

## Glyphs Recommended for a Khmer Unicode Font

Unicode <sup>1</sup>	Base glyph	Subscript glyph \u17D2 <sup>2</sup>	Ligature/Variants /Components <sup>3</sup>
\u1780 KHMER LETTER KA	ក	ក	ក ក ' ក
\u1781 KHMER LETTER KHA	ខ	ខ	ខ ខ ' ខ
\u1782 KHMER LETTER KO	គ	គ	គ គ ' គ
\u1783 KHMER LETTER KHO	ឃ	ឃ	ឃ ឃ ឃ ឃ ,
\u1784 KHMER LETTER NGO	ង	ង	ង ង ' ង
\u1785 KHMER LETTER CA	ច	ច	ច ច ' ច
\u1786 KHMER LETTER CHA	ឆ	ឆ	ឆ ច ' ឆ

<sup>1</sup> Unicode 3.0 (see Khmer subset in U1780.pdf; <http://www.bauhahn.m.clara.net/U1780.pdf>).

<sup>2</sup> Subscripts are formed by first typing KHMER SIGN COENG \u17D2 and then typing whatever base glyph should become a subscript. This applies both to consonants and independent vowels. All non-spacing vowels, signs, and subscripts should accept a variety of vertical spacings.

<sup>3</sup> Note that in ligatures there should not be a gap between the consonant and vowel, or between the slight curve on the top of the final part of vowel \u17C5 from its base part (unfortunately the small flag added in the font in this document is slightly shifted to the right **in some ligatures** so shows such an unsightly space). A ligating vowel should touch either at the right-most extreme of the 'hair' or 'hook' or at the right-most, top-most straight vertical part of the glyph (but not on the end of an upward swish as in \u1787, \u178A). Glyphs may be formed both from base consonant and spacing subscripts (those that come up beyond the base line to the right of the base consonant). Spacing subscripts may require variant forms for first and second subscript position.

\u1787 KHMER LETTER CO	ជ	េ	ជា ជាំ ' ជ៌ ជ័ <sup>4</sup>
\u1788 KHMER LETTER CHO	ឈ	ឺ	ឈា ឈាំ ឈ៌ ឈ័
\u1789 KHMER LETTER NYO	ញ <sup>5</sup>	្ល	ញា ញាំ ញ៌ ញ័ ញ
\u178A KHMER LETTER DA	ដ	ន	ដា ដាំ ' ដ៌
\u178B KHMER LETTER TTHA	ថ	ត	ថា ថាំ ' ថ៌
\u178C KHMER LETTER DO	ឌ	ឌ	ឌា ឌាំ ' ឌ៌
\u178D KHMER LETTER TTHO	ធ	ធ៌	ធា ធាំ ធ៌ ធ័
\u178E KHMER LETTER NNO	ណ <sup>6</sup>	ណ	ណា ណាំ ' ណ៌
\u178F KHMER LETTER TA	ត	ត	តា តាំ ' ត៌
\u1790 KHMER LETTER THA	ថ	ត	ថា ថាំ ' ថ៌

<sup>4</sup> These unusual superscript and subscript glyphs are used (in combination with those at \u17D4, \u17D3, and \u17E0 -> \u17E9) to represent lunar dates. **We need more documentation on the methods of representing lunar dates as well as accurate records/algorithms to convert between Gregorian and Khmer calendars past and present. These two glyphs may be misshapen...as they resemble the Khmer number eight more than anything else!**

<sup>5</sup> NYO drops its below-base glyph whenever a subscript (but not a below base dependent vowel) occurs below it. When NYO is a subscript to itself the **the** subscript form is shaped much like the base form (just smaller).

<sup>6</sup> Non-spacing subscripts to all other consonants line up to the right edge of the base consonant and for many in Cambodia that holds true for this consonant as well. On the other hand, in the Chhuan Nath dictionary all subscripts to the consonant NNO are left-right centered below the base consonant.

