

Os corpos da Linguateca na prática

Cláudia Freitas (Linguateca & PUC-Rio)

Programa

- Parte I – visão geral do AC/DC
 - o que é
 - como é feito
 - os corpos
 - a anotação
 - as etiquetas
- Parte II – Exemplos e exercícios
- Parte III – desdobramentos do AC/DC
 - Ensinador
 - VARRA

Linguateca

Há 10 anos atrás...

Projeto dedicado à

- Criação de recursos públicos destinados ao processamento computacional do português
 - Não adaptação direta das aplicações para o inglês
- Organização de avaliações conjuntas
- Criação e manutenção de um portal dedicado à área

Adaptado de Santos, 2009
Linguateca: Objectivos e resultados

Breve história dos corpos na Liguatca

- 1. tornar acessível (na rede) o que já existia
 - 2. criar e/ou melhorar (analizando) esses corpos
- Projecto AC/DC
- 3. adicionar a dimensão da tradução → COMPARA, CorTrad
 - 4. adicionar a dimensão da revisão humana → Floresta Sintá(c)tica, COMPARA
- 5. adicionar a possibilidade de criar corpos próprios
 - 6. adicionar a dimensão de corpos comparáveis
- Corpógrafo

Santos, 2009
Pesquisando corpos:
Foi você que pediu um Corpo Ferreira?

O projeto AC/DC

Acesso a corpos / Disponibilização de corpos

- ❑ Início em 1998 – 1999 Liderança: Diana Santos
- ❑ 20 corpos em português
- ❑ Cerca de 380 milhões de palavras
- ❑ Cerca de 17 milhões de frases
- ❑ Variantes do Brasil e de Portugal
- ❑ Diferentes tipos de texto (jornalístico, literário, técnico, blogs..)
- ❑ Anotados automaticamente pelo PALAVRAS (Bick, 2000) (análise morfossintática)
- ❑ Anotados automaticamente com campos semânticos de cor e de roupa (Corte-e-Costura (Mota & Santos 2009))

AC/DC – galeria dos corpos

Jornalístico geral

- CETEMPúblico
- NILC/São Carlos
- CHAVE

Jornalístico regional

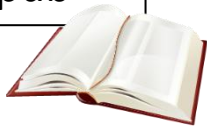
- NatMinho
- DiaCLAV

Jornalístico Específico

- Esportes : CONDIVport
- Política: Avante!
- Moda: CONDIVport
- Saúde: CONDIVport
- Ciências: CorTradjorn

Literário

- Vercial
- ClassLPPE
- ENPCpub



Transcrição de fala

- Museu da Pessoa
- ECI-EBR falado
- Selva falado



Textos Técnicos

- ECI-EE
- NILC/São Carlos
- Selva Ciência



Email

- Listas: ANCIB
- SPAM: CoNE



Recursos de Avaliação

CDHAREM



Blog

- Amazônia



Linguateca

[Estrutura](#)
[Equipa](#)

Apresentação

Acesso a recursos

- o AC/DC
- o [Procura](#)
- o [Frequência](#)
- o [PJR e exemplos](#)
- o [Corte-e-costura](#)
- o [Ensinador](#)
- o [VARRA](#)
- [CETEMPúblico](#)
- [CETENFolha](#)
- [CHAVE](#)
- [COMPARA](#)
- [Corpógrafo](#)
- [CorTrad](#)
- [Esfinge](#)
- [Floresta Sintá\(c\)tica](#)
- [GeoNetPT](#)
- [PAPEL](#)
- [ReEscreve](#)
- [REPENTINO](#)
- [Repositório](#)
- [WebSpell](#)
- [WPT](#)

Catálogo de recursos

Catálogo de ferramentas

Catálogo de atores

Catálogo de publicações

Informação

interessante

Fórum

Avaliação conjunta

Perguntas já respondidas

English

Norsk

Comentários e sugestões

Os apontadores na janela do lado direito não actualizam o lado esquerdo do ecrã. Sugerimos, pois, que naveguem

Acesso a corpos de português: Projecto AC/DC

[Linguateca](#)

[Information in English](#)

O projecto AC/DC (Acesso a corpos/Disponibilização de corpos), iniciado em 1999, surgiu da necessidade de juntar os poucos recursos disponíveis num único ponto na rede e dessa forma facilitar a comparação e a reutilização do material, permitindo ao mesmo tempo acesso a uma ferramenta poderosa de interrogação de corpos, o sistema [CWB](#) (versão nova do [IMS corpus workbench](#)), para o qual desenvolvemos esta interface.

Desde 2000, a anotação dos corpos tem sido feita automaticamente pelo [PALAVRAS](#) de Eckhard Bick, e convertida para o "formato AC/DC", descrito pormenorizadamente na página de [Anotação](#).

Uma descrição quantitativa inicial dos corpos servidos presentemente pelo AC/DC encontra-se na tabela abaixo. **Clique num dos corpos para o interrogar.** Para cada corpo, pode pedir concordâncias, distribuição e frequências simples e complexas, veja os [exemplos](#). Se é a primeira vez que visita o AC/DC e quer apenas experimentar, [procure no corpo Vercial](#).

Breve descrição dos corpos

Corpo	Tamanho (unidades)	Tamanho (palavras)	Tamanho (frases)	Variante(s)	Breve descrição
AmostRA-NILC	127.832	98.786	4.965	BR	AmostRA-NILC
ANCIB	1.690.376	1.258.764	80.992	BR	Correio electrónico correspondente ao tráfego na lista ANCIB
Avante!	7.766.309	6.501.146	204.414	PT	Semanário político Avante!, 1997-2002
CD HAREM	290.001	225.766	12.558	PT BR	Colecção dourada do HAREM
CETEMPúblico	232.543.379	189.575.095	7.665.410	PT	Jornal PÚBLICO, dividido em extractos, 1991-1998
CHAVE	123.936.528	99.355.321	4.740.448	PT BR	Jornais PÚBLICO e Folha de São Paulo, 1994-1995
CONDIVport	7.089.872	5.577.161	328.214	PT BR	Jornais desportivos e revistas de moda e saúde
CoNE	925.262	685.244	31.562	PT BR	Mensagens de correio electrónico não-endereçadas
DiaCLAV	7.758.469	6.651.523	232.152	PT	Diário de Coimbra, Diário de Leiria, Diário de Aveiro, Viseu Diário
ECI-EBR	917.124	724.008	44.381	BR	Texto do corpo Borba-Ramsey, compilado pelo ECI
ECI-EE	32.034	27.138	839	PT	Texto de chamada do programa europeu ESPRIT
ENPCPUB (parte em português)	92.688	72.375	4.371	PT BR	Literatura traduzida do inglês proveniente do ENPC
Floresta	7.252.252	5.950.072	327.179	PT BR	Floresta Sintá(c)tica
FrasasPB	23.313	19.162	653	BR	Frasas em português do Brasil
FrasasPP	20.048	16.232	594	PT	Frasas em português de Portugal
Museu da Pessoa	517.740	375.149	27.288	PT BR	Entrevistas realizadas pelo Museu da Pessoa
Natura/Minho	2.156.280	1.749.068	68.910	PT	Jornal regional Diário do Minho, antes da revisão
NILC/São Carlos	42.519.799	32.303.761	1.954.909	BR	Texto do corpo NILC, contendo maioritariamente texto jornalístico, mas também cartas comerciais e textos didácticos
todos juntos	18.851.915	14.313.624	596.869	todas	Todos os corpos juntos
Vercial	18.851.915	14.313.624	596.869	PT	Clássicos da literatura portuguesa, séculos XVI a XX
Total	473.363.136	379.793.019	16.923.577	PT BR	todos os corpos

Anotação dos corpos

Adaptado de Rocha (2007)

Cada corpo é anotado sintacticamente.



