



Procura-PALavras (P-PAL):

Uma aplicação *web* com índices
psicolinguísticos objectivos e subjectivos
para palavras do português europeu

Ana Paula **Soares**¹, Montserrat **Comesaña**¹, Álvaro **Iriarte**², José João de **Almeida**³,
Alberto **Simões**³, Ana **Costa**⁴, Patrícia **França**⁴ & João **Machado**⁴

¹Escola de Psicologia, ²Instituto de Letras e Ciências Humanas, ³Escola de Engenharia, ⁴Centro de Investigação em Psicologia
Universidade do Minho

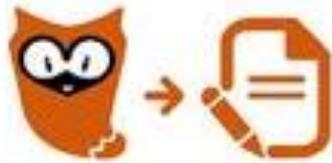
Motivações

- No PE as bases lexicais existentes são escassas e limitadas:
 - PORLEX (Gomes & Castro, 2003):
 - informação lexical de tipo ortográfico, fonológico, fonético, gramatical e de vizinhança para um total de 29.238 palavras;
 - informação de frequência para $\approx 5\%$ das entradas lexicais (importadas de um *corpus* oral de pequenas dimensões - Português Fundamental).
 - CORLEX (Bacelar-do-Nascimento *et al.*, 2000)
 - informação de frequência (absoluta) para 26.980 lemas e 140.976 formas [extraídas de um sub-*corpus* do *Corpus* de Referência do Português Contemporâneo (CRPC) do Centro de Linguística da Universidade de Lisboa];
 - informação de tipo morfo-sintáctico para essas palavras.

Características

- Aplicação *web* que disponibilizará informação objectiva e subjectiva sobre palavras do PE:
 - **frequência lexical**: ocorrência das palavras por milhão de palavras, logarítmica (base 10) e diversidade contextual;
 - **informação estrutural**: informação de tipo linguístico determinada pela análise da própria palavra (ex. extensão da palavra em letras e fonemas, classe gramatical);
 - **informação derivada**: informação que resulta da análise da relação da palavra com as restantes do léxico tanto a nível lexical (ex. vizinhança), como sublexical (ex. bigramas, bifones, sílabas, morfemas);
 - **informação subjectiva**: informação que reflecte as experiências dos indivíduos com o uso da própria língua (ex. familiaridade, imaginabilidade, concreteness, valência).

Funcionalidades



Obter palavras

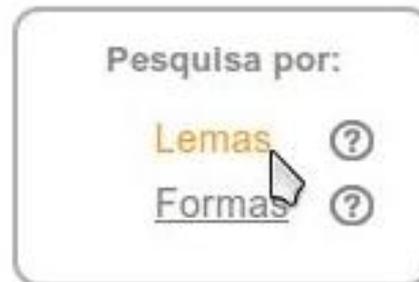


Analisar palavras

obter palavras que obedçam aos requisitos seleccionados do menu de análises (ex. palavras que possuam uma determinada categoria morfo-sintáctica ou um determinado valor de frequência).

analisar palavras pré-definidas pelo utilizador nos requisitos seleccionados do menu de análises (ex. índices de frequência, ponto de unicidade ortográfico, homógrafos ou vizinhos por transposição).

Tipos de pesquisa



Lema: forma escolhida convencionalmente para representar todas as flexões de uma palavra (ex. “*ser*” “*menino*”, “*bonito*”, “*homem*”, “*mulher*” “*animal*” ...).

Forma: cada uma das potenciais ocorrências de uma mesma palavra. Ex. “*cantar*”, “*canto*”, “*cantava*”, “*cantei*”, “*cantamos*” [...], são formas da mesma palavra, convencionalmente representadas pelo lema “*cantar*”.



