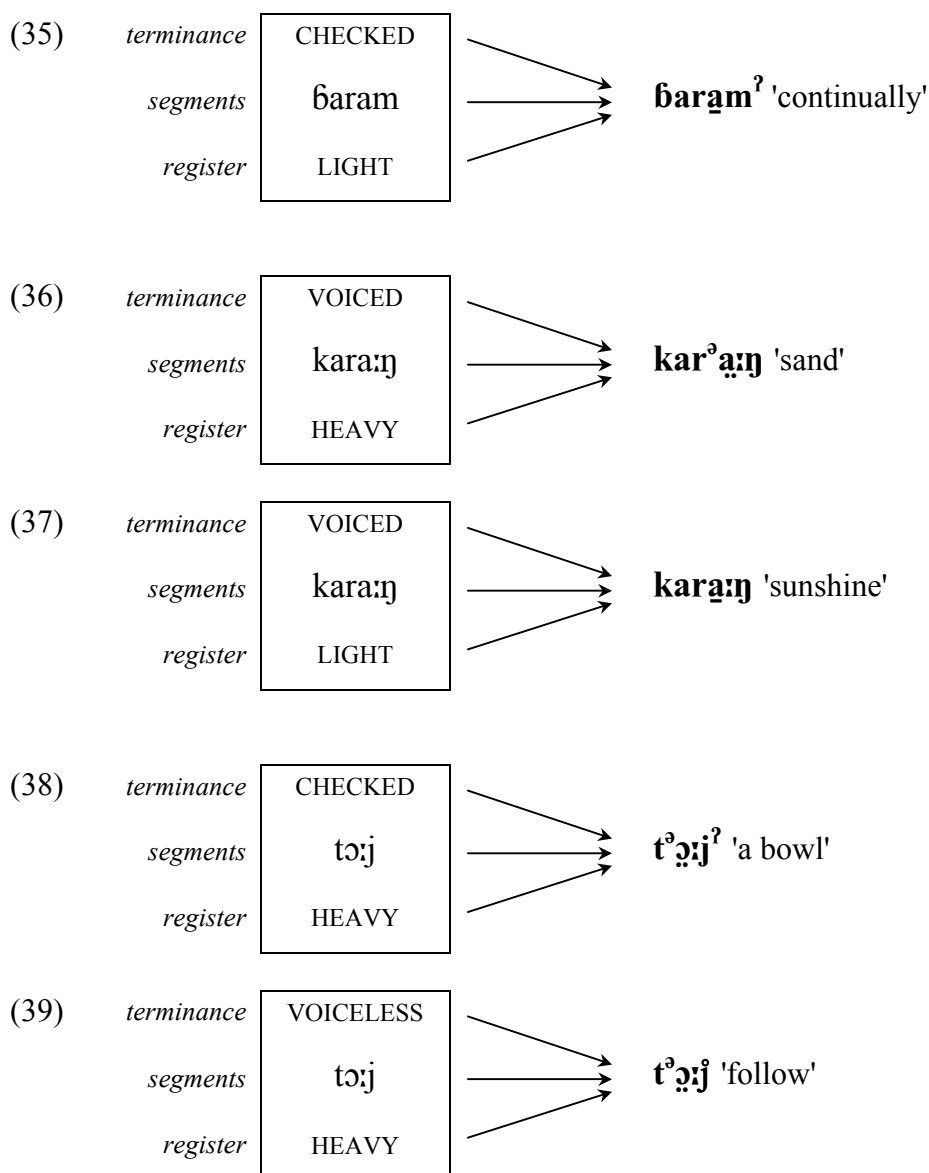


Figure 2. Word template (C=consonant, V=vowel, R=register, T=terminance, F=final)

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:



In this way, all words must have a value for terminance and register, and must have segmental material specified for at least the slots C and V. Remaining slots in the segmental template in Figure 2 may or may not be filled. To list the possibilities, we distinguish firstly between words with a single segment in the major-initial slot and those with a C₀C sequence:

(40) CV(C_F) type:

CV **ca:** '1pl.excl', **ca:**^ʔ 'dog', **z^ʰɔ:** 'dry', **ba:** 'to vomit'

CVC_F **ciŋ** 'rain', **ca:k^ʔ** '2SG', **z^ʰɔ:c^ʔ** 'belly, guts', **p^ʰɔ:!** 'tree sp.'

(41) $C_0CV(C_F)$ type:

C_0CV **kra:** 'old', **bra:** 'machete', **k^əd^əɛ:** 'sambar deer', **s^əma:** 'breed'

C_0CVC_F **praŋ** 'cross', **f^əla:j** 'hang net', **ble:ŋ** 'turtle sp.', **t^əbə** 'luck',

f^əhə:r 'hair on end'

Secondly, for both of these types, there may be a minor syllable preceding. Words of the type $CV(C_F)$ with a preceding minor syllable are common, as in the following examples:¹⁴

(42) $CV(C_F)$ type with minor syllable preceding:

$cvCV$ **?aka:** 'fish', **tara:** 'moon', **b^ərca:** 'tree sp.', **cadu:** 'to sit'

$cvCVC_F$ **kar^əɔ:ŋ** 'sand', **kafə:ŋ** 'seed', **t^ərfa:ŋ** 'walk', **pacə:r** 'to break off'

Words of the type $C_0CV(C_F)$ with preceding minor syllable are rare. In our current vocabulary of 2778 words, we have one word with the structure cvC_0CV , and three cases of cvC_0CVC_F :

(43) $C_0CV(C_F)$ type with minor syllable:

cvC_0CV **sakla:** 'fish scale'

cvC_0CVC_F **kapr^əɔ:t** 'beetle sp.', **kapla:ŋ** 'bamboo sheath', **kapl^əɔ:ŋ** 'rabbit'

2.3.2 Distributional constraints of segments

There are restrictions on the distribution and combination of segments within the template supplied in Figure 2. The situation regarding vowels is straightforward, as already described above: only **a**, **i**, and **u** may occur as V_{-1} , while any vowel may occur in the V slot. Note that if V is short, C_F must be realized. Recall that the laryngeals (glottal stop and **h**) are not final consonants in the segmental template, and accordingly there are no words with short vowel and glottal stop or **h** final.

