

Proposal for Khmer Script Root Zone Label Generation Rules (LGR)

LGR Version 2

Date: April 15, 2016

Document version: 1.5

Authors: Khmer Generation Panel

Contents

1	General Information/ Overview/ Abstract.....	3
2	Script for which the LGR is proposed.....	3
3	Background on Script and Principal Languages Using It.....	3
4	Overall Development Process and Methodology	4
5	Repertoire	5
5.1	Consonants.....	6
5.2	Independent Vowels.....	6
5.3	Dependent Vowels.....	7
5.4	Signs	7
5.4.1	YUUKALEAPINTU sign.....	7
5.4.2	SAMYOKSANNYA sign	8
5.4.3	NIKAHIT sign.....	8
5.4.4	REAHMUK sign	8
5.4.5	BANTOC sign	8
5.4.6	TOANDAKHIAT sign	8
5.4.7	ROBAT sign.....	8
5.4.8	COENG sign	8
5.5	Shifters	9
5.5.1	TRIISAP sign.....	9
5.5.2	MUUSIKATOAN sign.....	9
5.6	Shortlisted Repertoire.....	9
6	Variants	14
6.1	Khmer Variants	14
6.2	Cross-Script Variants.....	14
7	Whole Label Evaluation Rules (WLE)	14
7.1	No leading combining mark.....	14

7.2	Subscript Consonant	14
7.3	No More than Three Consonants in a Cluster.....	14
7.4	Context of COENG Sign (U+17D2).....	15
7.5	Context of Dependent Vowel.....	15
7.6	Context of Shifter - Khmer SIGN MUUSIKATOAN (U+17C9)	15
7.7	Context of Shifter - Khmer SIGN TRIISAP (U+17CA).....	15
7.8	Context of a Sign	15
7.9	Context of SAMYOKSANNYA sign (U+17D0)	15
7.10	Context of NIKAHIT SIGN (U+17C6)	15
7.11	Context of REAHMUK SIGN (U+17C7).....	15
7.12	Context of BANTOC SIGN (U+17CB).....	15
7.13	Context of TOANDAKHIAT SIGN (U+17CD)	16
8	Contributors	16
9	References	16
Appendix 1. Code Points Short-listed in the LGR Proposal		18
Appendix 2. Primary School Grade 1 [201] and Grammar [202] books by the Ministry of Education, Youth and Sports		22
Appendix 3: Initial Homoglyph Analysis by Khmer GP.....		26

1 General Information/ Overview/ Abstract

The purpose of this document is to give an overview of the proposed Khmer LGR in the XML format and the rationale behind the design decisions taken. It includes a discussion of relevant features of the script, the communities or languages using it, the process and methodology used and information on the contributors. The formal specification of the LGR can be found in the accompanying XML document:

- [Proposed-LGR-KhmerScript-20160415.xml](#)

Labels for testing can be found in the accompanying text document:

- [Labels-KhmerScript-20160415.txt](#)

2 Script for which the LGR is proposed

ISO 15924 – Code: Khmr

ISO 15924 – Number: 355

ISO 15924 - English Name: Khmer

Native Name: ខ្មែរ

Maximal Starting Repertoire (MSR) version: MSR-2 (based on Unicode 6.3.0 [3])

3 Background on Script and Principal Languages Using It

The Khmer script is an Abugida system used to write the Khmer language and some other languages spoken in Cambodia. It is also used to write Pali in the Buddhist liturgy of Cambodia and Thailand. Historically, it is believed to derive from the Pallava script, a variant of the Grantha alphabet, which descended from the Brahmi script, was used in southern India and South East Asia during the 5th and 6th centuries AD¹.

Khmer is written from left to right, without space between words, except to mark phrase or sentence boundaries. Originally there were 35 consonant characters, but modern Khmer uses only 33. Each consonant has an inherent vowel â or ô which can be explicitly overridden by writing a dependent vowel following it. In consonant clusters, in the onset or coda position of a syllable, the additional consonants, after the first one, are written in reduced form under it. Unicode refers these forms as subscript forms.

If there is no consonant in the onset of a syllable, the vowel is written in its independent form known as Independent Vowel, but when vowels come after the consonant(s) in the onset of a syllable, they are written in joined form around these consonant(s), with these forms known as Dependent Vowels. Most dependent vowels have two pronunciations, based on the inherent vowel (â or ô) of the consonant they

¹ Source: http://en.wikipedia.org/wiki/Khmer_alphabet.

attach with. There are additional diacritical marks used for further modifying the pronunciation. Additional characters in Khmer include those representing numerals and punctuation marks².

The Khmer script is used to write the Khmer language, which is the official language of Cambodia. In addition, it is also used to write a few other languages spoken in Cambodia, as listed in the table.

Language	ISO 639-3 Code(s)	Countries	EGIDS	Language Name in Khmer Script
Khmer	Khm	Cambodia, Laos, Thailand, Vietnam	1	ខ្មែរ
Tampuan	Tpu	Northeast border area, Central Rattanakiri province, Cambodia	6b	ទំពួន
Mnong	Cmo	Mondol Kiri province, Cambodia	5	គួង
Kuy	Kdt	North central Cambodia	6b	គួយ
Kru'ng	Krr	RattanaKiri and Mondol Kiri provinces, Cambodia	6b	ក្រៀង
Jarai	Jra	RattanaKiri province, Cambodia	5	ចារាយ

Note: There are few other languages spoken in Cambodia, but with small speaking population.

The Khmer script is used widely in Cambodia and Southern Vietnam and by the Cambodian expat community in France, Canada, Australia, USA and other countries.

Thai and Lao scripts are derived from the Khmer script. A similarity analysis is done by the Khmer Generation Panel for these scripts and is presented in the proposal.

4 Overall Development Process and Methodology

The panel formed a smaller working group of its members including those with technical, Unicode and linguistic expertise to conduct the initial analyses.

The working group had multiple face to face meetings in Phnom Penh to develop the proposal from June 2015 to January 2016. Where needed, the working group members also discussed linguistic matters with external experts from the Institute of National Language of Royal Academy of Cambodia which is an authoritative organization for the Khmer language and script. At each stage of the work, when the working group completed the task, the results were discussed with the members of the Khmer Generation Panel for review and were finalized based on feedback received.

The panel also got feedback of the interim work from the Integration Panel at ICANN to ensure that any concerns were addressed during the development of the proposal.

Finally, upon completion of the draft proposal, a public workshop was organized on 15 February 2016 in Phnom Penh where all the members of panel as well as other participants from private sector, ISP,

² ibid.

universities and relevant civil society were invited for collecting the feedback and recommendation of the proposal before it was finalized.

The cross-script variants were finalized after consultation with Lao and Thai GP representatives during face to face opportunity during ICANN55 meeting.

The remaining proposal represents the results of this process.

5 Repertoire

The Khmer script analysis for the Root zone Label Generation Rules takes code points shortlisted in MSR-2 as a starting point. The core references for this work are the Khmer language books used to teach Primary School Grade 1 [201] and Grammar [202] published by the Ministry of Education, Youth and Sports, Government of Cambodia, with relevant pages scanned and shown in Appendix 2. Also see [210] for a detailed linguistic description.

The current section includes a summary of analysis of the code points, based on which the repertoire has been selected. The repertoire and relevant properties of the code points are also discussed, based on which the label evaluation rules are derived in a later section.

The Khmer script organizes into grapheme clusters, which generally align with the phonological syllable, containing onset consonants, vowel nucleus and coda consonants. There can be zero to three consonants in onset and coda positions, as per the examples below. The written form may not conform strictly with the spoken form in some cases. Where there are no onset consonants, the syllable is written with independent vowels. These properties are similar to those of other Abugida scripts.

Phonological Structure	Word	Transliteration of Written Form	Transcription of the Spoken Form	Meaning in English
V (independent vowel)	ឯ	e	ae	at
CV (consonant + dependent vowel)	កា	ka	ka	cup
VC (independent vowel + consonant)	ឯក	ek	aek	one
CVC (consonant + dependent vowel + consonant)	មីង	ming	ming	aunt
CCV (consonant + consonant + dependent vowel)	ស្រែ	sre	srae	farm
CCCV (consonant + consonant + consonant + dependent vowel)	ស្រី	strei	strey	woman
CVCC (consonant + dependent vowel + consonant + consonant)	ព្រះ	poump	pum	template
CVCCC (consonant + dependent vowel + consonant + consonant + consonant)	ព្រះ	reastr	reas	citizen

Details about the writing system are also available at <http://www.omniglot.com/writing/khmer.htm>.

5.1 Consonants

There are 33 consonants that have been selected for inclusion. When more than one consonants occur together in the onset of a grapheme cluster, the additional consonants are written in subscript form, joined with the first consonant in the sequence. In Unicode this is done by additionally typing a subscripting mark given in the Unicode standard as ̣ COENG mark (U+17D2). For example, C₁C₂C₃V [្រ្រ្រ្រ] is written as C₁ subscript-mark C₂ subscript-mark C₃ V: 179F 17D2 178F 17D2 179A 17B8. Every other consonant can be subscripted except Khmer LETTER LA (U+17A1), because Khmer LETTER LO (U+179B) is used for spelling any word instead of Khmer LETTER LA (U+17A1) in subscript form. Unicode documents use of subscript version of 17A1 in Khmer script use in Thailand [100]. However, Khmer GP has not been able to determine the prevalence of this use. Thus, at this stage, U+17A1 has been tagged as base-only because it can occur before COENG mark (U+17D2) but not after it.

In the Khmer language, consonants can have light or strong pronunciation. This is orthographically indicated by shifter signs as discussed in more detail below. In addition, Khmer also uses six signs for phonological modifications as discussed in Section 5.4. The use of shifters and signs is sometimes restricted in context to different dependent vowels and/or consonants. These restrictions are also given in Sections 5.4 and 5.5.

