

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
0	0	0	0	0	ং: 17C8	ং 17B0	SP 0020	0 0030
0	0	0	1	1	ঁ 17C6	ঁ 005F	178A	1 0031
0	0	1	0	2	ং: 17C7	ং 178C	2 0032	ং 1797
0	0	1	1	3	ঃ 17A2	ঃ 17B1	# 0023	ঃ 0033
0	1	0	0	4	ঃ 17A4	ঃ 17B3	4 0034	ঃ 1799
0	1	0	1	5	ঃ 17A5	ঃ 1780	5 0035	ঃ 179A
0	1	1	0	6	ঃ 17A6	ঃ 1781	6 0036	ঃ 179B
0	1	1	1	7	ঃ 17A7	ঃ 1782	7 0037	ঃ 17A1
1	0	0	0	8	ঃ 17A9	ঃ 1783	8 0038	ঃ 17C1
1	0	0	1	9	ঃ 17AB	ঃ 1784	9 0039	ঃ 179C
1	0	1	0	A	LF 000A	ঃ 1785	* 002A	: 003A
1	0	1	1	B	ং 17AD	SS2 ----	+ 002B	; 003B
1	1	0	0	C	ং 1786	, 002C	ং 1792	ং 179D
1	1	0	1	D	CR 000D	ং 1787	S 1791	S 1793
1	1	1	0	E	ং 17AF	ং 1788	. 002E	ং 1794
1	1	1	1	F	ং 1789	/ 002F	? 003F	ং 17D0

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+FFFFD). 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).

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	ើ 17E4	ិ 17A8	 007C	P 0050	¼ 00BC	△ 26A0	
0	0	0	1	1	£ 00A3	= 003D	ី 17E5	ិ 17AA	A 0041	Q 0051	½ 00BD	∞ 221E	
0	0	1	0	2	\$ 0024	> 003E	ុ 17E6	ុ 17AC	B 0042	R 0052	¾ 00BE	² 00B2	
0	0	1	1	3	¥ 00A5	₹ 20B9	ុ 17E7	ុ 17AE	C 0043	S 0053	“ 201C	³ 00B3	
0	1	0	0	4	§ 00A7	^ 005E	ុ 17E8	ុ 17B2	D 0044	T 0054	” 201D	s 17DC	
0	1	0	1	5	" 0022	฿ 0E3F	ុ 17E9		E 0045	U 0055	¤ 00A4	៥ 17DB	
0	1	1	0	6	€ 20AC	₭ 20AD	០ 17D9	ិំ 17BF	F 0046	V 0056	Ω 03A9	៩ 17A3	
0	1	1	1	7	% 0025	(0028	០ 17DA	ិំ 17C0	G 0047	W 0057	μ 03BC		
1	0	0	0	8	& 0026) 0029	{ 007B	ិំ 17BE	H 0048	X 0058	° 00B0		
1	0	0	1	9	' 0027	៤ 17D4	} 007D	ិំ 17BD	I 0049	Y 0059	× 00D7		
1	0	1	0	A	FF 000C	៤ 17D5	ិំ 17E1,2044,17E4	≈ 2248	J 004A	Z 005A	÷ 00F7		
1	0	1	1	B	- 2212	SS3 ----	ិំ 17E1,2044,17E2	≠ 2260	K 004B	ጀ 17CD	± 00B1		
1	1	0	0	C	! 0021	០ 17E0	ិំ 17E3,2044,17E4	[005B	L 004C		% 2030		
1	1	0	1	D	CSI 009B	៤ 17E1	≤ 2A7D	~ 007E	M 004D		« 00AB		
1	1	1	0	E	- 002D	ិ 17E2	≥ 2A7E] 005D	N 004E		» 00BB		
1	1	1	1	F		ិ 17E3	\ 005C	.	O 004F	ጀ 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