

Khmer Character Specification/Usages

Author: Makara Sok

Contributors: Marc Durdin,

Todd Bequette,

Becky Bequette,

Seth Wilson, and

Martin Hosken

Last Update: April 09, 2021

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. [Encoding Order](#)
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. ខោូ /khoh/ ‘to be wrong’ + -មោ- /m/, កំប្មោន /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).

C₁ = { p t c k s ? m 1 }

C₂ = { p t c k ? b d m n n g w j l r s h }

C₃ =

4.1.3. Final Consonants

Aspirated plosives (p^h, t^h, c^h, k^h) , implosives (b, d, g), fricatives (f, s, z, ß; except glottal fricative /h/) and a flap (r) 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 ? h m n n g l w j }

Some foreign loanwords (i.e. English or French) may break this rule. For instance, ឃើសុបីក ‘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 ɿ [ɿ] ɿ 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, iə, 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/, /əj/ and /aj/.

V = { i ii ɪ ɨ u uu e ee ə əə o oo əə ɔɔ a aa ə əə iə ie ae ao aə iə uə ea 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.
Disyllables minor syllable + major syllable	A disyllabic word is made up of a minor syllable and a major syllable.

where:

- minor syllable is C(r)V(N)

- C, any consonant
- r, the flap /r/
- V, any vowel
- N, any nasal consonant

- major syllable is C₁C₂V(F)

- C₁, any C₁ consonant
- C₂, any C₂ consonant
- V, any vowel
- F, any final consonant

The minor syllable is susceptible to extreme syllable reduction (Huffman 1972:59-60). Here are some examples:

- /cam.kaa/ ‘farm’ > [cə.kaa]
- /kən.daal/ ‘middle’ > [kə.daal]
- /kaj.craeŋ/ ‘bamboo basket’ > [kə.craeŋ]
- /bəŋ.riən/ ‘to teach’ > [pə.riən]
- /rə.ɓaaŋ/ ‘fence’ > [lə.ɓaaŋ]
- /pra.hael/ ‘perhaps’ > [pə.hael]

Polysyllable multiple monosyllables

This syllable type is common in foreign loanwords, i.e. Pali/Sanskrit, English, French, etc.

4.5. Series Assimilation

Sok (2016:24) wrote:

Series assimilation, សំរួលប្រព័ន្ធឌី or ការអំទិញកើត "lit. sound adaptation" as proposed by Prom (2013:24) and ខ្សោតពិភ័យ "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,

- កែ 'neck' and កំ (ក+ំ) '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 ំ is /aa/.
- កុ 'to be mute' and កិ (ក+ិ) '(proper noun)' are pronounced as /koo/ and /kie/ respectively because they begin with the second series consonant symbol (ក), and the second series of the inherent vowel is /oo/ and the second series of vowel symbol ិ is /ie/.
- កុណាល (ក+ុ+ុ+ំ+ុ) 'head' is pronounced as /kbaal/ because ក /k/ and ុ /b/ are in the first series, and the first series of vowel symbol ំ is /aa/.
- កិន (ក+ិ+ស+ំ) 'first person singular (informal)' is pronounced as /knie/ because ក and ស are both in the second series, and the second series of the vowel symbol ំ 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.

IMPLOSIVE

1st Series ក ុ

2nd Series ៩ -

Phoneme ឃ ុ

1st Series ក ច ន - អ [Zuk]

UNASPIRATED PLOSIVE 2nd Series ព ជ ទ ធន -

Phoneme ក ច ត ប ?

1st Series ខ ស ប៊ ចំ ជំ

ASPIRATED PLOSIVE 2nd Series ឃ ិ ិ ិ ិ ិ

Phoneme ក^h ច^h ត^h ប^h ធម^h

1st Series ិ ិ ិ ិ ិ ិ

FRICATIVE 2nd Series - -

Phoneme ស ឯ

1st Series - - ិ ិ - - - ិ ិ -

SONORANT 2nd Series ឌ ឌ ឌ ឌ ឌ ឌ

Phoneme ឱ ឱ ឱ ឱ ឱ ឱ

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

ឲូង ‘papaya’

character sequence ឲ ូ ង ិ ិ

character mapping 1 h o[u] ិ ិ

series of character 2nd 1st 1st|2nd 2nd

series of syllable 1st

phonemic transcription /lhɔŋ/

ຂ້າວ ‘spur’ is pronounced with the 1st series of the vowel governed by the series of the least sonorous initial consonant, ຂ /k^h/.

ຂ້າວ ‘spur’

character sequence ຂ ແ ດ ວ

character mapping k^h n aa|je 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 (ງ /y/, ພ /p/, ສ /n/, ຖ /m/, ວ /j/, ຕ /r/, ດ /l/, and ຈ /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:

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

For instance, ດັບນີ ‘thorn’ is pronounced as /bun.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.

ດັບນີ ‘thorn’

character sequence	ឃ ស ា ង
character mapping	b n t 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, ឃន្ទីយ (ឃ /b/, ស /n/, ា /t/, ង /j/) 'barracks' is pronounced as /ban.tiej/ (not /ban.taaj/) and កន្ទីយ (ក /t/, ស /n/, ា /aa|je/, យ /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.

ឃន្ទីយ ‘barracks’

character sequence	ឃ ស ា ង យ
character mapping	b n t aa je y
series of character	1st 2nd 2nd 1st 2nd 2nd
series of syllable	1st 2nd
phonemic transcription	/ban.tiej/

កន្ទីយ ‘rabbit’

character sequence	ក ស ា ង យ
character mapping	b n t aa je y

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 សំឡូចក្បែង ‘small voice’ and សំឡូចធំ ‘big voice’ or សំឡូចព្រឹល ‘light voice’ and សំឡូចឆ្លួន ‘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 ា /aa/ 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) + ឃ /aa|ie/ > កឃ /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	យាត់ /k ^h oat/ ‘to forbid’
ង	U+1784	ŋ	2nd	ងារ /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)’
ឃ	U+1788	c ^h	2nd	ឃើញ /c ^h oog/ ‘to reach for (something)’
ញ	U+1789	n	2nd	ញាត់ /noat/ ‘to stuff (something) in tightly’
ដ	U+178A	d	1st	ដី /dai/ ‘hand/arm’
ឋ	U+178B	t ^h	1st	ឋាន /t ^h aan/ ‘place/location’
ឌ	U+178C	d	2nd	ឌី /di ^h / ‘to be obstinate’
ឃ	U+178D	t ^h	2nd	ឃាល់ /t ^h iel/ ‘shield’

ន	U+178E	n	1st	នាត /nat/ ‘to set (an appointment)’
ត	U+178F	t	1st	តាន /taoŋ/ ‘to hang on to’
ថ	U+1790	t ^h	1st	ថា /tʰaa/ ‘to say/tell’
ទ	U+1791	t	2nd	ទា /tie/ ‘duck’
ធ	U+1792	t ^h	2nd	ធាត /tʰoat/ ‘fat’
ន	U+1793	n	2nd	នោលី /niel/ ‘unit of weight (~12oz.)’
ប	U+1794	b	1st	បាត /bat/ ‘to disappear’
ផ	U+1795	p ^h	1st	ផាត /pʰat/ ‘to blow away (of the wind)’
ព	U+1796	p	2nd	ព៉ូត /poat/ ‘to besiege’
ក	U+1797	p ^h	2nd	កើក /pʰuək/ ‘mud’
ម	U+1798	m	2nd	មាត /moat/ ‘mouth’
យ	U+1799	j	2nd	យើរឃ /joɔŋ/ ‘to pull upward’
រ	U+179A	r	2nd	រោង /rooŋ/ ‘roofed structure’
ល	U+179B	l	2nd	លា /lie/ ‘donkey’
វ	U+179C	w	2nd	វិវ /wie/ ‘to crawl’
ស	U+179F	s	1st	សក /sak/ ‘hair’
ហ	U+17A0	h	1st	ហក /hak/ ‘to jump’
ឡ	U+17A1	l	1st	ឡាន /laan/ ‘car/truck’
អ	U+17A2	?	1st	អាន /ʔaan/ ‘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. កំឆោ and កំ are different and they are pronounced differently, /kaaj/ and /kjaa/ respectively.

A list of subscripts are included below. Thun (2011:17-18) wrote, “ឆោ and ឆូ must be written with their subscript (ឆ and ឆូ), but the subscript of ឆូ (ឆូ) /l/ is never used. [...] When subscript ឆ (ឆ) is placed under itself, it has to be replaced by its full form, but smaller in size, i.e. ឆ. Consonant ជ /d/ and ជ /t/ have the same subscript because most words written with ជ /t/ are pronounced as ជ /d/, for example, មាតាំបិតា /mie.daa.bej.daa/, តំណាង /dam.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
ឃ	ឃ	U+17D2	U+1788	c ^h	2nd
ញ	ញ	U+17D2	U+1789	n	2nd
ជ	ជ	U+17D2	U+178A	d	1st
ថ	ថ	U+17D2	U+178B	t ^h	1st

ং	ং	U+17D2 U+178C	d	2nd
ঃ	ঃ	U+17D2 U+178D	t ^h	2nd
঄	঄	U+17D2 U+178E	n	1st
অ	অ	U+17D2 U+178F	t	1st
ঈ	ঈ	U+17D2 U+1790	t ^h	1st
ঊ	ঊ	U+17D2 U+1791	t	2nd
ঋ	ঋ	U+17D2 U+1792	t ^h	2nd
ঌ	ঌ	U+17D2 U+1793	n	2nd
এ	এ	U+17D2 U+1794	b	1st
ও	ও	U+17D2 U+1795	p ^h	1st
ঔ	ঔ	U+17D2 U+1796	p	2nd
ক	ক	U+17D2 U+1797	p ^h	2nd
খ	খ	U+17D2 U+1798	m	2nd
ঘ	ঘ	U+17D2 U+1799	j	2nd
ঙ	ঙ	U+17D2 U+179A	r	2nd
ঠ	ঠ	U+17D2 U+179B	l	2nd
ঢ	ঢ	U+17D2 U+179C	w	2nd
ণ	ণ	U+17D2 U+179F	s	1st
ঠ	ঠ	U+17D2 U+17A0	h	1st
ঢ	ঢ	U+17D2 U+17A1	l	1st
ণ	ণ	U+17D2 U+17A2	?	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: ឃុំខុំ = ឃ ុ ុ ខុំ
- in disyllabic word: បន្ទុំ = ប ន ុ ុ ុ
- in final cluster: សិល្បូរុំ = ស ិ ល ុ រ ុ ុ
- after an independent vowel: ឃុំ = ឃ ុ

In rare occurrences, independent vowel can be a subscript, i.e. សុីវុំ /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 consonant 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
ឃ	U+1784 U+17C9	ឃ	1st ឃុំ /paj/ ‘day (colloquial)’
ឃ	U+1789 U+17C9	ឃ	1st ឃាំ /nam/ ‘to eat’
ម	U+1798 U+17C9	ម	1st មឹោះ /mœən/ ‘ten thousand’
យ	U+1799 U+17C9	យ	1st យុោដុំ /jaap/ ‘way, kind’
រ	U+179A U+17C9	រ	1st រុោកុំ /ruak/ ‘pulley’
វ	U+179C U+17C9	វ	1st វុោគុំ /waj/ ‘hit, fight (colloquial)’

ឃ	U+1794 U+17C9 p	1st	ពេស /pən/ ‘to be skillful’
ឃុ	U+1794 U+17CA ប	2nd	បុច្ចិត /bút/ ‘topaz’
ិស	U+179F U+17CA s	2nd	សេវិប៉ូ /səəp/ ‘to investigate’
ិង	U+17A0 U+17CA h	2nd	ហិនស /hien/ ‘to be bold’
ិអ	U+17A2 U+17CA ?	2nd	អិអុ /?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
ធម	U+17A0 U+17D2 U+179C	f	ធមុន /foon/ ‘phone’
ធបិ	U+17A0 U+17D2 U+179C U+17CA	f	ធបិល /fiil/ ‘film’
ធនិប	U+179F U+17D2 U+17A0	ʃ	ធនិបាប៊ី /ʃee.ɓap/ ‘Shebab’
ធនិអិត	U+179F U+17D2 U+17A0 U+17CA	ʃ	ធនិអិត /ʃii.?iit/ ‘Shiite’
ធគុ	U+17A0 U+17D2 U+179F	z	ធគុរ្យីនិទ /?aa.zaa.tin/ ‘Argentina’
ធបិធបិម	U+17A0 U+17D2 U+179F U+17CA	z	ធបិធបិម /zuum/ ‘zoom’
ធគុ	U+17A0 U+17D2 U+1780	g	ធគុណា /gaa.naa/ ‘Ghana’
ធគុ	U+17A0 U+17D2 U+1782	g	ធគុម /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. ຂ and ແ, ຂ and ແ, ວ and ແ, and ຖ and ແ do not make a cluster, which is why the intersecting cells are empty.

Note that ຈ /tʰ/ almost never have a subscript after it, only one instance that ຈ /tʰ/ is followed by a subscript as in ແຈ (ກ ອ ອຈ).

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

ຂຈ (or ແຈ, 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.

ຝ is the only instance that an independent vowel occurs as a subscript form. It is not commonly used though. For instance, ແຝ່ຍິ is usually written as ແບຸດ່ຍິ.

Both ຂຈ and ແຈ 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.

- ແກສ- ສ ແ ປູ as in ແກສ່ /strəj/ ‘woman’
- ແກຣ- ທ ແ ປູ as in ແກຣ່າ /franj/ ‘French currency’
- ແກຣ- ທ ແ ປູ as in ແກຣ່ກ /frag/ ‘brake (of vehicle)’

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:

- ស្រី- ស ុ ចេ as in ស្រីចេ /sprīc/ ‘spring’
- ស្រី- ស ុ ចេ as in ស្រីនុ /skriin/ ‘screen’
- ស្រី ស ុ ចេ as in ស្រីតិ /striit/ ‘street’
- ឃ្លា- ឃ ុ ចេ as in ឃ្លានុ /għriin/ ‘green’
- ឃ្លា- ឃ ុ ឬ ុ as in ឃ្លាខុ /flaah/ ‘flash’

Three consonant clusters in medial position:

- -ក្សី- ក ុ ឬ ុ as in លក្សី /leak.sməj/ ‘wellness, glory’
- -ម្ចោ- ម ុ ឬ ុ as in បិម្ចោជិកស្តុ /pət.tjat.t̥e?.ka?.sat/ ‘vertebrae’
- -ស្រី- ស ុ ចេ as in សំស្រីតិ /saŋ.skrot/ ‘Sanskrit’
- -ស្អី- ស ុ ឬ ុ as in បិស្អីតិ /bi.i.skwii/ ‘biscuit’
- -ក្រី- ក ុ ចេ as in អក្រីចិ /?aŋ.kraap/ ‘k.o. large red ant’
- -ម្ចី- ម ុ ឬ ុ as in សំម្ចី /saŋ.kja/ ‘counting’
- -ស្រី- ម ុ ចេ as in សំស្រីះ /saŋ.kruəh/ ‘to save/rescue’
- -ម្ចី- ម ុ ឬ ុ as in អម្ចីសុ /?aŋ.kleh/ ‘English’
- -ព្រី- ម ុ ចេ as in សិព្រីរ៉ូជិ /saŋ.kreac/ ‘monk chief’
- -ប្រី- ប ុ ឬ ុ as in ចិប្រីញំ /cəŋ.cram/ ‘to chop repeatedly’
- -ប្រី- ប ុ ចេ as in កប្រីញំ /kaŋ.crooŋ/ ‘fox’
- -ត្រី- ត ុ ចេ as in កត្រីញំ /kan.trieŋ/ ‘chief of a clan’
- -នីរី- ន ុ ឬ ុ as in សនីរី /san.tjie/ ‘twilight’
- -ត្រី- ន ុ ចេ as in កត្រីញំ /kan.traj/ ‘scissors’

Three consonant clusters are also found in the final position.

- -**ក្រោន** ស ុ ូ as in **ក្រោនទ្រុា** /kan.trua/ ‘wheelless pulley’
- -**ក្រោន** ស ុ ូ as in **តាមសាត្រូន្យ** /daa.raa.sah/ ‘astronomy’
- -**ក្រោន** ស ុ ូ as in **សុរីន្យ** /so?reen/ ‘Indra’
- -**ក្រោន** ន ុ ូ as in **ក្រោនបោក** /phreak/ ‘face’

ឯ /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 members, or it is not pronounceable. For instance, **ស្រីអូ** should be typed as it is spelled and pronounced, like (ស /s/+ ូ /p/+ ូ /r/+ ី /ii/+ អ /ŋ/ → /spriiŋ/), not (ស /s/+ ូ /r/+ ូ /p/+ ី /ii/+ អ /ŋ/ → /srpiiŋ/). This is how students are taught how to spell and pronounce words in school. Keyboard designers may use visual order (i.e. what you see first from left to right, you type first), phonetic order (i.e. you type as you would pronounce) or a combination of the two; as long as the outputs are consistent with the expected character order--the key is ូ /r/ has to occur last in the character sequence.

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, ន /k/ in **នៅក្យ** /s?aeck/ ‘tomorrow’ is a final consonant because it is in the coda position of the syllable and it is pronounced after the vowel ឯ /ae/. However, it is impossible to do syllable analysis algorithmically because of the ambiguities in determining where the syllable boundaries in a sentence occur, 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 ឆ /c^h/, ឆ /c^h/, ជ /p^h/, ហ /h/, ឌ /l/ and ឌ /ʔ/. There are instances which could be confusing whether they are a final consonant or an initial consonant. For instance, ក្រាង ‘jug’ is composed of three orthographic characters: ក រ ង, and it may be thought that ង is the final consonant. It is actually not the 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.

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

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

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

F = {ក ខ គ ឃ ច ន ញ ញ ជ ប ឌ ឈ ធម ឈ ត ម ទ ន ប ព ក ម យ ឯ ឱ ឱ} or

a random clusters of two or three consonants (See [Consonant Clusters](#))

It is important to note that, unlike old Khmer, a 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	ចំហាយ /cam.haaj/	‘steam (n)’
ចំហុយ	ចំហុយ /cam.hoj/	‘steam (v)’

Bernard (ibid) mostly used subscript ឃ after dependent vowel. But it is not always the case, for instance, ចំបូយ (with consonant យ) ‘to steam’ was not written as ចំបូឃ (with subscript ឃ).

