

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	4	5	6	7
0	0	0	0	0	@ 0040	- 2212	SP 0020	0 0030	i 00A1	P 0050	ç 00BF	p 0070
0	0	0	1	1	£ 00A3	— 005F	! 0021	1 0031	A 0041	Q 0051	a 0061	q 0071
0	0	1	0	2	\$ 0024	0300	" 0022	2 0032	B 0042	R 0052	b 0062	r 0072
0	0	1	1	3	¥ 00A5	0301	# 0023	3 0033	C 0043	S 0053	c 0063	s 0073
0	1	0	0	4	è 00E8	0309	€ 20AC	4 0034	D 0044	T 0054	d 0064	t 0074
0	1	0	1	5	é 00E9	0323	% 0025	5 0035	E 0045	U 0055	e 0065	u 0075
0	1	1	0	6	ù 00F9	0303	& 0026	6 0036	F 0046	V 0056	f 0066	v 0076
0	1	1	1	7	ì 00EC	0307	' 0027	7 0037	G 0047	W 0057	g 0067	w 0077
1	0	0	0	8	ò 00F2	0304	(0028	8 0038	H 0048	X 0058	h 0068	x 0078
1	0	0	1	9	ç 00E7	0302) 0029	9 0039	I 0049	Y 0059	i 0069	y 0079
1	0	1	0	A	LF 000A	030C	* 002A	: 003A	J 004A	Z 005A	j 006A	z 007A
1	0	1	1	B	Ø 00D8	SS2 ----	+ 002B	; 003B	K 004B	Ä 00C4	k 006B	ä 00E4
1	1	0	0	C	ø 00F8	Æ 00C6	, 002C	< 003C	L 004C	Ö 00D6	l 006C	ö 00F6
1	1	0	1	D	CR 000D	æ 00E6	- 002D	= 003D	M 004D	Ñ 00D1	m 006D	ñ 00F1
1	1	1	0	E	À 00C5	ß 00DF	. 002E	> 003E	N 004E	Ü 00DC	n 006E	ü 00FC
1	1	1	1	F	å 00E5	É 00C9	/ 002F	? 003F	O 004F	Š 00A7	o 006F	à 00E0

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 the extension of this table (subclause A.4.10.2). (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+00DF: Uppercase is U+1E9E (see the extension table).

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
					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	Œ	ª	º	À		Œ	¼	Δ	
					0150	00AA	01A1	0104	007C	0152	00BC	26A0	
0	0	0	1	1	đ	°	Ů	ą	Č	Ě	½	∞	
					0151	00BA	01AF	0105	0106	011A	00BD	221E	
0	0	1	0	2	Ů	ǧ	ı	Ę	ć	Ǧ	¾	²	
					0170	011F	01B0	0118	0107	011E	00BE	00B2	
0	0	1	1	3	ů	ş	˘	ę	ł	Ş	“	³	
					0171	015F	0306	0119	0141	015E	201C	00B3	
0	1	0	0	4	È	^	˘	þ	ı	ě	”	ı	
					00C8	005E	0328	00DE	0142	011B	201D	0131	
0	1	0	1	5	ı	Č	˘	þ	Ň	Ú	¤	ú	
					0069,0307	010C	0326	00FE	0143	00DA	00A4	00FA	
0	1	1	0	6	Ù	č	˘	Ð	ń	Í	Ω	í	
					00D9	010D	0327	00D0	0144	00CD	03A9	00ED	
0	1	1	1	7	İ	Š	˘	ð	Ś	Â	μ	â	
					00CC	0160	0308	00F0	015A	00C2	03BC	00E2	
1	0	0	0	8	Œ	š	{	Ð	ś	Ě	°	ê	
					00D2	0161	007B	0110	015B	00CA	00B0	00EA	
1	0	0	1	9	Ç	Ž	}	đ	Ž	Î	×	î	
					00C7	017D	007D	0111	0179	00CE	00D7	00EE	
1	0	1	0	A	FF	ž	˘	≈	ž	Û	÷	û	
					000C	017E	030B	2248	017A	00DB	00F7	00FB	
1	0	1	1	B	Ô	SS3	Ý	≠	Û	Ě	±	ë	
					00D4	----	00DD	2260	016E	00CB	00B1	00EB	
1	1	0	0	C	ô	Ä	ý	[ű	İ	‰	ï	
					00F4	0102	00FD	005B	016F	00CF	2030	00EF	
1	1	0	1	D	CSI	ă	≤	~	À	Ă	«	ă	
					009B	0103	2A7D	007E	00C0	00C3	00AB	00E3	
1	1	1	0	E	Á	ß	≥]	İ	Œ	»	ö	
					00C1	1E9E	2A7E	005D	0130	00D5	00BB	00F5	
1	1	1	1	F	á	Œ	\	·	œ	Ó	,	ó	
					00E1	01A0	005C	00B7	0153	00D3	2019	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				† 20BA		ƒ 0166		‡ 0167
0	1	0	1	5								
0	1	1	0	6								
0	1	1	1	7								
1	0	0	0	8					Ĥ 0126		ĥ 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					Ŋ 014A		ŋ 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+FFFD).

U+FFFD: 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.