¹⁴ For convenience here, we use lower case letters for the $C_{-1}V_{-1}$ of the minor syllable..

In all words, the C_F position may be filled by any of the final consonants (Table 2). In $CV(C_F)$ words, the C position may be filled by any of the major-initial consonants (Table 1). In words with a pre- C consonant, we have not yet found words with **h**, **j**, or **v** in either C_0 or C_{-1} position. A pre- C consonant will almost always be a voiceless unaspirated stop /**p**, **t**, **c**, **k**, **ʔ**/ or /**s**/. There are a few exceptions but the numbers are small. In pre- C position, aspirated stops are rare (with only a few cases of **kh-**), as are implosives, **ɗ-** and **ʄ-**. However, **ɓ-** is an exception, for which we have over 40 examples (mostly before **r** and **l**, also **s**). This is relevant to previous treatments of implosives in related languages as being clusters consisting of a glottal stop plus a stop, called 'preglottalized stops'. For example, Thompson (1987:23) describes the voiced implosives in Vietnamese as 'preglottalized and often imploded', analyzing [**ɓ-**] and [**ɗ-**] as /**ʔp-**/ and /**ʔt-**/. This would create problems in the Kariì case, due to the occurrence of initial strings like **ɓr-**, **ɓl-**, and **ɓs-** (e.g., **ɓraːʔ** 'machete', **ɓlɛːmʔ** 'to set (of the sun)', **ɓsɔːŋʔ** 'navel'), which would imply three-segment initial clusters /**ʔpr-**/, /**ʔpl-**/, and /**ʔps-**/.

There are very few cases in which aspirated stops occur non word-initially (i.e., in C position where there is pre- C material): our only cases are **ɕkʰaːnʔ** 'really', **ɕukʰɿlʔ** 'to lever', **ɕkʰaːl** 'cross-beam' (notably, all involving the same segments in pre- C and C positions).

Nasals in C_0 position are syllabic before homorganic stops, laterals, and nasals (note the resulting gemination in f-g):

- (44) (a) **ɱbaːtʔ** 'sweet'
 (b) **ɱlɔːʔ** 'palm sp.'
 (c) **ɱtɔː** 'oil'
 (d) **ɱcaːrʔ** 'mat'
 (e) **ɱkʰaːn** 'a wave'
 (f) **ɱmɛːnʔ** 'sweet'
 (g) **ɱnaːŋʔ** 'heavy'

In our only case in which a nasal in C_0 position precedes **h**, it is also syllabic:

(45) **nhɑ:mʔ** 'fragrant'

In other conditions, an epenthetic schwa appears between the C_0 nasal and **C**:

(46) (a) **ɲʷu:ʔ** 'fall through something'

(b) **ɲʷrɑŋ** 'River name'

(c) **mʷlɑ:cʰ** 'bland'

(d) **mʷlɛ:ŋ** 'person'

(e) **mʷrɛ:** 'in-laws'

2.3.3 Six rime types, as defined by terminance and register

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

		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 Kariì rime types, defined by intersections of terminance and register distinctions.

These six rime types show distinct type frequencies:

		register	
		<i>heavy</i> =	<i>light</i> =
		889 (32%)	1889 (68%)
terminance	<i>checked</i> = 1393 (50%)	400	993
	<i>voiced</i> = 1132 (41%)	412	720
	<i>voiceless</i> = 253 (9%)	77	176

Figure 4: Type frequencies of the different rime types from our list of 2778 words

These frequencies may appear to be unbalanced, but a closer look reveals that each category is well represented, once viewed within the context of the larger system. None of the rime types can be regarded as marginal.

First, as already noted above, there is an across-the-board imbalance between rimes of heavy and light register: that is, irrespective of terminance value, light rimes outnumber heavy rimes by about two to one. The total ratio of light rimes in our sample of 2778 words is 68% (n=1889). Within all checked terminance rimes, light register rimes make up 71% (n=993); within all voiced rimes, light register is 64% (n=720); within all voiceless rimes, it is 70% (n=176). So, the two-to-one ratio of light to heavy register is stable and insensitive to terminance type. This shows that register creates a fundamental dichotomy across the Kariï lexicon.

We may now ask why there are differences in absolute frequency between rimes of the three terminance values. The explanation is simple. As laid out in Table 5, above, checked terminance is a possible value for rimes with all 13 final consonants, whereas voiced terminance is possible for only 9 final consonants (stops are excluded), and voiceless terminance for only 5 (stops and nasals are excluded). Let us first compare the ratio of checked terminance rimes to voiced terminance rimes.

For checked terminance, at a total of 1393 cases with 13 finals possible, each final consonant occurs on average 107 times. For voiced terminance, at a total of 1132 cases with 9 finals possible, each final consonant occurs on average 125 times. Thus, the average occurrence of any final consonant for these two terminance values is of a similar order.

Turning to voiceless terminance, the situation looks different at first: there are 253 words with voiceless terminance, where five final consonants are possible. This would suggest an average frequency of 47 instances of voiceless terminance per consonant. However, as shown in Table 3, above, two of these five finals are vanishingly rare (one case of **-w̥** and three cases of **-ɲ**), and another two – **-r̥** and **-j̥** – have a single historical origin: ***-s**. So, not counting the handful of **-w̥** and **-ɲ** cases, and counting **-r̥** and **-j̥** together as a single consonant, we obtain a frequency of 127 instances of voiceless terminance per available final consonant, precisely in line with the frequencies for checked and voiced terminance (125 and 107), above.