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

នី	នី /nii/	‘female’
ដី	ដី /dej/	‘soil’
ក្រៀម	ក្រៀម /crej/	‘fig tree’

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

ជោយ “jaya”	ជោយ /cej/	‘victory’
វោយ “bhaya”	វោយ /p ^h ej/	‘fright’

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

ទាំង = ទាំង	ទាំង /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 /oə/. 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	/koo/
ខ	/kʰaa/	ខ	/kʰoo/
ឆ	/ŋaa/	ឆ	/ŋoo/
ច	/caa/	ច	/coo/
ឆ	/cʰaa/	ឆ	/cʰoo/
យ	/jaa/	យ	/joo/
ជ	/daa/	ជ	/doo/
ឋ	/tʰaa/	ឋ	/tʰoo/
ន	/naa/	ន	/noo/
ត	/taa/	ត	/too/
ច	/tʰaa/	ច	/tʰoo/
ប	/ɓaa/	ប	/booo/
ຜ	/pʰaa/	ຜ	/pʰoo/
ຍ	/jaa/	ຍ	/jooo/
រ	/raa/	រ	/rooo/

ឃា	/laa/	ឃូ	/loə/
ឃូ	/waa/	ឃូ	/woə/
ឃុ	/saa/	ឃុ	/soə/
ឃុ	/haa/	ឃុ	/hoə/
ឃុ	/?aa/	ឃុ	/?oə/

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, ឃុ /kaa/ and ឃុ /koə/ in Khmer are pronounced as /ka?/ and /ke?/ respectively.

5.2.4.2. Dependent Vowels

There are two types of vowels: dependent vowels and independent vowels. Dependent vowels must be attached to an initial consonant, whereas independent vowels can initiate a syllable. 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	កី	ក ី	/kee/
right	កា	ក ា	/kaa/
above	កិ	ក ិ	/ke?/
below	កូ	ក ូ	/ko?/
around	កីំ	ក ីំ	/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’	ិ /koɔ/ ‘to be mute’
ច U+17B6	តា /taa/ ‘grandfather’	ទី /tie/ ‘duck’
ច[៨] U+17B7	បិច្ច /pət/ ‘to cut obliquely’	ពិតិ /pit/ ‘to be true or real’
ស U+17B8	សើ /səj/ ‘shuttlecock’	សិី /sii/ ‘to eat (impolite)’
ទ U+17B9	ពីក /tək/ ‘10 cm’	ទីក /tik/ ‘water’
ខ U+17BA	កី /kəə/ ‘heart(s) (suit of ‘cards)’	គី /kii/ ‘to be’
ក U+17BB	កូក /kok/ ‘egret’	គូក /kuk/ ‘prison’
ក U+17BC	កូវ /koɔ/ ‘to stir’	គូវ /kuu/ ‘to draw’
ក U+17BD	កូវ /kuə/ ‘pod’	គូវ /kuə/ ‘should’
ទេ U+17BE	ពីឱ /taə/ ‘question marker’	ទីឱ /təə/ ‘to be held’
ធមី U+17BF	ពីឱ /tiə/ ‘dwarf’	ឯី /ciə/ ‘to believe’
ធមុ U+17C0	ពីឱប /tiəp/ ‘k.o. bowl’	ទីឱប /tiəp/ ‘custard apple’
ឱ U+17C1	កិឱ /keen/ ‘to cheat or exploit’	តិឱ /keej/ ‘to sleep (for kid)’
ឱ U+17C2	កិឱ /kae/ ‘to edit or adjust’	តិឱ /keə/ ‘craw (of birds)’
ឱ U+17C3	កិឱ /kaj/ ‘trigger’	តិឱ /kej/ ‘to be frugal’
ឱ U+17C4	កិឱ /kao/ ‘to shave’	តិឱ /koo/ ‘cow’
ឱ U+17C5	តិឱ /taw/ ‘unit of measurement’	ទិឱ /tiw/ ‘to go’
ិ U+17BB U+17C6	កិិ /kom/ ‘(don’t)’	ពិិ /kum/ ‘to hold a grudge’
ិ U+17C6	ចិ /cam/ ‘to be exactly at’	ជិ /cum/ ‘to discuss’
ិ U+17B6 U+17C6	មិ /cam/ ‘to remember’	ជិ /coam/ ‘to bruise’

ឃោះ	U+17C7	ពឃោះ /tah/ ‘hurriedly’	ឃោះ /teah/ ‘to slap’
ឃុំ	U+17BB U+17C7	ឃុំ /coh/ ‘to go down’	ឃុំ /cuh/ ‘to defecate’
ធមោះ	U+17C1 U+17C7	ធមោះ /ceh/ ‘to know’	ធមោះ /ceh/ ‘to chip (of skin/bark)’
ិកោះ	U+17C4 U+17C7	ិកោះ /kah/ ‘island’	ិកោះ /kuəh/ ‘to knock’

These vowels (i.e. ីី ីូី ីុំ ីុំ ីុំ ីុំ) are composed of more than one part of vowels with similar look, but they do not combine the quality of the two parts (Kin 2007:45). For instance, ីី is made up of ី /ee/ and ី /əjii/, but the realization of ីី 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, ឃោះ (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
ឃោះ	U+17C8	សិតិោះ /so?taʔ/ ‘listening (to)’	សិតិោះ /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
ីុំ	U+17C2 U+17C8	ីុំ /heʔ/ ‘(startle)’	ីុំ /?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	ពី៖ /teh/ ‘to insult’	ិ៖ /cih/ ‘to ride’
ើ៖	U+17B9 U+17C7	ឯើ៖ /ckeh/ ‘to pluck out’	ិឯើ៖ /cpih/ ‘to be rough (of fabrics)’
ុ៖	U+17BA U+17C7	ូុ៖ /cræh/ ‘to be tightly packed’	ូិុ៖ /priih/ ‘to be spirited’
ុ៊៖	U+17C2 U+17C7	ូុ៊៖ /craeh/ ‘rust’	ូុ៊៖ /creeh/ ‘to chop away’
ុំៗ	U+17B6 U+17C6 U+1784	ុំៗ /taŋ/ ‘from/since’	ុំៗ /teap/ ‘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
ី	U+17A5	ីផ្លូ /?ət/ ‘brick’ ីខ្សែវ /?əj.ləw/ ‘now’	75
ើ	U+17A6	ើសាន /?əj.saan/ ‘north east’	5
ុ	U+17A7	ុកប្តា /?ok.naa/ ‘tycoon’ ុមាត /?u?.pa?.maa/ ‘to be comparable to’	242
ុ	U+17A9	ុរុ /?uu.ru?/ ‘chest (royalty)’	7

ខ័រ	U+17AA	ខ័រ /?əw/ ‘father’	13
ឃុំ	U+17AB	ឃុំ /rik/ ‘behavior’	38
ធម្ម	U+17AC	ធម្ម /rii/ ‘or’	2
ិន្ទុ	U+17AD	ិន្ទុ /ro.lik/ ‘to be alert’	3
ិន្ទុ	U+17AE	ិន្ទុ /lii/ ‘to hear’	8
ឯក	U+17AF	ឯក /?aek/ ‘one’	28
ិន្ទុ	U+17B0	ិន្ទុ /?aj.daa/ ‘LOC’	7
ិន្ទុ	U+17B1	ិន្ទុ or ិន្ទុ ^[9] /?aoj/ ‘to give’	55
ិន្ទុ	U+17B2	ិន្ទុ	0
ិន្ទុ	U+17B3	ិន្ទុ /kra.?aw/ ‘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 ឃុំ being used as a subscript in three headwords (i.e. សុប្បទ, សែវប្បទ, and ហ្សុទីយ), plus more than 190 instances of Sanskrit transliteration in the Chuon Nath dictionary. Here are some examples:

Khmer Contemporary Spelling	IPA	Sanskrit Transliteration	Gloss
ព្រឹស (ក បូ ី ស)	/krəh/	क्रीषि (क रु ई श)	‘small, dwarf’
គ្រឿង (ត បូ ី ៩ ០៦)	/krih/	क्रीह (त रु ई ह)	‘house’
ព្រិក្យ (ព បូ ី ក ឲ្យ)	/pri?.saa/	प्रिक्ष (प रु ई क िंग)	‘tree’
មុសា (ម ុ ស ៣)	/mu?.saa/	मुह (म रु ई श)	‘to be false’
និន្ទិ (ន ុ ន ិន្ទិ ិ)	/wut.thi?/	वृष्टि (न रु न िन्धि ិ)	‘prosperity’
សាធិ (ស ុ ត ិ)	/sa?.te?/	सृष्टि (स रु त ि)	‘consciousness’
សំស្រើត (ស ា ស ុ បូ ិ ត)	/saj.skrət/	संस्रात (स ा ស ុ बु त)	‘Sanskrit’

ស្រែក (ស ក រ ី ង ុ ៗ ន ា)	/srəj.kie/	ស្រែក (ត ុ ង ុ ៗ ន ា)	'love, lover'
ខ្លិច្ចិដ (ខ ក រ ូ ិ ជ ុ)	/?u?.krət/	ខ្លិច្ចិដ (ខ ត ុ ម ុ)	'seriously, criminally'

5.3. Diacritics (signs)

Eight diacritical signs are currently used in the present-day Khmer. They are Muusikatoan (‘’), Triisap (‘), Bantoc (៥), Robat (ᾷ), Kakabat (៥), Ahsda (ᢃ), Samyok Sannya (ᢃ) and Toandakhiat (ᢃ). 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. However, it does occur occasionally in the actual usage.* (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	ᾷ	U+17CC	83
Kakabat	៥	U+17CE	15
Ahsda	ᢃ	U+17CF	10
Samyok Sannya	ᢃ	U+17D0	384
Toandakhiat	ᢃ	U+17CD	248

Diacritic Name Diacritic Character Code Points Frequency in KKD

Muusikatoan	“	U+17C9	722
Triisap	~	U+17CA	215
Bantoc	៥	U+17CB	1801
Robat	ᾷ	U+17CC	83
Kakabat	៥	U+17CE	15
Ahsda	ᢃ	U+17CF	10
Samyok Sannya	ᢃ	U+17D0	384
Toandakhiat	ᢃ	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. ថ្មូនកណ្តល ‘Thmenh Kandol’ or សម្បាប់ពី ‘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 /joo/ + " > ឃ /jaa/

ឃ /joo/ + " > ឃឃ /jaaa/

ម /moo/ + " > ម /maa/

យ /joa/ + " > យឃ /jaaa/

ឃ /raa/ + " > ឃ /raa/

ឃ /waa/ + " > ឃ /waa/

ឃ /baa/ + " > ឃឃ /paa/

5.3.1.2. Triisap (ែ)

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

ឃ /baa/ + " > ឃឃ /baaa/

ស៊ា /saa/ + ^ > ស៊ែ /sɔɔ/

ហោ /haa/ + ^ > ហោែ /hɔɔ/

អោ /?aa/ + ^ > អោែ /?ɔɔ/

5.3.1.3. Special Case of ុ

ុ 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 ុ from the first to the second series.

ុ /baa/

ុ^ ី /bɔɔ/

ុ" ូ /paa/

Another special case for ុ is that even though its consonant quality is /b/, it would be realized as /p/ when it precedes a subscript in an initial cluster, i.e. ូ as in ពូ (ុ ូ ី) ‘salty’ is pronounced as /praj/ not /braj/. In cases like this, Muusikatoan is not required to be placed on ុ /b/.

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 glyph that looks like ូ (not to be confused with U+17BB) 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 ូ 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 ូ (U+17B7), ី (U+17B8), ឹ (U+17B9), ួ (U+17BA), ើ (U+17BE) and ិំ (U+17B6 U+17C6) are the Above Vowels that fall into this special case.

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

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

ីក	ូ ិ ក	/bic/	‘pen’
ីន	ូ ិ ន	/bin/	‘cooler box’
ីន	ូ ិ ន	/pən/	‘to be keen on something’

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

ីម ^[10]	ូ ZWNJ ិ ម	/?im/	‘to be breastfed’
ី	ូ ិ	/?um/	‘uncle/aunty’

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

ីម	ូ ិ ម	/him/	‘(sigh)’
----	-------	-------	----------

ីម ូ ZWNJ ិ 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 ុ and ុ, Um & Seng (2012:52) rules out that the Triisap has to change its glyph to look like ូ, i.e. ីូ, ីូកីូក។
- Samyok Sannya (ូ) 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.

ប៉ាដ	ប៉ា ៩ ឌ	/paŋ/	'bread'
ប៉ូត	ប៉ូ ៩ រ	/bœə/	'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).

ក្រឡាង /cra.jaap/ 'scatter in the way'

not ក្រឡាង

ក្រឡិន /pra.hœən/ 'to be insolent'

not ក្រឡិន nor ក្រលិន

- When the Triisap is followed by this vowel (ឯ), one has to change the second part of the vowel (i.e. ឯ) to ឬ (Kul ibid:32).

○ សុឯ should be written as សុឬ

○ ហុឯ should be written as ហុឬ

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 determined by the least sonorous consonant or if the cluster is 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 ស ស្រីប ឬ /srɪp/ ‘sound of a heavy, solid object falling, thud’
as opposed to ស្រីប /srəp/
- ស្អួយ ស ស្អួយ ឬ យ /s?uj/ ‘to be fat (of the belly)’
as opposed to ស្អួយ /s?oj/
- ម៉ែ ម ម៉ែ ឬ /mpaj/ ‘one day’
as opposed to ម៉ែ /mnej/

When it comes to a cluster with ឯ as the second member and a consonant shifter is needed, Triisap stays the same, while Muusikatoan becomes a glyph which resembles ឯ.

- ស្រីម ស ស្រីម ឬ ម /spəm/ ‘to be stuck at the mouth or entrance’
- ស្រីង ស ស្រីង ឬ ង /sb̥ig/ ‘(a made up word given as an example)’

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

5.3.2. Bantoc (ោ)

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 "ោ". 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
- ឈ 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/, -នា /g/, -ច /c/, -ចា /p/, -នា /t/, -ស /n/, -សា /l/, -សា /s/, or -ុ /b/. [...]

No.	Syllable	1st Series	Syllable	2nd Series
	Structure		Structure	
1	a		b	
	CF°	/aa/ is shorted to /a/	CF°	/oo/ is changed to /uə/
	$C\ddot{O}F^{\circ}$	Where	$C\ddot{O}F^{\circ}$	Where $F = -\tilde{n}/-\dot{m}/-\tilde{g}/-\tilde{l}/ -\tilde{n}/-s/-\tilde{v}/-\dot{m}/-\tilde{u}$ /oo/ is changed to /u/
	$C\ddot{q}CF^{\circ}$	$F = -\tilde{n}/-\dot{m}/-\tilde{g}/-\tilde{l}/ -\tilde{n}/-s/-\tilde{v}/-\dot{m}/-\tilde{u}$	$C\ddot{q}CF^{\circ}$	(else where)
2	$C\ddot{O}F^{\circ}$		$C\ddot{q}C\ddot{O}F^{\circ}$	$F = -\dot{m}/-\tilde{g}/-\tilde{l}/-\tilde{m}/ -\tilde{n}/-s/-\tilde{v}/-\dot{m}/-\tilde{u}$
	$C\ddot{O}\ddot{O}F^{\circ}$	/aa/ is shorten to /a/	$C\ddot{O}F^{\circ}$	/ie/ is changed to /ea/
		Where	$C\ddot{O}\ddot{O}F^{\circ}$	Where
	$C\ddot{q}C\ddot{O}F^{\circ}$	$F = -\tilde{n}/-\dot{m}/-\tilde{g}/-\tilde{l}/ -\tilde{n}/-s/-\tilde{v}/-\dot{m}/-\tilde{u}$	$C\ddot{q}C\ddot{O}F^{\circ}$	/ie/ is changed to /oa/
	$C\ddot{q}C\ddot{O}\ddot{O}F^{\circ}$		$C\ddot{q}C\ddot{O}\ddot{O}F^{\circ}$	Where
				$F = -\tilde{n}/-s/-\tilde{v}/-\dot{m}/-\tilde{u}/-\dot{m}/-\tilde{u}$

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

និនិ 'to be harden/frozen'

character sequence	និ	និ
character mapping	k	aa oo
series of character	1	1/2
series of syllable	1	
phonemic transcription	/kaak/	

ឃុំ 'to wash (hair)'

<i>character sequence</i>	ឃុំ	ឃុំ	ី
<i>character mapping</i>	k	aa ɔɔ	k
<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 (-ឃុំ /k/ or -ឃុំ /s/), and it is changed to /u/ when the Bantoc is placed on any other final consonants. Therefore, for instance, ឃុំ 'to wash (clothes)' is transcribed as /kuək/ (See the illustration below)

ឃុំ 'to wash (clothes)'

<i>character sequence</i>	ឃុំ	ឃុំ	ី
<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 ឃុំ /aa/ is shorten to /a/ when the Bantoc is placed on the final consonant. Thus, as illustrate below, ឃុំ 'coin' is transcribed as /kak/.

ឃុំ 'coin'

<i>character sequence</i>	ឃុំ	ឃុំ	ី
<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 -ធន /-k/, -ឃ /-g/, -ឃ /-c/, or -ឃ /-p/ as in ទាក់ 'to trap' which is transcribed phonemically as /teak/ (see the illustration of ទាក់ below); and it is changed to /oa/ when the Bantoc is placed on any other final consonant as in កោត់ '3S' phonetically transcribed as /koat/ (see the illustration of កោត់ below).

ទាក់ ‘to trap’

<i>character sequence</i>	៩	ឃ	ធន	ំ
<i>character mapping</i>	t	aa ie		k
<i>series of character</i>	2	1/2		
<i>series of syllable</i>	2			

phonemic transcription /teak/

កោត់ ‘3S’

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

phonemic transcription /koat/

Bantoc never occurs after a consonant cluster.

