

Khmer Character Specification/Usages

Author: Makara Sok

Contributors: Marc Durdin,

Todd Bequette,

Becky Bequette,

Seth Wilson, and

Martin Hosken

Last Update: May 3, 2019

Table of Contents

1. [Introduction](#)
2. [Khmer Language and Khmer Script](#)
3. [The Linguistic Situation in Cambodia](#)
4. [Khmer Phoneme Inventory](#)
5. [Khmer Orthography](#)
6. [Ligatures](#)
7. [Unicode Encoding](#)
8. [Text Processing](#)
9. [Application of Khmer Script to Other Languages](#)
10. [Stone Inscriptions \(Pre-Angkor, Angkor, Post-Angkor Era\)](#)
11. [Sample Texts for Orthography Check](#)
12. [Wordlist for Orthography Check](#)
13. [Summary](#)

[References](#)

[Appendices](#)

1. Introduction

The Khmer language is the national language of the Kingdom of Cambodia and the Khmer script is the official script used not only in writing the Khmer language, but other ethnic minority languages. It is also used to write religious documents (i.e. Dharma).

This document seeks to describe how the Khmer script is used in various languages nowadays, namely:

- Khmer language,
- ethnic minority languages,
- Pali and
- Sanskrit.

Before getting into greater details, let's look at the evolution of the Khmer script.

2. Khmer Language and Khmer Script

According to the Unicode Consortium (2017:631), the Khmer script is the official script of Cambodia and it is descended from the Brahmi script. It has been in use for more than 1400 years. During this period of time, the script has evolved substantially.

According to a study on Khmer letters (Doek 2000:5), Khmer letters evolved from Brahmi script. Deok (2000:13) and Maspero (1915:48) claim, as noted by Kul (2008:179-181), that Khmer letters (Khmer Script) have undergone ten distinct evolutions since the 6th century. The list below gives the details on when the changes happened and inscription it was used on.

- 6th century, inscription found in the reign of Pavavarman
- early 7th century, in the reign of Mahendravarman Jetrasen
- 667 AD, in the reign of Jayavarman I
- 970 AD, in the reign of Jayavarman V
- 1002 AD, in the reign of Suryavarman I
- 1066 AD, in the reign of Utyadityavarman II
- 12th century, in the reign of Suryavarman II
- 13th century, in the reign of Jayavarman VII
- 1702 AD, in the mid of the Udong era
- Nowadays

During the French protectorate era, an attempt to romanize Khmer language was made, but it was not welcomed because it was seen as an attack on traditional learning and the Khmer society (Scheuren 2010:19). The process of Khmer romanization was then failed, and it gave rise to Khmerization.

3. The Linguistic Situation in Cambodia

3.1. Languages of Cambodia

There are 27 languages spoken in Cambodia, according to the Ethnologue (Eberhard et al. 2019).

Status	# Languages with Ethnologue Codes
national	1 Khmer [khm]
wider communication	3 Min Nan Chinese [nan], English [eng], French [fra]
dispersed	2 Thai [tha], Vietnamese [vie]
developing	3 Brao [brb], Jarai [jra], Central Mnong ^[1] [cmo]
vigorous	5 Western Cham [cja], Kaco' [xkk], Hakka Chinese [hak], Kraol [rka], Tampuan [tpu]
threatened	6 Kavet [krv], Krung [krr], Lao [lao], Lao Phuon [phu], Mel-Khaonh [hkn], Stieng Bulo [sti]
shifting	1 Kuay ^[2] [kdt]
moribund	3 Pear [pcb], Somray [smu], Su'ung [syo]
nearly extinct	1 Chung [scq]
dormant	2 Chong [cog], Samre [sxm]
Total	27

3.2. Languages Using Khmer Script

Other than the Khmer language, Khmer script is also used to write ethnic minority languages, such as: Kuay, Bunong, Tampuan, Jarai, Krung, Brao and Kavet.

Language Name	Population (in Cambodia)	Location (Provinces in Cambodia)	Language Status (EGIDS)
Kuay	37,700	Kampong Thom and Preah Vihear	7 (Shifting)
Bunong	37,500	Kratie, Mondolkiri	5 (Developing)
Tampuan	31,000	Ratanakiri	6a (Vigorous)
Jarai	20,800	Ratanakiri	5 (Developing)
Krung	20,700	Ratanakiri and Stung Treng	6b (Threatened)
Brao	9,030	Ratanakiri	5 (Developing)
Kavet	6,220	Ratanakiri and Stung Treng	6b (Threatened)

More information on each language can be found on the ethnologue website at <https://www.ethnologue.com/country/kh/languages>.

4. Khmer Phoneme Inventory

Khmer phoneme inventory varies from one scholar to another. The inventory hereunder is adapted from Huffman (1970:6-11) and Ehrman (1972:4-9) as stated in Sok (2016:11-13) and Headley (2014:x).

4.1 Consonants

4.1.1 Initial Consonants

There are 21 consonant phonemes shown in the table below. Twelve of them can only occur in the initial position (i.e. they are marked with a hyphen).

Bilabial Alveolar Palatal Velar Glottal

Plosives	/p/	/t/	/c/	/k/	/ʔ/
Asp. Plosives	/pʰ-/	/tʰ-/	/cʰ-/	/kʰ-/	
Implosives	/ɓ-/	/ɗ-/			
Fricatives		/s-/			/h/
Nasals	/m/	/n/	/ɲ/	/ŋ/	
Semi-vowels	/w/			/j/	
Lateral		/l/			
Flap		/ɾ-/			

There is still a discussion on whether the following phonemes are a combination of two phonemes: (unaspirated plosive + glottal fricative), i.e. (a) /ph, th, ch, kh/ versus (b) /pʰ, tʰ, cʰ, kʰ/. One might argue that (a) is likely to be more reasonable because the aspiration can be separated from the plosive when there is an infix. i.e. ខោខុំ /khoh/ ‘to be wrong’ + -ុំ- /m/, កំបានុំ /kam.hoh/ ‘mistake’.

For some foreign loan words, certain phonemes are used, yet they are not included in the table above as it is reserved for native phonemes. The phonemes are: labiodental fricative /f/, voiced alveolar fricative /z/, palatal fricative /ʃ/ and an alveolar implosive /g/. They are usually found at the beginning of a syllable.

4.1.2. Initial Consonant Clusters

Initial consonant clusters may be composed of two or three consonants (i.e. C₁C₂ or C₁C₂C₃).

$$C_1 = \{ p \ t \ c \ k \ ɓ \ m \ l \ }$$

$$C_2 = \{ p \ t \ c \ k \ ɓ \ d \ m \ n \ ɲ \ w \ j \ r \ s \ h \}$$

$$C_3 =$$

4.1.3. Final Consonants

Aspirated plosives (p^h, t^h, c^h, k^h) , implosives (ɓ, ɗ, g), fricatives (f, s, z, ʃ; except glottal fricative /h/) and a flap (ɾ) can never occur in the syllable final position. Each consonant in the syllable final position is pronounced without any audible release.

$$F = \{ p \ t \ c \ k \ ɓ \ h \ m \ n \ ɲ \ l \ w \ j \}$$

Some foreign loanwords (i.e. English or French) may break this rule. For instance, ឃើមុបូក ‘Facebook’ may be pronounced as /fees.buk/.

Also note that the orthographic representation may be different from the actual pronunciation when it comes to Pali/Sanskrit loanwords. See [Final Consonants](#) in [Khmer Orthography](#).

4.2. Vowels

Khmer vowels are divided into two main groups: monophthongs and diphthongs.

4.2.1. Monophthongs

There are 18 monophthongs: 10 long monophthongs and 8 short monophthongs.

Front Central Back

High	i ii	ɨ[3] ɨɨ	u uu
Mid	e ee[4]	ə əə[5]	o oo[6]
Mid-Low	ɛɛ	ɔɔ	
Low	a aa	ɑ aa	

4.2.2. Diphthongs

There are 9 diphthongs: / iə, ie, ae, ao, aə, ɨə, uə, ea, oa/. Vowels whose second member is a semi-vowel (i.e. /j/ or /w/) are not considered as diphthongs Ehrman (1972:9). A complete list of them is given below:

- /iə/, /əw/, /aw/
- /iŋ/, /əŋ/ and /aj/.

$$V = \{ i \text{ ii } ɨ \text{ ɨɨ } u \text{ uu } e \text{ ee } ə \text{ əə } o \text{ oo } əə \text{ əə } a \text{ aa } ɑ \text{ aa } iə \text{ ie } ie \text{ ae } ao \text{ aə } ɨə \text{ uə } ea \text{ oa } \}$$

4.3. Stress

Stress falls on the last syllable and it is not phonemic (Schiller 1994:312). Therefore it is not included in the phonemic transcription.

4.4. Syllable Structures

Khmer syllable structures are divided into three types: monosyllables, disyllables and polysyllables (Huffman 1970:11-12).

Syllable Structures	Description
Monosyllables C ₁ C ₂ C ₃ VF	The onset can have up to three consonants in a row. The final consonant is optional.
minor syllable + major syllable where: - minor syllable is C(r)V(N) <ul style="list-style-type: none"> • C, any consonant • r, the flap /r/ • V, any vowel • N, any nasal consonant 	A disyllabic word is made up of a minor syllable and a major syllable. The minor syllable is susceptible to extreme syllable reduction (Huffman 1972:59-60). Here are some examples: <ul style="list-style-type: none"> • /com.kaa/ ‘farm’ > [cə.kaa] • /kən.dəal/ ‘middle’ > [kə.dəal] • /kən.craeŋ/ ‘bamboo basket’ > [kə.craeŋ] • /pə.riən/ ‘to teach’ > [pə.riən] • /rə.ɓəaŋ/ ‘fence’ > [lə.ɓəaŋ] • /pra.hael/ ‘perhaps’ > [pə.hael]
Disyllables - major syllable is C ₁ C ₂ V(F) <ul style="list-style-type: none"> • C₁, any C₁ consonant • C₂, any C₂ consonant • V, any vowel • F, any final consonant 	This syllable type is common in foreign loanwords, i.e. Pali/Sanskrit, English, French, etc.
Polysyllable multiple monosyllables	

4.5. Series Assimilation

Sok (2016:24) wrote:

Series assimilation, ធម៌ខ្លួនឯង or មិនអំពីសាយ “lit. sound adaptation” as proposed by Prom (2013:24) and កែវតាមឯង “lit. sound inclination” by Chan (2010:100), is the core of Khmer pronunciation and it is regularly correlated with the writing system. The series system and the sonority hierarchy drive the series assimilation phenomenon. Huffman (1970:43) refers to this phenomenon as Vowel Governance, for the pronunciation of a vowel symbol, including the invisible inherent vowel, is determined or governed by the series of the initial consonant symbol it is attached to (see Table 6 above). Thus, it is simple to pronounce a Khmer word which begins with a single consonant and a consonant cluster whose members are from the same series. For instance,

- ក 'neck' and ក (ក+ឃ) 'cup' are pronounced as /kaa/ and /kaa/ respectively because these words begin with first series consonant symbol (ក), and the first series of the inherent vowel is /aa/ and the first series of vowel symbol ឃ is /aa/.
- ន 'to be mute' and ន (ន+ឃ) are pronounced as /kɔɔ/ and /kie/ respectively because they begin with the second series consonant symbol (ន), and the second series of the inherent vowel is /ɔɔ/ and the second series of vowel symbol ឃ is /ie/.
- ក្រាត (ក+ឃ+ឃ+ឃ) 'head' is pronounced as /kbaal/ because ក /k/ and ឃ /b/ are in the first series, and the first series of vowel symbol ឃ is /aa/.
- ឯ (ឯ+ស+ឃ) 'first person singular (informal)' is pronounced as /knie/ because ឯ and ស are both in the second series, and the second series of the vowel symbol ឃ is /ie/.

After describing three accounts on the series assimilation, Sok (ibid:26-27) categorized the orthographic consonants into five groups: IMPLOSIVE, UNASPIRATED PLOSIVE, ASPIRATED PLOSIVE, FRICATIVE, and SONORANT (See the table below). The least sonorous consonant of an initial cluster determines the series of the following vowel (Huffman (1970:44), Kin (2007:63) and Chan (2010:107)).

The table below reads IMPLOSIVE consonants are the least sonorous and SONORANT consonants are the most sonorous. Those in between are ordered by increasing sonority. For instance. IMPLOSIVE consonants are less sonorous than UNASPIRATED PLOSIVE, and UNASPIRATED PLOSIVE are less sonorous than ASPIRATED PLOSIVE, and so forth.

	1st Series ជ ឃ
IMPLOSIVE	2nd Series ៩ -
	Phoneme ឃ ី
	1st Series ក ច ត - ខ[៥]
UNASPIRATED PLOSIVE	2nd Series ឬ ជ ទ ិ ញ -
	Phoneme ក ច ត ិ ញ
	1st Series ៩ ឃ ុ ុ
ASPIRATED PLOSIVE	2nd Series ឃ ឃ ឃ ឃ ឃ
	Phoneme ក ^h ច ^h ព ^h ិ ^h ឃ ^h
	1st Series ិ ិ
FRICATIVE	2nd Series - -
	Phoneme ស ឃ
	1st Series - - ិ ិ - - ិ ិ -
SONORANT	2nd Series ឬ ឬ ឬ ឬ ឬ ឬ ឬ
	Phoneme ឱ ឱ ឱ ឱ ឱ ឱ ឱ

For examples, the vowel of ឬិឬ 'papaya' takes the 1st series because the least sonorous consonant of the initial cluster is in the 1st series.

ឬិឬ 'papaya'	
character sequence	ឬ ិ ឬ
character mapping	ឬ h o u ិ
series of character	2nd 1st 1st 2nd 2nd
series of syllable	1st
phonemic transcription	/lhog/

ឬិឬ 'spur' is pronounced with the 1st series of the vowel governed by the series of the least sonorous initial consonant, ឬ /k^h/.

ឬិឬ 'spur'	
character sequence	ឬ ិ ឬ ិ
character mapping	ឬ k ^h n aajie j
series of character	1st 2nd 1st 2nd 2nd
series of syllable	1st
phonemic transcription	/knaaj/

The series assimilation does not only occur in the monosyllabic words, but also disyllabic words. The rule is the same, the vowel of the second syllable (a.k.a. major syllable) is determined by the series of the least sonorous initial consonant of the first syllable (a.k.a. minor syllable). However, this comes with a caveat.

Sok (ibid:29) describes:

Khin (2007:63) wrote that, in disyllabic words, if the initial consonant of the major syllable is one of these (ບ /p/, ດ /p/, ສ /n/, ດ /m/, ຕ /j/, ຕ /r/, ດ /l/, and ດ /w/), the vowel of the major syllable should be in the same series as the series of the initial consonant of the minor syllable. [...]

Exceptions:

- ແກ້ວ ‘vegetable’ is pronounced as /?an.luək/, not /?an.lak/ (if assimilation occurs)
- ໄລົງ ‘pestle’ is pronounced as /?ay.ree/, not /?ay.rae/ (if assimilation occurs)

For instance, ດນ ‘thorn’ is pronounced as /ban.laa/ not /ban.lie/ because the minor syllable is in the 1st series, and the initial of the major syllable is one of the consonant listed above.

ດນ ‘thorn’

character sequence	ດ ນ ດ ອ
character mapping	ດ n ດ aa je
series of character	1st 2nd 2nd 1st 2nd
series of syllable	1st 1st
phonemic transcription	/ban.laa/

Sok (ibid:30) continues:

[...] if each syllable of a disyllable word begins with a plosive or fricative, the series of each vowel is determined independently. Thus, for example, ດນຕຍ (ດ /d/, ນ /n/, ຕ /t/, ອ /aajie/, ຕຍ /j/) ‘barracks’ is pronounced as /ban.tiej/ (not /ban.taaj/) and ຖນຕຍ (ຖ /t/, ນ /n/, ຕ /s/, ອ /aajie/, ຕຍ /j/) ‘rabbit’ is pronounced as /tun.saaj/ (not /tun.siej/). Khin (2007:63) and Chan (2010:107) also agree that if the initial consonant of the major syllable is a strong consonant group (i.e. implosives, plosives, or fricative) then the series of the vowel of the major and minor syllable should be independent from one another (i.e. no series assimilation between the two syllables), as show in the example below.

ດນຕຍ ‘barracks’

character sequence	ດ ນ ຕ ອ ຕຍ
character mapping	ດ n ຕ aa je ຕຍ
series of character	1st 2nd 2nd 1st 2nd 2nd
series of syllable	1st 2nd
phonemic transcription	/ban.tiej/

ຖນຕຍ ‘rabbit’

character sequence	ຖ ນ ຕ ອ ຕຍ
character mapping	ຖ n ຕ aa je ຕຍ
series of character	2nd 2nd 1st 1st 2nd 2nd
series of syllable	2nd 1st
phonemic transcription	/tun.saaj/

5. Khmer Orthography

5.1. Series System

“Series” is the term used by Huffman (1970:15) to refer to two distinct group of consonants which determine how a syllable should be pronounced, 1st series and 2nd series. They are also called ສົ່ງເຫຼືອ ‘small voice’ and ສົ່ງເຫຼືອໃໝ່ ‘big voice’ or ສົ່ງເຫຼືອພිණ ສົ່ງເຫຼືອຜູ້ສັກ ‘light voice’ and ສົ່ງເຫຼືອຜູ້ສັກ ‘heavy voice’ respectively. Henderson (1952:151) refers to this term as “registers”, 1st register and 2nd register. The pitch of the 2nd register tends to be lower than the 1st register.

In an essence, the vowel quality is changed in harmony with the series of the initial consonant or consonant cluster while its orthographic representation stays the same. For example, consonant ក /k/ is in the 1st series and consonant ក /k/ is in the 2nd series. Vowel ា is realized phonemically as /aa/ in the 1st series and as /ie/ in the 2nd series. When ក is concatenated with ា, it creates a syllable which is pronounced as /kaa/. When ក is concatenated with ឃ, the syllable is pronounced as /kie/.

ក /k/ (1st series) + ា /aa|ie/ > កា /kaa/

ក /k/ (2nd series) + ា /aa|ie/ > កី /kie/

When a word begins with an initial consonant cluster, the [Series Assimilation](#) comes in to play--the least sonorous consonant determines the series of the vowel.

5.2. Phoneme-Grapheme Correspondence

Phoneme-grapheme correspondence of initial consonants, modified initial consonants, initial consonant digraphs, initial consonant clusters and final consonants are illustrated in tables in the following sections. Glosses are adopted from the Khmer-English dictionary (Headley 1997).

5.2.1 Initial Consonants

5.2.1.1. Initial Consonant Monographs

The table below illustrates the grapheme-phoneme correspondences with their Unicode code points, series and examples.

Grapheme Code Points Phoneme Series Examples

ក	U+1780	k	1st	កាត់ /kat/ ‘to cut’
ខ	U+1781	k ^h	1st	ខាត់ /k ^h at/ ‘to polish’
គ	U+1782	k	2nd	គាត់ /koat/ ‘3P’
យ	U+1783	k ^h	2nd	យោត់ /k ^h oat/ ‘to forbid’
ង	U+1784	ŋ	2nd	ងារ់ /ŋie/ ‘role’
ច	U+1785	c	1st	ចាត់ /cat/ ‘to designate’
ឆ	U+1786	c ^h	1st	ឆ្លើត់ /c ^h at/ ‘umbrella’
ជ	U+1787	c	2nd	ជាត់ /coat/ ‘to drain off (liquid)’
ធម	U+1788	c ^h	2nd	ធមោង /c ^h oop/ ‘to reach for (something)’
ញ	U+1789	n	2nd	ញ្ចាត់ /noat/ ‘to stuff (something) in tightly’
ជ	U+178A	c̥	1st	ជោង /daj/ ‘hand/arm’
ប	U+178B	t ^h	1st	បីន /t ^h aan/ ‘place/location’
ឌ	U+178C	t̥	2nd	ឌី /d̥i/ ‘to be obstinate’
ញ	U+178D	t ^h	2nd	ញាល /t ^h iel/ ‘shield’
ណ	U+178E	n	1st	ណាត់ /nat/ ‘to set (an appointment)’
ក	U+178F	t	1st	កោង /taoŋ/ ‘to hang on to’
ថ	U+1790	t ^h	1st	ថា /t ^h aa/ ‘to say/tell’
ទ	U+1791	t	2nd	ទា /tie/ ‘duck’
ធ	U+1792	t ^h	2nd	ធាត់ /t ^h oat/ ‘fat’
ន	U+1793	n	2nd	នានី /niel/ ‘unit of weight (~12oz.)’
ប	U+1794	b	1st	បាត់ /bat/ ‘to disappear’
ផ	U+1795	p ^h	1st	ផាត់ /p ^h at/ ‘to blow away (of the wind)’
ព	U+1796	p	2nd	ពោទេ /poat/ ‘to besiege’
ក	U+1797	p ^h	2nd	កិក់ /p ^h uək/ ‘mud’
ម	U+1798	m	2nd	មាត់ /moat/ ‘mouth’
យ	U+1799	j	2nd	យូរុង /joop/ ‘to pull upward’
វ	U+179A	r	2nd	វុរុង /roong/ ‘roofed structure’
ល	U+179B	l	2nd	លុន /lie/ ‘donkey’
វ	U+179C	w	2nd	វិវ /wie/ ‘to crawl’

ស	U+179F	s	1st	សំកិ /sak/ ‘hair’
ហ	U+17A0	h	1st	ហាក់ /hak/ ‘to jump’
ឡ	U+17A1	l	1st	ឡាន /laan/ ‘car/truck’
អ	U+17A2	?	1st	អាន /?aan/ ‘to read’

5.2.1.2. Subscript Consonants

A subscript (a.k.a. Coeng, *lit.* ‘foot’) is an alternate form of a consonant which is usually placed after a consonant to form a consonant cluster. In Khmer orthography, if two consonants are put one after another, that means the first consonant is the initial consonant and the other one is the final consonant. កិន and ីន are different and they are pronounced differently, /kaaq/ and /kjaa/ respectively.

A list of subscripts are included below. Thun (2011:17-18) wrote, “ឡូ and ឡូ must be written with their subscript (ឡូ and ឡូ), but the subscript of ឡូ (ឡូ) /l/ is never used. [...] When subscript ឡូ (ឡូ) is placed under itself, it has to be replaced by its full form, but smaller in size, i.e. ឡូ. Consonant ដុ /d/ and ពុ /t/ have the same subscript because most words written with ពុ /t/ are pronounced as ដុ /d/, for example, មាតាបិតា /mie.daa.bej.daa/, ពំណាង /dum.naaj/ [...].”

Grapheme

Unicode Code Points of the Subscript Phoneme Series Consonant Subscript

ក	ក	U+17D2	U+1780	k	1st
ខ	ខ	U+17D2	U+1781	k ^h	1st
គ	គ	U+17D2	U+1782	k	2nd
យ	យ	U+17D2	U+1783	k ^h	2nd
ង	ង	U+17D2	U+1784	ŋ	2nd
ច	ច	U+17D2	U+1785	c	1st
ឆ	ឆ	U+17D2	U+1786	c ^h	1st
ជ	ជ	U+17D2	U+1787	c	2nd
ធម	ធម	U+17D2	U+1788	c ^h	2nd
ញ	ញ	U+17D2	U+1789	jn	2nd
ឌ	ឌ	U+17D2	U+178A	d	1st
ប	ប	U+17D2	U+178B	t ^h	1st
ឌ	ឌ	U+17D2	U+178C	d	2nd
ឌ	ឌ	U+17D2	U+178D	t ^h	2nd
ធន	ធន	U+17D2	U+178E	n	1st
ទ	ទ	U+17D2	U+178F	t	1st
ទ	ទ	U+17D2	U+1790	t ^h	1st
ទ	ទ	U+17D2	U+1791	t	2nd
ធ	ធ	U+17D2	U+1792	t ^h	2nd
ន	ន	U+17D2	U+1793	n	2nd
ធម	ធម	U+17D2	U+1794	b	1st
ជ	ជ	U+17D2	U+1795	p ^h	1st
ធន	ធន	U+17D2	U+1796	p	2nd
រ	រ	U+17D2	U+1797	p ^h	2nd
ម	ម	U+17D2	U+1798	m	2nd
ធម	ធម	U+17D2	U+1799	j	2nd
រ	រ	U+17D2	U+179A	r	2nd
ធន	ធន	U+17D2	U+179B	l	2nd
វ	វ	U+17D2	U+179C	w	2nd
ស	ស	U+17D2	U+179F	s	1st
ហ	ហ	U+17D2	U+17A0	h	1st
ឡ	ឡ	U+17D2	U+17A1	l	1st

Subscript(s) may occur in the initial clusters, disyllabic words and final clusters. They can be stacked up to two in a row; no more than that is allowed.