In sum, taking into account systemic facts about the cross-cutting distinctions of register and terminance, we see that the system of six distinct rime types summarized in Figures 3 and 4 shows an even balance across the Kariì lexicon. In this and other respects, this system of suprasegmental distinctions is strongly reminiscent of a lexical tone system. The Kariì system is obviously similar to those that set the stage for tonogenesis in Tai languages, Vietnamese, and varieties of Chinese (Maspero 1912, Haudricourt 1946, 1953, 1954, Matisoff 1973, Gage 1985, Norman 1988, Alves 1995, Ferlus 1997, Thurgood 2002, *inter alia*). However, we want to stress that there is no way to know how the Kariì system may develop in the future, nor can it be regarded as 'transitional' in any special sense. Furthermore, while the Kariì data will be extremely important for research in comparative Vietic, Kariì is a modern language which has evidently long been developing in its own way (*cf.* our discussion of voiceless terminance, above). The Kariì sound system should not be taken to be 'archaic' or otherwise equal to the system of proto-Vietic.

2.4 Orthography

Given the high number of contrasts in the Kariì sound system, it is inevitable that a practical orthography will involve non-standard characters and digraphs. The full inventory of major-initial consonants (see Table 1, above) will be transcribed as in Table 11:

	bilabial	alveolar	retroflex	palatal	velar	laryngeal
voiced (implosive) stops	b	d		qj		
voiceless aspirated stops	ph	th	tzrh		kh	
voiceless unaspirated stops	p	t	tzr	c	k	q
fricatives		s			gh	h
nasal sonorants	m	n		ñ	ng	
oral sonorants	v	l	r	j		

Table 11. Initial consonants, orthographically

The finals will be transcribed as in Table 12 (cf. Table 2, above):

		bilabial	alveolar	palatal	(post) velar
stops	checked	p	t	c	k
	nasal	m	n	ñ	ng
sonorants	oral	w	l r	j	

Table 12. Final consonants, orthographically.

Finals with checked terminance listed in Table 5, above, are represented in the orthography with **q** after the appropriate sonorant (giving checked-terminant **-Vq**, **-mq**, **-nq**, **-ñq**, **-ngq**, **-wq**, **-lq**, **-rq**, and **-jq** in contrast to voiced terminant **-V**, **-m**, **-n**, **-ñ**, **-ng**, **-w**, **-l**, **-r**, and **-j**). Finals with voiceless terminance listed in Table 5, above, are represented in the orthography with **h** after the appropriate oral sonorant (giving voiceless-terminant **-Vh**, **-wh**, **-lh**, **-rh**, and **-jh** in contrast to voiced terminant **-V**, **-w**, **-l**, **-r**, and **-j**). This is a purely practical representation, and we remind the reader that our analysis does not recognize final glottal stop and **h** (here as **-q** and **-h**) to be distinct segments. Rather, they are outcomes of particular terminance values.

The basic nine vowel symbols are equivalent to their values in Quốc Ngữ script used for Vietnamese (Thompson 1987:18ff):

i	u	u
ê	o'	ô
e	a	o

Table 13. Basic vowel symbols, orthographically.

The two-way split in phonological register (heavy versus light) across the nine vowels is marked by means of a diacritic (a grave accent above the vowel) for heavy register. Light register is left unmarked, since it is statistically predominant (see Figure 4, above). So, the vowel "a" will appear as heavy à or light a. This derives 18 distinct vowel symbols, shown in Table 14:

ì	ùr	ù
i	ur	u
è	ò'	ò
ê	o'	ô
è	à	ò
e	a	o

Table 14. Basic vowel symbols, orthographically, split into two registers with grave accent marking heavy register, and light register unmarked.

Long vowels are marked with double vowel symbol (i.e., short **a** versus long **aa**). Diphthongs are always marked as heavy, since they are usually pronounced that way (as **ìa**, **ùà** and **ùrà**). In subsequent sections of this article, all Karì words will be written using this orthography.

3 Morphology

Karì has no inflectional morphology (i.e., person, number, gender, case), but does have a modest amount of essentially non-productive derivational morphology. The derivational morphology of Karì is more elaborate than is found in heavily isolating neighboring

languages like Lao or Vietnamese, but significantly less than in morphologically rich cousins like Kmhmu (Northern Mon-Khmer, Suwilai 1987) or Semelai (Aslian, Kruspe 2004). We note tantalizing correspondences in the lexicon such as *paañq* 'sell' versus *maañq* 'borrow', and *slaac* 'slip' versus *haac* 'slippery' – while these seem to suggest erstwhile morphological relatedness, if there is any such relation it is no longer generally visible or productive. Other cases look like clearer candidates for morphological derivation but we only have one case: e.g., *pu-* in *putêngq* 'greeting' (cf. *têngq* 'arrive'). Yet other cases resemble otherwise existing morphology, but the semantic relations are not consistent with those cases: e.g., *kpôôh* 'a handwidth (across palm, including thumb)' versus *kapôôh* 'to speed up rice-pounding when working as a team' (cf. the two *-a-* infixes described below). In the following section, we list a few of the more regular-looking derivational morphemes, all of which are infixes.

3.1 Infixes

Infixes in Kariì have particular consequences for syllable structure of the resulting word: they always result in the creation of a minor syllable. We discuss three cases: in the first two, the process of infixation is only accessible to words with C_0C onsets.