PALAVRAS

```

$START
Cada      [cada] <quant> DET M S @>N
corpo     [corpo] N M S @SUBJ>
é         [ser] <fmc> V PR 3S IND VFIN @FAUX
anotado   [anotar] V PCP M S @IMV @#ICL-AUX<
sintacti  ente ALT sintaticamente      [sintático] ADV @<ADVL
$.
    
```

Formato AC/DC

forma

lema

classe gramatical
(PoS)

tempo

num/pes

gênero

função

forma	lema	classe gramatical (PoS)	tempo	num/pes	gênero	função
Cada	cada	DET_quant	0	S	M	>N
corpo	corpo	N	0	S	M	SUBJ>
é	ser	V_fmc	PR_IND	3S	0	FAUX
anotado	anotar	V	PCP	S	M	IMV_#ICL-AUX<
sintaticamente	sintático	ADV	0	0	0	<ADVL

	lema	classe gramatical (PoS)	tempo	num/pes	gênero	função
Cada	cada	DET_quant	0	S	M	>N
corpo	corpo	N	0	S	M	SUBJ>
é	ser	V_fm	PR_IND	3S	0	FAUX
anotado	anotar	V	PCP	S	M	IMV_#ICL-AUX<
sintaticamente	sintático	ADV	0	0	0	<ADVL

Projecto AC/DC: corpo todos juntos

AC/DC : [Linguateca](#)

Procurar:

- Resultado:**
- Concordância
 - Distribuição das formas
 - Distribuição dos lemas
 - Distribuição da categoria gramatical (PoS)
 - Distribuição do tempo verbal e/ou do caso pronominal
 - Distribuição de pessoa e/ou número
 - Distribuição do género morfológico
 - Distribuição da função sintáctica
 - Distribuição por variante do português
 - Distribuição por campo semântico
 - Distribuição por grupo (de cor, roupa, etc.)
 - Distribuição por corpo

Tipo	Diversos
Variante(s)	todas
Tamanho (unidades)	18.9 milhões
Tamanho (palavras)	14.3 milhões

[Página principal](#)
Procure noutros corpos:
[AmostRA-NILC](#) [ANCIB](#) [Avante!](#) [CD HAREM](#) [CETEMPúblico](#)
[CHAVE](#) [CONDIVport](#) [CoNE](#) [DiaCLAV](#) [ECI-EBR](#) [ECI-EE](#)
[ENPCPUB](#) (parte em português) [Floresta](#) [FrasesPB](#) [FrasesPP](#)
[Museu da Pessoa](#) [Natura/Minho](#) [NILC/São Carlos](#) **todos juntos**
[Vercial](#)

Alguns atributos variam conforme o corpo escolhido

Opções

- Resultados por ordem alfabética (só distribuições)
- Ignorar maiúsculas/minúsculas (não admite parâmetros)

Amostra aleatória de linhas.

Cada	cada	DET_quant	0	S	M	>N
corpo	corpo	N	0	S	M	SUBJ>
é	ser	V_fmc	PR_IND	3S	0	FAUX
anotado	anotar	V	PCP	S	M	IMV_#ICL-AUX<
sintaticamente	sintático	ADV	0	0	0	<ADVL

Como entender?

Explorando a página do projeto

Procurar: OK

- Resultado:**
- Concordância
 - Distribuição das formas
 - Distribuição dos lemas
 - Distribuição da categoria gramatical (PoS)
 - Distribuição do tempo verbal e/ou do caso pronominal
 - Distribuição de pessoa e/ou número
 - Distribuição do género morfológico
 - Distribuição da função sintáctica
 - Distribuição por variante do português
 - Distribuição por campo semântico
 - Distribuição por grupo (de cor, roupa, etc.)
 - Distribuição por corpo
- Opções**
- Resultados por ordem alfabética (só distribuições)
 - Ignorar maiúsculas/minúsculas (não admite parâmetros)
- Amostra aleatória de linhas.

Perguntas já respondidas!

Linguateca

[Estrutura](#)
[Equipa](#)

[Apresentação](#)
[Acesso a recursos](#)

- AC/DC
 - - [Procura](#)
 - - [Frequência](#)
 - - [PJR e exemplos](#)
 - - [Corte-e-costura](#)
 - - [Ensinador](#)
 - - [VARRA](#)
- [CETEMPúblico](#)
- [CETENFolha](#)
- [CHAVE](#)

Projecto AC/DC:

AC/DC : [Linguateca](#)

Procurar:

Resultado:

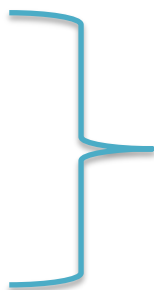
- Concordância
- Distribuição das formas
- Distribuição dos lemas
- Distribuição da categoria gramatical
- Distribuição do tempo verbal e/ou
- Distribuição de pessoa e/ou número
- Distribuição do género morfológico

Estrutura do corpo

Veja um [excerto do corpo](#) e [informação adicional](#).

Exemplo de pesquisas

	Procurar:	Resultado:
a palavra <i>inteiro</i>	inteiro	concordância
palavras terminando em <i>ístico</i>	".*ístico"	concordância
adjectivos antecedidos por uma preposição	[pos="PRP"] [pos="ADJ"]	concordância
formas do verbo <i>reunir</i>	[lema="reunir"]	concordância
substantivos modificados por <i>muito</i>	[lema="muito" & pos="DET" & func=">N"] @[pos="N.*"]	distribuição de lema



Dados quantitativos

- [Lista de formas deste corpo e sua frequência](#) (veja também [esta lista](#))
- [Lista de lemas deste corpo](#) (veja também [esta lista](#))
- [Dados sobre palavras gráficas](#)
- [Contagens de atributos estruturais](#)
- [Contagens de multipalavras](#)
- [Distribuição por categoria gramatical](#)

Para informação sobre como foram obtidos estes valores, consulte a página dos [Corpos](#)

Explicações e legendas da Anotação!

+ exemplos

INFORMAÇÕES ÚTEIS (1)

Lema

Substantivos (N): singular

Adjetivos (ADJ): masculino singular

[lema="menino"] → menino, meninos

[lema="menina"] → menina, meninas

[lema="menininho"] → menininho, menininhos

[lema="menininha"] → menininha, menininhas

[lema="bonito"] → bonito; bonitos, bonita; bonitas

[lema="bonitinho"] → bonitinho; bonitinhos, bonitinha; bonitinhas

outros atributos

classe gramatical → [pos="XXX"]

função sintática → [func="XXX"]

tempo verbal e/ou do caso pronominal → [temcagr="XXX"]

Mas não é preciso saber tudo!
Se quero saber como é o atributo "gênero morfológico" ...



Linguatca

[Estrutura](#)

[Equipa](#)

[Apresentação](#)

[Acesso a recursos](#)

- AC/DC
- - [Procura](#)
- - [Frequência](#)
- - [PJR e exemplos](#)
- - [Corte-e-costura](#)
- - [Ensinador](#)
- - [VARRA](#)
- [CETEMPúblico](#)
- [CETENFolha](#)
- [CHAVE](#)
- [COMPARA](#)
- [Corpógrafo](#)
- [CorTrad](#)
- [Esfinge](#)
- [Floresta Sintá\(c\)tica](#)
- [GeoNetPT](#)
- [PAPEL](#)
- [ReEscreve](#)
- [REPENTINO](#)
- [Repositório](#)
- [WebJspell](#)
- [WPT](#)

[Catálogo de recursos](#)

Projecto AC/DC: corpo todos juntos

AC/DC : [Linguatca](#)

Procurar: "/>

Resultado:

- Concordância
- Distribuição das formas
- Distribuição dos lemas
- Distribuição da categoria gramatical (PoS)
- Distribuição do tempo verbal e/ou do caso pronominal
- Distribuição de pessoa e/ou número
- Distribuição do género morfológico
- Distribuição da função sintáctica
- Distribuição por variante do português
- Distribuição por campo semântico
- Distribuição por grupo (de cor, roupa, etc.)
- Distribuição por corpo

Opções

- Resultados por ordem alfabética (só distribuições)
- Ignorar maiúsculas/minúsculas (não admite parâmetros)

Amostra aleatória de linhas.

- AC/DC
- Procura
- Frequência
- PJR e exemplos
- Corte-e-costura
- Ensinador
- VARRA
- CETEMPúblico
- CETENFolha
- CHAVE
- COMPARA
- Corpógrafo
- CorTrad
- Esfinge
- Floresta Sintá(c)tica
- GeoNetPT
- PAPEL
- ReEscreve
- REPENTINO

Resultados da procura

Sat Nov 5 16:54:17 WET 2011

Procura: [lema="menina"]

Distribuição de **gen**

Corpo: os corpos todos

10378 casos.

Distribuição

Houve 1 valores diferentes de género.

F 10378

Perguntas, comen...

Como representar o pronome "eu"?

[word="eu"]
Distribuição de pos

PERS

gen="F"

Projecto AC/DC: corpo todos juntos

AC/DC : Linguateca

Procurar: [lema="artista" & gen="F"]

OK

Resultado:

- Concordância
- Distribuição das formas
- Distribuição dos lemas
- Distribuição da categoria gramatical (PoS)
- Distribuição do tempo verbal e/ou do caso pronominal
- Distribuição de pessoa e/ou número

Como representar o sujeito?

