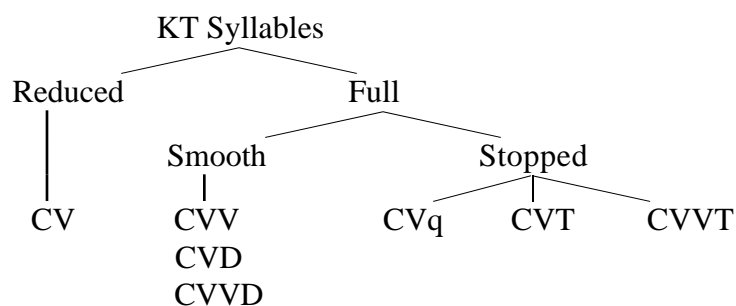


Chapter 2.

SEGMENTAL PHONOLOGY

2.1. Syllable structure

As seen in Chapter 1, words and morphemes are largely monosyllabic in KT. This is true both of lexical morphemes (nouns, verbs etc.) and grammatical morphemes (case, tense markers etc.). The different types of KT syllables are categorized according their rime. As schematized below, a syllable may be reduced or full, and the latter may be either smooth or stopped:



(C = consonant, V = vowel, D = sonorant, q = glottal stop, T = p or t)

Since each syllable is a morpheme or word, the following terminological distinctions will also be adopted: A **LEXICAL MORPHEME** refers to a member of any of the major category word classes: noun, verb, adjective, numeral. A **GRAMMATICAL MORPHEME** refers to any of the minor category word classes or particles which are typically limited in number: pronouns, demonstratives, tense markers etc. All lexical morphemes are phonological words, as are grammatical morphemes which are both full syllables and have a consonant onset. On the other hand, a preposed grammatical morpheme whose shape is CV or V will be referred to as a **PROCLITIC**, while a postposed grammatical morpheme whose shape is VV or VC will be referred to as an **ENCLITIC**. Illustrations of all three types of grammatical morphemes are seen in the sentence *à néé hlòn êe* ‘the two of them are eating’. The proclitic *à* is a third person subject marker, while the enclitic *êe* is the declarative marker. The grammatical morpheme *hlòn* marks the dual in KT. As seen from its CVD shape, it is both a grammatical morpheme and a phonological word.

2.1.1. Reduced syllables.

CV syllables which contain a short vowel and no coda are viewed as reduced because of their limited distribution. The person markers which occur before nouns and verbs all have this shape: *ká zòol* ‘my friend’, *ná zòol* ‘your friend’, *á zòol* ‘his/her friend’; *ká kàp êe* ‘I am crying’, *ná kàp êe* ‘you are crying’, *á kàp êe* ‘s/he is crying. The reflexive marker *ki* is another proclitic: *á kì mùu êe* ‘he sees himself, he is seen’ (cf. §XX). In addition, a CV syllable may be result from a vowel-shortening rule which applies to one class of smooth CVV syllables (§XX): *vâa* ‘bird’, *vá lièn* ‘big bird’; *zûu* ‘mouse’, *zú nù* ‘two mice’; *á hlùu êe* ‘s/he is falling’, *á hlù hlón !ê* ‘they (dual) are

falling’. Although vowel length on CVV vs. CV syllables is predictable in most cases, there are contrasts such as the following on the surface:

with vowel shortening: ká mù hlón ¹ê ‘we (excl, dual) see him/her/it’
 without vowel shortening: ká mùu hlón ¹á hii êe

Both of these sentences involve the verb *mùu* ‘see’ and are in most respects synonymous (see §XX).

The preposed bimorphemic demonstratives consist of a CV proclitic followed by a derived CV syllable *tsíe* *tsye*, while the postposed monomorphemic demonstratives consist of one CV syllable each:

	preposed	postposed	
near speaker	hítsye	hí	‘this’
near hearer	tsútsye	tsú	‘that’
far from both	khútsye	khú	‘that’

While the latter may be considered clitics, they are optionally long (híi, tsúu, khúu) when occurring medially in a sentence: *sáa ¹hí (~ ¹híi) á phà pòò êe* ‘this meat is not good’. The postverbal morpheme *ná*, which is both a nominalizer and instrumental/locative marker, has been found to occur only with a short vowel before pause: *á èn ná* ‘greenness’, *á nùu ná* ‘what he laughs with, where he laughs’ (cf. §XX). This morpheme is involved in several nominalizations, e.g. *lùpná* ‘bed’ (lit. place to lie down), *thìl zúoq nă* ‘market’ (lit. place to sell things), *thìná* ‘death’ (lit. manner of dying), *pà tsàn nâ* ‘success’ (lit. honorable distinction).

2.1.2. Smooth syllables.

Syllables are termed “smooth” which either have a long vowel (CVV) or end in a sonorant consonant (D) preceded by either a long or short vowel (CVD: *hlàa* ‘wing’, *hlàm* ‘untie’, *hlàm* ‘prune (branch)’. Vowels in lexical entries are redundantly long, as are the two diphthongs *ie* and *uo*:

Cii	tsii	‘salt’	thii	‘die’
Cee	bêe	‘bean’	nêe	‘eat’
Cuu	mũu	‘hawk’	gũu	‘steal’
Coo	tôo	‘bottom, end’	tsôo	‘buy’
Caa	ngâa	‘fish’	zâa	‘hear, smell’
Cie	lie	‘shadow’	tsiè	‘go, walk’
Cuo	khùo	‘foot, lower leg’	bũo	‘spill’

Note, however, that when diphthongs undergo the rule of vowel shortening (§XX) they are written *ye* and *wo*: *hítsyé tsápáng ¹hí* ‘this child’, *khwò hòo* ‘feet’ (cf. *khùo* ‘foot’).

The sonorant consonant coda of smooth syllables can be a nasal (m, n, ng [ŋ]), the liquid *l*, or a glide (w, y):

coda	CVC		CVVD	
-m	kǎm	‘mouth’	lām	‘dance’
-n	màn	‘price’	vàn	‘sky’
-ng	lǔng	‘heart’	kêeng	‘leg’
-l	mòl	‘stick’	zòol	‘friend’
-w	nôw	‘seedling’	thàaw	‘oil, fat’
-y	lěy	‘tongue’	hùuy	‘air, wind’

Long and short vowels contrast before sonorant consonants, thereby producing many minimal pairs:

coda	CVC		CVVD	
-m	hàm	‘wheat’	hām	‘language’
-n	bǎn	‘cut’	bāan	‘arm’
-ng	dǒng	‘collect’	dǒong	‘young, soft’
-l	hêl	‘stir, be uneasy’	hêel	‘woo’
-w	lôw	‘pick (fruit)’	lǒow	‘medicine’
-y	hlǎy	‘run’	hlāay	‘chew’

The diphthongs *ie*, *uo* are redundantly long before sonorant consonants:

-m	liem	‘welcome’	tsǔom	‘different’
-n	lièn	‘big’	hûon	‘cook, boil’
-ng	hìeng	‘yes’	sûong	‘stone, rock’
-l	kǐel	‘hungry’	mǔol	‘hill’
-y			ngûoy	‘droop, wilt’

There is one exception to this: When verbs with a *uoy* rime occur in their stem2 form, *uoy* shortens and is written *woy*:

stem1	stem2		stem1	stem2	
gùoy	gwòy	‘hire’	ngûoy	ngwòy	‘droop’
vũoy	vwòy	‘wrinkle’	nũoy	nwòy	‘rub (against)’

This vowel shortening is due to an earlier glottalization which affected -y in stem2, but has since been lost (§XX). A short diphthong also occurs in the interrogative word *kwòy* ‘who’.

2.1.3. Stopped syllables.

Syllables which end in an unreleased *p*, *t*, or *q* (glottal stop) are referred to as “stopped”:

	CVC		CVVD	
-p	nàp	‘nasal mucus’	khûup	‘knee’
-t	khùt	‘hand’	môot	‘banana’
-q	vòq	‘pig’		

Vowel length is contrastive before *p* and *t*. Minimal pairs include the following:

	CVC		CVVD	
-p	kàp	‘cry’	kàap	‘shooting’
	zèp	‘swim’	zèep	‘whipping’
-t	pàt	‘thread’	pâat	‘lake’
	pèt	‘bite’	pèet	‘going astray’

The diphthongs *ie* and *uo* are redundantly long before *p*, *t*: *phîet* ‘demolish’, *tsûop* ‘lung’, *khûot* ‘scratch’.

