

(Først på portugisisk, så på norsk)

1 Descrição de corpos

Compare os corpos LeMe e CiênciaViva

Ambos se referem a português de Portugal, mas o primeiro é um corpo de bulas (informação sobre medicamentos) e o outro contém textos de divulgação científica escritos por portugueses para portugueses, publicados na imprensa local.

Tamanhos: 2,6 milhões de palavras e 656 mil palavras

2 Expressões regulares

Crie uma expressão regular para obter datas do século dezanove.

18[0-9][0-9]

3 Crie uma expressão no AC/DC que obtenha palavras (formas) que contenham pelo menos três r.

`. * r . * . r . * r . *`

ou

`[word=" . * r . * . r . * r . * "]`

4 Crie uma expressão regular no AC/DC que obtenha lemas iniciados por "im", com mais de quatro letras.

`[lema="im... *"]`

ou

`[lema="im... +"]`

5 Escreva uma expressão de pesquisa no AC/DC que encontre um substantivo acabado em "o" seguido por um adjetivo acabado em "vel".

`[pos="N. *" & word=" . * o "][pos="ADJ. *" & word=" . * vel "]`

6 Procure no AC/DC frases que contenham um sintagma nominal que tenha algo (uma ou mais palavras) entre o substantivo e o adjetivo. (Por exemplo, *sopa muito boa*, *caso extremamente complicado*, *resposta nem sempre fácil*, etc.)

`[pos="N. *" ]+ [pos="ADJ. *" & func="N <"] within s`

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

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

8 O que é mais amarelo no corpo ConDiv?

Esta pergunta pode ser interpretada de várias maneiras. Admitindo que é a cor amarelo como cor e não como indicação de outra coisa,

[pos="N.*"] [pos="ADJ.*" & func="N<" & grupo="Amarelo"]
Distribuição de lema

dá cartão, lã e cor

Se for a palavra *amarelo*

[pos="N.*"] [pos="ADJ.*" & func="N<" & lema="amarelo"]
Distribuição de lema

dá cartão, lã e camisa

5 Qual o adjetivo (ou particípio passado) mais usado para adjetivar pernas, no corpo ConDiv?

[lema="perna"] @[pos="ADJ.*"]

Distribuição dos lemas
dá *direito*

Se juntar

[lema="perna"] [pos!="N.*"]* @[pos="ADJ.*" & func="N<"] within s

tb dá *direito*

ou

[lema="perna"] (@[pos="ADJ.*"]|@[pos="V.*" & temcagr="PCP"])

tem os 2 pontos, senão tem só 1.

6 Em que jornal do corpo Moçambula se fala mais do campo da saúde?

[sema="saude.*"]
Distribuição por fonte

É preciso fazer a contagem relativa.

74/24263=0,0030

28/15577=0,0018

24/7218=0,0033

7/2711=0,0025

4/11204=0,00035

É no jornal O País que há mais discussão sobre o campo da saúde.

7 Pesquisa sobre género

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

[sema="*.emo:saudade.*"]
Distribuição de sexo

Fazendo a distribuição relativa, são as mulheres

M: 0,000116405

F: 0,000211272

Se puser só a palavra ou o lema saudade, e o resto do processo estiver bem, damos 2 ou 2,5, respectivamente.

8 No corpo CHAVE, quais as três emoções mais vezes consideradas fortes?

(@[sema=".*emo.*" "forte"])(("forte" @[sema=".*emo.*"])

Distribuição dos lemas

sofrer

sentimento

emoção

Se repararem que dá um verbo, e refinarem para

(@[sema=".*emo.*" & pos="N" "forte"])(("forte" @[sema=".*emo.*" & pos="N"])

têm 1,5, senão têm 1.

sentimento

emoção

desejo

Também podem ter 1,5 se escolherem lema em vez de apenas a palavra forte:

(@[sema=".*emo.*" [lema="forte"]])([lema="forte" @[sema=".*emo.*"])

Existe diferença entre as duas variantes?