Análises



estor

- exactamente
- no início
- contém
- no fim



[Voltar à aplicação](#)

[Expandir opções](#)

- + Índices de frequência lexical
- + Índices morfológicos e morfo-sintácticos
- + Índices ortográficos
- + Índices fonológicos
- + Índices fonográficos
- + Índices silábicos
- + Índices semântico-lexicais
- + Índices subjectivos

Feedback

[Submeter](#)

Análises

[Voltar à aplicação](#)

estor

- exactamente
- no início
- contém
- no fim

admoestar	cestaria	emprestar	infestar	testar			
admoestar	contestar	emprestar	mal-estar	testar			
adoestar	contestar	entestar	mal-estar				
adoestar	crestar	entestar	manifestar				
aprestar	crestar	estar	manifestar				
aprestar	desaprestar	estar	molestar				
arrestar	desaprestar	estardalhaço	molestar				
arrestar	desempestar	estardalhaço	prestar				
assestar	desempestar	estarola	prestar				
assestar	desinfestar	estarola	protestar				
atestar	desinfestar	estarrecer	protestar				
atestar	detestar	estarrecer	requestar				
bem-estar	detestar	florestar	requestar				
bem-estar	empestar	florestar	restar				
cestaria	empestar	infestar	restar				

Feedback



Análises



[Voltar à aplicação](#)

[Expandir opções](#)

- + Índices de frequência lexical
- + Índices morfológicos e morfo-sintácticos
- + Índices ortográficos
- + Índices fonológicos
- + Índices fonográficos
- + Índices silábicos
- + Índices semântico-lexicais
- + Índices subjectivos

Feedback

[Submeter](#)

Voltar à aplicação

Colapsar opções

Índices de frequência lexical

Corpora

Frequência por milhão [freq_corp_mil] ⓘ
(mínimo actual: 000; máximo actual: 000; média actual: 000)

Mínimo: 2000 Máximo: 6000

Frequência logarítmica [freq_corp_log] ⓘ
(mínimo actual: 000; máximo actual: 000; média actual: 000)

Mínimo: Máximo:

Diversidade contextual [freq_corp_div] ⓘ
(mínimo actual: 000; máximo actual: 000; média actual: 000)

Mínimo: Máximo:

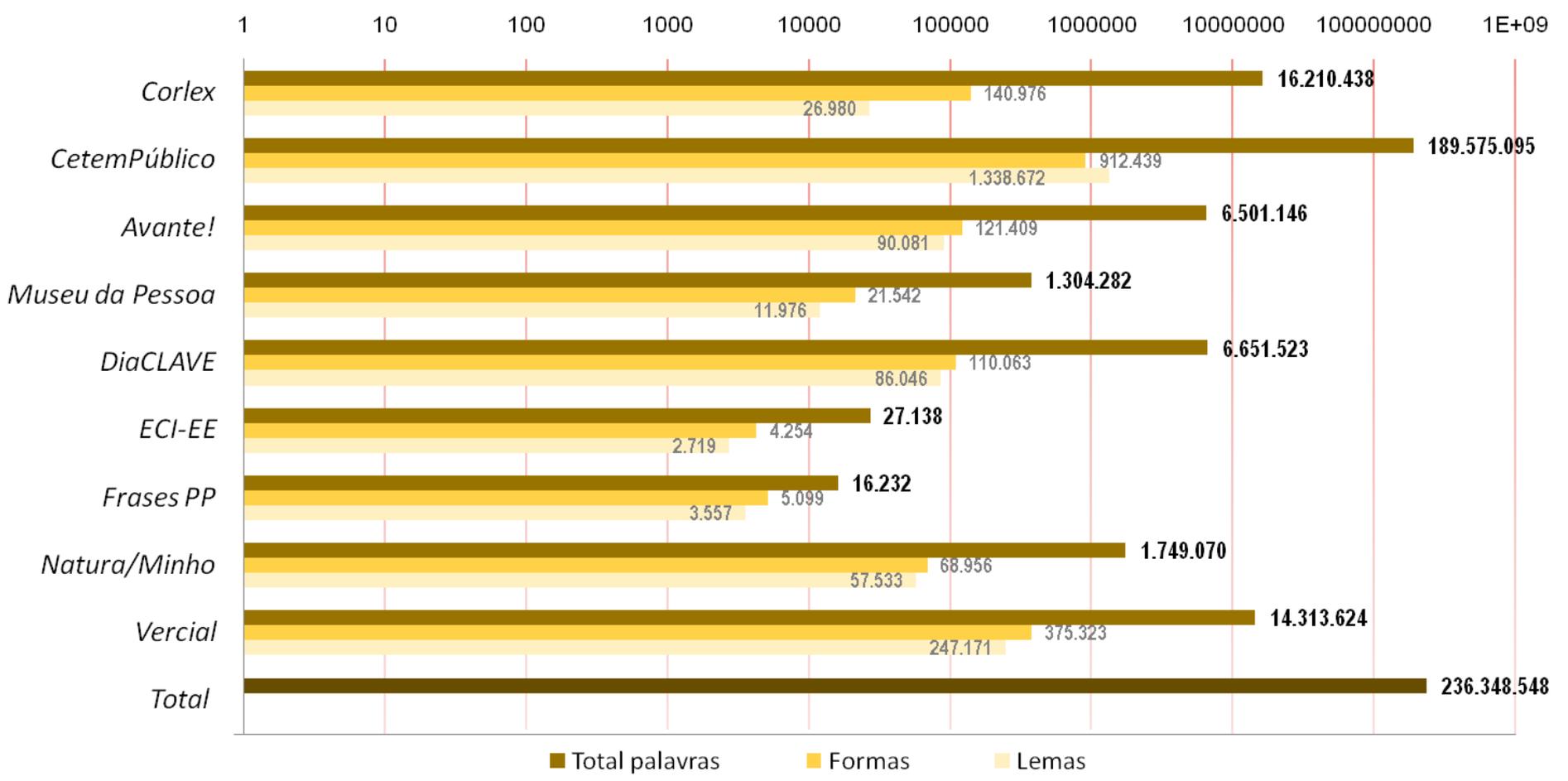
Legendas

Manuais Escolares

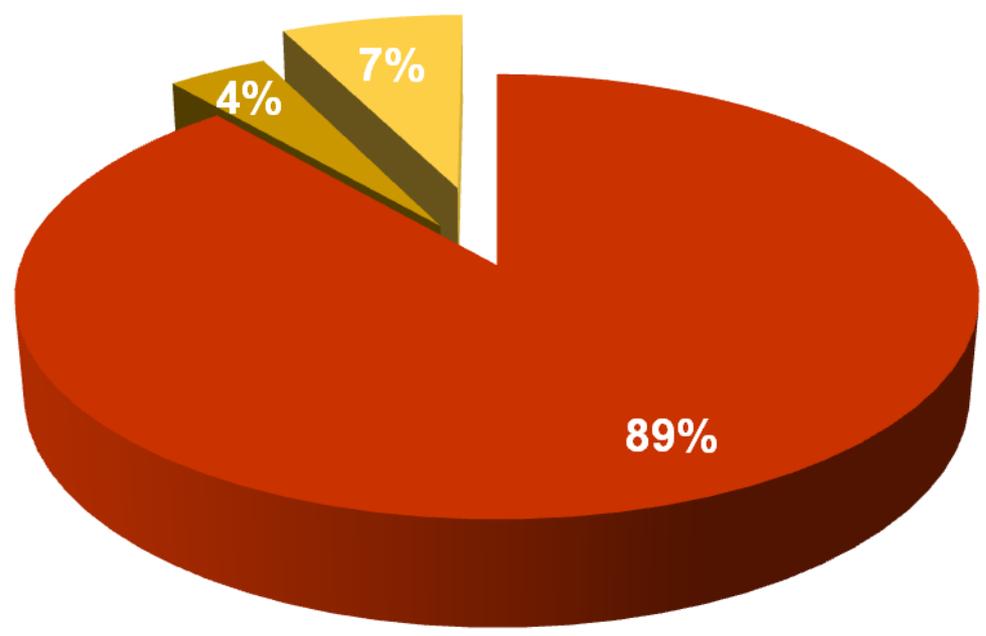
Feedback

Índices morfológicos e morfo-sintácticos