Bantoc is never used in Pali/Sanskrit loanwords.

* កែហុក /ko?.hak/ ‘to tell a lie’

not កែហុក * សុភិតិ /so?.kut/ ‘to die (royal register)’

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

* ពាំង /tag/ ‘to exhibit’

not ពាត់ * ពាំង /teay/ ‘both, all’

not ពាត់

5.3.3. Robat (៥)

According to the Khmer-English dictionary, Robat (៥) literally means ‘subscript ឃ (ឃួល)’; but it looks like the top part of the consonant (ឃ). It is a diacritical mark which represents an orthographic ‘៥’ in some Sanskrit loanwords. 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. no audible pronunciation). It also silences 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.

គ្រឿង /koə/ ‘pregnancy’

ធ្វី /h̥oa/ ‘Dharma’

មេគិតា /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. It is always written above the base consonant.

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
កែក !	/kook/	ក៥ ឲ្យ ៥ !	'sound made to catch others' attention'
កែកឃុ	/kook.kook/	ក៥ ឲ្យ ៥ ក៥ ឃុ	
ណា !	/naa/	ណា ឲ្យ !	
ណាដែ ^[11]	/nah/	ណា ៥ ឱះ !	'(affirmation)'
ឃ្លោះ !	/nah/	ឃ៥ ឲ្យ ៥ ឱះ !	
អីេ !	/?əəh/	អᅚ ឲ្យ ឱះ !	'(exclamation denoting surprise)'
អុក	/?ok/	អᅚ ៥ ិក	'to blame'
អែ !	/?ao/	អᅚ ឲោ ិ !	'(exclamation of distress)'
អិង !	/?ihih/	អᅚ ឲ៥ ZWNJ ិង ឲោ ិ ឱះ	'(exclamation of when having a burden)'
ចា	/c ^h aa/	ចᅚ ឲោ	'(a word used to refrain in a song)'

5.3.5. Ahsda (៥)

Ahsda (៥) is restrictedly used on five consonants: កᅚ ចᅚ សᅚ មᅚ បᅚ. They are pronounced with their respective inherent vowel. It is used to disambiguate the single character words

Khmer	IPA	Gloss
-------	-----	-------

កាំ	/kaa/	'also'
ដាំ	/daa/	'so, very'
នាំ	/naa/	'(usually used with ង់:)'
មាំ	/məa/	'to come (colloquial)'
ហាំ	/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: ទិ៍ ! /tiw/ 'go'. In the present day Khmer, Ahsda is never seen being 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
ជ៉ា	ជ ៉ា វ	/coa/	'rubber'
ប៉ាង	ប ៉ា ង	/pan/	'bread'
ត្រ៉ូយ	ត ្រ៉ូ ៉ូ យ	/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.

ខ៉ែវ ស៊ែវ ពោវ sword (royalty) ពោវ poa Pear ethnic group

លេក leak artificial dye

នឹង nej meaning or sense

ប៉ែង ប៉ែង pag bread

ទីនឹង tii.neap seat (royalty)

Samyok Sannya (សាំណុយ) 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
មោស្វ័រ	ម᠁ ស ត ូ ៩	/maa.stəə/	'master'
កុំព្យូទ័រ	ក ា ព ូ ទ ូ ៩	/kom.pjuu.təə/	'computer'
គុងទៅ	ក ុ ង ទ ូ ៩	/koŋ.təə/	'counter'
គុល់	គ ុ ល ៩	/kuu.ləə/	'color'

5.3.7. Toandakhiat (ទោនដកិតាទិត្យ)

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 is 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 Toandakhiat placement:

- after a subscript and vowel ៦៖ :

Khmer	Character	IPA	Gloss
កិត្តិ	ក ិ ន ូ ន ិ ៩	/keə/	'reputation'
កែត្តិ	ក េ ន ូ ន ិ ៩		

ឃុំស្ទឹក	ប៊ូ ឯ ស ូ ិ ី ៩	/poh/	'post office'
ប្រភិទ្ធិ	ប ប្រ ស ិ ទ ូ ិ ី ៩	/pra.sət/	'a place name'
រាមកិន្យា	រ ច ម ក ិ រ ូ ិ ី ៩	/riem.kee/	'Ramayanak'
ល្អឥត្តិថ្លែង	ស ូ ស ូ ិ ី ៩	/suah/	'glory'

- after a consonant and vowel ី :

Khmer	Character	IPA	Gloss
កាស្វោបី	ក ឆ្វ ោ ប ិ ៩	/ka?laop/	'a kind of basket'
តេនពី	ត េ ន ព ិ ៩	/taen/	'tent'
ទិបេតី	ទ ិ ប េ ព ិ ៩	/tii.bee/	'Tibet'
នាំអាមិន	ន ា ំ អ ា ម ិ ៩	/noam.?aat/	'to initiate'
ពោធិ៍	ព ោ ៗ ធម ិ ៩	/poo/	'banyan'
ពោធិ៍ជំ	ព ោ ៗ ធម ិ ៩ ធម ំ	/poo.t ^h um/	'a kind of plant'
ពោធិ៍ឆ្លែ	ព ោ ៗ ធម ិ ៩ ធម ូ ឱ	/poo.tlee/	'a kind of plant'
ពោធិ៍បាយ	ព ោ ៗ ធម ិ ៩ ប ១ យ	/poo.ñaaj/	'a kind of banyan tree'
ពោធិ៍សាត់	ព ោ ៗ ធម ិ ៩ ស ៣ ត ០	/poo.sat/	'Pursat province'
ព្យាជី	ព ូ ឲ ៣ ធម ិ ៩	/pjie/	'leprosy'
ប្រហស្ថគី	ប ្រ ហ ស ូ ុ គ ិ ៩	/prə.hoah/	'Thursday'
វិទអាគិត	វ ិ ទ អ ១ ៩ ិ ៩	/wiet.?aat/	'to try to gain power'
សកវាគិត	ស ក វ ិ ៣ ិ ៩	/sak.kak.waa/	'a kind of song'
ស្វាជី	ស ូ ឲ ៣ ធម ិ ៩	/smaat/	'meditate'
ស្វោគពោធិ៍	ស ូ ុ ី ក ៣ ៤ ធម ិ ៩	/slək.poo/	'banyan leaf'

- after vowel ឃ :

Khmer	Character	IPA	Gloss
រី	ឃ ឃ ឃ ី	/rie/	'name of a giant'
ចាត់	ច ឃ ឃ ី	/caa/	'small tubes made of gold'

It should be rendered right on the consonant. It has never 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 (៥)
- Atthacan (៥)
- Viriam (៥)
- Avakrahasanya (ᬸ)
- Obsolete consonants (៥ and ስ)
- Obsolete independent vowel (៥)

Here is a list of them with their code point.

Code Point Orthographic Character Name

U+17D3	៥	KHMER SIGN BATHAMASAT
U+17DD	៥	KHMER SIGN ATTHACAN
U+17D1	៥	KHMER SIGN VIRIAM

U+17DC	៩	KHMER SIGN AVAKRAHASANYA
U+179D	ិ	KHMER LETTER SHA
U+179E	ិ៥	KHMER LETTER SSO
U+17A8	ិ៥៥	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
ព្រៀកមាន់	Phnaek Muan	០	- mark the beginning of a poem
គោមូត្រា	Koomuut	៦៥	- mark the end of a poem
ល្បែង ូន	Lbah or Khan	១	- placed at the end of a sentence or a section of a poem
ឡណុល			- used in lunar dates
ល្បែងចំប់ ូន	Lbah Chob or Lbah	១៥	- 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 ឱ៥ with the same name.
បរិយោសាន់	Pareyaosan		
ជកូនា	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.

បែយកាល:	Beyyal or Lak លេខេ - used at the end of a list to denote that the list goes on
ឬ ឡាក់	or ឬ
	ឬ ។
ចំណុចពី	Camnuc Pii
គូស	Kuuuh or
ឬ ដំពី	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 Comnuc Pii Kuuuh for Comnuc Pii Kuuuh 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

នគេសញ្ញា	hyphen	-	<ul style="list-style-type: none"> - 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
នៃក្រចក	brackets	()	<ul style="list-style-type: none"> - denote extra information
ចុចមួយ	point	.	<ul style="list-style-type: none"> - place after abbreviation
ចុចពី	colon	:	<ul style="list-style-type: none"> - used in place of Chomnuc Pii Kuuuh - denote "etc"
ចំណុចរាយ / ពាតិភី	ellipsis	...	<ul style="list-style-type: none"> - denote unfinished phrase - denote a section of quote that is out of focus (usually place in side square brackets)

ក្រែង	comma	,	- separate items with spaces in between In Chuon Nath dictionary, it is used
ចុចេក្រែង	semicolon	;	- to denote the list has not ended yet - placed before បើន់នា ‘but’ - separate two or more related sentences
សញ្ញាស្ថរ	question mark	?	- end interrogative sentence
សញ្ញាមនុសា	exclamation mark	!	- end an interjection
អណ្តូវ	double quotes	“”	- mark a quote or direct quote
	Guillemets	«»	- encircle a title or name of a book or news article - focus point
យុប or តីអូប	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
បន្ទាត់ថ្វិក	forward slash	/	- separate numbers - used in place of ឬ ‘or’
ផ្លូវ	asterisk	*	- denote errors in a phrase
វត្ថុយុប	curly brackets	{}	- used in syntax to wrap around certain things
សញ្ញាស្ថី	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
សញ្ញាតុចជាង	smaller than; from	<	- denote less than - denote what is on its left is derived from what is on its right
សញ្ញាដូល	opposite	≠	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	១	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 if it ever gets combined 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: ខ្សែចុះ “waxing moon” and ខ្សែចុក “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, ៩១ 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 if above the sign, it denotes “waxing moon”; underneath, “waning moon”). For instance, ៩៩០៩១៩២ is read as “Sunday, 10th day of the waning moon,

5.8.1. Phase of the Moon

5.8.1.1. ខ្សែចុះ “Waxing Moon”

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

- ០ U+19E0 /pa?.t^haa.mie.saat/ ‘the first Ashadha (eighth month of the lunar calendar)’
- ១ U+19E1 /muəj.koət/ ‘the first day’
- ២ U+19E2 /pii.koət/ ‘the second day’
- ៣ U+19E3 /bəj.koət/ ‘the third day’
- ៤ U+19E4 /bəuən.koət/ ‘the fourth day’
- ៥ U+19E5 /pram.koət/ ‘the fifth day’
- ៦ U+19E6 /pram.muəj.koət/ ‘the sixth day’
- ៧ U+19E7 /pram.pii.koət/ ‘the seventh day’
- ៨ U+19E8 /pram.bəj.koət/ ‘the eighth day’
- ៩ U+19E9 /pram.bəuən.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.bəuən.koət/ ‘the fourteenth day’
- ៩៥ U+19EF /dəp.pram.koət/ ‘the fifteenth day’

5.8.1.2. និលុច “Waning Moon”

Waning moon, a.k.a. និលុច /rə.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).

- ៩៦ U+19F0 /tu?.ti?.jeak.saat/ ‘the second Ashadha during the Adhikameas leap year’
- ៩៧ U+19F1 /muaj.rooc/ ‘the first day’
- ៩៨ U+19F2 /pii.rooc/ ‘the second day’
- ៩៩ U+19F3 /bəj.rooc/ ‘the third day’
- ៩៩១ U+19F4 /buən.rooc/ ‘the fourth day’
- ៩៩២ U+19F5 /pram.rooc/ ‘the fifth day’
- ៩៩៣ U+19F6 /pram.muəj.rooc/ ‘the sixth day’
- ៩៩៤ U+19F7 /pram.pii.rooc/ ‘the seventh day’
- ៩៩៥ U+19F8 /pram.bəj.rooc/ ‘the eighth day’
- ៩៩៦ U+19F9 /pram.buən.rooc/ ‘the ninth day’
- ៩៩៧ U+19FA /dap.rooc/ ‘the tenth day’
- ៩៩៨ U+19FB /dap.muəj.rooc/ ‘the eleventh day’
- ៩៩៩ U+19FC /dap.pii.rooc/ ‘the twelfth day’
- ៩៩៩០ U+19FD /dap.bəj.rooc/ ‘the thirteenth day’
- ៩៩៩១ U+19FE /dap.buən.rooc/ ‘the fourteenth day’
- ៩៩៩២ U+19FF /dap.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
ម៉ិត្តសីរី	Mārgaśīrṣa (मार्गशीर्ष)	Mikeaser	៩
បុស្សា	Pauṣa (पौष)	Boss	៩២

មោង	Māgha (មាម)	Meakh	៩
ផែលុខ	Phālguna (फाल्गुन)	Phalkun	៥
ចៅត្រា	Caitra (चैत्र)	Chetr	៥
វិសាខ	Vaiśākha (वैशाख)	Pisakh	៦
យេស្តា	Jyeṣṭha (ज्येष्ठ)	Chesth	៧
អាសាតា	Aśadha (आषाढ)	Asath	៨
ស្រីរាយ	Śrāvaka (श्रावण)	Srapn	៩
បាប្រុបដ	Bhadrapada (भाद्रपद) Phutrabot	90	
អសុធន	Āśvina (अश्विन)	Assoch	99
កក្កិក	Kārtika (कार्तिक)	Kakdek	៩៩

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 may stay above the vowel (i.e. អិះ), but if the diacritic is a Triisap, the vowel should be placed above the Triisap (i.e. អិះិះ). 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. សិះកិ).

(៩)

(១) កិច្ចជីជញ្ជូបន្ត

(២)

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 a part of the base consonant to accommodate the subscript allowing the Below vowel to stay above its default position (see the images below).



កិម្ម

(a)



កិម្ម

(b)



A consonant consists of two parts: the body and the head or usually called hair. For instance, ឃ 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 ឃ stands on its own, the wavy line presides on the body fully; however, when a vowel^[14] that looks like ោ 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. The wavy line of the consonant and the vowel have to be merged (see the highlighted part).

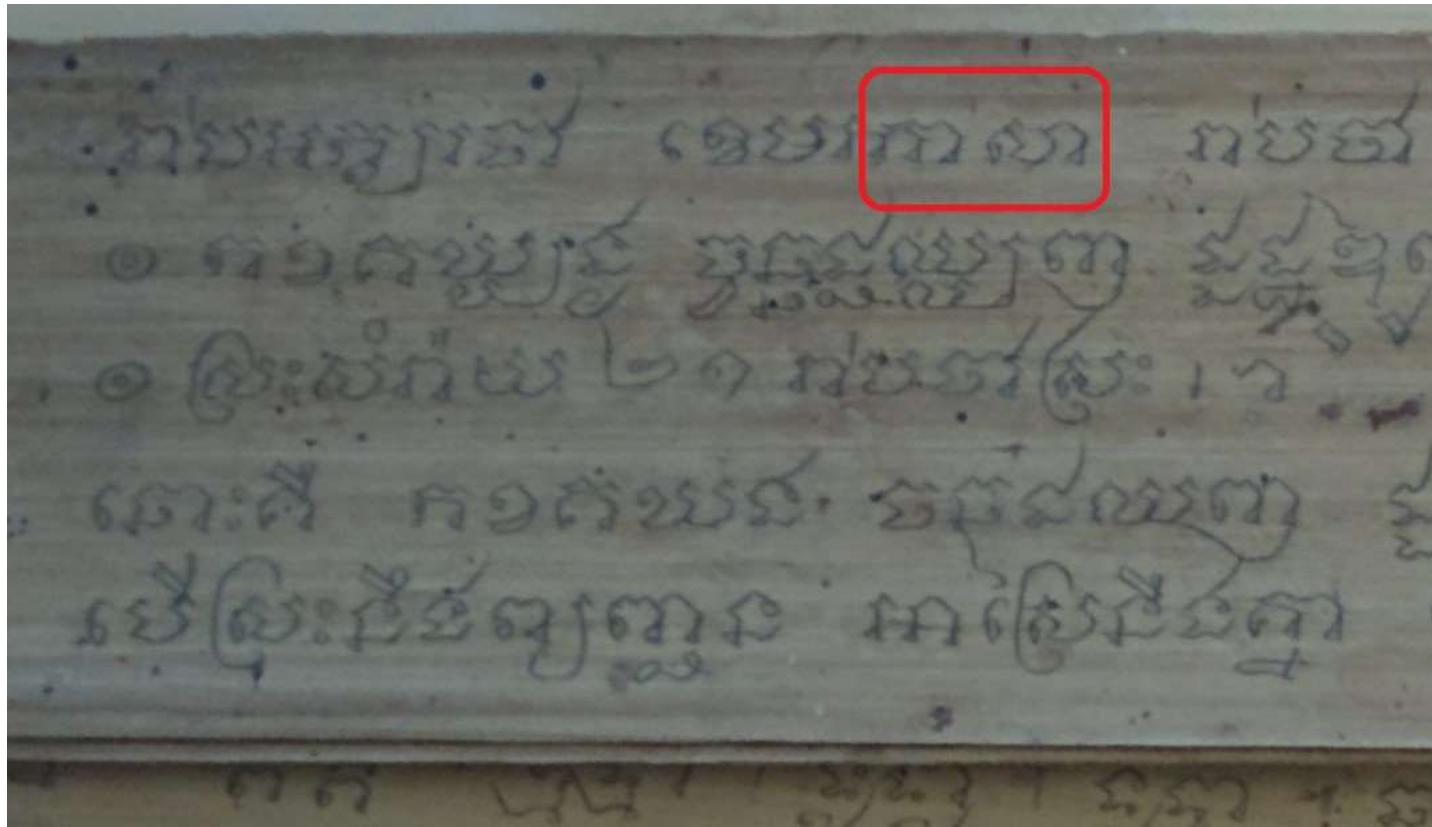
(a)



(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 នាមី (ន 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 ី.



NOTE: When ឃ is combined with ឃ ឃុ or ឃុំ, the shape of the consonant change to ឃុ ឃុិ and ឃុំ 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 ុ attached to it.



