

A.4.10 SMS/CBS 7-bit European (Latin script) alphabet (0x10) tables

This alphabet for Latin should be preferred over the 7-bit default alphabet (0x00). This (0x10) alphabet supports most European languages that are using the Latin script, and also Vietnamese and Tagalog. For some letters with marks, combining characters are used to represent them. This alphabet should be used instead of the default encoding (0x00) except when the default encoding is explicitly called for, and except for USSD.

A.4.10.1 SMS/CBS 7-bit European (Latin script)alphabet (0x10) 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	1	2	3
0	0	0	0	0	@ 0040	- 2212	SP 0020	0 0030
0	0	0	1	1	£ 00A3	005F	! 0021	1 0031
0	0	1	0	2	\$ 0024	0300	" 0022	2 0032
0	0	1	1	3	¥ 00A5	0301	# 0023	3 0033
0	1	0	0	4	è 00E8	0309	€ 20AC	4 0034
0	1	0	1	5	é 00E9	0323	% 0025	5 0035
0	1	1	0	6	ù 00F9	0303	& 0026	6 0036
0	1	1	1	7	í 00EC	0307	' 0027	7 0037
1	0	0	0	8	ò 00F2	0304	(0028	8 0038
1	0	0	1	9	ç 00E7	0302) 0029	9 0039
1	0	1	0	A	LF 000A	030C	* 002A	: 003A
1	0	1	1	B	Ø 00D8	SS2 ----	; 002B	; 003B
1	1	0	0	C	ø 00F8	Æ 00C6	, 002C	< 003C
1	1	0	1	D	CR 000D	æ 00E6	- 002D	= 003D
1	1	1	0	E	À 00C5	Ù 00DF	. 002E	> 003E
1	1	1	1	F	å 00E5	É 00C9	/ 002F	? 003F

A.4.10.2 SMS/CBS 7-bit European (Latin script) alphabet (0x10) extension (SS2) table

				b7	0	0	0	0	1	1	1	1	1
				b6	0	0	1	1	0	0	1	1	1
				b5	0	1	0	1	0	1	0	1	1
b4	b3	b2	b1	1B	0	1	2	3	4	5	6	7	
0	0	0	0	0	Ó 0150	ª 00AA	œ 01A1	À 0104	Í 007C	Œ 0152	¼ 00BC	À 26A0	
0	0	0	1	1	ő 0151	º 00BA	ú 01AF	à 0105	Ć 0106	Ě 011A	½ 00BD	∞ 221E	
0	0	1	0	2	Ű 0170	đ 011F	ë 01B0	€ 0118	ć 0107	đ 011E	¾ 00BE	² 00B2	
0	0	1	1	3	ű 0171	ş 015F	º 0306	€ 0119	ł 0141	ſ 015E	“ 201C	³ 00B3	
0	1	0	0	4	È 00C8	^ 005E	‘ 0328	þ 00DE	ł 0142	ě 011B	” 201D	ı 0131	
0	1	0	1	5	i 0069,0307	Č 010C	,	þ 00FE	ń 0143	ú 00DA	¤ 00A4	ú 00FA	
0	1	1	0	6	Ù 00D9	č 010D	,	đ 00D0	ń 0144	í 00CD	Ω 03A9	í 00ED	
0	1	1	1	7	Ì 00CC	Š 0160	,	ð 00F0	ś 015A	â 00C2	μ 03BC	â 00E2	
1	0	0	0	8	Ò 00D2	š 0161	{	đ 0110	ś 015B	Ê 00CA	° 00B0	ê 00EA	
1	0	0	1	9	Ç 00C7	Ž 017D	}	đ 0111	ż 0179	î 00CE	× 00D7	î 00EE	
1	0	1	0	A	FF 000C	ž 017E	,	≈ 2248	ż 017A	û 00DB	÷ 00F7	ú 00FB	
1	0	1	1	B	Ô 00D4	SS3 ----	Ý 00DD	≠ 2260	ü 016E	ë 00CB	± 00B1	ë 00EB	
1	1	0	0	C	ô 00F4	Ă 0102	ý 00FD	[005B	ú 016F	ï 00CF	% 2030	ï 00EF	
1	1	0	1	D	CSI 009B	ă 0103	≤ 2A7D	~ 007E	À 00C0	Ā 00C3	« 00AB	ã 00E3	
1	1	1	0	E	Á 00C1	Þ 1E9E	≥ 2A7E] 005D	í 0130	õ 00D5	» 00BB	õ 00F5	
1	1	1	1	F	á 00E1	ó 01A0	\ 005C	.	œ 0153	ó 00D3	,	ó 2019	ó 00F3
FF: FORM FEED; This code is defined as a Page Break character (with implied carriage return). A final recipient of the SMS/CBS message which does not have pagination shall convert FF to LF.													
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 (SS3) extension table (subclause A.4.10.3), automatically set along with setting this table. When splitting a message text into submessages, there must be no cut right after an SS3.													

Though labelled “European”, at least also Turkish, Azeri, Vietnamese, and Tagalog/Filipino can be written using this alphabet.

A.4.10.3 SMS/CBS 7-bit European (Latin script) alphabet (0x10) extension (SS3) 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			NNBSP 202F					
0	0	0	1	1								
0	0	1	0	2								
0	0	1	1	3								
0	1	0	0	4				ቂ 20BA		ቁ 0166		ቃ 0167
0	1	0	1	5								
0	1	1	0	6								
0	1	1	1	7								
1	0	0	0	8					ሮ 0126		ሮ 0127	
1	0	0	1	9								
1	0	1	0	A	UND ----	□ FFFD						
1	0	1	1	B		UND ----						
1	1	0	0	C								
1	1	0	1	D	UND ----		- 2010					
1	1	1	0	E					ঠ 014A		ঠ 014B	
1	1	1	1	F								

In the event that an MS receives a code where a symbol is not represented (including UND) in the above table then the MS shall display the REPLACEMENT CHARACTER (U+FFF4).

U+FFF4: Explicit REPLACEMENT CHARACTER (equivalent in function to SUBSTITUTE, U+001A).

NNBSP: NARROW NO-BREAK SPACE, for recommended space before/after some punctuation in French, as well as space between a number and a fraction (like in 2 ½), a number and %, a number and a unit symbol (like in 40 m), esp. to get a space but still avoid an automatic line break. This character is at this position in all of the 7-bit alphabets for SMS/CBS with reference code 0x10 or higher.

U+2010: HYPHEN, a true hyphen, in contrast to U+002D HYPHEN-MINUS. This character is at this position in all of the 7-bit alphabets for SMS/CBS with reference code 0x10 or higher.