

Perguntas já respondidas sobre o AC/DC: desde como começar até o uso complexo de funcionalidades poderosas

Cláudia Freitas, Diana Santos e Alice Gonçalves

07 de Novembro de 2011 – versão 2.1

Este documento foi redigido por Cláudia Freitas, maioritariamente a partir das respostas dadas ao longo dos anos por Diana Santos, e testado e verificado por Alice Gonçalves.

Sumário

PARTE I – VISÃO GERAL DO AC/DC

1. Algumas palavras iniciais.....	3
2. Por onde começar? Exemplos comentados	4
2.1. Mais exemplos de busca	11
3. Perguntas gerais sobre expressões de busca	12
4. Caracteres especiais	14
4.1. Exemplos de busca (comentados) com caracteres especiais	14
5. Visualização e apresentação dos resultados.....	16
5.1. Número de respostas	16
5.2. Contexto das ocorrências (aumentar ou diminuir)	17
5.3. Visualização de etiquetas nas concordâncias	18
5.4. Seleção de resultados por corpo, variante etc	19
5.5. Negritos	21
6. Como citar os corpos do AC/DC	21

PARTE II – PALAVRA

7. Buscas que envolvem a formação de palavras.....	23
8. Expressões multivocabulares/multipalavras; locuções	23
9. Tempos verbais.....	23
10. Pronomes	25
11. Listas de palavras e frequências.....	26

PARTE III – FRASES E/OU SENTENÇAS

12. Concordância verbal e nominal.....	27
13. Verbos com pronomes clíticos	27

14. Ordem de palavras.....	28
15. Argumentos.....	30
16. Orações com verbos no gerúndio.....	32
17. Orações adjetivas/ relativas	33
18. Orações substantivas / Complemento Oracional (e buscas com subjuntivo, infinitivo, etc).....	34
19. Orações adverbiais.....	37

PARTE IV – NÃO ENCONTRO O QUE PROCURO

20. Quando não sei em que seção das PJR procurar	38
21. Quando faço uma busca e obtenho zero resultados	38

PARTE I – VISÃO GERAL DO AC/DC

1. Algumas palavras iniciais...

O que é um corpo (ou corpus)? O AC/DC é um corpo?

O corpo, numa definição muito simples, é uma coleção de documentos. Os corpos são coleções bem estruturadas e com uma dimensão considerável, uma vez que são criados para a investigação em áreas como o processamento (computacional) de linguagem natural, ou extração de informação, ou, ainda, descrição da língua. Muitos, por vezes, podem ter anotações – informações linguísticas – dentro dos documentos, mediante a função que desempenha o próprio corpo.

O AC/DC é uma interface que permite realizar buscas linguísticas em corpos anotados. Isto é, todos os corpos do AC/DC já foram pré-processados automaticamente pelo analisador sintático PALAVRAS (Bick, 2000) e contêm, pelo menos, informação morfossintática associada às palavras/constituintes. Informação detalhada da anotação contida em cada corpo pode ser encontrada na página do próprio corpo.

Assim, o AC/DC é como um “Google linguístico”, em que é possível buscar ocorrências de verbos em determinados tempos ou pessoas, de sintagmas nominais com determinadas características, de orações relativas em que o pronome relativo exerce uma determinada função sintática, da ocorrência de adjetivos com determinados sufixos, o contexto de ocorrência de determinados advérbios, dentre uma série de outras possibilidades.

O AC/DC (como aliás todos os corpos que a Linguateca põe acessíveis) podem ser ótimas ferramentas para o estudo linguístico, mas os utilizadores têm de os saber utilizar, e compreender o que estão a perguntar, e a quê (ou seja, o tipo de texto e a quantidade de texto que está em causa). Além disso, relembramos que a anotação é automática (donde tem de ser passada pelo crivo do pesquisador, que não pode aceitá-la sem crítica e/ou confirmação).

Qual a diferença entre um corpus anotado e um corpus não anotado?

Um corpo anotado é aquele que possui alguma informação linguística a ele associado. Assim, um corpo pode ser anotado com relação a classes gramaticais, funções sintáticas, sintagmas nominais, nomes próprios (a que chamamos “entidades mencionadas”), relações semânticas, estrutura argumental, e o que mais se desejar. No AC/DC, a informação associada a cada corpo está disponível na página do próprio corpo.

Um corpo não anotado, por sua vez, é aquele em que não há qualquer informação linguística associada ao texto. É um texto “cru”.

O que é Linguística Computacional?