Vowels are necessarily short before *q*: *tûq* ‘hold in hand’, *gìq* ‘heavy’, *thàq* ‘itch’. This is true of vowels whose Vq rime derives Pre-KT *VVr or *VVk, as the following comparison with the Lai spoken in Hakha shows:

	KT	Hakha	
*VVr	păq	pâar	‘bloom, blossom’
	khâq	khâar	‘close’
	ăq	âar	‘chicken’
*VVk	nâq	hnàak	‘rib’
	vâq	vàak	‘crawl’
	ngâq	hngàak	‘wait’

The diphthongs *ie* and *uo* are also predictably short before *q*. These therefore can be written either *ieq/uoq* or *yeq/weq*:

ie:	zĭeq	=	zyĕq	‘reason’	lĭeq	=	lwĕq	‘lick’
uo:	sûoq	=	swôq	‘slave’	lûoq	=	lwôq	‘vomit’

The coda consonant *-q* differs from *-p* and *-t* in not allowing a length contrast in the preceding vowel. A second difference has to do with tone: a *CVq* syllable may carry any of the three lexical tones of KT, while *CVT* or *CVVT* are more limited (§XX). When *CVq* carries a rising tone, the vowel is slightly lengthened: *văq* [vă:ʔ] ‘feed’ vs. *à vâq* [vâʔ] *êe* ‘he is feeding (it)’.

2.1.4. Onsets.

Onsets, including those written with a digraph (*th*, *hl* etc.), consist of a single consonant. Thirty-nine lexical entries have been found which lack an onset:

shape	#	noun	verb
VV	11	ôo ‘voice’	ûu ‘yelp’ (puppies)
VD:	11	în ‘house’	ûm ‘exist’
		ûy ‘dog’	ěng ‘green’
VVD:	7	ôom ‘chest’	ôom ‘boil (tr.)’
		ùul ‘perspiration’	ùuy ‘stale’
Vq:	9	ăq ‘chicken’	ìq ‘belch’
VT:	Ø	---	---
VVT:	1	---	âat ‘cut meat, slash’

These show all of the rime possibilities except VT, although only one example of VVT has been identified. On the surface, onsets are usually required of KT syllables unless the preceding syllable ends in a long vowel. Words such as *ôo* ‘voice’, *în* ‘house’, and *ăq* ‘chicken’ have a weak glottal stop when occurring after pause or after a consonant. The glottal stop is particularly audible when preceded by a short vowel, e.g. *kà* [ʔ]ôo ‘my voice’, *vá* [ʔ]în ‘bird house’. When preceded by a long vowel, however, there is a smooth transition from one vowel to the next: *mèe* *ěng* ‘green curry’, *bùu* *în* ‘rice house’ (§XX). Since its presence is predictable, there is no need to write *q* in onset position.

2.2. The consonant system

KT distinguishes 22 consonant sounds, as indicated the following table:

	onset consonants				coda consonants		
voiceless unaspirated stops	p	t	k	(ʔ)	p	t	ʔ
voiceless aspirated stops	p ^h	t ^h	k ^h				
voiced stops	b	d	g				
voiceless affricate		ts					
voiceless fricatives	s	ʃ	h				
voiced fricatives	v	z					
approximants		l			w	l	y
nasals	m	n	ŋ		m	n	ŋ

As seen, counting the glottal stop, 20 consonants appear as syllable onsets, while nine appear as codas. The rows represent manner of articulation, while the columns refer to place, e.g. labial, alveolar, velar, and laryngeal, in the case of onset consonants.

There is one compound word with a voiceless nasal *hn*: *zàahnñi* ‘yesterday’, derived from *zàan* ‘night’ + *h* + *ñi* ‘day’.

2.2.1. Onset consonants.

The above table shows the broad phonetic properties of KT consonants. Orthographic conventions differing from the table are presented in this section.

2.2.1.1. Stop onsets. There is an opposition between unaspirated and aspirated stops in KT. Aspirated stops are written as *ph*, *th*, *kh*:

	unaspirated			aspirated	
p-	pěq	‘wag’	ph-	phěq	‘braid’
	pâat	‘lake’		phâat	‘time’
t-	tûu	‘now’	th-	thûu	‘word’
	tòw	‘go upwards’		thòw	‘fly’ (n.)
k-	kàa	‘burn’ (intr)	kh-	khàa	‘body lice’
	kôq	‘point’ (v.)		khôq	‘peel off’

As seen in the following triplets, unaspirated and aspirated stops also contrast with voiced stops:

p/ph/b		t/th/d		k/kh/g	
păa	‘thin’	tâa	‘about to’	kòy	‘put (down)’
phàa	‘good’	thâa	‘strength’	khòy	‘touch’
băa	‘borrow’	dâa	‘divorce’	gòy	‘grate against’
pûon	‘cloth’	tũm	‘aim’	kõong	‘basket’
phûon	‘revealing’	thũm	‘three’	khõong	‘weave into cloth’
bûon	‘wrestle’	dũm	‘blue’	gõong	‘skinny’

It should be noted that the aspirated velar stop can be pronounced [k^h] or [kx^h] and is in the process of becoming a velar fricative [x].

2.2.1.2. Fricative and affricate onsets. There are no palatal stops or fricatives in KT. There is a single affricate *ts* which could be argued to occupy the “ch” slot. Some speakers do in fact palatalize the affricate, *tsápáng* ~ *chápáng* ‘child’ and also show variation between [z], [dz], and occasionally [dʒ], e.g. *zûng* ~ *dzûng* ~ *jûng* ‘root’. We write *ts* and *z*, which were the most commonly recorded.

Although KT does not have *f*, it does have the voiced counterpart *v*, e.g. *vâa* ‘bird’, *věe* ‘look at’. KT has three voiceless fricatives: *s*, *hl*, and *h*. What is written *hl* is a voiceless lateral fricative [ɬ]. All three are quite common in the language. The following illustrates minimal pairs among involving these fricatives and the one affricate *ts*:

s/h		s/hl		h/hl	
săm	‘curse’ (v.)	sõom	‘summon’	hǎang	‘fierce’
hǎm	‘scoop out’	hlõom	‘few’	hlǎang	‘mountain’
h/l		ts/z		v/z	
lôw	‘field’	tsǎang	‘bangles’	vâa	‘bird’
hlôw	‘cultivate’	zǎang	‘plain’ (land)	zâa	‘hear’

2.2.1.3. Approximant onsets. The only approximant onset is *l*, which was seen to contrast with *hl* in the previous table. Other examples: *hlèq* ‘very’, *lèq* ‘with’; *hlêem* ‘cheat’, *lêem* ‘fake’; *hlîi* ‘sap’, *lîi* ‘four’. The consonants *w* and *y* do not occur in onset position. What is sometimes written *wo* is actually a short [ɤo] diphthong (§XX): *kwòy* ‘who’, *vwòy* ‘wrinkling’.

2.2.1.4. Nasal onsets. KT has three nasal onsets: *m*, *n*, *ng*. The digraph *ng* stands for a velar nasal [ŋ]. The following examples show the three nasals contrasting in onset position:

mǎa	‘front’	měq	‘mould’	mǎay	‘pumpkin’
nǎa	‘eardrum’	nêq	‘eating’	nǎay	‘pus’ (brown)
ngǎa	‘five’	ngèq	‘urge’	ngǎay	‘sap’

2.2.2. Coda consonants.

Nine consonants appear in coda position of a KT syllable.

2.2.2.1. Stop codas. The stops *p*, *t*, *q* appear contrastively in words such as following:

pàp	‘greedy’	sõop	‘wash (things)’	khùp	‘put upside down’
pàt	‘thread’	sôot	‘long’	khùt	‘hand’
pǎq	‘flower’	sûoq	‘appear, happen’	khôq	‘peel off, skin’

There is no aspiration contrast among stops in coda position. Instead, stops in coda position are voiceless, unaspirated, and unreleased. Except for *w* coda consonants do not resyllabify with a following vowel. Related to this fact, *p* and *t* optionally acquire a nasal release when a vowel immediately follows: *á kàp êe* ~ *á kàp^m êe* ‘he is crying’, *á pèt êe* ~ *á pèt^m êe* ‘it bit him’. This nasal release is blocked by any consonant, including epenthetic glottal stop: *pàt* [ʔ]əng ‘green thread’.

A comparison with closely related languages, e.g. Hakha Lai, reveals that the glottal stop derives from an earlier velar **k* or alveolar **r*, which first became velar, then glottal stop:

KT	Hakha		KT	Hakha	
thàq	thàk	‘be itchy’	thâq	thâr	‘new’
khôq	khòk	‘peel off’	khâq	khâar	‘close’
lêq	lìak	‘lick’	thũoq	tûar	‘suffer’

On the other hand, historical glottal stops have dropped out, as can be seen in the following comparisons, again with Hakha Lai:

KT	Hakha		KT	Hakha	
khûu	khǔq	‘cough’	nìi	hñiq	‘two’
khàa	thráq	‘body lice’	bùu	bǔq	‘rice’
gùu	rǔq	‘bone’	nàa	hnǎq	‘leaf’

What this means is that current CVV syllables come from two pre-KT sources: *CVV and *CVq. This has consequences for stem2 formation and the process of vowel shortening (§XX).