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



6.3. ឃ and ែ Combination

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

6.4. ឃ, ិ, ី, ុ, ុែ and ឃ, ិ, ី, ុ, ុែ

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

ក + ០ > ក

ក + ១ > ក

ក + ២ > ក

ក + ៣ > ក

ន + ០ > ន

ន + ១ > ន

ន + ២ > ន

ន + ៣ > ន

ក + ០ > ក

ក + ២ > ក

ក + ៣ > ក

ក + ៤ > ក

ក + ០ > ក

២

ក + ០ > ក

៣

ក + ០ > ក

៤

ក + ០ > ក

ិ + ០ > ិ

២

ិ + ០ > ិ

២

ិ + ០ > ិ

២

ិ + ០ > ិ

7. Encoding Order

7.1. Overview

There are discrepancies in character sequences posited by Unicode Standard (2018:646), Microsoft (2018) and the Open Forum (2004:7-14).

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

Unicode Standard (2019:654)

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

A zero width non-joiner or a zero width joiner can be placed before a vowel if needed.

Microsoft (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: Microsoft states ZWNJ is to be placed in between the RegShift and the AbvV to prevent the RegShift from changing its shape. However, its example has ZWNJ before the RegShift.

Open Forum

Open Forum of Cambodia (currently known as Open Institute) authored a document on “How to type Khmer Unicode” (2004) among other papers related to Khmer encoding issues as a part of their KhmerOS project whose aims were to localize and enable Khmer language in Linux and more. They have also created KhmerOS font family that has been adopted and used widely in Cambodia.

The Khmer character ordering was devised as:

- a. Consonant + Robat {+ Vowel} {+ Sign}

- b. 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+17C5]
- Sign
 - Above sign – [U+17C6, 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]

The following characters are implemented differently.

Unicode Standard	Microsoft	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, Microsoft breaks vowels into four subcategories based on where they are placed on the last character:

- (1) Pre-base Vowels (PreV): ଠ [U+17C1], ଢ [U+17C2], ତ୍ତ [U+17C3],
 - (2) Below-base Vowel (BlwV): ଙ [U+17BB], ଙ୍ଗ [U+17BC], ଙ୍କ [U+17BD],
 - (3) Above-base Vowel (AbvV): ଟ୍ଟ [U+17B7], ପ୍ର୍ତ୍ତ [U+17B8], ପ୍ର୍ତ୍ତିତ୍ତ [U+17B9], ପ୍ର୍ତ୍ତିତ୍ତା [U+17BA], ପ୍ର୍ତ୍ତିତ୍ତିତ୍ତ [U+17BE], and
 - (4) Post-base Vowels (PstV): ଇତ୍ତ [U+17C4], ଇତ୍ତିତ୍ତ [U+17C5], ଇତ୍ତିତ୍ତିତ୍ତ [U+17C0], ଇତ୍ତିତ୍ତିତ୍ତିତ୍ତ [U+17BF], $\text{ଇତ୍ତିତ୍ତିତ୍ତିତ୍ତିତ୍ତ}$ [U+17B6].

Nikahit (៥ U+17C6) and Reahmuk (៦ U+17C7) are used as vowels. These are the only two that can be attached to another vowel to create vowel combination. Here are the common ones: ៥៥ (U+17BB U+17C6), ៥ᦚ (U+17B6 U+17C6), ៥ᦚᦚ (U+17BB U+17C7), ៥ᦚᦚᦚ (U+17C1 U+17C7) and ៥ᦚᦚᦚᦚ (U+17C4 U+17C7). Khmer linguists usually consider them as one vowel unit. Unicode Standard, however, considers these as a combination of vowel and diacritics.

7.2. New Proposal

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]
- PstS: post-sign [U+17C8 or U+17C7]
- Coeng: the Coeng [U+17D2]
- RegShift: Register Shifter, Muusikatoan [U+17C9] or Triisap [U+17CA]
- AbvS: [U+17C6, U+17CB, U+17CD, U+17CE, U+17CF, U+17D0, U+17D1, U+17D3, U+17DD]
- Z: zero width non-joiner [U+200C] or zero width joiner [U+200D]
 - – Indicates 0 to 2 occurrences
 - [] – Indicates 0 or 1 occurrence
 - | – Exclusive OR
 - + – Cumulative AND
 - () – Indicates grouping

Partial clusters defined:

Base: BaseCommon | BaseOther

VowelGroup: AbvV | PreV | BlwV | PstV

CoengGroup: Coeng BaseCommon

TailGroup: [AbvS PstS AbvS] | CoengGroup

General cluster defined:

Base {[CoengGroup} | [Robat] [RegShift] [Z] [VowelGroup] [TailGroup]

Independent Cluster: All Khmer characters not classified above are considered clusters by themselves. Abvs occurs twice, one before and one after the PstS. This is needed for Jarai and Tampuan languages. The combinations of ៥ (U+17C6) & ៥ (U+17D0) and ៦ (U+17C7) and ៥ (U+17D0) are required in these languages.

Restrictions: In the present day Khmer, (1) digits are not allowed as bases and (2) Coeng Ro is not allowed to be placed before another subscript when there are two subscripts in a row and one of which is a Coeng Ro.

7.3. Character Distribution in an Orthographic Syllable

Reahmuk (៩), Nikahit (៥), Yuukaleapintu (៧), Toandakhiat (៥), Ahsda (៥) and Bantoc (៥) always occur 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 (៥ U+17D2), RegShift, Robat, Vowel, Samyok Sannya or Kakabat can never be followed by itself (i.e. duplicated).

The table below illustrates which character can be placed after another in the same orthographic syllable based on the proposed character order above. "v" denotes that characters in the top row (1) can be followed by those in the left most column (2) in the same syllable (as used in the Chuon Nath Dictionary published in 1967 which is taken as a blue print for the contemporary Khmer), "h" 1 can be followed by 2 (as used in the Cambodian-French Dictionary published in 1902) and "x" not applicable.

- "v" stands for "a valid contemporary use".
- "h" historic use.
- "x" not applicable.

1	Cons	IndV	Coeng	RegShift	Robat	V	Samyok	Sannya	Kakabat
---	------	------	-------	----------	-------	---	--------	--------	---------

2

Cons	x	x	v	x	x	x	x	x	x
IndV	x	x	h	x	x	x	x	x	x
Coeng	v	h	x	x	x	h[15]	x		x
RegShift	v	h	x	x	x	x	x	x	x
Robat	v	x	x	x	x	x	x	x	x
V	v	x	x	v	v	h	x		x
Samyok Sannya	v	x	x	v	x	x	x		x
Kakabat	v	x	x	v	x	v	x		x
Reahmuk	v	x	x	v	v	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	h	x		x

It is important to note that an orthographic syllable mentioned here is different from that used in linguistic context. An orthographic syllable start with a base character which can be a consonant (or an independent vowel), therefore unlike a syllable in linguistic sense, there is no final consonant in an orthographic syllable. For instance, កាត់ (ក ា ត ់) /kat/ 'to cut' has one linguistic syllable, but two orthographic syllables--ក (ក ា) and ត់ (ត ់).

Khmer spelling has evolved since the publication of the official dictionary. Six separate cases have been identified:

- An independent vowel can be followed by a Coeng, i.e. តីរ៉ាសិន (តី ឬ ស ា ន) 'north east'.
- An independent vowel can be followed by a register shifter, i.e. ខកញ្ចា (ខ ក ញ ៉ា) 'tycoon'.
- A Coeng can be followed by an independent vowel, i.e. បុទ្ធយ (ប ុ ទ ុ ទ ួ យ) 'heart'. This entry exists in the Khmer official dictionary, but it has rarely been used nowadays.
- A vowel can be followed by a subscript (i.e. Coeng + Consonant), i.e. ពីំ (ឬ" ព ី ំ) 'day'.
- A vowel can be followed by another vowel, i.e. ដីវិ (ជ ី វ ិ វ) 'to walk'.
- A vowel can be followed by a Bantoc, i.e. បុច (ប ុ ច ុ ច) 'to pass (sth to sb)'.

A noticeable placement of a register shifter is that the Muusikatoan can be placed at the bottom of base consonant as in the word for "to eat" (see the image below).



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 problems for the end users and developers.

- **Confusion:** 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. Consequently, users may be vulnerable of spoofing attack and/or unable to find what they are looking for.
- **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.

Consistency in text input can be improved by (1) showing the dotted circle (U+25CC) when users type out of a certain sets of character order rules or (2) automatically re-ordering characters at the input level. (2) cannot be done unless a set of character order rules is in place.

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.

- "Valid sequence" refers to when the character sequence abide by the New Proposal in Section 7.2, and "Invalid sequence" is the opposite.

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 in contemporary Khmer.

- Subscript + Vowel ៩ ឃ ុ > ឃុ៩ ‘Khmer’
- Vowel + Subscript ឃ ៩ ុ > ឃុ៩ **invalid sequence**, but it was used before the publication of the Khmer-Khmer dictionary in 1967 (See Scheuren (2010:42-43))

This issue is complicated and a solution is probably to distinguish between the old Khmer (prior to the Khmer-Khmer Dictionary publication) and the contemporary Khmer (after the dictionary was published). Unicode Standard allows both for the time being.

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] ៩. Placing either subscript before another would have the same visual output.

- Subscript + [U+17D2 U+179A] ស ុ ូ ី > ស្រី ‘woman’
- [U+17D2 U+179A] + Subscript ស ូ ុ ី > ស្រី **invalid sequence**

The dotted circle should be presented in the invalid sequence. By having it rendered exactly the same as the valid sequence only encourages more inconsistency as the users are not aware of the differences.

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 ូ” ៣ ធម > ម៉ែន ‘one way’
- Consonant shifter + Subscript ម” ូ ៣ ធម > ម៉ែន **invalid sequence**

On all popular web browsers (i.e. Edge, FireFox, Safari and Chrome), these two sequences are allowed and rendered almost identical. According to the proposal above, the Consonant Shifter should stand after the Subscript. When placed otherwise, the dotted circle should be shown.

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 ស” ី > សិ ‘to eat (vulgar)’
- Above vowel + Below vowel ស ី ោ > សិ **invalid sequence**
- Below vowel + Above vowel ស ោ ី > សិ **invalid sequence**
- Consonant Shifter (”) + Above vowel ស” ី > សិ **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 (ំ) it cannot be used with the 1st series consonants.

Firefox 72 allows the invalid sequences and renders them exactly the same as the valid one--which is not ideal.

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. Rendering engines should show dotted circles when invalid vowel combination is concatenated.

- [U+17BB] [U+17C6] ດ ດ ເ > ດ ‘don’t’
- [U+17C6] [U+17BB] ດ ເ ດ > ດ **invalid sequence**
- [U+17B6] [U+17C6] ຕ ຕ ເ > ຕ ‘to wait’
- [U+17C6] [U+17B6] ຕ ເ ຕ > ຕ **invalid sequence**

7.4.6. One Unicode Code Point

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

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

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

- [U+17BE] ុ ី > ុី ‘if’
- [U+17C1] [U+17B8] ុ ៩ ី > ុី **invalid sequence**
- [U+17B8] [U+17C1] ុ ី ៩ > ុ៩ី **invalid sequence**

A takeaway here is that two vowels should never be considered as a valid sequence; therefore, the rendering engine should not render them identically. Dotted circles are needed when invalid sequences are used.

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 ណ ល ិ ន ិ > កណលិនិ ‘Kandal province’
- [U+17D2] [U+178F] ក ណ ល ិ ន ិ > កណលិនិ **incorrect sequence**

This is a tricky issue which requires re-consideration of whether or not to equate these two identical characters.

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 (ិ), 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 and certain consonants (see section 6.4. ិ, ី, ឹ, ុ, ុី and ុ, ី, ុី, ុី). .

8. Text Processing

8.1. Keyboard

Below is the standard key arrangement approved by the National ICT Development Authority (NiDA). Their 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: « ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉ ၀ ၁ ၂ ၃
 - second row: ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉ ၀ ၁ ၂ ၃
 - third row: ၀ ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉ ၀ ၁ ၂ ၃
 - fourth row: ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉ ၀ ၁ ၂ ၃

(2) the Shift layer

- first row: » ! ຈ % ຕ ດ ດ () ~ = ໂ
 - second row: ໂ ດ ດ ດ ດ ດ ດ ດ ດ ດ ດ ດ ດ ດ
 - third row: ດ ດ ດ ດ ດ ດ ດ ດ ດ ດ ດ ດ ດ
 - fourth row: ໂ ພ ດ ດ ດ ດ ດ ດ ດ

(3) the AltGr layer

- first row: zwj zwnj @ ¯ \$ € ¤ ©™ * × ö \
- second row: ດ ປ ພ ຂ ປ ຂ ຂ
- third row: : ; :
- 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

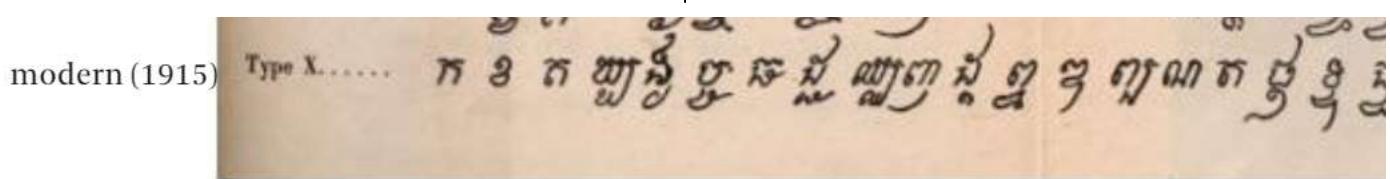
The Royal Academy of Cambodia was approached and asked for advice regarding the sorting in Khmer. It turns out that there is no documentation related to this topic that could help. The Choun Nath dictionary seems to deploy two ways of sorting: (1) alphabetical order of characters and (2) alphabetical 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 ស which sounds similar to the independent vowel. A paper done by PAN localization[\[16\]](#)(nd) 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 pronunciation, not spelling. For instance, ឃនាំ [U+1794 U+1784 U+17CB] is listed before បក្សិ [U+1794 U+1780 U+178F U+17B7] even though the second character of the second word (i.e. ន [U+1780]) appears before that of the first word (i.e. ង [U+1784]) in the alphabet chart.

More discussion is needed to come up with a solid sorting rule.

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).

	VALUE EN SANSKRIT.	ka kha ga gha na ña tha ja jha ña ja pha ña da ñha na ta tha ña ñ	VALUE EN KHMER MODERNE.
6th century	Type I.....	କ ଖ ଗ ଘ ଙ ଠ ଜ ଝ ଙ ପ ଫ ଙ ଢ ନ ତ ଥ ଙ	କ ଖ ଗ ଘ ଙ ଠ ଜ ଝ ଙ ପ ଫ ଙ ଢ ନ ତ ଥ ଙ
7th century	Type II.....	କେ ଦ ଦକ ମ	କେ ଦ ଦକ ମ
667	Type III.....	କୁ ତ ଦ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ	କୁ ତ ଦ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ
970	Type IV.....	କୁ ତ ଦ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ	କୁ ତ ଦ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ
1002	Type V.....	କୁ ତ ଦ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ	କୁ ତ ଦ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ
1066	Type VI.....	କୁ ତ ଦ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ	କୁ ତ ଦ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ ପିଲ କ ଷ୍ଟ
12th/13th century	Type VII.....	କ ଖ ଗ ଘ ଙ ଠ ଜ ଝ ଙ	କ ଖ ଗ ଘ ଙ ଠ ଜ ଝ ଙ
13th century	Type VIII....	କ ଖ ଗ ଘ ଙ ଠ ଜ ଝ ଙ	କ ଖ ଗ ଘ ଙ ଠ ଜ ଝ ଙ
1702	Type IX.....	କ ଖ ଗ ଘ ଙ ଠ ଜ ଝ ଙ	କ ଖ ଗ ଘ ଙ ଠ ଜ ଝ ଙ



8.3.1. Font Style

According to Scheuren (ibid:9-10) there are three main types of Khmer font styles: (a) មូល Mool ‘lit. round’, (b) ឯក្រៈជាគ 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.

Mool Style in Khmer OS Moul font face

នាល់ខ្លួន

Chrieng Style in Khmer OS Metalchrieng font face

កសារខ្មែរ

កសារខ្មែរ

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 character. Similarly, subscripts can be rendered to the left, right or below the base character. 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 character are shown in the table below.

left	above	below	right	around
ឃ (U+17C1)	ឃ (U+17B7)	ឃ (U+17BB)	ឃ (U+17B6)	ឃ (U+17BE)
ឃ (U+17C2)	ឃ (U+17B8)	ឃ (U+17BC)	ឃ (U+17C7)	ឃ (U+17BF)
ឃ (U+17C3)	ឃ (U+17B9)	ឃ (U+17BD)		ឃ (U+17C0)

❖ (U+17BA) ❖ (U+17C8) ❖ (U+17C4)
❖ (U+17C6) ❖ (U+17B6 U+17C6) ❖ (U+17C5)
❖ (U+17BB U+17C6)
❖ (U+17BB U+17C7)
❖ (U+17C1 U+17C7)
❖ (U+17C4 U+17C7)

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

8.4. Diacritic Position

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. However, a general character order can be posited as:

B {S} [C] [V] {O}

where:

- B – A base consonant character
- S – A subscript consonant (i.e. Coeng + Base)
- C – A consonant shifter
- V – A vowel sign
- O – Any other sign
- – Indicates 0 to 2 occurrences
- – Indicates 0 to 1 occurrence

This section describes 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 have a list of characters used in each ethnic minority language. The tables below show the orthographic characters 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. ឃិំ /ç/ and ុី /h/ always occurs at the final position.

