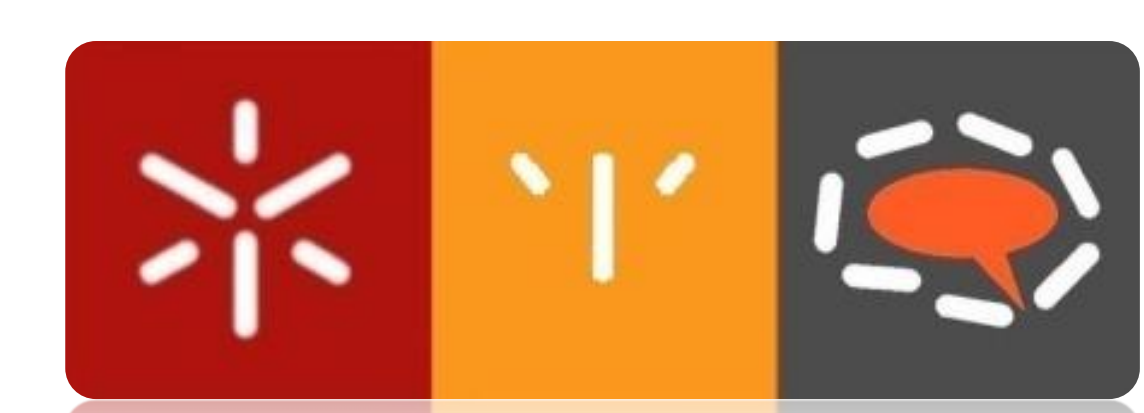


# Procura-PALavras (P-PAL): A web application for a new European Portuguese lexical database



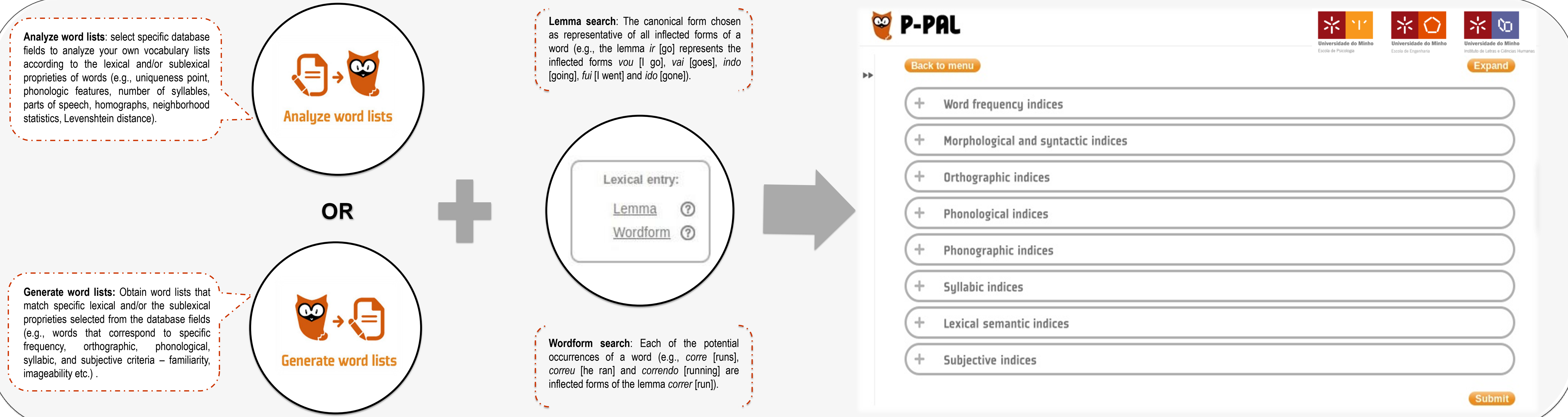
University of Minho, School of Psychology  
Research Group in Psycholinguistics



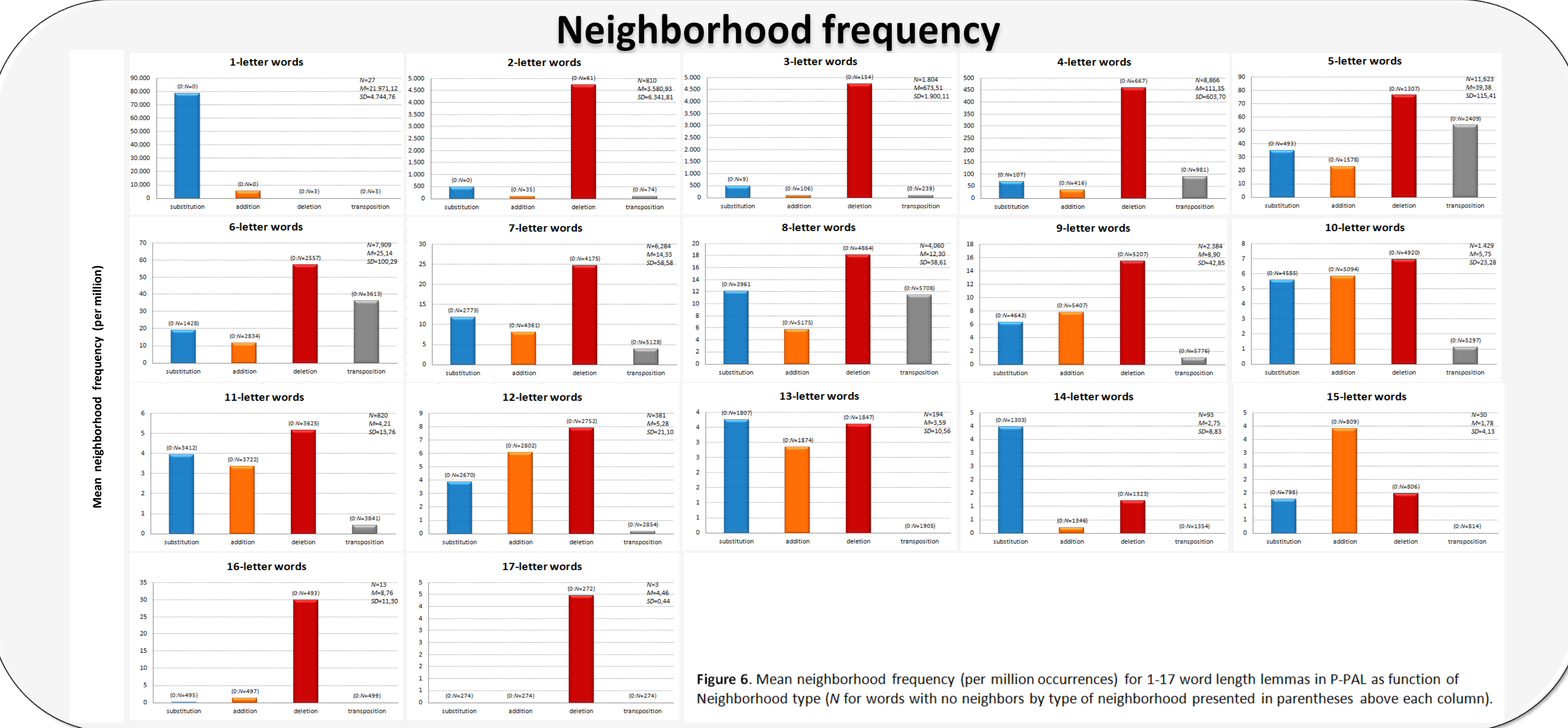
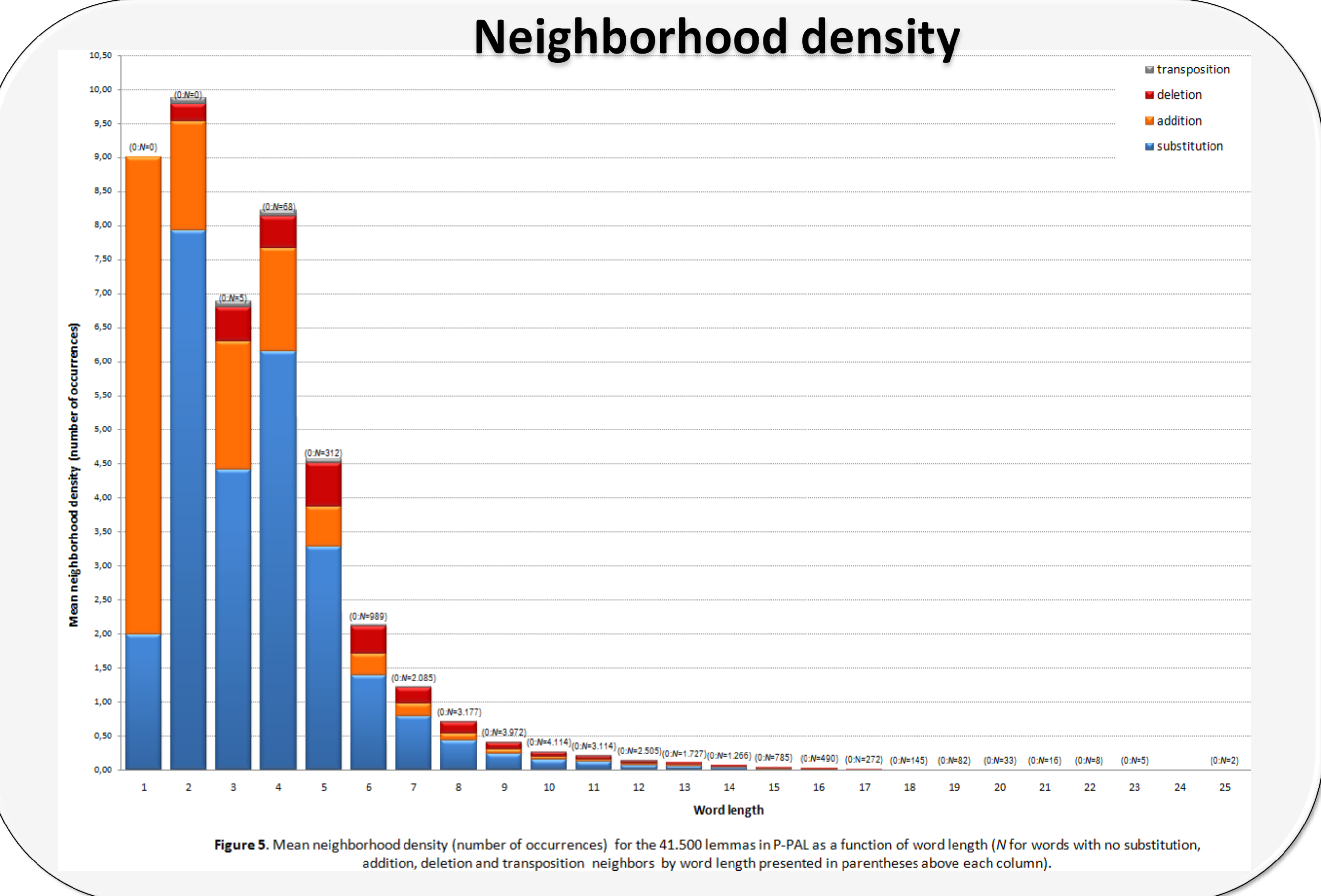
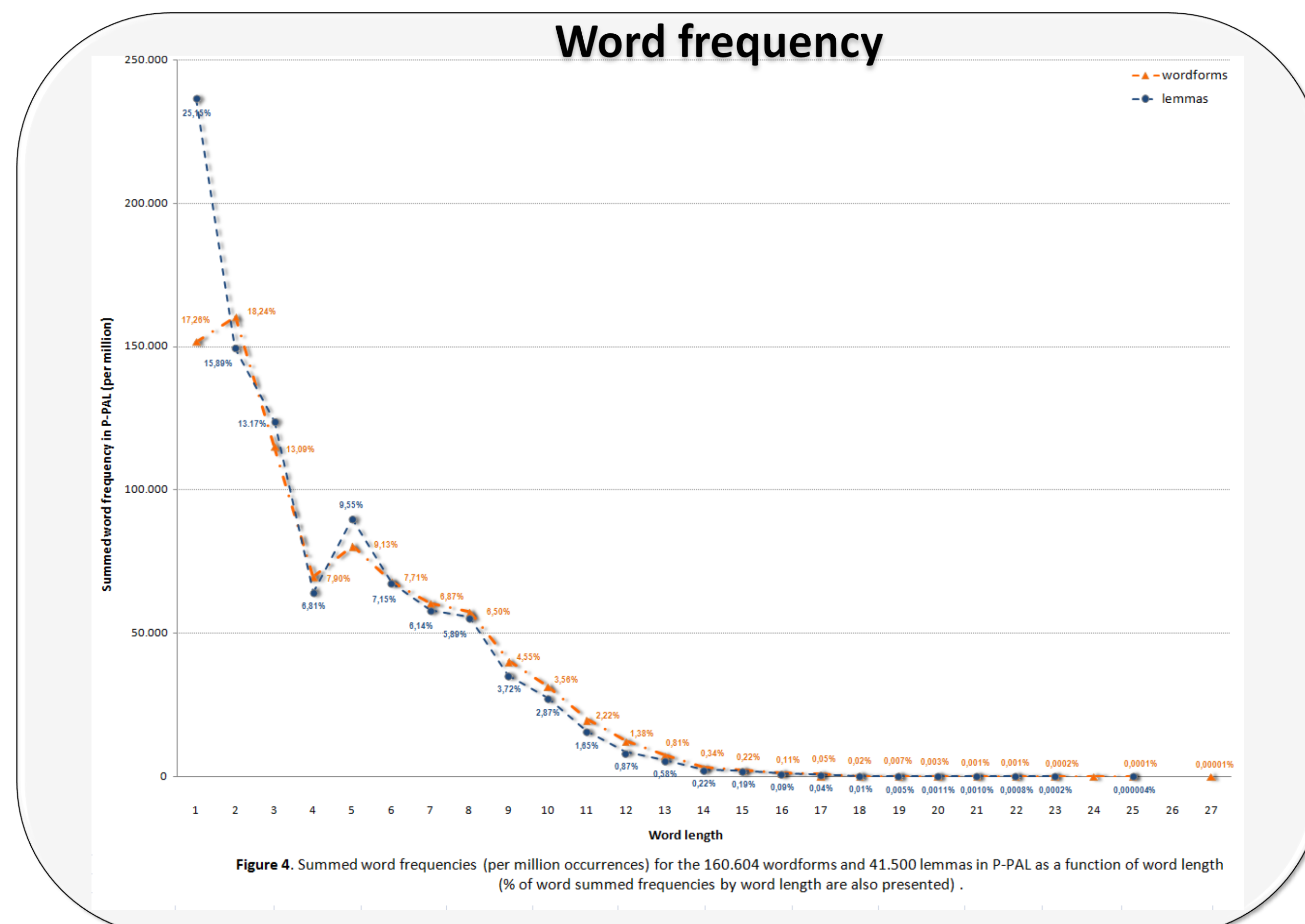