2.2.2.2. Approximant codas. All three consonants, *l*, *w*, *y*, can appear in coda position, as seen in the following contrasts:

gòl	‘line’	kòl	‘hug’	nǎal	‘slippery’
gòw	‘property’	kòw	‘call’	nǎaw	‘baby, offspring’
gòy	‘grate against’	kòy	‘keep’	nǎay	‘pus’ (brown)

Coda *w* usually acquires a *v* release when followed by an enclitic. *á kòw êe* ‘he is calling’ *á kòwv êe*, which may simplify to *á kòv êe* (cf. §XX).

2.2.2.3. Nasal codas. All three nasals contrast in coda position:

tâm	‘many’	kõom	‘corn husk’	zâam	‘flee’
tân	‘cut’	kõon	‘bent’	zâan	‘night’
tǎng	‘lizard’	kõong	‘basket’	zâang	‘light’ (weight)

It should be noted that short /e/ is in the process of merging with short /a/ before the coda consonant /n/: *ên ~ ân* ‘rice, food’, *sěn ~ sǎn* ‘red’.

2.3. The vowel system

As shown below, there are five orthographic vowels and two diphthongs in KT, which may appear either short or long. Long vowels are written as double:

Short vowels		Long vowels		Diphthongs			
i	u	ii	uu	ye	wo	ie	uo
e	o	ee	oo				
	a		aa				

2.3.1. Monophthongs

The rows in Table 2 refer to vowel height (high, mid, low). The vowels *i*, *ii*, *e*, *ee*, *ie* are front unrounded, while *u*, *uu*, *o*, *oo*, *uo* are back and rounded. The vowels *i*, *ii*, *u*, *uu*, *aa* have their usual phonetic values. While long *aa* is pronounced [a:], short *a* is realized as [ʌ] except before glottal stop, e.g. *lùpná* [lùpnʌ] ‘bed’, *kàp* [kʌp] ‘cry’, *thàt* [tʰʌt] ‘kill’, *kâl* [kʌl] ‘kidney’, *hlây* [ʰʌy] ‘run’, *gǎm* [gʌm] ‘land’, *màn* [mʌn], *mâng* [mʌŋ] ‘dream’, vs. *sàq* [sàʔ] ‘north’. The mid vowels *e*, *ee*, *o*, and *oo* are also pronounced open and lax: [ɛ], [ɛ:], [ɔ], [ɔ:].

2.3.2. Diphthongs

The vowels *ie* and *uo* vary between the diphthongized pronunciations [ie] and [uo] and the close mid vowel pronunciations [e:] and [o:]: *khìe* [kʰie] ~ [kʰè:] ‘fall’, *khùo* [kʰuo] ~ [kʰò:] ‘foot’. The same is true of their short variants: [tʰiěʔ] ~ [tʰéʔ] ‘sweep’, [tʰüoʔ] ~ [tʰòʔ] ‘endure’. Because of this latter variation, the vowel system of KT is best seen as being in transition:

Pre-KT		Conservative KT		Innovative KT	
i	u	i	u	i	u
ia	ua	ie	uo	e	o
ɛ	ɔ	ɛ	ɔ	ɛ	ɔ
	a		a		a

The two diphthongs were pronounced [ia] and [ua] in pre-KT, as seen from the following comparisons:

KT	Hakha		KT	Hakha	
fieng	fiang	‘shoulder’	hlúoq	thlùak	‘brain’
gîet	riat	‘eight’	kûo	kûa	‘nine’
bfieng	bfiang	‘cheek’	pûon	pûan	‘cloth’

Special note needs to be made of the realization of *ie* and *uo* in grammatical markers, e.g. declarative marker *êe*, e.g. *à núuy êe* ‘he is laughing’. While the orthography suggests that *êe* is pronounced [ê:] (cf. *èe* [è:] ‘splitting’), it is actually realized with the more close vowel [ê:]. Similarly, the corresponding negative marker and the polite imperative marker are written *pòo* and *ôo* but are pronounced [pò:] and [ô:]: *à núuy pòo !êe* ‘he isn’t laughing’, *núuy ôo* ‘laugh!’. These two markers plus the variability between *ie* ~ *e(e)* and *uo* ~ *o(o)* indicate that KT is gradually developing into a seven-vowel system: /i, e, ɛ, u, o, ɔ, a/, short and long.

2.3.3. Minimal pairs

For most syllable types, minimal or near-minimal pairs contrasting the KT vowels are not difficult to find:

	CVV		CVD		CVVD	
i/ii	khii	‘necklace’	hĩng	‘suffice’	ĩim	‘keep secret’
e/ee	khêe	‘spoon’	hlẽng	‘choose’	kẽen	‘steep’
u/uu	khùu	‘cover’	hlûng	‘arrive’	kûun	‘droop’
o/oo	khòo	‘throw’	hlòng	‘pick’ (beans)	kôon	‘bent’
a/aa	khàa	‘body lice’	hẵng	‘south, down’	kâan	‘cross’
ie	khìe	‘fall’	---		kĩem	‘decrease’
uo	khùo	‘foot’	---		kũong	‘boat’

	CVq		CVT		CVVT	
i/ii	thịq	‘iron’	khìt	‘tie’	---	
e/ee	tẽq	‘old’	khèt	‘striking’	tsêep	‘smoke’
u/uu	tûq	‘hold in hand’	khùt	‘hand’	sûut	‘untie’
o/oo	tòq	‘work’ (v.)	khòp	‘together’	tsôop	‘suck’
a/aa	tàq	‘very’	khàt	‘one’	sâat	‘cut’
ie	thyẽq	‘sweep’	---		sĩet	‘spit’
uo	thwỏq	‘endure’	---		tsûop	‘lung’

As indicated in §1, there is no lexical length contrast on stem vowels in open syllables. While these vowels and diphthongs are redundantly long in isolation, in close syntactic juncture, many CVV words become shortened: *khé lien* ‘big spoon’, *khà vòm* ‘black body lice’ (§XX). In addition, pronominal proclitics and other grammatical markers often have the shape CV: *ká hùon* ‘my garden’, *nà ìn* ‘your house’.

2.4. Segment frequencies and combinatorics

While KT has 22 consonants, and its five vowels and two diphthongs can appear long or short, not all combinations are attested. The counts reported in the following subsections are restricted to 926 monosyllabic words extracted from a lexicon of 1205 entries, the remainder being mostly compounds and bisyllabic words.

2.4.1. Segment frequencies

As a base line, this section establishes the relative frequencies of consonants and vowels in onset, coda, and nucleus positions.

2.4.1.1. Onset consonants. The following table provides the number of monosyllabic lexical entries having each of the indicated onset consonants (or Ø):

	Ø	p	t	k	ph	th	kh	b	d	g	ts
#	22	44	53	64	22	46	57	46	39	38	60
%	2.4	4.8	5.7	6.9	2.4	5.0	6.2	5.0	4.2	4.1	6.5

	s	hl	h	v	z	l	m	n	ng	totals
#	71	36	66	33	52	74	45	43	15	926
%	7.7	3.9	7.1	3.6	5.6	8.0	4.9	4.6	1.6	100.2

Average frequency (926 ÷ 20): 46.3 (5.0%)

As seen, the most frequent onset consonants are *l* (74), *s* (71), *h* (66), *k* (64) and *ts* (60). The least frequent onset consonants are *ng* (15), *ph* (22), and \emptyset (22). Because of the relative low number of *ph* onsets, unaspirated stops are slightly more frequent than aspirated (161 vs. 125). Other distributions by class include: voiced stops (123), voiceless fricatives (173), voiced fricatives (87), and nasals (103).

2.4.1.2. Coda consonants. The following table shows the frequencies of the nine coda consonants and \emptyset (in the case of CVV entries):

\emptyset	p	t	q	l	w	y	m	n	ng	totals
143	29	51	112	105	48	98	130	73	137	926
15.4	3.1	5.5	12.1	11.3	5.2	10.6	14.0	7.9	14.8	99.9

Average frequency (926 ÷ 10): 92.6 (10.0%)

The most common coda structures are \emptyset (143), *ng* (137), and *m* (130). The least common are *p* (29), *w* (48) and *t* (51). The high frequency of the velar nasal *ng* as a coda consonant is particularly noteworthy, as it has the lowest frequency as an onset consonant: 137 (14.8%) vs. 15 (1.6%).

