

Table 1. Frequencies of the Characters in the French 10M Sample.

Character	CODES		Character class code	Count	Percent in total sample	Percent in char. class	Occurrence in total sample	Occurrence in char. class
	ASCII	Unicode						
Blank	32	20	S	1,671,340	16.71	81.13	5.98	1.23
e	101	65	L	1,105,003	11.05	14.44	9.05	6.93
s	115	73	L	619,161	6.19	8.09	16.15	12.36
a	97	61	L	583,054	5.83	7.62	17.15	13.12
n	110	6e	L	581,078	5.81	7.59	17.21	13.17
i	105	69	L	575,826	5.76	7.52	17.37	13.29
r	114	72	L	544,959	5.45	7.12	18.35	14.04
t	116	74	L	543,162	5.43	7.10	18.41	14.09
u	117	75	L	434,721	4.35	5.68	23.00	17.60
o	111	6f	L	414,118	4.14	5.41	24.15	18.48
l	108	6c	L	413,598	4.14	5.40	24.18	18.50
d	100	64	L	325,449	3.25	4.25	30.73	23.51
c	99	63	L	250,923	2.51	3.28	39.85	30.50
p	112	70	L	221,640	2.22	2.90	45.12	34.53
m	109	6d	L	204,772	2.05	2.68	48.83	37.37
é	233	e9	L	198,226	1.98	2.59	50.45	38.60
,	44	2c	S	125,283	1.25	6.08	79.82	16.44
v	118	76	L	98,959	0.99	1.29	101.05	77.33

Table 1. (continued)

Character	CODES		Character class code	Count	Percent in total sample	Percent in char. class	Occurrence in total sample	Occurrence in char. class
	ASCII	Unicode						
'	39	27	S	94,105	0.94	4.57	106.26	21.89
f	102	66	L	80,013	0.80	1.05	124.98	95.64
g	103	67	L	79,722	0.80	1.04	125.44	95.99
q	113	71	L	77,043	0.77	1.01	129.80	99.32
.	46	2e	S	68,701	0.69	3.33	145.56	29.99
b	98	62	L	62,356	0.62	0.81	160.37	122.72
h	104	68	L	57,169	0.57	0.75	174.92	133.85
à	224	e0	L	36,662	0.37	0.48	272.76	208.73
x	120	78	L	33,916	0.34	0.44	294.85	225.62
"	34	22	S	33,492	0.33	1.63	298.58	61.51
-	45	2d	S	29,633	0.30	1.44	337.46	69.52
è	232	e8	L	26,507	0.27	0.35	377.26	288.69
M	77	4d	U	22,386	0.22	10.49	446.71	9.54
L	76	4c	U	21,332	0.21	9.99	468.78	10.01
j	106	6a	L	21,067	0.21	0.28	474.68	363.24
y	121	79	L	20,623	0.21	0.27	484.90	371.06
C	67	43	U	18,633	0.19	8.73	536.68	11.46
P	80	50	U	15,888	0.16	7.44	629.41	13.44
A	65	41	U	15,621	0.16	7.32	640.16	13.67
l	49	31	N	14,898	0.15	20.45	671.23	4.89
E	69	45	U	13,233	0.13	6.20	755.69	16.13
S	83	53	U	13,166	0.13	6.17	759.53	16.21
0	48	30	N	12,195	0.12	16.74	820.01	5.97
è	234	ea	L	11,185	0.11	0.15	894.05	684.16
(40	28	S	10,368	0.10	0.50	964.51	198.69
)	41	29	S	10,340	0.10	0.50	967.12	199.23
D	68	44	U	10,136	0.10	4.75	986.58	21.06
9	57	39	N	9,592	0.10	13.17	1,042.54	7.59
F	70	46	U	9,484	0.09	4.44	1,054.41	22.51
R	82	52	U	9,270	0.09	4.34	1,078.75	23.03
I	73	49	U	9,025	0.09	4.23	1,108.03	23.65
B	66	42	U	8,714	0.09	4.08	1,147.58	24.50
8	56	38	N	7,289	0.07	10.01	1,371.93	9.99
z	122	7a	L	7,238	0.07	0.09	1,381.60	1,057.24
2	50	32	N	6,876	0.07	9.44	1,454.33	10.59
:	58	3a	S	6,668	0.07	0.32	1,499.70	308.95
T	84	54	U	6,248	0.06	2.93	1,600.51	34.17
G	71	47	U	6,038	0.06	2.83	1,656.18	35.36
k	107	6b	L	5,931	0.06	0.08	1,686.06	1,290.22
J	74	4a	U	5,634	0.06	2.64	1,774.94	37.89