3.1.1 Causative infix *-a-*

The causative infix *-a-* can be inserted into words with C_0C onsets. The infix then occupies the v_{-1} slot and thereby creates a minor syllable. The erstwhile C_0 ends up in the new C_{-1} slot. The function of this infix is to turn an intransitive verb into a transitive verb, adding a controller/agent participant:

- (47) a. *trùùm* – to be on its face
tarùùm – to put something on its face
- b. *praang* – to cross over (e.g. a river)
paraang – to take someone across
- c. *slôôj* – to be washed away by flowing water
salôôj – to discard into flowing water, to let something be washed away
- d. *kliìh* – to go off (of a trigger or trap mechanism)
kaliìh – to set off a trigger or trap mechanism
- e. *bsot* – to go out (of light or fire)
basot – to turn off, put out

- f. *pdèèh* – to come untied
padèèh – to untie
- g. *kleer* – to fall off (e.g., a fruit, a leech)
kaleer – to pick off
- h. *krangq* – to be dried up, toasted
karangq – to toast, warm up
- i. *blang* – of the eyes, to become open (e.g. of a developing puppy)
balang – to open one's eyes

3.1.2 Verbalizing infix *-a-*

The verbalizing infix *-a-* is presumably related to the causative infix, just discussed. Further work is needed in order to clarify their relationship. This infix can be inserted into words with C_0C onsets. The infix then occupies the v_{-1} slot and thereby creates a minor syllable. The erstwhile C_0 ends up in the new C_{-1} slot. The function of the verbalizing infix *-a-* is to derive a verb from a noun. Most of our examples (48a-c) are from a series of forms derived from body measure terms (with the derivational pattern 'body measuring term' > 'action of measuring out using that body part'), but we also see verbalization from words for objects, as in (48d):

- (48) a. *ckaang* – a hand span
cakaang – to measure something by hand spans
- b. *plaajh* – an arm span
palaajh – to measure something by arm spans
- c. *cbuuc* – a unit of whatever can be picked up purse-hand (a 'pinch')
cabuuc – to pick up or measure out something purse-hand
- d. *ckhaal* – a diagonal cross-beam for support in base (e.g. in *cariit* 'back basket' or other woven artifact)
cakhaal – to place a diagonal cross-beam in base (e.g., of *cariit* 'back basket')

Both *-a-* infixes target only those words with C_0C onsets. For words with other types of onset ($CV(C_F)$ and $CvCV(C_F)$), the kinds of semantic relations which these infixes denote will have to be coded by other means (i.e., lexically or syntactically, as is the pattern for languages like Lao and Vietnamese).

3.2 Nominalizing infix *-rn-*

The nominalizing infix *-rn-* is inserted into words with simple C onsets. In the resultant syllable, the erstwhile C becomes C₋₁, and the sequence *-rn-* occupies the new V₋₁C slots. The function of this infix is to derive a noun from a verb:

- (49)
- a. *koorq* – to scoop out strips of a vegetable (e.g. cassava) for food preparation
krnoorq – an instrument to scoop out strips of a vegetable (e.g. cassava) for food preparation
 - b. *paang* – to signal one's ownership of a swidden
prnaang – a sign placed in a swidden to signal ownership
 - c. *sat* – to get one's foot stuck (e.g. in a snare)
srnat – a foothold
 - d. *kooq* – to live
krnooq – a house
 - e. *keep* – to pince
krneep – tongs, pincers
 - f. *tajh* – to tap, to strike
trnajh – a flint stone
 - g. *cooh* – to pierce
crnooh – a hung thread, pierced through multiple objects

For verbs which already have a pre-C consonant, the *-rn-* infix is reduced to *-r-*. The *-r-* infix then occupies the v₋₁ slot of the minor syllable, either by simply replacing the existing v₋₁ if the source form had a minor syllable, as in (50a-c), or by being inserted between C₀ and C of the source form, as in (50d), causing the erstwhile C₀ to end up in the C₋₁ slot, and creating a new minor syllable:

- (50)
- a. *kalêeng* – to shoulder-carry with an ad hoc constructed pole
krleêng – an ad hoc constructed shoulder-carrying pole
 - b. *kaleeq* – to point something out
krleeq – sign used for pointing out bee hives in trees
 - c. *kadôôlq* – to rest the head on something
krdôôlq – a pillow

- d. *tkap* – to pince
trkap – a pince

The 'pillow' case (50c) in particular shows that the *-r-* form is the derived one, since it has a very much more specific meaning (i.e., one can *kadôôlq* 'rest one's head on' anything, while a *krdôôlq* is something very specific for that purpose).

3.3 Reduplication

Reduplication is inherent to a good many verbs in the lexicon (by inherent we mean that there is no corresponding non-reduplicated form), where these all seem to involve repeated or prolonged action:

- (51) a. *kôôn-kôôn* – to shoulder-pole something, with only one side loaded
 b. *kuu-kuu* – to wash out the mouth and spit out, e.g. after eating
 c. *ceek-ceek* – to pick the teeth (as *ceek-ceek kasang*)
 d. *dec-dec* – to massage, repeatedly squeeze

In some cases, the root element may also be used on its own, denoting an event NOT repeated or prolonged.

- (52) a. *tanq* – to chop up (meat)
 a. *tanq-tanq* – to chop up (meat) into tiny pieces, finer than *tanq*

When the input form has a minor syllable, there is partial reduplication. Only the major syllable is doubled:

- (53) a. *careew* 'green'
 b. *careew-reew* 'greenish'

A number of lexical items show inherent reduplication, but where there is less clear semantic relation to notions like repetition or sustained action:

- (54) a. *boong-boong* 'window'
 b. *dii-diiq* 'small'

- c. *daa-daaq* ‘hurry’
- d. *doo-dooq* ‘sort of’
- e. *kôô-kôô* ‘tangled’
- f. *rôô-rôô* ‘large logs left over from burning swidden’
- g. *moo-moo* ‘follow, spy on’

4 Syntax

4.1 Constituent order

Basic constituent order appears to be 'SVO', but this may not be the best way of analyzing the syntax, since many other kinds of structures routinely occur. There is free ellipsis of nominals when the identity of their referents is known (i.e., ellipsis is not syntactically determined or controlled). Postposition and fronting of noun phrases is common. There is no case-marking and no cross-referencing/agreement. This means that a surface string will always have a verb, but links between arguments and predicates may be hard to determine from surface form alone. Like other languages of mainland Southeast Asia, Kariì requires a lot of dependence on context for resolution of grammatical relations.