(@[sema=".*emo.*" & variante="BR" "forte"])(("forte" @[sema=".*emo.*" & variante="BR"])

sentimento

emoção

expectativa

(@[sema=".*emo.*" & variante="PT" "forte"])(("forte" @[sema=".*emo.*" & variante="PT"])

sentimento

desejo

emoção

(Aqui, desde que os alunos tenham feito as duas procuras, e tenham posto a variante nos dois lados do |, não descontamos qual a forma que usaram.)

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

[lema="hábito" & pos="N.*"]

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).

10 Quais os (dois) adjetivos mais comuns como predicativos do sujeito, no corpo Literateca?

```
[pos="ADJ.*" & func=".*SC.*"]
```

Distribuição dos lemas

bom	7982
-----	------

preciso	6626
---------	------

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

```
[pos="ADJ.*" & lema="bom"]
```

>N

```
[pos="ADJ.*" & lema="preciso"]
```

<SC

Comentário: o adjetivo *bom* é um dos mais frequentes no corpo (e na língua toda), por isso não admira que seja frequente como predicativo. Pelo contrário, o adjetivo *preciso* é muito mais comum como predicativo do sujeito do que noutras configurações sintáticas.

NORSK:

1 Beskrivelse av korpus

Sammenlikn korpusene LeMe og CienciaViva.

Begge gjelder portugisisk fra Portugal, men det første omfatter pakkevedlegg (opplysninger om legemidler), og det andre inneholder tekster med vitenskapelig popularisering som er skrevet av portugisere for portugisere, og har blitt publisert i portugisiske aviser og blad. Omfang: 2,6 millioner ord og 656 tusen ord.

2 Regulære uttrykk

Skriv et regulært uttrykk for å finne datoer i det nittende århundre.

18[0-9][0-9]

3 Skriv et uttrykk i AC/DC for å finne ord(former) som inneholder minst tre r'er.

```
.*r.*r.*r.*
```

eller

```
[word=".*r.*r.*r.*"]
```

4 Skriv et regulært uttrykk i AC/DC for å finne lemma som begynner på "im", og som er på fire bokstaver.

```
[lema="im....*"]
```

eller

```
[lema="im...+"]
```

5 Skriv et søkeuttrykk i AC/DC for å finne et substantiv som ender på "o" fulgt av et adjektiv som ender på "vel".

[pos="N.*" & word=".*o"] [pos="ADJ.*" & word=".*vel"]

6 Søk i AC/DC etter setninger som inneholder et nominalsyntaxme som har ett eller flere ord mellom substantivet og adjektivet. (For eksempel *sopa muito boa, caso extremamente complicado, resposta nem sempre fácil.*)

[pos="N.*"] []+ [pos="ADJ.*" & func="N<"] within s

7 Forklar forskjellen mellom de to søkene [lema="caso"] og [word="caso"], der du viser tilfeller som opptrer i resultatet fra det ene søket og ikke i det andre og vice versa.

I det første tilfellet vil en få med ordformen *casos*, som ikke blir funnet i det andre søket, og i det andre søket vil en få med en form av verbet *casar*, som ikke blir funnet i det første søket.

8 Hva er det som er gulest i korpuset ConDiv?

Dette spørsmålet kan forstås på flere måter.

Hvis vi antar at det gjelder fargen gul og ikke gul som tegn på noe annet.

[pos="N.*"] [pos="ADJ.*" & func="N<" & grupo="Amarelo"]
Distribuição de lema

Svaret er *cartão, lâ* og *cor*.

Hvis det skulle være ordet *amarelo*

[pos="N.*"] [pos="ADJ.*" & func="N<" & lema="amarelo"]
Distribuição de lema

Svaret er *cartão, lâ* e *camisa*.

