

(Først på portugisisk, så på norsk)

1 Descrição de corpos

Compare os corpos Moçambula e Diáspora-TL

São ambos corpos pequenos. O primeiro contém cartas de leitores enviadas e publicadas em jornais moçambicanos, o segundo entrevistas a timorenses residentes em Portugal.

Tamanhos: 58 e 21 mil palavras respetivamente

2 Expressões regulares

Crie uma expressão regular para obter datas do século vinte e vinte e um.

(19|20)[0-9][0-9]

3 Crie uma expressão no AC/DC que obtenha palavras (formas) em português com pelo menos duas vogais.

`.*[aeiou].*[aeiou].*`

ou

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

4 Crie uma expressão regular no AC/DC que obtenha lemas que tenham a sequência *nh*, com mais de quatro letras.

`[lema=".*nh.*" & lema="....+"]`

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

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

6 Procure no AC/DC frases que contenham dois adjetivos separados por mais do que uma palavra ou sinal de pontuação. (Por exemplo, *calmo e discreto*, *amarelo mas muito encardido*, etc.)

`[pos="ADJ.*"] []+ [pos="ADJ.*"] within s`

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

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

8 Quais são as três coisas mais azuis no corpo ConDiv?

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

`[pos="N.*"] [pos="ADJ.*" & func="N<" & grupo="Azul"]`

Distribuição de lema

dá lã, algodão e seda

Se for a palavra *azul*

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

Distribuição dos lemas

dá lã, algodão e olho

5 Qual o adjetivo (ou particípio passado) mais usado para adjetivar pé(s), no corpo CHAVE?

[lema="pé"] @[pos="ADJ.*"]

Distribuição dos lemas

dá *esquerdo*

Se juntar

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

dá *direito*

ou

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

dá *esquerdo*

tem os 2 pontos, senão tem só 1.

6 Em que jornal do corpo Moçambula se fala mais de família?

[sema="familia.*"]

Distribuição por fonte

É preciso fazer a contagem relativa.

50/15577=0,0032

36/24263=0,0015

16/11204=0,0014

10/7218=0,0014

4/2711=0,0014

É no jornal Notícias que há mais uso de termos de família.

7 Pesquisa sobre género

Quem menciona mais o corpo humano no corpo Museu da Pessoa, são homens ou mulheres?

[sema="*.corpo.*"]

Distribuição de sexo

Fazendo a distribuição relativa, são as mulheres

M: 0,001428518

F: 0,001817865

8 No corpo Literateca, quais as três emoções mais vezes consideradas profundas?

(@[sema="*.emo.*"] [lema="profundo"])([lema="profundo"] @[sema="*.emo.*"])
Distribuição dos lemas

tristeza
amor
sentimento

Isto dá 1.5. Têm 1 se apenas puserem
[sema="*.emo.*"] [lema="profundo"]

Existe diferença se só considerarmos obras da escola literária do romantismo?

(@[sema="*.emo.*" & escola="*.romantismo"] [lema="profundo"])|
([lema="profundo" & escola="*.romantismo"] @[sema="*.emo.*"])

tristeza
amor
melancolia

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

[lema="pena" & 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 ECI-EBR?

[pos="ADJ.*" & func="*.SC.*"]
Distribuição dos lemas
bom 216
preciso 121

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

[pos="ADJ.*" & lema="bom"]
Distribuição de **func**
>N

[pos="ADJ.*" & lema="preciso"]
Distribuição de **func**

<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.

A bitola usada para atribuir as notas foi:

Reprova abaixo de 14, de 14 a 17 tem E, de 17 a 20 tem D, de 20 a 24 tem C, de 24 a 27 tem B, e tem A se tiver mais de 27 pontos.

NORSK:

1 Beskrivelse av korpus

Sammenlikn korpusene Moçambula og Diáspora-TL.