The two historic consonants Khmer LETTER SHA (U+179D) and Khmer LETTER SSO (U+179E) are excluded. The main reason for this exclusion is that they are not generally used in writing the Khmer language. Both of these are used only for Pali/Sanskrit transliteration. The historic nature of these characters is well documented in https://en.wikipedia.org/wiki/Khmer_alphabet.

5.2 Independent Vowels

The Khmer script includes multiple independent vowels. Of these, twelve have been included and two independent vowels have been excluded from the repertoire. The excluded independent vowels are Khmer INDEPENDENT VOWEL QUU (U+17A9) and Khmer INDEPENDENT VOWEL QOO TYPE TWO (U+17B2). Khmer INDEPENDENT VOWEL QUU (U+17A9) has been excluded because it is not commonly used in the Khmer dictionary, and few words using this letter now have alternate spellings without it. For example, ធុន្ត (ūdth, camel) is now spelled as គ្រួន្ត, which uses a consonant in place of the independent vowel. Therefore this is not included. Moreover, no new words use this code point. Further, the Khmer INDEPENDENT VOWEL QOO TYPE TWO (U+17B2) is not used. It is a variant for Khmer INDEPENDENT VOWEL QOO TYPE ONE (U+17B1). Khmer INDEPENDENT VOWEL QOO TYPE TWO (U+17B2) is used to write only one word, the verb “give” but now Choun Nath Khmer Dictionary, which is an authoritative source on Khmer, writes the word as ឲ្យ (ooy, give) with Khmer INDEPENDENT VOWEL QOO TYPE ONE (U+17B1) instead.

The Khmer consonants occur in subscript form when they occur in clusters in onset or coda position of a syllable. However, in Unicode 8.0 (Table 16-9) [100], the following Independent vowels are also given in subscript form. Currently, these subscripted independent vowels are not being used for any words in the Khmer language and therefore, have been excluded for use in Khmer script LGR for the Root zone.

Table 16-9. Khmer Subscript Independent Vowel Signs

Glyph	Code	Name
្ក	17D2 17A7	khmer independent vowel sign coeng qu
្ខ	17D2 17AB	khmer independent vowel sign coeng ry
្គ	17D2 17AC	khmer independent vowel sign coeng ryy
្ឃ	17D2 17AF	khmer independent vowel sign coeng qe

The included independent vowels cannot combine with any of the marks including dependent vowels and sign. They occur independently.

5.3 Dependent Vowels

The Unicode Khmer block contains 16 dependent vowels that are included in the Khmer LGR'. Some of the dependent vowels have further been classified into three subcategories namely dependent vowel 1, dependent vowel 2 and dependent vowel 3. This sub-categorization has been done as different signs (U+17C6, U+17C7 and U+ 17CD) can be preceded by a one of these restricted sets, as discussed in Section 5.4. The rest of the sign characters do not follow a dependent vowel.

Dependent vowels must follow a consonant, shifter or the Robat sign. They cannot occur independently.

5.4 Signs

Signs have different phonological functions in Khmer [210]. After relevant analysis by the Khmer GP, several signs have been included in the Khmer LGR, most of which are used to make phonological modifications in spoken Khmer language. These signs and the restrictions on their use are discussed below. Two signs have been excluded from the LGR. These are Khmer KAKABAT sign (U+17CE) and Khmer AHSDA (U+17CF). Khmer KAKABAT sign ្ក (U+17CE) is used with any word which is pronounced louder than other words. This diacritic sign is not used in repertoire because this sign is used to change the pronunciation louder only. As this sign only represents spoken form of the word and it is not normally used in written form, it is excluded from repertoire. Examples include ឃ្ក (‘‘aw’’ vocal expression to acknowledge) ឃ្កយ (‘‘oay’’ vocal expression to react to somebody hitting) ឃ្កា (yes (for female)) ឃ្កន់ (hey) ឃ្កះ (there). Khmer AHSDA sign ្ក (U+17CF) is used with consonants ក ខ គ ឃ ង to give stressed intonation ក្កដ៏នីមីហ្ក, e.g. for the imperative form. Examples include ក្កដោយ (however) ដ្កល្ក (so good), ឃ្កដ៏អាកាស (at the sky) នុះន្ក! (there) មីទៅផ្ទះ! (go home!) ហ្កយកចុះ! (take this!). As this sign only represents spoken form of the word and it is not normally used in written form, it is excluded from repertoire.

5.4.1 YUUKALEAPINTU sign³

Khmer SIGN YUUKALEAPINTU (U+17C8) is written after a consonant to indicate that it is to be followed by a short vowel and a glottal stop [200]. The sign is used with all the consonants.

³ Shifters can be used with YUUKALEAPINTU sign. Ex: ក្កហ្ក, ហ្កី, ឃ្កា, ឃ្កី, ឃ្កុះ, ឃ្កុះ, ឃ្កុះ. They are used with some consonants only - ‘‘ហ្ក, ឃ្ក, ឃ្ក’’. These words occur mainly in spoken language. The Khmer GP does not recommended supporting them as labels in the root zone.

5.4.2 SAMYOKSANNYA sign

Khmer SIGN SAMYOKSANNYA (U+17D0) is written above a consonant or a shifter to indicate that the syllable contains a particular short vowel [200]. The sign is used with all the consonants or all the shifters.

5.4.3 NIKAHIT sign

Khmer SIGN NIKAHIT (U+17C6) is written over a consonant, a dependent vowel, or a shifter, it nasalizes the inherent or dependent vowel; long vowels are also shortened [200]. The sign is used with all the consonants, all the shifters or the dependent vowels AA and U. A class “dependent-vowel-1” is defined for this set of dependent vowels, i.e. {្ហ ្ឡ}.

5.4.4 REAHMUK sign

Khmer SIGN REAHMUK (U+17C7) is written after a consonant or a dependent vowel or a shifter, it modifies and adds final aspiration to the inherent or dependent vowel [200]. The sign is used with all the consonants, all the shifters or the dependent vowels I, U, E, OO and Y. A class “dependent-vowel-2” is defined for this set of dependent vowels, i.e. {្អ ្ឣ ្ឤ ្ឥ ្ឦ ្ឧ}.

5.4.5 BANTOC sign

Khmer SIGN BANTOC (U+17CB) is written over the last consonant of a syllable, indicating shortening (and corresponding change in quality) of certain vowels [200]. The sign is used with the consonants KA, NGO, CA, NYO, TA, NO, BA, LO, and SA, so a “series-three” class is defined as a set of consonants, i.e. {ក ង ច ញ ត ឆ ប ល ស} [205].

5.4.6 TOANDAKHIAT sign

Khmer SIGN TOANDAKHIAT (U+17CD) is written over a final consonant to indicate that it is unpronounced. It is used with all consonants or Khmer vowel SIGN I (U+17B7). A class “dependent-vowel-3” is defined for this Khmer vowel SIGN (U+17B7).

5.4.7 ROBAT sign

The Unicode Khmer block contains KHMER SIGN ROBAT (U+17CC), also included in the Khmer LGR. Khmer ROBAT sign ្ហ (U+17CC) is used to write Khmer words derived from Sanskrit. Some examples include ទុក្ខិត (durgat, impoverished), ទុរ្យាន (durjan, bad man), ធម៌ (dharm, Dhama - prayer), ព័ត៌មាន (vartamāna, information), បូព៌ (pūrva, east), បូព៌ា (pūrva, east), មាគ៌ា (mārgā, path), ហានសួគ៌ (han-svarg, heaven), អាថ៌កំបាំង (āarthakampangam, secret), បរិវត (parvat, mountain), បរិបូណ៌ (paripūrna, a lot of), ពណ៌ (varna, color). These words are commonly used especially ព័ត៌មាន (vartamāna, Information), etc.

ROBAT SIGN must have a consonant before it. However, unlike other signs, it can occur before dependent vowels. This constraint has already been discussed in the previous section.

5.4.8 COENG sign

The COENG SIGN (U+17D2) is used to precede consonant when there are subscripted. For further details, please refer to Sections 5.1 and 5.2 of this proposal.

5.5 Shifters

The Khmer script contains two consonant shifters - Khmer SIGN MUUSIKATOAN (U+17C9) and Khmer SIGN TRIISAP (U+17CA) whose roles are to shift the base consonant between registers (i.e. phonological class that determines following vowel); see Consonants Registers page 615 in the [100] specification and <http://rshida.net/scripts/block/khmer>).

5.5.1 TRIISAP sign

Khmer SIGN TRIISAP (U+17CA) is written above a consonant, used to convert some a-series consonants to o-series. The sign is used with the consonants BA, SA, HA and QA. A “series-one” class is defined as a set of consonants, i.e. {ប ស ហ វ រ}.

5.5.2 MUUSIKATOAN sign

Khmer SIGN MUUSIKATOAN (U+17C9) is written above a consonant, used to convert some o-series consonants to a-series. The sign is used with the consonants NGO, NYO, BA, MO, YO, RO and VO. A “series-two” class is defined as a set of consonants, i.e. {ង ញ ម យ វ រ ល}.

5.6 Shortlisted Repertoire

The repertoire finalized for inclusion in the root zone LGR by the Khmer GP is given below.