4.2 Subordination and serialization

Kariì features an explicit marker of subordination (*kùù*) between verbs in series, where most neighboring languages would lack this, having bare serial verb constructions instead. Here are some examples:

- (55) *pdajh kùù vət*
 flick SUB discard
 ‘flicking (sth; e.g. a leech) off’
- (56) *qàvành kùù qooc*
 push.aside.vegetation.to.make.way.through SUB go
 ‘going (along) by pushing the vegetation aside to make way.’
- (57) *qalêng kùù cong*
 look SUB see
 ‘Look!’

- (58) *lêêq srdeel kùù sadeel*
 take a.plug SUB to.plug
 ‘taking a plug to plug (something)’
- (59) *qEEP kùù qalêêngq*
 be.quiet SUB look
 ‘Be quiet and look.’ (Mother to child, while looking at photographs.)
- (60) *qabeeh kùù tààjh*
 to.hook SUB pull
 ‘hooking something so as to pull it (towards oneself).’
- (61) *lêêq kùù paañq*
 take SUB sell
 ‘taking something (e.g. wildlife) so as to sell it.’
- (62) *romq kùù hanq ciinq*
 keep.to.ripen SUB 3SG ripe
 ‘putting aside green fruit so that it may ripen.’
- (63) *qEEP kùù têtèk*
 be.quiet SUB play
 ‘Be quiet and play.’ (Parent to child, asking child to play quietly.)
- (64) *lêêq kasêêq mee crnooh kùù cooh*
 take cord make tie.cord SUB thread
 ‘taking a cord to make a tie-cord, to thread it through (e.g. some fish).’

Adverbial type expressions in which one verb is a modifier of another do not involve the subordinator *kùù*. Two orders are observed:

- (65) a. *lêêq klêê* b. *klêê lêêq*
 take easy easy take
 ‘easy to take’ ‘easy to take’
- (66) a. *qooc lôômq* b. *lôômq qooc*
 go steal steal go
 ‘go secretly’ ‘secretly go’

Note that the subordinator *kùù* marks the difference between the adverbial constructions in (66) and a purposive construction in which two predicates are linked in temporal-conditional succession:

- (67) *qooc kùù lôômq*
 go SUB steal
 'go in order to steal (something)'

Sometimes verb-verb sequences are not marked with the subordinator, as in the following example where *sòòk tZRôôh* [seek-meet] may be a fixed idiomatic way of saying 'find':

- (68) *sòòk tZRôôh mleeng mee vêêk boo dêêh*
 seek meet person do work pcl neg
 'Have (you) found somebody to do the work?'

4.3 Three-place predicates

The following examples give a sense of how extra participants are packaged in the Kariì clause, showing patterns of verb serialization which closely resemble those found in neighboring languages (cf. Lao, Enfield 2007b). The Kariì case differs in showing the use of true (i.e., non-derived) prepositions (here, *quu* 'at' and *dêêwq* 'from'):

- (69) *book lêêq kààjh qùù tôôjq*
 scoop take stone loc bag/pocket
 'scooping/picking out a stone in pocket.' [PUT_Version_3_Tuut_2005_07_13_000138]
- (70) *pic sithiàn lêêq lôôh dêêwq nsoj*
 pull.out candle take exit from place
 'pulling the candle out from (its) place.' [PUT_Version_3_Tuut_2005_07_13_000217]
- (71) *lêêq kààjh *(lêêq) lôôh dêêwq vòòngq*
 take stone take exit from pot
 'taking a stone out from a pot.' [PUT_Version_3_Tuut_2005_07_13_000448]

(72) *lêêq lôôh kulôk dêêwq thang daak*
 take exit head from bucket water

‘taking the head out from a water bucket.’ [PUT_Version_3_Tuut_2005_07_13_000944]

(73) *qôôm-qôôm phap tabôôn qùù tôôq*
 carry.along.in.arms book place loc table

‘carrying along books (and) placing (them) on the table.’ [PUT_Version_3_Tuut_2005_07_13_001251]

5 Pronouns

Here is an inventory of Kariì pronouns:

	SG		DU		PL	
	BARE	POLITE				
		F M	INCL	EXCL	INCL	EXCL
1	<i>teeq</i>	<i>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 complex structures of the type PRONOUN+[TITLE+NAME], for referring to groups of people identified with a specific sub-member who is mentioned by name, as follows:

(74) *paa vòòk nôn*
 3pl grandpa N
 ‘they of whom Grandpa Non is one’

(75) *qaar qaj kham*
 3du non.resp.pref K
 ‘those two of which Kham is one’

(76) *ñaar saaw knin*
 1du.excl aunt K
 ‘we two (excl) of which the other is Aunt Knin.’

We do not yet understand the principles behind choosing who is the reference individual for the collective.

6 Kin terms and names

6.1 Kin terms

Karîi society is fixated on kinship. Everyone in this small community is related to everyone else, and addresses each other accordingly. There are several broad categories of kin, determining the correct pattern of pronoun use (for first and second persons), and of third person reference (specifically, determining whether you can utter the person's name or not):¹⁵

Kin category of addr	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. ySib or child)'	<i>pàànq</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 eSib'	<i>teeq</i> 1SG	kin term	No
6. <i>sdoong</i> ' <i>mree</i> of your own child' (i.e. someone to whom you are <i>mree</i> to their child) [reciprocal pronouns]	<i>khoojq</i> 1SG	<i>nôôq</i> 2SG	No

Note: For higher (i.e., older) *cià-maangq* you can't use *cak* as 2sg pronoun.

Table 15. Some kin categories, consequential for person reference.

¹⁵ If you "cannot use a person's name", then you refer to them by their first child's name (teknonymy); i.e. if a man's first child is named *kham* then he may be referred to as *pòòq qaj-kham* 'father of Kham', or simply may be directly referred to with the child's name, as long as there is a title prefix such as an appropriate classificatory kin term, or a title such as *qôông* 'Mr.' or *moo* 'Mrs'.