Here are some examples of when subscript(s) occur. The subscripts are highlighted

- in initial cluster: ស្រីដែ = ស្រ 50 ី ដែ 54
 - in disyllabic word: បន្ទាន = ប ន 20 ធន 01
 - in final cluster: សិក្សា = ស ិ 20 ក្សា 30 ូ 02
 - after an independent vowel: ឥឡូវ = ឥ 20 ឡូ 02 ុ 03

In rare occurrences, independent vowel can be a subscript, i.e. ໂປ່ງ /so?.rit/ ‘buddy’.

5.2.1.3. Modified Initial Consonants

Khmer consonants can be changed from one series to another if they do not have a counterpart. Those consonants are listed below. The characters used to modify these consonants are described in the following section ([Inherent Vowels](#) and [Consonant Shifters](#)).

Here is a list of consonants which can be modified.

Grapheme	Code Points	Phoneme Series	Examples
ࡔ	U+1784 U+17C9	ࡔ	1st ແࡔ /gaj/ ‘day (colloquial)’
ࡕ	U+1789 U+17C9	ࡕ	1st ແࡕ /nam/ ‘to eat’
ࡖ	U+1798 U+17C9	ࡖ	1st ແࡖ /məən/ ‘ten thousand’
ࡗ	U+1799 U+17C9	ࡗ	1st ແࡗ /jaap/ ‘way, kind’
ࡈ	U+179A U+17C9	ࡈ	1st ແࡈ /raak/ ‘pulley’
ࡉ	U+179C U+17C9	ࡉ	1st ແࡉ /waj/ ‘hit, fight (colloquial)’
ࡊ	U+1794 U+17C9	ࡊ	1st ແࡊ /pən/ ‘to be skillful’
ࡋ	U+1794 U+17CA	ࡋ	2nd ແࡋ /but/ ‘topaz’
ࡌ	U+179F U+17CA	ࡌ	2nd ແࡌ /saəɒp/ ‘to investigate’
ࡍ	U+17A0 U+17CA	ࡍ	2nd ແࡍ /hien/ ‘to be bold’
ࡎ	U+17A2 U+17CA	ࡎ	2nd ແࡎ /?um/ ‘aunt, uncle’

5.2.1.4. Initial Complex Consonants

Eight initial complex consonants are used to transliterate foreign loanwords where there is no parallel single orthographic consonant. To fit into the Khmer spelling convention, they also have to have their counterparts which, in an essence, are modified by Consonant Shifter.

Grapheme	Unicode		Phoneme Series	Examples
ឃ	U+17A0 U+17D2 U+179C	f	1st	ឃុន /foon/ ‘phone’
ិរិ	U+17A0 U+17D2 U+179C U+17CA		2nd	ិរិល /fiil/ ‘film’
ិត្ត	U+179F U+17D2 U+17A0	j	1st	ិត្តិប៉ាប់ /jee.bap/ ‘Shebab’
ិត្តិ	U+179F U+17D2 U+17A0 U+17CA	j	2nd	ិត្តិអិតិ /jii.?iit/ ‘Shiite’
ឲរូ	U+17A0 U+17D2 U+179F		1st	ឲរូហិន /?aa.zaq.tin/ ‘Argentina’
ឲរូ	U+17A0 U+17D2 U+179F U+17CA	z	2nd	ឲរូម /zuum/ ‘zoom’
ឲន	U+17A0 U+17D2 U+1780	g	1st	ឲនាន /gaa.naa/ ‘Ghana’
ឲណ	U+17A0 U+17D2 U+1782	g	2nd	ឲណិម /geem/ ‘game’

5.2.2. Consonant Clusters

5.2.2.1. Two Consonant Clusters

This [spreadsheet](#) shows the instances of consonant clusters occurring in word-initial position (in transparent cells, for examples see [Appendix A](#)), in word medial position (in grey cells, for examples see [Appendix B](#)) and in final position (in yellow cells, for examples see [Appendix C](#)). Consonants on the top rows are the first member of the clusters (C_1), and those on the far left column are the second member (C_2), which are in subscript forms. The IPA corresponding to each consonant is included in the row and column next to C_1 and C_2 . The orthographic realization of the clusters are placed in the intersecting cell. For example, the first cluster (ຂ) is composed of ຂ and ແ. ກ and ແ, ຂ and ແ, ຕ and ແ, ເວ and ແ, and ແ and ແ do not make a cluster, which is why the intersecting cells are empty.

Note that ຕ /tʰ/ almost never have a subscript after it, only one instance that ຕ /tʰ/ is followed by a subscript as in ດັງ (ດ ຖ ດັງ).

Subscript ແ /l/ is not included here, as mentioned above that it is never been used in the contemporary Khmer.

ຂີ (or ຂົງ, as an alternative spelling) “to give” is a commonly used word that does not follow the Khmer spelling convention due to its historical background. It is the one and only instance that a subscript occurs after an independent vowel in contemporary Khmer.

ໜີ is the only instance that an independent vowel occurs as a subscript form. It is not commonly used though. For instance, ໜິກິພິ is usually written as ນິກິພິ.

Both ຂີ and ໜີ are found as headwords in the Chuon Nath Dictionary, the official Khmer-Khmer dictionary published in 1967.

5.2.2.2. Three Consonant Clusters

Initial consonant clusters composed of three consonants are rarely found in Khmer. They usually occur in foreign loanwords. There are three instances of three consonants in a row in the initial position in the headwords of the Khmer-Khmer dictionary.

- ປິນ- ສ ອ ປູ as in ປິນອຸ່ນ /strəj/ ‘woman’
- ປິນ- ບ ອ ປູ as in ປິນອຸ່ນ /frəŋ/ ‘French currency’

In the present-day Khmer, there are more instances of words with the initial clusters of three consonants. These are very rare as well. Here are a few examples:

- ປິນ- ສ ອ ປູ as in ປິນອຸ່ນ /sprinj/ ‘spring’
- ປິນ- ສ ອ ປູ as in ປິນອິນ /skriin/ ‘screen’
- ປິນ ສ ອ ປູ as in ປິນຕີຕີ /striit/ ‘street’
- ປິນ- ບ ອ ປູ as in ປິນອິນ /griin/ ‘green’
- ຫຼື- ບ ອ ອ as in ຫຼາສ ອ /flaah/ ‘flash’

Three consonant clusters in medial position:

- -ກົມ- ກ ອ ອ as in ດົກົມ້ /leak.sməj/ ‘wellness, glory’
- -ຜົມ- ຜ ອ ອ as in ດົມຜົມກົມ້ /pət.tjat.tʰe?.ka?.sat/ ‘vertebrae’
- -ສິນ- ສ ອ ປູ as in ສິນສິນຕີ /saŋ.skrot/ ‘Sanskrit’
- -ສິນ- ສ ອ ອ as in ດິນສິນ /bii.skwii/ ‘biscuit’
- -ຜິນ- ຜ ອ ປູ as in ດິນຜິນ /?aŋ.kraap/ ‘k.o. large red ant’
- -ຜິນ- ຜ ອ ອ as in ດິນຜິນ /saŋ.kjaap/ ‘counting’
- -ຜິນ- ຜ ອ ປູ as in ສິນຜິນ /saŋ.kruəh/ ‘to save/rescue’
- -ຜິນ- ຜ ອ ອ as in ດິນຜິນ /?aŋ.kleh/ ‘English’
- -ຜິນ- ຜ ອ ປູ as in ດິນຜິນ /saŋ.kreac/ ‘monk chief’
- -ຜິນ- ຜ ອ ປູ as in ດິນຜິນ /saŋ.cram/ ‘to chop repeatedly’
- -ຜິນ- ຜ ອ ປູ as in ດິນຜິນ /kan.crooŋ/ ‘fox’
- -ໃນ- ອ ອ ປູ as in ດິນໃນ /kan.trieŋ/ ‘chief of a clan’
- -ໃນ- ອ ອ ອ as in ດິນໃນ /san.tjie/ ‘twilight’
- -ໃນ- ອ ອ ປູ as in ດິນໃນ /kan.traj/ ‘scissors’

Three consonant clusters are also found in the final position.

- -ໃນ ອ ອ ປູ as in ດິນໃນ /kan.traa/ ‘wheelless pulley’
- -ໄນ ອ ອ ປູ as in ດິນໄນ /daa.raa.sah/ ‘astronomy’
- -ໃນ ອ ອ ປູ as in ດິນໃນ /so?.reen/ ‘Indra’
- -ໃນ ກ ອ ປູ as in ດິນໃນ /pʰeak/ ‘face’

ສ /r/ seems to play an important role in the three consonant clusters. According to the Chuon Nath dictionary data, it usually occurs as the last member of the cluster, but never to be found in between the first and the last member, or it is not pronounceable. For instance, ស្រីរ should be typed as it is spelled, like (ស /s/+ រ /p/+ ឃ /r/+ ឃ /ii/+ ឃ/y/ → /spriij/), not (ស /s/+ ឃ /r/+ ស /p/+ ឃ /ii/+ ឃ/y/ → /srpiij/). For a list of 50 instances of words with three consonant in the final position (i.e. one consonant followed by two subscripts), please refer to [Appendix D](#).

5.2.3. Final Consonants

A final consonant occurs after a vowel in an orthographic syllable. For instance, ក /k/ in សំខាន់ /s?ae/ ‘tomorrow’ is a final consonant because it is in the coda position of the syllable and it is pronounced after the vowel េ /ae/. However, it is impossible to do syllable analysis algorithmically because of the ambiguities in determining where the syllable boundaries in a sentence given that spaces are not used in between words in Khmer.

According to the data in the Khmer-Khmer dictionary, it is observed that six orthographic consonants are not found in the coda position of orthographic words. They are ឃ /c^h/, ុន /ch^h/, ធម /p^h/, ុន /h/, ុន /l/ and ុន /ʔ/. There are instances which could be confusing whether they are a final consonant or an initial consonant. For instance, ក្រាលូ ‘jug’ is composed of three orthographic characters: ក ូ លូ, and it may be thought that លូ is the final consonant. It is actually not a case. It is, instead, the initial consonant of the second syllable with an inherent vowel. The word is pronounced as /kra.laa/, not /kraal/.

Consonant clusters are found in the coda position of Pali/Sanskrit loanwords. These are unpredictable and the combination can be composed of any two or three consonants which are always left unpronounced, unless an orthographic vowel is placed after them.

សាស្ត្រិយ /sah/ ‘science or subject’ (no vowel after the final consonant cluster)

សាស្ត្រិយិ /sah.straa/ ‘palm leaf manuscript’ (with a vowel after it)

សាស្ត្រិយិ does not have a vowel after the cluster -ិយិ, which is why it is left out of the pronunciation; សាស្ត្រិយិ, however, has ិ after the cluster (-ិយិ). In this case, another syllable is created--ិយិ /straa/.

F = {ក ច ឈ ឯ ធម ុ ុន ុន ុន ុន ុន ុន} or
a random consonant clusters of two or three consonants (See [Consonant Clusters](#))

It is important to note that, unlike old Khmer, subscript never occurs after a dependent vowel in the present day Khmer. Here are some examples from Bernard (1902:40-42) and their corresponding words from Khmer Choun Nath Dictionary (1967).

Old Khmer	Present Day Khmer	Gloss
ចក្រិយ	ចក្រិយ /ca?.krej/	‘chakri’
ចំហាង	ចំហាង /cam.haaj/	‘steam (n)’
ចំហុយ	ចំហុយ /cam.hoj/	‘steam (v)’

Bernard (ibid) mostly used subscript ិយ after dependent vowel. For instance, ចំហុយ (with consonant ុយ) ‘to steam’ was not written as ចំហុយ (with subscript ិយ).

Maspero (1915:38) explained in his book entitled “Grammar of Khmer Language” that the subscript ិយ after a dependent vowel is a semi-vowel. For examples:

នី	នី /nii/	‘female’
ដី	ដី /dej/	‘soil’
ផ្លូវ	ផ្លូវ /cej/	‘fig tree’

He further gave more examples that Sanskrit loanwords with “aya” at the end are written with ិយ and a final subscript ិយ. For examples:

ជ័យ “jaya”	ជ័យ /cej/	‘victory’
ក្រុយ “bhaya”	ក្រុយ /p ^h ej/	‘fright’

Maspero (ibid:70,138) also described the used of subscript ិ can also be placed after a dependent vowel, and it is equivalent to the final consonant. For example

ទៅ = ទៅពេ /teap/ ‘all’

5.2.4. Vowels

Vowels are divided into three types in this paper: Inherent Vowels, Dependent Vowels and Independent Vowels.

5.2.4.1. Inherent Vowels

Inherent vowels are invisible, so they do not have their orthographic representation. Two inherent vowels are used in the present day Khmer, consonants in the 1st series take /aa/, and those in the 2nd series take /ɔə/. Therefore each Khmer orthographic consonant should be pronounced as shown in the table below. The grey cells shows the modified consonants in correspondent with their counterparts. Explanation as to how they are modified, see [Consonant Shifters](#).

1st Series		2nd Series	
Orthographic	Phonemic	Orthographic	Phonemic
ក	/kaa/	ក	/kɔə/
ខ	/kʰaa/	ខ	/kʰɔə/
ឆ	/jaa/	ឆ	/jɔə/
ច	/caa/	ច	/cɔə/
ឆ	/cʰaa/	ឆ	/cʰɔə/
ញ	/jaaa/	ញ	/jɔɔ/
ជ	/daa/	ជ	/dɔə/
ថ	/tʰaa/	ថ	/tʰɔə/
ណ	/naa/	ណ	/nɔə/
ព	/taa/	ព	/tɔə/
ថ	/tʰaa/	ថ	/tʰɔə/
ប	/baa/	ប	/pɔə/
ផ	/pʰaa/	ផ	/pʰɔə/
យ	/jaa/	យ	/jɔə/
រ	/raa/	រ	/rɔə/
ឃុំ	/laa/	ឃុំ	/lɔə/
ឃ	/waa/	ឃ	/wɔə/
ស	/saa/	ស	/sɔə/
ហ	/haa/	ហ	/hɔə/
អ	/?aa/	អ	/?ɔə/

Inherent vowels in Pali/Sanskrit loanwords are different from those of Khmer. They are /a?/ in 1st series and /ea?/ in the 2nd series. Therefore, ក /kaa/ and ក /kɔə/ in Khmer are pronounced as /ka?/ and /ke?/ respectively.

5.2.4.2. Dependent Vowels

There are two types of them: dependent vowels and independent vowels. The first type is always attached to an initial consonant, and the second one can start a syllable on its own. It is important to note that different vowels stand in various places around the base consonant. It can go to the left, right, above, below and around the base. Please see the examples below.

Position Example Character Combination IPA

left	កើ	ក ើ	/keə/
right	កា	ក ា	/kaa/
above	កិ	ក ិ	/ke?/
below	កុ	ក ុ	/ko?/
around	កីរូ	ក ីរូ	/kiə/

The table below shows the vowels as taught in school (Um & Seng 2012:1).

Dependent Vowels

Grapheme	Unicode	1st Series	2nd Series
(inherent) -		ក /kaa/ ‘neck’	ព /koo/ ‘to be mute’
ត	U+17B6	ពា /taa/ ‘grandfather’	ទា /tie/ ‘duck’
ច[8]	U+17B7	បិត /pət/ ‘to cut obliquely’	ពិត /pit/ ‘to be true or real’
ស	U+17B8	សី /səj/ ‘shuttlecock’	សិី /sii/ ‘to eat (impolite)’
ទ	U+17B9	ពីក /tək/ ‘10 cm’	ទីក /tik/ ‘water’
កែ	U+17BA	កី /kəə/ ‘heart(s) (suit of ‘cards)’	កី /kiə/ ‘to be’
កូ	U+17BB	កូក /kok/ ‘egret’	គូក /kuk/ ‘prison’
កូ	U+17BC	កូវ /koɔ/ ‘to stir’	គូវ /kuu/ ‘to draw’
កូ	U+17BD	កូវ /kuə/ ‘pod’	គូវ /kuə/ ‘should’
តី	U+17BE	តី /taə/ ‘question marker’	ទី /təə/ ‘to be held’
ជីវី	U+17BF	ជីវី /tiə/ ‘dwarf’	ដីវី /ciə/ ‘to believe’
ទីរី	U+17C0	ទីរីប /tiəp/ ‘k.o. bowl’	ទីរីប /tiəp/ ‘custard apple’
កើង	U+17C1	កើង /keenj/ ‘to cheat or exploit’	តើង /keenj/ ‘to sleep (for kid)’
កើន	U+17C2	កើន /kae/ ‘to edit or adjust’	តើន /keə/ ‘craw (of birds)’
កើន	U+17C3	កើន /kaj/ ‘trigger’	តើន /kej/ ‘to be frugal’
ការ	U+17C4	ការ /kaə/ ‘to shave’	តោ /koo/ ‘cow’
តោ	U+17C5	តោ /taw/ ‘unit of measurement’	ទោ /tiw/ ‘to go’
កុំ	U+17BB U+17C6	កុំ /kom/ ‘(don’t)’	តុំ /kum/ ‘to hold a grudge’
កុំ	U+17C6	កុំ /cam/ ‘to be exactly at’	ជុំ /cum/ ‘to discuss’
កាំ	U+17B6 U+17C6	កាំ /cam/ ‘to remember’	ជាំ /coam/ ‘to bruise’
កោះ	U+17C7	កោះ /tah/ ‘hurriedly’	ឆោះ /teah/ ‘to slap’
កោះ	U+17BB U+17C7	កោះ /coh/ ‘to go down’	ជោះ /cuh/ ‘to defecate’
កោះ	U+17C1 U+17C7	កោះ /ceh/ ‘to know’	ជោះ /ceh/ ‘to chip (of skin/bark)’
កោះ	U+17C4 U+17C7	កោះ /kah/ ‘island’	តោះ /kuoh/ ‘to knock’

These vowels (i.e. នី នីវី ទីរី កោះ កោះ កោះ) are composed of more than one part of vowels with similar look, but they do not have the quality of the each part (Kin 2007:45). That is to say, when writing, we write two parts of each of them by lifting up the pen. For instance, នី is made up of នី /ee/ and វី /əjii/, but the realization of នី is /aə/, not /ee+əj/.

Vowels whose second member is ំ (a.k.a. Nikahit) or ោះ (a.k.a. Reahmuk) always have /m/ or /h/ in the final position. Nikahit and Reahmuk act as a final /m/ and /n/ respectively. It marks the end of the syllable and cannot be followed by a vowel.

A sign behaves like a vowel, ោះ (a.k.a. Yuukaleapintu), yet it is not counted as a vowel. It has its 1st and 2nd series realizations, and it never takes a final consonant.

Grapheme	Code Points	1st Series	2nd Series
ោះ	U+17C8	ស្ថិតិះ /so?taʔ/ ‘listening (to)’	ស្ថិតិះ /sot.teaʔ/ ‘laborer’

Yuukaleapintu can also be used to combine with another vowel just like Nikahit and Reahmuk.

Grapheme	Code Points	1st Series	2nd Series
ខោះ	U+17C2 U+17C8	ខោះ /heʔ/ ‘(startle)’	ខោះ /?heʔ/ ‘(interjection)’

These are the vowels that do not exist in the official alphabetical order.

Grapheme	Code Points	1st Series	2nd Series
ិះ	U+17B7 U+17C7	ពិះ /teh/ ‘to insult’	ពិះ /cih/ ‘to ride’
ិះ	U+17B9 U+17C7	ផិះ /ckəh/ ‘to pluck out’	ផិះ /cpih/ ‘to be rough (of fabrics)’
ិះ	U+17BA U+17C7	ឃិះ /creəh/ ‘to be tightly packed’	ឃិះ /priih/ ‘to be spirited’

ឃោះ	U+17C2 U+17C7	ឃោះ /craeh/ ‘rust’	ឃោះ /creəh/ ‘to chop away’
តាំង	U+17B6 U+17C6 U+1784	តាំង /taŋ/ ‘from/since’	តាំង /teŋ/ ‘together with’

5.2.4.3. Independent Vowels

Independent vowels are able to start a syllable without an initial consonant or initial consonant cluster (Um & Seng 2012:3). They are usually used in Pali/Sanskrit loanwords. The corresponding phonemic transcription to each grapheme are in **bold** next to the example words. Independent vowel are rarely used in any new words in the contemporary Khmer language. Even in the 1990s, there were not many either. See how often they are used in the headwords in the Chuon Nath dictionary below.

Independent Vowels

Grapheme Code Points Examples Frequency

ឃ	U+17A5	ឃិច្ច /?ət/ ‘brick’ ឃិច្ចីវ /?əj.ləw/ ‘now’	75
ឃី	U+17A6	ឃីសាន /?əj.saan/ ‘north east’	5
ឃុ	U+17A7	ឃុកញ្ញា /?ok.naa/ ‘tycoon’ ឃុបមា /?uʔ.paʔ.maa/ ‘to be comparable to’	242
ឃុ	U+17A9	ឃុឃ /?uu.ru?/ ‘chest (royalty)’	7
ឃុ	U+17AA	ឃុ /?əw/ ‘father’	13
ឃុ	U+17AB	ឃុក /rik/ ‘behavior’	38
ឃុ	U+17AC	ឃុ /rii/ ‘or’	2
ឃុ	U+17AD	ឃុកិ /rɔ.lik/ ‘to be alert’	3
ឃុ	U+17AE	ឃុ /lii/ ‘to hear’	8
ឃុ	U+17AF	ឃុក /?aek/ ‘one’	28
ឃុ	U+17B0	ឃុីវ /?aj.daa/ ‘LOC’	7
ឃុ	U+17B1	ឃុីវ or ឃុ[9] /?aoj/ ‘to give’	55
ឃុ	U+17B2	ឃុីវ /?aoj/ ‘to give’	0
ឃុ	U+17B3	ឃុកិ /kra.ʔaw/ ‘edible root of the lotus’	6

In the present day, there is no evidence of an instance where an independent vowel is used as a subscript in the contemporary Khmer, even though there are three instances of ឃ being used as a subscript in three headwords (i.e. សុព្រឹទ, សែវព្រឹទ, and បុទ្ទិយ), plus more than 190 instances of Sanskrit transliteration in the Chuon Nath dictionary. Here are some examples:

Khmer Contemporary Spelling	IPA	Sanskrit Transliteration	Gloss
ក្រិះ (ក ឃ ុ ះ ស)	/krəh/	ក្រិះ (ក ឃ ុ ិ ន)	‘small, dwarf’
គ្រឿង (គ ឃ ុ ិ ៗ ៖)	/krih/	គ្រឿង (គ ឃ ុ ិ ហ)	‘house’
ព្រើស្ស (ព ឃ ុ ិ ក ឃ ុ)	/priʔ.saa/	ព្រើស្ស (វ ឃ ុ ិ ក ឃ ុ)	‘tree’
មុស (ម ឃ ុ ស ៣)	/muʔ.saa/	មុស (ម ឃ ុ ិ ស)	‘to be false’
វិនិយិ (វ ឃ ុ ិ ន ិ យ ិ)	/wut.hiʔ?/	វិនិយិ (វ ឃ ុ ិ ន ិ យ ិ)	‘prosperity’
សាតិ (ស ឃ ុ ិ)	/saʔ.teʔ?/	សាតិ (ស ឃ ុ ិ ត ិ)	‘consciousness’
សំស្រួត (ស ឃ ុ ិ ស ឃ ុ ិ ត)	/sap.skrət/	សំស្រួត (ស ឃ ុ ិ ស ឃ ុ ិ ត)	‘Sanskrit’
ស៊ិនិក (ស ឃ ុ ិ ស ឃ ុ ិ ក ិ)	/srəŋ.kie/	ស៊ិនិក (ឥ ឃ ុ ិ ស ឃ ុ ិ ក ិ)	‘love, lover’
ខ្មោះ (ខ ឃ ុ ិ ម ឃ ុ)	/?uʔ.krot/	ខ្មោះ (ខ ឃ ុ ិ ម ឃ ុ)	‘seriously, criminally’

5.3. Diacritics (signs)

Eight diacritical signs are currently used in the present-day Khmer. They are Muusikatoan (〃), Triisap (˜), Bantoc (៥), Robat (˜), Kakabat (៥), Ahsda (៥), Samyok Sannya (៥) and Toandakhiat (៥). These play different roles in the Khmer spelling conventions. The following sections describe the usage of each of these. The following table shows the frequency of each diacritic as occurred in the headwords of the Chuon Nath Dictionary. *Usually, there is no occurrence of two diacritics in a row, except a combination of Consonant Shifter and a Samyok Sannya.* (See [Special Treatment of Consonant Shifters](#))

Diacritic Name Diacritic Character Code Points Frequency in KKD

Muusikatoan	〃	U+17C9	722
Triisap	˜	U+17CA	215
Bantoc	៥	U+17CB	1801
Robat	˜	U+17CC	83
Kakabat	៥	U+17CE	15
Ahsda	៥	U+17CF	10
Samyok Sannya	៥	U+17D0	384
Toandakhiat	៥	U+17CD	248

5.3.1. Consonant Shifters

Consonant Shifters are also known as “Series Shifters” or “Register Shifters”. As the name suggests, Consonant Shifters change the series of the consonant from the 1st to the 2nd or vice versa. This is done in order that the Khmer consonant chart is complete and each consonant has their corresponding counterpart as you can See Inherent Vowels above.