[word="ele"][word="disse"]
Distribuição de func

SUBJ>

INFORMAÇÕES ÚTEIS (2)

Tudo o que está entre colchetes corresponde a um elemento/uma unidade/um item lexical

- Procura: [lema=".*udo" & pos="ADJ"] -> *agudas...*
 - **Um** item: palavra que termina em -udo & que seja ADJ)
- Procura: [lema=".*udo"] [pos="ADJ"] -> *estudo microscópico*
 - **Dois** itens: palavra que termina em -udo +(imediatamente seguida por) palavra que é adjetivo
- Adjetivos no masculino plural imediatamente seguidos de nomes no masculino plural

[pos="ADJ" & gen="M" & pessnum="P"] @[pos="N" & gen="M" & pessnum="P"]

Quais são os adjetivos?

Quais são os substantivos?

Subst masc + subst femin + participio no feminino → “os critérios e as opções selecionadas”

[pos="N" & gen="M"] [pos="KC.*"] [pos="N" & gen="F"] [temcagr="PCP" & gen="F"]

Resultado: concordância

*Os **esforços e propostas apresentadas** para a revisão...
mas em **laboratórios ou instituições situados** em países
diferentes,...
consumira um **alimento ou água contaminada** no México...
em **artigos e resenhas enviados** por internautas...
waypoints ou rotas criadas em casa...*

... e para saber como é feita a concordância de gênero
nesses casos?

[pos="N" & gen="M"] [pos="KC.*"] [pos="N" & gen="F"] @ [temcagr="PCP"]

Distribuição de **gen**

Corpo: os corpos todos 3666 casos.

Distribuição

Houve **2** valores diferentes de **género**.

F 2148

M 1518

EBRALC - 2011

INFORMAÇÕES ÚTEIS (3)

CARACTERES ESPECIAIS

. (um caractere)

!= (negação)

* (opcionalidade/ um elemento pode aparecer ou não)

Alternância: | []

.* (qualquer coisa)

@

[]* (zero ou mais palavras) (“coringa”)

[].* (uma ou mais palavras) (“coringa”)

palavras que terminam em *-udo* ou *-ento* e que NÃO são ADJ nem N

```
[lema=".*udo|.*ento" & pos!="ADJ|N"]
```

```
[lema=".*[aei]ndo" & temcagr!="GER" & variante="BR"]
```

palavras que termina em *-ando* ou *-endo* ou *-indo* e que NÃO estão no gerúndio e apenas no português do Brasil

AC/DC – alguns exercícios simples

- Quais os verbos que começam com des- ?

– Procura: [lema="des.*" & pos="V"]

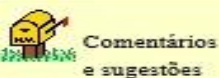
– Resultado: Distribuição de lema

NILC / São Carlos – 36239 casos	
desenvolver	3123
destinar	2417
descobrir	2256
desejar	1538
destruir	1497
descrever	1363
destacar	900
desconhecer	864
desaparecer	863

AC/DC – alguns exercícios simples

- Encontre, no corpo Amostra NILC, os advérbios terminados em -mente:
 - PASSO 1: *Como os advérbios foram anotados pelo PALAVRAS?*
 - Procura: [word=".*mente"]
 - Resultado: Distribuição de pos

- AC/DC
- Procura
- Frequência
- PJR e exemplos
- Corte-a-costura
- Ensinador
- VARRA
- CETEMPúblico
- CETENFolha
- CHAVE
- COMPARA
- Corpógrafo
- CorTrad
- Esfinge
- Floresta Sintá(c)tica
- GeoNetPT
- PAPEL
- ReEscreve
- REPENTINO
- Repositório
- WebJspell
- WPT



Os apontadores na janela do lado direito não actualizam o lado esquerdo do ecrã. Sugerimos, pois, que naveguem exclusivamente através dos apontadores do lado esquerdo.

Procura: [word=".*mente"]

Distribuição de pos

Corpo: os corpos todos

1314282 casos.

Distribuição

Houve 43 valores diferentes de pos.

ADV	1211383
ADV_quant	84805
N	4979
PROP	4688
V_fmc	2709
ADJ	2547
V	1597
ADV_kc	297
IN	217
V-REGIDO	213
V-REGENTE_fmc	205
ADJ_n	175
V_DERS_fmc	119
V-REGENTE	118
NUM_card	31
V+ADV	29
V-REGIDO_fmc	29
V-REGENTE-NEG_fmc	23
ADV_NUMord	22
N_prop	16
ADJ+ADV	10
V_DERP_fmc	10
ADV_quant_dem+ADV	7
V-REGENTE-NEG	7
ADJ_prop	7
PERS+ADV	6
V_fmc+ADV	6
PRP	6
V-REGENTE_DERS_fmc	4
PERS+ADV_quant	2
V-REGIDO_DERS_fmc	2
PRP+ADV	2
DET_quant+ADV	1
V-REGENTE+ADV	1

Todos esses valores podem ser investigados por meio do atributo "pos"

[word=".*mente" & pos="V-REGIDO"]

[word=".*mente" & pos="PERS+ADV"]

etc

etc

...

A visualização dos resultados deve ser "Concordância"!

- AC/DC
- Procura
- Frequência
- PJR e exemplos
- Corte-a-costura
- Ensinador
- VARRA
- CETEMPúblico
- CETENFolha
- CHAVE
- COMPARA
- Corpógrafo
- CorTrad
- Esfinge
- Floresta Sintá(c)tica
- GeoNetPT
- PAPEL
- ReEscreve
- REPENTINO
- Repositório
- WebJspell
- WPT



Os apontadores na janela do lado direito não actualizam o lado esquerdo do ecrã. Sugerimos, pois, que naveguem exclusivamente através dos apontadores do lado esquerdo.

Procura: [word=".*mente"]

Distribuição de pos

Corpo: os corpos todos

1314282 casos.

Distribuição

Houve 43 valores diferentes de pos.

ADV	1211383
ADV_quant	84805
N	4979
PROP	4688
V_fmc	2709
ADJ	2547
V	1597
ADV_kc	297
IN	217
V-REGIDO	213
V-REGENTE_fmc	205
ADJ_n	175
V_DERS_fmc	119
V-REGENTE	118
NUM_card	31
V+ADV	29
V-REGIDO_fmc	29
V-REGENTE-NEG_fmc	23
ADV_NUMord	22
N_prop	16
ADJ+ADV	10
V_DERP_fmc	10
ADV_quant_dem+ADV	7
V-REGENTE-NEG	7
ADJ_prop	7
PERS+ADV	6
V_fmc+ADV	6
PRP	6
V-REGENTE_DERS_fmc	4
PERS+ADV_quant	2
V-REGIDO_DERS_fmc	2
PRP+ADV	2
DET_quant+ADV	1
V-REGENTE+ADV	1

Advérbios terminados em -mente...

[word=".*mente" & pos="ADV.*"]

A visualização dos resultados deve ser "Distribuição dos lemas"!

2. a. No corpo Amostra-NILC, quais as palavras terminadas em –mente que não são advérbios?

Procura: [lema=".*mente" & pos!="ADV.*"]

Resultado: Distribuição de lemas

2. b. Qual a classe gramatical das palavras encontradas em 2 a. ?

Procura: [lema=".*mente" & pos!="ADV.*"]

Resultado: Distribuição de Pos

ACDC – exercícios

3. Encontre, no corpo CHAVE, os adjetivos que modificam (e sucedem) “mulher”

Modificam ≠ co-ocorrem!
volume de *dinheiro sujo*
Manter o *dinheiro disponível* para saques

```
[pos="N"& lema="mulher"] @[pos="ADJ"]
```

Resultado: Distribuição de lemas

4. Encontre, no corpo CHAVE, os adjetivos que modificam (e antecedem) “mulher”

Distribuição

Houve 3 valores diferentes de **func.**

>N 578

N< 12

<OC 1

```
[pos="ADJ"] [pos="N"& lema="mulher"]
```

Distribuição da função sintática

```
[pos="ADJ" & func=">N"] [pos="N"& lema="mulher"]
```

Distribuição de lemas

Procura: [pos="N" & lema="mulher"] @[pos="ADJ" & func="N<"]

Distribuição de lema

Corpo: Coleção CHAVE v. 4.3
3410 casos.

grávido	185
negro	119
brasileiro	108
bonito	96
português	82
branco	77
nu	74
jovem	73
casado	72
fatal	57
americano	44
pobre	42
idoso	34
morto	33
muculman	...

Procura: [pos="ADJ" & func=">N"] [pos="N" & lema="mulher"]

Pedido de uma concordância em contexto

Corpo: Coleção CHAVE v. 4.3

578 ocorrências.

... e para comparar
entre variantes?

... & variante="BR"
... & variante="PT"

único	80
jovem	61
belo	45
futuro	27
respectivo	24
novo	23
grande	18
actual	17
lindo	14
velho	14
pobre	12
certo	11
próprio	11
só	10
atual	0

Quais os verbos mais usados em orações relativas?

Qual a distribuição da função sintática dos pronomes relativos em orações relativas?

- Como identificar orações relativas?