2.4.1.3. Vowel nuclei. The following table shows the total number of each (short and long) vowel quality throughout the 926 entries:

	i	e	u	o	a	ii	ee	uu	oo	aa	ie	uo	totals
#	72	85	71	100	111	27	60	56	98	139	41	66	926
%	7.8	9.2	7.7	10.8	12.0	2.9	6.5	6.0	10.6	15.0	4.4	7.1	100.0

Average frequency (926 ÷ 12): 77.2 (8.3%)

As seen, the most frequent vowels are *aa* (139), *a* (111), *o* (100) and *oo* (98). Particularly striking is the high frequency of *aa* (15%) and the low frequency of *ii* (2.9%). In terms of distribution, there appears to be a split between the low, back vowels *o*, *oo*, *a*, *aa* vs. non-low and front vowels: Short *o* and *a* account for 211 entries, while *i*, *e*, *a* together account for 228. Among the long vowels and diphthongs, *oo* and *aa* entries total 237, while the *ii*, *ee*, *uu*, *ie*, and *uo* entries total 250. The following distribution of long vowels in CVV entries confirms this division:

	Cii	Cee	Cuu	Coo	Caa	Cie	Cuo	totals
#	17	12	28	19	49	5	12	142
%	12.0	8.5	19.7	13.4	34.5	3.5	8.5	100.1

Average frequency (142 ÷ 7): 20.3 (14.3%)

As seen, the *Coo* and *Caa* account for 68 or 47.9% of the entries, whereas *Cii*, *Cee*, *Cuu*, *Cie* and *Cuo* entries total 74 or 52.1% of the entries.

2.4.2. Combinatorics

Besides having overall differences in distribution, certain vowels are highly restricted (or prohibited) before specific coda consonants (§2.4.2.1). Possible dependencies between onsets and vowels and between onsets and codas are also examined in §2.4.2.2 and §2.4.2.3.

2.4.2.1. VC restrictions. There are a number of restrictions between the vowel nucleus and coda of a syllable. The following table shows the attested distributions:

	m	n	ng	l	w	y	p	t	q	#
i	11	8	12	19	2*	∅	3	6	10	71
e	4	6	7	9	1	26	6	1	27	87
u	18	10	16	4	∅	4	2*	4	13	71
o	11	5	9	13	29	11	∅	5	17	100
a	21	6	30	9	∅	4	7	5	29	111
#	65	35	74	54	32	45	18	21	96	440
ii	4	2	4	∅	∅	∅	∅	∅	∅	10
ee	9	8	13	11	1	∅	2	4	∅	48
uu	4	1	∅	6	∅	12	2	3	∅	28
oo	24	9	13	12	1	9	3	8	∅	79
aa	10	8	15	6	14	25	3	8	∅	89
ie	9	1	7	7	∅	∅	∅	6	6	36
uo	5	9	11	9	∅	7	1	1	10	53
#	65	38	63	51	16	53	11	30	16	343

This table is interpreted as follows: Where there is ∅, we assume that the gap is systematic. Where there is only one entry, it could be an exception, e.g. *khyèw* ‘tickle (s.o.)’, *nêew* ‘small’. An asterisk indicates that the two entries suspiciously involve the same root: *kìw* ‘knock’ = ‘elbow’, *khùp* ‘put upside down’ = ‘cover’ (n.).

The following gaps or near gaps appear to be systematic concerning coda consonants and the vowels that may precede them:

• *uu* is not found before *ng*, although there is a compound *vá hǔung* ‘peacock’ which attests the rime *uung*. There is only one example each of *uun* and *ien*. However, two more instances of the latter occur in derived stem2 forms (§XX): *pǎeng pìen* ‘be born’, *thǐeng thien* ‘clean’.

• *l* appears to allow all vowels to precede it except *ii*. Since short *i* is the most common vowel before *-l*, there may have been a historical shortening of *ii* to *i* before *l*.

• *w* allows only *o* and *aa* to precede: *hòw* ‘religion’, *kòw* ‘call’, *tsàaw* ‘bangles’, *pâaw* ‘proud’. It is likely that *Cow* should be analyzed underlyingly as /Caw/ with /a/ becoming rounded before the coda *w*.

• *y* disallows preceding *i*, *ii*, *ee*, and *ie*. The syllable structure *Cey* is abundantly attested, but it is possible that many of these derive from /Cay/, which is rare. Still, there are a few *Cay* forms and near-minimal pairs such as *bày* ‘easy’ vs. *běy* ‘used up’, *hlây* ‘run’ vs. *hêy* ‘axe’, *lêy* ‘earth’.

Concerning the final stops, it would appear that there are no lexical entries of the shape *Cop*, *Ciip*, *Ciep* or *Ciit*. In addition, there is only one root of the shape *Cup*, *Cuop*, *Cet* or *Cuot*. However, as the following tables show, some of these gaps are filled by derived forms—specifically by the stem2 forms of verbs (see §XX):

	ip	ep	up	op	ap	iip	eep	uup	oop	aap	iep	uop	#
lex	3	6	2*	∅	7	∅	2	2	3	3	∅	1	29
der	3	7	3	8	9	∅	2	∅	3	2	2	2	41
#	6	13	5	8	16	∅	4	2	5	5	2	3	69

	it	et	ut	ot	at	iit	eet	uut	oot	aat	iet	uot	#
lex	6	1	4	5	5	∅	4	3	8	8	6	1	51
der	2	6	5	5	11	1	6	7	8	15	6	5	77
#	8	7	9	10	16	1	10	10	16	23	12	6	128

For example, while there are no lexical entries of the shape *Cop*, there are eight derived stem2 forms in the lexicon that have exactly this shape, e.g. *zòm zòp* ‘connect’, *tsòm tsòp* ‘jump’, *lòm lòp* ‘celebrate’. Similarly, while there was only one example of *Cet* in the lexicon (*pèt* ‘bite’), there are six stem2 forms of this shape: *věe vèt* ‘look at’, *nèen nèt* ‘dirty’, *khèen khèt* ‘hit, strike’. When both lexical and derived forms are taken into consideration, only *Ciip* is missing entirely from KT.

2.4.2.2. CV restrictions. The following table indicates how many lexical entries are found for each combination of onset + vowel.

	i	e	u	o	a	ii	ee	uu	oo	aa	ie	uo	#
p	1	2	4	2	9	2	5	4	3	7	2	3	44
t	7	3	6	8	21	2	2	3	3	4	2	5	66
k	7	5	4	8	5	1	4	2	10	8	3	6	63
ph	∅	3	4	2	2	1	3	1	∅	3	1	2	22
th	7	5	1	4	7	3	1	2	3	7	3	2	45
kh	1	3	6	6	7	1	6	6	5	7	2	7	57
b	1	4	7	3	6	∅	4	3	6	4	1	6	45
d	5	4	1	6	5	1	1	∅	11	4	∅	1	39
g	8	2	∅	3	4	∅	1	4	3	3	3	7	38
ts	7	6	3	6	6	3	3	2	10	8	3	3	60
s	4	6	10	2	4	4	3	4	15	12	5	5	74
hl	1	5	5	6	4	1	2	1	2	10	∅	∅	37
h	3	9	4	9	8	3	4	2	6	14	4	3	69
v	1	7	3	6	3	1	4	1	1	6	∅	1	34
z	4	1	3	9	3	4	3	6	1	11	3	1	49
l	4	6	3	8	9	4	7	7	4	10	5	5	72
m	5	5	4	5	9	5	2	3	3	5	∅	3	49
n	2	2	1	3	8	1	5	3	2	9	2	4	42
ng	∅	2	∅	2	2	2	∅	∅	1	6	1	1	17
#	68	80	69	98	122	39	60	54	89	138	40	65	922

While some cells are empty (or filled by only one item), there do not appear to be any robust cooccurrence restrictions between onset and vowel nucleus. There also is no evidence of CV assimilatory effects (e.g. palatalization).

2.4.2.3. C...C restrictions. The following table indicates how many lexical entries are found for each combination of onset + coda.

	m	n	ng	l	w	y	p	t	q	#
p	4	3	4	6	3	1	1	6	5	33
t	6	5	8	7	3	6	1	2	9	47
k	8	7	9	13	3	8	2	4	4	58
ph	∅	3	4	3	1	1	∅	2	4	18
th	7	2	7	3	4	3	1	2	7	36
kh	6	4	6	9	1	5	3	5	4	43
b	2	5	12	5	1	3	∅	3	8	39
d	5	5	9	4	∅	5	2	1	3	34
g	5	∅	6	8	1	6	1	2	3	32
ts	12	3	14	7	3	2	3	1	6	51
s	21	6	8	7	3	3	3	4	6	61
hl	8	4	7	∅	1	3	1	∅	6	30
h	9	5	8	8	5	10	2	3	10	60
v	1	4	2	5	∅	7	∅	5	4	28
z	5	4	9	5	8	1	3	3	5	43
l	15	2	10	2	5	11	5	2	9	61
m	1	7	7	9	2	6	∅	3	4	39
n	11	1	3	3	3	8	1	1	3	34
ng	∅	1	2	∅	∅	5	∅	∅	4	12
#	126	71	135	104	47	94	29	49	104	759

While there are few generalizations that refer to all members of a class of segments, the above distributions suggest at least some resistance to having a labial in both onset and coda position:

- Onset = *ph*: There are no words of the shapes *phVm*, *phVVm*, or *phVp*, either among the lexical entries or among derived forms. There is only one word with a *w* coda: *phôw* ‘spread out to dry’.

- Onset = *b*: There are no words of the shape *bVp* or *bVVp* and only one word with a coda *w*: *bòw* ‘only’. There are two words with an *m* coda: *bôm* ‘net’, *bêm* ‘circle’.

