

# Procura-PALavras (P-PAL): Uma aplicação web com uma nova base lexical para o português europeu



Universidade do Minho, Escola de Psicologia  
Grupo de Investigação Psicolinguística

7º ENCONTRO NACIONAL DA ASSOCIAÇÃO PORTUGUESA DE PSICOLOGIA EXPERIMENTAL

Ana Paula Soares<sup>1</sup>, Ana Costa<sup>1</sup>, João Machado<sup>1</sup>, Álvaro Iriarte<sup>2</sup>, Alberto Simões<sup>3</sup>, José João de Almeida<sup>3</sup>, Patrícia França<sup>1</sup> & Montserrat Comesaña<sup>1</sup>

{<sup>1</sup>Escola Psicologia, <sup>2</sup>Instituto de Letras e Ciências Humanas, <sup>3</sup>Departamento de Informática}, Universidade do Minho

O Procura-PALavras (P-PAL) é uma aplicação web que permite a obtenção de uma grande diversidade de métricas lexicais e sublexicais (incluindo diversas medidas *type* e *token* de frequência, morfossintáticas e de estrutura e de similaridade ortográfica, fonológica, fonográfica e silábica) para ≈208.000 formas e ≈53.000 lemas do português europeu contemporâneo extraídos de um *corpus* de grandes dimensões (superior a 225 milhões de palavras). Ao usar o P-PAL o utilizador poderá fazer uma utilização personalizada do programa ao seleccionar as métricas que se adequem aos propósitos da sua investigação entre duas funcionalidades (obter palavras que obedecem a determinados requisitos ou avaliar palavras em determinados requisitos) e tipos de pesquisa (formas ou lemas). Neste trabalho apresentamos algumas das métricas de estrutura e de similaridade fonético-fonológicas já disponíveis na aplicação.

**Analisar palavras:** nesta funcionalidade o utilizador submete listas de palavras à aplicação (ficheiro .txt, .csv ou .xls) com vista à análise/obtenção das suas propriedades objectivas (lexicais/sublexicais) e/ou subjectivas seleccionadas do menu de análises.

**Obter palavras:** nesta funcionalidade o utilizador obtém listas de palavras que obedecem às propriedades objectivas (lexicais/sublexicais) e/ou subjectivas seleccionadas do menu de análises. Poderá ainda descarregar o ficheiro em formato .csv ou.xls.

