

1 Comparação de corpos

Descreva e compare os corpos DisPR e Museu da Pessoa

DisPR é um corpo de discursos presidenciais, ditos por presidentes portugueses e brasileiros, por ocasião da sua tomada de posse, ou dos festejos do 25 de abril, em Portugal, correspondendo ao período de 1889 a 2024.

O corpo Museu da Pessoa contém entrevistas a pessoas do povo, em Portugal e no Brasil, feitas na década de 90 do século XX.

São ambos corpos de linguagem oral, mas, enquanto o DisPR se refere a linguagem formal, o Museu da Pessoa tem uma linguagem informal.

Tamanhos: 275 mil palavras e 1,4 milhões de palavras respetivamente

2 Expressões regulares

Escreva uma expressão regular que aceite datas no formato 1/3/62 ou 01/03/62 para indicar 1 de março de 1962. (Mais exemplos: 28/02/46, ou 31/1/65, ou 03/12/74.)

`(0[1-9])|([1-9])|([1-2][0-9])|(3[0-1])\V(0[1-9])|([1-9])|(1[0-2])\V[0-9][0-9]`

3 Expressões regulares 2

Escreva uma expressão regular no AC/DC que obtenha lemas com exatamente dois f.

`[lema="^[^f]*f[^f]*f[^f]*"]`

4 Expressões regulares no AC/DC

Escreva uma expressão regular no AC/DC que obtenha conjunções de adjetivos, como "simpático, bonito e elegante", ou "verde e azul" ou "cansado, deprimido, pessimista e inseguro".

`([pos="ADJ.*"] ",")* [pos="ADJ.*"] [word="e"] [pos="ADJ.*"]`

5 Expressões regulares no AC/DC 2

Escreva uma expressão de pesquisa no AC/DC que encontre um substantivo no plural acabado em *a* modificado por um adjetivo acabado em *vel*. Exemplos: *problemas muito consideráveis*, *pessoas amáveis*, *caixas de papel e desdobráveis*.

`[pos="N.*" & lema=".*a" & pessnum="P"] [pos!="N.*"]* [pos="ADJ.*" & lema=".*vel" & func="N<"] within s`

6 Expressões regulares no AC/DC 3

Procure nomes próprios que incluam as palavras *Brasil* ou *Portugal*, como em *Biblioteca Nacional de Portugal*, ou *Banco do Brasil*, ou *"Comitê Brasil - Noruega"* ou *"Portugal no coração"*.

`[lema="(.*)*(Brasil|Portugal)(=.*)*" & lema=".*=.*)" ]`

7 Procuras no AC/DC

Quais são as três coisas mais azuis no corpo CHAVE? Considere a palavra azul, e não membros do grupo semântico da cor do grupo Azul. (2)

[pos="N.*"] [pos!="N.*"]* [pos="ADJ.*" & func="N<" & lema="azul"] within s
Distribuição dos lemas

dá capacete, olho e bandeira

Se só fizerem

[pos="N.*"] [pos="ADJ.*" & func="N<" & lema="azul"]
têm 1.5

Se só fizerem

[pos="N.*"] [pos="ADJ.*" & lema="azul"]
têm 1.0

Existe diferença entre a variante portuguesa e a brasileira? (1)

Por variante

[pos="N.*" & variante="BR"] [pos="ADJ.*" & func="N<" & lema="azul"]
dá olho, capacete e céu

[pos="N.*" & variante="PT"] [pos="ADJ.*" & func="N<" & lema="azul"]
dá capacete, bandeira e olho

8 Diferença entre lema e forma

Explique a diferença entre as duas procuras [lema="rio"] e [word="rio"] indicando casos que apareceriam numa resposta e não noutra.

No primeiro caso encontraria a palavra/forma *rios*, que não era encontrada no segundo; e no segundo caso encontraria a forma do verbo *rir*, que não seria encontrada no primeiro.

9 Emoção

Qual a palavra (lema) de emoção mais usada no corpo DisPR? E no Diaspora-TL?

[sema="*.emo:.*"]

Distribuição por lema

querer e amigo

10 Corpo humano

Usa-se o lema *olho* mais frequentemente no plural ou no singular (1)? Use o corpo Literateca.

[lema="olho"]

Distribuição de pessoa e/ou número

P 40731

S 2065

No plural

E o substantivo *olhar*? (1)

[lema="olhar" & pos="N.*"]

S 8931

P 1475

No singular.

11 Família

Quem menciona mais a família no corpo Museu da Pessoa, são homens ou mulheres?

[sema=".*familia.*"]

Distribuição de sexo

Fazendo a distribuição relativa, são as mulheres

M: 0,00579

F: 0,01187

12 Medo

No corpo NOBRE, quais os três lemas mais usados sobre a emoção medo? (1)

[sema=".*emo:medo.*"]

Distribuição dos lemas

medo

terrível

receio

Estude os casos dos verbos referentes a medo. Em que percentagem se encontram negados?
(2)

[sema=".*emo:medo.*" & pos="V.*"] : 4179

[lema="ninguém|nada|nenhum|não|nunca|jamais"] [pos="(V-AUX|ADV).*"]*

[sema=".*emo:medo.*" & pos="V.*"] : 275

[lema="nenhum"] [pos="N.*"] [pos="(V-AUX|ADV).*"]* [sema=".*emo:medo.*" & pos="V.*"] : 3

6,65%

13 Classificação

Através de uma amostra aleatória de 15 casos no corpo OBRAS, verifique que sentidos consegue obter da palavra língua. (Diga as traduções para norueguês dos diferentes sentidos.)

língua

Concordância

Amostra aleatória de 15 linhas.

O aluno deve copiar os 15 casos obtidos e dar uma tradução/classificação a cada...

E eventualmente depois dizer qual a mais frequente e que outros sentidos obteve (fazer um resumo).

14 Passiva

Dos cinco autores de Prosa com mais material na Literateca, qual o que usa mais passiva?

Primeiro é preciso identificar os cinco autores de prosa om mais texto na Literateca:

Qual a sua função sintática mais frequente, para cada um deles?

[classe="Prosa.*"]

Distribuição por autor

CamCBra	3101754	3490533
EcaQue	2784560	2784520
MacAss	2175466	2198452
TeoBra	1640558	1730103
AleHer	1469019	1490939

Depois é preciso estudar a passiva nestes autores

[classe="Prosa.*" & autor="CamCBra|EcaQue|MacAss|TeoBra|AleHer" & temcagr=".*PASSIVA.*"]

CamCBra	8468	3490533
EcaQue	7175	2784520
TeoBra	6760	1730103
MacAss	6324	2198452
AleHer	5401	1490939

E depois usar os valores relativos na prosa: ou seja

CamCBra 8468 em 3101754 = $2,73 \times 10^{-3}$

EcaQue 7175 em 2784560 = $2,57 \times 10^{-3}$

TeoBra 6760 em 1640558 = $4,12 \times 10^{-3}$

MacAss 6324 em 2175466 = $2,90 \times 10^{-3}$

AleHer 5401 em 1469019 = $3,67 \times 10^{-3}$

É Teófilo Braga.

A bitola usada para atribuir as notas é:

Reprova abaixo de 14, de 14 a 17 tem E, de 17 a 20 tem D, de 20 a 23 tem C, de 23 a 27 tem B, e tem A se tiver mais de 27 pontos.