	Unicode Code Point	Glyph	Unicode Code Point Name	Unicode Code Point General Category	Category/ Tag	EGIDS and Language	Reference
1	1780	ក	KHMER LETTER KA	Lo	consonant series-three	1 Khmer	203, 205
2	1781	ខ	KHMER LETTER KHA	Lo	consonant	1 Khmer	203
3	1782	គ	KHMER LETTER KO	Lo	consonant	1 Khmer	203
4	1783	ឃ	KHMER LETTER KHO	Lo	consonant	1 Khmer	203
5	1784	ង	KHMER LETTER NGO	Lo	consonant series-two series-three	1 Khmer	203, 205, 210
6	1785	ច	KHMER LETTER CA	Lo	consonant series-three	1 Khmer	203, 205
7	1786	ឆ	KHMER LETTER CHA	Lo	consonant	1 Khmer	203
8	1787	ជ	KHMER LETTER CO	Lo	consonant	1 Khmer	203
9	1788	ឈ	KHMER LETTER CHO	Lo	consonant	1 Khmer	203

10	1789	ញ	KHMER LETTER NYO	Lo	consonant series-two series-three	1 Khmer	203, 205, 210
11	178A	ឆ	KHMER LETTER DA	Lo	consonant	1 Khmer	203
12	178B	ត	KHMER LETTER TTHA	Lo	consonant	1 Khmer	203
13	178C	ដ	KHMER LETTER DO	Lo	consonant	1 Khmer	203
14	178D	ធុ	KHMER LETTER TTHO	Lo	consonant	1 Khmer	203
15	178E	ណ	KHMER LETTER NNO	Lo	consonant	1 Khmer	203
16	178F	ត្រ	KHMER LETTER TA	Lo	consonant series-three	1 Khmer	203, 205
17	1790	ថ	KHMER LETTER THA	Lo	consonant	1 Khmer	203
18	1791	ត	KHMER LETTER TO	Lo	consonant	1 Khmer	203
19	1792	ថ	KHMER LETTER THO	Lo	consonant	1 Khmer	203
20	1793	ន	KHMER LETTER NO	Lo	consonant series-three	1 Khmer	203, 205
21	1794	ប	KHMER LETTER BA	Lo	consonant series-one series-two series-three	1 Khmer	203, 205, 210
22	1795	ផ	KHMER LETTER PHA	Lo	consonant	1 Khmer	203
23	1796	ព	KHMER LETTER PO	Lo	consonant	1 Khmer	203
24	1797	ភ	KHMER LETTER PHO	Lo	consonant	1 Khmer	203
25	1798	ម	KHMER LETTER MO	Lo	consonant series-two	1 Khmer	203, 210
26	1799	យ	KHMER LETTER YO	Lo	consonant series-two	1 Khmer	203, 210

27	179A	រ	KHMER LETTER RO	Lo	consonant series-two	1 Khmer	203, 210
28	179B	ល	KHMER LETTER LO	Lo	consonant series-three	1 Khmer	203, 205
29	179C	វ	KHMER LETTER VO	Lo	consonant series-two	1 Khmer	203, 210
30	179F	ស	KHMER LETTER SA	Lo	consonant series-one series-three	1 Khmer	203, 205, 210
31	17A0	ហ	KHMER LETTER HA	Lo	consonant series-one	1 Khmer	203, 210
32	17A1	ឃ	KHMER LETTER LA	Lo	base-only consonant	1 Khmer	203
33	17A2	ង	KHMER LETTER QA	Lo	consonant series-one	1 Khmer	203, 210
34	17A5	ឡ	KHMER INDEPENDENT VOWEL QI	Lo	independent-vowel	1 Khmer	206
35	17A6	ឺ	KHMER INDEPENDENT VOWEL QII	Lo	independent-vowel	1 Khmer	206
36	17A7	ឺ	KHMER INDEPENDENT VOWEL QU	Lo	independent-vowel	1 Khmer	206
37	17AA	ឺ	KHMER INDEPENDENT VOWEL QUUV	Lo	independent-vowel	1 Khmer	206
38	17AB	រ្យ	KHMER INDEPENDENT VOWEL RY	Lo	independent-vowel	1 Khmer	206
39	17AC	រ្យ	KHMER INDEPENDENT VOWEL RYY	Lo	independent-vowel	1 Khmer	206
40	17AD	ល្យ	KHMER INDEPENDENT VOWEL LY	Lo	independent-vowel	1 Khmer	206
41	17AE	ល្យ	KHMER INDEPENDENT VOWEL LYY	Lo	independent-vowel	1 Khmer	206
42	17AF	ឡ	KHMER INDEPENDENT VOWEL QE	Lo	independent-vowel	1 Khmer	206
43	17B0	ឡ	KHMER INDEPENDENT VOWEL QAI	Lo	independent-vowel	1 Khmer	206
44	17B1	ឡ	KHMER INDEPENDENT VOWEL QOQTYPEONE	Lo	independent-vowel	1 Khmer	206

45	17B3	្រ	KHMER INDEPENDENT VOWEL QAU	Lo	independent-vowel	1 Khmer	206
46	17B6	្រ	KHMER VOWEL SIGN AA	Mc	dependent-vowel dependent-vowel-1	1 Khmer	204
47	17B7	្រ	KHMER VOWEL SIGN I	Mn	dependent-vowel dependent-vowel-2 dependent-vowel-3	1 Khmer	204
48	17B8	្រ	KHMER VOWEL SIGN II	Mn	dependent-vowel	1 Khmer	204
49	17B9	្រ	KHMER VOWEL SIGN Y	Mn	dependent-vowel dependent-vowel-2	1 Khmer	204
50	17BA	្រ	KHMER VOWEL SIGN YY	Mn	dependent-vowel	1 Khmer	204
51	17BB	្រ	KHMER VOWEL SIGN U	Mn	dependent-vowel dependent-vowel-1 dependent-vowel-2	1 Khmer	204
52	17BC	្រ	KHMER VOWEL SIGN UU	Mn	dependent-vowel	1 Khmer	204
53	17BD	្រ	KHMER VOWEL SIGN UA	Mn	dependent-vowel	1 Khmer	204
54	17BE	្រ	KHMER VOWEL SIGN OE	Mc	dependent-vowel	1 Khmer	204
55	17BF	្រ	KHMER VOWEL SIGN YA	Mc	dependent-vowel	1 Khmer	204
56	17C0	្រ	KHMER VOWEL SIGN IE	Mc	dependent-vowel	1 Khmer	204

57	17C1	៊០	KHMER VOWEL SIGN E	Mc	dependent-vowel dependent-vowel-2	1 Khmer	204
58	17C2	៊០	KHMER VOWEL SIGN AE	Mc	dependent-vowel	1 Khmer	204
59	17C3	៊០	KHMER VOWEL SIGN AI	Mc	dependent-vowel	1 Khmer	204
60	17C4	៊០្រ	KHMER VOWEL SIGN OO	Mc	dependent-vowel dependent-vowel-2	1 Khmer	204
61	17C5	៊០្រ	KHMER VOWEL SIGN AU	Mc	dependent-vowel	1 Khmer	204
62	17C6	៊០	KHMER SIGN NIKAHIT and Vowel Am	Mn	sign	1 Khmer	204
63	17C7	៊០៖	KHMER SIGN REAHMUK and Vowel Ah	Mc	sign	1 Khmer	208
64	17C8	៊០៖	KHMER SIGN YUUKALEAPINTU	Mc	sign	1 Khmer	207, 208, 209
65	17C9	៊០	KHMER SIGN MUUSIKATOAN	Mn	shifter	1 Khmer	207, 208, 209, 210
66	17CA	៊០	KHMER SIGN TRIISAP	Mn	shifter	1 Khmer	207, 208, 109, 210
67	17CB	៊០	KHMER SIGN BANTOC	Mn	sign	1 Khmer	205, 207, 208, 209
68	17CC	៊០	KHMER SIGN ROBAT	Mn	robat	1 Khmer	207, 208, 209
69	17CD	៊០	KHMER SIGN TOANDAKHIAT	Mn	sign	1 Khmer	207, 208, 209
70	17D0	៊០	KHMER SIGN SAMYOKSANNYA	Mn	sign	1 Khmer	207, 208, 209

71	17D2	◌	KHMER SIGN COENG	Mn	coeng	1 Khmer	100
----	------	---	------------------	----	-------	---------	-----

6 Variants

6.1 Khmer Variants

Khmer GP considers two code points are variants if they are visually same to each other in either full consonant form or the subscript form. The code points 178F and 178A are visually identical in subscript form (i.e. when they follow 17D2), so the two sequences “17D2 178F” and “17D2 178A” are variants. As only one is correct for a given word, there is no need to allocate both. Therefore, they are proposed as blocked variants of each other.

6.2 Cross-Script Variants

Khmer Generation Panel has analyzed Khmer consonants and vowels compared to Thai, Lao and Burmese consonants and vowels. Some of consonants and vowels are very similar to Thai, Lao and Burmese (Appendix 3). However, when they occur in combined form in a label their shape changes. Based on this, and follow up discussion with other GPs (Khmer, Thai, Lao and Myanmar) there are no cross-script variants proposed.

7 Whole Label Evaluation Rules (WLE)

The following rules define restriction on grapheme cluster in Khmer. One or more grapheme clusters can be concatenated to form Khmer words and could also form a domain label.

7.1 No leading combining mark

The default rule in MSR-2 also applied to Khmer, where a label may not start with a combining mark, i.e. those code points with the General Category of Mn and Mc in the repertoire table.

7.2 Subscript Consonant

As discussed, all consonants except Khmer LETTER LA (U+17A1) can occur in subscript form. Subscript form is generated by preceding COENG mark (U+17D2). Thus, COENG mark (U+17D2) may be followed by any consonant except for Khmer LETTER LA (U+17A1), which is also tagged as “base-only” as this occurs only as the base consonant and not as a subscript consonant.

7.3 No More than Three Consonants in a Cluster

Khmer does not have a consonant cluster larger than three consonants. Thus, if there are more than three consonants in a cluster, it must not be valid. As a cluster is formed by joining subsequent consonants as

subscripts to the first one, no more than two consonants can be subscripted. Thus, if there are more than two occurrences of the subscripting COENG mark (U+17D2), it would form an invalid sequence.