Consonant Shifters can also be used to change the series of consonant clusters. It is usually placed after the cluster, for it changes the series of the cluster, not either of the first nor the second alone (See [Consonant Shifters with Consonant Clusters](#)).

5.3.1.1. Muusikatoan (〃)

Muusikatoan (〃), a.k.a. ព្រៃកណូន ‘Thmenh Kandol’ or សមលប់ពីី ‘Samlab Pii’, is used to change the series of the consonants -- from the 2nd series to 1st series. Not any consonant can be used with Muusikatoan. It is only applicable with the ones that does not have their counterpart in the 1st series.

ឃ /gəo/ + 〃 > ឃ /gaa/
 ញ /ŋəo/ + 〃 > ញ /ŋaa/
 ម /məo/ + 〃 > ម /maa/
 យ /jəo/ + 〃 > យ /jaa/
 ឃ /rəo/ + 〃 > ឃ /raa/
 ឃ /wəo/ + 〃 > ឃ /waa/
 ឃ /θəa/ + 〃 > ឃ /θaa/

5.3.1.2. Triisap (˜)

Triisap (˜), a.k.a. សកោក ‘Sakka’, is used to change the series of the consonants -- from the 1st series to the 2nd series.

ឃ /θaa/ + ˜ > ឃ /θəo/
 ស /saa/ + ˜ > ស /səo/
 ហ /haa/ + ˜ > ហ /həo/
 ឃ /?aa/ + ˜ > ឃ /?əo/

5.3.1.3. Special Case of ឃ

ឃ is the only consonant that either Muusikatoan or Triisap can be attached to, and Muusikatoan does not change the series of it, it changes the consonant instead; On the other hand, Triisap does changes the series of ឃ from the first to the second series.

ឃ /θaa/
 ឃ ˜ ឃ /θəo/

ឃ៊ ឃុ /paa/

Another special case for ឃុ is that even though its consonant quality is /b/, it would be realized as /p/ when it precedes subscript ូ in a cluster, i.e. ព្រឃុ ‘salty’ is pronounced as /praj/ not /braj/.

5.3.1.4. Special Treatment of Consonant Shifters

According to the spelling convention (Khin 2017:44), Muusikatoan or Triisap has to be turned into a glype that looks like ឃុ when there is a vowel on top of the base consonant (a.k.a. Above Vowels). Khin (ibid) does not give a list of these vowels, but other scholars do include them. The ឃុ is not U+17BB, but it is a rendering version of the consonant shifter. Um & Seng (2012:52) and Kul (2008:28-31) state that ឃិ (U+17B7), ឃី (U+17B8), ឃឹ (U+17B9), ឃុ (U+17BA), ឃុំ (U+17BE) and ឃុំំ (U+17B6 U+17C6) are the Above Vowels that fall into this special case.

Kul (ibid) and Nuon (1954:សំណើរបាយ) describes a few exceptional cases of when Shifters should or should not turn into the character that looks like ឃុ.

- One caveat is that when Triisap is attached to ឃុ and they followed by one of the Above Vowels mentioned above, Triisap does not change its orthographic realization to ឃុ. It should stay the same. If changing, it would get confused with Muusikatoan.

ឃិក	ឃែ ឃិ ក	/b̥ic/	‘pen’
ឃិន	ឃែ ឃិ ន	/b̥in/	‘cooler box’
ឃិន	ឃែ ឃិ ន	/pən/	‘to be keen on something’

- Triisap may stay the same when attached to ឃុ even though there is an above vowel after it.

ឃិម៉ា[10]	ឃែ ឃិ ម៉ា	/?im/	‘to be breastfed’
ឃុំ	ឃែ ឃុំ	/?um/	‘uncle/aunty’

- Triisap may stay the same or be changed when there is an above vowel attached to it. This is only with ឃុ (U+17A0).

ឃិម	ឃែ ឃិ ម	/him/	‘(sigh)’
-----	---------	-------	----------

ឃិម ឃែ ZWNJ ឃុ ម (The ZWNJ is placed in between the Triisap and the vowel in order to prevent the default rendering from happening.)

- For the above two cases of Triisap~ being used with ឃុ and ឃុំ, Um & Seng (2012:52) rules out that the Triisap has to change its glype to look like ឃុ, i.e. ឃិម, ឃិកអិក។
- Samyok Sannya (ឃុំ) is seen to influence the orthographic realization of the consonant shifter in the same way the above vowels do. However, no reference has been found to describe this phenomenon.

ឃិដ្ឋ	ឃែ ឃិ ឌុ	/pag/	‘bread’
ឃិវ	ឃែ ឃិ វ	/bəə/	‘butter’

- Disyllabic words should never take a consonant shifter in the second syllable. The series assimilation does the job of changing the series of the second syllable which results in changing the vowel of the syllable (Sok 2016:28-30).

ត្រូវឃិដ្ឋ /cra.gyaap/ ‘scatter in the way’

not ត្រូវឃិដ្ឋ

ត្រូវឃិវ /pro.həən/ ‘to be insolent’

not ត្រូវឃិវ nor ត្រូវឃិវីន

- When the Triisap is followed by this vowel (ឃុំ), one has to change the second part of the vowel (i.e. ឃុំ) to ឃុ (Kul ibid:32).
 - ឃុំ should be written as ឃិដ្ឋ
 - ឃិវ should be written as ឃិវីន

However, both spellings are found in everyday use. It may be because this rule has been overlooked.

5.3.1.5. Consonant Shifters with Consonant Clusters

Consonant Shifters can also be used with consonant clusters when the desired vowel is not the default one that determine by the least sonorous consonant or the cluster composed of consonants from the same series. The consonant shifter is usually placed in between the subscript and the vowel. To some, it should go after the base consonant, but in fact the shifter does not merely change the series of one of the two members. It changes the series of the cluster as a whole.

- ស្រីប សំ ចូវេះ ឬ ឧ /srip/ ‘sound of a heavy, solid object falling, thud’
as opposed to ស្រីប /srəp/
- ស្អួល សំ ឃុំ ឬ ឃុំ /s?uj/ ‘to be fat (of the belly)’
as opposed to ស្អួល /s?oj/
- ម៉ែ ម៉ែ ឬ ម៉ែ /mjaj/ ‘one day’
as opposed to ម៉ែ /mnej/
- ស្អិម សំ ឃុំ ឬ ម៉ែ /spəm/ ‘to be stuck at the mouth or entrance’
as opposed to ស្អិម /sbəm/

For list of all possible initial consonant clusters with their corresponding counterparts, please see [Appendix E](#).

5.3.2. Bantoc (៥)

Sok (2016:66-68) called this phenomenon “Vowel Modification”. Here is a quote of the account:

Two vowels could be modified by adding a diacritic to it. The diacritic used is called Bantoc "៥". The Bantoc is used to shorten as well as change the vowel quality completely. The two vowels are:

- Inherent vowels in the first series and the second series
- Ꭻ in the first series and the second series

Bantoc is usually placed on the final consonant, and only certain orthographic consonants could have the Bantoc. They could be one of these nine final consonants: -ន /k/, -នុ /ŋ/, -ធន /c/, -ញ /p/, -និ /t/, -ស /n/, -នុវ /l/, -នុវ /s/, or -ធន /b/. [...]

No.	Syllable Structure	1st Series	Syllable Structure	2nd Series
1	a		b	
	CF៥	/aa/ is shortened to /a/	CF៥	/ɔə/ is changed to /uə/
	៥F៥	Where	៥F៥	Where F = -និ/-នុ/-ធន /ɔə/ is changed to /u/
	៥CCF៥	F = -និ/-នុ/-ធន/-ញ/-ញិ/-ធន/-ធន	៥CCF៥	(else where)
2	C៥F៥		៥C៥F៥	F = -និ/-នុ/-ធន/-ញ/-ញិ/-ធន
	៥C៥F៥	/aa/ is shortened to /a/	៥C៥F៥	/ie/ is changed to /ea/
	Where		៥C៥F៥	Where
	៥CC៥F៥	F = -និ/-នុ/-ធន/-ញ/-ញិ/-ធន/-ធន	៥CC៥F៥	/ie/ is changed to /oa/
	៥C៥៥F៥		៥C៥៥F៥	Where
				F = -និ/-ស/-ុ/-នុ/-នុវ/-ធន

1a. The first series inherent vowel /aa/ is shortened to /a/ when the Bantoc is placed on the final consonant. For example, ននា ‘to be harden/frozen’ is transcribed as /kaak/, while ននា ‘to wash (hair)’ is transcribed as /kak/.

ននា ‘to be harden/frozen’

character sequence	ន	ន
character mapping	k aa ɔə	k
series of character	1 1/2	
series of syllable	1	
phonemic transcription	/kaak/	
ននា ‘to wash (hair)’		
character sequence	ន	ន៥
character mapping	k aa ɔə	៥

<i>series of character</i>	1	1/2
<i>series of syllable</i>	1	
<i>phonemic transcription</i>	/kak/	

1b. The second series inherent vowel /ɔɔ/ is changed to /uə/ when the Bantoc is placed on one of the these two final consonants (-ឃ /k/ or -ឃុ /s/), and it is changed to /u/ when the Bantoc is placed on any other final consonants. Therefore, for instance, ឃឃា 'to wash (clothes)' is transcribed as /kuək/ (See the illustration below)

ឃឃា 'to wash (clothes)'

<i>character sequence</i>	ឃ	ឃ	ោ
<i>character mapping</i>	k	aa ɔɔ	k
<i>series of character</i>	2	1/2	
<i>series of syllable</i>	2		
<i>phonemic transcription</i>	/kuək/		

2a. The first series vowel ឃ /aa/ is shorten to /a/ when the Bantoc is placed on the final consonant. Thus, as illustrate below, ឃឃា 'coin' is transcribed as /akak/.

ឃឃា 'coin'

<i>character sequence</i>	ឃ	ឃ	ោ
<i>character mapping</i>	k	aa ie	k
<i>series of character</i>	1	1/2	
<i>series of syllable</i>	1		
<i>phonemic transcription</i>	/akak/		

2b. The second series vowel ឃ /ie/ is changed to /ea/ when the Bantoc is placed on -ឃ /-k/, -ឃុ /-g/, -ឃុុ /-c/, or -ឃុុុ /-j/ as in ឃឃា 'to trap' which is transcribed phonemically as /teak/ (see the illustration of ឃឃា below); and it is changed to /oa/ when the Bantoc is placed on any other final consonant as in ឃឃាគ '3S' phonetically transcribed as /koat/ (see the illustration of ឃឃាគ below).

ឃឃា 'to trap'

<i>character sequence</i>	ឃ	ឃ	ោ
<i>character mapping</i>	t	aa ie	k
<i>series of character</i>	2	1/2	
<i>series of syllable</i>	2		
<i>phonemic transcription</i>	/teak/		

ឃឃាគ '3S'

<i>character sequence</i>	ឃ	ឃ	ោ
<i>character mapping</i>	k	aa ie	t
<i>series of character</i>	2	1/2	
<i>series of syllable</i>	2		
<i>phonemic transcription</i>	/koat/		

Bantoc never occurs after a consonant cluster.

Bantoc is never used in Pali/Sanskrit loanwords.

* ឃឃាគ /ko?.hak/ 'to tell a lie'

not ឃឃាគ * ឃឃាគ /so?.kut/ 'to die (royal register)'

Note: The combination of ឃ ឃ ោ is not allowed in Khmer spelling convention (Kol 2008:33). In cases uəlike that, ឃោ ឃ is used instead.

- * ព័ត៌ម្លេ /tag/ ‘to exhibit’
- not ពាត់ * ព័ជ្រ /teay/ ‘both, all’
- not ពាត់

5.3.3. Robat (៥)

According to the Khmer-English dictionary, Robat (៥), which literally means ‘subscript ៥ (៥)’; however, it looks like the top part of the consonant (៥), is a diacritical mark which indicates an orthographic ‘r’ in some words borrowed from Sanskrit. It is never used in Pali loanwords nor Khmer words. It is used in the same way as Repha in Sanskrit, but it is silent in Khmer (i.e. It does not have any audible pronunciation). It also silents the final consonant which has no vowel after it, but when a vowel is attached to the consonant, the consonant should be pronounced as normal.

Robat is always put on the consonant, but never on a subscript. It may be followed by a vowel. It never occurs after a consonant cluster. No sign or vowel should be put above it.

ធនី /koə/	‘pregnancy’
ធម៌ /tʰoa/	‘Dharma’
មាតិ /mie.kie/	‘road, way’

5.3.4. Kakabat (៥)

There is no reference on where Kakabat (៥) should be placed. In a manual on how to type Khmer Unicode characters, it is said to be placed after a consonant, a subscript or a dependent vowel. Words with Kakabat have to be pronounced with loud voice and the speaker may use high or low pitch (Khin 2007:72).

Here are some examples of words (from the Choun Nath Dictionary) written with Kakabat.

Khmer	IPA	Character	Gloss
កំក !	/kook/	ក៥ ក !	‘sound made to catch others’ attention’
កំកុំ	/kook.kook/	ក៥ ក៥ ក៥ ក៥	
ណាំ !	/naa/	ណ៥ ា៥ ំ !	
ណាំ:[1]	/nah/	ណ៥ ា៥ ០៩ : !	‘(affirmation)’
ហ្លែះ !	/nah/	ហ៥ ្លែ៥ ះោះ !	
អ៊ែះ !	/?oəh/	អᅕ ែᅕ ះោះ !	‘(exclamation denoting surprise)’
អំក	/?ok/	អᅕ ំᅕ ក	‘to blame’
អ៉ែ !	/?ao/	អᅕ ៉ែᅕ ែ !	‘(exclamation of distress)’
អូី !	/?hih/	អᅕ ូᅕ ZWNJ ីᅕ ះោះ ០៩ :	‘(exclamation of when having a burden)’
ងាំ	/c ^h aa/	ងᅕ ាᅕ ំ	‘(a word used to refrain in a song)’

5.3.5. Ahsda (៥)

Ahsda (៥) is restrictedly used on five consonants: កᅕ ដᅕ សᅕ មᅕ បᅕ. They are pronounced with their respective inherent vowel. It is used to disambiguate the single character words

Khmer	IPA	Gloss
កᅕ	/kaa/	‘also’
ដᅕ	/daa/	‘so, very’
សᅕ	/naa/	‘(usually used with ឌᅕ:)’
មᅕ	/moo/	‘to come (colloquial)’
បᅕ	/haa/	‘(say this when passing something to someone)’

Ahsda never occurs after a consonant cluster.

There is only one occurrence of when Ahsda is placed after a dependent vowel: ទីវា ! /tiw/ ‘go’. In the present day Khmer, Ahsda has never been used after any vowel, but if it does, it should still be rendered above the consonant. If there is an above vowel, it should be rendered right above the vowel above the consonant.

5.3.6. Samyok Sannya (េ)

Samyok Sannya (េ) is usually placed after a consonant, a subscript or after a consonant shifter. \

Khmer	Character	IPA	Gloss
ដោរ	ជោរេ	/coa/	‘rubber’
បុំដុំ	បុំដុំេ	/pag/	‘bread’
ត្រូវយ៉ា	ត្រូវយ៉េយ	/traj/	‘threefold’

Samyok Sannya is equivalent with vowel /a/ when the consonant preceding it is in the 1st series, but it can realize as /ea/, /e/ or /oa/ in other environments.

ខេះ ខែសកេន	sword (royalty)	ពោនា	Pear ethnic group
លីខេះ	leak	អិខេះ	artificial dye
នីយេ	nej	នីយេ	meaning or sense
បុំដុំបុំដុំ	pag	បុំដុំបុំដុំ	bread
	tii.neap	ពីឱិនីង	seat (royalty)

Samyok Sannya (េ) in the present-day Khmer has another phonetic property, it is [əə]. Most words borrowed from English with the rhotacization at the end usually are transliterated into a consonant with Samyok.

Khmer	Character	IPA	Gloss
មាស្ថ្ទៃ	ម៉ាស្ថ្ទៃេ	/maa.stəə/	‘master’
កំពួច្ច់រ	កំពួច្ច់រេ	/kom.pjuu.təə/	‘computer’
កំដួច្ច់រ	កំដួច្ច់រេ	/kon.təə/	‘counter’
គូលេរ	គូលេរេ	/kuu.ləə/	‘color’

5.3.7. Toandakhiat (េ)

Toandakhiat is used to silence the character, which usually is a consonant) it is placed on. It is found to be used after a subscript once in the Choun Nath dictionary (i.e. អ្នកសិល្បៈ). It is also found after certain vowels with a few instances of when it preceded by a subscript. It gets ambiguous as to which character is silenced when Toandakhiat is placed after a consonant cluster and/or vowel.

Examples of words with Toandakhiat placed:

- after a subscript and a vowel:

Khmer	Character	IPA	Gloss
កិត្តិវេរ	កិត្តិវេរេ	/kee/	‘reputation’
កិត្តិវេរេ	កិត្តិវេរេេ	/keə/	‘reputation’
បុំផ្លូវ	បុំផ្លូវេ	/poh/	‘post office’
ប្រសិទ្ធភី	ប្រសិទ្ធភីេ	/pra.sət/	‘a place name’
រោមកិត្តិវេរ	រោមកិត្តិវេរេ	/riem.kee/	‘Ramayanak’
សុំសិល្បេ	សុំសិល្បេេ	/suah/	‘glory’

- after a consonant and vowel ី

Khmer	Character	IPA	Gloss
កាហ្វោប៉ី	កាហ្វោប៉ីេ	/ka?laop/	‘a kind of basket’
តែនតី	តែនតីេ	/taen/	‘tent’
ទីបឹកី	ទីបឹកីេ	/tii.bee/	‘Tibet mount’
នាំអាតី	នាំអាតីេ	/noam.?aat/	‘to initiate’

ពោធិ៍	ព ហោ ធ ិ ៍	/poo/	'banyan'
ពោធិ៍ចំ	ព ហោ ធ ិ ៍ ច ំ	/poo.tum/	'a kind of plant'
ពោធិ៍ផ្ទៃ	ព ហោ ធ ិ ៍ ផ ុ ែ	/poo.tlee/	'a kind of plant'
ពោធិ៍បាយ	ព ហោ ធ ិ ៍ ប ា យ	/poo.baaj/	'a kind of banyan tree'
ពោធិ៍សាត់	ព ហោ ធ ិ ៍ ស ា ត ៗ	/poo.sat/	'Pursat province'
ឈ្មោះ	ឈ ុ ២ ឈ ិ ៍	/pjie/	'leprosy'
ព្រហស្បទ័រ	ព ្រ ហ ស ុ ន ក ិ ៍	/prə.hoah/	'Thursday'
វិទអាទី	វ ិ ទ អ ា ទ ិ ៍	/wiet.?aat/	'to try to gain power'
សកវាទី	ស ក វ ា ទ ិ ៍	/sak.kak.waa/	"
ស្វាទី	ស ្វ ុ ៣ ិ ៍	/smaat/	'meditate'
ស្វៀកពោធិ៍	ស ្វ ុ ៩ ក ព ហោ ធ ិ ៍	/slək.poo/	'banyan leaf'

- after ឈ

Khmer	Character	IPA	Gloss
រី	រ ឲ ឱ	/rie/	'name of a giant'
ចានី	ច ោ ន ឲ ឱ	/caa/	'small tubes made of gold'

It should be rendered right on the consonant. It has never been seen being used with any other vowel besides the ones mentioned above.

5.4. Obsolete Characters

Some signs are no longer used in the present-day Khmer texts. They have been known to be used in Pali/Sanskrit loanwords (see sections on [Pali](#), [Sanskrit](#) and [Inscriptions](#)).

- Bathamasat (េ)
- Atthacan (ែ)
- Viriam (ៅ)
- Avakrahasanya (៧)
- Obsolete consonants (ិ and ិ៍)
- Obsolete independent vowel (ឯ)

Here is a list of them with their code point.

Code Point Orthographic Character Name

U+17D3	េ	KHMER SIGN BATHAMASAT
U+17DD	ែ	KHMER SIGN ATTHACAN
U+17D1	ៅ	KHMER SIGN VIRIAM
U+17DC	៧	KHMER SIGN AVAKRAHASANYA
U+179D	ិ	KHMER LETTER SHA
U+179E	ិ៍	KHMER LETTER SSO
U+17A8	ឯ	KHMER INDEPENDENT VOWEL QUK

5.5. Punctuation

Two types of punctuation are used in Khmer: (1) native punctuation and (2) foreign punctuation. According to Khin (2007:87-90), 6 native punctuations were created by Khmer people and they are used in ancient and modern texts; 10 foreign punctuations were borrowed from French to be used in writing and printing. Prom (2006:55-69) lists more than that.

5.5.1. Native Punctuation

Name in Khmer	Name in English	Sign	Usage
ផ្លូវកម្មាធ័ណ្ឌ	Phnaek Muan ០	-	mark the beginning of a poem

តែម្បក្រិត	Koomuut	៧៩០	- mark the end of a poem
ឈុំ ឬ ខណ្ឌ	Lbah or Khan	១	- placed at the end of a sentence or a section of a poem - used in lunar dates
ឈុំចប់ ឬ ឈុំ	Lbah Chob or Lbah	១ ៩០	- Khin (2007:88) wrote it was used in the old days to mark the end of an article or a book, which is different from Prom (20006:55) who presented ៩ with the same name.
បរិយាសាន	Pareyaosan		
ជក្យា	White Space		- mark a pause in the speech, not to separate words (Open Forum of Cambodia 2004:15). - used in the same way as a comma and a semicolon in French.
បេយ្យាល់ ឬ ផ្លក់	Beyyal or Lak or	១ ៨១	
ចំណុចពីរ	Camnuc Pii		- used at the end of a list to denote that the list goes on
គូន	Kuuuh or	៩	
ឬ ដំពីរ	Dompii		- placed before a list or a direct quote in the same way that French and English uses colon (:). Colon is used in place of Commuc Pii Kuuh for Commuc Pii Kuuh is hard to find and type in the computer (Prom 2006:58).