OU

**Analisar palavras**      **Obter palavras**

+

**Pesquisa por:**

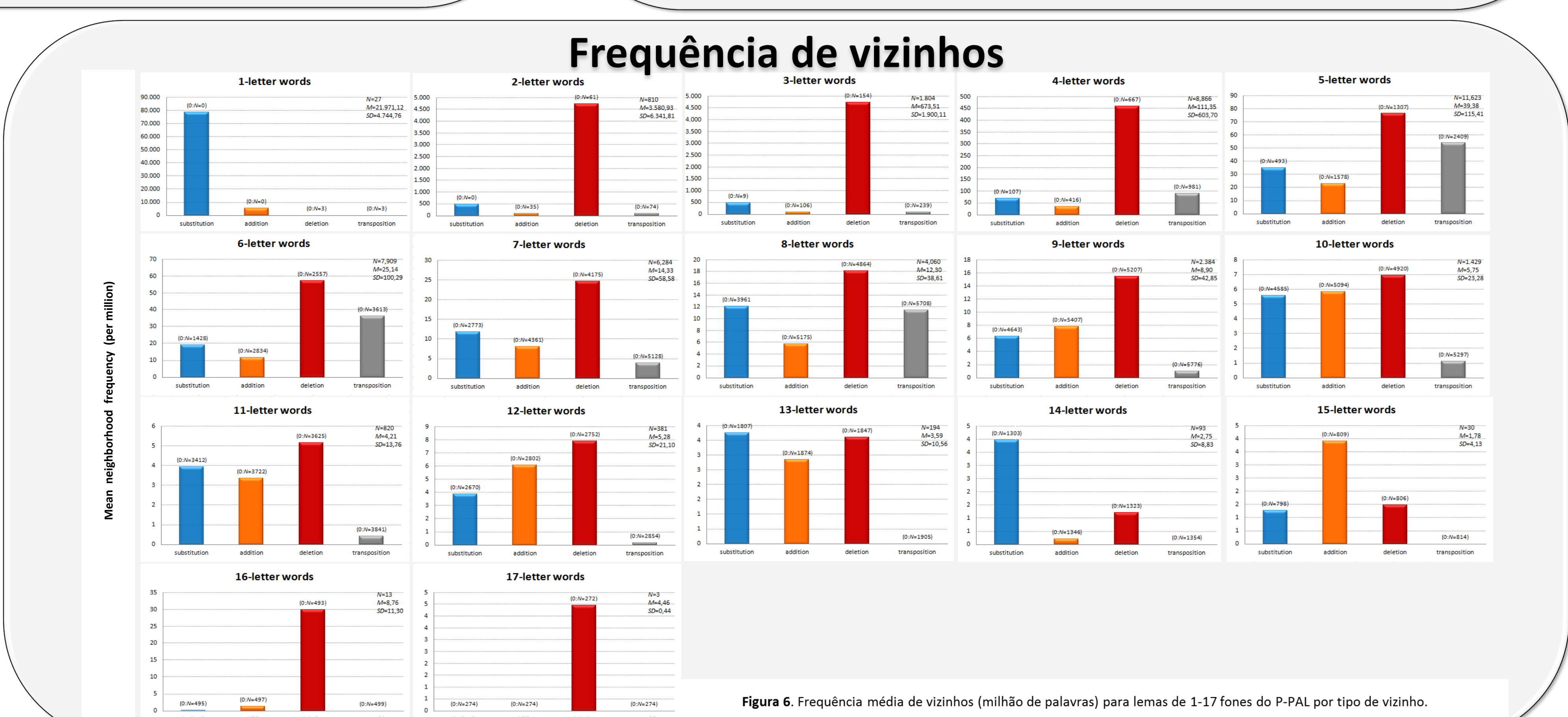
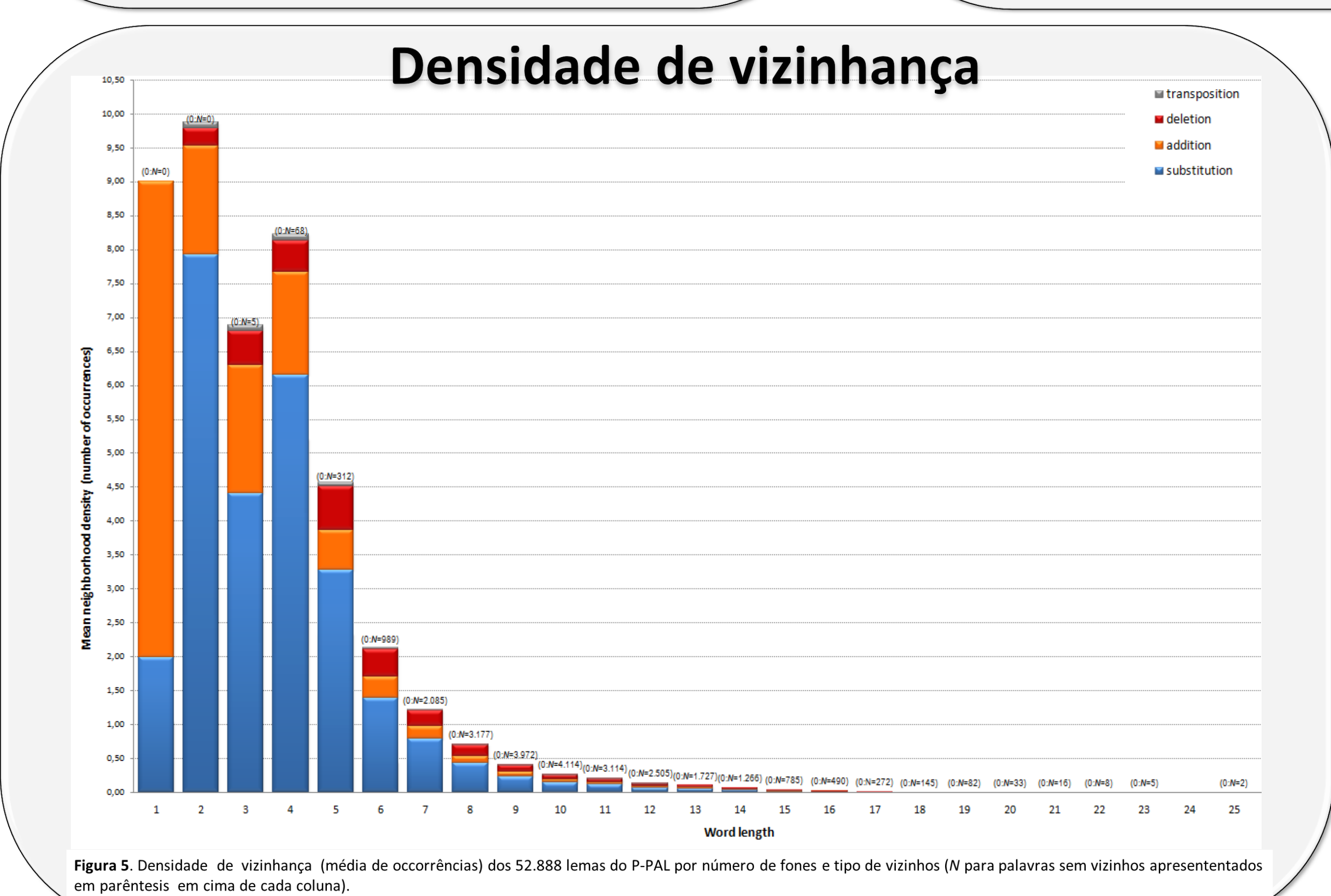