Table 1. (*continued*)

Character	CODES		Character class code	Count	Percent in total sample	Percent in char. class	Occurrence in total sample	Occurrence in char. class
	ASCII	Unicode						
U	85	55	U	5,382	0.05	2.52	1,858.05	39.67
5	53	35	N	5,376	0.05	7.38	1,860.12	13.55
N	78	4e	U	5,363	0.05	2.51	1,864.63	39.81
3	51	33	N	4,808	0.05	6.60	2,079.87	15.15
O	79	4f	U	4,710	0.05	2.21	2,123.14	45.33
ç	231	e7	L	4,687	0.05	0.06	2,133.56	1,632.66
ô	244	f4	L	4,432	0.04	0.06	2,256.32	1,726.60
7	55	37	N	4,124	0.04	5.66	2,424.83	17.66
4	52	34	N	3,876	0.04	5.32	2,579.98	18.79
6	54	36	N	3,814	0.04	5.24	2,621.92	19.10
V	86	56	U	3,769	0.04	1.77	2,653.22	56.64
H	72	48	U	3,323	0.03	1.56	3,009.33	64.24
%	37	25	S	3,132	0.03	0.15	3,192.85	657.74
;	59	3b	S	3,056	0.03	0.15	3,272.25	674.10
?	63	3f	S	2,558	0.03	0.12	3,909.30	805.34
â	226	e2	L	2,315	0.02	0.03	4,319.65	3,305.53
ù	249	f9	L	2,186	0.02	0.03	4,574.57	3,500.59
û	251	fb	L	2,150	0.02	0.03	4,651.16	3,559.21
w	119	77	L	1,902	0.02	0.02	5,257.62	4,023.29
K	75	4b	U	1,682	0.02	0.79	5,945.30	126.92
Q	81	51	U	1,554	0.02	0.73	6,435.01	137.38
Ô	211	d3	O	1,321	0.01	99.85	7,570.02	1.00
W	87	57	U	1,150	0.01	0.54	8,695.65	185.64
!	33	21	S	972	0.01	0.05	10,288.07	2,119.39
Y	89	59	U	701	0.01	0.33	14,265.34	304.54
Z	90	5a	U	449	0.00	0.21	22,271.71	475.47
ë	235	eb	L	358	0.00	0.00	27,932.96	21,375.12
X	88	58	U	284	0.00	0.13	35,211.27	751.71
É	201	c9	U	267	0.00	0.13	37,453.18	799.57
+	43	2b	S	213	0.00	0.01	46,948.36	9,671.60
/	47	2f	S	138	0.00	0.01	72,463.77	14,927.91
ü	252	fc	L	93	0.00	0.00	107,526.88	82,282.72
ö	246	f6	L	55	0.00	0.00	181,818.18	139,132.60
*	42	2a	S	44	0.00	0.00	227,272.73	46,819.34
Ç	199	c7	U	43	0.00	0.02	232,558.14	4,964.77
ä	228	e4	L	25	0.00	0.00	400,000.00	306,091.72
=	61	3d	S	8	0.00	0.00	1,250,000.00	257,506.38
ï	239	ef	L	6	0.00	0.00	1,666,666.67	1,275,382.17
î	238	ee	L	3	0.00	0.00	3,333,333.33	2,550,764.33
æ	230	e6	O	2	0.00	0.15	5,000,000.00	661.50
Total				10,000,000				