7.4 Context of COENG Sign (U+17D2)

The COENG sign (U+17D2) used for subscripting consonants must occur between two consonants. If it occurs between any other categories, it is not in a valid context so the label is not well formed. Further, the consonant following it must not include Khmer LETTER LA (U+17A1), as already discussed. Finally, this constraint must also be checked in the context of variant sequences given in section 6.1, as they also have a COENG sign.

7.5 Context of Dependent Vowel

The dependent vowels in Khmer can only follow a consonant, a shifter or a Robat sign, thereby excluding the possibility of two consecutive dependent vowels. It cannot follow an independent vowel or a sign.

7.6 Context of Shifter - Khmer SIGN MUUSIKATOAN (U+17C9)

A shifter - Khmer SIGN MUUSIKATOAN (U+17C9) can only be preceded by one of the subset of consonants which have been tagged “series-two” in the repertoire table (ក ក្រ ម ឃ ង ឆ ជ ឈ ឍ).

7.7 Context of Shifter - Khmer SIGN TRIISAP (U+17CA)

A shifter - Khmer SIGN TRIISAP (U+17CA) can only be preceded by one of the subset of consonants which have been tagged “series-one” in the repertoire table (ក ស ហ វ រ).

7.8 Context of a Sign

Most signs in Khmer language can only occur after any consonant (which also has an inherent vowel). Some signs, such as SAMYOKSANNYA, NIKAHIT, REAHMUK, TOANDAKHIAT, and BANTOC have different rules described in the following rules.

7.9 Context of SAMYOKSANNYA sign (U+17D0)

The SAMYOKSANNYA sign (U+17D0) can only be preceded by a consonant or a shifter. It includes all consonants and all shifters

7.10 Context of NIKAHIT SIGN (U+17C6)

The NIKAHIT SIGN (U+17C6) can only be preceded by a consonant or a shifter or one of the subset of dependent vowels tagged “dependent-vowel-1” in the repertoire table (ា ុ), i.e. vowel signs AA and U.

7.11 Context of REAHMUK SIGN (U+17C7)

The REAHMUK SIGN (U+17C7) can only be preceded by a consonant or a shifter or one of the subset of dependent vowels tagged “dependent-vowel-2” in the repertoire table (ិ ី ុ េ ែ ៃ), i.e. vowel signs I, Y, U, E, and OO.

7.12 Context of BANTOC SIGN (U+17CB)

The BANTOC SIGN (U+17CB) can only be preceded by one of the subset of consonants tagged “series-three” in the repertoire table (ក ង ច ញ ត ឆ ឃ ឈ ស), i.e. consonants KA, NGO, CA, NYO, TA, NO, BA, LO, and SA.

7.13 Context of TOANDAKHIAT SIGN (U+17CD)

The TOANDAKHIAT SIGN (U+17CD) can only be preceded by a consonant or the Khmer Vowel SIGN I (U+17B7).

8 Contributors

Dr. Sopheap Seng, Chair

Mr. Rapid SUN, Secretary

Mr. Daro Chin, Member

Mrs. Yatal Lim, Member

Mr. Ken Rangsey, Member

Mr. Makara Than, Member

Mrs. Sopheap Say, Member

Mr. Hong Danh, Member

Mr. Ra An, Member

Mr. Khemara Mok, Member

Mr. Chhan Kimsoeun, Member

ICANN Staff

Sarmad Hussain

Rida Hijab Basit

9 References

3. The Unicode Consortium. The Unicode Standard, Version 6.3.0, (Mountain View, CA: The Unicode Consortium, 2013. ISBN 978-1-936213-08-5
100. The Unicode Consortium. The Unicode Standard, Version 8.0.0, (Mountain View, CA: The Unicode Consortium, 2015. ISBN 978-1-936213-10-8), Chapter 16: Southeast Asia, section 16.4: Khmer, pages 616-618. <http://www.unicode.org/versions/Unicode8.0.0/ch16.pdf>
101. Internet Corporation for Assigned Names and Numbers, "Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels." (Los Angeles, California: ICANN, March, 2013) <http://www.icann.org/en/resources/idn/variant-tlds/draft-lgr-procedure-20mar13-en.pdf>

102. Integration Panel "Requirements for LGR Proposals from Generation Panels" available online at <https://www.icann.org/en/system/files/files/Requirements-for-LGR-Proposals-20150424.pdf>
103. The Unicode Consortium, Unicode Character Database, available online as <http://www.unicode.org/Public/UCD/latest/>
200. Khmer alphabet, https://en.wikipedia.org/wiki/Khmer_alphabet
201. PRIMARY SCHOOL GRADE 1, MOEYS, ISBN 9-789-995-001-674, 2015
202. Khmer Grammar, MOEYS, ISBN 13-978-99963-33-16-3, 2015
203. PRIMARY SCHOOL GRADE 1, MOEYS, ISBN 9-789-995-001-674, 2015. See Figure 1 in Appendix C below.
204. PRIMARY SCHOOL GRADE 1, MOEYS, ISBN 9-789-995-001-674, Publication 2015. See Figure 2 in Appendix C below.
205. Prum Mol, *Grammar of Modern Khmer Language*, National Institute of Language, Royal Academy of Cambodia, 2006. See Figure 10 in Appendix C below.
206. PRIMARY SCHOOL GRADE 1, MOEYS, ISBN 9-789-995-001-674, 2015. See Figure 6 in Appendix C below.
207. PRIMARY SCHOOL GRADE 1, MOEYS, ISBN 9-789-995-001-674, 2015. See Figure 7 in Appendix C below.
208. PRIMARY SCHOOL GRADE 1, MOEYS, ISBN 9-789-995-001-674, 2015. See Figure 8 in Appendix C below.
209. PRIMARY SCHOOL GRADE 1, MOEYS, ISBN 9-789-995-001-674, 2015. See Figure 9 in Appendix C below.
210. Franklin E. Huffman, *Cambodian System of Writing and Beginning Reader*, Yale University, 1970, reprinted 1987. <http://www.pratyeka.org/csw/hlp-csw.pdf>. See particularly chapters III "Consonants" and VII "Diacritics and Punctuation."

Appendix 1. Code Points Short-listed in the LGR Proposal

1780

Khmer

17FF

	178	179	17A	17B	17C	17D	17E	17F
0	ក 1780	ថ 1790	ហ 17A0	ឡ 17B0	ៀ 17C0	័ 17D0	័ 17E0	័ 17F0
1	ខ 1781	ទ 1791	ឡ 17A1	ឱ 17B1	េ 17C1	័ 17D1	័ 17E1	័ 17F1
2	ត 1782	ធ 1792	អ 17A2	ច 17B2	ៃ 17C2	័ 17D2	័ 17E2	័ 17F2
3	យ 1783	ន 1793	អ 17A3	ឱ 17B3	ៃ 17C3	័ 17D3	័ 17E3	័ 17F3
4	ង 1784	ប 1794	អ 17A4	័ 17B4	ៃ 17C4	័ 17D4	័ 17E4	័ 17F4
5	ច 1785	ផ 1795	ត 17A5	័ 17B5	ៃ 17C5	័ 17D5	័ 17E5	័ 17F5
6	ឆ 1786	ព 1796	ឡ 17A6	ា 17B6	័ 17C6	័ 17D6	័ 17E6	័ 17F6
7	ជ 1787	ភ 1797	ខ 17A7	ិ 17B7	័ 17C7	័ 17D7	័ 17E7	័ 17F7
8	ឈ 1788	ម 1798	ឱ 17A8	ិ 17B8	័ 17C8	័ 17D8	័ 17E8	័ 17F8
9	ញ 1789	យ 1799	ខ 17A9	ិ 17B9	័ 17C9	័ 17D9	័ 17E9	័ 17F9
A	ង 178A	រ 179A	ឱ 17AA	ិ 17BA	័ 17CA	័ 17DA		
B	ថ 178B	ល 179B	ប 17AB	ិ 17BB	័ 17CB	័ 17DB		
C	ខ 178C	រ 179C	ប 17AC	ិ 17BC	័ 17CC	័ 17DC		
D	ផ 178D	ត 179D	ឡ 17AD	ិ 17BD	័ 17CD	័ 17DD		
E	ណ 178E	ថ 179E	ឡ 17AE	ិ 17BE	័ 17CE			
F	ត 178F	ស 179F	ង 17AF	ៀ 17BF	័ 17CF			

The Unicode Standard 8.0, Copyright © 1991-2015 Unicode, Inc. All rights reserved.

1780

Khmer

17C6

Consonants

1780	ក	KHMER LETTER KA
1781	ខ	KHMER LETTER KHA
1782	គ	KHMER LETTER KO
1783	ឃ	KHMER LETTER KHO
1784	ង	KHMER LETTER NGO
1785	ច	KHMER LETTER CA
1786	ឆ	KHMER LETTER CHA
1787	ជ	KHMER LETTER CO
1788	ឈ	KHMER LETTER CHO
1789	ញ	KHMER LETTER NYO
178A	ដ	KHMER LETTER DA
178B	ត	KHMER LETTER TTHA
178C	ឌ	KHMER LETTER DO
178D	ឍ	KHMER LETTER TTHO
178E	ណ	KHMER LETTER NNO

- as this character belongs to the first register, its correct transliteration is nna, not nno

178F	ក	KHMER LETTER TA
1790	ថ	KHMER LETTER THA
1791	ទ	KHMER LETTER TO
1792	ធិ	KHMER LETTER THO
1793	ណ	KHMER LETTER NO
1794	ប	KHMER LETTER BA
1795	ផ	KHMER LETTER PHA
1796	ព	KHMER LETTER PO
1797	ភ	KHMER LETTER PHO
1798	ម	KHMER LETTER MO
1799	យ	KHMER LETTER YO
179A	រ	KHMER LETTER RO
179B	ល	KHMER LETTER LO
179C	វ	KHMER LETTER VO

179D ឃ KHMER LETTER SHA

- used only for Pali/Sanskrit transliteration

179E ឆ KHMER LETTER SSO

- used only for Pali/Sanskrit transliteration
- as this character belongs to the first register, its correct transliteration is ssa, not sso

179F	ស	KHMER LETTER SA
17A0	ហ	KHMER LETTER HA
17A1	ឡ	KHMER LETTER LA
17A2	អ	KHMER LETTER QA

- glottal stop

Deprecated independent vowels for transliteration

These cloned independent vowels were originally intended only for Pali/Sanskrit transliteration.