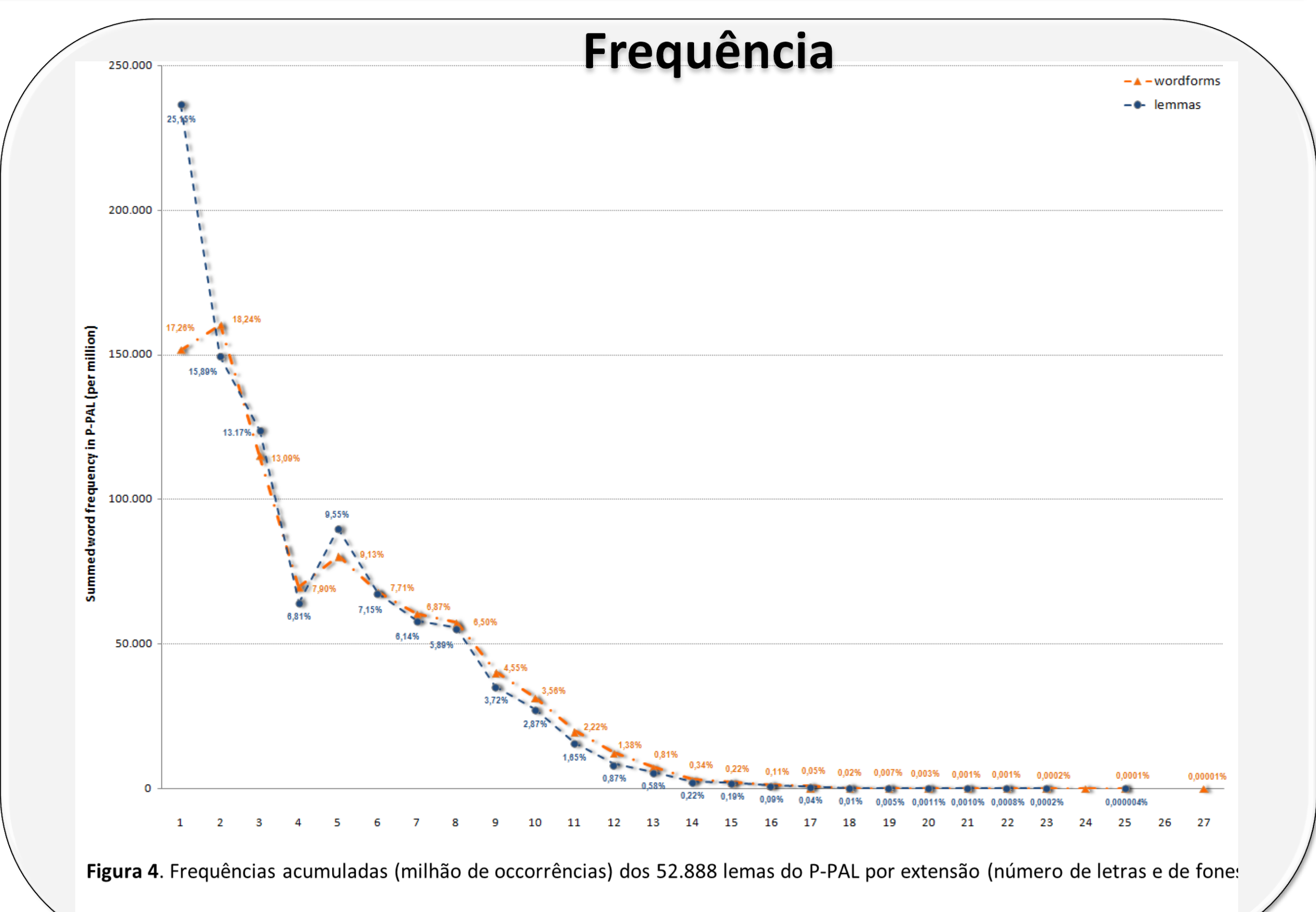
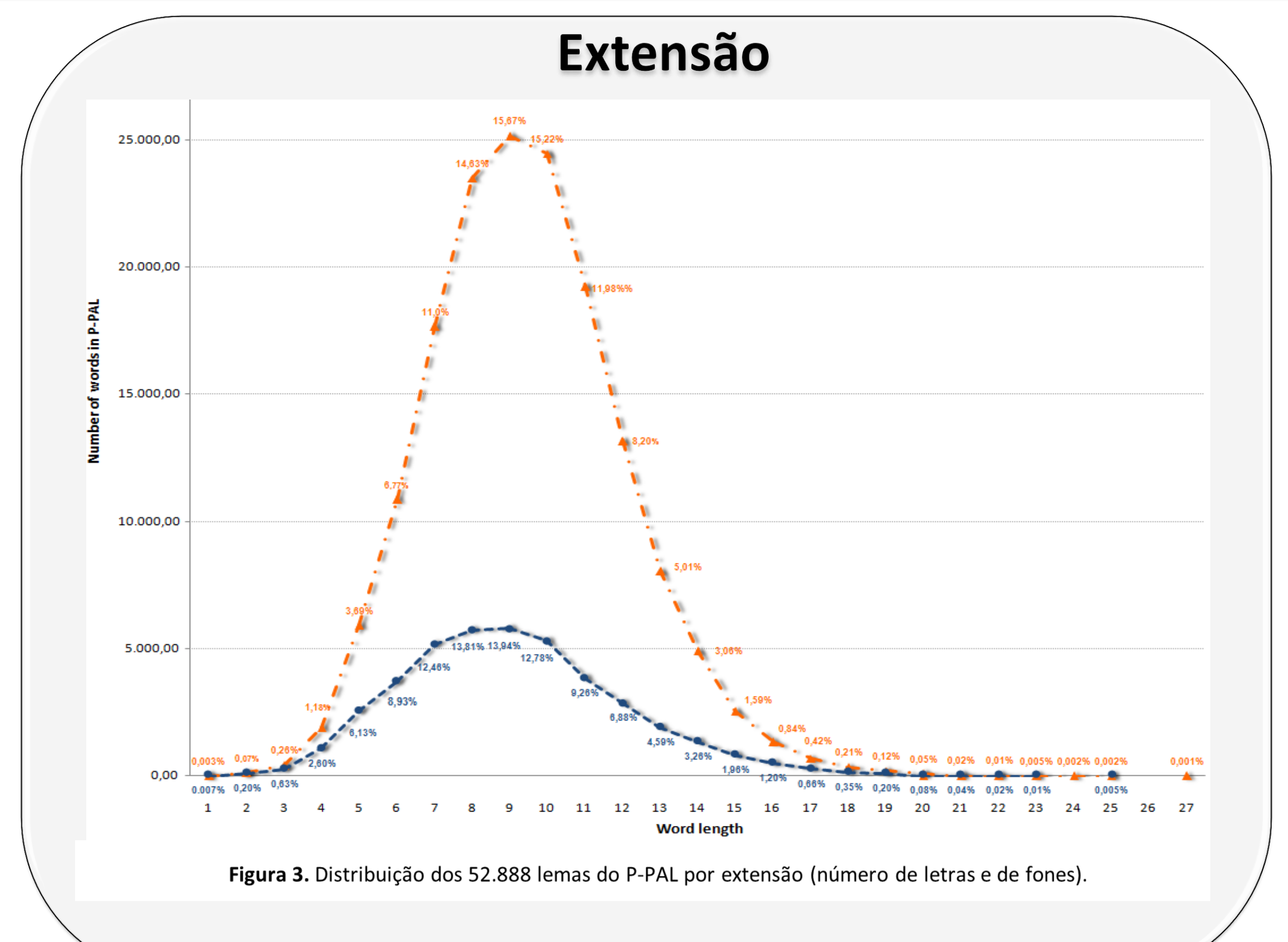