- Onset = *v*: There are no words of the shape *vVw*, *vVVw*, *vVp* or *vVVp*. There is only one word with an *m* coda: *vòm* ‘black’.

- Onset = *m*: There are no words of the shape *mVp* or *mVVp*, and only one word whose coda is *m*: *môm* ‘spider’. Two words have a *w* coda: *mốw* ‘bride, groom’, *mắaw* ‘bamboo (sp.)’.

- Onset = *p*: There doesn’t seem to be any avoidance of an *m* or *w* coda, but there is only one word with coda *p*: *pàp* ‘greedy’.

Evidence of resistance to an alveolar consonant occurring in both onset and coda positions is considerably weaker. The most suggestive evidence concerns coda *t* which occurs in only one word each with the onsets *d* and *n*: *dòt* ‘pierce, stab’, *nìt* ‘observe a holiday’. There also is only one lexical entry with an *n* in both onset and coda position: *nèen* ‘dirty’. To this we can add a single additional word, *nùng* ‘obey’, whose stem2 form *nùn* has this shape.

Another striking gap is the absence of words of the shapes *hlVl* or *hlVVl*. This suggests an avoidance of co-occurring lateral consonants. However, two words appear with voiced *l* in both onset and coda positions: *lèl* ‘lose, be defeated’, *lǒol* ‘throat’.

Assuming that *z* is from an earlier palatal consonant, it may also be significant that only one *z*-initial word occurs with a *y* coda: *zùuy* ‘follow’.

Other gaps may be accidental, due to historical factors, or attested in via derived forms. Thus, although there are no lexical entries of the shapes *gVn* or *gVVn*, these are found in the stem2 form of several verbs: *gìng gìn* ‘ring’, *gàng gàng* ‘fast’, *gùong gùn* ‘prepare’, *gǒong gòn* ‘skinny’. The lexical forms of these verbs amply demonstrate that velar onsets readily occur with the coda *ng*. Since both coda **k* and **r* became *q* in KT (cf. §XX), only comparative evidence will reveal how many of the *q* codas of present-day *kVq*, *khVq*, *gVq*, and *ngVq* derive from **k*. Some combinations of velar onset + *w* coda are also not very common.

2.5. Segmental alternations

There are relatively few across-the-board segmental phonological rules in KT. Other than tonal modifications (§XX) and a process of vowel shortening (§2.5.1), when lexical words occur in sequence in an utterance, they generally keep their isolation form. The segmental changes that do take place are summarized in the following table:

	+ C	lexical -VV, -VC	gram -VV, -VC
C +	∅	-V ?V	p, t p ^m , t ⁿ w v
VV +	VV V	VV V	-in n -êe ì, y -ùu- ùv
V +	∅	-V ?V	-V ?V

The C, VV and V in the first column refer to the ending of the first of lexical or grammatical morpheme. Except when the plural enclitic *ùu* becomes *ùv* before another enclitic, grammatical and lexical morphemes behave identically when positioned before another element. The lexical vs. grammatical function of the morpheme similarly does not matter when the second element begins with a consonant. Differences are observed, however, when the second element is vowel-initial.

2.5.1. Vowel shortening (VS). By a rule of vowel shortening, CVV syllables often become CV in context. Whether VS will occur depends on three factors: (i) the lexical vs. grammatical status of a following vowel-initial morpheme; (ii) the syntactic configuration in which the CVV is found; (iii) the underlying form (or historical origin) of the CVV.

2.5.1.1. VS within the noun phrase. As seen in the following examples, CVV nouns shorten before an adjective, a numeral, and the plural marker *hòo*, as well as in noun-noun possessives and compounds:

vâa	‘bird’	mũu	‘hawk’	khàa	‘body lice’
vá lien	‘big bird’	mú gâng	‘fast hawk’	khà tâm	‘many lice’
vá nìi	‘two birds’	mú thúm	‘three hawks’	khà gùup	‘six lice’
vá hòo	‘birds’	mú hoo	‘hawks’	khà hòo	‘lice’
vá hlàa	‘bird’s wing’	mú lúu	‘hawk’s head’	khà pèe	‘lice’s biting’
vá ¹ bũu	‘bird nest’	mú túy	‘hawk egg’	khà gũu	‘lice poison’

CVV adjectives also shorten their vowel in the same contexts: *vá phà nìi* ‘two good birds’, *vá phà hòo* ‘good birds’.

Preposed demonstratives undergo VS, and the diphthong *ie* must therefore be written *ye*: *hítsyé vâa* ‘this bird’, *tsútsyé múu* ‘that hawk’ (near hearer), *khútsyé khâa* ‘those lice’ (far). Recall from §2.1.1 that postposed demonstratives are short before pause but variable in context: *vòq hí* ‘this pig’, *vòq hí* (~ *hĩi*) *ká tsòo êe* ‘I bought this pig’.

Vowel shortening does not occur before the postposed demonstratives: *vâa ¹hĩ* ‘this bird’, *múu ¹tsú* ‘that hawk (near hearer)’, *khâa khú* ‘those lice’ (far). The numerals *nìi* ‘two’ and *lìi* ‘four’ and the plural marker *hòo* never shorten: *vá nìi hòo* ‘two birds’, *vá lìi hòo* ‘four birds’.

Multiple CVV nouns in a possessive or compound relation, if unmodified, will shorten: *vá hlà gũu* ‘bird’s wing-bone’, *mú lú khâa* ‘hawk’s head-lice’ (*lũu* ‘head’). An adjective modifier of the first noun will similarly undergo VS: *vòq phà màn* ‘price of a good pig’, *théy ¹khá khũong* ‘bowl of bitter fruit’ (*khâa* ‘bitter’) XX. While these last examples show that a modifier on the first noun can undergo shortening, VS will never apply if the modifier is a numeral or plural *hòo*: *tsápáng nìi nũu* ‘the two children’s mother’ XX, *vá hòo bũu* ‘the birds’ nest’.

A noun or adjective which is final in its noun phrase will not shorten before a subject marker or another noun phrase: *vâa ¹á hlàa êe* ‘a bird fell (from a height)’, *nà khúo ká mùu êe* ‘I see your village’, *múu ¹ngáa ¹ká pìe êe* ‘I gave the hawk a fish’.

If the syntactic conditions are met, VS will apply whether the following lexical morpheme begins with a consonant (as in the above examples) or with a vowel: *vá ¹in* ‘bird house’, *ngá ěng* ‘green fish’, *sá ¹ôo* ‘animal’s voice’. On the other hand, VS will not apply before a vowel-initial enclitic such as the plural possessor marker *ũu*: *kà vâa ùu* ‘our bird’, *nà khũo ùu* ‘your (pl.) village’, *á khâa ùu* ‘their body lice’. VS also does not apply before case markers: *zũu in* ‘mouse’ (ergative), *lũu in* ‘with the head’ (instrumental), *khũo âq* ‘in the village’ (locative), *thũ ¹áa tũy* ‘water in the blood (oblique). It is hard to know whether the non-application of VS is because such enclitics begin with a vowel, or because they, like postposed demonstratives, occur at the end of a noun phrase (§XX). The last example cited also shows that the oblique marker *áa* does not shorten.

The following low tone nouns are exceptions to VS:

bũu	‘food, rice’	bũu màn	‘food price’	cf. Hakha bũq
mèe	‘curry’	mèe hòo	‘curries’	cf. Hakha mẽq
nàa	‘leaf’	nàa lien	‘big leaves’	cf. Hakha nãq
lìe	‘shadow’	lìe hòo	‘shadows’	cf. Hakha hlyãq

As seen in the rightmost column, these L CVV nouns correspond to CVq with rising tone in Hakha Lai. low tone nouns which correspond to Hakha CVV, on the other hand, always undergo VS in the appropriate context:

hlàa	‘wing’	hlà hòo	‘wings’	cf. Hakha thlǎa
kùo	‘hole’	kwò nìi	‘two holes’	cf. Hakha kǎa

As seen in the next section, low tone CVV verbs which derive from pre-KT *CVq systematically fail to undergo VS. Such verbs are numerous, since pre-KT *-q figured importantly in stem2 formation. On the other hand, only a few CVV nouns correspond to pre-KT *CVq, and their exceptionality will simply have to be learned. For example, while *nàa* means both ‘leaf’ and ‘work’, it is only in its second meaning that VS applies: *nàa màn* ‘leaf price’, *nà màn* ‘work price’. On the other hand, there is evidence that the distinction between *CVV and *CVq is being lost. Forms such as *bù màn* ‘food price’ are also occasionally heard, and the following two nouns, which are expected to resist VS, are actually variable, usually produced short in VS contexts:

gàa	‘fruit’	á gà màn	‘its fruit price’	cf. Hakha ǎq
gùu	‘bone’	gù nìi	‘two bones’	cf. Hakha ǎq

CVV nouns which have either falling or rising tone in KT systematically undergo VS where the syntactic conditions are met: *kìi* ‘horn’, *kí lièn* ‘big horn’; *hǎa* ‘tooth’, *há thúm* ‘three teeth’.