ក	ខ	គ	ង	ច	ន	ជ	ញ	ជ	ត
/k/	/kʰ/	/g/	/ŋ/	/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

ច	ឆ	ស	ប	ផ	ព	ម	យ	ឃិំ	ី
/tʰ/	/d/	/n/	/b/	/pʰ/	/p/	/m/	/j/	/ç/	/r/

U+1799

U+1790 U+1791 U+1793 U+1794 U+1795 U+1796 U+1798 U+1799 U+17D2 U+179A

U+179F

ល	វ	ធម៌	ិំ	ហ	ន	អិំ	អី
/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

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

ꝣ	Ꝥ	ꝥ	Ꝧ	ꝧ
/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

ꝑ	Ꝓ	ꝓ	Ꝕ	ꝕ	ꝗ	Ꝙ	ꝙ	Ꝛ	ꝛ	Ꝝ	ꝝ
/a/	[i]	/i/	[ɛ]	/i/	/ɯ/	/u/	/ə/	/ɪ/	/e/		
U+17B6	U+17B7	U+17B8	U+17B9	U+17BA	U+17BB	U+17BC	U+17BE	U+17C0	U+17C1		
ꝑ	Ꝓ	ꝓ	Ꝕ	ꝕ							
/ɛ/	/aj/	/o/	/aw/	/ɔ/							
U+17C2	U+17C3	U+17C4	U+17C5	U+17DD							

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

Ꝓ /aj/ [17C3] and ꝓ /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

ꝑ:	or	Ꝓ:	ꝓ:	Ꝕ:	ꝕ:	ꝗ:	Ꝙ:	ꝙ:	Ꝛ:	Ꝝ:	ꝝ:
----	----	----	----	----	----	----	----	----	----	----	----

/ã/	/ɔ/	[ɛ]	/ɛ/	[ð]	/ʒ/	/ɸ/
U+17B6					U+17DD	
U+17C8				U+17C8		
or	U+17BA	U+17C1	U+17C2	U+17C4	or	U+17BF
U+17B6	U+17C8	U+17C8	U+17C8	U+17C8	U+17DD	
U+17CB					U+17CB	

ពីរក្នុង /ɸ/ [U+17BF] ត្រូវបានប្រើបាយនៅពេលសម្រាប់ពាណិជ្ជកម្មខ្មែរ។

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. ៗ ៖ ៖ ៖ ៖) and others are borrowed from English or French.

9.1.6. How is the writing system different from Khmer?

- The use of two Khmer obsolete characters:
 - ប /b/ [U+179E] as in បរៀ (ប រី ឯ) /bar/ ‘two’
 - ក /ɔ/ [U+17DD] as in កិន (ក ិ ន) /kɔn/ ‘child by birth; son’
- ច [U+179B] can be placed on consonant ស [U+179A] as in ពិច្ច (ព ិ ច ច) /pɔr/ ‘to burn’.
- ឃ: [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 ឃ: [U+17C8] is usually used to attach to the consonant without any vowel intervening it.
- ឃ [U+17DD] behaves like a vowel, and it can be followed by a final consonant with or without ច as in ពិង (ព ិ ង) /gɔŋ/ ‘k.o. gong’ and ពិង (ព ិ ង ច) /gɔŋ/ ‘to roast’.
- Sequences uniquely used in Bungong (i.e. ឃូង, ឃូង, ឃូងូល).
- 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

ក	ខ	គ	ម	យ	ឃ	ច	ជ	ឈ	ឃ	ឃ	ឃ
/k/	/kh/	/k/	/kh/	/y/	/ŋ/	/c/	/c/	/ɲ/	/?d/	/?d/	/?d/

U+1780 U+1781 U+1782 U+1783 U+1784 U+1785 U+1787 U+1789 U+178A U+178C

ឃូង ឃូង

/n/ /t/ /th/ /t/ /th/ /n/ /?b/ /ph/ /p/ /ph/

U+178E U+178F U+1790 U+1791 U+1792 U+1793 U+1794 U+1795 U+1796 U+1797

ឃ ឃុ ឃ ឃុ ឃ ឃ ឃ ឃុ ឃ ឃុ ឃ ឃុ

/m/ /j/ /r/ /l/ /w/ /ç/ /h/ /l/ /?/ /?j/

U+17A2

U+1798 U+1799 U+179A U+179B U+179C U+179F U+17A0 U+17A1 U+17A2 U+17D2

U+1799

ឃុ ឃុ

/cj/ /hp/

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 /j/ for ឃុ [U+1789] and /d/ instead of /?d/ for ឃ [U+178A] and ឃ [U+178C].

9.2.2. Tampuan Subscripts

៥ ៥ ៥ ៥ ៥ ៥៥ ៥៥ ៥ ៥

/?/ /j/ /j/ /n/ /m/ /j/ /r/ /l/ /w/

U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2

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

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

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

9.2.3. Tampuan Vowels

inherent	ঠ [17]	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ
/ɒ:/	/a:/	/ɛ/	/æ/	/ə/	/ə̄/	/o/	/ɔ:ɔ:ou/	/uə/	/aə/	
/ɔ:/	N/A	/i/	/ɪ/	/i/	/ī/	/u/	/u:/	/uə/	/ə:/	
U+17B6 U+17B7 U+17B8 U+17B9 U+17BA U+17BB U+17BC U+17BD U+17BE										
ଠୋଁ	ଠୋଁ	ଠୋ	ଠୋ	ଠୋ	ଠୋ	ଠୋ	ଠୋ	ଠୋ	ଠୋ	ଠୋ
/iə/	/iə/	/e:/	N/A	/ai/	/ao/	N/A	/om/	/om/	/am a/	
/jə/	/jə/	/e:ɛ:/	/e:/	N/A	/ɔ:/	/i:w/	/u:m/	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										
ଠୋ:	ଠୋ:	ଠୋ:	ଠୋ:	ଠୋ:	ଠୋ:	ଠୋ:	ଠୋ:	ଠୋ:	ଠୋ:	ଠୋ:
/ah/	/oh/	/eh/	/ph/	N/A	/əh/	/aəh/	N/A	/ouh/		
N/A	/u:h/	N/A	N/A	/i:h/	/i:h/	N/A	N/A	/u:h/		
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										

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

- ឃ ឃ ឃ ឃ ឃ ឃ
- ស ស ស ស ស

9.2.5. How is the writing system different from Khmer?

- The ោ [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 ោ 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 ោ: [U+17C7] with ោ, it has to be encoded before the vowel. If placing ោ after ោ:, 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 ឃ ឃ to indicate that ឃ sounds [p], not [b]. Therefore, ឃឃ is [br] and ឃឃ is [pr]; ឃឃ is [bl] and ឃឃ is [pl].
- ោ 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 ែ [U+17D2 U+179C] is placed after subscript ឃ [U+17D2 U+179A] in words with a presyllable, i.e. ក្រែក (ក ឃ ែ ក ឃ) 'sound of door opened' is pronounced as [krwa:k]. In Khmer, this cannot happen because of the restriction of the pronunciation.

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	ច	ន	ជ	ឃ	ញ	
/k/	/kʰ/	/k/	/kʰ/	/g/	/c/	/ɟ/	/c/	/ɟ/	/p/

U+1780 U+1781 U+1782 U+1783 U+1784 U+1785 U+1786 U+1787 U+1788 U+1789

ដ	ប	ឌ	ឃុ	ណ	ត	ច	ទ	ធ	ស
/d/	/?d/	/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	ផ	ព	រ	ម	យ	វ	ល	វ	ស
/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	ឃុ	ឌ	ឃុ	ន	ច
/h/	/l/	/?/	/?j/	/g/	/?b/

U+17A2

U+17A0 U+17A1 U+17A2 U+17D2 U+179D U+179E

U+1799

9.3.2. Brao Subscripts

ឃ ី
/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, ឃ is realized as /ɛ/ in the 1st series and /i/ for the 2nd series.

inherent ឃ	ឃ	ី	ី	ី	ី	ី	ី	ី	ី
/ə/	/aa/	/ɛ/	/ij/	/ə/	N/A	/o/	/oo/	/uə/	/iə/
N/A	N/A	/i/	/ii/	/ɨ/	/ɨ/	/u/	/uu/	N/A	N/A

U+17B6 U+17B7 U+17B8 U+17B9 U+17BA U+17BB U+17BC U+17BD U+17BF

ឃ	ី	ី	ី	ី	ី	ី	ី	ី
/iə/	/ə/	N/A	/aj/	/əm/	/ah/	/əh/	/əm/	/am/
N/A	N/A	/ɛɛ/	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

ឃី

/aj/

N/A

U+17B6

U+17C6

U+1784

9.3.4. Brao Diacritics

ጀ ጀጀ ጀጀጀ ጀጀጀጀ ጀጀጀጀጀ

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. ጀ and ጀጀ) are used as consonants.
- Only two subscripts are used in Brao.
- Khmer does not have ጀጀ as a cluster, nor ጀጀጀ /?ɔr?jiip/ ‘very black’.
- ጀ is placed on any final consonant whose vowel (either /aa/ and /ɔɔ/) before it is shortened.
- A white space is used in between each word to denote a word boundary.
- ጀ lengthens the vowel in the /ɔɔh/ 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.

ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ	ጀ
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

ធម៌	ធប់	ធគា	ធនា	ធបា	ធតា	ធប៉ា	ធប៉ា	ធបា	ធប់
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

ធបា	ធរា	ធបា	ធយា	ធគា	ធលា	ធគា	ធសា	ធបា	ធវា
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

ធអា	ធបា	ធប់
q	g	qb

U+17A2 U+179D U+179E

9.4.2. Krung Subscripts

៥	៥	៥	៥	៥	៥
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	៥	៥	៥	៥	៥	៥	៥	៥	៥៥
/៥៥/	/aa/	/៥៥/	/៥៥/	/៥៥/	N/A	/៥៥/	/៥៥/	/៥៥/	/៥៥/
N/A	N/A	/i/	/ii/	/៥៥/	/៥៥/	/u/	/uu/	/ue/	N/A

U+17B6 U+17B7 U+17B8 U+17B9 U+17BA U+17BB U+17BC U+17BD U+17BE

ឃុំ	ីោ	ីោ	ឈោះ	ឈោះ	ិោះ	ិោះ
/ie/	N/A	/ay/	/ah/	/ah/	/eh/	/ang/
/ie/	/èè/	N/A	N/A	N/A	N/A	N/A
U+17C0 U+17C2 U+17C3 U+17C7	U+17C4 U+17C7	U+17B6 U+17C6 U+1784	U+17B7 U+17C7			

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. ឃ and ុ) are used as consonants.
- The consonant shifters are used with the obsolete characters (i.e. ឃ 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

ក	ខ	ច	យ	ជ	ច	ន	ជ	យ	ញ
/k/	/kʰ/	/k/	/kʰ/	/g/	/c/	/cV̄/	/c/	/cV̄/	/ɲ/

U+1780 U+1781 U+1782 U+1783 U+1784 U+1785 U+1786 U+1787 U+1788 U+1789

ជ	ប	ន	ឆ	ណ	ច	ទ	ធ	ស	
/tV̄/	/ʔd/	/tV̄/	/ʔd/	/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

ប	ផ	ព	រ	ម	យ	វ	ល	វ	ស
/bV̄/	/pʰ/	/p/	/pʰ/	/m/	/j/	/ɣ/	/l/	/w/	/s/

U+1794 U+1795 U+1796 U+1797 U+1798 U+1799 U+179A U+179B U+179C U+179F

ហ	ឃ	ន	ជ	ច
/h/	/l/	/ʔ/	/kV̄/	/ʔb/

U+17A0 U+17A1 U+17A2 U+179D U+179E

9.5.2. Jarai Subscripts

េ	ែ	ៃ	ោ	ៅ	ំ	ៈ	៉
/k/	/g/	/ŋ/	/n/	/m/	/j/	/l/	/w/

U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2

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

9.5.3. Vowels

inherent	ឃ	ង	ច	ឆ	ជ	ឈ	ញ	ដ	ឋ
/ə/	/a:/	/ɛ?/	N/A	/ə?/	N/A	/o?/	/ɔ:/	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									
ឃែក	ងែក	ចែក	ឆែក	ជែក	ឈែក	ញែក	ដែក	ឋែក	ឃែកែ
/ia/	/ie/	N/A	N/A	/am/	/ah/	/a?/	/ah/	/eh/	N/A
/ia/	N/A	/e:/	/ɛ:/	N/A	N/A	N/A	N/A	/ih/	/ɛh/
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									
ឃែក	ងែក	ចែកែ	ឆែកែ	ជែកែ	ឈែកែ	ញែកែ	ដែកែ	ឋែកែ	ឃែកែែ
/om/	/am/	/ap/	/əh/	/oh/					
/um/	N/A	N/A	/ih/	/uh/					
U+17BB	U+17B6	U+17B6	U+17B9	U+17BB					
U+17C6	U+17C6	U+17C6	U+17C7	U+17C7					
			U+1784						

9.5.4. Jarai Diacritics

Four diacritics are used.

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

9.5.5. How is the writing system different from Khmer?

- Two obsolete characters (i.e. ឃ and ឃ) 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.
 - ត្តិំ ធា” ឲ្យ ៥ ៥
 - ស៊ូម ន ៥ ី ៥ ម [hiyūm] ‘needle’
 - ឃ៊េ៥ ? [būa?] ‘work’
 - ឃ៊៥ ? [wā?] ‘oil’
 - ក្លេ៥ ? [klaap] ‘difficult’
 - ៥៉េ៥ ? [gūah] ‘morning’

9.6. Kuay

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

9.6.1. Kuay Consonants

ក	ខ	គ	ឃ	ឃ	ច	ឃ	ឃ	ឃ	ឃ
/k/	/kʰ/	/k/	/kʰ/	/g/	/c/	/cʰ/	/c/	/cʰ/	/j/

U+1780 U+1781 U+1782 U+1783 U+1784 U+1785 U+1786 U+1787 U+1788 U+1789

ធម៌	ធគា	ធបា	ធតា	ធប់	ធទា	ធពា	ធនា	ធរា	ធផា
/d/	/d/	/n/	/t/	/t ^h /	/t/	/t ^h /	/n/	/θ/	/p ^h /

U+178A U+178C U+178E U+178F U+1790 U+1791 U+1792 U+1793 U+1794 U+1795

ធតា	ធមា	ធយា	ធនា	ធបា	ធភា	ធសា	ធបិ	ធវា	ធអា
/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

ធឃ

/ʃ/

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]	[n]	[d]	[d ^h]
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
ត	ត	ត	ត	ត	ត	ត	ត	ត	ត
[t]	[t ^h]	[t]	[n]	[6]	[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	ិ	ី	ុ	ុ
[i]	[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	ិ	ី	ុ	ុ	ុ	ុ	ុ	ុ
/e:/	/a:/	/ɛ/	[ej]	/ɜ/	/ə:/	/ɔ o/	/o:/	/uə/	/ɜ:/
N/A	/ia/	/i/	/j/	/ə/	/ɹ/	/ɥ/	/ɯ/	N/A	/ɜ:/

U+17B6 U+17B7 U+17B8 U+17B9 U+17BA U+17BB U+17BC U+17BD U+17BE

ឃី	ិី	ីី	ុី	ុី	ុោ	ុោ	ុោ	ុោ	ុោ
/iə/	/iə/	/e:/	/ɛ/	[aj]	/ə:/	N/A	[əm]	[ah]	[ɛh]
/iə/	/iə/	/ɛ:/	N/A	[ej]	/ə:/	[əw]	[əm]	[ah]	[ɛh]

U+17BF U+17C0 U+17C1 U+17C2 U+17C3 U+17C4 U+17C5 U+17C6 U+17C7 U+17C1
U+17C7

ឃោះ	ិោះ	ីោះ	ុោះ	ុោះ	ុោះ
[əh]	[əh]	N/A	[əh]	[əm]	[əŋ]
[uh]	[uh]	[ih]	N/A	[əm]	[əŋ]

U+17C4 U+17BB U+17B7 U+17B9 U+17B6 U+17B6
U+17C7 U+17C7 U+17C7 U+17C7 U+17C6 U+1784

9.6.4. Kuay Diacritics

Four diacritics are used.

៥ ም ሚ ሪ

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

Punctuation are used in the same as in Khmer language (i.e. ៩ ៤ ២ ៧ ៨ ១ ៣ ៤ ៦ ៧ ៨ ៩ ៩ ៩ ៩).

9.6.5. How is the writing system different from Khmer?

- ။ gets to be used on ៥, a thing which never occurs in Khmer.
- ။ can be used on a final consonant preceded by a Samyok Sannya or ၊.

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 that 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	យ	ង	ច	ឆ	ជ	យ	ព
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

ជ	ប	ຂ	ຜ	ល	ព	ម	ទ	ສ	ធម៌
t	th	d	dh	ɳ	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

ុ	ិ	ុ	ី	ុ	ិ	ុ	ិ	ុ	ិ
u	ü	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

ុ	ិ	ី
h	l	m

U+17A0 U+17A1 U+17C6

9.7.1.2. Pali Subscripts

៥	៥	៥	៥	៥	៥	៥	៥	៥	៥
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

៥	៥	៥	៥	៥	៥	៥	៥	៥	៥
t	th	d	dh	ɳ	t	th	d	dh	n

U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 U+17D2 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	s

U+17D2									
U+1794	U+1795	U+1796	U+1797	U+1798	U+1799	U+179A	U+179B	U+179C	U+179F

ঃ
h
U+17D2
U+17A0

9.7.1.3. Pali Independent Vowels

ା	ା	ି	ି	ୟ	ୟ	ୟ	ୟ	ୟ	ୟ
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	ା	ି	ି	ୟ	ୟ	ୟ	ୟ	ୟ	ୟ
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

៥	៥	៥
m̥	im̥	um̥

U+17B7 U+17BB
U+17C6 U+17C6

9.7.1.6. How is the writing system different from Khmer?

- Nikahit (៥) 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: ៥ ៥ ៥ (i.e. ៥៥៥).
- Nikahit is used to combine with two vowels: ៥ and ៥ to make ៥ and ៥. ៥ is exceptionally found in Pali, and never in Khmer spelling convention. (It may be replaced by ៥ 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 (2005:២,៣៧,៤៨) includes three additional consonants: ៥៥, ៥៥៥ and ៥៥៥៥. 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

៥	៥	៥	៥	៥	៥	៥	៥	៥	៥
k	kh	g	gh	ṅ	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

ং	ঃ	ঙ	ঝ	ঘ	ঁ	ঁ	ঁ	ঁ	ঁ
t	th	d	dh	ñ	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

ঃ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ
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

ঁ	ঁ	ঁ
s	s	h

U+179E U+179F U+17A0

9.7.2.2. Sanskrit Subscripts

ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ
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

ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ
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

ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ
---	---	---	---	---	---	---	---	---	---

p	ph	b	bh	m	y	r	l	v	ç
U+17D2									
U+1794	U+1795	U+1796	U+1797	U+1798	U+1799	U+179A	U+179B	U+179C	U+179D
§	§	§							
s	s		h						
U+17D2	U+17D2		U+17D2						
U+179E	U+179F		U+17A0						

9.7.2.3. Sanskrit Independent Vowels

ା	ାମ	ି	ିମ୍	ୟ	ୟି	ର୍ଯୁ	ୟୁ	ିର୍ଯୁ	ୟେ
a	ā	i	ī	u	ū	r	īr	īl	e
U+17A3	U+17A4	U+17A5	U+17A6	U+17A7	U+17A9	U+17AB	U+17AC	U+17AD	U+17AF
ାଇ	ଓ	ାଉ							
ai	o	au							
U+17B0	U+17B1	U+17AA							

9.7.2.4. Sanskrit Dependent Vowels

inherent	ା	ି	ୟ	ୟି	ୟୁ	ୟେ	ାଇ	ୟାଇ	ୟୋଇ
a	ā	i	ī	u	ū	e	ai	o	au
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

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

៥	៥	៥ ^[18]
---	---	-------------------

k៥	k៥	k៥
----	----	----

U+1780 \ U+17D2 \ U+17AB U+1780 \ U+17D2 \ U+17AC U+1780 \ U+17D2 \ U+17AD

៥ is rendered incorrectly in some application, it should be rendered as shown below

Correct rendering:

Wrong rendering on Chrome for Mobile v75.0.3770.143: the two characters do not combine to be a cluster unit.