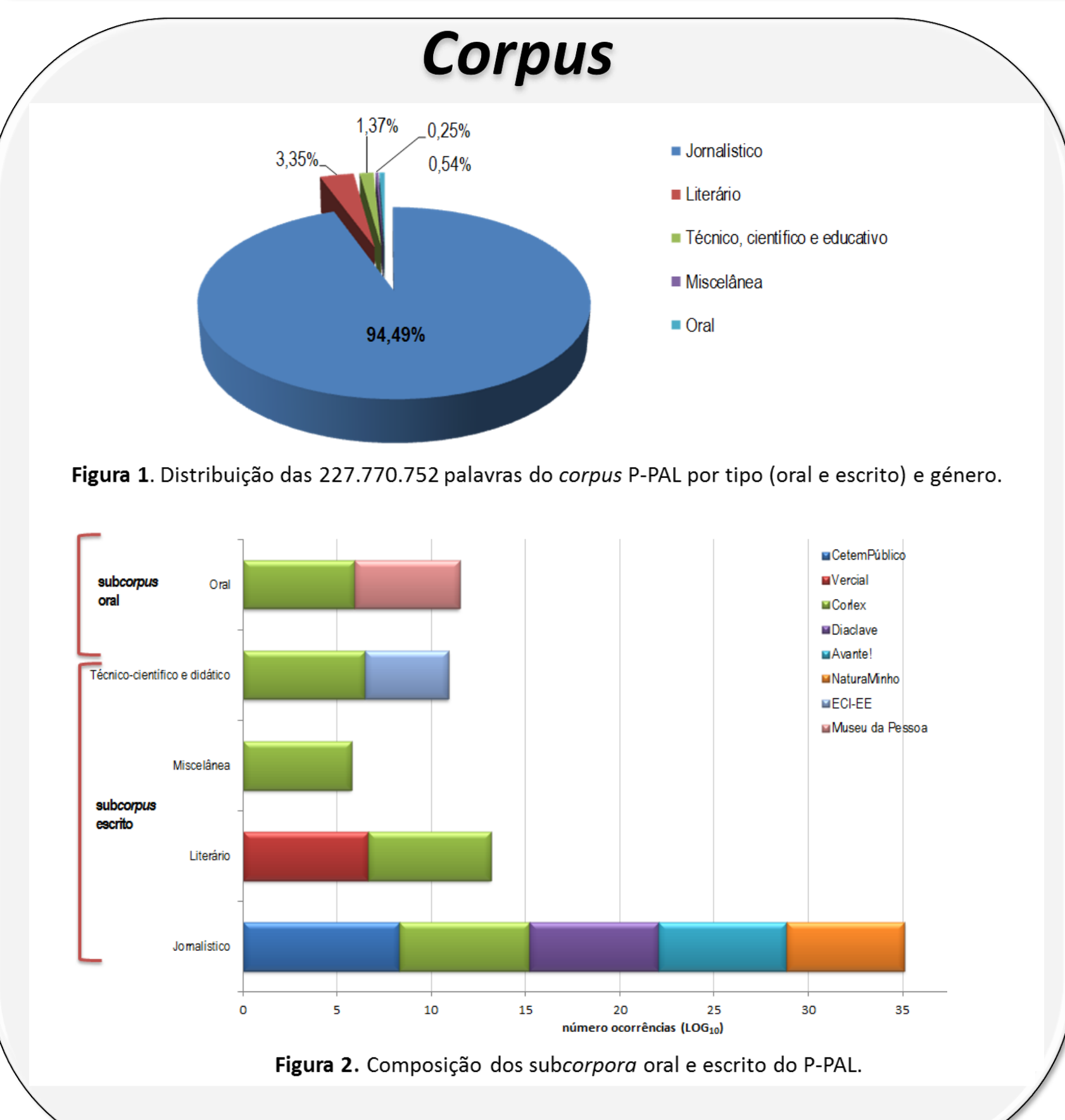
Lemas      ?