In addition to the low tone CVV nouns which resist VS is a class of F and R tone personal nouns, e.g. *nâu* ‘mother’, *pǎu* ‘grandfather’. The following shows that there is shortening when these nouns are followed by an adjective or number, whether preceded by a possessive proclitic or not:

nú phàa	‘good mother’	kà nú phàa	‘my good mother’
nú nìi	‘two mothers’	kà nú nìi ùu	‘our two mothers’
pù lièn	‘big grandfather’	à pù lièn	‘his big grandfather’
pù nìi	‘two grandfathers’	à pù ní ùu	‘their two grandfathers’

As the first member of a simple noun + noun possessive/compound construction, VS is variable: *núú bùu* ~ *nú bùu* ‘mother’s food’, *púu ìn* ~ *pú ìn* ‘grandfather’s house’. However, the second variant sounds more like it is the name of the food or house: ‘mother food’, ‘grandfather house’. As seen in the following table, these two and other CVV personal nouns resist VS when the noun + noun combination is in turn preceded by a possessive proclitic:

nâu	‘mother’	kà núu !thâa	‘my mother’s strength’	cf. Hakha nâu
pâa	‘father’	kà páa !tîeng	‘my father’s spear’	cf. Hakha pâa
pǎu	‘grandfather’	kà púu ìn	‘my grandfather’s house’	cf. Hakha pâu
pǎi	‘grandmother’	kà píi mâay	‘my grandmother’s face’	cf. Hakha pài
zǐi	‘spouse’	kà zǐi mêeng	‘my spouse’s cat’	

mîi	‘person’	hítsyé mîi ¹ vâa	‘this person’s bird’	cf. Hakha mîi
nùu	‘female’	kà tsá ¹ núu ¹ ôo	‘my daughter’s voice’	cf. Hakha nùu
pàa	‘male’	kà tsá ¹ páa ¹ lôw	‘my son’s field’	cf. Hakha pàa
úu	‘older sibling’	kà úu nûu	‘my older sister’	cf. Hakha úu

As seen, these nouns correspond to CVV in Hakha, but fail to undergo VS in this construction. On the other hand, *tsâa* ‘child’, does shorten: *kà tsá ¹thêy* ‘my child’s fruit’. Compare also the long form *tsápâng* ‘child’, as well as *tsá pàa* ‘son’ and *tsá nùu* ‘daughter’ in the above examples. Similarly, VS applies in the lexicalized compound *nù pîi* ‘married woman’ (from *nùu* ‘female’ + *pîi* ‘great, major’). Finally, note the differences in vowel length of /úu/ in the following words: XX

úu pâa	‘older brother’	ù pâa	‘married man, leader, elder’
úu nûu	‘older sister’		

The above variation indicates that VS may be in a state of flux for this class of nouns. The greatest resistance occurs when the noun is the final element of a left-branching noun phrase: ‘my mother’ + ‘strength’. Other nouns, as well as adjectives, will shorten even in this environment: *mîi phàa* ‘good person’, *mîi phà bûu* ‘good person’s food’. Although the genitive high tone discussed in §XX appears in this latter phrase, /mîi + phàa + ^ˈ + bûu/, this does not seem to have any bearing on whether VS will apply or not.

2.5.1.2. VS within the verb complex. Parallel to the noun phrase, a verb undergoes VS when meeting the appropriate conditions within the verb complex. The following examples show that stem1 verbs such as *tsôo* ‘buy’ and *gûu* ‘steal’ undergo VS when a following postposition is a full word, but not when it is an enclitic (*êe*, *ùu*):

VS:	à tsó hlòn êe	à gú hlòn êe	‘they (dual) bought/stole it’
	tsó ðing á hîi êe	gú ðing á hîi êe	‘he will buy/steal it’
	à tsó zìng êe	à gú zìng êe	‘he often buys/steals it’
	à tsó ¹ núom ¹ êe	à gú ¹ núom ¹ êe	‘he wants to buy/steal it’
	à tsó ¹ kít ¹ êe	à gú ¹ kít ¹ êe	‘he stole/bought it again’
No VS:	à tsóo ¹ êe	à gúu ¹ êe	‘he buys/steals it’
	à tsóo ùvêe	à gúu ùvêe	‘they (pl.) buy/steal it’

The following sentences show that stem1 CVV verbs of all three tones (falling, rising, low) potentially undergo VS, e.g. before the dual marker *hlòn*:

tone	stem1	stem2		example	
F:	nêe	nêq	‘eat’	à né hlòn êe	‘they (dual) ate it’
	zâa	zâq	‘hear’	à zá hlòn êe	‘they (dual) heard it’
	lâa	lâq	‘take’	à lá hlòn êe	‘they (dual) took it’
R:	băa	bât	‘borrow’	à bà hlón ¹ êe	‘they (dual) borrowed it’

	về	vết	‘watch’	à zè hlón 'êe	‘they (dual) watched it’
	pửu	pụt	‘carry’	à pù hlón 'êe	‘they (dual) carried it’
L:	vòo	vòq	‘hit, beat’	á vò hlón 'êe	‘they (dual) beat it’
	sùu	sùq	‘pound’	á sù hlón 'êe	‘they (dual) pound it’
	sàa	sàq	‘pay back’	á sà hlón 'êe	‘they (dual) paid it back’

All falling and rising tone stem1 verbs undergo VS. However, there are many low tone CVV stem1 verbs like *tìi* ‘bind’ and *tèe* ‘measure’ which fail to shorten:

no VS:	á <i>tìi</i> hlón 'êe	á <i>tèe</i> hlón 'êe	‘they (dual) bound/measured it’
	<i>tìi</i> dǐng 'á hìi êe	<i>tèe</i> dǐng 'á hìi êe	‘he will bind/measure it’
	à <i>tìi</i> zǐng êe	à <i>tèe</i> zǐng êe	‘he often binds/measures it’
	á <i>tìi</i> núom 'êe	á <i>tèe</i> núom 'êe	‘he wants to bind/measure it’
	à <i>tìi</i> kít 'êe	à <i>tèe</i> kít 'êe	‘he bound/measured it again’
No VS:	á <i>tìi</i> ê	á <i>tèe</i> ê	‘he binds/measures it’
	á <i>tìi</i> ùvêe	á <i>tèe</i> ùvêe	‘they (pl.) bind/measure it’

Of the 45 low tone CVV verbs in the lexicon, 18 fail to undergo shortening. Further examples seen in the following table:

tone	stem1	stem2			
L:	<i>tìi</i>	<i>tìi</i>	‘bind’	á <i>tìi</i> hlón 'êe	‘they (dual) bound it’
	<i>tèe</i>	<i>tèe</i>	‘measure’	á <i>tèe</i> hlón 'êe	‘they (dual) measured it’
	<i>nùu</i>	<i>nùu</i>	‘rub’	á <i>nùu</i> hlón 'êe	‘they (dual) rubbed it’
	<i>dòo</i>	<i>dòo</i>	‘hold’	á <i>dòo</i> hlón 'êe	‘they (dual) held it’
	<i>hlàa</i>	<i>hlàa</i>	‘let go’	á <i>hlàa</i> hlón 'êe	‘they (dual) let it go’

A comparison of the stem2 forms of low tone CVV verbs shows that VS can be predicted as follows: verbs which have a *CVq* or *CVt* stem2 form undergo VS, while those which have a stem2 CVV form do not.

Related to this last observation is the fact that stem2 CVV verbs never undergo VS:

tone	stem1	stem2		example	
F:	<i>thôw</i>	<i>thòo</i>	‘get up’	á <i>thòo</i> nǔng	‘after he got up’
	<i>khâq</i>	<i>khàa</i>	‘close, lock’	á <i>khàa</i> nǔng	‘after he closed it’
	<i>hôw</i>	<i>hòo</i>	‘shave’	á <i>hòo</i> nǔng	‘after he shaved’
R:	<i>kěq</i>	<i>kèe</i>	‘break’	á <i>kèe</i> nǔng	‘after it broke’
	<i>hlỏq</i>	<i>hlòo</i>	‘soft’	á <i>hlòo</i> nǔng	‘after it was soft’
	<i>thỉi</i>	<i>thìi</i>	‘die’	á <i>thìi</i> nǔng	‘after he died’
L:	<i>tèp</i>	<i>tèe</i>	‘taste’	á <i>tèe</i> nǔng	‘after he tasted it’
	<i>pèt</i>	<i>pèe</i>	‘bite’	á <i>pèe</i> nǔng	‘after he bit him’

tsòq tsòo ‘stir’ á tsòo nǔng ‘after he stirred’

In the above forms the stem1 form ends in a consonant, while the stem2 forms have the shape CVV with low tone. The examples involve clauses ending with postposition *nǔug* ‘after’, which requires a preceding stem2 verb. Note that all CVV stem2 forms have low tone in KT.