[word="que"]

- Como o PALAVRAS anota o pronome relativo no “que”?

- “Distribuição de Pos”

- Como é estrutura muito frequente, usar corpos pequenos.

Procura: [word="que"]
Distribuição de **pos**
Corpo: AmostRA-NILC v. 3.2

1929 casos.

Distribuição

Houve 7 valores diferentes de **pos**.

SPEC_rel	893
KS	874
SPEC_interr	80
ADV_foc	55
DET_interr	15
ADV_quant_interr	10
ADV	2

+ frequentes...
O que significam?

Procura: [word="que" & pos="SPEC_rel"]
Concordância em contexto

Procura: [word="que" & pos="KS"]
Concordância em contexto

Procura: [word="que" & pos="SPEC_rel"].

par=1: Antes de iniciarmos o estudo da origem da vida, é necessário conhecer alguns caracteres **que** distinguem os seres vivos dos

par=10: ART única exceção à regra da organização celular dos seres vivos são os vírus, organismos muito pequenos **que** só foram

par=11: Em geral, sua estrutura é composta de um envoltório protéico **que** abriga o material genético (cromossomos), não apresent

par=13: Os indivíduos **que** apresentam organização celular podem ser unicelulares (formados por uma única célula) multicelulares (

par=29: Presença de Ácido Nucléico: molécula orgânica complexa **que** contém informações vitais para os seres vivos .

par=39: Estes seres, **que** apareciam inexplicavelmente no lodo e no lixo, eram encarados como manifestações da vontade dos deus

par=47: Uma delas é a fórmula para se obterem ratos: «enche-se de trigo e fermento um vaso, **que** é fechado com uma camisa suj

par=49: Como já se sabe, os ratos **que** aparecem não se formam a partir da receita dada por Von Velmont, mas são atraídos pela r


Procura: [word="que" & pos="KS"].

par=17: É importante lembrar **que** o mecanismo de diferenciação celular dos seres vivos exige

par=19: O termo metabolismo pode ser aplicado tanto às atividades químicas **que** ocorrem em u

par=27: Neste caso, diz-se **que** o crescimento é de fora para dentro .

par=38: Antigas doutrinas da Índia, da Babilônia e do Egito ensinavam **que** rãs, cobras e crocod



Confirmação: o
que mais pode
estar marcado
como SPEC_rel?

Procura: [pos="SPEC_rel"]
Distribuição de lema
Corpo: AmostRA-NILC v. 3.2

951 casos.

Distribuição

Houve 2 valores diferentes de lema.

que 893
quem 58

... e "cujo"?
... e "qual"?

Mesmo processo

[lema="cujo"] [lema="qual"]
Distribuição de PoS Distribuição de PoS

OU...

Pedir concordâncias e usar o comando "mostrar categoria gramatical" → show +pos;[lema="qual|cujo"]

[pos="SPEC_rel"] [pos=".*rel"]
[pos="DET_rel"] Distribuição de pos
....rel???

Procura: show +pos; [lema="cujo|qual"].

par=30: Reprodução/N :/PU processo/N através/ADV do/PRP+DET_artd qual/DET_rel os/DET_artd organismos/N c

par=31: Evolução/N :/PU processo/N através/ADV do/PRP+DET_artd qual/DET_rel grupos/N de/PRP organismos/N
originando/V grupos/N diferentes/ADJ de/PRP organismos/N ./PU

par= DV do/PRP+DET_artd qual/DET_rel as/DET_artd células/N mo

par= iu/V_fmc a/DET_artd teoria/N da/PRP+DET_artd geração/N espe
a/DE T_artd seres/N vivos/ADJ originam-se/V+PERS_refl da/PRP+DE

par= nterr é/V_fmc o/DET_artd assunto/N ?/PU

show +func;
show +temcagr;
...ou qualquer outro atributo!
...ou mais de um atributo!!
show +func; show +temcagr;

Procura: [pos="*.rel"]
 Distribuição de pos
 Corpo: AmostRA-NILC v. 3.2
 1572 casos.

Distribuição

Houve 8 valores diferentes de pos.

SPEC_rel	951
ADV_prp_rel	273
ADV_ks_rel	201
DET_rel	69
ADV_rel	65
ADV_quant_ks_rel	7
DET_quant_rel	3
ADV_quant_rel	3

... quais os verbos mais usados em orações relativas?

[pos="*.rel"] [pos!="V.*"]* @[pos="V.*"]
 Distribuição dos lemas

Garante que entre o pron. relativo e o verbo pode haver outros elementos (isto é, outro(s) elemento(s) que não seja(m) um verbo)

Corpo: AmostRA-NILC v. 3.2



Distribuição

Houve 548 valores diferentes de lema

ser	172
estar	52
ter	48
poder	41
haver	31
fazer	29
ir	21
ocorrer	21
apresentar	19
vir	18
dar	18
saber	15

Qual a distribuição da função sintática dos pronomes relativos em orações relativas?

Distribuição da função sintática

Procura: [pos="*.rel"] [pos!="V.*"]* [pos="V.*"]
 Distribuição de func
 Corpo: AmostRA-NILC v. 3.2

1522 casos.

Distribuição

Houve 77 valores diferentes de func.

SUBJ>_#FS-N<	592
ACC>_#FS-N<	189

Retirar o sinal "@" !!!

Sujeitos e objetos do verbo **CHOVER**, quando está flexionado na 3ª pessoa do plural?
chovem/choveram dólares
As nuvens choveram (?)

[lema="chover" & pessnum="3P"]

Como saber como estão marcados os **sujeitos e os objetos** de chover na 3ª pess plural?

Basta acrescentar

show +O_TIPO_DE_ATRIBUTO_QUE_SE_DESEJA_VER;

Exibir função
(concordância)

show +func; [lema="chover" & pessnum="3P"]

Como estão marcados os sujeitos e os objetos de CHOVER?

Resultados da procura

show +func; [lema="chover" & pessnum="3P"]

Mon Nov 7 00:55:40 WET 2011

Procura: show +func; [lema="chover|nevar|ventar|trovejar|gear|garoar" & pessnum="3P"]

Pedido de uma concordância em contexto

Corpo: os corpos todos

97 ocorrências.

Concordância

Procura: show +func; [lema="chover|nevar|ventar|trovejar|gear|garoar" & pessnum="3P"].

par=1149: A/>N Light/SUBJ> apela/FMV para/<ADVL a/>N população/P< /PONT **chovam**/FMV todos/<ACC /PONT por/ADVL favor/P< /PONT

par=2704: (/PONT P/SUBJ> em/N< plano/P< superior/N< joga/FMV água/PRED>)PONT Sobre/<ADVL este/>N vidro/P< **ventam**/FMV todos/>N as/>N ventos/PRED> que/SUBJ> atravessam/FMV esta/>N cidade/<

par=fut2035: Hão-de/FAUX+PRT-AUX< **chover**/IMV_#ICL-AUX< /PONT após/<ADVL este/>N desafio/P< /PONT as/>N censuras/<ACC à/<ADVL+>N equipa/P< seleccionada/N< /PONT

par=fut4767: Ruas/NPHR atalhadas/N< de/N< gente/P< /PONT bandeiras/NPHR /PONT filarmónica/NPHR a/N< tocar/IMV_#ICL-P< /PONT pétalas/NPHR de/N< flores/P< a/ADVL «PONT **choverem**/IMV_#ICL- todas/>N as/>N janelas/P< /PONT foguetório/NPHR /PONT «PONT vivas/FMV»PONT /PONT palmas/<PRED etc./<ADVL /PONT etc/<ADVL /PONT

par=fut5452: O/>N árbitro/<ACC apita/FMV o/>N público/<SUBJ protesta/FMV insultando/IMV_#ICL-<ADVL o/>N árbitro/<ACC /PONT os/>N fiscais/SUBJ> de/N< linha/P< não/ADVL> podem/FAUX assinalar/IM mais/>A insignificante/>N falta/<ACC /PONT e/CO não/ADVL> vai/FMV longe/<ADVS o/>N tempo/<SUBJ em/<ADVL que SUB_#FS-P< as/>N pedras/SUBJ> e/CO outros/>N objectos/SUBJ> caíam/FMV no/<PIV+ como/COM_#AS-<ADVL se/SUB_#FS-A< estivessem/FAUX **chovendo**/IMV_#ICL-AUX< /PONT

par=fut9042: Depois/ADVL>.../PONT tudo/ACC> segue/FMV no/<ADVL+>N mesmo/>N ritmo/P< vivo/N< e/CO as/>N bolas/SUBJ> **chovem**/FMV na/<ADVL+>N baliza/P< da/N<+>N «PONT reserva/P<»/PONT...