17A3 អ KHMER INDEPENDENT VOWEL QAA

- use of this character is strongly discouraged; 17A2 អ should be used instead

17A4 អ៊ KHMER INDEPENDENT VOWEL QAA

- use of this character is strongly discouraged; the sequence 17A2 អ 17B6 ៊ should be used instead

Independent vowels

17A5 ក្រ KHMER INDEPENDENT VOWEL QI

17A6 ក្រ្រ KHMER INDEPENDENT VOWEL QII

17A7 ខ្រ KHMER INDEPENDENT VOWEL QU

17A8 ខ្រ្រ KHMER INDEPENDENT VOWEL QUK

- obsolete ligature for the sequence 17A7 ខ 1780 ក
- use of the sequence is now preferred

17A9	ខ្រ	KHMER INDEPENDENT VOWEL QUU
17AA	ខ្រ្រ	KHMER INDEPENDENT VOWEL QUUV
17AB	ឃ្រ	KHMER INDEPENDENT VOWEL RY
17AC	ឃ្រ្រ	KHMER INDEPENDENT VOWEL RYY
17AD	ឃ្រ	KHMER INDEPENDENT VOWEL LY
17AE	ឃ្រ្រ	KHMER INDEPENDENT VOWEL LYY
17AF	ង្រ	KHMER INDEPENDENT VOWEL QE
17B0	ង្រ	KHMER INDEPENDENT VOWEL QAI
17B1	ង្រ្រ	KHMER INDEPENDENT VOWEL QOO TYPE ONE
17B2	ង្រ្រ	KHMER INDEPENDENT VOWEL QOO TYPE TWO
		• this is a variant for 17B1 ង្រ្រ, used in only two words
		• 17B1 ង្រ្រ is the normal variant of this vowel
17B3	ង្រ្រ	KHMER INDEPENDENT VOWEL QAU

Inherent vowels

These are invisible combining marks for phonetic transcription to distinguish Indic language inherent vowels from Khmer inherent vowels. These characters are included solely for compatibility with particular applications; their use in other contexts is discouraged.

17B4 ៊ KHMER VOWEL INHERENT AQ

17B5 ៊ KHMER VOWEL INHERENT AA

Dependent vowel signs

17B6	៊	KHMER VOWEL SIGN AA
17B7	៊	KHMER VOWEL SIGN I
17B8	៊	KHMER VOWEL SIGN II
17B9	៊	KHMER VOWEL SIGN Y
17BA	៊	KHMER VOWEL SIGN YY
17BB	៊	KHMER VOWEL SIGN U
17BC	៊	KHMER VOWEL SIGN UU
17BD	៊	KHMER VOWEL SIGN UA

Two-part dependent vowel signs

These two-part dependent vowel signs have glyph pieces which stand on both sides of the consonant. These vowel signs follow the consonant in logical order, and should be handled as a unit for processing.

17BE	៊	KHMER VOWEL SIGN OE
17BF	៊	KHMER VOWEL SIGN YA
17C0	៊	KHMER VOWEL SIGN IE

Dependent vowel signs

17C1	៊	KHMER VOWEL SIGN E
17C2	៊	KHMER VOWEL SIGN AE
17C3	៊	KHMER VOWEL SIGN AI

Two-part dependent vowel signs

These two-part dependent vowel signs have glyph pieces which stand on both sides of the consonant. These vowel signs follow the consonant in logical order, and should be handled as a unit for processing.

17C4	៊	KHMER VOWEL SIGN OO
17C5	៊	KHMER VOWEL SIGN AU

Various signs

17C6	៊	KHMER SIGN NIKAHIT
		= srak am
		= anusvara
		• final nasalization
		• this character is usually regarded as a vowel sign am, along with om and aam
		→ OE4D ៊ thai character nikhahit
		→ 1036 ៊ myanmar sign anusvara

17C7	Khmer	17F6
<p>17C7 ៖ KHMER SIGN REAHMUK = srak ah = visarga → 1038 ៖ myanmar sign visarga</p> <p>17C8 ៖ KHMER SIGN YUUKALEAPINTU • inserts a short inherent vowel with abrupt glottal stop • the preferred transliteration is yukaleapintu</p> <p>Consonant shifters <i>These signs shift the base consonant between registers.</i></p> <p>17C9 ៖ KHMER SIGN MUUSIKATOAN • changes the second register to the first • the preferred transliteration is muusekatoan</p> <p>17CA ៖ KHMER SIGN TRIISAP • changes the first register to the second • the preferred transliteration is treisap</p> <p>Various signs</p> <p>17CB ៖ KHMER SIGN BANTOC • shortens the vowel sound in the previous orthographic syllable • the preferred transliteration is bantak</p> <p>17CC ៖ KHMER SIGN ROBAT • a diacritic historically corresponding to the repha form of ra in Devanagari</p> <p>17CD ៖ KHMER SIGN TOANDAKHIAT • indicates that the base character is not pronounced → 0E4C ៖ thai character thanthakhat</p> <p>17CE ៖ KHMER SIGN KAKABAT • sign used with some exclamations</p> <p>17CF ៖ KHMER SIGN AHSDA • denotes stressed intonation in some single-consonant words</p> <p>17D0 ៖ KHMER SIGN SAMYOK SANNYA • denotes deviation from the general rules of pronunciation, mostly used in loan words from Pali/Sanskrit, French, and so on</p> <p>17D1 ៖ KHMER SIGN VIRIAM • mostly obsolete, a “killer” • indicates that the base character is the final consonant of a word without its inherent vowel sound</p> <p>17D2 ៖ KHMER SIGN COENG • functions to indicate that the following Khmer letter is to be rendered subscripted • shape shown is arbitrary and is not visibly rendered</p> <p>Lunar date sign</p> <p>17D3 ៖ KHMER SIGN BATHAMASAT • originally intended as part of lunar date symbols • use of this character is discouraged in favor of the complete set of lunar date symbols → 19E0 ៖ khmer symbol pathamasat</p> <p>Various signs</p> <p>17D4 ៖ KHMER SIGN KHAN • functions as a full stop, period → 0E2F ៖ thai character paiyannoi → 104A ៖ myanmar sign little section</p>	<p>17D5 ៖ KHMER SIGN BARIYOOSAN • indicates the end of a section or a text. → 0E5A ៖ thai character angkhankhu → 104B ៖ myanmar sign section</p> <p>17D6 ៖ KHMER SIGN CAMNUC PII KUUH • functions as colon • the preferred transliteration is camnoc pii kuuh → 00F7 ៖ division sign → 0F14 ៖ tibetan mark gter tsheg</p> <p>17D7 ៖ KHMER SIGN LEK TOO • repetition sign → 0E46 ៖ thai character maiyamok</p> <p>17D8 ៖ KHMER SIGN BEYYAL • et cetera • use of this character is discouraged; other abbreviations for et cetera also exist • preferred spelling: 17D4 ៖ 179B ៖ 17D4 ៖</p> <p>17D9 ៖ KHMER SIGN PHNAEK MUAN • indicates the beginning of a book or a treatise • the preferred transliteration is phnek moan → 0E4F ៖ thai character fongman</p> <p>17DA ៖ KHMER SIGN KOOMUUT • indicates the end of a book or treatise • this forms a pair with 17D9 ៖ • the preferred transliteration is koomoot → 0E5B ៖ thai character khomut</p> <p>Currency symbol</p> <p>17DB ៖ KHMER CURRENCY SYMBOL RIEL</p> <p>Various signs</p> <p>17DC ៖ KHMER SIGN AVAKRAHASANYA • rare, shows an omitted Sanskrit vowel, like an apostrophe • the preferred transliteration is avakraha sannya → 093D ៖ devanagari sign avagraha</p> <p>17DD ៖ KHMER SIGN ATTHACAN • mostly obsolete • indicates that the base character is the final consonant of a word with its inherent vowel sound → 17D1 ៖ khmer sign viriam</p> <p>Digits</p> <p>17E0 ៖ KHMER DIGIT ZERO</p> <p>17E1 ៖ KHMER DIGIT ONE</p> <p>17E2 ៖ KHMER DIGIT TWO</p> <p>17E3 ៖ KHMER DIGIT THREE</p> <p>17E4 ៖ KHMER DIGIT FOUR</p> <p>17E5 ៖ KHMER DIGIT FIVE</p> <p>17E6 ៖ KHMER DIGIT SIX</p> <p>17E7 ៖ KHMER DIGIT SEVEN</p> <p>17E8 ៖ KHMER DIGIT EIGHT</p> <p>17E9 ៖ KHMER DIGIT NINE</p> <p>Numeric symbols for divination lore <i>These characters have numeric values 0-9, respectively, but are not used for calculation.</i></p> <p>17F0 ៖ KHMER SYMBOL LEK ATTAK SON</p> <p>17F1 ៖ KHMER SYMBOL LEK ATTAK MUOY</p> <p>17F2 ៖ KHMER SYMBOL LEK ATTAK PII</p> <p>17F3 ៖ KHMER SYMBOL LEK ATTAK BEI</p> <p>17F4 ៖ KHMER SYMBOL LEK ATTAK BUON</p> <p>17F5 ៖ KHMER SYMBOL LEK ATTAK PRAM</p> <p>17F6 ៖ KHMER SYMBOL LEK ATTAK PRAM-MUOY</p>	