Here is a list of basic kin terms:¹⁶

(77)	F	-	<i>pòòq</i>
	M	-	<i>méêq</i>
	eB	-	<i>maangq</i>
	eZ	-	<i>cùq</i>
	yG	-	<i>qeem</i>
	FyZ	-	<i>qoo</i>
	FyB/HyB	-	<i>pòòq</i>
	MyB/WyB	-	<i>kùùq</i>
	MyZ	-	<i>muruq</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>
	CCC	-	<i>cec</i>
	CCCC	-	<i>celavec</i>
	CCCCC	-	<i>calavor</i>

Some of the terms are 'skewed', extending in reference across generations, and across classes of kin. In the (78a-b) examples, one counts relatives of one's spouse as if one's spouse were one's parent (or, as if taking one's own child's perspective):

(78)	a.	<i>taa</i>	=	[MeB, MFeBS] _{cià-maangq} , [WeB, WFeBS] _{mree}
	b.	<i>pòòq</i>	=	[FyB, FeBS, FFyBS] _{cià-maangq} , [HyB, HeBS] _{kmoon}
	c.	<i>matààm</i>	=	[DH, yZH, FyBDH]

¹⁶ Abbreviations are: F=father, M=mother, B=brother, Z=sister, G=sibling, H=husband, W=wife, C=child, y=younger, e=elder.

6.2 Titles as prefixes for address

When addressing people, names can be prefixed by the relevant kin term (e.g., *vòòk-nôn* 'grandpa Nôn'), or by another kind of title, such as the following:

(79) Titles

- a. *qôông* – ‘sir, Mr.’
- b. *mooq* – ‘ma’am, Mrs.’
- c. *sìàng* – ‘ex-novice’
- d. *caan* – ‘ex-monk’
- e. *lùùng* – ‘uncle’ (Lao)
- f. *màjòòq* – ‘the late’: a prefix for any dead person ‘deceased’; also a term of insult for living people

7 Elaborative language

7.1 Echo-formation

There are lexically specified elaborative couplets, where a standard word has a lexically specified double (an 'echo-formative'), which appears nowhere else but in the couplet:

- | | | | |
|------|----|----------------------|--------------------------------------|
| (80) | a. | <i>rvaajq</i> | ‘spirit/soul’ |
| | | <i>rvaajq rlôông</i> | ‘spirit/soul and that sort of thing’ |
| | b. | <i>thừang</i> | ‘stuff’ |
| | | <i>thừang thừàw</i> | ‘stuff and that sort of thing’ |
| | c. | <i>pừn</i> | ‘grass’ |
| | | <i>pừn pòòt</i> | ‘grass and that sort of thing’ |
| | d. | <i>hưrt</i> | ‘tobacco’ |
| | | <i>hưrt hooj</i> | ‘tobacco and that sort of thing’ |

Less often, we find echo-formatives for verbs (with less predictable semantics):

- | | | | |
|------|----|-----------------|------------------------|
| (81) | a. | <i>rooh</i> | ‘dry’ |
| | | <i>rooh rap</i> | ‘sort of dry, dry-ish’ |

- a. *sangòòc* 'quiet'
sangòòc sangeec 'dead quiet'

7.2 Expressives

The expressive word class is found in most Eastern Mon-Khmer languages (Diffloth 1972), and in many Southeast Asian languages more generally. These are phonologically elaborate, 'sound symbolic' in nature, and tend to convey narrowly defined perceptual experience. As the following examples show, they often occur with a more everyday class of expression (e.g., a verb) over which they have descriptive scope:

- (82) a. *cùng* 'raining'
cùng phiñ-phiñ 'raining very lightly, drizzling, unceasingly'
- b. *sphiiq* 'all different, mixed up'
sphiiq lang-ñang 'all different, mixed up, very much so'
- c. *ñông* 'bow down'
ñông khununuh 'bow down with back bent over'
- d. *tùngq* 'stand'
kdong-doongq 'stand leaning forward with bum sticking out'
- e. *satùm* 'dark'
satùm ngiñq 'dark, so dark that you literally can't see a thing'
- f. *thòng thèl* 'swinging big'
- g. *teew-ngeewq* 'with head to side'
- h. *tool-qoolq* 'head tilted back, looking up, chest out'
- i. *vòò vòòh* 'shaking a torch up and down as one makes one's way through the dark'
- j. *vee veeh* 'shaking a tiny torch side to side (or up and down) as one makes one's way through the dark'
- k. *vông vajh* 'shaking a torch around in large circles as one makes one's way through the dark'

These are distinct from echo-formatives (see above).

8 Nominals

8.1 Demonstratives

There are five distinctions in exophoric uses of demonstratives (i.e., references to physically present persons, places, or things):

- (83)
- a. *nì* general ('this')
 - b. *naaq* external ('that', away, far)
 - c. *seeh* distal ('yon', across, far)
 - d. *còòh* external down, downstream, below
 - e. *lêeh* external up, upstream, above

At least (83a-c) have endophoric uses (i.e., discourse and anaphoric uses). In addition, there are two further demonstrative-type items, which only have endophoric functions:

- (84)
- a. *tààh* proximal
 - b. *nààq* external (manner) 'thus'

Demonstratives may occur as complements to nominal heads, postposed:

- (85) *longq naaq*
 CLF dem.external
 'that one'

Demonstratives may occur as adverbs, postposed:

- (86) *cìih quu tìk còòh*
 descend loc below dem.down
 'to go down to the ground'
- (87) *qooc qu trààh lêeh*
 go loc swidden dem.up
 'to go up to the swidden'

- (88) *mòòj phàn seeh*
 ten thousand dem.yon
 ‘ten thousand’

The term *tààh* occurs in a number of contexts which suggest greater productivity previously: e.g., *kòòh-tààh* ‘today’, *qatààh-tààh* ‘just this minute’, and the common sentence-final particle *tààh*.