par=fut21735: É/FMV certo/<SC que/SUB_#FS-<SUBJ vão/FAUX **chover**/IMV_#ICL-AUX< a/<ADVL cântaros/P< as/>N teorias/<ACC da/N<+>N pressão/P< psicológica/N< sobre/N< a/>N equipa/P< /PONT mas/C mesma/<SC a/>N questão/SUBJ> dos/N<+>N grandes/>N campeões/P< /PONT serem/IMV_#ICL-<SUBJ ou/CO não/>A capazes/<SC de/A< ultrapassar/IMV_#ICL-P< as/>N pressões/<ACC e/CO as/>N ansiedades/<A sobretudo/<ADVL quando/ADVL>_#FS-<ADVL se/ACC> trata/FMV de/<PIV vencer/IMV_#ICL-P< um/>N só/>N jogo/<ACC /PONT seja/FMV ele/<SC qual/SUBJ>_#FS-<SUBJ for/FMV /PONT para/<ADVL se/SU campeão/<SC /PONT

par=fut26347: Mal/ADVL> /PONT o/>N juiz/SUBJ> António/N< Camacho/N< deu/FMV o/>N jogo/<ACC por/<OC concluído/P< e/CO logo/ADVL> «PONT **choveram**/FMV»PONT as/>N declarações/<ACC de/N< intervenientes/N< no/N<+>N encontro/P< /PONT

Como estão marcados os sujeitos e os objetos de CHOVER?

Resultados da procura

show +func; [lema="chover" & pessnum="3P"]

Mon Nov 7 00:55:40 WET 2011

Procura: show +func; [lema="chover|nevar|ventar|trovejar|gear|garoar" & pessnum="3P"]

Pedido de uma concordância em contexto

Corpo: os corpos todos

97 ocorrência

Concordân

[lema="chover" & pessnum="3P"] [pos!="v.*"]* [func="<ACC"]

Procura: show +func; [lema="chover|nevar|ventar|trovejar|gear|garoar" & pessnum="3P"]

par=1149: A/>N Light/SUBJ> apela/FMV para/<ADVL a/>N população/P< /PONT chovam/FMV todos/<ACC /PONT por/ADVL favor/P< /PONT

par=2704: (/PONT P/SUBJ> em/N< plano/P< superior/N< joga/FMV água/PRED>)PONT Sobre/<ADVL este/>N vidro/P< ventam/FMV todos/>N os/>N ventos/PRED> que/SUBJ> atravessam/FMV esta/>N cidade/<

par=fut2035: Hão-de/FAUX+PRT-AUX< chover/IMV_#ICL-AUX< /PONT após/<ADVL este/>N desafio/P< /PONT as/>N censuras/<ACC à/<ADVL+>N equipa/P< seleccionada/N< /PONT

par=fut4767: Ruas/NPHR atalhadas/N< de/N< gente/P< /PONT bandeiras/NPHR /PONT filarmônica/NPHR a/N< tocar/IMV_#ICL-P< /PONT pétalas/NPHR de/N< flores/P< a/ADVL «PONT choverem/IMV_#ICL- todas/>N as/>N janelas/P< /PONT foguetório/NPHR /PONT «PONT vivas/FMV»PONT /PONT palmas/<PRED etc./<ADVL /PONT etc/<ADVL /PONT

par=fut5452: O/>N árbitro/<ACC apita/FMV o/>N público/<SUBJ protesta/FMV insultando/IMV_#ICL-<ADVL o/>N árbitro/<ACC /PONT os/>N fiscais/SUBJ> de/N< linha/P< não/ADVL> podem/FAUX assinalar/IMV_#ICL- mais/>A insignificante/IMV_#ICL- objectos/SUBJ> caíam/FMV no/<PIV- como/COM_#AS-<ADVL /PONT

par=fut9042: Depois/A /PONT «PONT reserva/P<»/PONT... /PONT

par=fut21735: É/FMV /PONT mas/C /PONT sobe/N< a/>N equipa/P< /PONT mas/C /PONT mesma/<SC a/>N questão/SUBJ> dos/N<+>N grandes/>N campeões/P< /PONT serem/IMV_#ICL-<SUBJ ou/<CO não/>A capazes/<SC de/A< ultrapassar/IMV_#ICL-P< as/>N pressões/<ACC e/CO as/>N ansiedades/<A sobretudo/<ADVL quando/ADVL>_#FS-<ADVL se/<ACC> trata/FMV de/<PIV vencer/IMV_#ICL-P< um/>N só/>N jogo/<ACC /PONT seja/FMV ele/<SC qual/SUBJ>_#FS-<SUBJ for/FMV /PONT para/<ADVL se/SUBJ> campeão/<SC /PONT

par=fut26347: Mal/ADVL> /PONT o/>N juiz/SUBJ> António/N< Camacho/N< deu/FMV o/>N jogo/<ACC por/<OC concluído/P< e/CO logo/ADVL> «PONT choveram/FMV»PONT as/>N declarações/<ACC de/N< intervenientes/N< no/N<+>N encontro/P< /PONT

Procura: [lema="chover" & pessnum="3P"] [pos!="V.*"]* @[func="<ACC"]

Distribuição de lema

Corpo: os corpos todos

356 casos.

Procura: [func="SUBJ>"] [pos!="V.*"]* [lema="chover" & pessnum="3P"]

Distribuição de lema

Corpo: os corpos todos

89 casos.

crítica	38
acusação	17
telefonema	15
protesto	15
queixa	13
pedra	11
carta	9
aplausos	7
elogio	6
calúnia	6
palma	6
pergunta	6
pedido	6
ameaça	4
insulto	4

crítica	11
acusação	7
pergunta	4
queixa	3
telefonema	2
pedido	1
quaquilhão	1

Este tipo de construção é mais frequente no singular ou plural?

SE QUIER TRABALHAR COM APENAS UMA DAS VARIANTES?

variante="..."

SE QUIER ELIMINAR UM CORPO DO CORPO TODOS?

corpo!="..."

SE QUIER USAR SOMENTE UM CORPO?

Exercícios

- Padrões de relações semânticas: como encontrar a expressão “X é um Y que” (ex: “*sonambulismo é um fenômeno que...*”)?

[pos="N"] [lema="ser"& pos="V"] [lema="um"& pos="DET_arti"]
[pos="N"] [word="que"]

www.linguateca.pt

Linguateca

[Estrutura](#)
[Equipa](#)

[Apresentação](#)
[Acesso a recursos](#)

- AC/DC
- - Procura
- - Frequência
- - PJR e exemplos
- - Corte-e-costura
- - Ensinador
- - VARRA
- CETEMPúblico
- CETENFolha
- CHAVE
- COMPARA
- Corpógrafo
- CorTrad
- Esfinge
- Floresta
- Sintá(c)tica
- GeoNetPT
- PAPEL
- ReEscreve
- REPENTINO

Resultados da procura

Sat Nov 5 19:09:46 WET 2011

Procura: [pos="N"] [lema="ser"&pos="V"] [pos=""]
Pedido de uma concordância em contexto
Corpo: os corpos todos

584 ocorrências.

Concordância

Procura: [pos="N"] [lema="ser"&pos="V"] [pos=""]

: Conseguimos coletar material para fazer análise do pr deslocamento, nós olhamos com mais atenção à segura

: Eu ganhava bem porque o **patrão era um senhor que**

: Conseguimos coletar material para fazer análise do pr deslocamento, nós olhamos com mais atenção à segura

...a **toxicod dependência é um problema que afecta...**

...o meu **pai foi uma pessoa que estudou...**

...que **dançarina é uma mulher que dança...**

...o **homem era um animal que possuía,...**

...a **avaliação é uma actividade que...**

...o **cientista é um indivíduo que precisa se isolar...**

...o **ciberespaço é um ambiente que nos permite...**

...porque o **cancro é uma doença que assume...**

...a **insulina era uma droga que não devia...**

...o **intestino é um órgão que reflecte...**



Reposi

WebJst

WPT

[Catálogo](#)

[recurso](#)

[Catálogo](#)

[ferram](#)

[Catálogo](#)

[atores](#)

[Catálogo](#)

[publica](#)

[Inform](#)

[interess](#)

[Fórum](#)

[Avaliac](#)

[conjunt](#)

[Pergun](#)

[respond](#)

English

Norsk

...porque a questão da **segurança foi um aspecto que pesou...**

...constatar que a criação de uma **obra é um processo que depende...**

O turismo na **favela é uma experiência que garante...**

...considerou ainda que a contestação ao **processo é uma obsessão que só pode ter...**

A cedência deste **terreno é um assunto que vinha a ser mastigado...**

...se a participação e a intervenção mais constante dos **cidadãos é um objectivo que merece...**

...a formação de platéia para esse tipo de **música é um processo que exige...**

10/11/2011 EBRALC - 2011 34
...ensino das ciências com crianças a partir dos seis anos de **idade é uma área que tem...**



Resultados da procura

Sat Nov 5 18:49:32 WET 2011

Procura: [pos="N"] [lema="ser"&pos="V"] [lema="um" & pos="DET_arti"] [pos="N"] [word="que"]

Distribuição de **func**

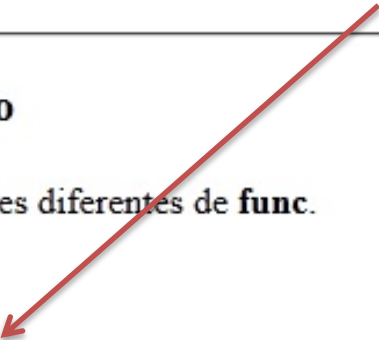
Corpo: os corpos todos

584 casos.