Sugerimos a leitura de do artigo abaixo, disponível na página de publicações da Linguateca:

Santos, Diana. "Introdução ao processamento de linguagem natural através das aplicações", in Elisabete Ranchhod (ed.), Tratamento das Línguas por Computador. Uma introdução à linguística computacional e suas aplicações, Lisboa: Caminho, pp.229-259.

Com que modelo teórico o AC/DC trabalha?

Em geral, o “problema” do AC/DC e/ou de qualquer serviço que alguém venha a dar sobre corpos anotados, é que 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.

Por isso, a maneira geral que alguém pode ter, no AC/DC ou noutra serviço qualquer, de ver se há suficiente semelhança entre as teorias gramaticais usadas na anotação e as necessidades de um dado usuário é experimentar ver como é que o sistema analisa e/ou contrasta casos que acha que devem (ou não) ser analisados da mesma forma (ou diferente). Peça a distribuição dos casos de cada atributo que marcamos (POS, TEMCAGR, FUNC, etc) e veja -- nos contextos que lhe interessam -- se permitem distinguir o que pretende. Esta é a nossa sugestão geral (e é o que fazemos também para responder às perguntas ou para realizar nossos próprios estudos)

O funcionamento de todos os corpos disponibilizados no AC/DC é o mesmo?

Sim! O funcionamento é uniforme em todos os corpos. O que pode variar ligeiramente são os marcadores estruturais e atributos posicionais disponíveis em cada um deles. A informação específica de cada um dos corpos está disponível nas páginas dedicadas a cada um deles, e é possível também consultar <http://www.linguateca.pt/acesso/contabilizacao.php>

Como escolher um corpo para a minha pesquisa? Posso começar com qualquer um?

Aconselhamos a usar primeiro, por exemplo, um corpo pequeno para se familiarizar com a sintaxe da procura e o tipo de resultados que podem ou não interessar-lhe, e apenas depois usar os maiores, que, por serem muito grande, levam muito mais tempo a produzir os resultados. No entanto, é importante lembrar que alguns fenômenos podem não ser observados em corpos muito pequenos.

O que são as unidades nos corpos do AC/DC?

Unidades são, além das palavras, os sinais de pontuação e outros casos como numerais e separadores, que se encontram nos textos dos corpos, mas que não correspondem a verdadeiras palavras. Essa contagem/discriminação é feita pelo nosso atomizador.

2. POR ONDE COMEÇAR? Exemplos comentados

Basta digitar uma palavra qualquer?

Para fazer pesquisa no AC/DC não é preciso escolher uma palavra. Pode escolher um tempo verbal, uma função sintática, um grupo de lemas, ou uma combinação de tudo isso... A seguir listamos alguns exemplos simples, comentados e diversificados, para que se tenha uma visão geral das possibilidades do AC/DC.

a) Como encontrar a forma casa do verbo CASAR?

Procura: [word="casa" & pos="V.*"]

Resultado: Concordância

Comentário: Com a informação pos="V.*" garantimos que só aparecerão nos resultados as ocorrências em que “casa” é verbo. (“pos” faz referência à classe gramatical, e significa “Part of speech”. Ou seja, queremos uma palavra cuja classe gramatical é um verbo. Para saber os diferentes tipos de pos (ou, em outras palavras, as diferentes siglas para as classes gramaticais), consulte <http://www.linguateca.pt/acesso/anotacao.html>)

b) Como obter as ocorrências de “mente” relativas ao verbo “mentir”, não ao substantivo “mente”?

Para os casos em que “mente” é um verbo:

Procura: [word="mente" & pos="V.*"]

Para os casos em que “mente” é um substantivo:

[word="mente" & pos="N"]

c) Como encontrar as ocorrências do verbo DESEJAR?

Procura: [lema="desejar"]

Comentário: Quando indicamos lema, dizemos que estamos interessados em todas as ocorrências do verbo, independente de tempo e pessoa.

Para ver as frases com “desejar”, faça uma busca por Concordância.

Para ver quais os tempos verbais usados com “desejar”, faça a busca por “Distribuição do tempo verbal e/ou do caso pronominal”

Para ver as ocorrências de “desejar” em termos de pessoa verbal, faça a busca por “Distribuição de pessoa e/ou número”

Para incluir mais de um elemento na mesma procura, por exemplo construções com o verbo desejar ou almejar, basta usar o sinal |.

Procura: [lema="desejar|almejar"]

d) Qual a distribuição da categoria gramatical da forma CARA? (ou: “cara” é mais frequente como substantivo ou adjetivo?)

