

A.4.1F SMS/CBS 7-bit Khmer Alphabet (0x1F)

A.4.1F.1 SMS/CBS 7-bit Khmer Alphabet (0x1F) base table

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	៖ 17C8	ញ 17B0	SP 0020	0 0030	ផ 1795	ា 17B6	័ 17CB	p 0070
0	0	0	1	1	័ 17C6	— 005F	ដ 178A	1 0031	ព 1796	័ 17B7	a 0061	q 0071
0	0	1	0	2	័ 17C7		ឌ 178C	2 0032	ក 1797	័ 17B8	b 0062	r 0072
0	0	1	1	3	អ 17A2	ឌ 17B1	# 0023	3 0033	ម 1798	័ 17BB	c 0063	s 0073
0	1	0	0	4	អា 17A4	ឌ 17B3		4 0034	យ 1799	័ 17BC	d 0064	t 0074
0	1	0	1	5	ក 17A5	ក 1780	ប 178B	5 0035	រ 179A	័ 17B9	e 0065	u 0075
0	1	1	0	6	ក 17A6	ខ 1781	ឈ 178D	6 0036	ល 179B	័ 17BA	f 0066	v 0076
0	1	1	1	7	ខ 17A7	គ 1782	ណ 178E	7 0037	ឡ 17A1	័ 17C9	g 0067	w 0077
1	0	0	0	8	ខ 17A9	យ 1783	ត 178F	8 0038		័ 17C1	h 0068	x 0078
1	0	0	1	9	ឡ 17AB	ង 1784	ថ 1790	9 0039	រ 179C	័ 17C2	i 0069	y 0079
1	0	1	0	A	LF 000A	ច 1785	* 002A	: 003A	ស 179F	័ 17C3	j 006A	z 007A
1	0	1	1	B	ញ 17AD	SS2 ----	+ 002B	; 003B	ប 179E	័ 17CA	k 006B	័ 17CE
1	1	0	0	C		ង 1786	, 002C	ផ 1792	គ 179D		l 006C	័ 17CF
1	1	0	1	D	CR 000D	ង 1787	ទ 1791	ន 1793	ហ 17A0	័ 17C4	m 006D	័ 17DD
1	1	1	0	E	ង 17AF	ឈ 1788	. 002E	ប 1794	័ 17CC	័ 17C5	n 006E	័ 17D6
1	1	1	1	F		ញ 1789	/ 002F	? 003F	័ 17D0	័ 17D2	o 006F	ញ 17D7

In the event that an MS receives a code where a symbol is not represented in the above table then the MS shall display the REPLACEMENT CHARACTER (U+FFFD). Note that for the Khmer multipart vowels, not all components are encoded as characters, and all Khmer multipart vowels lack canonical decompositions.

LF: LINE FEED; moves to the next line (with implied carriage return).

CR: CARRIAGE RETURN; CR is not used as CR but is used as a filler after the actual SMS/CBS message text. CR should not occur inside a message text, but if it does, it must be converted to LF. Note though that <SS2,CR> is actually CSI (see next table), and that "CR" is not a filler, nor LF.

SS2: SINGLE SHIFT TWO; This code shifts the next 7-bit code unit to refer to an extension of this table (subclause A.4.1F.2). (Note: setting, in the SMS and CBS protocols, base or extension table with reference value 0x10 or larger automatically sets the other tables with the same reference value.) When splitting a message text into submessages, there must be no cut right after an SS2.

U+17D2: KHMER SIGN COENG: This is a control character for forming conjuncts. If COENG is not between two Khmer consonants, U+17D2 is a combining character without nominal glyph (but use an underscript + as suggested in charts).

A.4.1F.2 SMS/CBS 7-bit Khmer Alphabet (0x1F) extension (SS2) table

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1	1B	0	1	2	3	4	5	6	7
0	0	0	0	0	@ 0040	< 003C	ឆ 17E4	ឌ 17A8	 007C	P 0050	¼ 00BC	△ 26A0
0	0	0	1	1	£ 00A3	= 003D	ឝ 17E5	ឍ 17AA	A 0041	Q 0051	½ 00BD	∞ 221E
0	0	1	0	2	\$ 0024	> 003E	ប 17E6	ឡ 17AC	B 0042	R 0052	¾ 00BE	² 00B2
0	0	1	1	3	¥ 00A5	₹ 20B9	ព 17E7	ឱ 17AE	C 0043	S 0053	“ 201C	³ 00B3
0	1	0	0	4	§ 00A7	^ 005E	ឝ 17E8	ឱ 17B2	D 0044	T 0054	” 201D	§ 17DC
0	1	0	1	5	" 0022	฿ 0E3F	ឝ 17E9		E 0045	U 0055	¤ 00A4	₹ 17DB
0	1	1	0	6	€ 20AC	₭ 20AD	◌ 17D9	ឱ 17BF	F 0046	V 0056	Ω 03A9	អ 17A3
0	1	1	1	7	% 0025	(0028	◌ 17DA	ឱ 17C0	G 0047	W 0057	μ 03BC	
1	0	0	0	8	& 0026) 0029	{ 007B	ឱ 17BE	H 0048	X 0058	° 00B0	
1	0	0	1	9	' 0027	្ក 17D4	} 007D	ឱ 17BD	I 0049	Y 0059	× 00D7	
1	0	1	0	A	FF 000C	្ក 17D5	្ក 17E1,2044,17E4	≈ 2248	J 004A	Z 005A	÷ 00F7	
1	0	1	1	B	- 2212	SS3 ----	្ក 17E1,2044,17E2	≠ 2260	K 004B	◌ 17CD	± 00B1	
1	1	0	0	C	! 0021	◌ 17E0	្ក 17E3,2044,17E4	[005B	L 004C		‰ 2030	
1	1	0	1	D	CSI 009B	្ក 17E1	≤ 2A7D	~ 007E	M 004D		« 00AB	
1	1	1	0	E	- 002D	្ក 17E2	≥ 2A7E] 005D	N 004E		» 00BB	
1	1	1	1	F		្ក 17E3	\ 005C	· 00B7	O 004F	◌ 17D1	' 2019	

In the event that an MS receives a code where a symbol is not represented in the above table then the MS shall display the REPLACEMENT CHARACTER (U+FFFD).

FF: FORM FEED; This code is defined as a Page Break character (with implied carriage return). Any mobile station which does not have pagination shall treat it as a LF (with implied carriage return).

CSI: CONTROL SEQUENCE INTRODUCER; enables decimal character references, enables styling. An MS which does not support CSI shall convert it to REPLACEMENT CHARACTER.

SS3: SINGLE SHIFT THREE; This code is reserved for the extension to another extension table. When splitting a message text into submessages, there must be no cut right after an SS3.

U+2044: FRACTION SLASH; the sequence may need to be isolated from adjacent digits by ZWNJ.

U+17D1: KHMER SIGN VIRIAM: Virama for final consonants. U+17D1 is a combining character with nominal glyph