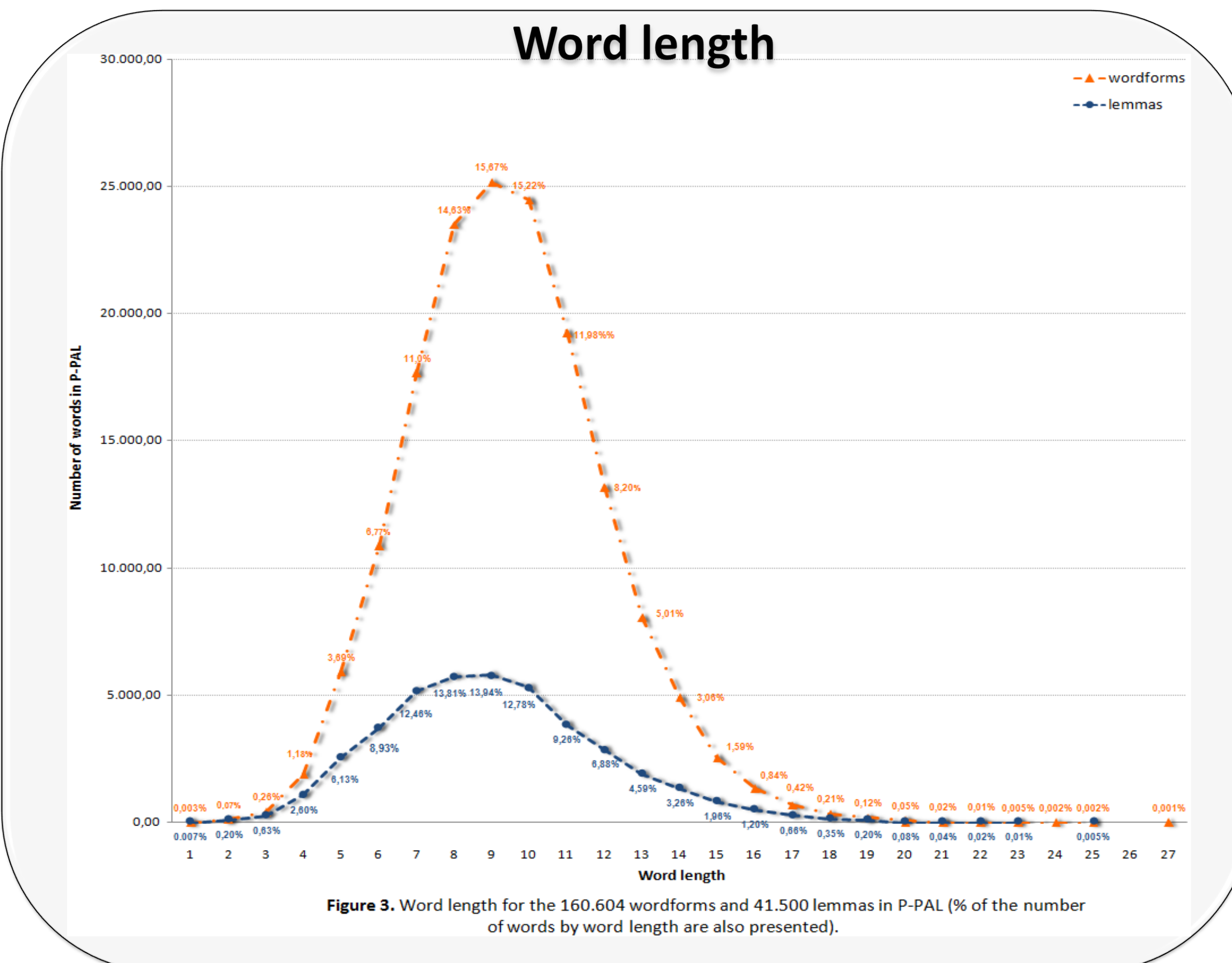
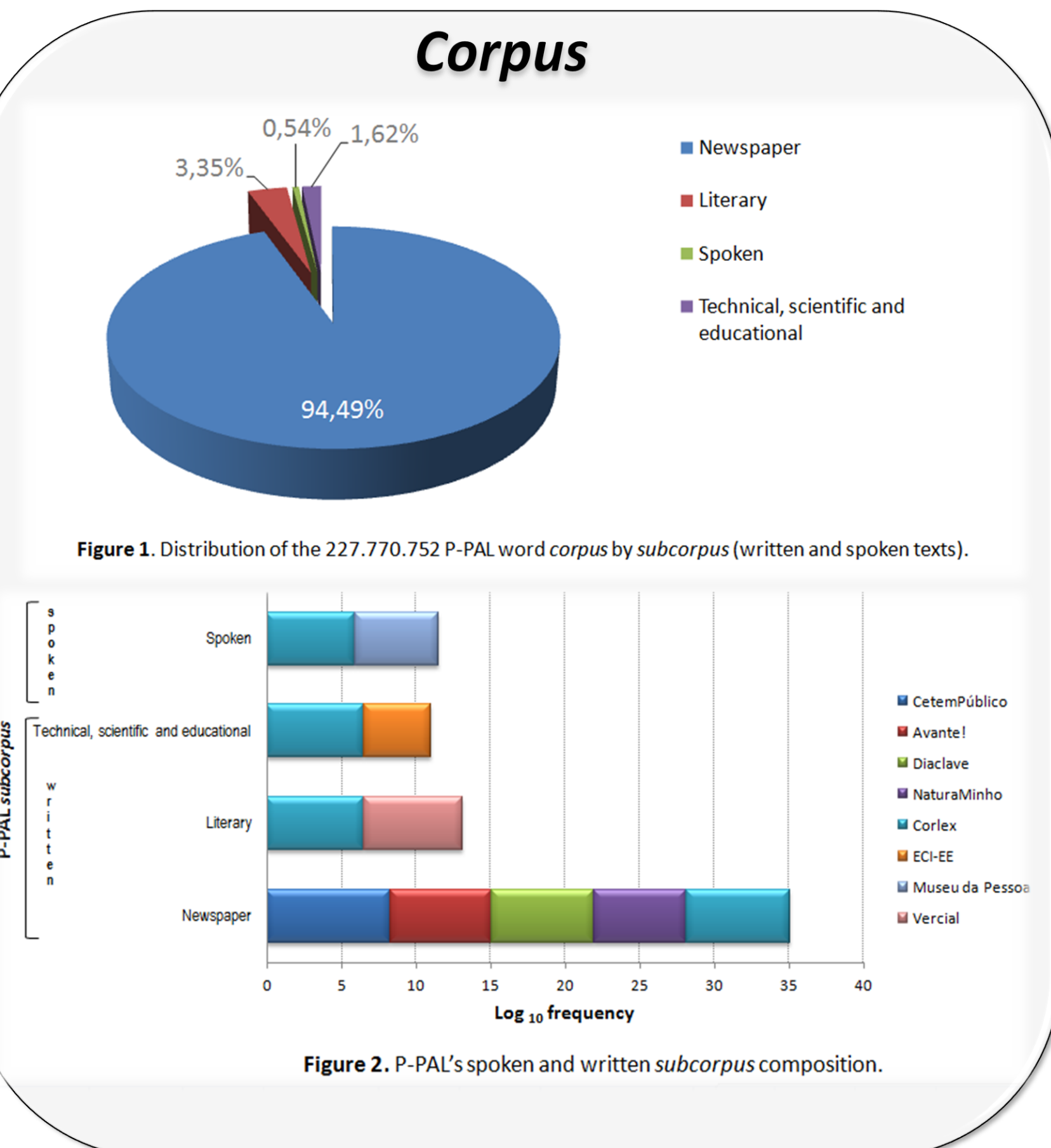
Ana Paula Soares<sup>1</sup>, Montserrat Comesaña<sup>1</sup>, Álvaro Iriarte<sup>2</sup>, José João de Almeida<sup>3</sup>, Alberto Simões<sup>3</sup>, Ana Costa<sup>1</sup>, Patrícia França<sup>1</sup> & João Machado<sup>1</sup>

{<sup>1</sup>Psycholinguistics Research Group, <sup>2</sup>Centre for Humanistic Studies, <sup>3</sup>Computer Science and Technology Center}, University of Minho

**Procura-PALavras (P-PAL)** is a web application for a new European Portuguese (EP) lexical database that provides a series of objective (lexical and sublexical) and subjective indices for  $\approx 160.000$  non-lemmatized and  $\approx 42.000$  lemmatized EP words. Based on a *corpus* of over 225 million EP words, the P-PAL enables users to obtain a broad range of statistics concerning the properties of word stimuli, including several measures of word frequency, syllable frequency, bigram and biphone frequency, orthographic and phonological structure, morphological and syntactic structure, orthographic and phonological similarity, lexical semantic indexes, concreteness, familiarity, imageability, valence, arousal, and dominance measures. In order to obtain these statistics the user should decide between a lemma or wordform search in the application and between two word-based queries: (i) analyze word lists in specific characteristics (objective and/or subjective); or (ii) generate lists of words with specific characteristics (objective and/or subjective). In this work we present the lemma and wordform frequency indices available, as well as some structure and similarity orthographic measures such as word length, neighborhood density and frequency. Letters and bigram type and token frequencies are also presented.