To summarize, a stem2 CVV never undergoes VS, while a stem1 CVV will not undergo VS if both (i) it has low tone, and (ii) its stem2 form is CVV. The historical generalization is that CVV forms which come from pre-KT *CVq do not undergo VS. All stem2 CVV forms are from *CVq, and those which are low tone CVV in both stem1 and stem2 were invariantly *CVq in pre-KT. There are two special verbs worthy of note:

	KT			Hakha	
stem1	mùu	‘see’	á mù hlón !êe	‘they (dual) saw him’	hmǔu
stem2	mùu		á mùu hlón nǔng	‘after they (dual) saw him’	hmǔq
stem1	thǐi	‘die’	à thǐ hlón !êe	‘they (dual) died’	thǐi
stem2	thìi		à thǐi hlón !êe	‘after they (dual) died’	thǐq

mùu ‘see’ has a low tone CVV shape in both stem1 and stem2. As such it is expected to resist VS in both forms. However, as the first example shows, the stem1 verb exceptionally undergoes VS. The Hakha Lai form to the right shows why: Whereas other low tone CVV/CVV verbs reconstruct as *CVq in both forms, *mùu* reconstructs as *CVV in stem1, but *CVq in stem2. The same is seen with respect to the verb *thǐi* ‘die’, except that the stem1 tone is rising rather than low. Again, VS applies to the stem1 form, but not the stem2.

The following table presents verb compounds consisting either of two verbs or a verb plus postposition:

compound		verb1		verb2/postposition	
sù thiǎng	‘clean up’	sùu/sùq	‘pound’	thiǎng/thiǎn	‘clean’
hé zìng	‘remember’	hée/héet	‘know’	zìng	‘always’
lá dòq	‘pick up’	lâa/lâq	‘take’	dòq/dòo	‘stick out’
dâa hlâa	‘abandon’	dâa/dâa	‘divorce’	hlâa/hlâq	‘drop’ (intr)
nòo khùm	‘attack’	nòo/nòo	‘attack’	khùm/khùm	‘lock in’
thúu !ân	‘judge’	thúu/thúut	‘judge’	tân/tàn	‘cut’

In the first group of compounds, VS applies to the first verb, as expected. In the second group, VS fails to apply since the first verb has a CVV structure in both stem1 and stem2.

As seen in the following examples, just like single verbs, verb compounds often show different lexical vs. derived forms:

lexical	derived		verb1		verb2/postposition	
vâq lèe	vâa lèe	‘wander’	vâq/vâa	‘wander’	lèe/lèe	‘transpose’

bòt èq	bòo èq	‘tear’	bòt/bòo	‘pull’	èq/èq	‘tear’
ôq dèn	òo dèn	‘stuck’	òq/òo	‘caught’	dèn	‘still’

The lexical forms are as these compounds would appear in stem1-inducing contexts, while the derived forms show how these compounds are realized in a stem2-inducing context: à váq lèe êe ‘he wanders’, á vâa lèe nǔng ‘after he wanders’. As seen, VS fails to apply to the derived stem2 CVV forms, as expected.

Another kind of verb compound consists of a noun + verb combination:

lexical	derived		noun		verb	
tí zàa	tí zàat	‘startled’	tîi	‘body’	zâa/zâat	‘feel’
ná ngóng	ná ngòn	‘deaf’	năa	‘eardrum’	ngóng/ngòn	‘deaf’
thá sîe	thá sîet	‘lazy’	thâa	‘strength’	sîe/sîet	‘break down’

As seen, the noun undergoes VS in both lexical and derived forms.

2.5.1.3. No VS across major constituents. VS does not apply to words that come together across major constituents. A CVV noun will therefore remain long if followed by a subject marker or by a noun in another noun phrase: *váa* ¹*ká mùu êe* ‘I saw a bird’, *váa* ¹*môot kà píe* ¹*êe* ‘I gave a bird a banana’.

2.5.2. Glottal stop insertion (GSI). Morphemes which begin with a vowel often acquire an initial glottal stop by GSI. This process depends on the preceding environment and affects lexical and grammatical morphemes differently. There is some variability, and except when preceded by a short vowel, the glottal stop is not always clearly audible.

2.5.2.1. Postpausal GSI. Any vowel-initial morpheme that follows pause will acquire a glottal stop: [ʔ]á hlàa êe ‘he fell (from a height)’, [ʔ]úy ¹ná pèt êe ‘a dog bit you’, [ʔ]iim in ‘keep it secret!’. However, if these same morphemes are preceded by another word, whether within the same major constituent or not, GSI may not apply, depending on the rules in the following paragraphs: *váa* ¹á hlàa êe ‘a bird fell (from a height)’.

2.5.2.2. Postvocalic GSI. GSI will apply to a lexical morpheme or proclitic that is immediately preceded by a short vowel. The glottal stop will be quite noticeable in this case: *ká lùpnà* [ʔ]á ¹tsím ¹êe ‘the bed collapsed’, *nà* [ʔ]ín ‘your house’, *khwó* [ʔ]úy ‘village dog’. *khí* [ʔ]éng ‘green bead’, *á kî* [ʔ]àq êe ‘he put on a shirt’. GSI will apply before long vowels as well: *vá* ¹[ʔ]ôo ‘bird’s voice’, *zú* [ʔ]ôoy ‘mouse’s belly’, *sá* [ʔ]àat ‘meat cutting’. Since VS will not apply before an enclitic, the latter only rarely is preceded by a short vowel. When this does occur one of two things can happen. If the enclitic has the shape in, there is variation between GSI and coalescence (§XX): *lùpná* [ʔ]ín ~ *lùpnán* ‘bed + instrumental’. If the enclitic has the shape àq or áa, there is obligatory GSI: *lùpnà* [ʔ]àq ‘bed + locative], *lùpná* [ʔ]áa ‘bed + oblique’.

It should be noted that unless the following word is specifically emphasized (e.g. by placing a pause before it), GSI will not apply after a CVV word. In case the second word is a lexical

morpheme, there will be a smooth transition between the vowels: *bùu ùuy* ‘smelly rice’, *nàa ǎng* ‘green leaf’. In cases where a CVV word alternates in vowel length, *móot gàa ǎng* ~ *móot gà* [ʔ]ǎng ‘green banana fruit’. When the following vowel-initial morpheme is an enclitic, there is either a smooth transition or coalescence: *sàa ôo* ‘please pay him back!’, *á hlòo êe* (~ *á !hlóêy*) ‘he is trampling it’, *váa ìn* (~ *vâan*) *á pèt ê* ‘the bird bit him’.

2.5.2.3. Postconsonantal GSI. A glottal stop will not be inserted between a consonant and a vowel-initial enclitic: *dón !ôo* ‘please babysit!’, *hóol ìn* ‘look for it!’, *á kòon êe* ‘it is bent’. GSI optionally applies to a post-consonant, vowel-initial lexical morpheme: *hùon* [ʔ]ǎng ‘green garden’, *hòw* [ʔ]ín ‘church’ (*hòw* ‘religion’, *ín* ‘house’), *á sùon* [ʔ]ùm *!êe* ‘he is trustworthy’ (*sùon* ‘trust’, *ùm* ‘exist’). GSI also applies between a lexical morpheme and a vocalic proclitic: *hítsyé móot* [ʔ]à *hlúm êe* ‘this banana is sweet’, *kà méeng ìn tsápáng* [ʔ]à *khúot !êe* ‘my cat scratched the child’, *tsíní* [ʔ]í *sù mìl êe* ‘you and I forgot the sugar’. In fast speech, postconsonantal GSI often does not apply.

2.5.3. Enclitic vowel coalescence. Rather than GSI, whenever a CV or CVV morpheme is followed by a VV or VC enclitic, coalescence can result. In some cases the resulting coalescence is more common than the corresponding non-coalesced form. Unless otherwise noted, the latter form is always available in slow or hyperarticulated speech.

2.5.3.1. Coalescence of *in* markers. Enclitics of the shapes *ìn* and *ǎn* occur throughout KT grammar. Low tone *ìn* denotes ergative case on subject noun phrases and marks imperatives and conjoined clauses: *tsápáng ìn éy mùu êe* ‘the child saw me’, *líut ìn* ‘enter!’, *à húng ìn, éy mùu êe* ‘he came and saw me’. Rising tone *ǎn* marks instruments and temporals: *kháaw ǎn bóong ká !héen êe* ‘I attached the cow with a rope’, *á hlùn núng ǎn* ‘after he arrives’, *túu tàq ǎn* ‘right now’. Both markers optionally lose the vowel [i] when preceded by a vowel. Each of the seven combinations of CVV + imperative *ìn* is illustrated in the following table:

underlying	full form	coalesced form	
/tsǎi ìn/	tsǎi ìn	[tsǎi:̀̀n]	‘sneeze!’
/nêe ìn/	nêe ìn	[nêe:̀̀n]	‘eat it!’
/gúu ìn/	gúu ìn	[gúu:̀̀n]	‘steal it!’
/tsóo ìn/	tsóo ìn	[tsóo:̀̀n]	‘buy it!’
/hlàa ìn/	hlàa ìn	[hlàa:̀̀n]	‘let it go!’
/pêe ìn/	pêe ìn	[pêe:̀̀n]	‘give it!’
/púo ìn/	púo ìn	[púo:̀̀n]	‘carry it!’