Distribuição

Houve **6** valores diferentes de **func**.

SUBJ>	358
P<	207
<SC	8
<ACC	5
PRED>	5
AS<	1



Resultados da procura

Sat Nov 5 18:53:18 WET 2011

Procura: [pos="N" & func="P<"] [lema="ser"&pos="V"] [lema="um" & pos="DET_arti"] [pos="N"] [word="que"]

Pedido de uma concordância em contexto

Corpo: os corpos todos

207 ocorrências.

[pos="N" & func!="P<"]

Eliminar esses casos!

Concordância

Procura: [pos="N" & func="P<"] [lema="ser"&pos="V"] [lema="um" & pos="DET_arti"] [pos="N"] [word="que"].

: Conseguimos coletar material para fazer análise do projeto, que infelizmente, não foi aprovado, porque a questão da **segurança foi um aspecto que** pesou muito na análise da em
olhamos com mais atenção à segurança do empreendimento, que era muito falha .

: Conseguimos coletar material para fazer análise do projeto, que infelizmente, não foi aprovado, porque a questão da **segurança foi um aspecto que** pesou muito na análise da em
olhamos com mais atenção à segurança do empreendimento, que era muito falha .

: Mas, mesmo assim, há tanta gente que acredita no futuro que parece que o negócio de **adivinho é um negócio que** dá dinheiro .

: Quantos de nós simplesmente estancaram, ou há muito abandonaram um lindo projeto, que na **verdade era um sonho que** o Próprio D-us colocou em seus corações por se achar

<p>: Quanto ao público da banda, Jorge afirma que a formação de platéia para esse tipo de **música é um processo que** exige paciência e muito esforço para manter a identidade m
anos 80, com belas e sofridas melodias, e letras que lidam com temas como solidão e depressão, " uma dança mais interior ", nas palavras do vocalista .

<p>: Basta conversar com qualquer artista para constatar que a criação de uma **obra é um processo que** depende do acesso a outras obras .

<p>: O turismo na **favela é uma experiência que** garante a você uma compreensão do que é a cultura brasileira, porque está tudo ali, diz Bianca .

<p>: A primeira investida para tentar politizar a entrevista, exibida nesta segunda, 24/9/2007 -- já que o que se espera ali no **programa é uma discussão que** renda posições sobre
comportamento humano...),

par=DA-N1389-1: Por outro lado, hoje é um facto inequívoco que o ensino das ciências com crianças a partir dos seis anos de **idade é uma área que** tem que ser fortemente desen
Martins .

par=DC-N1987-1: Os restantes saíram do Tribunal de Coimbra com a absolvição, mas serrar um automóvel ao **meio é um comportamento que** não é muito correcto e espero que
carros, sublinhava o juiz .

par=DC-N3115-1: José Cavalheiro considerou ainda que a contestação ao **processo é uma obsessão que** só pode ter entendimento do ponto de vista psiquiátrico .

Resultados da procura

Sat Nov 5 18:59:16 WET 2011

Procura: [pos="N" & func!="P<"] [lema="ser"&pos="V"] [lema="um" & pos="DET_arti"] [pos="N"] [word="que"]

Pedido de uma concordância em contexto

Corpo: os corpos todos

377 ocorrências.

Concordância

Procura: [pos="N" & func!="P<"] [lema="ser"&pos="V"] [lema="um" & pos="DET_arti"] [pos="N"] [word="que"].

: Eu ganhava bem porque o **patrão era um senhor que** morava mesmo no Muro dos Bacalhoeiros, os escritórios também eram lá, e ele tinha pena dos meus pais e dáva-me 50 escudos por semana

: A minha mãe é mais liberal, é outra pessoa idêntica ao meu pai só que o meu **pai foi uma pessoa que** estudou, no tempo dele quase que podia ser um professor, e a minha mãe é analfabeta .

: Eu que sempre fui uma pessoa que gostei de estudar, não era capaz de mudar de página enquanto não tivesse percebido a página anterior, era uma pessoa que o **conhecimento era uma coisa que**

: Eu ganhava bem porque o **patrão era um senhor que** morava mesmo no Muro dos Bacalhoeiros, os escritórios também eram lá, e ele tinha pena dos meus pais e dáva-me 50 escudos por semana

par=4228: O senhor ficará sabendo, além do mais, que **dançarina é uma mulher que** dança, que o pintinho diz piu-piu- piu e que o Rio de Janeiro é um porto que está no Oceano Atlântico .

par=6019: O próprio ato de escrever não é senão um modo de falar sozinho, segundo aquela definição do gordo Alexander Woollcott, para quem o **escritor é um homem que** passa a vida inteira o que a conversa, com o correr dos anos fica cada vez mais interessante .

par=11482: «Por muito tempo acreditou-se que o **homem era um animal que** possuía, de maneira espontânea e natural, o dom de ser sociável .

: Estamos plenamente conscientes de que a **avaliação é uma actividade que** exige muito trabalho e que constitui por si só uma área de investigação muito interessante .

: Não devemos deixar persistir o mito de que o **cientista é um indivíduo que** precisa se isolar nele mesmo mas alertar sobre a necessidade do treinamento de pessoas capacitadas no Brasil para ger

: Conclui-se que o **ciberespaço é um ambiente que** nos permite inúmeras possibilidades de mundo “ real ” .

par=3996: O facto de o relatório da CRES estar disponível na Internet e não ter suscitado grande interesse por parte da população foi lamentado por Daniel Serrão, alegando que a **saúde é um tem** geral .

par=13149: Mesquita Machado referiu em conferência de imprensa que esta **redução é uma medida que** era necessária aplicar, porque vai abranger várias pessoas da mesma família, ao invés do descontos beneficiam só o próprio utente .

par=30305: É como, segundo a mesma, uma história vive ambientes, há que ter em consideração os espaços, dado que a **televisão é um meio que** amplifica este aspecto .

- Verbos associados à preposição “a” ?

[pos="V.*"] [word="a" & func="<PIV"]

- Que palavras podem aparecer depois da preposição “a”, completando os verbos?

[pos="V.*"] [word="a" & func="<PIV"] @[word=".*"]

- Quais são os argumentos mais frequentes do verbo assistir?

[lema="assistir"] [func="<PIV"] [func=">N"]* @[pos="N.*"]

Ou

[lema="assistir"] [pos!="V.*"]* [func="<PIV"]
[func=">N"]* @[pos="N.*"]

assistira **absolutamente impassível e sério** a todas as minhas afirmações
Trinta mil pessoas assistiram **sucessivamente** a este singular desafio .
assistir **lá dentro** a esplêndidos morticínios

- Do que é que se tem medo?

```
[lema="medo"] [func="N<"]* [func="N<" & pos="PRP"] [func=">N"]*  
@[pos="N.*"]
```

... e o que se teme?

represália

retaliação

assalto

coisa

atentado

falar

avião

ataque

Um (?)

morte

dizer

sequestro

ameaça

água

fantasma

Procura: "de" a:[pos="N.*"] "em" @[word=a.word] within s

Distribuição de lema

Corpo: os corpos todos
4950 casos.