9.7.2.6. Sanskrit Diacritics

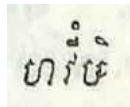
ጀ	ጀ	ጀ	ጀ	ጀ
---	---	---	---	---

Virama \ ិរោមា:	Avakraha	Anusvara	Visarga	Athisvara	Anuneaseka ^[19]
អវក្រាបា:		អនុស្សា:	ិស្សា:	អធិស្សា:	អនុនិកា:

U+17D1	U+17DC	U+17C6	U+17C8	U+17CC	N/A
VIRIAM	AVAKRAHASANYA NIKAHIT YUUKALEAPINTU ROBAT	-			

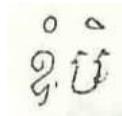
- ៥ Virama (Huot 1956:20-21)
- ៩ Avakraha (Huot 1956:21)
- ៥ Anusvara or Nikahit (Huot 1956:18-19)
- Anusvara is placed on a vowel (ibid:117)

ឧវិសារា



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

ឧវិសារា

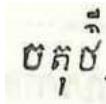


- ិ៍ Visarga or Visachani (Huot 1956:19-20)
- Anuneaseka (Huot 1956:19)



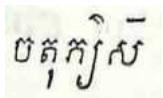
- [~]Athisvara RO in a cluster as the second member (Huot 1956:26). It is not listed as one of the 5 diacritics. It is instead an alternative appearance of ឈ when occurring after another consonant in an initial cluster.
- AbvV can be placed above the Athisvara (ibid:133)

ចតុវិស្ស



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

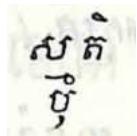
ចតុក្សែស



9.7.2.7. How is the writing system different from Khmer?

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

សិរិយា



- Consonant clusters which never exist in Khmer language:

- សត្វ sattva ‘animal’
- មត្វ matsya ‘fish’
- វ្រត्र vrata ‘buddhist temple’

មត្វ may not render correctly in some applications.

Correct rendering:

មត្វ

Incorrect rendering on Chrome for Mobile v75.0.3770.143: the two subscripts overlaps.

មត្វ

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/1jxv9xxrWNPd0U7j1wDW838cIUQMUMrV0?usp=sharing>.

Consonant Subscript Vowels Special Signs

ក	ឃ	ង	ំ
ខ	ច	នា	ែះ
ច	ឃ	ព់	េ[20]
យ	ឃុ	ុវ៉	
ជ	ឃ	៩	
ចេ	ឃ	៩	
ន	ឃ	ុ	
ជ	ឃ	ុរ៉	
យេ	ឃុ	ី	
ញ	ឃ	ី	
ជេ	ឃ	ីរ៉	
ប៊	ឃ	ីប៉	
ន៊	ឃ	ីន៉	
ញ៊	ឃ	ីញ៉	

ବା ତ ପ ଦ ନ ବ ମ ଯ ର ଲ କ ଶ ସ ଗ

11. Sample Texts for Orthography Check

This section includes sample text of each ethnic minority language.

11.1. Bunong Sample Text

ច្បាស់? ហាលសនប្រព័ន្ធគិង្វំពាក់ឆ្លាត់ដៃសក៍ពីកំ ការព្យួរតិចប្រវិទ្យាសក៍ជីថុ ។ លាសនប្រព័ន្ធដែកនាំប្រព័ន្ធប្រព័ន្ធមួយ មេដីកនាំប្រព័ន្ធប្រព័ន្ធបែងប្រជុំ នប្រព័ន្ធបែងប្រជុំ” ។ ការព្យួរតិចប្រព័ន្ធគិង្វំពាក់ឆ្លាត់រាយលាសនប្រព័ន្ធ មានឈើសម្រួលាសហានំអាចចំណាត់បានទូទាត់ដីភាពច្បាស់។

11.2. Tampuan Sample Text

(A drowning boy)

11.3. Brao Sample Text

លេខ ៨៤ បិះ

ບົນດີ ເລື່ອງ ມຍ ໄສ ຂໍມ້າຕ ທໍມ້າຕ ເລື່ອງ ຢ້າຕ ອາຕ່າກ ພ

យ៉ាត់ អាត់ព្រោះ នេ ត្រី បិច ក្នុង ប្រាប់ ឃី៖ វា តែង មួយ យ៉ាត់ អាត់ព្រោះ ត្រី ជីក ក្នុង ឈាម ណាក ហើ មី ដឹង ទីនឹង ត្រី ក្នុង ឈាម ណាក ត្រី ក្រុម ឈាម ។ ដឹង ត្រី ព័ត៌មូក ប្រយ ដឹង ឈាម ណាក តែ ដឹង ឃាយ យ៉ាវ លំក្នុង ឈាម ណាក ដឹង ត្រី ជី ហើ ហន់ឈាម ។ ដឹង ត្រី ជី កិ ក្រសិរ កខិប គើី ដឹង ត្រី ណាក ត្រី ឱក តាមៗ៖ «អើយ នើក ណាក ថា អីមួយ ឯុះ ? នៅ ហាយី ឯុះ ? » តែ ក្រោមីវ បានឱ្យ ។ ដឹង ត្រី ឱក តាមៗ ត្រី៖ អនំឈាម៖ «អើយ ហាយី ឯុះ ដឹង បក អនំឈាម៖ » ។ ដឹង ក្នុង ណាក ត្រី ក្រោមីវ៖ «អីមួយ អីមួយ អាត់ព្រោះ ដឹង ឈាម នៅ ត្រី ចាត់ ពី៖ តែ ដឹង ឃាយ យ៉ាវ អងក្នុង ឈាម នៅ ។ ដឹង ចាត់ អិតិ ដឹង ត្រី ណាក តែបី ប្រយ ដឹង ឈាម ណាក ទីនឹង មី ។ ដឹង ឃី៖ សាត តែ តែ អីមួយ ប្រសាធ កតាម ដឹង ឃាយ កតាម «ចាត់ មន់ពេ ដឹង អនំអា ក្នុង ប្រាប់ អនំសច ណាក អនំនៃ៖ » ។

និង ធ្វើ បាយ ដើម ថី សិរិក រៀន៖ «មេ បាយ មន់តែ ជួង នឹង ដើម ថី នឹង អនំអាំ ក្នុង ប្រាង ដើម និង» ។ និង មើល សាត ប្រយ ត្រូវ ក្រាល មើល ដក ថី អី តែ មើល តែ អីប៊។

នៃ ឡើ កញ្ញី ឡើ ដី សាត ឆី៖ អនុវារ តុំ កណ្តុំ ឆ្នង លាត ឆី៖ អនុវារ ឈាត់ តែ ឡើ តែ ។ ដើ ឡើ កញ្ញី ប្រយោះ «ឆ្នា អុ ប្រេ ប្រី យ៉ង បង់ដែរ ជីវិ អវក់ ប្រី ដាក គ្រប់ដែរ ចំពោ វន់សាត ជួង នៃ អុ ឈាត់ អនុឆ្នាប់ អំ កូន ប្រាង ហើ ហនុធមាម បាត់ មន់តែ ជួង នៃ ឈាត់ អនុឆ្នាប់ អំ កូន ប្រាង អនុសុច ហើ ហនុធមាម តែ ។

11.4. Jarai Sample Text

- ហុវ ហាកំងមនុស ញ្ច ូ ណារ ចេរ បាន ញ្ច ូ ពត ចល
 - ប្រជ ិទល នើយ ធម៌ ញ្ច ូ ក ឯ លើយ បាន ញ្ច ូ ញ្ច ូ ផ្លូត ណារ ពត សាង ញ្ច ូ ម៉ែ សយ ។
 - វិគ មិះង នន អីយ មង បាត យ ញ្ច ូ ប៉ែត ហិ បាន ។
 - ញ្ច ូ ហ្មាប់ច ូ បុរី បាន ឡ្វក ឯ «យ៉ា យិត អីៗ លួយកម្ព នាក ឯ ប្រើយ មនុស ចុក ខែ អីៗ?»
 - ញ្ច ូ អ្និតអ្និត ដើច អីៗ ខត បល នន យិត អីៗ ប្រើយ ញ្ច ូ ចុក ខែ អីៗ?»
 - បាន ឡ្វក ផ្លូត ឯ «អីៗ អ្ន ឡ្វក អី! មនុស ថ្ង វ ឯ ណារ! កតាំង ណារ! តាម ឃិន ញ្ច ូ អី!» «ហិម យិត តញ្ច ូ ? ពត ប៊ ញ្ច ូ ណារ?»
 - បាន នន ឡ្វក ផ្លូត ឯ «ញ្ច ូ ផ្លូត ណារ ពត សាង បស ញ្ច ូ ណារ ម៉ែ សយ ញ្ច ូ ។
 - តើញ្ច ូ ញ្ច ូ បស នើយ ធម៌ តើញ្ច ូ ញ្ច ូ ប៉ែត គឺន មើយ្យ ។»

11.5. Kuay Sample Text and the Translation in English

(The Story about Angels and Men))

ເນື້ອເຕີມ ເນື້ອງ

ເງື່ອນີ້ມ ຖຸຍແກວປ້າຍ ເຖິງກົນເຖິງຕັກດັບບໍ່ແຫຼ້ມ ເກີດເຮັດວຽກມາຮືນຶກູ້ຍຸກມາຕະບ ບັກຜູ້ບໍ່ແຫຼ້ມ ເພົ່າຍຸເປົ່າເຊີ້ນບັນດີ ປຶ້ມືດແນະກູ້ຍຸ້ມູ້ຍຸ້ນີ້ໃສ້ ເກີດມາເລີ້ມສີ້ນ ເຊົາມາ
ນາມໄສວຣອກຄູ່ຍຸກມາຕະບ ມອນເລັດລະບອບນັກວຽກລ້າຕີ ເບີບີ່ນບໍລິມະຍ ມອນນັກວຽກລ້າມະຍ ເບີບີ່ນລົກິພາວຕະກິມອນນັກວຽກລ້າມະຍ ۱

កីឡាចោរដូចមិនអាមេរិក កីឡានេរតាតាយ “លោស្សរៀបគារាបច្ចុប្បន្ន លោស្សរៀបគារាបច្ចុប្បន្ន ថាគ្រោះព្រៃ ថាគ្រោះប៉ូលទា” កីឡានេរតាតាយ “តួនកិចិតម៉ែនវី កិលបិនី សែងចោរ” ឡើតាតាយ ”ធមលគ្រោកខិមយោច្ចឹង ធមលគ្រោកខិមយោច្ចឹង ថាគ្រោះសុំ ថាគ្រោះចុង” បីនិច្ចិកនៅឱ្យតាក់ទា មិត្តភក្រាប់ជី ។

កីឡាឌ្ឋានដីដែលបានរាយការណ៍នៅក្នុងខេត្តសៀមរាប ជាផ្លូវការប្រើប្រាស់ដើម្បីបង្កើតការងារសំខាន់សំខាន់នៃក្រសួងពេទ្យ និងក្រសួងសំគាល់សំខាន់សំខាន់នៃក្រសួងពេទ្យ និងក្រសួងសំគាល់។

11.6. Pali Sample Text

(The Verses of the Buddha's Auspicious Victories)

(៤) សច្ខោរ វិហាយមធិ សច្ចកវាទកេតុ រាជកិណោ បិតមនិ អតិអន្ត ភ្នំ បញ្ហបទិ

បង្កើត ដឹកវា មុនិន្ទា តាមដសា ករពុ ទេ ដយមង្គលិ ។

(ព) នគ្គបាលនគរូបតាំ វិញ្ញុដំ មហិត្តិ បុគ្គល ចំរកុដគេន ទមាបយន្តា តីឡបទេស

វិធីនា ជិត្យវា មុនិក្រា តារេដសា កវត្ថុ នៅ ដយមង្គលានិ ។

(៤) ទូរាប់ខ្លួន រាជធានី សេចក្តីបាត់ ព្រៃយ៉ា វិស្វកិ ជាតិមិនិ ពាក្យិធានំ ព្រោះអតិថិជន

វិធីនា ជិត្យវា មួនឱ្យ ពារិនិត្យ ការងារ នៅក្នុងបណ្តុះបណ្តាល និង ដំឡើង សម្រាប់បាន ការងារ នៅក្នុងបណ្តុះបណ្តាល និង ដំឡើង សម្រាប់បាន

(៤) ឯការណ៍ធម្មជនអង្គភាពថា យោ វាទេន្ចាន់ ទិន្នន័យ សរពត មតិនឹង ហិតានសេរី

វិធានី ចេបទក្រឡិន មោកំ ស្បែខំ អធិតមេយ្យ ននា សបតោះ ។

11.7. Sanskrit Sample Text

(To be obtained)

11.8. Inscription Sample Text

(K.557 Inscription, Mungkol Borey, Takeo)

๙- น้ำดื่มที่เคียงบบดีต่อสุขภาพอย่างไร ดีท็อกซ์ได้เท่าไหร่บ้างและดีอย่างไร

๒- รัฐธรรมนูญไทยได้กำหนดให้เป็นวันสำคัญของชาติไทย คือ ๑๕ พฤษภาคม พ.ศ.๒๕๖๒

๓- រៀបចំក្រសួងនៃគោលទេរសភាទីផ្លូវការនៅក្នុងក្រសួង

ចុះត្រូវបែនព្យ័រដោយក្រុងមិនពាយ

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 ច ុ). This paper seeks to describe each character in use in the Khmer script, including usage in *ligatures, encoding order, 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). វិធីបង្កើតពាក្យ កម្មពាក្យ និង ការរំប្រាស់ពាក្យ “Creating Words, Borrowing Words and Using Words.” Phnom Penh: Royal Academy of Cambodia.

- Chin, C., Ung, S., Men, P., Em, U., & Kheang, S. (2012). ភាសាថ្មី ឆ្នាំទី១ “Pali Language, Grade 1.” National Buddhist Studies.
- Chuon, N. (1967) វចនាឌីក្រមខ្សោយ “Khmer Dictionary”. Phnom Penh: Buddhist Institute.
- Developing OpenType Fonts for Khmer Script - Typography. (2018, February 8). Retrieved from <https://docs.microsoft.com/en-us/typography/script-development/khmer>.
- 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. Retrieved from https://www.sil.org/system/files/reapdata/15/34/62/153462537465381054623906304930919921193/Spoof_Vulnerable_Rendering_in_Khmer_Uni
- 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). មេណ្ឌលភាសាសំព្លើតិ “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). សំគ្គរណិសំព្លើតិ “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). នៅយករាយភាសាខ្មែរ “Khmer Grammar.” Royal Academy of Cambodia. Retrieved from <http://www.elibraryofcambodia.org/veyeakor-peasa-khmer/>.
- Kul, S. (2008). ភាសាខ្មែរ “Khmer Language.” Nikrotheawoan Pagoda.
- Maspero, G. (1915). Grammaire de la langue khmère (cambodgien). Paris, Impr. nationale. Retrieved from <http://archive.org/details/grammairedelala00maspgoog>.
- Muang, P. (2012). ព័ត៌មានខ្លួនក្រមទំពុន-ខ្មែរ “Tampuan-Khmer Dictionary.” (P. Tuy & S. Chhuk, Eds.) (First Draft). Phnom Penh. Retrieved from <http://www.tampuanreader.com/en/dictionary>.
- Nhok, T. (1962). បែបង្កើត ភាសាពាណិជ្ជកម្មថែង “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). អក្សរនគរបាលខ្មែរ “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>.
- Pawley, M.-S., & Pawley, E. (2013). Cambodian Jarai Khmer Based Writing System.
- Pech, Y. (2006). ស៊ូវនៅក្រោមក្រុមទំពុន “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/typfacedesign/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 12.1.0, (Mountain View, CA: The Unicode Consortium, 2019. ISBN 978-1-936213-25-2). Retrieved October 14, 2019, from <http://www.unicode.org/versions/Unicode12.1.0/>.
- Thun, H. (2011). វិសាវករណ៍ខ្មែរ “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). សិលាបនវិកទេសប្រព័ន្ធសកម្មជាសម័យមូនអង្គរ “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 ក, look this pattern “^ក([^\u00e1])\$” up in an online dictionary (i.e. <http://dictionary.tovnah.com/search>). This could give a list of words which begin with ក followed by either any character beside a subscript or nothing at all.