Formas      ?

**Lemas:** nesta pesquisa o utilizador analisa/obtém palavras que representam todas as formas flexionadas que uma palavra pode apresentar no *corpus* P-PAL (ex. lema "ir" representa as formas "vou", "vai", "indo", "fui" e "ido").

**Formas:** nesta pesquisa o utilizador analisa/obtém todas as potenciais ocorrências de uma palavra no *corpus* P-PAL (ex. "correr", "corre" e "correu" são formas do lema "correr" que podem ser obtidas neste tipo de pesquisa).

## Procura-PALavras: <http://p-pal.di.uminho.pt/>



### Fones e Bifones

Letters	Type frequency	Token frequency
A	32.804	359.768,11
B	5.353	32.574,83
C	17.112	147.254,22
D	12.587	183.401,81
E	25.869	509.508,92
F	4.805	33.884,89
G	6.039	43.528,67
H	3.330	23.606,45
I	24.389	197.173,62
J	1.060	10.492,29
K	130	274,52
L	14.277	96.368,29
M	12.054	168.904,10
N	18.200	153.943,86
O	26.748	350.639,13
P	8.637	108.547,98
Q	1.309	41.112,76
R	25.800	288.526,39
S	14.678	181.643,79
T	18.696	173.419,57
U	9.949	144.371,92
V	5.473	46.730,38
W	54	89,37
X	1.164	8.806,88
Y	71	140,35
Z	2.160	15.439,20