5.5.2. Foreign Punctuation

Name in Khmer Name in English Sign Usage

វគ្គសញ្ញា	hyphen	-	- begin a list - placed after a number list - begin a dialogue exchange - denote the continuation of a word at the end of a line - denote "to" or "till" - connect related two items - syllable break - denote omitted syllable
វគ្គចចក	brackets	()	- denote extra information
ចុចមួយ	point	.	- place after abbreviation
ចុចពីរ	colon	:	- used in place of Chomnuc Pii Kuuh - denote "etc"
ចំណុចរាយ / ពង្រីក ellipsis	...		- denote unfinished phrase - denote a section of quote that is out of focus (usually place in side square brackets)
ក្មែស	comma	,	- separate items with spaces in between In Chuon Nath dictionary, it is used
ចុចក្មែស	semicolon	;	- to denote the list has not ended yet - placed before ប៉ូន្តិន្ត 'but' - separate two or more related sentences
សញ្ញាស្ថរ	question mark	?	- end interrogative sentence
សញ្ញាមទល	exclamation mark	!	- end an interjection - mark a quote or direct quote
អណ្តូន	double quotes	""	- encircle a title or name of a book or news article - focus point
ឃ្លាប or តិច្ចុប	square brackets	[]	- used in place of the normal brackets when there is brackets-in-brackets to avoid confusion - used for phonetic transcription
សញ្ញា "និង"	ampersand	&	- used to emphasize the items on its left and right
បន្ទាត់គ្រឹះ	forward slash	/	- separate numbers - used in place of ឬ 'or'
ផ្ទាយ	asterisk	*	- denote errors in a phrase
វត្ថុឲ្យប	curly brackets	{}	- used in syntax to wrap around certain things
សញ្ញាសី	equal; that is	=	- denote 'that means' or 'equal to'
សញ្ញាដំជាង	greater than; to	>	- denote greater than

	- denote what is on its left becomes what is on its right
សញ្ញាត្រូចជាង សញ្ញាត្រូចបិយ	smaller than; from opposite
<	- denote less than - denote what is on its left is derived from what is on its right
=	the word on its left and its right are opposite

5.6. Numerals

The table below shows Khmer numerals in the first row, Unicode code points in the second, and their gloss in the third row.

០	៩	២	៣	៤	៥	៦	៧	៨	៩
U+17E0	U+17E1	U+17E2	U+17E3	U+17E4	U+17E5	U+17E6	U+17E7	U+17E8	U+17E9
‘zero’	‘one’	‘two’	‘three’	‘four’	‘five’	‘six’	‘seven’	‘eight’	‘nine’

Note that Thai and Lao numerals are derived from Khmer's which is why they are very similar.

Numerals can never be used as a base of an orthographic syllable. It does not have a consonant quality, thus it is impossible to pronounce it when if it ever gets to combine with a vowel.

5.7. Divination Lore (a.k.a. លេកអត្ត “Lek Attak”)

Divination lore (a.k.a. លេកអត្ត “Lek Attak”) is used in Old Khmer to calculate and foretell the future of an event, but not in the present-day Khmer. The table below shows the divination lore in the first row and their Unicode code points.

០	១	២	៣	៤	៥	៦	៧	៨	៩
U+17F0	U+17F1	U+17F2	U+17F3	U+17F4	U+17F5	U+17F6	U+17F7	U+17F8	U+17F9
‘zero’	‘one’	‘two’	‘three’	‘four’	‘five’	‘six’	‘seven’	‘eight’	‘nine’

5.8. Lunar Dates

Khmer traditional date system follows the moon's phase. It is divided into two main categories: នឹងពិ “waxing moon” and នឹងនាម “waning moon”. The proper way of writing the date is using the Khmer numerals with a Khmer punctuation character “៯”. Khin (2007:87-88) illustrates that the number to the left of the Khan (៯) denotes days of the week (i.e. ៩ for Sunday, ២ for Monday and so forth), to the right of it is the number value of Khmer months as given in [Khmer Months of the Year](#), and the number that goes above or underneath the sign corresponding with the [Phase of the Moon](#) (i.e. if number is above the sign, it denotes “waxing moon”; underneath, “waning moon”). For instance, ៩២៥ is read as “Sunday, 10th day of the waning moon,

5.8.1. Phase of the Moon

5.8.1.1. នឹងពិ “Waxing Moon”

Waxing moon, a.k.a. នឹងពិ /knaot/ in Khmer, is the first 15 days of the lunar month.

- ៩០ U+19E0 /pa?.t̥aa.mie.saat/ ‘the first Ashadha (eighth month of the lunar calendar)’
- ៩១ U+19E1 /muəj.koət/ ‘the first day’
- ៩២ U+19E2 /pii.koət/ ‘the second day’
- ៩៣ U+19E3 /bəj.koət/ ‘the third day’
- ៩៤ U+19E4 /bəən.koət/ ‘the fourth day’
- ៩៥ U+19E5 /pram.koət/ ‘the fifth day’
- ៩៦ U+19E6 /pram.muəj.koət/ ‘the sixth day’
- ៩៧ U+19E7 /pram.pii.koət/ ‘the seventh day’
- ៩៨ U+19E8 /pram.bəj.koət/ ‘the eighth day’
- ៩៩ U+19E9 /pram.bəən.koət/ ‘the ninth day’
- ៩០ U+19EA /dəp.koət/ ‘the tenth day’
- ៩១ U+19EB /dəp.muəj.koət/ ‘the eleventh day’
- ៩២ U+19EC /dəp.pii.koət/ ‘the twelfth day’
- ៩៣ U+19ED /dəp.bəj.koət/ ‘the thirteenth day’
- ៩៤ U+19EE /dəp.bəən.koət/ ‘the fourteenth day’
- ៩៥ U+19EF /dəp.pram.koət/ ‘the fifteenth day’

5.8.1.2. នៅច “Waning Moon”

Waning moon, a.k.a. នៅច /ro.nooc/ in Khmer, is a period of 15 days in the lunar calendar counting from the day after the full moon back to the new moon (i.e. complete darkness).

- ៩០ U+19F0 /tu?.ti?.jeak.saat/ ‘the second Ashadha during the Adhikameas leap year’
- ៩១ U+19F1 /muəj.rooc/ ‘the first day’
- ៩២ U+19F2 /pii.rooc/ ‘the second day’
- ៩៣ U+19F3 /bəj.rooc/ ‘the third day’
- ៩៤ U+19F4 /bəən.rooc/ ‘the fourth day’
- ៩៥ U+19F5 /pram.rooc/ ‘the fifth day’
- ៩៦ U+19F6 /pram.muəj.rooc/ ‘the sixth day’
- ៩៧ U+19F7 /pram.pii.rooc/ ‘the seventh day’
- ៩៨ U+19F8 /pram.bəj.rooc/ ‘the eighth day’
- ៩៩ U+19F9 /pram.bəən.rooc/ ‘the ninth day’
- ៩០ U+19FA /dəp.rooc/ ‘the tenth day’
- ៩១ U+19FB /dəp.muəj.rooc/ ‘the eleventh day’
- ៩២ U+19FC /dəp.pii.rooc/ ‘the twelfth day’
- ៩៣ U+19FD /dəp.bəj.rooc/ ‘the thirteenth day’
- ៩៤ U+19FE /dəp.bəən.rooc/ ‘the fourteenth day’
- ៩៥ U+19FF /dəp.pram.rooc/ ‘the fifteenth day’

5.8.2. Khmer Months of the Year

Khmer months of the year is used in fortune telling and they have their own numeric value which is also used in Lunar date described in a section on [Lunar dates](#).

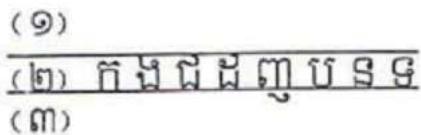
Khmer Name	Sankrit Name ^[12]	English Transliteration	Numeric Value
មិត្តសីរ	Mārgaśīrṣa (मार्गशीर्ष)	Mikeaser	៩
បុណ្យ	Pauṣa (पौष)	Boss	៩១
មាម	Māgha (माघ)	Meakh	៩២
ផែលុន	Phālguna (फाल्गुन)	Phalkun	៩៣
ចេត្រ	Caitra (चैत्र)	Chetr	៩៤
ពិសាខ	Vaiśākha (वैशाख)	Pisakh	៩៥
ជោឆ្នាំ	Jyeṣṭha (ज्येष्ठ)	Chesth	៩៦
អាសាច	Ashadha (आषाढ)	Asath	៩៧
ស្រవានា	Śrāvana (श्रावण)	Srapn	៩៨

ក្រុបទេស្ត	Bhadrapada (भाद्रपद) Phutrabot	90
អស्विन	Āśvina (अश्विन) Assoch	99
កាទិកា	Kārtika (कार्तिक) Kakdek	9៥

6. Ligatures

6.1. Consonant and Vowel Combination

Khin (2007:41) illustrates the anatomy of Khmer consonants and how Khmer characters should be written. There are three levels to consider: (1) the top level is reserved for a vowel, Consonant Shifters and/or a diacritic, (2) the middle level is for consonants and (3) the bottom level is for subscripts and/or a vowel (see the image below). In level (1), if a vowel and a diacritic occur together, the diacritic stay above the vowel (i.e. ឥ), but if the diacritic is a Triisap, the vowel should be placed above the Triisap (i.e. ឥិត). In level (3), if a subscript is used with a Below Vowel^[13], the Below vowel should be placed right below the subscript (i.e. សុក).



However, given that the space is limited when below subscript and below vowel stacked together. Two solutions have been done: (a) put the below subscript and below vowel sideways and (b) shrink the base consonant to accommodate the subscript allowing the below vowel to stay above its default position.

(a)

(b)

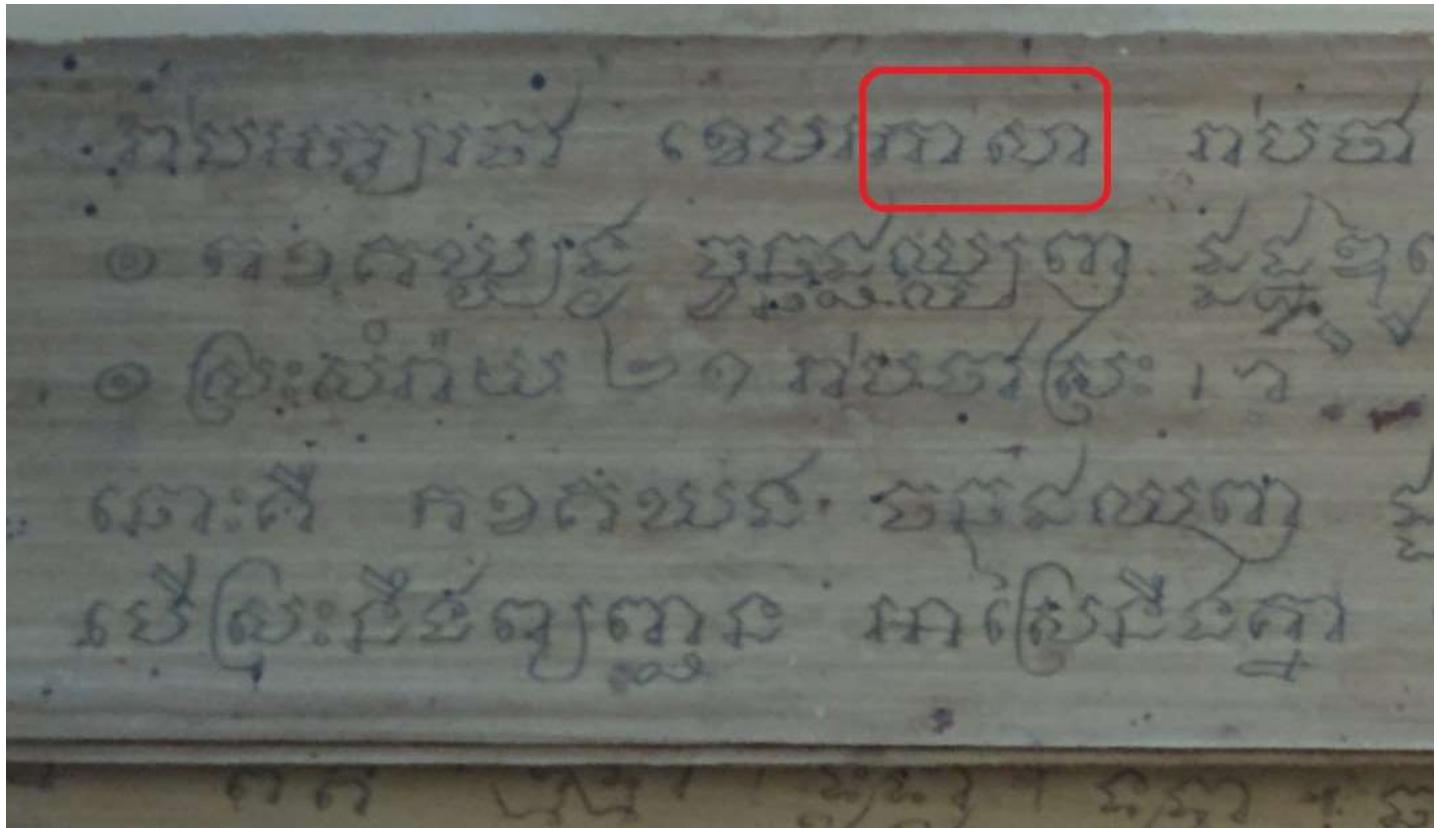
A consonant consists of two parts: the body and the head or usually called hair. For instance, ន is composed of an upside down U shape and a wavy line above it. The upside down U shape is the body and the wavy line is the hair. When ន stands on its own, the wavy line presides on the body fully; however, when a vowel^[14] that looks like ឃ is attached to it. The wavy line goes beyond the body and gets merged with the vowel as illustrated in (b). (a) is not recommended as it goes against the consonant and vowel composition convention.

(a)

(b)



A proof of how Khmer characters should be presented when combining a consonant with a wavy line and a ឃ looking vowel exists in Khmer manuscript written on palm leaves (see image below). The word highlighted is នាមា (ក ឃ ម ា) and you may notice the difference in how the wavy line looks when the consonant stands alone and when it is attached to ឃ.



NOTE: When ឃ is combined with ឃ ឃី or ឃីំ , the shape of the consonant change to ឃោ ឃិោ and ឃិោំ respectively (see the image below).



6.2. ឃ with a Subscript

ឃ has a little curvy line at the bottom level. When a subscript is placed underneath it, that curvy line should disappear and be replaced by the expected subscript.

This is the character when it stands alone.



This is the character with ឃ attached to it.



If the character (ឃ) needs to be combined with its own subscript (ឃ), the combination should look like this:



6.3. ឃ and ឃ Combination

When vowel ឃ [U+17B7] is followed by diacritic ឃ [U+17CD], they get connected (ឃ).

6.4. ក, ក, ក, ក, ក and ឃ, ឃ, ឃ

In some font, Khmer OS Muol for example, the combination of consonants ក (U+1780), ក (U+1782), ក (U+178F), ក (U+1797) and ក (U+179C) and above vowels (i.e. ឃ (), ឃ (), ឃ (), ឃ ()) becomes a glyph whose top line (i.e. the hair, the wavy line) got straighten and be in a shape similar to the vowels.

ឃ	+	ឃ	>	ក
				

ឃ	+	ឃ	>	ក
				

ឃ	+	ឃ	>	ក
				

ឃ	+	ឃ	>	ក
				

ក្រ + ០ វក្ស
ក្រ + ០ វក្ស

ឃ + ០ > ឃ

២

ឃ + ០ > ឃ

៤

ឃ + ០ > ឃ

៤

ឃ + ០ > ឃ

7. Unicode Encoding

7.1. Overview

There are discrepancies in character sequences posited by Unicode Standard (2018:646), OpenType (2018) and the Open Forum (2004:7-14). Only the conflicting characters are in the table below.

	Unicode Standard	OpenType	Open Forum
RegShift after the base consonant between a consonant and a subscript			
Robat before or after a subscript	considered as an Above Sign which goes after an Above vowel		between a subscript and a vowel
Subscript at the end	n/a		in between the base consonant and a subscript
IndV can be the base	can't be the base		n/a
Nikahit sign	sign		can be the base
Reahmuk sign	sign		vowel/sign?
			vowel/sign?

Unlike Unicode Standard and Open Forum, the OpenType breaks vowels into four subcategories:

- (1) Pre-base Vowels (PreV): ឃ [U+17C1], ឃ០ [U+17C2], ឃ១ [U+17C3],
- (2) Below-base Vowel (BlwV): ឃ៣ [U+17BB], ឃ៤ [U+17BC], ឃ៥ [U+17BD], ឃ៥៦ [U+17BB U+17C6],
- (3) Above-base Vowel (AbvV): ឃ៥៧ [U+17B7], ឃ៥៨ [U+17B8], ឃ៥៩ [U+17B9], ឃ៥៩៦ [U+17BA], ឃ៥៩៧ [U+17BE], and
- (4) Post-base Vowels (PstV): ឃ៧១ [U+17C4], ឃ៧២ [U+17C5], ឃ៧៣០ [U+17C0], ឃ៧៣៥ [U+17BF], ឃ៧៦ [U+17B6].

In Khmer language grammar textbook, Nikahit (៥ U+17C6) and Reahmuk (៣៦ U+17C7) are included in the vowel inventory, even though they are described elsewhere that they are diacritics or signs. The public usually consider these two as vowels. Plus when these two characters are used with other vowels, i.e. ឃ៣ (U+17BB) as in ឃ៣៦, ឃ១ (U+17C1) as in ឃ១៦ and ឃ៧១ (U+17C4) as in ឃ៧៣៦, Khmer linguists usually consider them as one vowel unit. Unicode Standard, however, considers these as a combination of vowel and diacritics (Nikahit and Reahmuk).

Below are the character sequences posited by Unicode Standard, OpenType and Open Forum:

Unicode Standard (2018:639)

B {R | C} {S {R}}* {(Z) V} {O} {S}

where

- B is a base character (consonant character, independent vowel character, and so on)
- R is a Robat C is a consonant shifter
- S is a subscript consonant or independent vowel sign
- V is a dependent vowel sign
- Z is a zero width non-joiner or a zero width joiner
- O is any other sign

OpenType (2018)

Cons + {COENG + (Cons | IndV)} + [PreV | BlwV] + [RegShift] + [AbvV] + {AbvS} + [PstV] + [PstS]

where

- Cons – Consonant character
- IndV – Independent vowel character
- COENG – The COENG code
- PreV – Vowel that is positioned before the base glyph
- BlwV – Vowel that is positioned below the base glyph
- RegShift – Triisap or Muusikatoan character
- AbvV - Vowel that is positioned above the base glyph
- AbvS – A sign character that is positioned above the base glyph
- PstV – Vowel that is positioned after the base glyph
- PstS – Sign character that is positioned after the base glyph
- – Indicates 0 to 2 occurrences
- [] – Indicates 0 or 1 occurrence
- | – Exclusive OR
- + – Cumulative AND

Note: ZWNJ and ZWJ are to be placed in between the RegShift and the AbvV to prevent the RegShift from changing its shape.

Open Forum

Open Forum of Cambodia wrote a document on “How to type Khmer Unicode” (2004). Khmer character ordering was devised as:

Consonant + Coeng consonant(s) + Vowel + Sign(s)

Consonant + Coeng consonant(s) + Consonant-shifter + Vowel + Sign(s)

Consonant + Robat {+ Vowel} {+ Sign}

Consonant + Coeng consonant(s) + Consonant-shifter + Vowel + Above signs + After signs

where

- Consonant – [U+1780..U+17A2] or [U+17A5..U+17B3]
- Coeng consonant – [U+17D2] + [U+1780..U+17A2] or [U+17A5..U+17B3]
- Vowel – [U+17B6..U+17C7]
- Sign – [U+17CB..U+17D1, U+17D3]
 - Above sign – [U+17CB, U+17CD..U+17D1, U+17D3, U+17DD]
 - After sign – [U+17C7, U+17C8]
- Consonant shifter – U+17C9, U+17CA
- Robat – U+17CC

7.2. New Proposal

Character orderings posited below are an attempt made to eliminate the discrepancy in the three schemes mentioned above. Though the proposal here has been discussed and contributed from various sources and individual, there may still be room of improvement. Kanjahn (2012:2) posited a character ordering that allows Register Shifter in two places, one before the subscript (right after the base) and another after the subscript. He allows Post Signs (which he called Space Signs) (◊: [U+17C8] and ◊◊: [U+17C7]) after a subscript.

Here is what we propose.

- BaseCommon: Consonant [U+1780..U+17A2] or Independent Vowel [U+17A5..U+17B3]
- BaseOther: Dotted Circle [U+25CC] or non-break space [U+00A0]
- Robat: the Robat [U+17CC]
- Ahsda: the Ahsda [U+17CF]
- AbvV: above-base vowel [U+17B7..U+17BA, U+17BE]
- PreV: pre-base vowel [U+17C1..U+17C3]
- BlwV: below-base vowel [U+17BB..U+17BD]
- PstV: post-base vowel [U+17B6, U+17BF, U+17C0, U+17C4, U+17C5]
- Coeng: the Coeng [U+17D2]
- RegShift: Register Shifter, Muusikatoan [U+17CA] or Triisap [U+17C9]
- AbvS: [U+17C6, U+17CB, U+17CD, U+17CE, U+17D0, U+17D1, U+17D3, U+17DD]
- Z: zero width non-joiner [U+200C] or zero width joiner [U+200D]

Partial clusters defined:

Base: BaseCommon | BaseOther

VowelGroup: (Z + AbvV) | PreV | BlwV | PstV

CoengGroup: Coeng + Base

Robat/Ahsda cluster defined:

Base [Robat | Ahsda] [PreV | BlwV | PstV] [PstS]

General cluster defined:

Base {CoengGroup} [RegShift] [VowelGroup] [AbvS] [PstS | CoengGroup]

Independent Cluster: All Khmer characters not classified above are considered clusters by themselves.

Restrictions: Here are some restrictions in the present day Khmer which have been observed. It:

- doesn't allow digits as bases,
- doesn't allow Coeng, consonant shifters, or above-base vowels or signs in robat syllables (but Jarai does, i.e. ៩៩ [U+17C6 U+17D0], a combination of two above-base signs),
- allows Coeng Ro only as the last Coeng when there are two subscripts in a row and one of which is a Coeng Ro,
- allows at most two subscripts in a syllable, and
- allows at most one vowel, and at most one above-base and one post-base sign.

7.3. Character Distribution in an Orthographic Syllable

Reahmuk (៩៩), Nikahit (៩), Yuukaleapintu (៩៩), Toandakhiat (៩៩), Ahsda (៩៩) and Bantoc (៩) always occurs at the syllable final position. They can be considered as orthographic syllable boundary.

Not any character can be weaved together with another one. An independent vowel can not be placed after another independent vowel in the same syllable. Coeng (U+17D2), RegShift, Robat, Vowel, Samyok Sannya or Kakabat can never be followed by itself.

The table below illustrates which character can be placed after another in the same orthographic syllable. "v" denotes that characters in the top row (1) can follow those in the left most column (2), "x" 1 can't follow 2, and "v" 1 can follow 2 in theory in the same syllable.

	1	Cons	IndV	Coeng	RegShift	Robat	V	Samyok Sannya	Kakabat
2		x	x	v	x	x	x	x	x
Cons		x	x	v'	x	x	x	x	x
IndV		x	x	v'	x	x	x	x	x
Coeng		v	v'	x	x	x	v [15]	x	x
RegShift		v	x	x	x	x	x	x	x
Robat		v	x	x	x	x	x	x	x
V		v	x	x	v	v	x	x	x

Samyok Sannya	v	x	x	v	x	x	x	x	x
Kakabat	v	x	x	v'	x	v	v'	x	x
Reahmuk	v	x	x	v	x	v	x		v
Nikahit	v	x	x	v	x	v	x		v
Yuukaleapintu	v	x	x	v'	v	v	x		v
Toandakhiat	v	x	x	x	x	v	x		x
Ahsda	v	x	x	x	x	v	x		x
Bantoc	v	x	x	x	x	x	x		x

7.4. Rendering Issues

According to Horton et al. (2017), there are at least eight cases of rendering issues in Khmer Unicode implementation. These eight issues lead to other problem for the end users and developers.

- **Confusability:** users are not able to make a sound judgement on which is the right way to type a word. Oftentimes, different sequences of the same characters are shown up exactly the same on the screen.
- **Vulnerability:** users can be spoofed and taken advantage of.
- **Searchability:** users are not able to find what they are looking for because they use different character sequences in the query.
- **Compatibility:** when a user uses android to send/share text documents with a windows or mac user, the text may look different because of the different implementation in the rendering engines.