Begge er små korpora. Det første omfatter leserbrev sendt og publisert i mosambikanske aviser, og det andre inneholder intervjuer med folk fra Øst-Timor bosatt i Portugal. Omfang: 58 og 21 tusen ord.

2 Regulære uttrykk

Skriv et regulært uttrykk for å finne datoer i det tjuende og det tjueførste århundret.

(19|20)[0-9][0-9]

3 Skriv et uttrykk i AC/DC for å finne ord(former) på portugisisk som inneholder minst to vokaler.

.*[aeiou].*[aeiou].*

eller

[word=".*[aeiou].*[aeiou].*"]

4 Skriv et regulært uttrykk i AC/DC for å finne lemma som inkluderer *nh*, og som er på minst fire bokstaver.

[lema=".*nh.*" & lema="....+"]

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

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

6 Søk i AC/DC etter setninger som inneholder to adjektiver som har ett eller flere ord eller skilletegn mellom seg. (For eksempel *calmo e discreto*, *amarelo mas muito encardido*.)

[pos="ADJ.*"] []+ [pos="ADJ.*"] within s

7 Forklar forskjellen mellom de to søkene [lema="baixo"] og [word="baixo"], 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 *baixa*, som ikke blir funnet i det andre søket, og i det andre søket vil en få med en form av verbet *baixar*, som ikke blir funnet i det første søket.

8 Hva er de tre ting som er blåest i korpuset ConDiv?

Dette spørsmålet kan forstås på flere måter.

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

[pos="N.*"] [pos="ADJ.*" & func="N<" & grupo="Azul"]

Distribuição de lema

Svaret er *lã*, *algodão* og *seda*.

Hvis det skulle være ordet azul

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

Distribuição de lema

Svaret er *lã*, *algodão* og *olho*.

9 Hva for adjektiv (eller perfektum partisipp) blir mest brukt om føtter i korpuset CHAVE?

Hvis studenten svarer bare slik:

[lema="pé"] @[pos="ADJ.*"]

Distribuição dos lemas

Svaret er *esquerdo*.

Får han eller hun ett poeng.

Hvis studenten tilføyer dette

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

(blir svaret nå *direito*)

eller dette

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

Svaret er *esquerdo*.

får studenten 2 poeng.

10 I hvilken avis i korpuset Moçambula skriver en mest om familie?

[sema="familia.*"]

Distribuição por fonte

De må telle relativ hyppighet.

50/15577=0,0032

36/24263=0,0015

16/11204=0,0014

10/7218=0,0014

4/2711=0,0014

Det er i avisen *Notícias* som en drøfter helse mest.

11 Søk etter kjønn

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

[sema="*.corpo.*"]

Distribuição de sexo

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

M: 0,001428518

F: 0,001817865

12 I korpuset Literateca: Hvilke tre emosjoner blir oftest modifisert ved ordet *profundo*?

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

Distribuição dos lemas

tristeza

amor

sentimento

Hvis studenten bare skriver

[sema="*.emo.*"] [lema="profundo"]

får hun eller han 1. Ellers 1,5.

Blir resultatet forskjellig hvis en bare leter i tekster fra romantikken?

(@[sema="*.emo.*" & escola="*.romantismo"] [lema="profundo"])|
([lema="profundo" & escola="*.romantismo"] @[sema="*.emo.*"])

tristeza

amor

melancolia

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

[lema="pena" & 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).

14 Hvilke to adjektiv forekommer hyppigst som subjektspredikativ i korpuset ECI-EBR?

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

Distribuição dos lemas

bom	216
-----	-----

preciso	121
---------	-----

Hvilken syntaktisk funksjon har hvert av dem som oftest?

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

Distribuição de **func**

>N

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

Distribuição de **func**

<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.

Terskelverdiene var som følgende:

Består ikke med mindre enn 14 poeng. Får E mellom 14 og 17 poeng. Får D med mellom 17 og 20 poeng. Får C med 20 til 24 poeng, B med 24 til 27 poeng, og A med flere enn 27 poeng.