As seen, the vowel length of the CVV syllable is maintained. The low tone of *ìn* is realized on the nasal, which sounds syllabic. The other *in* markers show the same coalescence possibilities when preceded by a CVV syllable. In addition, the grammatical morphemes *kwòy* ‘who’ and *kêy* ‘I, me’ have special contracted forms when followed by ergative *ìn*: *kwòy ìn* ~ *kòn*, *kêy ìn* ~ *kên*. Lexical morphemes do not simplify in this way.

2.5.3.2. Coalescence of êe. The declarative marker *êe* readily coalesces with a preceding CVV syllable:

underlying	full form	coalesced form	
/á ðii íe/	á ðii êe	[á 'dí:ì]	'it's in pieces'
/á khèe íe/	á khèe êe	[á 'kxé:ì]	'he peeled it'
/á khùu íe/	á khùu êe	[á 'kxú:ì]	'he covered it'
/á hlòo íe/	á hlòo êe	[á 'hlóé:ì]	'he trampled it'
/á phàa íe/	á phàa êe	[á 'phá:ì]	'it is good'
/á tsìe íe/	á tsìe êe	[á 'tsíé:ì]	'he walked'
/á khùo íe/	á khùo êe	[á 'kxúó:ì]	'he rinsed it'

The coalesced forms show the raising of *êe* to *i*, which may even be desyllabified to *y*. Two frequent (and slightly irregular) coalescences occur with the copula *hii* and the negative marker *pòo*: *lów 'mú 'á hii ê ~ lów 'mú 'á 'hii* 'he is a farmer', *lów 'mú 'á hii pòo êe ~ lów 'mú 'á hii pòoi* 'he is not a farmer'. It is possible, but rare to hear *hii êe* in its non-coalesced form. Since *hii + êe* frequently occurs in expressing the copula, and in marking verb tenses and cleft constructions, it may best be written *hii*.

2.5.3.3. Other enclitics. The remaining enclitics do not coalesce with a preceding CVV word. This is seen in the following examples with respect to the enclitic *ùu*, which expresses plurality of one of the arguments of the verb or of the possessor of a noun:

à thii ùu	'their blood'	à pháa ùu	'their father'
á phèe ùu	'their flesh'	á lie ùu	'their shadow'
á bùu ùu	'their rice'	à khùo ùu	'their village'
à thòo ùu	'their fence'		

However, when *ùu* is followed by another enclitic it has a special realization, e.g. *ùu + êe ùvêe*, *ùu + in ùvin* (§XX).

Additional non-coalescing enclitics include the polite imperative marker *ôo*, the locative marker *àq*, and the oblique marker *ãa*:

tsíi 'ôo	'please sneeze!'	kà píi áa	'my grandmother's'
née 'ôo	'please eat it!'	kà khée àq	'in/on my spoon'
gúu 'ôo	'please steal it!'	kà púu áa	'my grandfather's'
tsóo 'ôo	'please buy it!'	kà óo àq	'in my voice'
hlàa ôo	'please let it go!'	kà páa ãa	'my father's'
píe 'ôo	'please give it!'	ká lie àq	'in/on my shadow'
púo 'ôo	'please carry it!'	ká khùo àq	'in/on my foot'

2.5.4. Consonant release into an enclitic. While the preceding section dealt with cases where an enclitic vowel fuses into a preceding CVV syllable, there are also two processes where a consonant has a special realization before a VV or VC enclitic.

2.5.4.1. Nasal release. The two coda stops *p* and *t* frequently acquire a nasal release when an enclitic vowel follows:

noun + enclitic			verb + enclitic		
khûup	khúup ^m ĩn	‘with the knee’	kâap	káap ^m in	‘shoot!’
	khúup ^m àq	‘in the knee’		à káap ^m !êe	‘he shot’
	khúup ^m ăa	‘of the knee’		á kâap ^m ăa	‘he shot &’
khùt	khùt ⁿ ĩn	‘with the hand’	lûut	lúut ⁿ in	‘enter!’
	khùt ⁿ àq	‘in the hand’		à lúut ⁿ !êe	‘he entered’
	khùt ⁿ ăa	‘of the hand’		á lùut ⁿ ăa	‘he entered &’

Glottal stop does not acquire a nasal release, nor do *p* and *t* acquire it when the following vowel-initial word is a lexical morpheme. Instead, a glottal stop is typically inserted: *khúup* [ʔ]ũul ‘knee perspiration’. In general KT resists resyllabifying a consonant with the vowel of a following morpheme. When this morpheme is lexical, there is a glottal stop which demarcates the break. When it is an enclitic, the consonant is freer to be syllabified with the following vowel. The nasal release acts a buffer, allowing a smooth release of *p* and *t* but not resyllabification.

2.5.4.2. Coda lengthening. The sonorant codas *m*, *n*, *ng*, and *l* undergo slight lengthening when followed by an enclitic.

noun + enclitic			verb + enclitic		
gám ăa	[gám: ă:]	‘of the land’	à tsôm !êe	[λ tsóm: ê:]	‘he is short’
súm ín	[súm: ín]	‘with money’	à sím êe	[λ sím: ê:]	‘he is reading’
hùon àq	[hùon: âʔ]	‘in the garden’	á lien êe	[Á lien: ê:]	‘it is big’
zàan:ĩn	[zà:n:ĩn]	‘at night’	à dóon êe	[λ dó:n: ê:]	‘he is drinking’
bòong ín	[bɔ:ŋ: ín]	‘cow + ergative’	à híng êe	[λ híŋ: ê:]	‘he is alive’
tíeng ĩn	[tíeŋ: ĩn]	‘with a spear’	à thíeng êe	[λ thíeŋ: ê:]	‘it is clean’
bèel âq	[bè:l: àʔ]	‘in the pot’	à vél êe	[λ vél: ê:]	‘it is shining’
mòl ĩn	[mòl: ĩn]	‘with a stick’	à bóol êe	[λ bó:l: ê:]	‘he is doing it’

This lengthening, which is less than in languages with true geminates, is presumably also a “buffer effect”, since resyllabification with the following vowel is avoided.

2.5.4.3. w → ww. There is one exception to the prohibition against the resyllabification of a consonant across a morpheme boundary: Whenever *w* is followed by an enclitic vowel, it acquires a *v* release. As seen in the following examples, the *v* appears to function as the onset of the enclitic syllable:

noun + enclitic			verb + enclitic		
kìw	kìw vǐn	‘with an elbow’	kìw	á kìw vêe	‘he is knocking’
lôw	lów vậq	‘in the field’	tốw	à tów vêe	‘he is sitting’
thàaw	thàaw vǐn	‘with oil’	tâaw	à táaw 'vêe	‘he is praying’

In casual speech, the *w* may disappear entirely: *kì vǐn* ‘with an elbow’, *à tó vêe* ‘he is sitting’.

The plural enclitic *ùu* obligatorily shortens and acquires a *v* when followed by an enclitic vowel. The resulting fused enclitics are written without a space:

underlying	pronounced	
/á kàp ùu êe/	á kàp ùvêe	‘they are crying’
/à ìn ùu àq/	à ìn ùvậq	‘in/on their house’
/à sủoq ùu áa/	à sủoq ùvǎa	‘their slave’s’
/núuy ùu ìn/	nùuy úvìn	‘laugh!’ (pl.)
/núuy ùu ôo/	nùuy ú'vôo	‘please laugh!’ (pl.)

There are two ways to interpret the *ù(u) ~ ùv* alternation: The first is that a *w* transition is created that becomes *v* because of the following vowel. The second is that the underlying form of this marker is /ùw/, a rime shape that otherwise does not exist in the language (§2.4.2.1). In this latter interpretation, whenever the putative /w/ fails to become *v*, /ùw/ is realized *ùu* (finally) or *ù* (non-finally). A pronunciation with *wv* is unattested.

Finally, it should be noted that speakers do not apply *w v* in two syllables in sequence:

underlying	pronounced	
/á kàp ùu êe/	á hòw ùvêe	‘they are worshipping’
/à ìn ùu àq/	à lów ùvậq	‘in/on their field’
/à mớw ùu áa/	à mòw úvǎa	‘their fiances’
/thôw ùu ìn/	thów ùvìn	‘get up!’ (pl.)
/thôw ùu ôo/	thów ùvôo	‘please get up!’ (pl.)

Pronunciations such as *á hòw vùvêe* or *á hò vùvêe* (etc.) are strongly dispreferred.