The following sections list the issues one by one. The examples are adopted from the paper (Horton et al. 2017). The rendered text were done in Google Chrome 58.0 and Android 6.0.1.

7.4.1. Subscript and Vowel Concatenation

This is a case of when a “subscript” and a “vowel” are combined. Typing either one before another does not make any difference in the visual output on the screen, but it is invalid to place a Vowel before a Subscript.

- Subscript + Vowel ស ិ ី > សិី ‘Khmer’
- Vowel + Subscript ិ ស ិ ី > ិសិី **invalid sequence**

7.4.2. Concatenation of Two Subscripts

This is a case of when having two subscripts after a base consonant, and one of the subscript is [U+17D2 U+179A] ន. Placing either subscript before another would have the same visual output.

- Subscript + [U+17D2 U+179A] ន ន ុ ុ > នុុ ‘woman’
- [U+17D2 U+179A] + Subscript ន ុ ន ុ > ុនុ **invalid sequence**

7.4.3. Concatenation of a Subscript and a Consonant Shifter

This is a case of combining a subscript with a consonant shifter. The Khmer spelling convention (Kol 2008:28-32) does not mention which one should come first, but Open Forum of Cambodia (2004:11) claims that the consonant shifter should always come after the subscript.

- Subscript + Consonant Shifter ម ូ ំ ំ > មំូំ ‘one way’
- Consonant shifter + Subscript ម ំ ូ ំ > មំូំ **invalid sequence**

7.4.4. Concatenation of A Consonant Shifter and a Vowel

This is a special case of Khmer text rendering. The consonant shifter, either “[U+17CA] or” “[U+17C9], has to be rendered as a glyph that looks like ុ, when there is an above vowel attached to it (See [Special Treatment of Consonant Shifters](#)).

- Consonant Shifter (ំ) + Above vowel ន ុ > នុំ ‘to eat (vulgar)’
- Above vowel + Below vowel ន ុ ន ុ > នុុ **invalid sequence**
- Below vowel + Above vowel ន ុ ន ុ > នុុ **invalid sequence**
- Consonant Shifter (ំ) + Above vowel ន ុ > នុំ **incorrect sequence**

The last example is considered as incorrect rather than invalid because the character sequence is valid, but according to the usage of Muusikatoan (ធមេសិកតុលាន) it cannot be used with the 1st series consonants.

7.4.5. Vowels of Two Unicode Code Points

This is one of the most common confusing character sequences. Which should come first is not a concern among users though because, like other issue, the output on the screen does not look any different.

- [U+17BB] [U+17C6] ក ឃ ី > កី ‘don’t’
- [U+17C6] [U+17BB] ឃ ី ក > ីក **invalid sequence**
- [U+17B6] [U+17C6] ច ឃ ី > ីច ‘to wait’
- [U+17C6] [U+17B6] ឃ ី ច > ីច **invalid sequence**

7.4.6. One Unicode Code Point

ឃុំ [U+17C4] can be confused with a combination of ឃ [U+17C1] and ុំ [U+17B6] because the outputs of the three encoding on the screen look the same.

- [U+17C4] ឃ ុំ ក > ឃុំក ‘Mr.’
- [U+17C1] [U+17B6] ឃ ុំ ក > ឃុំក **invalid sequence**
- [U+17B6] [U+17C1] ឃ ុំ ក > ុំឃុំក **invalid sequence**

Similarly, ឃី [U+17BE] can be confused with a combination of ឃ [U+17C1] and ី [U+17B8].

- [U+17BE] ឃ ី > ឃី ‘if’
- [U+17C1] [U+17B8] ឃ ី ី > ឃី **invalid sequence**
- [U+17B8] [U+17C1] ី ី ឃ > ីីឃ **invalid sequence**

7.4.7. Similar Subscripts

This is the case of identical subscript. The users do not see any different in the outlook of these two.

- [U+17D2] [U+178A] ក ុំ ឃ ី ុ > កុំឃីុ ‘Kandal province’
- [U+17D2] [U+178F] ក ុំ ឃ ី ុ > កុំឃីុ **incorrect sequence**

7.4.8. Spaces and Joiners

- Zero Width Space is an invisible space (i.e. no width) which is usually put in between words in a sentence. It is helpful for text processing tools because it tells where the word boundaries are (Open Forum of Cambodia 2004:16-17).
- Zero Width non-Joiner can be inserted before consonant shifters in order to prevent them from being rendered as subscript (ឃ), and it can also be inserted directly before vowels in order to prevent the formation of ligatures between the base character and the vowel. (Kanjahn 2012:3).
- Zero Width Joiner can also be used before vowels to force a ligature between above vowels (see the section on [Special Treatment of Consonant Shifters](#)) and certain consonants.

8. Text Processing

8.1. Keyboard

Below is the standard key arrangement approved by the National ICT Development Authority (NiDA). There keyboard layout is divided into three layers: (1) the main layer, (2) the Shift layer and (3) the AltGr layer. A normal press on any key could output the character on the main layer. To output the character on Shift layer or AltGr layer, press and hold the Shift or AltGr key and then press any key of the expected character. Here is a list of character on each layer:

(1) the main layer

- first row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០
- second row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០
- third row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០
- fourth row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០

(2) the Shift layer

- first row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០
- second row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០
- third row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០
- fourth row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០

(3) the AltGr layer

- first row: zwj zwnj @ ០ \$ € ០ ០ * × ០ \
- second row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០
- third row: ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០ ៩ ៨ ៧ ៦ ៥ ៤ ៣ ២ ១ ០
- fourth row: , . /

Here is how the layout looks:



For smartphone, there is no standard keyboard layout in place when it comes to the number of rows on each layer and the number of characters on each row. However, most keyboards use 4x10 on each layer, meaning 4 rows and 10 characters on each row.

8.2. Sorting

The Royal Academy of Cambodia was approached and asked for advice regarding the sorting in Khmer. It turns out that there is no documentation related to this topic that could help. The Choun Nath dictionary seems to deploy two ways of sorting: (1) alphabetical order of characters and (2) alphabetical order of sounds. You may find words written with similar initial consonant sounds listed next to another. For instance, words begin with independent vowel ឃ are listed after the consonant ឃ which sounds similar to the independent vowel. A paper done by PAN localization^[16] entitled “Khmer Collation Development” suggests a solution that the sorting used in the Chuon Nath dictionary has to be adapted. Chuon Nath dictionary sorts entries based on they are pronounced, not the spelling. For instance, ឃង [U+1794 U+1784 U+17CB] is listed before បក្ខិត [U+1794 U+1780 U+178F U+17B7] even though the second character of the second word (i.e. ឃ [U+1780]) appears before that of the first word (i.e. ង [U+1784]) in the alphabet chart.

8.3. Fonts

Khmer typefaces have changed significantly since the 6th century. The following image shows how Khmer consonants evolved over time as quoted by Scheuren (2010:8) from Maspero (1915:48).

	VALUEUR EN SANSKRIT	VALUEUR EN KHMER MODERNE	
6th century	Type I.....	ક ક ગ ગ્ન ઙ ઙ	ક ક ગ ગ્ન ઙ ઙ
7th century	Type II.....	ર ર દ દ	ર ર દ દ
667	Type III.....	ફ ફ પ પ	ફ ફ પ પ
970	Type IV.....	સ સ શ શ	સ સ શ શ
1002	Type V.....	ષ ષ ષ ષ	ષ ષ ષ ષ
1066	Type VI.....	ષ ષ ષ ષ	ષ ષ ષ ષ
12th/13th century	Type VII.....	ક ક ક ક	ક ક ક ક
13th century	Type VIII....	ક ક ક ક	ક ક ક ક
1702	Type IX.....	ક ક ક ક	ક ક ક ક
modern (1915)	Type X.....	ક ક ક ક	ક ક ક ક

8.3.1. Font Style

According to Scheuren (ibid:9-10) there are three main types of Khmer font styles: (a) ឃុំ Mool ‘lit. round’, (b) ឃ្លឹង Chrieng ‘lit. slanted’ and (c) Upright which is the standing version of Chrieng. Mool style was commonly found in inscriptions, while Chrieng was used in palm-leaf manuscript (a.k.a. Sastra) which existed before the printing types introduced in 1877.

To date, the Mool style is usually used in banners and titles of books or articles; while the Upright, the standing Chrieng, is of everyday uses. The table below shows how the three styles look.

នាវ៉ាទេវ

Mool Style in Khmer OS Moul font face

Chrieng Style in Khmer OS Metalchrieng font face

កសាខ់ខ្មែរ

កសាខ់ខ្មែរ

Upright style in Khmer OS system font face

8.3.2. Font Rendering

Khmer font rendering is complex because characters are not rendered in a linear order. Vowels are not always found after the base consonant. They can go to the left, right, above, below or even around the base. Similarly, subscripts can be rendered to the left, right or below the base. These make it confusing to the users as to what should be typed when and placed where.

Lists of vowel according to their positions when collating with the base are shown in the table below.

left	above	below	right	around
				ឃំ (U+17BE)
				ឃី (U+17BF)
ឃោ (U+17B7)				ឃួ (U+17C0)
ឃៅ (U+17C1)	ឃះ (U+17B8)	ឃ៉ (U+17BB)	ឃៈ (U+17B6)	ឃុ (U+17C4)
ឃំ (U+17C2)	ឃះ (U+17B9)	ឃ៉ (U+17BC)	ឃ៊ (U+17C7)	ឃុំ (U+17C5)
ឃំ (U+17C3)	ឃះ (U+17BA)	ឃ៉ (U+17BD)	ឃ៊ (U+17C8)	ឃុំ (U+17BB U+17C6)
ឃោ (U+17C6)			ឃំ (U+17B6 U+17C6)	ឃ៊ (U+17BB U+17C7)
				ឃុះ (U+17C1 U+17C7)
				ឃុះ (U+17C4 U+17C7)

Here is a list of subscripts and where they should be rendered in their respective categories.

left	below	right
ឃោ	ឃ៉	ឃៈ
ឃៅ	ឃ៊	ឃុ
ឃំ	ឃុំ	ឃុះ
ឃះ	ឃុះ	
ឃ៉		
ឃៈ		
ឃុ		
ឃុំ		
ឃុះ		

8.4. Diacritic Position

All diacritics are placed on top of the base: "ក ស ា ខ ់ ខ ្ម ែ រ". Only one diacritic is usually found on a consonant in a syllable, except 'a Consonant Shifter + Samyok Sannya'. There is also a common occurrence of a consonant shifter with Nikahit "Consonant Shifter + Nikahit," but be aware that Nikahit functions as a vowel in that environment, despite its name.

9. Application of Khmer Script to Other Languages

Khmer alphabet is not only used to write Khmer language, but it is also used to write at least six ethnic minority languages (i.e. Bunong, Tampuan, Brao, Krung, Jarai and Kuay) and two dead languages (i.e. Pali and Sanskrit). Each language requires unique syllable configuration and character sequences. This section describes the two main points in each language:

- a list of consonants, subscripts, vowels, diacritics, symbols and punctuations if applicable
- how the writing system is different from Khmer

See [Appendix F](#) for a side by side comparison of the characters existing in Khmer and each ethnic language.

The following sections has a list of characters used in each ethnic minority language if the form of table where the orthographic characters are in the first row, phonemic representations corresponding with the orthographic characters in the second row and the unicode code points in the third row.

9.1. Bunong

According to the Bunong-Khmer Bilingual Dictionary (2011:៩-៨), 53 Khmer characters are used in writing Bunong.

9.1.1. Bunong Consonants

There are 28 orthographic consonants in Bunong. ឃិ /ç/ and ុុ /h/ always occurs at the final position.

ក	ខ	ច	ឆ	ប	ន	ជ	ញ	ជ	ត
/k/	/kʰ/	/g/	/ɣ/	/c/	/cʰ/	/ɟ/	/ɲ/	/d/	/t/
U+1780 U+1781 U+1782 U+1784 U+1785 U+1786 U+1787 U+1789 U+178A U+178F									
ថ	ស	ស	ប	ជ	ព	ម	យ	ឃិ	រ
/tʰ/	/d/	/n/	/b/	/pʰ/	/p/	/m/	/j/	/ç/	/r/
U+1799									
ុ	ី	ួ	ូ	ុុ	ុុ	ុុ	ុុ	ុុ	ុុ
/l/	/w/	/b/	/h/	/h/	/ʔ/	/ʔj/	/ʔw/		
U+17A2 U+17A2									
ុុ	ុុ	ុុ	ុុ	ុុ	ុុ	ុុ	ុុ	ុុ	ុុ
U+179B U+179C U+179E U+179F U+17A0 U+17A2 U+17D2 U+17D2									
U+179A U+179C									

9.1.2. Bunong Subscripts

៥ [U+17D2 U+17A2] is not listed in the table because it is used not in the Bunong Khmer Dictionary, however, it is used by some speakers.

៥	៥	៥	៥	៥
/j/	/r/	/l/	/w/	/h/
U+17D2	U+17D2	U+17D2	U+17D2	U+17D2
U+1799	U+179A	U+179B	U+179C	U+17A0

9.1.3. Bunong Vowels

ា	ី	ើ	ើ	ី	ុ	ុ	ុ	ុ	ុ
/a/	[i]	/i/	[ɛ]	/i/	/ʊ/	/u/	/ə/	/i/	/e/
U+17B6 U+17B7 U+17B8 U+17B9 U+17BA U+17BB U+17BC U+17BE U+17C0 U+17C1									
ើ	ើ	ើ	ើ	ើ	ើ	ើ	ើ	ើ	ើ
/ɛ/	/äj/	/o/	/aw/	/ɔ/					
U+17C2 U+17C3 U+17C4 U+17C5 U+17DD									

ើ /i/ [U+17C0] is used in certain words borrowed from Khmer.

እ ዘ /aj/ [17C3] and እው /aw/ [U+17C5] are phonological compounds of a vowel and a consonant.

Bunong does not have an inherent vowel.

9.1.4. Bunong Additional Vowels

ଠ: or ଠାଁ ଠ୍ୟ:	ଠୋ:	ଠୁଁ:	ଠାଁ:	ଠ୍ୟ: or ଠିଁ ଠୁଁ
/ଠ/	/ଠ୍ୟ/	[ଠ]	/ଠୁଁ/	[ଠ୍ୟ]
U+17B6				U+17DD
U+17C8				U+17C8
or	U+17BA	U+17C1	U+17C2	U+17C4
U+17B6	U+17C8	U+17C8	U+17C8	U+17C8
U+17CB				U+17CB

ឃុំ /p/ [U+17BF] is used in certain words borrowed from Khmer.

9.1.5. Bunong Symbols and Punctuation

In Bunong writing, four symbols are used in the same way as those of Khmer language(i.e. េ ែ ោ and ៉) and other which are borrowed.

9.1.6. How is the writing system different from Khmer?

- The use of two Khmer obsolete characters:
 - ប /b/ [U+179E] as in បរ (ប ា រ) /bar/ ‘two’
 - ក /ə/ [U+17DD] as in កិន (ក ិ ន) /kən/ ‘child by birth; son’
 - ែ [U+17CB] can be placed on consonant ឃ [U+179A] as in ពីរ (ព ិ រ ែ) /pər/ ‘to burn’.
 - ែ: [U+17C8] cannot occur with a consonant on its own, it has to be preceded by a vowel as shown in “Additional Vowels” list above. This is the opposite of the Khmer spelling convention where ែ: [U+17C8] is usually used to attach to the consonant without any vowel intervening it.
 - ុ [U+17DD] behaves like a vowel, and it can be followed by a final consonant with or without ែ as in តិង (ត ិ ង) /gəŋ/ ‘k.o. gong’ and តិង់ (ត ិ ង ់) /gəŋ/ ‘to roast’.
 - Sequences uniquely used in Bungong (i.e. ឃ្លាំង, ឃ្លាំង, ឃ្លាំង).
 - The writing system does not follow the Khmer two series system where one vowel symbol can represent two sounds, depending on which consonant series precedes it. In Bunong, each vowel symbol represents only one sound in every instance.

9.2. Tampuan

The number of consonants and vowels in the Tampuan Alphabet book (Pech 2006) and the Tampuan-Khmer Dictionary (Muang 2012) varies. The first lists 30 consonants and 27 vowels. The latter lists 29 consonants and 28 vowels. The following details are adapted from the Tampuan Alphabet book (*ibid*), the Tampuan-Khmer dictionary (*ibid*) and other documents obtained from SIL members.

9.2.1. Tampuan Consonants

ឃុ ឃុ
/cj/ /hɲ/

U+1785 U+17A0
U+17D2 U+17D2
U+1799 U+1789

“Tampuan Khmer English Dictionary with English Khmer Tampuan Glossary” (2007) uses /ñ/ instead of /ŋ/ for ឃុ [U+1789] and /d/ instead of /?d/ for ឌុ [U+178A] and ឌុ [U+178C].

9.2.2. Tampuan Subscripts

ឃុ								
/ʔ/	/g/	/ŋ/	/n/	/m/	/j/	/r/	/l/	/w/
U+17D2								
U+1782	U+1784	U+1789	U+1793	U+1798	U+1799	U+179A	U+179B	U+179C

ឃុ [U+17D2 U+1784] and ឃុ [U+17D2 U+1789] are used to write proper names only. They do not usually used in common Tampuan words.

ឃុ [U+17D2 U+1782] only occurs word finally under ឃ and ឃុ.

9.2.3. Tampuan Vowels

	ឃុ ^[17]	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ
/ɑ:/ /ɔ:/	/a:/	/ɛ/	/əi/	/ə/	/əi/	/o/	/ɔ:ɔ:ou	/uə/	/aə/
N/A	/i/	/ɪ/	/i:/	/ɪ/	/ɪ/	/ʊ/	/ʊ/	/eə/	/ə/
U+17B6	U+17B7	U+17B8	U+17B9	U+17BA	U+17BB	U+17BC	U+17BD	U+17BE	
ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ
/iə/	/iə/	/e:/	N/A	/ai/	/ao/	N/A	/om/	/əm/	/am a/
/iə/	/iə/	/ɛ:ɛ:/	/ɛ:/	N/A	/ɔ:/	/ɔ:/	/um/	N/A	N/A
U+17BF	U+17C0	U+17C1	U+17C2	U+17C3	U+17C4	U+17C5	U+17BB	U+17C6	U+17B6
U+17C6							U+17C6	U+17C6	U+17C6
ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ
/ah/	/oh/	/eh/	/ph/	N/A	/əh/	/aəh/	N/A	/ouh/	
N/A	/uh/	N/A	N/A	/ih/	/ih/	N/A	N/A	/uuh/	
U+17C7	U+17BB	U+17C1	U+17C4	U+17B7	U+17B9	U+17BE	U+17C2	U+17BC	
U+17C7		U+17C7	U+17C7	U+17C7	U+17C7	U+17C7	U+17C7	U+17C7	U+17C7

ឃុ: exists in the Tampuan-Khmer dictionary (ibid), but not in the Tampuan alphabet book (ibid). One of the reference documents states that ឃុ: is not used.

The empty cell in the first column of the first vowel table above is the inherent vowel.

9.2.4. Tampuan Diacritics

The Tampuan Alphabet book (ibid) illustrates examples of the usage of ំ (U+17C9) and ែ (U+17CA). They are used to change the series of the consonant.

- ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ
- ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ ឃុ

9.2.5. How is the writing system different from Khmer?

- The \circ [U+17D0] symbol indicates that the main vowel is pronounced with a breathy phonation. The symbol only ever occurs word finally. It is only used in situations where the vowel would otherwise be tense. In most cases a second series consonant will indicate breathy phonation on the following vowel too. It should be noted that the \circ is supposed to stand at the word boundary (i.e. usually on the final consonant), but since there is restriction in the Khmer Unicode character ordering when combining $\circ:$ [U+17C7] with \circ , it has to be encoded before the vowel. If placing \circ after $\circ:$, the text would not look right (i.e. the dotted circle appears in between the two characters).
- Only nine consonants have subscript forms.
- " [U+17C9] is used with clusters like $\text{ឃ} \text{ឃ}$ to indicate that ឃ sounds [p], not [b]. Therefore, ឃ is [br] and ឃ is [pr]; ឃ is [bl] and ឃ is [pl].
- \circ can be put on any final consonant. It does not have the same restriction as in Khmer.
- A white space is used in between each word to denote a word boundary.
- Subscript ៥ [U+17D2 U+179C] is placed after subscript ០ [U+17D2 U+179A].

9.3. Brao

The following is the character inventory of Brao language (a.k.a. Brao Ombaa). There are 36 consonants, 2 subscripts, 19 vowels and 5 diacritics.

9.3.1. Brao Consonants

ក	ខ	គ	ឃ	ឃ	ច	ឆ	ជ	ធម	ឣ
/k/	/kʰ/	/k/	/kʰ/	/g/	/c/	/ʒ/	/c/	/ʒ/	/p/
U+1780 U+1781 U+1782 U+1783 U+1784 U+1785 U+1786 U+1787 U+1788 U+1789									
ជ	ថ	ន	ញ	ណ	ត	ថ	ន	ជ	ស
/d/	/ʔd/	/d/	/n/	/n/	/t/	/tʰ/	/t/	/tʰ/	/n/
U+178A U+178B U+178C U+178D U+178E U+178F U+1790 U+1791 U+1792 U+1793									
ឃ	ធម	ឃ	ី	ឃ	ឃ	ី	ី	ី	ី
/b/	/pʰ/	/p/	/pʰ/	/m/	/j/	/r/	/l/	/w/	/ç/
U+1794 U+1795 U+1796 U+1797 U+1798 U+1799 U+179A U+179B U+179C U+179F									
ឃ	ធម	ឃ	ី	ី	ី	ី	ី	ី	ី
/h/	/l/	/ʔ/	/ʔʃ/	/g/	/ʔb/				
U+17A2									
U+17A0 U+17A1 U+17A2 U+17D2 U+179D U+179E									
U+1799									

9.3.2. Brao Subscripts

០	៥
/r/	/l/
U+17D2 U+179A U+17D2 U+179B	

9.3.3. Brao Vowels

Some vowels in Brao have two phonemic representations: one for the 1st series and another is for the 2nd series. For instance, \hat{o} is realized as /ɛ/ in the 1st series and /i/ for the 2nd series.

inherent	ា	េ	ែ	ោ	ៅ	៉	ៈ	េះ	ែះ
/aa/	/aa/	/ɛ/	/i/	/ə/	N/A	/o/	/oo/	/uə/	/iə/
N/A	N/A	/i/	/ii/	/i/	/ii/	/u/	/uu/	N/A	N/A
U+17B6 U+17B7 U+17B8 U+17B9 U+17BA U+17BB U+17BC U+17BD U+17BF									
ឃេ	ឃែ	ឃោ	ឃៅ	ឃំ	ឃៈ	ឃ៉	ឃេះ	ឃែះ	ឃោះ
/iə/	/oo/	N/A	/aj/	/əm/	/ah/	/əh/	/om/	/am/	
N/A	N/A	/ee/	N/A	N/A	N/A	/uəh/	/um/	N/A	

U+17C0 U+17BE U+17C2 U+17C3 U+17C6 U+17C7 U+17C4 U+17BB U+17B6
 U+17C7 U+17C6 U+17C6

កំស

/əŋ/

N/A

U+17B6

U+17C6

U+1784

9.3.4. Brao Diacritics

៥ ៦ ៧ ៨ ៩ ០

U+17CB U+17C9 U+17CA U+17DD U+17CE

9.3.5. How is the writing system different from Khmer?

- Two obsolete characters (i.e. ុ and ុ) are used as consonants.
- Only two subscripts are used in Brao.
- Khmer does not have ឌី as a cluster, nor ឃីុ /?ɔr?jiiip/ ‘very black’.
- ០ is placed on any final consonant whose vowel (either /aa/ and /oo/) before it is shortened.
- A white space is used in between each word to denote a word boundary.
- ០ lengthens the vowel in the /ooh/ sequence.

9.4. Krung

In Krung language, there are 33 consonants, 6 subscripts, 16 vowels and 3 diacritics. Krung series system mostly conforms to Khmer writing system.

