

Table 11. Frequencies of the Characters in the Portuguese 2.5M Sample.

Character	CODES		Character class code	Count	Percent in total sample	Percent in in char. class	Occurrence in total sample	Occurrence in char. class
	ASCII	Unicode						
Blank	32	20	S	432,686	17.31	83.97	5.78	1.19
a	97	61	L	241,302	9.65	12.48	10.36	8.02
e	101	65	L	233,395	9.34	12.07	10.71	8.29
o	111	6f	L	205,115	8.20	10.61	12.19	9.43
s	115	73	L	153,257	6.13	7.92	16.31	12.62
r	114	72	L	130,778	5.23	6.76	19.12	14.79
i	105	69	L	125,351	5.01	6.48	19.94	15.43
n	110	6e	L	103,682	4.15	5.36	24.11	18.65
d	100	64	L	101,475	4.06	5.25	24.64	19.06
t	116	74	L	90,146	3.61	4.66	27.73	21.45
m	109	6d	L	86,902	3.48	4.49	28.77	22.26
u	117	75	L	77,522	3.10	4.01	32.25	24.95
c	99	63	L	67,463	2.70	3.49	37.06	28.67
l	108	6c	L	53,534	2.14	2.77	46.70	36.13
p	112	70	L	50,638	2.03	2.62	49.37	38.19
,	44	2c	S	36,116	1.44	7.01	69.22	14.27
v	118	76	L	29,109	1.16	1.51	85.88	66.44
.	46	2e	S	26,198	1.05	5.08	95.43	19.67
g	103	67	L	22,023	0.88	1.14	113.52	87.82
q	113	71	L	20,218	0.81	1.05	123.65	95.66
f	102	66	L	18,680	0.75	0.97	133.83	103.54
b	98	62	L	17,683	0.71	0.91	141.38	109.37
h	104	68	L	17,303	0.69	0.89	144.48	111.78

Table 11. (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						
â	226	e2	L	827	0.03	0.04	3,022.97	2,338.64
Q	81	51	U	811	0.03	1.73	3,082.61	57.96
l	49	31	N	755	0.03	20.72	3,311.26	4.83
U	85	55	U	672	0.03	1.43	3,720.24	69.94
H	72	48	U	623	0.02	1.33	4,012.84	75.45
0	48	30	N	601	0.02	16.49	4,159.73	6.06
É	201	c9	U	598	0.02	1.27	4,180.60	78.60
9	57	39	N	357	0.01	9.80	7,002.80	10.21
2	50	32	N	355	0.01	9.74	7,042.25	10.26
5	53	35	N	340	0.01	9.33	7,352.94	10.72
y	121	79	L	299	0.01	0.02	8,361.20	6,468.40
8	56	38	N	279	0.01	7.66	8,960.57	13.06
6	54	36	N	252	0.01	6.92	9,920.63	14.46
ü	252	fc	L	248	0.01	0.01	10,080.65	7,798.60
4	52	34	N	241	0.01	6.61	10,373.44	15.12
k	107	6b	L	239	0.01	0.01	10,460.25	8,092.27
7	55	37	N	236	0.01	6.48	10,593.22	15.44
3	51	33	N	228	0.01	6.26	10,964.91	15.98
Z	90	5a	U	223	0.01	0.47	11,210.76	210.78
X	88	58	U	183	0.01	0.39	13,661.20	256.85
'	39	27	S	133	0.01	0.03	18,796.99	3,874.44
w	119	77	L	118	0.00	0.01	21,186.44	16,390.28
K	75	4b	U	106	0.00	0.23	23,584.91	443.42
W	87	57	U	89	0.00	0.19	28,089.89	528.12
/	47	2f	S	47	0.00	0.01	53,191.49	10,963.83
_	95	5f	S	46	0.00	0.01	54,347.83	11,202.17
◊	186	ba	S	33	0.00	0.01	75,757.58	15,615.15
Á	193	c1	U	31	0.00	0.07	80,645.16	1,516.23
Y	89	59	U	29	0.00	0.06	86,206.90	1,620.79
%	37	25	S	23	0.00	0.00	108,695.65	22,404.35
\$	36	24	S	17	0.00	0.00	147,058.82	30,311.76
À	192	c0	U	17	0.00	0.04	147,058.82	2,764.88
Ô	212	d4	U	8	0.00	0.02	312,500.00	5,875.38
ä	170	aa	S	7	0.00	0.00	357,142.86	73,614.29
Ó	211	d3	U	5	0.00	0.01	500,000.00	9,400.60
=	61	3d	S	5	0.00	0.00	500,000.00	103,060.00
+	43	2b	S	3	0.00	0.00	833,333.33	171,766.67
[91	5b	S	3	0.00	0.00	833,333.33	171,766.67
]	93	5d	S	3	0.00	0.00	833,333.33	171,766.67
Ú	218	da	U	2	0.00	0.00	1,250,000.00	23,501.50