\u1791 KHMER LETTER TO	ទ	៤	ទា ទា ' ៧
\u1792 KHMER LETTER THO	ធ	៥	ធា ធា ' ៧
\u1793 KHMER LETTER NO	ន	៥	នា នា ' ៧
\u1794 KHMER LETTER BA	ប	៧	បា បា ៧៧ ៧៧
\u1795 KHMER LETTER PHA	ផ	៥	ផា ផា
\u1796 KHMER LETTER PO	ព	៥	ពា ពា
\u1797 KHMER LETTER PHO	ភ	៥	ភា ភា
\u1798 KHMER LETTER MO	ម	៥	មា មា
\u1799 KHMER LETTER YO	យ	៧	យា យា ៧៧៧៧ ៧៧
\u179A KHMER LETTER RO	រ <sup>៧</sup>	៧	រា <sup>៧</sup>
\u179B KHMER LETTER LO	ល	៥	លា លា
\u179C KHMER LETTER VO	វ	៥	វា វា
\u179D KHMER LETTER SHA	ឆ	៥	ឆា

<sup>7</sup> \u179A is unusual in that the subscript displays to the immediate left of the base consonant (even though its logical order is following that base consonant).

\u179E KHMER LETTER SSO	ថ	ថ	<sup>8</sup> មា
\u179F KHMER LETTER SA	ស	ស្ង	សា សា្ង ស្ង្ង
\u17A0 KHMER LETTER HA	ហា	ហ	ហា ហា
\u17A1 KHMER LETTER LA	ឡ		ឡា ឡា
\u17A2 KHMER LETTER QA	អ	អ	អា អា
\u17A3 KHMER INDEPENDENT VOWEL QAA	អ	អ	Independent vowels do not have ligatures or subscripts with <b>two exceptions</b> ឡ្ង <b>ឡ្ង</b>
\u17A4 KHMER INDEPENDENT VOWEL QAA	អា	អា	
\u17A5 KHMER INDEPENDENT VOWEL QJ	ត	ត	
\u17A6 KHMER INDEPENDENT VOWEL QII	ឡ្ង		
\u17A7 KHMER INDEPENDENT VOWEL QU	ឡ	ឡ	

---

<sup>8</sup> I have not seen examples of a subscript SSO, so am not sure which shape is appropriate.

\u17A8 KHMER INDEPENDENT VOWEL QUK	ខ្នំ	ខ្នំ	
\u17A9 KHMER INDEPENDENT VOWEL QUU	ខ្នុ	ខ្នុ	
\u17AA KHMER INDEPENDENT VOWEL QUUV	ខ្នួ	ខ្នួ	
\u17AB KHMER INDEPENDENT VOWEL RY	រ្យ	រ្យ	
\u17AC KHMER INDEPENDENT VOWEL RYY	រ្យ្យ	រ្យ្យ	
\u17AD KHMER INDEPENDENT VOWEL LY	ល្យ	ល្យ	
\u17AE KHMER INDEPENDENT VOWEL LYY	ល្យ្យ	ល្យ្យ	
\u17AF KHMER INDEPENDENT VOWEL QE	ឝ	ឝ	
\u17B0 KHMER INDEPENDENT VOWEL QAI	ឞ	ឞ	
\u17B1 KHMER INDEPENDENT VOWEL QOO TYPE ONE	ខ្នំ	ខ្នំ	

\u17B2 KHMER INDEPENDENT VOWEL QOO TYPE TWO			
\u17B3 KHMER INDEPENDENT VOWEL QAU	្រ	្រ	
\u17B4 KHMER VOWEL INHERENT AQ	្រ		Exceptional Pali/Sanskrit short inherent vowel <b>exceptional</b> (normally invisible [optionally visible])
\u17B5 KHMER VOWEL INHERENT AA	្រ		Regular Khmer long inherent vowel, <b>usual</b> (normally invisible [optionally visible])
\u17B6 KHMER VOWEL SIGN AA	្រ		្រ ្រ
\u17B7 KHMER VOWEL SIGN I	្រ		្រ
\u17B8 KHMER VOWEL SIGN II	្រ		
\u17B9 KHMER VOWEL SIGN Y	្រ		
\u17BA KHMER VOWEL SIGN YY	្រ		

---

<sup>9</sup> Ligatures or attachments of AA's left extremity varies according to the height of the connection point on the preceding consonant.