9.4.1. Krung Consonants

The source does not provide phonemic representation, but the romanized version of each orthographic consonant. They are listed in the second rows.

ក	ខ	គ	ឃ	ញ	ច	ឆ	ជ	ឈ	ឃ	ញ	ឃ
k	kh	k	kh	ng	c	j	c	j	nh		
U+1780 U+1781 U+1782 U+1783 U+1784 U+1785 U+1786 U+1787 U+1788 U+1789											
ដ	ឃុ	ឃ	ឃុ	ឃុ	ឃ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ	ឃុ
d	qd	d	qd	n	t	t	n	b	ph		
U+178A U+178B U+178C U+178D U+178E U+178F U+1791 U+1793 U+1794 U+1795											
ព	ឃុ	ឃ	ឃុ								
p	ph	m	j	r	l	w	s	h	l		
U+1796 U+1797 U+1798 U+1799 U+179A U+179B U+179C U+179F U+17A0 U+17A1											
ឃ	ឃុ	ឃ	ឃុ								
q	g	qb									
U+17A2 U+179D U+179E											

9.4.2. Krung Subscripts

៤	៥	៦	៧	៨	៩	០
ng	n	m	r	l	q	
U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	
U+1784	U+1793	U+1798	U+179A	U+179B	U+17A2	

9.4.3. Krung Vowels

Inherent	ঠ	ঃ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ
/àà/	/aa/	/è/	/uy/	/e/	N/A	/o/	/oo/	/ue/	/ee/
N/A	N/A	/i/	/ii/	/ù/	/ùù/	/u/	/uu/	/ue/	N/A
	U+17B6	U+17B7	U+17B8	U+17B9	U+17BA	U+17BB	U+17BC	U+17BD	U+17BE
ঔ	ঔ	ঔ	ঔ	ঔ	ঔ	ঔ	ঔ	ঔ	ঔ
/ie/	N/A	/ay/	/ah/	/ah/	/èh/	/ang/			
/ie/	/èè/	N/A	N/A	N/A	N/A	N/A			
	U+17C0	U+17C2	U+17C3	U+17C7	U+17C4	U+17B7	U+17B6	U+17C6	
					U+17C7	U+17C7	U+1784		

9.4.4. Krung Diacritics

Three diacritics are used:

- ̄ shortens the vowel length.
- ̄ changes the consonant to the 1st series
- ̄ changes the consonant to the 2nd series

ঁ ̄ ̄

U+17CB U+17C9 U+17CA

9.4.5. How is the writing system different from Khmer?

- Two obsolete characters (i.e. ঃ and ঁ) are used as consonants.
- The consonant shifters are used with the obsolete characters (i.e. ঃ and ঁ).
- A white space is used in between each word to denote a word boundary.

9.5. Jarai

In Jarai language, there are 35 consonants, nine subscripts, 22 vowels and four diacritics.

9.5.1. Jarai Consonants

କ	କ	କ	ଯ	ଯ	କ	କ	କ	କ	କ
/k/	/kʰ/	/k/	/kʰ/	/g/	/c/	/cV/	/c/	/cV/	/j/
U+1780	U+1781	U+1782	U+1783	U+1784	U+1785	U+1786	U+1787	U+1788	U+1789
ଚ	ଚ	ଚ	ଚ	ଚ	ଚ	ଚ	ଚ	ଚ	ଚ
/tV/	/d/	/tV/	/d/	/n/	/t/	/tʰ/	/t/	/tʰ/	/n/
U+178A	U+178B	U+178C	U+178D	U+178E	U+178F	U+1790	U+1791	U+1792	U+1793
ପ	ପ	ପ	ପ	ପ	ପ	ପ	ପ	ପ	ପ
/bV/	/pʰ/	/p/	/pʰ/	/m/	/j/	/V/	/l/	/w/	/s/
U+1794	U+1795	U+1796	U+1797	U+1798	U+1799	U+179A	U+179B	U+179C	U+179F
ହ	ହ	ହ	ହ	ହ	ହ	ହ	ହ	ହ	ହ
/h/	/l/	/?/	/kV/	/b/					
U+17A0	U+17A1	U+17A2	U+179D	U+179E					

9.5.2. Jarai Subscripts

ଙ	ଙ	ଙ	ଙ	ଙ	ଙ	ଙ	ଙ	ଙ
/k/	/g/	/p/	/n/	/m/	/j/	/l/	/w/	
U+17D2								

U+1782 U+1784 U+1789 U+1793 U+1798 U+1799 U+179B U+179C

9.5.3. Vowels

inherent	၁	၂	၃	၄	၅	၆	၇	၈	၉	၁၀
/ə/	/a:/	/ɛ?/	N/A	/ə?/	N/A	/o?/	/o:/	N/A	N/A	
/ɔ/	N/A	/i?/	/i/	/i?/	/i/	/u?/	/u:/	/uə/	/ə:/	
U+17B6 U+17B7 U+17B8 U+17B9 U+17BA U+17BB U+17BC U+17BD U+17BE										
၁၂	၁၃	၁၄	၁၅	၁၆	၁၇	၁၈	၁၉	၁၁၀	၁၁၁	၁၁၂
/ia/	/ie/	N/A	N/A	/um/	/ah/	/a?/	/ah/	/eh/	N/A	
/ia/	N/A	/e:/	/ɛ:/	N/A	N/A	N/A	N/A	/ih/	/ɛih/	
U+17BF U+17C0 U+17C1 U+17C2 U+17C6 U+17C7 U+17C8 U+17C4 U+17B7 U+17C2 U+17C7 U+17C7 U+17C7										
၁၁၃	၁၁၄	၁၁၅	၁၁၆	၁၁၇	၁၁၈	၁၁၉	၁၁၁၀	၁၁၁၁	၁၁၁၂	၁၁၁၃
/om/	/am/	/ap/	/ah/	/oh/						
/um/	N/A	N/A	/ih/	/uh/						
၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂
၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂
၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂	၁၁၇၂

9.5.4. Jarai Diacritics

Four diacritics are used.

- ៥ [U+17CB] marks short vowel. *
- ៥” [U+17C9] changes second series consonant to the first series.
- ៥~ [U+17CA] changes the first series consonant to the second series.
- ៥៥ [U+17D0] marks nasalization, but since it can be confused in function with the Khmer Samyok Sannya, another form is proposed: ៥៥, placing before the syllable it modifies.

9.5.5. How is the writing system different from Khmer?

- Two obsolete characters (i.e. ៥៥ and ៥៥) are used as consonants.
- A white space is used in between each word to denote a word boundary.
- The placement of ៥ is unclear, but it seems to be inconsistent with how it is used in Khmer.
 - ៥៥ ៥៥” ៥៥ ៥៥ ៥៥
 - ៥៥ ៥៥ ៥៥ ៥៥ ៥៥ [jivum] ‘needle’
 - ៥៥ ៥៥ ? [buə?] ‘work’
 - ៥៥ ៥៥ ? [wā?] ‘oil’
 - ៥៥ ៥៥ ? [klaap] ‘difficult’
 - ៥៥ ៥៥ ? [gūah] ‘morning’

9.6. Kuay

In Kuay language, there are 34 consonants, 21 subscripts, 25 vowels, and 4 diacritics.

9.6.1. Kuay Consonants

၁	၂	၃	၄	၅	၆	၇	၈	၉	၁၀	၁၁
/k/	/kʰ/	/k/	/kʰ/	/g/	/c/	/cʰ/	/c/	/cʰ/	/p/	
U+1780 U+1781 U+1782 U+1783 U+1784 U+1785 U+1786 U+1787 U+1788 U+1789										
၁၂	၁၃	၁၄	၁၅	၁၆	၁၇	၁၈	၁၉	၁၁၀	၁၁၁	၁၁၂
/d/	/d/	/n/	/t/	/tʰ/	/t/	/tʰ/	/n/	/b/	/pʰ/	
U+178A U+178C U+178E U+178F U+1790 U+1791 U+1792 U+1793 U+1794 U+1795										

ព	ម	យ	រ	ល	វ	ស	ុ	ឡ	ន
/p/	/m/	/j/	/r/	/l/	/w/	/s/	/h/	/l/	/ʔ/

U+1796 U+1798 U+1799 U+179A U+179B U+179C U+179F U+17A0 U+17A1 U+17A2

ខ

/f/

U+17A2

U+17D2

U+1785

9.6.2. Kuay Subscripts

ក	ក	ក	ក	ក	ក	ក	ក	ក	ក
[k]	[k ^h]	[k]	[k ^h]	[g]	[c]	[c]	[p]	[d ^f]	[d]
U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2
U+1780	U+1781	U+1782	U+1783	U+1784	U+1785	U+1787	U+1789	U+178A	U+178C
ខ	ខ	ខ	ខ	ខ	ខ	ខ	ខ	ខ	ខ
[t]	[t ^h]	[t]	[n]	[θ]	[p]	[p ^h]	[m]	[j]	[r]
U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2	U+17D2
U+178F	U+1790	U+1791	U+1793	U+1794	U+1796	U+1797	U+1798	U+1799	U+179A
ឃ	ឃ	ឃ	ឃ	ឃ					
[β]	[s]	[h]	[?]						
U+17D2	U+17D2	U+17D2	U+17D2	U+17D2					
U+179B	U+179C	U+179F	U+17A0	U+17A2					

9.6.3. Kuay Vowels

inherent	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ
/e:/	/a:/	/ɛ/	[ej]	/ɜ/	/ə:/	/ɔə/	/o:/	/uə/	/ɜ:/
N/A	/ia/	/i/	/j/	/ə/	/ɛ/	/ɪ/	/ʊ/	N/A	/ɔ:/
U+17B6	U+17B7	U+17B8	U+17B9	U+17BA	U+17BB	U+17BC	U+17BD	U+17BE	
ឃី	ឃី	ឃី	ឃី	ឃី	ឃី	ឃី	ឃី	ឃី	ឃី
/iə/	/iə/	/e:/	/ɛ:/	[aj]	/ə:/	N/A	[əm]	[ah]	[əh]
/jə/	/jə/	/ɔ:/	N/A	[ej]	/ə:/	[əw]	[əm]	[ah]	[əh]
U+17BF	U+17C0	U+17C1	U+17C2	U+17C3	U+17C4	U+17C5	U+17C6	U+17C7	U+17C1
ឃោះ	ឃោះ	ឃោះ	ឃោះ	ឃោះ	ឃោះ	ឃោះ	ឃោះ	ឃោះ	ឃោះ
[əh]	[əh]	N/A	[əh]	[əm]	[əh]				
[ah]	[uh]	[ih]	N/A	[əm]	[əh]				
U+17C4	U+17BB	U+17B7	U+17B9	U+17B6					U+17B6
U+17C7	U+17C7	U+17C7	U+17C7	U+17C6					U+17C6
									U+1784

9.6.4. Kuay Diacritics

Four diacritics are used.

៥	"	~	៥
---	---	---	---

U+17CB U+17C9 U+17CA U+17D0

Punctuation are used in the same as in Khmer language (i.e. ៖ “ ” « » ? ! : ; ... ॥ ែែ).

9.6.5. How is the writing system different from Khmer?

- ጀ gets to be used on ጀ, a thing which never occurs in Khmer.
- ጀ can be used on a final consonant preceded by a Samyok Sannya or ጀ.

9.7. Sastras

9.7.1. Pali

Nhok (1962:1-2) wrote that there are 41 characters in the Pali alphabet inventory--8 vowels and 33 consonants. Note that vowels are divided into two groups: independent and dependent. It is observed independent vowels usually start a syllable, while the dependent ones always attached to a initial consonant.

No diacritic is used in Pali.

The tables below present Khmer characters used in Pali in the first row, romanization of them in the second and the unicode code points in the third row.

9.7.1.1. Pali Consonants

ក	ខ	គ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ
k	kh	g	gh	ñ	c	ch	j	jh	ñ		
U+1780 U+1781 U+1782 U+1783 U+1784 U+1785 U+1786 U+1787 U+1788 U+1789											
ឋ	ឋ	ឋ	ឋ	ឋ	ឋ	ឋ	ឋ	ឋ	ឋ	ឋ	ឋ
ঢ	ঢ	ঢ	ঢ	ঢ	ঢ	ঢ	ঢ	ঢ	ঢ	ঢ	ঢ
U+178A U+178B U+178C U+178D U+178E U+178F U+1790 U+1791 U+1792 U+1793											
ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ
p	ph	b	bh	m	y	r	l	v	s		
U+1794 U+1795 U+1796 U+1797 U+1798 U+1799 U+179A U+179B U+179C U+179F											
ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ	ឃ
h	l	m									
U+17A0 U+17A1 U+17C6											

9.7.1.2. Pali Subscripts

ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ
k	kh	g	gh	ñ	c	ch	j	jh	ñ		
U+17D2											
U+1780	U+1781	U+1782	U+1783	U+1784	U+1785	U+1786	U+1787	U+1788	U+1789		
ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ
t	th	d	dh	ñ	t	th	d	dh	n		
U+17D2											
U+178A	U+178B	U+178C	U+178D	U+178E	U+178F	U+1790	U+1791	U+1792	U+1793		
ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ
p	ph	b	bh	m	y	r	l	v	s		
U+17D2											
U+1794	U+1795	U+1796	U+1797	U+1798	U+1799	U+179A	U+179B	U+179C	U+179F		
ጀ											
h											
U+17D2											
U+17A0											

9.7.1.3. Pali Independent Vowels

ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ
ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ

a	ā	i	ī	u	ū	e	o
U+17A3	U+17A4	U+17A5	U+17A6	U+17A7	U+17A9	U+17AF	U+17B1

9.7.1.4. Pali Dependent Vowels

inherent	ၠ	၃	၄	၅	၆	၈	၉
a	ā	i	ī	u	ū	e	o
N/A	U+17B6	U+17B7	U+17B8	U+17BB	U+17BC	U+17C1	U+17C4

9.7.1.5. Pali Vowel Combination

၁	၂	၃
m	im	um
U+17C6	U+17B7	U+17BB
	U+17C6	U+17C6

9.7.1.6. How is the writing system different from Khmer?

- Nikahit (၁) is in the consonant chart and used in the same way as a consonant. Chin et. al. (2012:2) wrote that it is not only used to put on top of a consonant, but also on one of these three independent vowels: ၏ ၒ ၔ (i.e. ၁၏ ၒ၏ ၔ၏).
- Nikahit is used to combine with two vowels: ၃ and ၅ to make ၁၃ and ၁၅. ၁၃ is exceptionally found in Pali, and never in Khmer spelling convention. (It may be replaced by ၄ in Khmer alphabet.)
- Consonants can be stacked together but not pronounced as a cluster. The subscript is pronounced as an initial consonant of the next syllable whether or not there is a vowel after it.

9.7.2. Sanskrit

This inventory is adapted from Huot (1956:1-5,18-21). In a more recent book, Hum (2005:၂,၄၀,၄၇) includes three additional consonants: ၃၂ ၃၃ ၃၄. Sanskrit has 33 consonants, 33 subscripts, nine dependent vowels, 13 independent vowels and six diacritics.

The tables that follow present Khmer characters used in Sanskrit in the first row, romanization of them in the second and the unicode code points in the third row.

9.7.2.1. Sanskrit Consonants

၁	၂	၃	၄	၅	၆	၇	၈	၉	၁၀	၁၁
k	kh	g	gh	ñ	c	ch	j	jh	ñ	
U+1780	U+1781	U+1782	U+1783	U+1784	U+1785	U+1786	U+1797	U+1788	U+1789	
၁၂	၁၃	၁၄	၁၅	၁၆	၁၇	၁၈	၁၉	၁၁၀	၁၁၁	၁၁၂
t	th	d	dh	ṇ	t	th	d	dh	n	
U+178A	U+178B	U+178C	U+178D	U+178E	U+178F	U+1790	U+1791	U+1792	U+1793	
၁၁၂	၁၁၃	၁၁၄	၁၁၅	၁၁၆	၁၁၇	၁၁၈	၁၁၉	၁၁၁၀	၁၁၁၁	၁၁၁၂
p	ph	b	bh	m	y	r	l	v	ç	
U+1794	U+1795	U+1796	U+1797	U+1798	U+1799	U+179A	U+179B	U+179C	U+179D	
၁၁၃	၁၁၄	၁၁၅	၁၁၆	၁၁၇	၁၁၈	၁၁၉	၁၁၁၀	၁၁၁၁	၁၁၁၂	
s		h								
U+179E	U+179F	U+17A0								

9.7.2.2. Sanskrit Subscripts

၁	၂	၃	၄	၅	၆	၇	၈	၉	၁၀	၁၁
k	kh	g	gh	ñ	c	ch	j	jh	ñ	
U+17D2										
၁၁၃	၁၁၄	၁၁၅	၁၁၆	၁၁၇	၁၁၈	၁၁၉	၁၁၁၀	၁၁၁၁	၁၁၁၂	
U+1780	U+1781	U+1782	U+1783	U+1784	U+1785	U+1786	U+1797	U+1788	U+1789	

ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ
t	th	d	dh	n	t	th	d	dh	n		
U+17D2											
U+178A	U+178B	U+178C	U+178D	U+178E	U+178F	U+1790	U+1791	U+1792	U+1793		
ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ
p	ph	b	bh	m	y	r	l	v	ঃ		
U+17D2											
U+1794	U+1795	U+1796	U+1797	U+1798	U+1799	U+179A	U+179B	U+179C	U+179D		
ঃ	ঃ	ঃ									
ঃ	s	h									
U+17D2	U+17D2	U+17D2									
U+179E	U+179F	U+17A0									

9.7.2.3. Sanskrit Independent Vowels

ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ
a	ā	i	ī	u	ū	r	ī	l	e		
U+17A3	U+17A4	U+17A5	U+17A6	U+17A7	U+17A9	U+17AB	U+17AC	U+17AD	U+17AF		
ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ		
ai	o	au									
U+17B0	U+17B1	U+17AA									

9.7.2.4. Sanskrit Dependent Vowels

inherent	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ
a	ā	i	ī	u	ū	r	ī	l	e		
N/A	U+17B6	U+17B7	U+17B8	U+17BB	U+17BC	U+17C1	U+17C3	U+17C4	U+17C5		

9.7.2.5. Sanskrit Combinations of Consonant and Independent Vowel

ঃ [U+1780] serves as a placeholder for any consonant.

ঃ	ঃ	ঃ
kঃ	kঃ	kঃ
U+1780 \ U+17D2 \ U+17AB U+1780 \ U+17D2 \ U+17AC U+1780 \ U+17D2 \ U+17AD		[18]

ঃ is rendered incorrectly here, it should be rendered as shown below

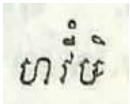


9.7.2.6. Sanskrit Diacritics

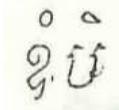
ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ	ঃ
Virama \ ৰিগ্রাম:	Avakraha	Anusvara	Visarga	Athisvara	Anuneaseka [19]						
ঃ	ঃ	ঃ	ঃ	ঃ	ঃ						
U+17D1	U+17DC	U+17C6	U+17C8	U+17CC	N/A						

VIRIAM AVAKRAHASANYA NIKAHIT YUUCALEAPINTU ROBAT -

- ጀ Virama (Huot 1956:20-21)
- ៩ Avakraha (Huot 1956:21)
- ጀ Anusvara or Nikahit (Huot 1956:18-19)
- Anusvara is placed on a vowel (ibid:117)
ហ្មិច



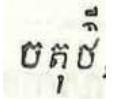
- Anusvara is placed on an independent vowel (ibid:115)
ខ្មិច



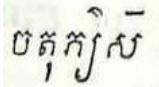
- ០៯ Visarga or Visachani (Huot 1956:19-20)
- Anuneaseka (Huot 1956:19)



- ~ Athisvara RO in a cluster as the second member (Huot 1956:26). It is not listed as one of the 5 diacritics. It is instead an alternative appearance of ឫ when occurring after another consonant in an initial cluster.
- AbvV can be placed above the Athisvara (ibid:133)
ចតុកិនី

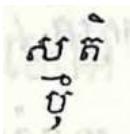


- The Athisvara is placed on a subscript (ibid:160)
ចតុកិនីស



9.7.2.7. How is the writing system different from Khmer?

- The use of the two obsolete characters (i.e. ឃ ង ួ).
- Anuneaseka doesn't exist in the current version of Khmer Unicode character inventory.
- Independent vowels are seen to be used like subscripts.
 - ក្បុង kṛṣṇa ‘black’ (Huot 1956:12)
 - ស្រូបា sprha ‘wish (n)’ (Huot 1956:12)
- Independent vowel which is used like a subscript can be placed under a subscript (ibid:87)
ស្វែក



- Consonant clusters which never exist in Khmer language:
 - សំពើ sattva ‘animal’
 - មេត្យុ matsya ‘fish’
 - វ្រេត្ត vrata ‘buddhist temple’

មេត្យុ is not rendered correctly here, it should be rendered as shown below:



10. Stone Inscriptions (Pre-Angkor, Angkor, Post-Angkor Era)

The oldest stone inscriptions written in Khmer language dated back to the 5th century. In stone inscriptions, there are 33 consonants, 14 vowels and three diacritics (Vong 2011:15). These are used to transcribe inscriptions as they were carved in the stones. The table below shows the equivalence of each of them. For more details on character comparison, see

<https://drive.google.com/drive/folders/1jxv9xxrWNPD0U7j1wDW838elUQMUMrV0?usp=sharing>.

Consonant Subscript Vowels Special Signs

ន	០	អ	៩
៩	០០	អី	៩៩
ច	០៥	ថ	៩២[20]
យ	០១	ឃ	
ច	០២	ា	
ច	០៣	ាំ	
ច	០៤	ឃំ	
ឃ	០៥	ឃី	
ឃ	០៧	ឃុ	
ឃ	០៨	ឃុំ	
ឃ	០៩	ឃុំំ	
ឃ	០៩០	ឃុំ៦	
ឃ	០៩១	ឃុំ៧	
ឃ	០៩២	ឃុំ៨	
ឃ	០៩៣	ឃុំ៩	
ឃ	០៩៤	ឃុំ៩៦	
ឃ	០៩៥	ឃុំ៩៧	
ឃ	០៩៦	ឃុំ៩៨	
ឃ	០៩៧	ឃុំ៩៩	
ឃ	០៩៨	ឃុំ៩៩៦	
ឃ	០៩៩	ឃុំ៩៩៧	
ឃ	០៩៩០	ឃុំ៩៩៨	
ឃ	០៩៩១	ឃុំ៩៩៩	
ឃ	០៩៩២	ឃុំ៩៩៩៦	
ឃ	០៩៩៣	ឃុំ៩៩៩៧	
ឃ	០៩៩៤	ឃុំ៩៩៩៨	
ឃ	០៩៩៥	ឃុំ៩៩៩៩	
ឃ	០៩៩៦	ឃុំ៩៩៩៩៦	
ឃ	០៩៩៧	ឃុំ៩៩៩៩៧	
ឃ	០៩៩៨	ឃុំ៩៩៩៩៨	
ឃ	០៩៩៩	ឃុំ៩៩៩៩៩	

11. Sample Texts for Orthography Check

- (៤) មាត្រិនកម្មភីយដិតសព្វតិ យោរមន្ទា ទ្វាក មកមចនូយកំ ខនិសនេ

វិធីនា ជិត្យា មនិនា តនេដសា រវត កៅ ជយមងលានិ ។

- (iii) ນາງຄົກລົງ ຕັ້ງວ່າ ມີຕືບຕົກກຳ ຖາວທີ່ຕົກມະນີ້ ສູງຄາດນີ້ ເພື່ອສະເໜີ

វិធីនា ជិត្យវា មនិនោ តាមដសា ករព ទៅ ដយមងលានី ។

- (៤) ឧភិត្យគត មតិបាតសរុបនា ជានីយោ ដនបចេងលិមាល លវង់ តទិកិសង

តម្រនា ដិតវា មនិនា តនេជសា រវាទ តែ ដយមងលានី ។

- (៥) កតាន កដុមទ្វាំ តីវ តពិនិយ ចិត្តាយ ទដវចន់ ជនកាយ មធ្វើ សន្លែន

សេចក្តីថ្លែងការណ៍ និត្យរា មនិត្យ តាមដៃសា កវត តែ ជំយមងលានិ ។

- (៤) សច្ច័ន់ វិហាយមក្តី សង្គមភាពពេទ្យកំរាល រាជការណ៍ បិតមនំ អគ្គនៃ ក្រោម បញ្ហាបន្ទិទ្ធេ

បជលិកោ ដីតវ មុនិន្តា តនេដសា កវត្ថុ ពេ ដយមងូលាទិ ។

- (ព) នគ្របនខ្លួនការដំឡើង មហិតី បរកេន ថែរការជោគន ទមាបយន្ត តុលបន្ទះស៊ី

វិធីនា ជិត្យា មុនិនោ តាមដែល ករពុ ទៅ ដឹម្មួលនឹង ។

- (៤) ទូទាត់ទិន្នន័យ រាជក្រឹតា ស្ថិស្សហកំ ព្រៃា វិស៊ី ជតិមិនិ ពករិធានំ ពាណាពទេន

វិធីនា ដីត្រា មនិត្រា តាមជសារ ករែត ទៅ ជយមង្គលានី ។

- (៤) ឯការណ៍ ពួកជយមង្គលអធិភាពា យោ វាទេនា ទិន្នន័យ សរុប មកនឹង ពីតាមនេរក្សា

វិធានី ចុបទភាពី មោកំ សុខ អធិតមេយ្យា នៅ សប្តាហ៍ “

11.7. Sanskrit Sample Text

(To be obtained)

11.8. Inscription Sample Text

(K.557 Inscription, Mungkol Borey, Takeo)

๗- ផ្នែកគេត្តុបាត់គ្រឿងប្រព័ន្ធដែលមានការរំភោគនៅក្នុងប្រព័ន្ធ ។

๒- วิธีการท่องเที่ยวทางการค้าและเชิงพาณิชย์ท่องเที่ยวที่นักท่องเที่ยวต่างประเทศให้ความสนใจมากที่สุด

๓- รัฐธรรมนูญแห่งราชอาณาจักรไทยจะเป็นกฎหมายที่สำคัญที่สุดในประเทศไทย

ចេះតែវរោះដោនវេដោទុកនំជម្រាវបានបាន

12. Wordlist for Orthography Check

The database of the Chuon Nath dictionary is available at: <https://code.google.com/archive/p/khmer-dictionary-tools/downloads>. This is a complete dictionary with headwords, subentries, part of speech, Khmer pronunciation of certain borrowed words, meanings and cross references.