17F7

Khmer

17F9

- 17F7 ្ក KHMER SYMBOL LEK ATTAK PRAM-PII
- 17F8 ្ខ KHMER SYMBOL LEK ATTAK PRAM-BEI
- 17F9 ្គ KHMER SYMBOL LEK ATTAK PRAM-BUON

Appendix 2. Primary School Grade 1 [201] and Grammar [202] books by the Ministry of Education, Youth and Sports

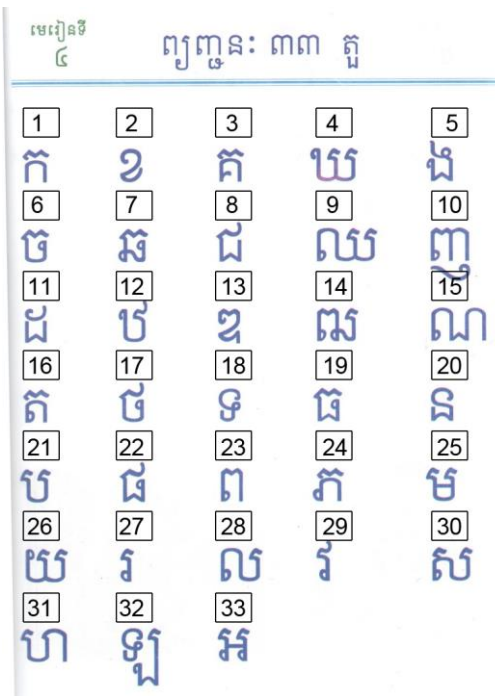


Figure 1 – Consonants

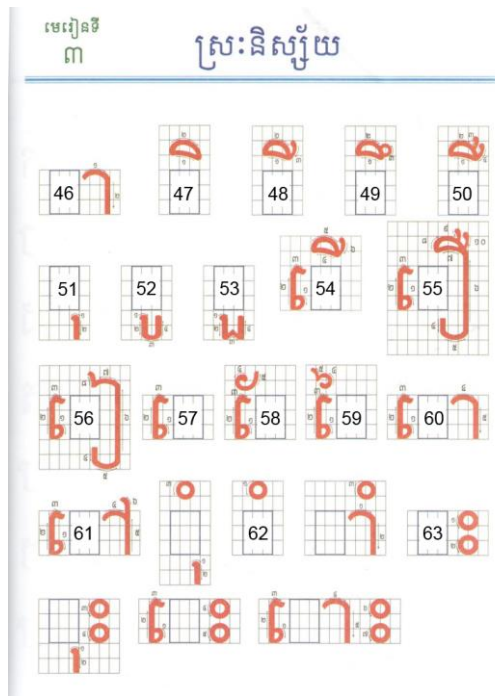


Figure 2 – Dependent Vowels

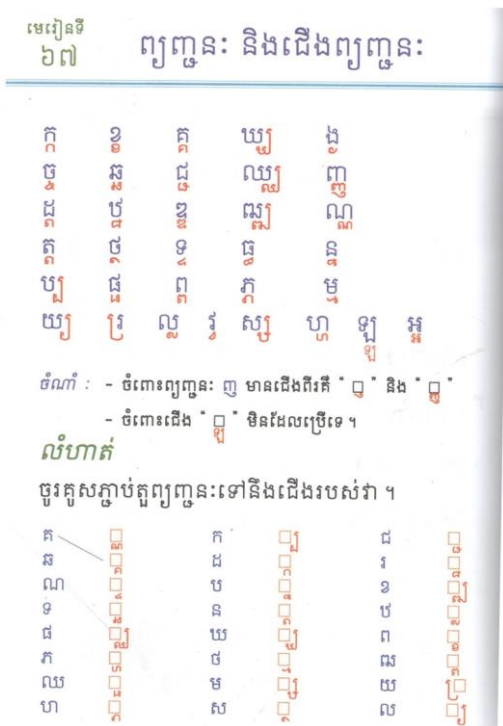


Figure 3 – Sub Consonants

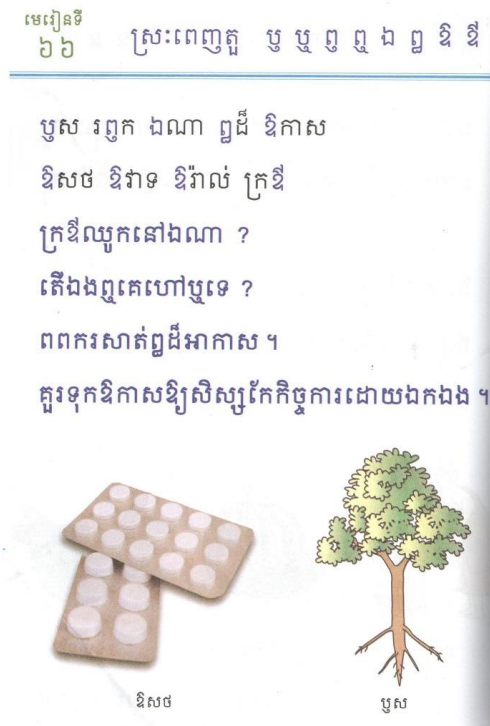


Figure 4 – Independent Vowels

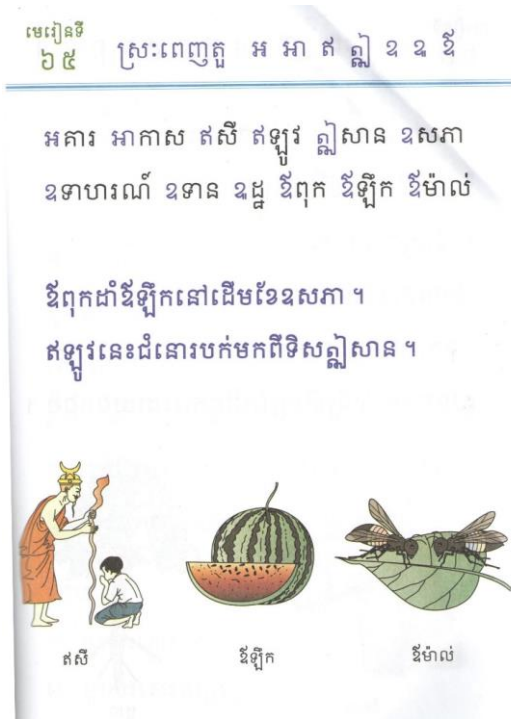


Figure 5 -- Independent Vowels

មូលកម្ម ឬ ធាតុកណ្តាល “ ៉ ” [65]

ជាសញ្ញាប្រើសម្រាប់សម្លាប់, បំបែកសំឡេងព្យញ្ជនៈ ៧ តួ ដែលមានសំឡេងយោសៈ ឲ្យទៅជាអយោសៈ គឺ ង, ញ, ប, ម, យ, រ, វ, ឲ្យទៅជាសំឡេងអយោសៈវិញ ង, ញ, ប, ម, យ, រ, វ, ។ ប្រែកតែតួ “ ប ” មួយដែលមានសំឡេងអយោសៈស្រាប់ ប៉ុន្តែគ្រាន់តែលម្អៀងសូរពី ប ទៅ ប៉ ប៉ុណ្ណោះ ។ អក្សរច្រើនសំឡេងទាំងនេះ ត្រូវបានប្រើក្នុងពាក្យ ងកង, ញៀងញៀង, បំប៉ាបំប៉ាច, ម៉ែឌី, យ៉ាប់យ៉ែន, រឿងកំរ, រឿងកំរង ។

- បំណាំ ៖ កាលណាពាក្យទាំងនេះទៅប្រទះនឹងស្រៈ ចិ ចិ ចិ ចិ ឆើ ឧទាហរណ៍ ៖ ផ្ទះចិត្ត យាប់យ៉ែន, ល្បែងចាប៉ុក... ។ គេត្រូវយកសញ្ញា (១) មកជំនួសធាតុកណ្តាល ដែលដាក់ពីលើ ដើម្បីឲ្យស្រៈយុត្តិធម៌ ទើបទាំងគ្នាជាមួយស្រៈខាងលើ ។

ត្រីសំព័ន្ធ (ិ) [66]

ជាសញ្ញាសម្រាប់ដាក់លើព្យញ្ជនៈ ៨ តួ ជាអយោសៈ ប, ស, ហ, អ, ឲ្យទៅជាអយោសៈវិញ ៖ បិ, សិ, ហិ, អិ, ។ ព្យញ្ជនៈមួយសំឡេងទាំងនេះ ត្រូវបានយកទៅប្រើក្នុងពាក្យ សាព្វិ, សម្ពុស្តិប, ហ៊ាន, ទ្រុឌិ... ។ កាលណាសញ្ញាត្រីសំព័ន្ធនេះទៅប្រទះនឹងស្រៈ ចិ ចិ ចិ ចិ ឆើ ក៏ត្រូវប្រែប្រួលរួមមកជា (១) ដូចករណីខាងលើដែរ ឧទាហរណ៍ សុំបាយ, លេងហ៊ុំ... លើកលែងតែពាក្យ ប៊ឹក, អ៊ុច៖អ៊ុច៖... ។

វស្សសញ្ញា ឬ បន្តក់ (ំ) [67]

សម្រាប់ប្រើដាក់លើព្យញ្ជនៈ ដែលកើតឡើងពីស្រៈ “ អ ” និងស្រៈ “ អា ” ហើយកាត់សំព័ន្ធជាឲ្យទៅជាខ្លី ដែលត្រូវបានលើស្រៈ ឧទាហរណ៍ ៖ កាត = សំព័ន្ធជែង, កាត់ = សំព័ន្ធជែង, ចប់ = សំព័ន្ធជែង