Frequência lexical

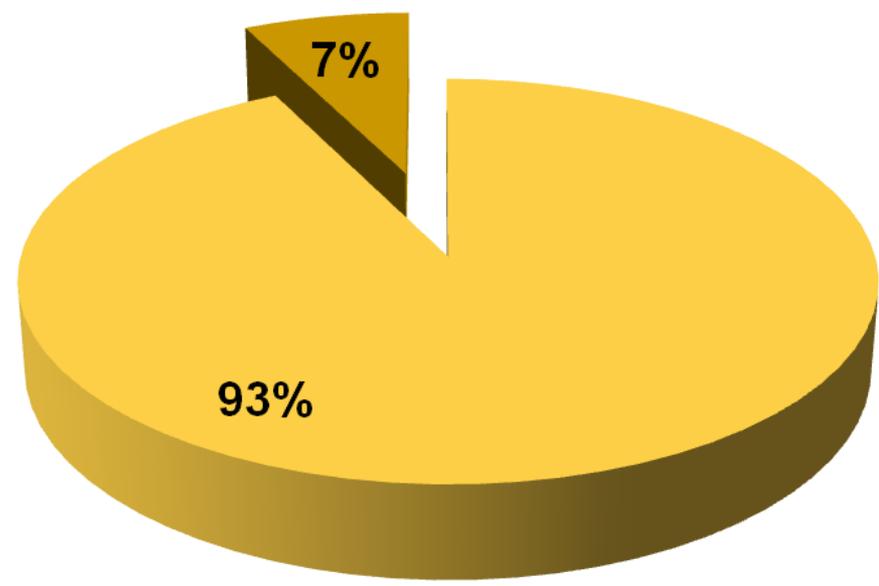


Frequência lexical



■ Jornalístico ■ Literário ■ Vários géneros (Técnico-científico e outros)