Procura: "cara"

Resultado: Distribuição da categoria gramatical (POS)

Comentário: Como estamos interessados apenas na forma cara, e não em “caro/caras/caros” (no caso em que **car**a é adjetivo), não fazemos o pedido por lema, mas apenas com a palavra entre aspas. E notamos que dá exatamente no mesmo colocar a palavra entre aspas ou usar a expressão [word="cara"]

- e) Como encontrar adjetivos seguidos de nomes no masculino plural?** (como essa busca retorna muitos resultados, sugerimos, como teste, evitar corpos muito grandes, como o CHAVE ou NILC/São Carlos e Floresta):

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

Comentário: Cada informação entre [] indica uma unidade somente. Assim, a primeira expressão entre [] corresponde ao primeiro elemento da busca (adjetivos & do gênero masculino & no plural). A segunda expressão entre [] corresponde ao segundo elemento da busca (substantivo no masculino e no plural).

Para ver as frases em que aparecem “adjetivos seguidos de nomes no masculino plural”, escolha Concordância.

Para ver apenas os lemas dos adjetivos que são “seguidos de nomes no masculino plural”, escolha “distribuição dos lemas”.

Para ver apenas os lemas dos substantivos aparecem depois dos adjetivos (o segundo elemento da busca), acrescente @ imediatamente antes da expressão referente ao segundo elemento da busca:

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

Resultado: Distribuição dos lemas.

- f) Como encontrar adjetivos terminados em -udo?**

Procura: [pos="ADJ.*" & lema=".*udo"]

Resultado: Distribuição dos lema

Comentário: O uso de .* significa “qualquer coisa” (ver secção “carateres especiais”).

- g) -eiro é um sufixo mais produtivo na formação de adjetivos ou substantivos?**

Procura:[lema=".*eiro"]

Resultado: Distribuição da categoria gramatical (PoS)

Comentário: Diferentemente do exemplo anterior, agora não sabemos a classe gramatical (pos) das palavras que procuramos – e por isso não há o atributo “pos” na expressão de busca. Aliás, a classe é justamente o que queremos saber, e, por isso, pedimos o resultado em termos de Distribuição da categoria gramatical.

- h) Como encontrar frases com verbos da 3ª conjugação no gerúndio? E com verbos da 3ª conjugação no gerúndio seguidos de artigo singular seguido de substantivo?**

Para encontrar os verbos da 3ª conjugação no gerúndio:

Procura: [temcagr="GER" & word=".*indo"]

Resultado: Concordância

Comentário: Na expressão de busca estão os nomes dos atributos usados na criação dos corpos:

temcagr significa "tempo, caso e grau",

GER é um dos vários valores para tempo,

word significa a forma.

Os nomes dos atributos usados encontram-se nas páginas de anotação (<http://www.linguateca.pt/acesso/anotacao.html>), assim como os valores possíveis.

Para encontrar os verbos da 3ª conjugação no gerúndio seguidos de artigo singular seguido de substantivo:

Procura: [temcagr="GER" & word=".*indo"] [pos="DET.*" & pessnum="S"] [pos="N"]

No exemplo, cada informação dentro de [] indica que estamos à procura de uma unidade somente, que tem que obedecer às duas condições sobre os atributos (temcagr e word) que especificamos. (& significa conjunção).

i) Como encontrar expressões do tipo "fazer + V" como em "fazer aquecer"?

Como a categoria gramatical do verbo não é só "V", mas sim "V qualquer coisa"..., a expressão deve ser

Procura: [lema="fazer"] [pos="V.*"]

(e deve, além disso, pedir que o verbo seguinte esteja no infinitivo, ou, ainda melhor, que tenha a função IMV ("infinitive main verb")):

Procura: [lema="fazer"] [pos="V.*" & func="IMV.*"]

j) Que adjetivos modificam uma determinada palavra? (por exemplo, a palavra *pedra*?)

Aqui temos duas hipóteses: considerar que "palavra" se refere ao paradigma inteiro (pedra, pedras, pedrinha, pedrinhas, etc.) ou apenas à forma singular pedra.

Escolhendo a primeira hipótese, ainda temos a questão da modificação posterior ou anterior. Para modificação posterior:

Procura: [lema="pedra"] @[pos="ADJ.*"]

Resultado: distribuição de lemas

Para modificação anterior:

Procura: [pos="ADJ.*"] [lema="pedra"]

Resultado: distribuição de lemas

Caso tenha interesse em verificar qual a distribuição de "pedra" em termos do número, e ver se isto muda quando as pedras estão sozinhas ou qualificadas:

Procura: [lema="pedra"]

Resultado: distribuição de pessoa ou número

k) Como pesquisar os nomes que antecedem uma determinada palavra?