- កំណត់សម្គាល់
- ក/ ចំពោះព្យញ្ជនៈប្រកបនឹងអក្សរ “ ម ” ត្រង់សំឡេងប្រកបដោយទី២នោះ មិនបាច់ប្រើសញ្ញា “ ឲ្យ ” ឬបន្តក់ទេ ឧទាហរណ៍ ៖ ចាប់ ។ ក្នុងករណីនេះ គេប្រើនិមិត្តហ៊ុត() ឬស្រៈ “ ចាប់ ” ចាប់ “ ចាប់ ” ជំនួសវិញ ។
- ខ/ វិញព្យញ្ជនៈប្រកបនឹងអក្សរ “ យ ” ត្រង់សំឡេងប្រកបដោយទី២នោះ ក៏មិនបាច់ប្រើ “ វស្សសញ្ញា ” ឬបន្តក់ដែរ, ដូចជាពាក្យ “ ចាប់ ” ក្នុងករណីនេះគេយកស្រៈ “ ៃ ” មកជំនួសវិញ ព្រោះមានសំឡេងដូចគ្នា ។
- គ/ ចំពោះពាក្យទាំងឡាយណា ដែលក្លាយមកពីភាសាបាលី ឬសំស្ក្រឹត មិនត្រូវប្រើសញ្ញា (៦) នេះឡើយ ដូចជាពាក្យ បទ, សព, ភព, កុសល, ផល កិច្ចកល ធនធាន រសជាតិ សាមក ... ពាក្យទាំងនេះមានសំឡេងខ្លីស្រដៀងគ្នា តាមទម្លាប់អានបែបបាលីនិងសំស្ក្រឹត ។

សំយោគសញ្ញា (ំ) [70]

ក/ គេប្រើសញ្ញា (៦) នេះសម្រាប់លម្អៀងសំព័ន្ធជាឲ្យស្រដៀងគ្នាមកពីភាសាបាលី ឬសំស្ក្រឹត ដូចជា ៖ កយ > ក័យ, ជយ > ជ័យ, នយ > ន័យ, អាស្រយ > អាស្រ័យ[៦], អាសយ > អាស័យ ។

Figure 7 – Diacritics

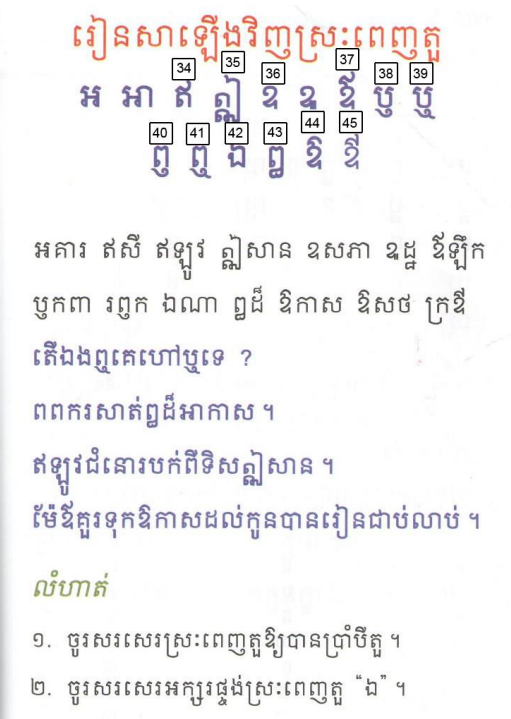


Figure 6 – Independent Vowels

ខ/ សញ្ញានេះក៏ប្រើលើពាក្យខ្លះខ្លះ ដែលមានសំឡេងស្រដៀងស្រៈ “ អ ” និង ស្រៈ “ អា ” ដើម្បីឲ្យមានសូរសំឡេងមួយត្រូវចំនឹងសំឡេងពិតនៃពាក្យនោះ គឺ ញី, ជី, ទុំ, វី... ។

គ/ សម្រាប់ប្រើសម្រួលសំឡេងក្លាយពីពាក្យបរទេសខ្លះៗ ដូចជា ដុកទំ, ប៉ាស្ទ័រ, អាំងហ្សេនីញ៉័រ, កុំព្យូទ័រ, ប្រាក់ចំរ [7] ។

រពាទ (ំ) [68]

វណ្ណយុត្តិធម៌ ដាក់លើពាក្យ “ រ ” ដែលគេកាត់យកក្បាលខាងលើរបស់វា មកក្បាច់ក្នុងសញ្ញា (៦) ។ គេប្រើវណ្ណយុត្តិធម៌ សម្រាប់ដាក់លើពាក្យដែលក្លាយមកពីភាសាសំស្ក្រឹត ដើម្បីសម្រួលសំឡេងអានតាមបែបសំឡេងខ្មែរ ។ គេកាត់សម្លាប់ថា ពាក្យបាលីណា ដែលសរសេរត្រូវនោះ កាលណាយើងសរសេរជាពាក្យសំស្ក្រឹតក្លាយខ្មែរ គេត្រូវដាក់ (៦) រពាទនេះជំនួសវិញ ។

អស្តា (ំ)

វណ្ណយុត្តិធម៌ គឺមានទ្រង់ទាញជាលេខ ៨ ប៉ុន្តែប្រែឲ្យតូច ដើម្បីងាយយកមកដាក់លើព្យញ្ជនៈ “ ក ” និង “ ផ ” ដែលមាននិមិត្តសំព័ន្ធ ។ ឧទាហរណ៍ ៖ ខ្ញុំក៏ទៅ, អ្នកក៏ទៅ, ទោះបីយ៉ាងណាក៏ដោយ ... ។ ក្រៅពីនេះគេប្រើលេខអស្តាសម្រាប់ដាក់លើពាក្យខ្មែរខ្លះខ្លះ ជាពាក្យត្រាស់ ស្នើចំ រឺ ភើប... និមិត្ត ដុំ៖ឃ្លី ។

ទណ្ណយោគ ឬ បដិសេធ (ំ) [69]

វណ្ណយុត្តិធម៌ មានរូបសណ្ឋានដូចសក់លេខ “ ៩ ” ។ គេប្រើទណ្ណយោគ ដើម្បីដាក់លើអក្សរព្យញ្ជនៈណា ដែលគេមិនអាន ប៉ុន្តែគេត្រូវតែរក្សាទុកព្យញ្ជនៈនោះជាចាំបាច់ ក្នុងន័យរក្សាអត្ថកិច្ចរបស់ពាក្យ និងជាគ្រឿងសម្គាល់ជាតិសំព័ន្ធរបស់ពាក្យ ។ ឧទាហរណ៍ ៖ ទូរទស្សន៍, ប្រយោជន៍, គមនាគមន៍, ត្រីក្រិកាណ៍, ពាក្យពេទ័ន៍, វេយ្យាករណ៍, ត្រ័យរតន៍ ។ គេប្រើទណ្ណយោគលើពាក្យក្លាយពីបាលី និងសំស្ក្រឹត ។ សញ្ញានេះ ក៏យកមកប្រើចំពោះពាក្យខ្លីពីបរទេសខ្លះ សម្រាប់រក្សាសំនាំអត្ថកិច្ចដើមរបស់ពាក្យទាំងនោះ ដូចជា ស៊ីម៉ងត៍, អេដស៍, អាស័រិច... ។

វិសដិទី (ំ) ឬ រះមុខ

គេប្រើសញ្ញានេះ សម្រាប់ប្រកបជំនួសឱ្យលំ “ ហ ” ជាពាក្យកាត់កំបុតខ្លី ហើយអានដោយលើកចំហោមដូចជា តិះ, វិះ, ចុះ, ចង្អុល (ចង្អុល) ឈ្មោះ, ទះ, ដោះ ដូរ, លោកផ្លោះ ។ ល។ វាមានសូរសំឡេងស្រដៀងនឹងពាក្យដែលប្រកបនឹងព្យញ្ជនៈ “ ស ” ដែរ ។

Figure 8 - Diacritics

64

យុគលពិន្ទុ(០:) ឬហៅថា “ ចុចពីរ ” ក៏បាន [៨]

គេប្រើវណ្ណយុគ្គិន្ទន្ទៈសម្រាប់ ៖ ដាក់នៅចុងពាក្យមានកំណើតពីបាលី ឬ សំស្ក្រឹត ដែលមានសូរ “ អាត ” ឬ “ អាក ” ដូចជា លោកៈ, មោហៈ, ទោសៈ...