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

1	ក-	kg	1st	21	ក្រាក	kjaok	peacock
2	ក-	kd	1st	44	ក្តាម	kdaam	crab
3	ក-	kd	2nd	5	ក្តុក	kdfuk	sound of something falling
4	ក-	kt	2nd	2	ក្តុម	ktum	tree of paradise
5	ក-	kn	1st	6	ក្តួង	knog	in
6	ក-	k6	1st	32	ក្តាចល	kbaal	head
7	ក-	km	1st	4	ក្រើច	kmeej	young
8	ក-	kr	1st	313	ក្រាម	kraom	under
9	ក-	kl	1st	36	ក្រូក	kliok	armpit
10	ក-	kw	1st	4	ក្រេន	kwaen	agile
11	ក-	ks	1st	17	ក្រាន់	ksaan	serene
12	ក-	k?	1st	24	ក្តុម	k?aam	clay pot
13	ខ-	kc	1st	16	ខោវ	kcaw	snail
14	ខ-	kc	2nd	23	ខិល	kcil	lazy
15	ខ-	kp	1st	32	ខិ	knog	ginger
16	ខ-	kt	1st	1	ខត	ktaat	to cackle (of hen)

17	ខ-	kt	2nd	30	ខំ	ktɔɔm	hut
18	ក-	kn	1st	34	ក្នុង	knaaŋ	back (of a person)
19	ក-	kp	2nd	16	ក្បស់	kpuəh	tall, high
20	ខ-	km	1st	34	ខ្មែរ	kmae	Khmer
21	ខេ-	kj	1st	3	ខេម្ខ	kjaɑŋ	mollusc
22	ខ-	kl	1st	25	ខ្សោ	klaa	tiger
23	ខ-	kw	1st	53	ខ្លិត	kwət	wood apple
24	ខេ-	ks	1st	33	ខ្សោច់	ksac	sand
25	ឆ្វើ-	kn	2nd	6	ឆ្វី	knie	1S, 3S (informal)
26	ប្រើ-	kr	2nd	143	ប្រាប់	kroap	seed
27	ឆ្វើ-	kw	2nd	1	ឆ្វាម	kwaam	text, composition (Thai loanword)
28	ឃើ-	kŋ	2nd	1	ឃើធរ	kŋəəc	in a jerking manner
29	ឃើ-	kn	2nd	7	ឃើប	kniep	pincers
30	ឃើ-	km	2nd	10	ឃើុ	kmum	bee
31	ឃើ-	kl	2nd	27	ឃើធរ	kluəŋ	leprosy
32	ឃើ-	kw	2nd	9	ឃើល	kwiel	to herd
33	ជ-			0			
34	ប្រួ-	cn	1st	4	ប្រួច	cnaj	to polish a gem
35	ប្រួ-	cŋ	1st	23	ប្រួង	cŋaŋ	to do battle
36	ប្រួ-	cj	1st	2	ប្រួត	cjot	to die (of angel)
37	ប្រួ-	cr	1st	132	ប្រួន	creən	much/many/a lot
38	ឆ្វិ-	ck	1st	20	ឆ្វិត	ckae	dog

39	ក្រ-	ck	2nd	2	ក្រោង	ckœŋj	to be awkward
40	ឆ្ល-	cj	1st	12	ឆ្លាយ	cjaaj	far
41	ឆ្ន-	cd	1st	4	ឆ្នាំ	cdœao	k.o. fish
42	ឆ្ន-	cn	1st	21	ឆ្នុក	cnok	cork, stopper
43	ឆ្ន-	cp	2nd	7	ឆ្និន	cpin	k.o. fish
44	ឆ្ន-	cm	1st	11	ឆ្នា	cmaa	cat
45	ឆ្ន-	cl	1st	26	ឆ្នីយ	claœj	to answer
46	ឆ្ន-	cw	1st	16	ឆ្នឹង	cweej	left (side)
47	ឆ្ន-	c?	1st	15	ឆ្នីង	c?œj	bone
48	ប្រ-	cr	2nd	83	ប្រុក	cruuk	pig
49	ច្ប-	cw	2nd	2	ច្បាត់	cwie	Java
50	លូរ-	cj	2nd	5	លូរក	cjook	to bow the head
51	លូរ-	cn	2nd	13	លូរត	cnuet	headband
52	លូរ-	cm	2nd	11	លូរល	cmoool	male (non-human)
53	លូរ-	cl	2nd	19	លូរីង	cleœŋj	leech
54	លូរ-	cw	2nd	4	លូរីង	cweej	to be clear, pure
55	ពុ-			0			
56	ផុ-			0			
57	បុ-			0			
58	ផុ-			0			
59	ណុ-			0			
60	ណុ-	nh	1st	2	ណុះយ	nhaœj	Don't bother

61	ពិ-	tj	1st	1	តោក	tjaok	neck-fetter (for criminals)
62	ផ្ទុ-	tn	1st	1	ត្រូត	tnaot	sugar-palm tree
63	គ្នា-	t6	1st	15	គ្នាល់	tbal	mortar
64	ឆ្នុ-	tm	1st	6	ឆ្នុត	tmaat	vulture
65	ត្រី-	tr	1st	230	ត្រីល់	tral	weaver's shuttle
66	ឆ្នូ-	tl	1st	9	ឆ្នូក	tlok	clown
67	ឆ្និ-	tw	1st	3	ឆ្និ	twaa	beef sausage
68	ឆ្និ-	t?	1st	7	ឆ្និក	t?aok	to hiccup
69	ច្បែ-	tk	1st	14	ច្បែល់	tkal	to prop up
70	ច្បែ-	tk	2nd	2	ច្បែម	tkiem	molar (tooth)
71	ច្បេ-	tj	1st	13	ច្បេស	tjaah	forehead
72	ច្បេ-	tf	1st	4	ច្បេក	tdfaok	small wooden bells
73	ច្បេ-	tn	1st	22	ច្បេល	tnaal	seed bed
74	ច្បេ-	tp	2nd	9	ច្បេល់	tpoal	cheek(s)
75	ច្បេ-	tm	1st	11	ច្បេ	tmaa	rock, stone, concrete
76	ច្បេ-	tl	1st	20	ច្បេម	tlaom	liver
77	ច្បេ-	tw	1st	15	ច្បេយ	twaaaj	to give, offer (to royalty, clergy or deities)
78	ទំរំ-	tj	2nd	1	ទំរំតិស្សាបន្ត	tjoo.te?. sah	a sacred script foretelling future
79	ប្រ-	tr	2nd	129	ប្រ	trwo	Cambodian stringed fiddle
80	ទូ-	tw	2nd	30	ទូវ	twie	door
81	ច្បេ-	tj	2nd	8	ច្បេន់	tjuon	to be heavy
82	ផ្ទុ-	tn	2nd	13	ផ្ទុរ	tnoe	shelf

83	ព្រ-	tm	2nd	21	ធ្វុរា	tmip	tooth
84	ចូរ-	tj	2nd	9	ឆ្វេង	tjuuŋ	charcoal
85	ចូ-	tl	2nd	21	ឆ្លាក់	tleak	to fall (unintentionally)
86	ចិ-	tw	2nd	5	ឆ្វី	twœə	to do
87	និស-	nr	2nd	2	នរោយ	nriej	Narayana (epithet of Vishnu)
88	និ-	n?	1st	1	នាមនិង	n?aal.niŋ	in order that
89	បុ-	pd'	1st	4	បុី	pdəj	husband
90	ប្រ-	pr	1st	517	ប្រាក់	prak	silver
91	បុ-	pl	1st	39	បែក	plack	to be odd
92	បុ-	p?	1st	2	បុណ្ណ	p?oon	younger sibling
93	ផិ-	pk	1st	20	ផ្តាយ	pkaaj	star
94	ផិ-	pk	2nd	13	ផំ	pkum	to group/assemble
95	ផិ-	pj	1st	9	ផ្សារ	pjaa	to be face up
96	ផិ-	pc	1st	13	ផ្សាក់	pcanj	to defeat
97	ផិ-	pp	1st	7	ផ្លើ	pnaə	to send
98	ផិ-	pd'	1st	33	ផិត	pdət	to pat dry
99	ផិ-	pt	1st	7	ផិល	ptəl	copper/silver bowl for water
100	ផិ-	pt	2nd	30	ផែះ	pteah	house
101	ផិ-	pn	1st	21	ផ្លូវ	pnoo	tomb
102	ផិ-	pl	1st	28	ផ្លូវ	pləw	road, street, path
103	ផូរ-	ps	1st	32	ផែរុង	psaeŋ	smoke (n)
104	ផិ-	p?	1st	31	ផ្លូម	p?aem	to be sweet

105 ພ-	pk	1st	1	ពុល	pkol	k.o. large tree
106 ພ-	pn	2nd	8	ពូង	pnwang	Bunong ethnic group
107 ພ៥-	pj	2nd	34	ព្យេះ	pjuh	storm
108 ພ្ល-	pr	2nd	201	ព្រោន	prien	hunter
109 ພ-	pl	2nd	3	ពូក	pluk	ceremonial leader in cremating a corpse
110 ພ-	p?	2nd	1	ពិះ	p?ih	tightly
111 ພ៥-	pj	2nd	10	ព្យួត	pjuut	to give someone a bath
112 ພ៥-	pc	2nd	12	ពិត	pcit	to seal
113 ພ៥-	pŋ	2nd	7	ព្យាក់	pneak	to wake up
114 ພ៥-	pn	2nd	28	ផ្លូក	pneek	eyes
115 ພ៥-	pm	2nd	1	ព្លាស	pmieh	entirely
116 ພ៥-	pl	2nd	49	ពូក	pluk	tusk
117 ພ្លិ-	pr	2nd	4	ព្រិន	pruun	stomach worm
118 ພ៥-	ps	2nd	1	ព្យែម	psoam	to accustom to
119 ພ-	mk	1st	2	ម្មាក់	mkak	k.o. fruit tree
120 ພ-	mk ^h	1st	1	ម្មាច	mk ^h aaj	one side
121 ພ-	mŋ	2nd	2	ម៉ែ	mŋaj	one day
122 ພ-	mc	1st	6	ម្មាស់	mcah	owner
123 ພ-	mc	2nd	2	ម្មូល	mcul	needle
124 ພ-	mŋ	2nd	2	ម៉ែម៉ែ	mŋae.mŋaa	to be always making excuses
125 ພ-	md	1st	5	ម្មង	mdaag	one time

126 ម-	mn	2nd	14	ម្នក់	mneak	one person
127 ម-	mp ^h	2nd	1	ម៉ែ	mp ^h ej	twenty
128 មូ-	mj	2nd	6	ម៉ាង	mjaaj	one way
129 មូ-	mr	2nd	23	ម៉ូន	mriem	finger/toe
130 ម-	ml	2nd	10	ម៉ុប	mlup	shade
131 មូ-	ms	1st	7	ម៉ោវ	msaw	flour
132 ម-	mh	1st	10	ម៉ូប	mhoop	food
133 ម-	m?	1st	2	ម៉ុម	m?aam	k.o. aromatic grass
134 ម-			0			
135 ខ-			0			
136 ល-	lk	1st	2	ល្អម	lkaam	to be very tender
137 ល-	lk ^h	1st	1	ល្អណ៍	lk ^h aon	theatrical performance
138 ល-	lk	2nd	6	លីក	lkik	if only
139 ល-	lj	2nd	10	លូ	ljeo	sesame
140 លូរ-	lø	1st	29	លូរយេ	løaaaj	solution, blended
141 ល-	lp	2nd	2	លំប់	lpiw	pumpkin
142 ល-	lm	2nd	23	លូតិ	lmut	sapodilla tree
143 លូរ-	lj	2nd	1	លីរំ	ljum	to be dangling
144 ល-	lw	2nd	16	លីដី	lwiij	bitter
145 ល-	lh	1st	16	លូដី	lhuuj	papaya
146 ល-	l?	1st	21	លាង	l?aaaj	cave
147 ខ-	wh	1st	3	ខ៊ី	whoj	in a daze

148 ធម្ម-	sk	1st	29	ស្វោរ	skaa	sugar
149 ធម្ម-	sk	2nd	19	ស្វោរ	skoə	drum
150 ធម្ម-	sj	1st	12	ស្វាប	sjaap	to yawn
151 ធម្ម-	sj	1st	15	ស្វើប	sjaəp	to shudder (in fear)
152 ធម្ម-	sf	1st	29	ស្វាំ	sfam	right (hand side)
153 ធម្ម-	st	1st	16	ស្វក	stok	stock
154 ធម្ម-	st ^h	1st	14	ស្វាន	st ^h aan	place
155 ធម្ម-	st	2nd	54	ស្វង់	stoɔj	bunch (of banana)
156 ធម្ម-	sn	1st	60	ស្វា	snaa	crossbow
157 ធម្ម-	sb	1st	33	ស្វវេ	səew	k.o. coarse grass
158 ធម្ម-	sp	2nd	28	ស្វាន	spien	bridge
159 ធម្ម-	sm	1st	69	ស្វា	smaa	shoulder
160 ធហុ-	sr	1st	318	ស្រែះ	srah	pond
161 ធម្ម-	sl	1st	74	ស្វីក	slək	leaf
162 ធម្ម-	sw	1st	67	ស្វូយ	swaaj	mango
163 ធម្ម-	s?	1st	25	ផ្សែក	s?ack	tomorrow
164 ធម-	n	1st	6	ប្លឹង	nəŋ	this, these
165 ធម-	m	1st	14	ប្លត់	mat	all gone
166 ធហុ-	r	1st	3	ប្រញួ	hrah.swa?	dwarf
167 ធម-	l	1st	10	ប្លង់	luəŋ	king
168 ធម-	w	1st	10	ប្លិយ	fił	film
169 ធម-			0			

170 ឃ-	?g	1st	1	ឃោង	?gaeng	2P (to younger girls)
171 ឃ-	?n	1st	15	ឃុក	neak	2P
172 ឃ-	?w	1st	2	ឃិ	?wəj	what
173 ឃ-	?h	1st	17	ឃោង	?haeng	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 “[^\u00e1]C\u00e1C[^\u00e1]” 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, [^\u00e1]\u00e1.[^\u00e1] matches any word containing \u00e1 and \u00e1 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

-ក្រ-	430	តក្រមា /tak.ka?.maa/ ‘to be stupefied’
-ក្រ-	46	អក្ររា /?ak.kʰa?.raa/ ‘letters’
-ក្រ-	117	ពាក្យកំព្យ /piek.kam.prie/ ‘a single character word’
-ក្រ-	4	ក្រី /pʰea?.kdəj/ ‘loyalty’
-ក្រ-	[22]	អបសក្រិ /?ap.sak/ ‘powerless’
-ក្រ-	9	ទក្រិណ៍ /teak.sən/ ‘the south’
-ក្រ-	61	

-៥.-	68	(All are similar to word-initial consonant clusters.)
-៥៥.-	157	បុគ្គល /bok.kul/ ‘individual’
-៥៥-	50	យោគេរាជ /joo.kjea?.p ^h iep/ ‘intelligence’
-៥៥-	2	ខេត្តក្រសួង /?uk.k ^h oo.sa?.naa/ ‘public address’
-៥៥-	2	អង់គ្លេស /?aŋ.kleh/ ‘English’
-៥៥-	2	
-៥៥.-	16	វិប្បុភាព /wi?.krea?.p ^h iep/ ‘destruction’
-៥៥-	1	
-៥៥.-	640	កំតុប /kaŋ kaep/ ‘frog’
-៥៥-	172	ជំថ្វី /dəŋ.k ^h aw/ ‘head of a commercial house’
-៥៥-	33	ជំបំ /dəŋ.kum/ ‘bunched together (of trees)’
-៥៥-	153	សង្កែម /saŋ.k ^h im/ ‘hope (n)’
-៥៥-	27	បង្កែលឃើ /baŋ.θaoj/ ‘to swing the arms while walking’
-៥៥-	1	ពង្កើក /puŋ.riik/ ‘to magnify’
-៥៥-	59	កង្វាល់ /kaŋ.wal/ ‘worry (n)’
-៥៥-	58	សង្កែរឃើ /saŋ.saj/ ‘to suspect’
-៥៥-	30	កង្វារ /kaŋ.haa/ ‘fan’
-៥៥-	62	បង្កែមឃើ /baŋ.?aem/ ‘dessert’
-៥៥-	46	
-៥៥.-	167	បច្ចីឃើ /pac.caj/ ‘suffix’
-៥៥-	53	មច្ចាន /mac.c ^h aa/ ‘fish’
-៥៥-	88	
-៥៥.-	28	(All are similar to the word-initial consonant cluster.)
-៥៥.-	120	បញ្ជាន /pdac.naa/ ‘to commit’

-ពិ-	6	ពាណិជ្ជកម្ម /pie.nic.cea?.kam/ ‘commercial’
-ផុំ-	64	មផ្សែត្រន /mac.c ^h ea?.t ^h aan/ ‘habitat’
-ផុំ-	23	រដ្ឋរាជក្រឹង /rieck.jieŋ.kea?/ ‘royal administration’
-ផុំ-	2	
-ណូ-	6	(All are similar to the word-initial consonant cluster.)
-កា-	320	កញ្ចក់ /kap.caŋ/ ‘mirror, glass’
-កា-	71	កញ្ញា /kap.ca ^h aa/ ‘marijuana’
-កា-	33	កប់ក្រី /kap.caə/ ‘basket’
-កា-	91	កញ្ចូល /kap.ca ^h uh/ ‘to scrape the foot on the ground’
-កា-	9	កញ្ចា /kap.caaa/ ‘young woman’
-កា-	113	បញ្ចា /pan.haa/ ‘problem’
-កា-	1	
-សិ-	114	សេចិី /see.t ^h əj/ ‘wealthy man’
-សិ-	109	អដ្ឋប្រទិភី /?at.da? ^h .pra? ^h .te? ^h .p ^h uu/ ‘advocate’
-សិ-	5	
-បុ-	0	(no result found)
-គ-	18	វគ្គិ /wut.t ^h i?/ ‘prosperity’
-គ-	16	លើចាតិ /leet.du?.baat/ ‘the fall of a clump of earth’
-គ-	2	
-នុ-	1	អាមុជ្រុ /?aa.tjea?.caa/ ‘to have been formerly wealthy’
-នុ-	1	
-ណុ-	138	កណ្ឌិត្យ /kan.dəŋ/ ‘bell’
-ណុ-	7	សណ្ឌភាព្យ /san.t ^h aa.kie/ ‘hotel’

- ិណ -	75	ពិណុ /?ən.duu/ ‘Hindu’
- ិន -	1	ិនុណី /tʰun.tʰi?/ ‘Genesha (son of Shiva)’
- ិណ្ឌ -	11	បិណ្ឌាគារ /pan.naa.kie/ ‘bookstore’
- ិណ -	32	ិណ្ឌទិន /bon.jea?.tin/ ‘holiday’
- ិណ -	1	តិណា /tan.haa/ ‘desire, passion’

- ិណ -

479

- ិត - [23]	236	កិតិយស /kət.te?.juəh/ ‘reputation’
- ិត -	7	ខត្តម /?ut.dam/ ‘excellent’
- ិត -	17	លព្យា /sa?.tjaa/ ‘true words’
- ិត -	82	វត្ថុ /woat.tʰo?/ ‘thing’
- ិត -	4	ទសវត្ថុវ /tuə.sa?.woat/ ‘a period of ten years’
- ិត -	4	ចត្តា /cat.twaa/ ‘four’

- ិច -

25 ិច្ចាបាន /mi?.tjaa.caa/ ‘adultery’

- ិច -

1

- ិទ -

235 ពិទិ /put.tʰi?/ ‘intellect’

- ិទ -

102 សិទាបិតិ /sat.tea?.ciet/ ‘sound, voice, speech’

- ិទ -

40

- ិទ -

30 (All are similar to the word-initial consonant cluster.)