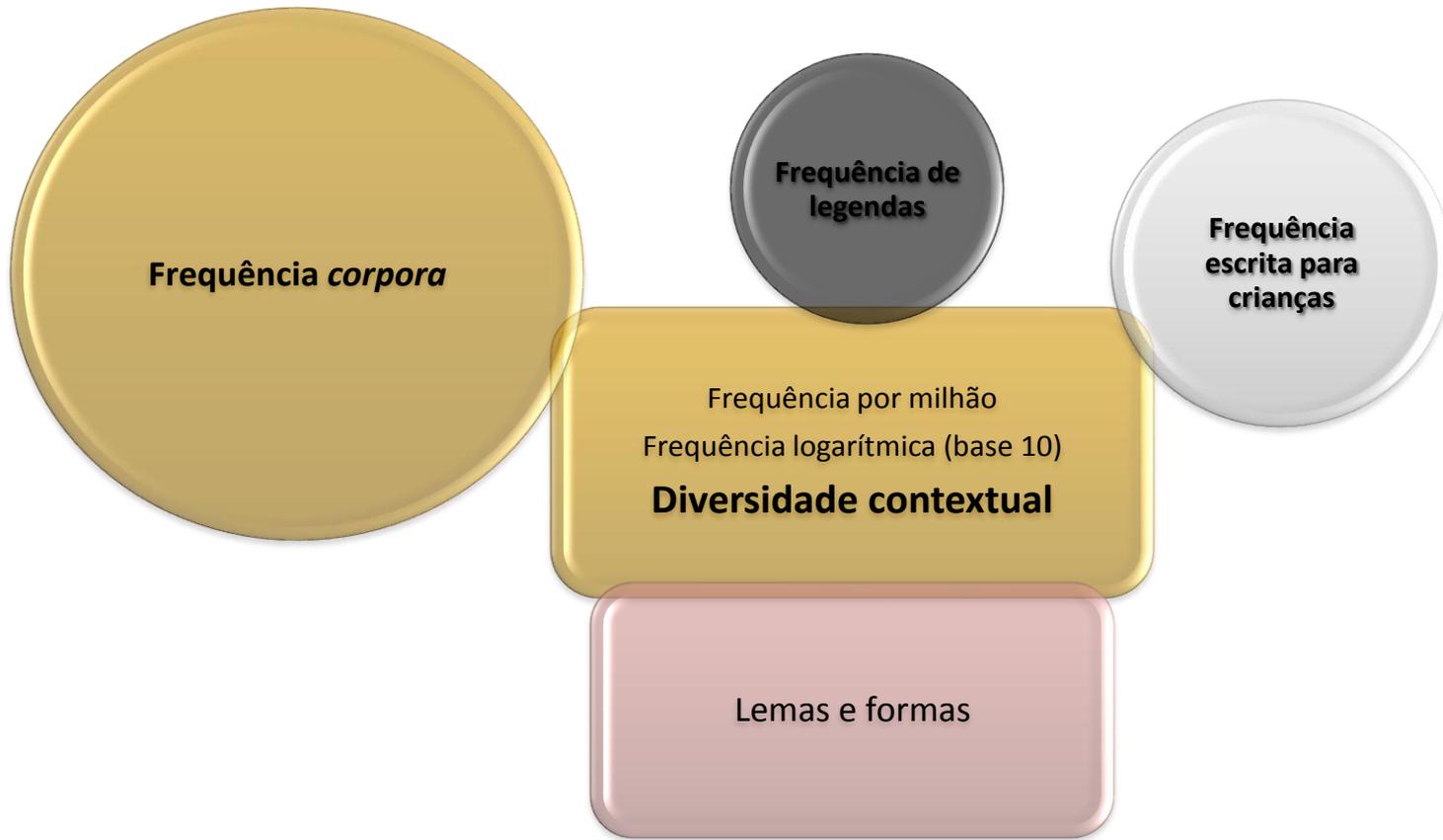
Frequência lexical



■ Escrito

■ Oral

Frequência lexical



Análises

Voltar à aplicação

Colapsar opções

Índices de frequência lexical

Corpora

- Frequência por milhão [freq_corp_mil] ⓘ Minimo: 2000 Maximo: 6000
(mínimo actual: 000; máximo actual: 000; média actual: 000)
- Frequência logarítmica [freq_corp_log] ⓘ Minimo: Maximo:
(mínimo actual: 000; máximo actual: 000; média actual: 000)
- Diversidade contextual [freq_corp_div] ⓘ Minimo: Maximo:
(mínimo actual: 000; máximo actual: 000; média actual: 000)

Legendas

Manuais Escolares

+ Índices morfológicos e morfo-sintácticos

Pesquisar

- exactamente
- no início
- contém
- no fim

Ordenação dos resultados ⓘ

- freq_tot_mil ✕
- freq_tot_log ✕



Feedback

Resultados

Frequência logarítmica

Lema	freq_corp_mil	freq_corp_log
ano	2174.0419	3.3400
como	3694.4046	3.5700
dizer	2363.5698	3.3700
esta	2366.1126	3.3700
estar	3734.7347	3.5700
este	2387.0464	3.3800
fazer	2793.6000	3.4500
haver	2516.5522	3.4000
ir	2689.7588	3.4300
já	2101.7788	3.3200
mais	4046.9679	3.6100
mas	3287.2005	3.5200
o	5574.9591	3.7500
ou	3017.8994	3.4800
poder	2491.8337	3.4000

Resultados

Lema	freq_corp_mil	freq_corp_log
ano	2174.0419	3.3400
como	3694.4046	3.5700
dizer	2363.5698	3.3700
esta	2366.1126	3.3700
estar	3734.7347	3.5700
este	2387.0464	3.3800
fazer	2793.6000	3.4500
haver	2516.5522	3.4000
ir	2689.7588	3.4300
já	2101.7788	3.3200
mais	4046.9679	3.6100
mas	3287.2005	3.5200
o	5574.9591	3.7500
ou	3017.8994	3.4800
poder	2491.8337	3.4000

Feedback

Dê-nos a sua opinião

o meu e-mail:

mensagem:

Cancelar

Enviar

Obrigada!



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