Procura: "de" a:[pos="N.*"] "em" @[word=a.word & variante="BR"] within s

Distribuição de lema

Corpo: os corpos todos
788 casos.

dois	559
três	321
quatro	258
mão	204
porta	185
geração	176
15	171
cinco	149
boca	143
seis	139
tempo	120
dez	110
meio	100
casa	95
terra	79
quinze	70

hora	54
30	42
cidade	35
20	32
vitória	31
10	29
oito	27
bar	26
mesa	26
sete	25
crise	24

tempo	95
porta	55
dois	45
mão	41
quatro	27
casa	26
hora	24
geração	23
15	19
três	19
boca	18
cinco	15
pé	13
dez	13
cidade	11
seis	9
galho	9

Problemas típicos com utilizadores avançados do AC/DC

- As categorias com que trabalham não são as que estão marcadas (ou que são subjacentes à anotação)
não há (e se calhar não pode haver) uma terminologia gramatical 100% consensual -- cada linguista / gramático faz as distinções que lhe parecem pertinentes e de certa forma divide a língua de forma diferente.
- Há erros na anotação ou no próprio texto do corpo
É humanamente impossível rever 350 milhões de palavras 😞
- O género textual do corpo com que estão a trabalhar não é apropriado para tirar conclusões

Santos, 2008

- Linguatca
- Estrutura
- Equipa
- Apresentação
- Acesso a recursos
 - AC/DC
 - Procura
 - Frequência
 - PJR e exemplos
 - Corte-e-costura
 - Ensinador
 - VARRA
 - CETEMPúblico
 - CETENFolha
 - CHAVE
 - COMPARA
 - Corpógrafo
 - CorTrad
 - Esfinge
 - Floresta Sintá(c)tica
 - GeoNetPT
 - PAPEL
 - ReEscreve
 - REPENTINO
 - Repositório
 - WebSpell
 - WPT
- Catálogo de recursos
- Catálogo de ferramentas
- Catálogo de atores
- Catálogo de publicações

Ensinador

Procurar para a secção

Corpo "CHAVE"

Sobre: O corpus CHAVE contém textos jornalísticos dos jornais [Público](#) e [Folha de São Paulo](#) dos anos de 1994 e 1995. Este corpus foi criado no âmbito da participação da língua portuguesa no [CLEF](#).

Tipo	Jornalístico
Variante(s)	PT BR
Tamanho (unidades)	123936 mil
Tamanho (palavras)	99355 mil
Anotado	sim

Marcadores estruturais: DOC [documento], s [frase], CATEGORY (secção do jornal), AUTHOR (autor do texto), NE (lista de nomes próprios contidos no documento, usando a anotação do PALAVRAS).

Veja um [excerto do corpo e informação adicional](#).

Agradecimentos: Agradecemos ao PÚBLICO e à Folha de São Paulo a gentil autorização de disponibilização dos textos.

Procure noutros corpos: [AmostrA-NILC](#) | [ANCB](#) | [Avante!](#) | [CD HAREM](#) | [CETEMPúblico](#) | [CHAVE](#) | [CONDIVport](#) | [CoNE](#) | [DiaCLAV](#) | [ECI-EBR](#) | [ECI-EE](#) | [ENPCPUB \(parte em português\)](#) | [Floresta](#) | [FrasesPB](#) | [FrasesPP](#) | [Museu da Pessoa](#) | [Natura/Minho](#) | [NILC São Carlos](#) | [todos juntos](#) | [Vercial](#)

Verbos *ir, estar* ou *fazer* no futuro do subjuntivo:
[temcagr="FUT_SUBJ" & lema="ir|fazer|estar"]

Escolher o corpo

Ensinador

A procurar “[temcagr="FUT_SUBJ" & lema="ir|fazer|estar"]” no corpus NILC/São Carlos v. 9.2 [2758 entradas.]

Selecione as concordâncias que deseja usar.

OK

- par=259*: Assim, por exemplo, se uma pessoa **estiver** numa floresta e vir folhas em movimento, ouvir um ruído e sentir o cheiro de um animal.
- par=341*: Se a mulher **estiver** na fase do seu ciclo sexual em que haja a produção do óvulo, este se unirá a um espermatozóide .
- par=346*: O estudante de sétima série que **estiver** usando este manual para estudo poderá observar em seu próprio corpo as transformações pro
- par=1983*: Dessa forma, se **estivermos** trabalhando no Sistema Internacional, em que a unidade de tempo é o segundo (s) , veremos que a unid
- par=2242*: Após liberar o calor, o sal será estocado a uma temperatura de 287,78 oC em outro tanque com isolantes térmicos e retornar ao prin
- par=2427*: Nossos movimentos respiratórios estão adaptados à pressão atmosférica normal; não conseguimos respirar ar normal, se **estivermo**
- par=2852*: puxar a língua dele para fora, se **estiver** retraída ;
- par=2886*: Essa massa será equivalente a 1 kg quando os pratos **estiverem** em equilíbrio .
- par=2965*: Quando ela **estiver** alguns graus abaixo de zero, retire o tudo de ensaio com cuidado e coloque-o num copo vazio .
- par=3012*: A força é o agente capaz de produzir o movimento de um corpo, se ele **estiver** em repouso, de modificá-lo, se ele estiver em movim
- par=3012*: A força é o agente capaz de produzir o movimento de um corpo, se ele estiver em repouso, de modificá-lo, se ele **estiver** em movim
- par=3077*: Porém, quanto mais distantes **estiverem** entre si, menor será a atração entre eles .
- par=3155*: Enquanto você **estiver** com o dedo no botão da campainha, esse constante ligar e desligar fará o martelo bater várias vezes na camp
- par=5409*: Se daqui a 10 anos a árvore **estiver** com 10 metros de altura, como ela cresce pela «ponta», a marca continuará no mesmo nível, ist
- par=7497*: Se o Brasil não **fizer** a reforma agrária, a situação vai ficar ainda pior» .
- par=7725*: Aliás, quem **estiver** interessado pode mandar seus CDs para nós .
- par=8037*: Perguntado se voltaria a ligar para o serviço de videntes no futuro, Fabrice respondeu: «Só se **for** para agradecer pelo fato de eu ter
- par=8136*: Quem **for** ao Sharp vai ter a oportunidade de assistir a um show de feras .
- par=8177*: Mas quem **for** assistir ao Rio cena contemporânea também vai ter a oportunidade de conferir trabalhos experimentais de diretores b
- par=8785*: Sei também que aquela pessoa que **estiver** representando meu pai na tela não será meu pai, mas a interpretação que outras pessoas
- par=8814*: -- Se eles **estiverem** interessados em mim, ótimo !

[Estrutura](#)
[Equipa](#)

[Apresentaçã](#)

[Acesso a recursos](#)

■ [AC/DC](#)
○ - [Procura](#)

○ - [Frequência](#)

○ - [PJR e exemplos](#)

○ - [Corte-e-costura](#)

○ - [Ensinador](#)
○ - [VARRA](#)

[CETEMPúblico](#)

[CETENFolha](#)

■ [CHAVE](#)
■ [COMPARA](#)
■ [Corpógrafo](#)

■ [CorTrad](#)

■ [Esfinge](#)

■ [Floresta](#)

[Sintá\(c\)tica](#)

■ [GeoNetPT](#)

■ [PAPEL](#)

■ [P.F.](#)

1. *par=259*: Assim, por exemplo, se uma pessoa _____ numa floresta
2. *par=341*: Se a mulher _____ na fase do seu ciclo sexual em que haja
3. *par=2965*: Quando ela _____ alguns graus abaixo de zero, retire o t
4. *par=3077*: Porém, quanto mais distantes _____ entre si, menor s
5. *par=3155*: Enquanto você _____ com o dedo no botão da campainh
6. *par=7497*: Se o Brasil não _____ a reforma agrária, a situação vai ficar
7. *par=8814*: -- Se eles _____ interessados em mim, ótimo !
8. *par=8889*: Quem não _____ à apresentação de hoje, vai precisar de muita s
9. *par=11647*: Enquanto a Secretaria de Segurança não _____ nada, vai co
10. *par=13266*: Assim, enquanto o mundo _____ assistindo ao julgame
11. *par=14492*: O edital exige que a consultoria que _____ a modelagem es
12. *par=18106*: Quando _____ as malas, não se esqueça do guarda-chuva .
13. *par=18688*: Se _____ considerando isso, o substituto terá de ser alg
14. *par=20292*: Diz que se o governo _____ isso, pode dar a impressão de c
15. *par=21031*: O pedestre que _____ atravessando a rua se assusta, o n
16. *par=21782*: Sei que, se _____ isso, amanhã vai faltar siri nos rios para e