13. Summary

Khmer script is used to write Khmer, minority languages (i.e. Bunong, Tampuan, Brao, Jarai and Kuay), Pali, Sanskrit and inscription. When working with Khmer script on computer, according to the Unicode Standard, it is important to be mindful of the order of characters within each word. To compensate for sounds which do not exist in Khmer, the minority languages make use of characters that are no longer used in modern Khmer. (i.e. ន ុ ី). This paper seeks to describe each character in use in the Khmer script, including usage in ligatures, unicode encoding, text processing and how usage of the characters may differ in minority languages.

It is observed (1) that some obsolete characters are commonly used in minority languages, Sastra and inscriptions. One has to consider the outlook and rendering of those characters. (2) The character ordering is also another issue when it comes to one size fit all scenario. Khmer language uses certain characters for one purpose in one way, while they may be used differently for different purposes in minority languages.

References

- Bernard, J. B. (1902). Dictionnaire cambodgien-français. Hongkong: Imprimerie de la Société des missions étrangères.
- Bunong-Khmer Bilingual Dictionary. (n.d.). Retrieved from http://icc.org.kh/download/Bunong-Khmer_Bilingual%20_Dictionary.pdf.
- Chan, S. (2010). វិធាវើតារក្ស កម្មពាក្យ និង ការប្រើប្រាលពាក្យ “Creating Words, Borrowing Words and Using Words.” Phnom Penh: Royal Academy of Cambodia.
- Chin, C., Ung, S., Men, P., Em, U., & Kheang, S. (2012). ភាសាបាស់ថ្វាក់ទី១ “Pali Language, Grade 1.” National Buddhist Studies.
- Chuon, N. (1967) វចនាឌក្ខមខ្មែរ “Khmer Dictionary”. Phnom Penh: Buddhist Institute.
- Doek, K. (2000). A Study on the Evolution of Khmer Letters. Apsara. Retrieved from <http://www.elibraryofcambodia.org/ka-seuk-sa-ompi-viwat-ney-aksor-khmer-ebook/>.
- Ehrman, M. E., & Kem, S. (1972). Contemporary Cambodian: A Grammatical Sketch. Washington. DC: Foreign Service Institute.
- Ethnolinguistic Groups of Cambodia. (2011, December). Retrieved January 24, 2018, from http://www.unesco.org/new/fileadmin/MULTIMEDIA/FIELD/Phnom_Penh/pdf/ethnolinguistic_groups_of_cambodia_poster.pdf.
- Headley, R. K., Rath Chim, Ok Soeum. (1997). Modern Cambodian-English Dictionary. Kensington, Md: Dunwoody Pr.
- Headley, R. K., Rath Chim. (2014). Modern Cambodian-English Dictionary, Second Edition. Hyattsville, Md: Dunwoody Pr.
- Henderson, E. J. A. (1952). The Main Features of Cambodian Pronunciation. Bulletin of the School of Oriental and African Studies, University of London, 14(1), 149–174.
- Horton, J., Sok, M., Durdin, M., & Ty, R. (2017). Spoof-Vulnerable Rendering in Khmer Unicode Implementations. Presented at the ACIS2017.
- Huffman F.E. (1970). Cambodian System of Writing and Beginning Reader with Drills and Glossary. Retrieved from <http://archive.org/details/CambodianSystemOfWritingAndBeginningReader>.
- Hum, S. (2005). មេណៃកសាសសំគ្លើត “Sanskrit Language Lessons.” Retrieved February 14, 2018, from <http://www.elibraryofcambodia.org/mereang-pheasa-somskrert-prer-lout-proyouk-ning-somrourl-veyeakor-banthem-ebook/>.
- Huot, T. (1956). ដៃយករណីសំគ្លើត “Sanskrit Grammar” (2nd ed.). Phnom Penh: Buddhist Institute. Retrieved February 14, 2018, from <http://www.elibraryofcambodia.org/veyeakor-sam-skert-ebook>.
- Jordi, J. (n.d.). Brao Ombaa Writing System. ICC & SIL.
- Kanjahn, D. (2012, November 1). The Mondulkiri Font Family. Retrieved April 30, 2019, from http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=mondulkiri#23f7dd25.
- Keller, C. E. (2005). Two Systems for Writing Krung. SIL.
- Khin S. (2007). ដៃយករណីភាសខ្មែរ “Khmer Grammar.” Royal Academy of Cambodia. Retrieved from <http://www.elibraryofcambodia.org/veyeakor-peasa-khmer/>.
- Kul, S. (2008). ភាសខ្មែរ “Khmer Language.” Nikrotheawoan Pagoda.
- Maspero, G. (1915). Grammaire de la langue khmère (cambodgien). Paris, Impr. nationale. Retrieved from <http://archive.org/details/grammairedelala00maspgooog>.
- Muang, P. (2012). វចនាឌក្ខមទីពុនិត្យ “Tampuan-Khmer Dictionary.” (P. Tuy & S. Chhuk, Eds.) (First Draft). Phnom Penh. Retrieved from <http://www.tampuanreader.com/en/dictionary>.
- Nhok, T. (1962). បែបរៀនថ្មី ភាសាបាសជានជូន “New Method to Learn Basic Pali.” Retrieved February 14, 2018, from <http://www.elibraryofcambodia.org/beb-rean-thmey-pheasa-balei-jon-dombong/>.
- Nuon, B. (1954). អក្សរនក្រមខ្មែរ “Khmer Lexicon.” Phnom Penh.
- Open Forum of Cambodia. (2004) How to Type Khmer Unicode, Version 1.0:7–14. Retrieved from <http://khmeros.info/download/KhmerUnicodeTyping.pdf>.
- Developing OpenType Fonts for Khmer Script - Typography. (2018, February 8). Retrieved from <https://docs.microsoft.com/en-us/typography/script-development/khmer>.
- Pawley, M.-S., & Pawley, E. (2013). Cambodian Jarai Khmer Based Writing System.

- Pech, Y. (2006). សេវីវិភាគក្រុមចូល “Tampuan Alphabet Book.” (T. Way, W. Wang, K. Sraen, S. Thieng, L. Thoung, T. Lan, ... S. Chhuk, Eds.). Retrieved from <http://tampuanreader.com/en/mlsp-dl/481/28096/76388?r=0>.
- Scheuren, Z. Q. (2010). Khmer Printing Types and the Introduction of Print in Cambodia: 1877–1977 (Dissertation). University of Reading. Retrieved from https://issuu.com/typefacedesign/docs/zachary_scheuren_matd_dissertationw.
- SCHILLER, E. (1994). Khmer nominalizing and causativizing infixes. Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society, 309–326.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2019. Ethnologue: Languages of the World. Twenty-second edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Sok, M. (2016). Phonological principles and automatic phonemic and phonetic transcription of Khmer words (Master’s Thesis). Payap University, Thailand. Retrieved from http://inter.payap.ac.th/wp-content/uploads/linguistics_students/Makaras-Thesis.pdf.
- The Unicode Consortium. The Unicode Standard, Version 10.0.0, (Mountain View, CA: The Unicode Consortium, 2017. ISBN 978-1-936213-16-0). Retrieved January 23, 2018, from <http://www.unicode.org/versions/Unicode10.0.0/>.
- Thun, H. (2011). ឯករាងក្រុមចូល “Khmer Grammar.” Retrieved January 16, 2019, from <http://www.elibraryofcambodia.org/veyeakor-khmer-somrab-krob-phom-siksa-ebook/>.
- Um, B., & Seng, T. (2012). Khmer Grammar for Primary School. Phnom Penh: Publishing and Distribution House.
- Vong, S. (2011). សិលាបារីកនៃប្រព័ន្ធកម្ពុជាសម័យមួនអង្គរ “Pre-Angkor Inscriptions of Cambodia” (2nd ed., Vol. 1). Phnom Penh: Editions Angkor.

Appendices

A. Word-Initial Consonant Clusters

All possible word-initial consonant clusters as found in the headwords in Khmer-Khmer Dictionary together with their phonemic representations, series, number of instances in the dictionary and the examples are illustrated in the table below. In the official Khmer-Khmer dictionary, 4397 instances of word-initial consonant clusters are found and there are at least 164 possible unique clusters. Nine consonants (#33, #55, #56, #57, #58, #59, #134, #135 and #169) do not take a subscript in word-initial position.

The number of instances found in the Khmer-Khmer Dictionary (KKD) is obtained by doing RegEx^[21] searches on a pattern matching any words whose first member in the word-initial position is a consonant followed by a subscript without any trailing subscript after it. For instance, to find words beginning with ក្រ, look this pattern “^ក្រ([_])\$” up in an online dictionary (i.e. <http://dictionary.tovnah.com/regex-search>). This could give a list of words which begin with ក្រ followed by either any character beside a subscript or nothing at all.

No Cluster Phoneme Series Instances in KKD Grapheme Example Phonemic Transcription Gloss

No	Cluster	Phoneme	Series	Instances	Grapheme	Example	Phonemic	Transcription	Gloss
1	ក្រ-	kg	1st	21	ក្រោក	kjaok		peacock	
2	ក្ត-	kd'	1st	44	ក្តាម	kfaam		crab	
3	ក្ន-	kd'	2nd	5	ក្នក	kduk		sound of something falling	
4	ក្ន-	kt	2nd	2	ក្នអ	ktum		tree of paradise	
5	ក្ន-	kn	1st	6	ក្នឃ	knog		in	
6	ក្រ្យ-	k&	1st	32	ក្រ្យាល	kbaal		head	
7	ក្រ-	km	1st	4	ក្រោង	kmeeŋ		young	
8	ក្រ-	kr	1st	313	ក្រាម	kraom		under	
9	ក្រ-	kl	1st	36	ក្រឹក	kliak		armpit	
10	ក្រ-	kw	1st	4	ក្រេន	kwaen		agile	
11	ក្រ-	ks	1st	17	ក្រាន	ksaan		serene	
12	ក្ត-	k?	1st	24	ក្តាម	k?dam		clay pot	
13	ក្ន-	kc	1st	16	ក្នី	kcaw		snail	
14	ក្ន-	kc	2nd	23	ក្និល	keil		lazy	
15	ក្ន-	kj	1st	32	ក្និ	knoj		ginger	
16	ក្ន-	kt	1st	1	ក្នត	ktaat		to cackle (of hen)	
17	ក្ន-	kt	2nd	30	ខ្នុម	ktoom		hut	

18	ក-	kn	1st	34	ក្បង	knaaqj	back (of a person)
19	ខ-	kp	2nd	16	ខ្លែង	kpuəh	tall, high
20	ខ-	km	1st	34	ខ្លីរ	kmae	Khmer
21	ខេ-	kj	1st	3	ខ្សំដំបូង	kjaaqj	mollusc
22	ខ-	kl	1st	25	ខ្នា	klaa	tiger
23	ខ-	kw	1st	53	ខ្លិត	kwət	wood apple
24	ខេ-	ks	1st	33	ខ្សោច	ksac	sand
25	ខ្ល-	kn	2nd	6	ខ្លាត	knie	1S, 3S (informal)
26	ក្រុ-	kr	2nd	143	ក្រាប់	kroap	seed
27	ក្រុ-	kw	2nd	1	ក្រាម	kwaam	text, composition (Thai loanword)
28	ក្យ-	kp	2nd	1	ក្យើច	kneəc	in a jerking manner
29	ក្យ-	kn	2nd	7	ក្យាប	kniep	pincers
30	ក្យ-	km	2nd	10	ក្យំ	kmum	bee
31	ក្យ-	kl	2nd	27	ក្យួង	kluəqj	leprosy
32	ក្យ-	kw	2nd	9	ក្យាល	kwiel	to herd
33	ក្យ-			0			
34	ក្រ-	cn	1st	4	ក្រុះ	cnaj	to polish a gem
35	ក្រុ-	c6	1st	23	ក្រោង	cbaŋj	to do battle
36	ក្រុ-	cj	1st	2	ក្រុត	cjot	to die (of angel)
37	ក្រុ-	cr	1st	132	ក្រុនិន	creən	much/many/a lot
38	ក្រ-	ck	1st	20	ក្រុះ	ckae	dog
39	ក្រ-	ck	2nd	2	ក្រុង	ckəŋj	to be awkward
40	ក្រ-	cŋ	1st	12	ក្រាយ	cŋaaj	far
41	ក្រ-	cd	1st	4	ក្រាហ	cdaø	k.o. fish
42	ក្រ-	cn	1st	21	ក្រុក	cnok	cork, stopper
43	ក្រ-	cp	2nd	7	ក្រុនិន	cpin	k.o. fish
44	ក្រ-	cm	1st	11	ក្រុ	cmaa	cat
45	ក្រ-	cl	1st	26	ក្រើយ	claoj	to answer
46	ក្រ-	cw	1st	16	ក្រុង	cweenj	left (side)
47	ក្រ-	c?	1st	15	ក្រុង	c?əŋj	bone
48	ក្រុ-	cr	2nd	83	ក្រុក	cruuk	pig
49	ក្រ-	cw	2nd	2	ក្រា	cwie	Java
50	ក្យ-	cŋ	2nd	5	ក្យាក	cŋook	to bow the head
51	ក្យ-	cn	2nd	13	ក្យុត	cnuət	headband
52	ក្យ-	cm	2nd	11	ក្យាល	cmoool	male (non-human)
53	ក្យ-	cl	2nd	19	ក្យើង	cleəŋj	leech
54	ក្យ-	cw	2nd	4	ក្យុង	cweenj	to be clear, pure
55	ក្រ-			0			
56	ក្រ-			0			
57	ក្រ-			0			
58	ក្រ-			0			
59	ក្រ-			0			
60	ក្យ-	nh	1st	2	ក្យិយ	nhaəj	Don't bother
61	ក្រ-	tŋ	1st	1	ក្រោក	tjaok	neck-fetter (for criminals)
62	ក្រ-	tn	1st	1	ក្រោត	tnaot	sugar-palm tree
63	ក្រុ-	t6	1st	15	ក្រាល	tbal	mortar
64	ក្រ-	tm	1st	6	ក្រាត	tmaat	vulture

65	ព្រិ-	tr	1st	230	ត្រូល់	tral	weaver's shuttle
66	ធ្វុ-	tl	1st	9	ធ្វុក	tlok	clown
67	ធ្វុ-	tw	1st	3	ធ្វុា	twaa	beef sausage
68	ធ្វុ-	t?	1st	7	ធ្វុក	t?aok	to hiccup
69	ច្បែ-	tk	1st	14	ច្បែល់	tkal	to prop up
70	ច្បែ-	tk	2nd	2	ច្បោម	tkiem	molar (tooth)
71	ច្បែ-	tj	1st	13	ច្បោស	tjaah	forehead
72	ច្បែ-	td	1st	4	ច្បោក	tdaok	small wooden bells
73	ច្បែ-	tn	1st	22	ច្បោល	tmaal	seed bed
74	ច្បែ-	tp	2nd	9	ច្បោល់	tpoal	cheek(s)
75	ច្បែ-	tm	1st	11	ច្បែ	tmaa	rock, stone, concrete
76	ច្បែ-	tl	1st	20	ច្បោម	tlaom	liver
77	ច្បែ-	tw	1st	15	ច្បោយ	twaaj	to give, offer (to royalty, clergy or deities)
78	ទូរ-	tj	2nd	1	ទូរភីសាប្បន្ធ	tjoo.te?. sah	a sacred script foretelling future
79	ប្រ-	tr	2nd	129	ប្រី	trwo	Cambodian stringed fiddle
80	ខ្ល-	tw	2nd	30	ខ្លា	twie	door
81	ឆ្ល-	tj	2nd	8	ឆ្លន់	tjuon	to be heavy
82	ឆ្ល-	tn	2nd	13	ឆ្លី	tneo	shelf
83	ឆ្ល-	tm	2nd	21	ឆ្លញ្ញ	tmeuy	tooth
84	ឆ្ល-	tj	2nd	9	ឆ្លង	tjuuj	charcoal
85	ឆ្ល-	tl	2nd	21	ឆ្លក់	tleak	to fall (unintentionally)
86	ឆ្ល-	tw	2nd	5	ឆ្លី	twao	to do
87	ប្រិ-	nr	2nd	2	ប្រាយ	nriej	Narayana (epithet of Vishnu)
88	នឹ-	n?	1st	1	នឹលនិង	n?aal.niŋ	in order that
89	បុ-	pd	1st	4	បុី	pdej	husband
90	ប្រ-	pr	1st	517	ប្រាក់	prak	silver
91	ប្រ-	bl	1st	39	ប្រែក	plack	to be odd
92	ប្រ-	b?	1st	2	ប្រុន	p?oon	younger sibling
93	ជ្រ-	pk	1st	20	ជ្រាយ	pkaaj	star
94	ជ្រ-	pk	2nd	13	ជ្រំ	pkum	to group/assemble
95	ជ្រ-	pj	1st	9	ជ្រារ	pjaah	to be face up
96	ជ្រ-	pc	1st	13	ជ្រាច្រ	pcanj	to defeat
97	ជ្រ-	pp	1st	7	ជ្រី	pjae	to send
98	ជ្រ-	pd	1st	33	ជ្រិក	pdet	to pat dry
99	ជ្រ-	pt	1st	7	ជ្រិល	ptol	copper/silver bowl for water
100	ជ្រ-	pt	2nd	30	ជ្រេះ	pteah	house
101	ជ្រ-	pn	1st	21	ជ្រូវ	pnoo	tomb
102	ជ្រ-	pl	1st	28	ជ្រុវ	plew	road, street, path
103	ជ្រូ-	ps	1st	32	ជ្រូង	psaej	smoke (n)
104	ជ្រូ-	p?	1st	31	ជ្រូម	p?aem	to be sweet
105	ពុ-	pk	1st	1	ពុល	pkol	k.o. large tree
106	ពុ-	pn	2nd	8	ពុង	pnooŋ	Bunong ethnic group
107	ពុរ-	pj	2nd	34	ពុរេះ	pjuh	storm
108	ព្រា-	pr	2nd	201	ព្រាន	prien	hunter
109	ពុ-	pl	2nd	3	ពុក	pluk	ceremonial leader in cremating a corpse
110	ពុ-	p?	2nd	1	ពុះ	p?ih	tightly
111	ពុ-	pj	2nd	10	ពុត	pjuut	to give someone a bath

112 ព-	pc	2nd	12	ពិត	p̄it	to seal
113 ឆ-	p̄p	2nd	7	ឆាក់	p̄neak	to wake up
114 ន-	pn	2nd	28	នៅកិ	pneek	eyes
115 ន-	pm	2nd	1	នាល	pmich	entirely
116 ន-	pl	2nd	49	នុក	pluk	tusk
117 ព្រ-	pr	2nd	4	ព្រឹន	pruun	stomach worm
118 ស-	ps	2nd	1	សំណើ	psoam	to accustom to
119 ម-	mk	1st	2	ម្នាក់	mkak	k.o. fruit tree
120 ម-	mk ^h	1st	1	ម្នាច់	mk ^h aap	one side
121 ម-	m̄j	2nd	2	ម៉ែ	m̄aj	one day
122 ម-	mc	1st	6	ម្នាស់	mcah	owner
123 ម-	mc	2nd	2	ម្នល	mcul	needle
124 ម-	mp	2nd	2	ម៉ែម៉ែ	mpae.mpaa	to be always making excuses
125 ម-	md	1st	5	ម្នង	mdsaaj	one time
126 ម-	mn	2nd	14	ម្នាក់	mneak	one person
127 ម-	mp ^h	2nd	1	ម៉ែ	mp ^h ej	twenty
128 ម-	m̄j	2nd	6	ម៉ាង	m̄jaaj	one way
129 ម-	mr	2nd	23	ម្រាម	mriem	finger/toe
130 ម-	ml	2nd	10	ម្បច	mlup	shade
131 ម-	ms	1st	7	ម៉ោរ	msaw	flour
132 ម-	mh	1st	10	ម្បប	mhoop	food
133 ម-	m?	1st	2	ម្បម	m?aam	k.o. aromatic grass
134 ម-			0			
135 ឌ-			0			
136 ល-	lk	1st	2	លូម	lkaam	to be very tender
137 ល-	lk ^h	1st	1	លូន	lk ^h aon	theatrical performance
138 ល-	lk	2nd	6	លីក	lkik	if only
139 ល-	lj	2nd	10	លូ	ljeo	sesame
140 ល-	l̄b	1st	29	លូយ	l̄baaj	solution, blended
141 ល-	lp	2nd	2	លូវី	lp̄iw	pumpkin
142 ល-	lm	2nd	23	លូតិ	lmut	sapodilla tree
143 ល-	lj	2nd	1	លីំ	ljum	to be dangling
144 ល-	lw	2nd	16	លីង	lwiij	bitter
145 ល-	lh	1st	16	លីង	lhuū	papaya
146 ល-	l?	1st	21	លូង	l?aaaj	cave
147 ន-	wh	1st	3	នៅ	whoj	in a daze
148 ស-	sk	1st	29	សូវ	skaa	sugar
149 ស-	sk	2nd	19	សូវ	skoɔ̄	drum
150 ស-	sj	1st	12	សូប	sjaap	to yawn
151 ស-	sj	1st	15	សូីប	snaəp	to shudder (in fear)
152 ស-	sf	1st	29	សូំ	sfam	right (hand side)
153 ស-	st	1st	16	សូក	stok	stock
154 ស-	st ^h	1st	14	សូន	st ^h aan	place
155 ស-	st	2nd	54	សូង	stoɔ̄j	bunch (of banana)
156 ស-	sn	1st	60	សូណិ	snaa	crossbow
157 ស-	s̄b	1st	33	សូវី	s̄bəw	k.o. coarse grass
158 ស-	sp	2nd	28	សូន	spien	bridge

159 ស្រ-	sm	1st	69	ស្រា	smaa	shoulder
160 ស្រិ-	sr	1st	318	ស្រិ៖	srah	pond
161 ស្អ-	sl	1st	74	ស្អីក	slək	leaf
162 ស្អ-	sw	1st	67	ស្អាយ	swaj	mango
163 ស្អ-	s?	1st	25	ស្អែក	s?æk	tomorrow
164 ម្ន-	n	1st	6	ម្នីង	nəŋ	this, these
165 ម្ន-	m	1st	14	ម្នត់	mat	all gone
166 ប្រ-	r	1st	3	ប្រាស	hrah.swa?	dwarf
167 ព្រ-	l	1st	10	ព្រឹង	luəŋ	king
168 ព្រ-	w	1st	10	ព្រឹល	fił	film
169 ទ្រ-			0			
170 ន្វ-	?y	1st	1	ន្វីង	?jaeŋ	2P (to younger girls)
171 ន្វ-	?n	1st	15	ន្វុក	neak	2P
172 ន្វ-	?w	1st	2	ន្វី	?woj	what
173 ន្វ-	?h	1st	17	ន្វីង	?haeŋ	2P (to younger people)

B. Word-Medial Consonant Clusters

The table below show a list of all possible consonant clusters occurs in word-medial position. To do so, a RegEx pattern is used to filter for them. The pattern is “[↑]CC[↑]” which guarantees that there is no preceding or trailing subscript. Figures and examples are obtained from an online dictionary (<http://dictionary.tovnah.com/reg-search>). For example, [↑]ក្ត.ក[↑] matches any word containing ក្ត and ក which is followed by a consonant which has no subscript after it.