Procedessor	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
A	4	1.039	3.530	5.370	73	642	1.075	10	479	128	7	4.535	2.972	4.856	3.016	978	165	9.027	1.495	2.415	521	1.295	1	86	14	282	
B	1.206	3	30	28	882	5	3	0	1.064	39	0	141	19	9	771	6	1	859	180	28	404	12	1	0	14	1	
C	6.374	0	148	2	1.645	0	0	845	2.752	0	38	389	1	45	5.168	0	0	842	1	656	950	1	0	0	0	0	
D	2.841	1	1	1	4.657	0	3	0	2.071	30	1	3	34	1	4.318	0	5	384	6	0	437	23	2	0	0	6	
E	887	331	1.818	956	194	420	832	2	1.961	251	1	2.598	1.758	6.783	355	556	218	4.116	4.810	1.495	257	475	3	683	7	232	
F	876	0	0	0	816	19	0	0	1.272	0	1	250	0	0	753	0	0	497	1	18	356	0	0	0	0	1	
G	1.595	2	0	2	945	0	5	6	918	0	0	140	53	172	700	0	0	875	2	1	730	0	0	0	0	0	
H	1.510	1	0	0	549	0	0	0	371	0	0	0	0	0	763	0	0	1	0	9	145	0	0	0	0	0	
I	3.404	455	4.205	2.284	394	635	726	2	6	55	9	1.712	1.751	4.359	2.429	455	109	2.416	3.118	1.882	56	1.475	1	173	0	1.289	
J	405	0	0	0	165	0	0	0	15	0	0	0	0	0	213	0	0	0	0	0	266	0	0	0	0	0	
K	18	2	0	0	27	0	1	2	25	0	0	1	1	1	8	3	0	1	3	1	2	0	2	0	2	0	
L	2.911	39	190	145	1.589	100	151	1.019	3.218	5	5	28	439	9	1.610	76	17	102	514	590	181	0	0	0	0	3	3
M	2.534	605	0	0	4.059	0	0	0	1.600	0	2	2	23	2.210	1.029	0	0	0	459	0	0	0	0	0	0	0	
N	2.650	1	2.002	1.381	1.266	491	710	761	1.846	131	10	30	0	3	1.534	2	131	72	1.319	6.622	296	277	0	32	2	68	
O	3.68	574	978	458	192	344	923	4	311	71	12	2.051	1.360	3.309	75	600	96	4.730	1.750	871	371	390	17	92	6	66	
P	1.903	0	44	0	1.835	0	1	0	929	0	0	479	0	20	1.385	9	0	1.700	85	208	406	0	0	0	1	0	
Q	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1.308	0	0	0	0	0	
R	6.310	220	551	579	4.794	119	360	0	4.260	25	8	63	618	442	3.575	171	106	1.345	257	965	615	237	1	5	5	13	
S	2.064	103	1.313	24	1.744	210	108	32	1.774	5	13	84	1.025	68	1.776	854	139	25	1.229	3.346	965	76	3	0	1	0	
T	4.992	3	13	1	5.571	2	0	9	3.753	0	0	19	20	27	3.667	1	0	2.500	16	21	1.063	1	2	0	1	9	
U	756	476	394	413	334	101	197	1	923	37	0	1.374	540	717	109	349	22	1.499	680	804	0	99	1	50	0	0	
V	1.239	0	0	0	2.085	0	0	0	1.240	0	0	0	1	987	0	0	48	0	0	0	60	0	0	0	1	0	
W	16	2	1	0	5	0	0	1	7	0	0	1	0	3	6	0	0	1	2	1	0	0	0	0	0	0	
X	233	0	66	0	109	2	0	0	234	0	0	0	0	136	152	0	0	0	159	48	0	0	0	0	1	0	
Y	4	3	0	0	7	0	0	0	2	0	0	2	0	1	3	1	0	3	3	3	2	0	1	0	0	0	
Z	1.427	0	0	0	199	0	0	0	186	0	0	1	11	0	109	0	0	0	0	0	1	38	0	0	0	0	7