\u17BB KHMER VOWEL SIGN U	៊		10
\u17BC KHMER VOWEL SIGN UU	៊		
\u17BD KHMER VOWEL SIGN UA	៊		
\u17BE KHMER VOWEL SIGN OE	៊ ៊		៊ ៊
\u17BF KHMER VOWEL SIGN YA	៊ ៊		៊ ៊ ៊
\u17C0 KHMER VOWEL SIGN IE	៊ ៊		៊ ៊ ៊
\u17C1 KHMER VOWEL SIGN E	៊		
\u17C2 KHMER VOWEL SIGN AE	៊		
\u17C3 KHMER VOWEL SIGN AI	៊		
\u17C4 KHMER VOWEL SIGN OO	៊ ៊		៊ ៊
\u17C5 KHMER VOWEL SIGN AU	៊ ៊		៊ ៊ ៊
\u17C6 KHMER SIGN NIKAHIT	៊		
\u17C7 KHMER SIGN REAHMUK	៊		

<sup>10</sup> Note that a glyph like this may define one of three characters \u17BB, \u17C9, or \u17CA (as the latter two change to this shape and position when there is a surmounting sign or vowel above them in their regular configuration). Only the context will tell which is which (a significant problem in conversion of existing glyph-based text encodings).

\u17C8 KHMER SIGN YUUKALEAPIN TU	៖		
\u17C9 KHMER SIGN MUUSIKATOAN	័		័
\u17CA KHMER SIGN TRIISAP	័		័
\u17CB KHMER SIGN BANTOC	័		
\u17CC KHMER SIGN ROBAT	័		
\u17CD KHMER SIGN TOANDAKHIA T	័		័ <sup>11</sup>
\u17CE KHMER SIGN KAKABAT	័		
\u17CF KHMER SIGN AHSDA	័		
\u17D0 KHMER SIGN SAMYOK SANNYA	័		
\u17D1 KHMER SIGN VIRIAM	័		
\u17D2 KHMER SIGN COENG	័		(normally invisible [optionally or momentarily visible?])
\u17D3 KHMER SIGN BATHAMASAT	័		

---

<sup>11</sup> This is an unusual ligature which displays when the characters \u17B7 and \u17CD occur together.



\u17D4 KHMER SIGN KHAN	១		១ ១
\u17D5 KHMER SIGN BARIYOOSAN	១១		
\u17D6 KHMER SIGN CAMNUC PII KUUH	៖		
\u17D7 KHMER SIGN LEK TOO	១		
\u17D8 KHMER SIGN BEYYAL	១៧១		១៧១ ១ - ៧១-
\u17D9 KHMER SIGN PHNAEK MUAN	៖		
\u17DA KHMER SIGN KOOMUUT	៖		
\u17DB KHMER CURRENCY SYMBOL RIEL	៛		៛ ៛ ៛
\u17DC KHMER SIGN AVAKRAHASA NYA	៖		
\u17E0 KHMER DIGIT ZERO	០		០ ១០ ១០ ០
\u17E1 KHMER DIGIT ONE	១		១ ១១ ១១ ១
\u17E2 KHMER DIGIT TWO	២		២ ១២ ១២ ២
\u17E3 KHMER DIGIT THREE	៣		៣ ១៣ ១៣ ៣
\u17E4 KHMER DIGIT FOUR	៤		៤ ១៤ ១៤ ៤

\u17E5 KHMER DIGIT FIVE	៥		៥      ១៥      ១៥      ១៥
\u17E6 KHMER DIGIT SIX	៦		៦                      ៦
\u17E7 KHMER DIGIT SEVEN	៧		៧                      ៧
\u17E8 KHMER DIGIT EIGHT	៨		៨                      ៨
\u17E9 KHMER DIGIT NINE	៩		៩                      ៩
\u0030 DIGIT ZERO	0		
\u0031 DIGIT ONE	1		
\u0032 DIGIT TWO	2		
\u0033 DIGIT THREE	3		
\u0034 DIGIT FOUR	4		
\u0035 DIGIT FIVE	5		
\u0036 DIGIT SIX	6		
\u0037 DIGIT SEVEN	7		
\u0038 DIGIT EIGHT	8		
\u0039 DIGIT NINE	9		
\u0020 SPACE			(Character for phrase breaks)
\u2008 PUNCTUATION SPACE			(Same width as a period)
\u00A0 NO- BREAK SPACE			
\u2009 SPACE, THIN			