8.2 Possession

Possession may be explicitly marked by the classifier *longq*, but need not be:

- (89) *cluu (longq) maangq teeq*
 buffalo (CLF) eB 1SG
 ‘my elder brother’s buffalo’

8.3 Modifiers

Modifying elements are postposed:

- (90) *tukùùp qjòòñ*
 hill high
 ‘tall hill’

8.4 Classifiers

There are constructions which look like numeral classifier constructions, well described for mainland Southeast Asian languages. In Kariì, as in many languages of the area, multiple orders are possible: both NUM+CLF+N (Sinitic/Vietnamese Pattern, as in (91)) and N+NUM+CLF (Tai pattern, as in (92)) are possible:

- (91) *haar longq kadeeq*
 two CLF child
 ‘two children’
- (92) *kadeeq haar longq*
 child two CLF
 ‘two children’

9 Verbal marking

9.1 Negation

There are three forms for negation:

- (93) a. *dêêh* b. *laa* c. *cìrì*

These particles show distinct grammatical behavior. The form *cìrì* is distinct from *dêêh* and *laa* in that it appears before the subject, not immediately before the verb:

- (94) a. *cìrì hanq kooq*
 NEG 3SG have
 ‘S/he doesn't have (any).’
 b. **hanq cìrì kooq*
 3SG NEG have
 (S/he doesn't have (any).)
- (95) a. *hanq dêêh/laa kooq*
 3SG NEG have
 ‘S/he doesn't have (any).’
 b. **dêêh/laa hanq kooq*
 NEG 3SG have
 (S/he doesn't have (any).)

The form *dêêh* can be distinguished from *laa* in that *dêêh* may appear on its own as a full utterance (e.g., as a simple 'no' answer to a polar question), while *laa* may not:

- (96) Q *cak qooc boo*
 2SG go PCL
 ‘Are you going?’
 A1. *dêêh* A2. **laa*
 NEG (NEG)
 ‘No.’

9.2 Time, aspect

Here is a set of common modifiers of temporal, aspectual, and modal meaning in the verb:

- (97) a. *V kadooj* ‘often, all the time’
 b. *kadang V* ‘just V-ed’
 b'. *kanaq V* ‘just V-ed’ (also sometimes *khanaq*)
 c. *V cdang* ‘only V’
 d. *thòòm V* ‘progressive’
 e. *khòòj V* ‘have ever’
 f. *V qà-tàh-tàh* ‘just then’
 g. *V baramq* ‘keep doing V like that’
 h. *V pù tùj* ‘V as before’
 i. *samêng V* ‘almost V-ed’
 j. *V lòj* ‘V without ado’
 k. *manêêt V* ‘probably V’

A very common general predicative marker is the particle *ci*, which appears before the verb, and often before the subject (if there is pre-subject material, suggesting a kind of ‘second position’ placement):

- (98) *teeq ci dêêh ruuq qa-ni*
 1SG PRED NEG know nom-this
 ‘I don’t/wouldn’t know (about) this.’
- (99) *qa-ni ci teeq dêêh ruuq*
 nom-this PRED 1SG NEG know
 ‘This, I don’t/wouldn’t know (about).’

9.3 Adjectives

Adjectives behave mostly like verbs, but can be distinguished as a sub-class by their inaccessibility to a special emphatic construction involving *tôô tàn* [be true]. Both verbs and adjectives may take *tàn* 'true, really' as an intensifier:

(100) *vààl* *tàn*
 move.around true
 'really moving around'

(101) *qjòòñ* *tàn*
 tall true
 'really tall'

However, only verbs allow the more elaborate expression *tôô tàn*:

(102) *vààl* *tôô tàn*
 move.around be true
 'really moving around'

(103) **qjòòñ* *tôô tàn*
 tall be true
 (really tall)

10 Final Remark

This brings to an end our introductory report on Kariì, a previously undocumented Vietic language of Laos. We hope to have made a contribution not only to what is known about this language, but also to what is known about this branch of Mon-Khmer. Current knowledge of Vietic is dominated by work on Vietnamese (and Muong to a lesser extent). As is the case for many language families, the best described members are often the least representative in typological terms. This is true for heavily Tai-ized Muong, and could hardly be more dramatic than in the case of Vietnamese, thanks to 1000 years of Chinese influence, and a long period of national language status in addition (Nguyen 1980). Most other members of the Vietic branch, spoken by upland minorities in Laos and Vietnam (cf. Ferlus 1996), are closer in typological structure to Kariì. Kariì does, however, display a number of apparently

exceptional features for a Vietic language, including full sets of terminance contrasts with de-voicing and post-glottalization, a robust final **-r**, an initial palatal implosive **ɟ**-, unusual combinations in initial clusters (e.g., **ɟ^h**- and **ɲ^h**-), a lack of lexical tone, a set of derivational morphemes, an explicit marker of subordination in verb serialization, and a pre-sentential negation marker. In order to judge the degree to which these and other features of Kariì are exceptional, much more work needs to be done on Kariì and related languages. Minority languages of the Vietic group remain seriously endangered, and in urgent need of description.

References

- ADB/UNEP, 2004. *Greater Mekong Subregion Atlas of the Environment*. Asian Development Bank (ADB), Manila, Phillipines, and United Nations Environment Programme (UNEP) Regional Resources Centre for Asia and the Pacific, Pathumthani, Thailand. Manila/Pathumthani: ADB/UNEP.
- Alves, M. J. 1995. Tonal features and the development of Vietnamese tones. *Working papers in Linguistics, University of Hawaii at Manoa* 27:1-13.
- Chamberlain, James R. 1997. *Nature and culture in the Nakai – Nam Theun Conservation Area*. Unpublished manuscript, Vientiane 1997.
- Costello, Nancy. 1993. *Katu folktales and society*. Vientiane: Institute of Research on Lao Culture and Society.
- Davis, John J. 1968. *Nyaheun phonemes*. Overseas Missionary Fellowship. (Cited in Sidwell and Jacq 2003.)
- Diffloth, G. 1972. Notes on expressive meaning. *Chicago Linguistics Society* 8:440-447.
- Diffloth, G. 1974. The KmMu' principle. In *Parmentier Festschrift*, ed. E. Hamp. Chicago: University of Chicago Press.
- Diffloth, G. 1984. *The Dvaravati-Old Mon Language and Nyah Kur*. Bangkok: Chulalongkorn University Press.
- Enfield, N. J. 1999. "Lao as a national language," in *Laos: Culture and Society*. Edited by G. Evans, pp. 258-290. Chiang Mai: Silkworm Books.
- Enfield, N. J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual review of anthropology* 34:181-206.