- ិន -

735 ិនិ /kʰan.təj/ ‘tolerance’

- ិន -

141 កិន្ទារ /kan.tʰao/ ‘spittoon’

- ិន -

25 កិន្ទូយ /kan.tuj/ ‘tail’

- ិន -

224 កិន្ទូយ /kan.tʰiej/ ‘soft-shelled turtle’

-နှ-	93	ခို၌နှံယ် /tin.nea?.nej/ ‘data’
-နှ-	34	ပုံဗြာင် /pon.maan/ ‘how much/many’
-နှဲ-	7	ကန္ဂီ /kon.jaa/ ‘unmarried girl’
-နှို-	21	ပုံဗြာယ် /pun.riej/ ‘bright and shining with different colors’
-နှဲ-	1	ကန္ဂာတ /kon.laat/ ‘cockroach’
-နှဲ-	141	ဇန္ဂာယ် /tun.saaj/ ‘rabbit’
-နှို-	34	
-ပှ-	319	
-ပှါ-	116	ကပ္ပါဆ် /kap.bah/ ‘cotton’
-ပှိ-		ပူဗြာ /bop.p ^h aa/ ‘flower’
-ပှို-	9	ပြရှေ့ဗြီ /pra.nap/ ‘prohibition’
-ပှိ-	8	ပြာဗြာဒီလာဂ် /praa.pdaa.p ^h i?.liep/ ‘full of luck’
-ပှိ-	2	အပြုဂ် /?ap.sa?.raa/ ‘Apsara’
-ပှို-	4	အီးပြောဗြီ /?eo.roo.pjag/ ‘european’
-ပှို-	1	
-ဖှို-	76	(All are similar to the word-initial consonant cluster.)
-ပါ-	225	
-ပါ-		နီ၌ဗြာင် /nip.pien/ ‘Nirvana’
-ပါ-	53	နှုံဗြာပီဒိ /tu?.p ^h ie.sət/ ‘bad language/advice’
-ပါ-	25	ဒေါ်ဗြာ /tep.daa/ ‘angel’
-ပါ-	1	
-ဘါ-	63	
-ဘါ-		အကျိုတမ် /?a?.pjie.kum/ ‘war’
-မှု-	3	
-မှု-	660	စမ္မုလ် /sam.kul/ ‘to be swollen and ugly’
-မှု-	5	တမ္မုံဗြီဒိ /tam.b̥iət/ ‘tweezers’
-မှု-	81	စမ္မုနှုံး /sam.p ^h oah/ ‘perception, sensation’

-ម៉ែ-	9	សំឡូនិត /sam.piət/ ‘school bag’
-ម៉ា-	83	អារម្មណិ /?aa.rum/ ‘feeling’
-ម៉ុ-	61	ធម្មិជ /rom.ləəŋ/ ‘to uproot’
-ម៉ុ-	66	
-មួយ.-	51	ទេយ្យទាន /tej.jea?.tien/ ‘gift given to a Buddhist monk’
-មួយរ-	50	គុយុខ័ណី /kuj.hak/ ‘genitals’
-មួ-	1	
-កៅ.-	18	កៅកី /kee/ ‘honor’
-វោ-	5	វោនុស /wea?.rea?.man/ ‘Varman (title used by Khmer kings)’
-ស្វោ-	2	ស្វោរានិត /soo.rea?.kaan/ ‘k.o. mythical stone’
-ស្វោ-	7	អង្គតិវិប្បី /?an.dee.ra?.tʰəj/ ‘non-Buddhist’
-ស៊ែ-	1	អាសិនិពី /?aa.see.rea?.pih/ ‘poisonous snake’
-ស៊ែ-	1	ទុកិញ្ជុ /tu?.rea?.pʰik/ ‘famine’
-ស៊ែ-	1	ពុជ្ជិថ្នោន /poa.səə.laen/ ‘porcelain’
-លុ.-	68	ដុល្លារ /dol.laa/ ‘dollar’
-លុ-	26	
-ចុ.-	4	ជុវិ /ci?.whaa/ ‘tongue’
-ចុ-	4	
-សិ.-	442	និស្សិត /ni?.sət/ ‘student’
-សិស្ស-	110	ពិស្សិនិភាព /pi?.snu?.look/ ‘rank given to provincial chievers’
-សិស្ស-	5	
-សិ-	41	បុស្សិ /poh/ ‘post’
-សិ-	7	អស្សិរី /?ah.caa/ ‘wonderful’

-ពុ. -	40	ស៊ូរិត /so?.rit/ '(close) friend'
-ពុ-	2	
-ទី. -	12	ទាតីករណី /tea?.lhaj.kaa/ 'proof'
-ទី-	12	
-នុ. -	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

-ពិ	6	សិកិ
-ពិ	37	ទិភិ
-ពិ	1	បុកលិកិ
-ពិ	4	លិកិ
-ក្ស	8	ពាក្ស
-ក្ស	6	សុក្ស
-ក្ស	39	ប្រគ្សក្ស
-ខុ	1	ប្រមុខុ
-ធិ	20	វិធិ
-ធិ	3	អនុញ្ញ
-ធិ	2	ពិទិកិ
-គុ	3	អាណគុ

-ព្រៃ	ស្តីព្រៃ
-ប្រៃ	បណ្តុះប្រៃ
-អ្រៃ	អង្វៃ
-ស្រៃ	សង្រៃ
-ព្រៃ	ពង្រៃ
-ប្រៃ	ហង្ឃៃ
-ស្រៃ	សិត្រៃ
-ច្រៃ	កិច្ចៃ
-ប្រៃ	អប្បិច្ចៃ
-ប្រៃ	ពាយិច្ចៃ
-ស្រៃ	សង្គៃ
-ប្រៃ	ប្រាងៃ
-ប្រៃ	រាងៃ
-ព្រៃ	ពេប្រៃ
-ប្រៃ	ប្រីប្រៃ
-ប្រៃ	សាមប្រៃ
-ប្រៃ	វជ្ជៃ
-ប្រៃ	វជ្ជៃ
-ខ្លៃ	ខំខ្លៃ
-ខ្លៃ	អភិវឌ្ឍខ្លៃ
-ល្អៃ	វណ្ណៃ
-ល្អៃ	ខំល្អៃ

-ុំ	វុំ
-ុយ	បុុយ
-ុែ	ខុែ
-ុត្រ	មិត្រ
-ុត្រ	បរមត្រ
-ុត្រ	នានីត្រ
-ុត្រ	អាត្រ
-ុព្យ	ពិនិព្យ
-ុព្យ	ក្បុព្យ
-ុព្យ	សព្យ
-ុព្យ	យេវិព្យ
-ុច្ច	ពពុលច្ច
-ុទ្ធ	សទ្ធ
-ុទ្ធ	ប្រយុទ្ធទ
-ុទ្ធ	រចនប័ទ្ធ
-ុទ្ធ	ពទ្ធ
-ុទ្ធ	សុទ្ធ
-ុព្យ	សុព្យ
-ុន្ល	យុន្ល
-ុន្ល	និគ្រុន្ល
-ុន្ល	សុវិន្ល
-ុន្ល	ប្រពន្ល

-ន	13	បច្ចុប្បន្ន
-ន្ត	2	ជន្ត
-និរ	5	សុនិរ
-នេ	1	ជនេ
-ប	5	ប្រពេប
-ប្រ	1	យល់សប្ត
-ប្រូ	11	កិប្រូ
-ណ្ឌ	20	ទួសណ្ឌ
-ណុ	3	ប្រានុ
-ណុ	1	អណុ
-ណុរ	7	ទ្រណុរ
-ពុ	14	សពុ
-ព្រិតិ	1	អ៊ព្រិតិ
-ពុម្ព	6	ពុម្ព
-ម្ព	9	បារម្ព
-ម្ព	103	កសិកម្ព
-ម្យ	6	មនោរម្យ
-ម្រ	4	កម្រ
-ម្រ	3	សម្រ
-ម្រ	2	លម្រ
-ម្រ	1	លុម្រ
-យ្យ	19	គណនឹយ្យ

-ឃ	22	អាច្រួយ
-ិរិ	1	សិរិ
-ិនិ	1	សិនិ
-ិលិ	1	បំណើកលិ
-ិលិ	3	កលិ
-ិលិ	1	ពាយុគុលិ
-ិលិ	4	គុរុកាសិលិ
-ិលិ	6	កោសិលិ
-ិនិ	1	សិនិរិនិ
-ិសិ	2	សិវិសិ
-ិសិ	4	ដិសិ
-ិសិ	2	ត្រូវិសិ
-ិសិ	1	ប្រពិសិ
-ិសិ	2	អង្គាបុសិ
-ិសិ	1	សោសិ
-ិសិ	24	សិសិ
-ិបិ	2	ព្រិបិ
-ិវិ	2	អវិវិទិ

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 following a consonant. See the table below. Instances of three subscripts following a consonant does not exist in the KKD.

Word-Initial

[^] .Q.Q.	4	
ស្រែ-	2	ស្រែ /strɔj/ ‘woman’
ប្រែ-	2	ប្រែង /fraj/ ‘brake’
ស្រែ-	24	

Word-Medial

-Q.Q-	170	លក្ខី /leak.sməj/ ‘prosperous, wealthy’
-ក្រែ-	3	ប្រក្រោប /baŋ.kraap/ ‘to defeat’
-ក្រោ-	14	សក្រោក /saŋ.kjaa/ ‘counting’
-ក្រោះ-	2	អក្រោះសិ /?aq.kleh/ ‘English’
-ក្រោ-	1	ចក្រោង /caŋ.krug/ ‘wide open’
-ក្រោះ-	21	សក្រោះនិ /saŋ.kriec/ ‘head master (Buddhist monk)’
-ក្រោះំ-	1	ចិក្រោះំ /caŋ.cram/ ‘to chop’
-ក្រោះំ-	10	បក្រោះំសិ /θaŋ.cruəh/ ‘to purposely avoid (of travelling)’
-ក្រោះំ-	21	បិក្រោះំកសត្វ /pot.tjat.tʰeʔ.ka?.sat/ ‘vertebrate’
-ក្រោះំ-	1	កក្រោះំខោះ /kon.raoŋ/ ‘to jump up (to get something)’
-ក្រោះំ-	33	កក្រោះំកុ /kan.truək/ ‘to be worn out’
-ក្រោះំ-	43	សក្រោះំ /san.tjie/ ‘twilight’
-ក្រោះំ-	1	

-ស៊ូ-	1	បិសីតិ៍ /bih.kwii/ ‘biscuit’
-អ្នរ-	3	អាមេរិក /?aa.frik/ ‘Africa’

Word-Final

.ខ.ខ.ខ.	50	ភាគ /peak/ ‘face’
ន្ល-	1	នង្វែន /noo.reen/ ‘reigning king’
ន្យ-	13	នាំងុំ /rich/ ‘citizen’
ព្រៃ-	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	ក្ត	kŋ	ជ	ភ	pn
ក្ប	ក្ប / ក្ប?	kdf	ជ្រ	ភ្រ	pl
ក្ត	ក្ត	kt	ជ្រ	ភ្រ	ps
ក្ន	ក្ន	kn	ជ្រ	ជ្រ / ជ្រ?	p?
ក្បវ	ក្បវ / ក្បវ?	kɓ	ណ	ណ	pk
ក្ប	ក្ប	km	ជ្រ	ណ	pn
ក្រ	ក្រ	kr	ជ្រ	ណូ	pj

ក	ភ	kl	ប	ព	pr
ក	ធម្ម	kw	ប	ល	pl
ក្តុំ	ស្តូរ	ks	ប	ន	p?
ក	កិ / គិ?	k?	ប	រ	pj
ក	ធម្ម	kc	ប	ទ	pc
ក	យ	kŋ	ប	វ	pŋ
ក	ធម្ម	kt	ប	ធ	pn
ក	យ	kn	ប	ធម្ម	pm
យ	ធម្ម	kp	ប	រ	pr
យ	ធម្ម	km	ប	វ	pl
យ	យុ	kj	ប	ន	ps
យ	ធម្ម	kl	ប	ម	mk
យ	ធម្ម	kw	ប	ម្ប	mk ^h
យ	ធម្ម / យុ?	ks	ប	ឃ	mj
យ	ធម្ម	cn	ប	ជ	mc
យ	ធម្ម / ធម្ម?	cb	ប	ឃជ	mjh
យ	ធម្ម	cj	ប	ឃុ	md
យ	ធម្ម	cr	ប	ឃុ	mn
យ	ធម្ម	ck	ប	ឃុ	mp ^h
យ	ធម្ម	cŋ	ប	ឃុ	mj
យ	ធម្ម	cd'	ប	ប្រ	mr
យ	ធម្ម	cn	ប	ឃ	ml

ក	ក	cp	ឃ	ឃ	ឃ	ms
ក	ក	cm	ធម្ម	ធម្ម	ធម្ម	mh
ក	ក	cl	ធម្ម	ធម្ម	ធម្ម	m?
ក	ក	cw	ធម្ម	ធម្ម	ធម្ម	lk
ក	ក	c? / ធម្ម?	ធម្ម	ធម្ម	ធម្ម	lk ^h
ក	ក	cw	ធម្ម	ធម្ម	ធម្ម	lg
ឃ	ឃ	nh	ឃ	ឃ	ឃ	l6
ឃ	ឃ	tj	ឃ	ឃ	ឃ	lp
ឃ	ឃ	tn	ឃ	ឃ	ឃ	lm
ឃ	ឃ	t6	ឃ	ឃ	ឃ	lj
ឃ	ឃ	tm	ឃ	ឃ	ឃ	lw
ឃ	ឃ	tr	ឃ	ឃ	ឃ	lh
ឃ	ឃ	tl	ឃ	ឃ	ឃ	l?
ឃ	ឃ	tw	ឃ	ឃ	ឃ	wh
ឃ	ឃ	t? / ធម្ម?	ឃ	ឃ	ឃ	sk
ឃ	ឃ	tk	ឃ	ឃ	ឃ	sg
ឃ	ឃ	tj	ឃ	ឃ	ឃ	sp
ឃ	ឃ	td̚	ឃ	ឃ	ឃ	sd̚
ឃ	ឃ	tn	ឃ	ឃ	ឃ	st
ឃ	ឃ	tp	ឃ	ឃ	ឃ	st ^h
ឃ	ឃ	tm	ឃ	ឃ	ឃ	sn
ឃ	ឃ	tl	ឃ	ឃ	ឃ	sb

ស	ស	tw	ពិរ្យ	ស្តី	sp
ន	ន	tj	ស្តុ	ស្តុ	sm
ព	ព	tj	ស្រ	ស្រ	sr
នៃ / ប្រើ	ប្រើ	nr	ស្តុ	ស្តុ	sl
ន	ន	n?	ស្តុ	ស្តុ	sw
ប	ប	pd	ស្តុ	ស្តុ	s?
ប្រ	ប្រ	pr	ប្រ	ប្រ	n
ប្រ	ប្រ	pl	ប្រ	ប្រ	m
ប្រ	ប្រ	p?	ប្រ	ប្រ	r
ប្រ	ប្រ	pk	ប្រ	ប្រ	l
ប្រ	ប្រ	pŋ	ប្រ	ប្រ	w
ប្រ	ប្រ	pc	ប្រ	ប្រ	?ŋ
ប្រ	ប្រ	pɲ	ប្រ	ប្រ	?n
ប្រ	ប្រ	pd	ប្រ	ប្រ	?w
ប្រ	ប្រ	pt	ប្រ	ប្រ	?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 /uː/.

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

[5] Huffman (1970:8-9) uses /əɪ/

[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, ឃ is realized as /eʔ/ (Huffman 1970:26)

[9] ឃ is discouraged in formal writing.

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

[11] ែ (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. ឃែ ែែ ែែ៥ = ឃែែ៥).

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

[13] Below Vowels are: ឃ ឃ៥ ឃ៥៥.

[14] The vowel can be any of these: ឃ, ឃ៥ and ឃ៥៥.

[15] This is only found in documents published for the fifth edition of the Chuon Nath dictionary, i.e. ឃាយ (ឃ ឃ ឃ) ‘rice’ in present day Khmer is written as ឃ៥ (ឃ ឃ ឃ៥) 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 show up like the other two in the table. Please note this may vary from one application to another. You may refer to the images of both correct and incorrect renderings.
- [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 -ិ- look identical, but they are different.
- [23] -ឃ- and -ិ- looks identical, but they are different.
- [24] The data in the Khmer-Khmer online dictionary has different character sequences. Instead of ឃិ (ុ ិ ិ), its sequence is ិឃ (ិ ុ ិ). The one used in this documentation is ឃិ (ុ ិ ិ).