\u200B ZERO WIDTH SPACE			(normally invisible [optionally or momentarily visible as a dotted vertical line])(for invisible word break [optional visibility of dotted vertical line])
\u007E TILDA	~		
\u002D HYPHEN-MINUS	-		
\u002B PLUS SIGN	+		
\u002F SOLIDUS, SLASH	/		
\005C REVERSE SOLIDUS=BAC KSLASH	\		
(\u00F7) DIVISION SIGN	÷		
\u00D7 MULTIPLICATION SIGN	x		
(\u002A) ASTERISK	*		
\u2015 HORIZONTAL BAR	—		
(\u00AD) SOFT HYPHEN	-		
\u00AB GUILLEMET, LEFT POINTING	«		

\u00BB GUILLEMENT, RIGHT POINTING	»		
(\u2039) GUILLEMENT, LEFT POINTING SINGLE	‹		
(\u203A) GUILLEMENT, RIGHT POINTING SINGLE	›		
\u2018 LEFT SINGLE QUOTATION MARK	‘		
\u2019 RIGHT SINGLE QUOTATION MARK	’		
\u201C LEFT DOUBLE QUOTATION MARK	“		
\u201D RIGHT DOUBLE QUOTATION MARK	”		
\u0028 LEFT PARENTHESIS	(		
\u0029 RIGHT PARENTHESIS	)		
(\u005B) LEFT SQUARE BRACKETS	[		
(\u005D) RIGHT SQUARE BRACKETS	]		
\u2022 BULLET	•		
\u002E FULL STOP	.		

\u002C COMMA	,		
(\u003A) COLON	:		(Not to be confused with \u17C8)
\u003B SEMICOLON	;		
\u0021 EXCLAMATION MARK	!		
\u003F QUESTION MARK	?		
\u0040 COMMERCIAL AT	@		
\u0023 NUMBER SIGN	#		
\u0026 PERCENT SIGN	%		
\u0024 DOLLAR SIGN	\$		
\u00A5 YEN SIGN	¥		
\u00A3 POUND SIGN	£		
\u0E3F CURRENCY SYMBOL BAHT, THAI			
\u20AD CURRENCY, LAOTIAN			
\u20AB CURRENCY, VIETNAMESE			

For corrections and additions please contact:  
Maurice Bauhahn, 2 Meadow Way; Dorney Reach; MAIDENHEAD SL6  
ODS; U.K. Email: [bauhahn@clara.net](mailto:bauhahn@clara.net);  
Tel: +44(0)1628 626068  
3 February 2001 Version 0.2 beta

I would particularly appreciate comments on how to avoid excessive foreign glyphs...while maintaining a set of the most commonly used ones in Khmer text. A complex **scenario**: In Khmer geometry texts a circumflex-like sign is spread over two or more Latin alphabet capitals to indicate an angle. Any idea of how to encode that?

Note that there is only one Khmer vowel in each cluster/quantum (all the characters surrounding and including a base character). Hence there should be only one keystroke for characters `\u17A3-\u17C5`. The Khmer signs `\u17C6-\u17C7` should always follow any existing vowels. It is an anomaly of Khmer that these vowel-sign constructs function as unified vowels in many different combinations (at least for sorting purposes).

Once a complete set of characters is represented, a keyboard can be constructed. Necessarily it will be quite different from existing glyph-encoding based keyboards (**See my paper, Principles of Khmer Keyboards**). As much as possible the modifications should strive to put the most frequently used characters (such as `\u17D2`) in the home row or other easily accessible keys. I have done extensive statistical studies of Khmer letter frequencies...and need to either extract those or do a new set. I believe common usage would dictate that the resultant keyboard would resemble the Query phonetic layout. Not my preference, incidentally, just reality.

Maurice Bauhahn; 2 Meadow Way; Dorney Reach; MAIDENHEAD SL6 0DS; U.K. Email: [bauhahnm@clara.net](mailto:bauhahnm@clara.net) 21 February 2001