- Enfield, N. J. 2007a. Everyday ritual in the residential world. (Paper delivered at Wenner-Gren conference on "Ritual Communication", Sintra, Portugal, March 2007. To be published in proceedings, Berg Press.)
- Enfield, N. J. 2007b. Encoding three-participant events in the Lao clause. *Linguistics* 45.3, 509-538.
- Enfield, N. J. in press. *A grammar of Lao*. Berlin: Mouton de Gruyter. (2007)
- Ferlus, Michel. 1996. Langues et peuples Viet-Muong. *Mon-Khmer Studies*, 26:7-28.
- Ferlus, M. 1997. Problèmes de la formation du système vocalique du vietnamien. *Cahiers de Linguistique Asie Orientale* 26.
- Ferlus, M. 1998. Les systèmes de tons dans les langues viet-muong. *Diachronica* 15.1, 1-27.
- Gage, W. W. 1985. "Vietnamese in Mon-Khmer perspective," in *Southeast Asian Linguistic Studies presented to Andre-G. Haudricourt*. Edited by S. Ratanakul, D. Thomas, and S. Premssirat, pp. 493-524. Bangkok: Mahidol University.
- Gedney, W. J. 1993. *The Saek language: glossary, texts, and translations*. Michigan: Center for South and Southeast Asian Studies.
- Gregerson, K. J. 1976. "Tongue-root and register in Mon-Khmer," in *Austroasiatic studies, Part I*. Edited by P. N. Jenner, L. C. Thompson, and S. Starosta, pp. 323-369. Honolulu: University of Hawaii Press.
- Haudricourt, A.-G. 1946. Propagation phonétique ou évolution phonologique? Assourdissement et sonorisation d'occlusives dan l'Asie du Sud-Est. *BSLP* 43:82-92.
- Haudricourt, A.-G. 1953. La place du vietnamien dan les langues austroasiatique. *BSLP* 49:122-8.
- Haudricourt, A.-G. 1954. De l'origine des tons en viêtnamien. *Journal Asiatique* 242:69-82.
- Haupers, R. 1979. *Stieng-English Dictionary*. Dallas: SIL Publications.
- Henderson, E. 1952. The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies* 14:149-174.
- Henderson, E. 1965. The topography of certain phonetic and morphological characteristics of South East Asian languages. *Lingua* 15:400-434.
- Huffman, F. E. 1970. *Modern spoken Cambodian*. Ithaca, NY: Cornell University, Southeast Asia Program.
- Kruspe, N. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press.
- Maspero, H. 1912. Phonétique historique de la langue annamite: les initiales. *Bulletin de l'Ecole Française d'Extrême-Orient* 12:1-27.

- Matisoff, J. A. 1973. "Tonogenesis in Southeast Asia," in *Southern California occasional papers in Linguistics, No. 1*. Edited by L. M. Hyman, pp. 72-95. Los Angeles: University of Southern California.
- Nguyen, D. H. 1980. *Language in Vietnamese society*. Carbondale, Illinois: Asia Books.
- Norman, J. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Quy, T.V., 2002. The Quy Hop archive: Vietnamese-Lao relations reflected in border-post documents dating from 1619 to 1880. In: M. Ngaosrivathana and K. Breazeale (Editors), *Breaking new ground in Lao history: essays on the seventh to twentieth centuries*. Silkworm Books, Chiang Mai, pp. 239-260.
- Ramsey, S. R. 1987. *The languages of China*. Princeton: Princeton University Press.
- Rivière, Captain. 1902 'Voyages au Laos et en Annam', pp215-294 of *Mission Pavie Indochine 1879-1895 – Géographie et voyages, IV Voyages au Centre de l'Annam et du Laos et dans les régions sauvages de l'est de l'Indo-chine*, by Captain De Malglaive and Captain Rivière. Paris: Ernst Leroux, 1902.
- Sapir, E. 1921. *Language: an introduction to the study of speech*. Orlando/San Diego/New York/London: Harcourt Brace Jovanovich.
- Schliesinger, Joachim. 2003. *Ethnic groups of Laos. Volume 2. Profile of Austro-Asiatic-Speaking Peoples*. Bangkok: White Lotus.
- Shorto, H. 2006. *A Mon-Khmer comparative dictionary*. Canberra: Pacific Linguistics.
- Sidwell, P. 2005. *The Katuic Languages: classification, reconstruction and comparative lexicon*. Munich: Lincom Europa.
- Sidwell, P., and P. Jacq. 2003. *A Handbook of comparative Bahnaric: Volume 1, West Bahnaric*. Canberra: Pacific Linguistics.
- Suwilai Premsrirat. 1987. *Khmu, a minority language of Thailand*. Canberra: Pacific Linguistics.
- Swantesson, J.-O., D. Tayanin, K. Lindell, T. Kingsada, and S. Sayavong. 1994. *Khmumu-Lao dictionary*. Vientiane: Ministry of Public Health.
- Theraphan L-Thongkum. 2001. *Languages of the tribes in Xekong province, southern Laos: a foundation for research and development*. Bangkok: The Thailand Research Fund.
- Thompson, L. C. 1987(1965). *A Vietnamese grammar*, (Second edition.) edition. Honolulu/Seattle: University of Hawaii Press/ University of Washington Press.
- Thurgood, G. 2002. Vietnamese and tonogenesis: revising the model and the analysis. *Diachronica* 19:333-363.