In order not to show the same sequences of the consonant clusters, only the ones whose sequences are different from the ones occurs in the word-initial position are presented.

-CS- Instances Example

-ក្ត.-	430	ពក្តមា /tak.kaʔ.maa/ ‘to be stupefied’
-ក្ត-	46	អក្តណ /?ak.kʰaʔ.raa/ ‘letters’
-ក្តរ-	117	ពក្តរកំពូ /piek.kam.prie/ ‘a single character word’
-ក្តរ-	4	ក្តី /pʰeaʔ.kdəj/ ‘loyalty’
-ក្ត [22]	27	អបសក្តី /?ap.sak/ ‘powerless’
-ក្ត-	9	ទក្តិណា /teak.sən/ ‘the south’
-ក្តរ-	61	
-ខ.-	68	(All are similar to word-initial consonant clusters.)
-គ.-	157	បគ្គល /bok.kul/ ‘individual’
-គ-	50	យោគ្គរាល /joo.kjeaʔ.pʰiep/ ‘intelligence’
-គរ-	2	ខគ្គរាល /?uk.kʰoo.saʔ.naa/ ‘public address’
-គរ-	2	អគ្គគ្រោន /?aŋ.kleh/ ‘English’
-គ-	2	
-ឃ.-	16	វិឃរាល /wiʔ.kreaʔ.pʰiep/ ‘destruction’
-ឃរ-	1	
-ឃ-	640	កងឃប /kaŋ.kaep/ ‘frog’
-ឃ-	172	ជងឃ /caŋ.kʰaw/ ‘head of a commercial house’
-ឃ-	33	ជងី /caŋ.kum/ ‘bunched together (of trees)’
-ឃ-	153	សងឃម /saŋ.kʰim/ ‘hope (n)’
-ឃរ-	27	បងឃរួយ /baŋ.baoj/ ‘to swing the arms while walking’
-ឃរ-	1	បងឃក /buŋ.riik/ ‘to magnify’
-ឃ-	59	កងឃល /kaŋ.wal/ ‘worry (n)’
-ឃ-	58	សងឃយ /saŋ.saj/ ‘to suspect’

-ឆ-	30	កំឆារ /koŋ.haa/ ‘fan’
-ប-	62	បេប៉ុម /baŋ.?aem/ ‘dessert’
-ច-	46	
-ច-	167	បច្ចុយ /pac.caj/ ‘suffix’
-ច-	53	មច្ចាត /mac.c ^h aa/ ‘fish’
-ច-	88	
-ន-	28	(All are similar to the word-initial consonant cluster.)
-ន-	120	
-ន-	6	ល្អជ្រើន /pdac.naa/ ‘to commit’
-ន-	64	ពាណិជ្ជកម្ម /pie.nic.cea?.kam/ ‘commercial’
-ន-	23	មជ្ជភាស /mac.c ^h ea?.t ^h aan/ ‘habitat’
-ន-	2	រដ្ឋរាជក្រឹត /rieck.jieq.kea?/ ‘royal administration’
-ល-	6	(All are similar to the word-initial consonant cluster.)
-ល-	320	
-ល-	71	កញ្ចក់ /kap.caŋ/ ‘mirror, glass’
-ល-	33	កញ្ចា /kap.chaa/ ‘marijuana’
-ល-	91	កញ្ចី /kap.cəə/ ‘basket’
-ល-	9	កញ្ចូស /kap.c ^h uuuh/ ‘to scrape the foot on the ground’
-ល-	113	កញ្ចា /kap.naa/ ‘young woman’
-ល-	1	បញ្ចា /pan.haa/ ‘problem’
-ព-	114	
-ព-	109	សេដ្ឋី /see.t ^h ɔj/ ‘wealthy man’
-ព-	5	អត្ថប្រពិក្ស /?at.da?.pra?.te?.p ^h uu/ ‘advocate’
-ព-	0	(no result found)
-វ-	18	វិវឌ្ឍ /wut.t ^h i?/ ‘prosperity’
-វ-	16	លេខបាតិ /leet.du?.baat/ ‘the fall of a clump of earth’
-វ-	2	
-ន-	1	អាមេរីក្រុណ /?aa.tjea?.caa/ ‘to have been formerly wealthy’
-ន-	1	
-ណ-	138	កណ្ឌីង /kan.dəŋ/ ‘bell’
-ណ-	7	សណ្ឌាគារ /san.t ^h aa.kie/ ‘hotel’
-ណ-	75	តណ្ឌ /?ən.duu/ ‘Hindu’
-ណ-	1	ឯណ្ឌី /t ^h un.t ^h i?/ ‘Genesha (son of Shiva)’
-ណ-	11	បណ្ឌាគារ /pan.naa.kie/ ‘bookstore’
-ណ-	32	បុណ្យទិន /bon.jea?.tin/ ‘holiday’
-ណ-	1	តណ្ឌុ /ton.haa/ ‘desire, passion’
-ណ-	479	
-ណ-[23]	236	កិត្តិយស /kət.te?.juəh/ ‘reputation’
-ណ-	7	ឧត្តម /?ut.dəm/ ‘excellent’
-ណ-	17	សព្វា /sa?.tja/ ‘true words’
-ណ-	82	វត្ថុ /woat.t ^h o?/ ‘thing’
-ណ-	4	ទសវត្ថុយ /tuə.sa?.woat/ ‘a period of ten years’
-ណ-	4	ចត្តា /cat.twaa/ ‘four’
-ច-	25	មិច្ចាបារ /mi?.tja.caa/ ‘adultery’
-ច-	1	
-ថ-	235	ពុទិ /put.t ^h i?/ ‘intellect’
-ថ-	102	សទ្ធផាគិ /sat.tea?.ciet/ ‘sound, voice, speech’
-ថ-	40	
-ថ-	30	(All are similar to the word-initial consonant cluster.)

-ខ-	735	ខេត្ត /kʰan.təj/ ‘tolerance’
-ខ-	141	កង្ហារ /kan.tʰao/ ‘spittoon’
-ខ-	25	កន្ទុយ /kan.tuj/ ‘tail’
-ខ-	224	កន្ទាយ /kan.tʰiey/ ‘soft-shelled turtle’
-ខ-	93	ទិន្ទន័យ /tɪn.nea?.nej/ ‘data’
-ខ-	34	បុន្ទាន /pon.maan/ ‘how much/many’
-ខ-	7	កន្ទុ /kon.jaa/ ‘unmarried girl’
-ខូ-	21	ពុន្ទាយ /pun.riej/ ‘bright and shining with different colors’
-ខូ-	1	កន្ទាត /kan.laat/ ‘cockroach’
-ខូ-	141	ទុន្ទាយ /tun.saaj/ ‘rabbit’
-ខូ-	34	
-បុ-	319	កបុបាល /kap.bah/ ‘cotton’
-បុ-	116	បុច្ចា /bop.pʰaa/ ‘flower’
-បុ-	9	ប្រញូបិ /pra.nap/ ‘prohibition’
-បុ-	8	ប្រាបុគិលាក /praa.pdaa.pʰi?.liep/ ‘full of luck’
-បុ-	2	អប្បរា /?ap.sa?.raa/ ‘Apsara’
-បុ-	4	អីរប់រោច /?eo.roo.pjaj/ ‘european’
-បុ-	1	
-ចុ-	76	(All are similar to the word-initial consonant cluster.)
-ពុ-	225	និញ្ញាន /nip.pien/ ‘Nirvana’
-ពុ-	53	ទុញ្ញសិតិ /tu?.pʰie.sət/ ‘bad language/advice’
-ពុ-	25	ទេញ្ញា /tep.daa/ ‘angel’
-ពុ-	1	
-ពុ-	63	អក្សរតម /?a?.pjie.kum/ ‘war’
-រុ-	3	
-មុ-	660	សមុល /sam.kul/ ‘to be swollen and ugly’
-មុ-	5	តម្លៃវិក /tam.biət/ ‘tweezers’
-មុ-	81	សម្បសួយ /sam.pʰoah/ ‘perception, sensation’
-មុ-	9	សម្បសិតិ /sam.piət/ ‘school bag’
-មុ-	83	អាមុលិ /?aa.rum/ ‘feeling’
-មុ-	61	វម្លើដឹង /rom.ləoŋ/ ‘to uproot’
-មុ-	66	
-យុ-	51	ទេយុជាន /tej.jea?.tien/ ‘gift given to a Buddhist monk’
-យុ-	50	គុយ៉ែ /kuj.hak/ ‘genitals’
-យុ-	1	
-វុ-	18	កុវិ /kee/ ‘honor’
-វុ-	5	វុះ /wea?.rea?.man/ ‘Varman (title used by Khmer kings)’
-វុ-	2	សុរាជាន /soo.rea?.kaan/ ‘k.o. mythical stone’
-វុ-	7	អនុគិតិយ /?an.dec.ra?.tʰəj/ ‘non-Buddhist’
-វុ-	1	អាសិនិស /?aa.see.rea?.pih/ ‘poisonous snake’
-វុ-	1	ទុនិញ្ញ /tu?.rea?.pʰik/ ‘famine’
-វុ-	1	ពុលិញ្ញន /poa.səo.laen/ ‘porcelain’
-វុ-	1	
-លុ-	68	ដុល្លារ /dol.laa/ ‘dollar’
-លុ-	26	
-ធម-	4	ឯធមុ /ci?.whaa/ ‘tongue’
-ធម-	4	
-ធម-	442	និស្សិត /ni?.sət/ ‘student’
-ធម-	110	ពិស្សុលាក /pi?.snu?.look/ ‘rank given to provincial chievers’
-ធម-	5	បុស្សី /poh/ ‘post’
-ធម-	41	អស្សុយ /?ah.caa/ ‘wonderful’
-ធម-	7	

- ពិ. -	40	សំបូទ /so?.rit/ ‘(close) friend’
- ពុ. -	2	
- ទី. -	12	ទទឹករណី /tea?.lhaj.kaa/ ‘proof’
- នី. -	12	
- នួ. -	2	(All are similar to the word-initial consonant cluster.)

C. Word-Final Consonant Clusters

Words ending in consonant clusters are Pali/Sanskrit loanwords.

-CS Instances Example

- ក្ត	6	សក្ត
- ក្រ	37	ខក្រ
- ក្ន	1	បុកលក្ន
- ក្ប	4	លក្ប
- ក្ស	8	ពាក្ស
- ក្រក	6	សក្រក
- ក្រក្ស	39	ប្រគ្គក្ស
- វក្ស	1	ប្រមុខ
- ក្នុ	20	វត្ថុ
- ក្នុក្ស	3	អនក្នុ
- ក្នុក្ស	2	ពិទេក្នុ
- ក្នុក្ស	3	អាណក្នុ
- ក្នុក្ស	2	ស្វេក្នុ
- ក្នុក្ស	6	បណ្តុក្នុ
- ក្នុក្ស	50	អង្គ
- ក្នុក្ស	4	សក្នុ
- ក្នុក្ស	1	ពក្នុ
- ក្នុក្ស	19	ហក្នុ
- ក្នុក្ស	1	សិក្នុ
- ក្នុ	32	កិច្ច
- ក្នុ	2	អបិច្ច
- ក្នុ	8	ពាណិក្នុ
- ក្នុក្ស	1	សក្នុ
- ក្នុ	6	ប្រាក្នុ
- ក្នុ	9	រក្នុ
- ក្នុ	1	ពព្រេ
- ក្នុ	4	ប្រ៉ែក្នុ
- ក្នុ	15	សាមក្នុ
- ក្នុ	5	វក្នុ
- ក្នុ	34	ដក្នុ
- នី	1	ខ៊ី
- នួរ	3	អរិវនួរ
- នួរ	4	វណ្ណ
- នួរ	37	ខណ្ឌ
- នួរ	16	វណ្ណ
- នួរ	7	បុណ្យ
- នួរ	3	ខណ្ឌ

-ព្រ	93	មិត្ត
-ន្ទ	10	បរមត្ត
-ន្ទិ	5	នានិត្ត
-ន្ទ	1	អាត្ត
-ព្យ	19	ពិនិស្ស
-ព្រិ	72	ក្បុព្រ
-ព្រិ	15	សវ្ត
-ព្យ	2	យ្យវិស្ស
-ព្រិ	1	ពពួលច្ចេ
-ខ្ល	4	សខ្ល
-ខ្ល	40	ប្រយុខ្ល
-ខ្លិ	2	រចនាប់ខ្លិ
-នីរ	13	ពនីរ
-នូ	4	សុនូ
-នូរ	1	សុនូរ
-នី	62	យុនី
-នី	2	និគ្រីនី
-នី	37	សុវិនី
-នី	33	ប្រពនី
-នី	13	បង្កុប្បនី
-នី	2	ជនី
-នីរ	5	សុនីរ
-នី	1	ជនី
-នូ	5	ប្រញូនូ
-នូ	1	យល់សណ្ឌ
-ប្រុ	11	កប្រុ
-ព្រុ	20	ទូរសព្វ
-ព្រុ	3	ប្រារព្វ
-ព្រុ	1	អព្វ
-ព្យរ	7	ច្រព្យរ
-ព្រុ	14	សព្វ
-ព្រុ	1	អំព្រុ
-នូម	6	ពុនុម
-នូម	9	បានុម
-នូម	103	កសិកុម
-មូរ	6	មនោរមូរ
-ប្រម	4	កប្រម
-នូម	3	សុនុម
-នូម	2	លុនុម
-នូម	1	លុនុម
-យូរ	19	គណនេយូរ
-រូរ	22	អារូរ
-រូរ	1	សិរូរ
-នូរ	1	សុនូរ
-នូរ	1	បំណោកូរ
-លូរ	3	កលូរ
-នូរ	1	ពាយុគូល

-ឃ្លោ 4	គុរកាសឃ្លោ
-ឃ្លោ 6	កោសឃ្លោ
-ឃ្លោ 1	សិលារីស្លោ
-ឃ្លោ 2	ឈ្មោះស្លោ
-ឃ្លោ 4	ដែស្លោ
-ឃ្លោ 2	ត្រូវស្លោ
-ឃ្លោ 1	ប្រពិស្លោ
-ឃ្លោ 2	អង្គាបុស្លោ
-ឃ្លោ 1	ស្វោស្លោ
-ឃ្លោ 24	សិស្លោ
-ឃ្លោ 2	ព្រហ្ម
-ឃ្លោ 2	អវិវឌ្ឍនោ

D. Three Consonant Clusters

The consonant clusters with two subscripts occurs mostly in word-medial position (i.e. 170 instances found in the Khmer-Khmer Dictionary), then less frequent in the word-final position (i.e. 50 instances), and it rarely occurs in the initial position (i.e. only 4 instances found). There is no instance of when there are three subscripts follows a consonant. See the table below. Instances of three subscripts following a consonant does not exist in the KKD.

Word-Initial

^#:	4	ស្រី /strɔj/ ‘woman’
	2	ប្រំដែង /frap/ ‘brake’
ឃ្លោ-[24]	2	

Word-Medial

-៩-	170	លក្ខី /leak.smaj/ ‘prosperous, wealthy’
-ក្យោ-	3	បក្យាប /baŋ.kraap/ ‘to defeat’
-ក្យោ-	14	សក្យាំ /saŋ.kjaap/ ‘counting’
-ក្យោ-	2	អក្យោសិន /?aŋ.kleh/ ‘English’
-ក្យោ-	1	ចក្យោដុង /caŋ.krug/ ‘wide open’
-ក្យោ-	21	សក្យោភាទី /saŋ.kriec/ ‘head master (Buddhist monk)’
-ក្យោ-	1	ចិក្យោាំ /ceŋ.cram/ ‘to chop’
-ក្យោ-	10	បក្យោស /baŋ.cruəh/ ‘to purposely avoid (of travelling)’
-ក្យោ-	21	បិក្យោិកសត្វ /pət.tjat.tʰeʔ.kaʔ.sat/ ‘vertebrate’
-ក្យោ-	1	កក្យោង /kan.traoŋ/ ‘to jump up (to get something)’
-ក្យោ-	33	កក្យោក /kan.truək/ ‘to be worn out’
-ក្យោ-	43	សក្យាំ /saŋ.tjie/ ‘twilight’
-ក្យោ-	1	បិក្យោិក /bih.kwii/ ‘biscuit’
-ក្យោ-	1	អាក្យោក /?aa.frik/ ‘Africa’
-ក្យោ-	3	

Word-Final

៩:\$	50	រ៉ូក /peak/ ‘face’
ត្រូ-	1	នន្ត្រូ /nəə.reen/ ‘reigning king’
ត្រូ-	13	រ៉ូក /rieħ/ ‘citizen’
ក្យោ-	36	

E. Initial Consonant Clusters in Both Series

The following table is an extended list of [Initial Consonant Clusters](#). The clusters modified by [Consonant Shifters](#) are included next to their respective pair, however they are not officially used or recognized by the public yet. They are there as a reminder that they might be needed in the future, especially with transliterated words.

No attempt has been made to create a list of the more complex initial consonant clusters (i.e. three consonant clusters).

1st Series 2nd Series Phoneme 1st Series 2nd Series Phoneme

ក	ក	kj	ក	ក	pn
ក	ក / ក?	kf	ក	ក	pl
ក	ក	kt	ក	ក	ps
ក	ក	kn	ក	ក / ក?	p?
ក	ក / ក?	kб	ក	ក	pk
ក	ក	km	ក	ក	pn
ក	ក	kr	ក	ក	pj
ក	ក	kl	ក	ក	pr
ក	ក	kw	ក	ក	pl
ក	ក	ks	ក	ក	p?
ក	ក / ក?	k?	ក	ក	pŋ
ក	ក	kc	ក	ក	pc
ក	ក	kŋ	ក	ក	pŋ
ក	ក	kt	ក	ក	pn
ក	ក	kn	ក	ក	pm
ក	ក	kp	ក	ក	pr
ក	ក	km	ក	ក	pl
ក	ក	kj	ក	ក	ps
ក	ក	kl	ក	ក	mk
ក	ក	kw	ក	ក	mk ^h
ក	ក / ក?	ks	ក	ក	mŋ
ក	ក	cn	ក	ក	mc
ក	ក / ក?	cб	ក	ក	mŋ
ក	ក	cj	ក	ក	md
ក	ក	cr	ក	ក	mn
ក	ក	ck	ក	ក	mp ^h
ក	ក	cŋ	ក	ក	mj
ក	ក	cd	ក	ក	mr
ក	ក	cn	ក	ក	ml
ក	ក	cp	ក	ក	ms
ក	ក	cm	ក	ក	mh
ក	ក	cl	ក	ក	m?
ក	ក	cw	ក	ក	lk
ក	ក / ក?	c?	ក	ក	lk ^h
ក	ក	cw	ក	ក	lŋ
ក	ក	nh	ក	ក	lб
ក	ក	tŋ	ក	ក	lp
ក	ក	tn	ក	ក	lm
ក	ក / ក?	tб	ក	ក	lj
ក	ក	tm	ក	ក	lw
ក	ក	tr	ក	ក	lh
ក	ក	tl	ក	ក	l?

ព	ព	tw	ព	ព	wh
ក	ក / ខ?	t?	ក	ក	sk
ក	ក	tk	ក	ក	sg
ក	ក	tj	ក	ក	sj
ក	ក	tdf	ក	ក	sf
ក	ក	tn	ក	ក	st
ក	ក	tp	ក	ក	st ^h
ក	ក	tm	ក	ក	sn
ក	ក	tl	ក	ក	sb
ក	ក	tw	ក	ក	sp
ក	ក	tj	ក	ក	sm
ក	ក	tj	ក	ក	sr
ក	ក	nr	ក	ក	sl
ក	ក	n?	ក	ក	sw
ក	ក	pd	ក	ក	s?
ក	ក	pr	ក	ក	n
ក	ក	pl	ក	ក	m
ក	ក	p?	ក	ក	r
ក	ក	pk	ក	ក	l
ក	ក	pj	ក	ក	w
ក	ក	pc	ក	ក	?j
ក	ក	pjn	ក	ក	?n
ក	ក	pd	ក	ក	?w
ក	ក	pt	ក	ក	?h

F. Orthographic comparison [charts](#)

G. Khmer signs distribution [chart](#)

Notes

[1] Mnong is usually referred to as “Bunong” in Cambodia.

[2] Kuay is usually referred to as “Kuy” or “Kui” in Cambodia.

[3] Headley (2014:x) uses /u/.

[4] Headley (2014:x) and Huffman (1970:8-9) uses /ei/.

[5] Huffman (1970:8-9) uses /ɔi/

[6] Headley (2014:x) and Huffman (1970:8-9) uses /ou/.

[7] There is still a discussion over whether ឃ is a consonant or a vowel. It is widely considered as a consonant though.

[8] When there is no final orthographic consonant, ែ is realized as /e?/ (Huffman 1970:26)

[9] ឃី is discouraged in formal writing.

[10] ZWNJ is used to prevent the default rendering from happening.

[11] ឃោ (Reahmuk) in many way behaves like a vowel, but since the Unicode has it as an After Diacritic, no other diacritic should be placed after it, or it would not be rendered correctly (i.e. ឃង ឃោ = ឃងោ).

[12] The Sanskrit name of these months are adopted from https://en.wikipedia.org/wiki/Month#Hindu_calendar.

[13] Below Vowels are: ឃ ឃ ឃ.

[14] The vowel can be any of these: ឃ, ឃ and ឃ.

[15] This is only found in documents published for the fifth edition of the Chuon Nath dictionary, i.e. ឃឃឃ (ឃ ឃ ឃ) ‘rice’ in present day Khmer is written as ឃឃ (ឃ ឃ ឃ) in Khmer-French dictionary (Bernard 1902).

[16] The paper can be accessed at: [<http://www.panl10n.net/english/final%20reports/pdf%20files/Cambodia/CAM01.pdf>]
(<http://www.panl10n.net/english/final%20reports/pdf%20files/Cambodia/CAM01.pdf>)

[17] The alphabet book only includes the first series of the vowels.

[18] Chrome does not render this sequence correctly, it should be shown up like the other two in the table.

[19] This sign should be added to the Unicode character inventory.

[20] Viriam is no longer used due to readability issue (ibid:16).

[21] RegEx is a shorthand for [Regular Expression](https://en.wikipedia.org/wiki/Regular_expression).

[22] -ឃ- and -ឃ- look identical, but they are different.

[23] -ឃ- and -ឃ- looks identical, but they are different.

[24] The data in the Khmer-Khmer online dictionary has different character sequences. Instead of ឃឃឃ (ឃ ឃ ឃ), its sequence is ឃឃឃ (ឃ ឃ ឃ). The one used in this documentation is ឃឃឃ (ឃ ឃ ឃ).