- **ចំណាំ ៖** បើពាក្យទាំងនេះនៅខាងដើមគេ ដូចជា ពលកម្ម, វណ្ណកម្ម សច្ចភាព, ខេមរភាសា នោះ គេមិនចាំបាច់ដាក់យុគលពិន្ទុទេ ព្រោះពាក្យទាំងនេះជាសមាសនាម ។
- **លើកលែងតែពាក្យ ៖ គណៈ** ដែលទោះបីវានៅដើមគេ ឬក៏ពាក្យគេក្តី ក៏ត្រូវតែប្រើយុគលពិន្ទុជា ដលប ឧទាហរណ៍ គណៈប្រតិភូ, គណៈកម្មការ, គណៈសង្ឃ, គណៈរដ្ឋមន្ត្រី, គណៈប្រធានរដ្ឋ... ព្រោះថាក្នុងពាក្យទាំងនេះ “គណៈ” មិនចូលក្នុងសមាសនាមតាមក្បួនខ្នាតវេយ្យាករណ៍បាលី ។
- **លើកលែងតែពាក្យ ៖ គណបក្ស, គណនា, គណនេយ្យ...** ដែលគេមិនត្រូវប្រើយុគលពិន្ទុទេ ។
- **រឹងពាក្យ ៖** កាលៈទេសៈ គេត្រូវប្រើយុគលពិន្ទុទាំងពីរម៉ាកតែម្តង ។ ក៏ប៉ុន្តែ បើគេសរសេរពាក្យ នេះ “ បក្រាស ” មកវិញ “ ទេសកាល ” នោះ គេមិនត្រូវប្រើយុគលពិន្ទុឡើយ ។
- **ក្នុងសម័យនេះ** គេយកយុគលពិន្ទុមក ប្រើក្លាយក្នុងប្រយោគខ្លះដែរ សម្រាប់រាប់រាយឈ្មោះ អ្វីៗ ដូចពាក្យថា ទង់ជាតិបារាំងមាន ៣ ពណ៌គឺ ៖ ស, ខៀវ, ក្រហម ។ ប៉ុន្តែជាទូទៅ អ្នកនិពន្ធទាំង ឡាយតែងតែប្រើ “ ទ្វិពិន្ទុលេខ ” ក្នុងករណីរៀបរាប់, ឧទេស បង្ហាញ... ព្រោះការប្រើសញ្ញាខណ្ឌបែប នេះត្រឹមត្រូវ និងសមហេតុសមផលជាង ។ យើងនឹងបានឃើញព្រមទាំងអធិប្បាយពន្យល់ អំពី ខណ្ឌសញ្ញានេះនៅទំព័រខាងមុខ ។

Figure 9 – Diacritics

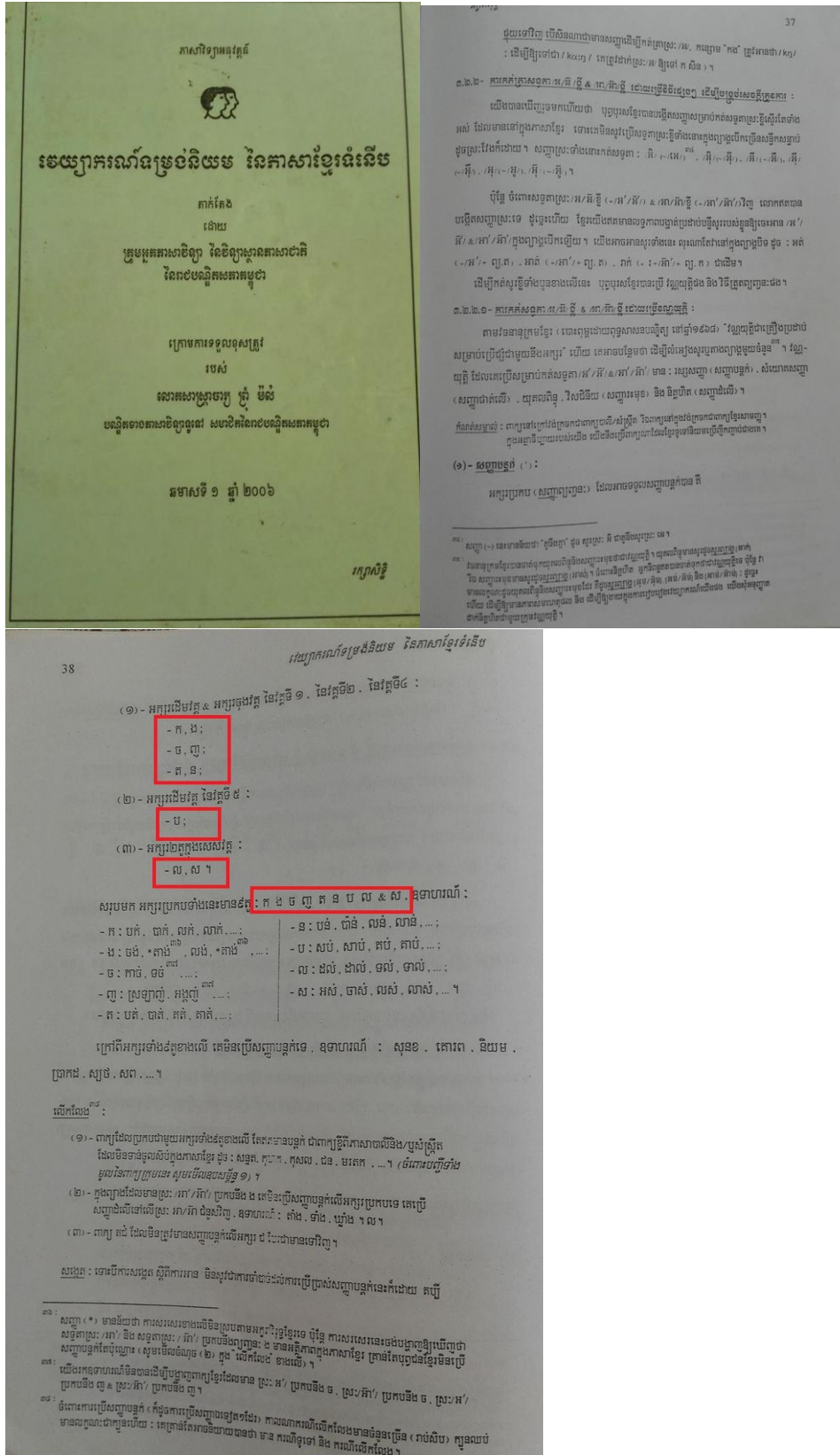


Figure 10 – Context of Bantoc Sign

Appendix 3: Initial Homoglyph Analysis by Khmer GP

Khmer-Thai homoglyphs

Khmer Letter	Thai Letter
ឥ U+ 178B Khmer letter TTHA	ช U + OE0A Thai letter CHO CHANG
ឥ U+ 178D Khmer letter TTHO	ฌ U + OE12 Thai letter THO PHUTHAO
ព U+ 1796 Khmer letter PO	ต U + OE15 Thai letter TO TAO
រ U+ 179A Khmer letter RO	ร U + OE23 Thai letter RO RUA
◌ U+ 17DO Khmer sign samyoksannya	◌ U + OE31 Thai letter MAIN HAN-AKAT
◌ U+ 17B6 Khmer vowel sign AA	า U + OE32 Thai letter SARA AA
◌ U+ 17B7 Khmer vowel sign I	◌ U + OE34 Thai letter SARA I
◌ U+ 17B8 Khmer vowel sign II	◌ U + OE35 Thai letter SARA II
◌ U+ 17B9 Khmer vowel sign Y	◌ U + OE36 Thai letter SARA UE
◌ U+ 17BA Khmer vowel sign YY	◌ U + OE37 Thai letter SARA UEE
◌ U+ 17BB Khmer vowel sign U	◌ U + OE38 Thai letter SARA U
◌ U+ 17BC Khmer vowel sign UU	◌ U + OE39 Thai letter SARA UU
◌ U+ 17C1 Khmer vowel sign E	เ U + OE40 Thai letter SARA E
◌ U+ 17CF Khmer sign AHSDA	◌ U + OE47 Thai letter MAITAIKHU
◌ U+ 17CB Khmer sign BANTOC	◌ U + OE48 Thai letter MAI EK
◌ U+ 17CD Khmer sign TOANDAKHIAT	◌ U + OE4C Thai letter THANTHAKHAT
◌ U+ 17C6 Khmer sign NIKAHIT	◌ U + OE4C Thai letter NIKAHIT

Khmer-Lao homoglyphs

Khmer Letter	Lao Letter
⊖ U+ 1781 Khmer letter KHA	ຂ U + OE82 Lao letter KHO SUNG
ប U+ 1794 Khmer letter BA	ບ U + OE9A Lao letter BO
ព U+ 1796 Khmer letter PO	ຕ U + OE95 Lao letter TO
ឃ U+ 1783 Khmer letter KHO	ຜ U + OE9E Lao letter PHO TAM
ຫ U+ 17A0 Khmer letter HA	ທ U + OE97 Lao letter THO TAM
ល U+ 179B Khmer letter LO	ລ U + OEA5 Lao letter LO
រ U+ 179A Khmer letter RO	ຣ U + OEA3 Lao letter RO
◌ U+ 17A7 Khmer independent vowel QU	ຊ U + OE8A Lao letter SO TAM
◌ U+ 17DO Khmer sign samyoksannya	◌ U + OEB1 Lao vowel sign MAI KAN
◌ U+ 17BB Khmer vowel sign U	◌ U + OEB8 Lao vowel sign U
◌ U+ 17BC Khmer vowel sign UU	◌ U + OEB9 Lao vowel sign UU
◌ U+ 17C1 Khmer vowel sign E	ເ U + OEEO Lao vowel sign E

◌ U+ 17CB Khmer sign BANTOC	◌ U + OEC8 Lao tone MAI EK
◌ U+ 17CD Khmer sign TOANDAKHIAT	◌ U + OECC Lao cancellation MARK
◌ U+ 17C6 Khmer sign NIKAHIT	◌ U + OECD Lao NIGGAHITA

Khmer-Burmese homoglyphs

Khmer Letter	Myanmar Letter
ប U+ 1794 Khmer letter BA	◻ U + 100E Myanmar letter DDHA
ក U+ 1783 Khmer letter KHO	◻ U + 1003 Myanmar letter GHA
យ U+ 1799 Khmer letter YO	◻ U + 101A Myanmar letter YA
ល U+ 179B Khmer letter LO	◻ U + 101C Myanmar letter LA
◌ U+ 17B6 Khmer vowel sign AA	◌ U + 106B Myanmar Sign Western Pwo Karen Tone 3
◌ U+ 17C6 Khmer sign NIKAHIT	◌ U + 1036 Myanmar Sign ANUSVARA
◌ U+ 17C7 Khmer sign REAHMUK	◌ U + 1038 Myanmar Sign VISARGA