Ver Solução

Aleatorizar

Descarregar:

enunciado

solução

Procurar

[itemcagr="FUT_SUBJ" & lema="]

para a secção

no corpus

SAOCARLOS

Adicionar

Preencha as lacunas com as formas adequadas dos verbos ser, estar e fazer

1. *par=3155*: Enquanto você _____ com o dedo no botão da campainha, esse constante ligar e desligar fará o martelo bater várias vezes na campânula, produzindo
2. *par=Empregos--94a-1*: «Para cada crítica que _____ ao desempenho de um funcionário, faça-lhe quatro elogios» .
3. *par=18688*: Se _____ considerando isso, o substituto terá de ser alguém de estrita confiança e que, além de tudo, não veja problema em ser ministro-tampão .
4. *par=8889*: Quem não _____ à apresentação de hoje, vai precisar de muita sorte (e algum dinheiro) para conseguir vê-los em alguns shows pequenos e isolados que ainda participações especiais no festival Lolapaloosa, quando não farão o show completo -- até o fim do ano, antes de uma merecida, porém triste para seus fãs, aposentadoria .
5. *par=Brasil--94b-1*: Se as partes _____ seu contrato em URV, daí para frente não muda .
6. *par=21782*: Sei que, se _____ isso, amanhã vai faltar siri nos rios para eu tirar minha sobrevivência», garante Edgar, que acorda diariamente antes do sol nascer e ca
7. *par=2965*: Quando ela _____ alguns graus abaixo de zero, retire o tudo de ensaio com cuidado e coloque-o num copo vazio .
8. *par=22026*: Sendo assim, um jogador de pólo que _____ se preparando para os Jogos Olímpicos que acontecerão daqui a oito anos deve ter hoje em torno de 19
9. *par=8814*: -- Se eles _____ interessados em mim, ótimo !
10. *par=341*: Se a mulher _____ na fase do seu ciclo sexual em que haja a produção do óvulo, este se unirá a um espermatozóide .
11. *par=21031*: O pedestre que _____ atravessando a rua se assusta, o motorista freia bruscamente .
12. *par=20292*: Diz que se o governo _____ isso, pode dar a impressão de que há algo a ser escondido .
13. *par=14492*: O edital exige que a consultoria que _____ a modelagem exija do concessionário menor tarifa que a atual e melhor prestação de serviço .
14. *par=21958*: Mesmo se os Jogos Olímpicos de 2004 _____ parar na Conchinchina .
15. *par=11647*: Enquanto a Secretaria de Segurança não _____ nada, vai continuar assim», disse o advogado .

Descarregar:**Procurar****para a secção****no corpus**

10/11/2011

45

Lamentar, confessar, afirmar que...

```
[lema="afirmar|lamentar|confessar"] [pos!="V.*"] [word="que" & pos="KS"]
```

Afirmar, lamentar ou confessar?

1. Alguns analistas _____ a crise dos liberais será um permanente factor de instabilidade no futuro executivo .
2. E _____ só quer ganhar o filho como resultado deste acerto de contas .
3. O técnico Carlos Alberto Parreira _____ a seleção brasileira precisará tocar mais a bola contra um adversário que ele prevê muito defensivo .
4. Ele _____ o pedido de prisão feito pelo Ministério Público não tem nenhuma ligação com a fiscalização feita pela Sunab no Makro .

Foi ou era? Estava ou esteve?

- Enquanto ____ disponível em caráter experimental, a Biblioteca Virtual Carlos Chagas chegou a receber mensagens de portadores da doença, que tiveram suas dúvidas esclarecidas por especialistas
- Este enunciado que só Jesus é a verdadeira alegria, é um enunciado que sempre ____ presente, sempre-já lá, como diz Pêcheux, na memória discursiva dos fiéis enquanto dogma religioso
- No ano em que, segundo Balzac, o poeta Dante Alighieri ____ em Paris, ele poderia ter lido todos os 1.338 volumes da biblioteca da universidade (que era, então, a maior da França)

Santos, 2008

VARRA – Validação de relações semânticas no AC/DC

Usar os recursos dos corpos do AC/DC para validar
relações semânticas na forma de triplos
palavra RELAÇÃO palavra

Podem vir de recursos lexicais externos:
ontologias, tesouros, léxicos

O VARRA gera dossiês com frases do AC/DC com as palavras-alvo
Os dossiês serão validados por pessoas
Validação 100% confiável 😊

Serve também para testar e descobrir padrões (usando a sintaxe do
AC/DC)

10/11/2012 a partir de relações prototípicas (feio / bonito)

Já existem, hoje, dossiês prontos para serem validados;

Disponíveis na página do VARRA;

As relações semânticas vêm do PAPEL.

PAPEL → recurso lexical público

→ parceria entre a Porto Editora e a
Linguateca

www.linguateca.pt/PAPEL

NOME:

Leia os exemplos e complete as colunas **Resposta** e **Comentário**.

Os textos dos exemplos ilustram a relação entre as duas palavras apresentada na primeira coluna?

- 1: Sim
- 2: Não. É compatível com a relação, mas não a exemplifica
- 3: Não. O texto é completamente não relacionado
- 4: Não. Pelo contrário, invalida-a
- 5: Não sei mesmo

Exemplo de um dossiê –
formato RTF

Para cada linha, escolha uma das possibilidades 1 a 5, e comente se achar necessário.

Relação	Procura	Exemplo	Resposta (1-5)	Comentário
chefe HIPERONIMO_DE reformador	MU meet chefe reformador s	<i>par=ext172808-nd-91b-1</i> : O ex-primeiro ministro, Moulod Hamrouche, líder dos reformadores , e o chefe de fila dos históricos, Mohammed Salah Yahiaoui, acabaram com efeito por integrar a nova comissão política, após a luta interna entre as duas facções ter provocado o arrastamento da reunião do comité central durante 20 dias .		
chefe HIPERONIMO_DE reformador	MU meet chefe reformador s	<i>par=ext655990-pol-98a-3</i> : A REPUBLIKA Srpska (RS, entidade sérvia da Bósnia) tem desde ontem um governo do agrado ocidental liderado pelo reformador Milorad Dodik chefe do Partido dos Sociais		

Exploração do comportamento
semântico das palavras em contexto!

VARRA: procurar e validar relações semânticas

VARRA: Validação, Avaliação e Revisão de Relações semânticas no [AC/DC](#), usando as relações do [PAPEL](#)

1) Procurar uma relação específica entre duas palavras, usando [padrões](#) já implementados:

- SINONIMO_DE
- HIPERONIMO_DE
- PARTE_DE
- CAUSADOR_DE
- FINALIDADE_DE
- ANTONIMO_DE
- INSTRUMENTO_PARA

ou 2) Procurar uma relação entre duas palavras, sem fazer recurso a padrões pré-estabelecidos:

Todas as relações possíveis

justiça sentimento

Restrição (opcional) da categoria gramatical:

Indique (opcionalmente) qual o triplo pretendido para validar, na forma palavra RELACAO palavra, para que a coluna respectiva no dossiê seja preenchida:

Opções

Não apresentar cabeçalho no dossiê

Tenho um recurso em que a relação/triplo justiça TIPO_DE sentimento foi obtida automaticamente...

TIPO_DE não é uma relação originalmente prevista no VARRA.. Mas Ok!

Obrigada!

... e visitem

<http://www.linguateca.pt>

AC/DC - <http://www.linguateca.pt/ACDC/>

Luís Costa, Diana Santos & Paulo Alexandre Rocha. "Estudando o português tal como é usado: o serviço AC/DC", STIL 2009. STIL 2009, The 7th Brazilian Symposium in Information and Human Language Technology (São Carlos, Brasil, 8-11 de Setembro de 2009).

Cristina Mota & Diana Santos (2009) . "Corte e costura no AC/DC: auxiliando a melhoria da anotação nos corpos".

<http://www.linguateca.pt/aceso/corte-e-costura.pdf>

Diana Santos. "Pesquisando corpos: Foi você que pediu um Corpo Ferreira?". *Escola de Verão Belinda Maia (Edv 2009)* (FLUP, Porto, Portugal, 29 de Junho - 3 de Julho 2009)

Diana Santos. "Corpos linguísticos da Linguateca: apresentação". *TaLC at TaLC: Teaching and Linguateca's (Portuguese language) Corpora* (ISLA, Lisboa, 3 de Julho de 2008).

Diana Santos & Elisabete Ranchhod. "Ambientes de processamento de corpora em português: Comparação entre dois sistemas". In Irene Rodrigues & Paulo Quaresma (eds.), *Actas do IV Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR'99)* (Évora, 20-21 de Setembro de 1999), pp. 257-268.

Diana Santos. "Breves explorações num mar de língua". *Ilha do Desterro* 52 (2007), pp. 127-150.

ENSINADOR - <http://www.linguateca.pt/Ensinador>

Alberto Simões & Diana Santos. "Ensinador: corpus-based Portuguese grammar exercises". *Procesamiento del Lenguaje Natural* **47** (2011), pp. 301-309.

PAPEL - <http://www.linguateca.pt/PAPEL>

Hugo Gonçalo Oliveira, Diana Santos & Paulo Gomes. "Extracção de relações semânticas entre palavras a partir de um dicionário: o PAPEL e sua avaliação". *Linguamática* **2.1** (2010), pp. 77-93.

VARRA - <http://www.linguateca.pt/ACDC/>

Cláudia Freitas, Diana Santos, Hugo Gonçalo Oliveira & Violeta Quental. "VARRA: Validação, Avaliação e Revisão de Relações semânticas no AC/DC". In *Atas do ELC2010 2011*.

Agradecimentos