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



## Letters and Bigrams

Letters	Type frequency	Token frequency
A	32.804	399.768,11
B	5.353	32.574,83
C	17.112	147.254,22
D	12.567	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	288.525,39
S	14.678	181.643,79
T	18.698	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	14.439,20

Predecessor	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	4.056	3.016	978	185	9.027	1.495	2.415	521	1.295	1	88	14	282		
B	1.206	3	30	28	682	5	3	1.064	39	0	141	19	9	771	6	1	859	160	28	404	12	1	0	14	1	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	960	1	0	0	0	0	0	
D	2.841	1	1	4.857	0	3	0	2.071	30	1	3	34	1	4.318	0	5	384	6	0	437	23	2	0	0	0	0	0	
E	887	371	1.818	956	164	420	832	2	1.961	251	1	2.598	1.758	6.783	353	556	218	4.116	4.610	1.465	257	475	3	683	7	232		
F	876	0	0	0	816	19	0	0	1.272	0	1	250	0	0	763	0	0	497	1	18	356	0	0	0	0	0	0	
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	0	
H	1.510	1	0	0	549	0	0	0	371	0	0	8	2	763	0	0	0	1	0	9	145	0	0	0	0	0	0	
I	3.404	455	4.205	2.264	394	635	726	2	6	65	9	1.712	1.751	4.359	2.429	455	109	2.416	3.118	1.882	56	1.475	1	173	0	1.296		
J	405	0	0	0	165	0	0	0	15	0	0	0	0	213	0	0	0	0	0	266	0	0	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	0	
L	2.911	39	190	146	1.589	100	151	1.019	3.218	5	5	28	439	9	1.610	76	17	20	102	514	690	181	0	0	3	3	3	
M	2.534	605	0	0	4.059	0	0	0	1.600	0	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	6.222	296	277	0	32	2	68		
O	3.658	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	0	0	0	
Q	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
R	6.310	220	551	579	4.794	119	360	0	4.280	25	8	63	618	442	3.575	171	106	1.345	257	965	615	237	1	5	5	13		
S	2.054	103	1.313	24	1.744	210	108	32	1.774	5	13	84	1.025	68	1.176	854	139	25	1.229	3.346	965	76	3	0	1	0	0	
T	4.992	3	13	1	5.571	2	0	0	3	3.753	0	0	19	20	27	3.667	1	0	2.500	16	1	21	1.063	1	2	0	1	9
U	756	476	394	413	834	101	197	1	923	37	0	1.374	540	717	109	349	22	1.499	680	804	0	99	1	60	0	98		
V	1.239	0	0	0	2.085	0	0	0	1.240	0	0	0	0	1	987	0	0	48	0	80	0	0	0	0	0	0	0	
W	16	2	1	0	5	0	0	0	1	7	0	0	0	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	0	136	152	0	0	0	159	48	0	0	0	0	0	0	
Y	4	3	0	0	7	0	0	0	2	0	0	2	0	1	3	1	0	0	3	3	2	0	1	0	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	0	1	38	0	0	0	0	

For more information: [asoares@psi.uminho.pt](mailto:asoares@psi.uminho.pt)