Table 11. (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						
ã	227	e3	L	16,911	0.68	0.87	147.83	114.37
ç	231	e7	L	11,890	0.48	0.61	210.26	162.66
á	225	e1	L	9,262	0.37	0.48	269.92	208.82
é	233	e9	L	9,084	0.36	0.47	275.21	212.91
z	122	7a	L	7,699	0.31	0.40	324.72	251.21
-	45	2d	S	7,417	0.30	1.44	337.06	69.48
í	237	ed	L	6,218	0.25	0.32	402.06	311.04
x	120	78	L	5,214	0.21	0.27	479.48	370.93
A	65	41	U	5,100	0.20	10.85	490.20	9.22
j	106	6a	L	4,995	0.20	0.26	500.50	387.20
E	69	45	U	4,819	0.19	10.25	518.78	9.75
ó	243	f3	L	4,212	0.17	0.22	593.54	459.18
è	234	ea	L	4,025	0.16	0.21	621.12	480.51
P	80	50	U	3,697	0.15	7.87	676.22	12.71
C	67	43	U	3,539	0.14	7.53	706.41	13.28
S	83	53	U	3,267	0.13	6.95	765.23	14.39
"	34	22	S	3,190	0.13	0.62	783.70	161.54
O	79	4f	U	3,040	0.12	6.47	822.37	15.46
M	77	4d	U	2,984	0.12	6.35	837.80	15.75
N	78	4e	U	2,626	0.11	5.59	952.02	17.90
õ	245	f5	L	2,421	0.10	0.13	1,032.63	798.87
D	68	44	U	2,264	0.09	4.82	1,104.24	20.76
)	41	29	S	1,921	0.08	0.37	1,301.41	268.25
à	224	e0	L	1,901	0.08	0.10	1,315.10	1,017.39
(40	28	S	1,850	0.07	0.36	1,351.35	278.54
ú	250	fa	L	1,834	0.07	0.09	1,363.14	1,054.55
B	66	42	U	1,736	0.07	3.69	1,440.09	27.08
!	33	21	S	1,616	0.06	0.31	1,547.03	318.87
T	84	54	U	1,608	0.06	3.42	1,554.73	29.23
?	63	3f	S	1,590	0.06	0.31	1,572.33	324.09
V	86	56	U	1,535	0.06	3.27	1,628.66	30.62
:	58	3a	S	1,530	0.06	0.30	1,633.99	336.80
R	82	52	U	1,341	0.05	2.85	1,864.28	35.05
F	70	46	U	1,326	0.05	2.82	1,885.37	35.45
J	74	4a	U	1,283	0.05	2.73	1,948.56	36.64
I	73	49	U	1,177	0.05	2.50	2,124.04	39.93
G	71	47	U	1,175	0.05	2.50	2,127.66	40.00
L	76	4c	U	1,088	0.04	2.31	2,297.79	43.20
ô	244	f4	L	1,079	0.04	0.06	2,316.96	1,792.45
;	59	3b	S	865	0.03	0.17	2,890.17	595.72

Table 11. (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						
Ê	202	ca	U	1	0.00	0.00	2,500,000.00	47,003.00
è	232	e8	L	1	0.00	0.00	2,500,000.00	1,934,053.00
>	62	3e	S	1	0.00	0.00	2,500,000.00	515,300.00
Total				2,500,000				