5 Hva for adjektiv (eller perfektum partisipp) blir mest brukt om bein i korpuset ConDiv?

[lema="perna"] @[pos="ADJ.*"]
Distribuição dos lemas

Svaret er *direito*.

Hvis studenten tilføyer dette

[lema="perna"] [pos!="N.*"]* @[pos="ADJ.*" & func="N<"] within s

blir svaret også *direito*.

eller

[lema="perna"] (@[pos="ADJ.*"]|@[pos="V.*" & temcagr="PCP"])

Får studenten 2 poeng, ellers bare 1.

6 I hvilken avis i korpuset Moçambula skriver en mest om helse?

[sema="saude.*"]

Distribuição por fonte

De må telle relativ hyppighet.

74/24263=0,0030

28/15577=0,0018

24/7218=0,0033

7/2711=0,0025

4/11204=0,00035

Det er i avisen *O País* som en drøfter helse mest.

7 Søk etter kjønn

Hvem er det som nevner lengsel oftest i korpuset Museu da Pessoa, menn eller kvinner?

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

Distribuição de sexo

Når en ser på den relative fordelingen, er det kvinner.

M: 0,000116405

F: 0,000211272

Hvis studentene skriver bare ordet eller lemmaet *saudade*, og resten av prosessen går som den skal, gir vi henholdsvis 2 og 2,5.

8 I korpuset CHAVE: Hvilke tre emosjoner blir oftest modifisert ved ordet *forte*?

(@[sema=".*emo.*"] "forte")(("forte" @[sema=".*emo.*"])

Distribuição dos lemas

sofrer

sentimento

emoção

Dersom studentene skulle legge merke til at søket gir et verb, og innskrenker søket til dette:

(@[sema=".*emo.*" & pos="N"] "forte")(("forte" @[sema=".*emo.*" & pos="N"])

får de 1,5, ellers får de bare 1.

sentimento

emoção

desejo

Dessuten kan en få 1,5 hvis en velger lemmaet i stedet for bare ordet *forte*:

(@[sema=".*emo.*"] [lema="forte"])([lema="forte"] @[sema=".*emo.*"])

Er det en forskjell mellom de to variantene av portugisisk her?

(@[sema=".*emo.*" & variante="BR"] "forte")(("forte" @[sema=".*emo.*" & variante="BR"])

sentimento

emoção

expectativa

```
(@[sema=".*emo.*" & variante="PT"] "forte")("forte" @[sema=".*emo.*" &
variante="PT"] )
sentimento
desejo
emoção
```

(For å få full pott her trenger de bare å ha gjort de to søkene og å ha satt uttrykket med «variante» på begge sider av |.)

9 Ved hjelp av et tilfeldig utvalg av 15 tilfeller fra korpuset NOBRE, skal du finne ut hvilke betydninger som ordet *hábito* har. (Nevn de norske oversettelsene av de forskjellige betydningene.)

```
[lema="hábito" & pos="N.*"]
Concordância
Amostra aleatória de 15 linhas.
```

Studentene må kopiere de 15 tilfellene de får som resultat, og gi en oversettelse/klassifisering av hvert av dem...

Og deretter eventuelt nevne hva som er den hyppigste, og hva for andre betydninger som han/hun har funnet (lage et sammendrag).

10 Hvilke to adjektiv forekommer hyppigst som subjektspredikativ i korpuset Literateca?

```
[pos="ADJ.*" & func=".*SC.*"]
Distribuição dos lemas
bom          7982
preciso      6626
```

Hvilken syntaktisk funksjon har hvert av dem som oftest?

```
[pos="ADJ.*" & lema="bom"]
>N
```

```
[pos="ADJ.*" & lema="preciso"]
<SC
```

Merknad: Adjektivet *bom* er ett av dem som forekommer flest ganger i korpuset (og i språket). Derfor er det ikke merkelig at det opptrer ofte som predikativ. På den annen side er adjektivet *preciso* mye mer vanlig som subjektspredikativ enn i andre syntaktiske sammenhenger.