Para procurar os nomes que antecedem uma determinada palavra, por exemplo a palavra "ontem", a expressão de busca é

Procura: [pos="N"] "ontem"

Neste caso, não é preciso usar o caractere @ porque o que se quer é a distribuição da primeira palavra da expressão. Em Resultados, pode escolher Distribuição dos lemas, ou Distribuição das formas, ou Concordância...

l) Como encontrar um verbo seguido de um advérbio de negação, (por exemplo "decidiu não" "resolvi nunca")?

Procura: [pos="V"] [lema="nada|não|ninguém|nunca|nenhum|nem"]

A primeira parte indica que buscamos um verbo (pos significa Part of Speech). Caso queira limitar a um determinado tempo verbal (por exemplo, para evitar os participios), deve usar o campo "temcagr" (pode encontrar exemplos detalhados na página de anotação - <http://www.linguateca.pt/acesso/ anotacao.html>). Por exemplo, para só encontrar verbos no perfeito:

[temcagr="PS.*"] [lema="nada|não|ninguém|nunca|nenhum|nem"]

Os advérbios de negação não estão explicitamente marcados, pelo que a única maneira é listá-los separados por uma barra vertical. É melhor usar o campo lema porque algumas palavras (como "nenhum") podem ser flexionadas.

m) Qual a construção mais comum com o verbo dizer?

Aqui o importante é definir "construção". Admitindo que estamos apenas interessados em três "construções", descritas informalmente por "dizer que", "dizer N" e "dizer a alguém".

A primeira ("dizer que") é

Procura: [lema="dizer"] "que"

ou [lema="dizer(\+.*)*"] "que"

A segunda ("dizer N"), é

Procura: [lema="dizer"][pos="N.*"]

ou [lema="dizer(\+.*)*"][pos="N.*"]

A terceira ("dizer a alguém"), é

Procura: [lema="dizer"] [word="a" & pos="PRP"] [pos="N.*"]

Para juntar os três e obter a distribuição num só passo, poderíamos escrever (é tudo uma linha só, e os parênteses fazem parte da expressão):

Procura: (([lema="dizer(\+.*)*"] @ "que") | ([lema="dizer(\+.*)*"] @[pos="N.*"] | ([lema="dizer"] @[lema="a(\+.*)*" & pos="PRP(\+.+)*"] [pos="N.*"]

E pedir a Distribuição por categoria gramatical (pos), porque os três casos correspondem a diferentes categorias.

n) Que verbos são mais usados em orações relativas?

Em primeiro lugar, cabe identificar orações relativas: o que não é, naturalmente, trivial. Várias sugestões de forma a ir devagarinho:

Em primeiro lugar, para procurar de que forma é que o PALAVRAS marca pronome relativo na palavra **que**, pode pedir-se a distribuição por pos (ou por função sintática). (Dada a alta ocorrência do **que**, não aconselhamos essa busca em corpos grandes como o CHAVE, NILC/SãoCarlos ou Floresta)

Procura: [word="que"]

Resultado: Distribuição da categoria gramatical (PoS)

E depois ver se alguns dos nomes são imediatamente claros, ou se correspondem àquilo que imaginamos. No caso de termos pedido a distribuição em termos de categoria gramatical, poderíamos procurar KS, e SPEC_rel (as duas categorias mais frequentes para que), e simplesmente pedir esses casos:

[word="que" & pos="KS"]

Resultado: Concordância

Isto mostra que KS indica “que” como conjunção integrante. A expressão

[word="que" & pos="SPEC_rel"]

Resultado: Concordância

mostra que SPEC_rel era ou parecia ser o que queríamos. Podemos ver se há mais palavras marcadas com esta categoria

Procura: [pos="SPEC_rel"]

Pedido: Distribuição de lemas

Com isso, veríamos que os lemas de SPEC_rel seriam, inesperadamente, apenas **que** e **quem**. Admitindo que também gostaríamos de encontrar orações com **cujo** e com **qual** (e variações), teríamos de procurar as categorias destas palavras. Outra maneira de fazer isto (além de pedir a distribuição por pos, como fizemos para o **que**) é pedir concordâncias e assinalar a opção "mostrar categoria gramatical", o que pode ser feito com o comando “show +” antes da expressão de busca:

Procura: show +pos; "qual|cujo|quanto"

Resultado: Concordância

(e se quiséssemos ver a função sintática do qual|cujo|quanto, a expressão seria

Procura: show +func; "qual|cujo|quanto")

Daí veríamos que também existe a categoria (pos) DET_rel, e que possivelmente há mais categorias com o sufixo "rel". Fazendo

Procura: [pos="*.rel"]

Resultado: Distribuição da categoria gramatical (Pos)

e pedindo a distribuição por categoria gramatical e por lema, vemos que existem SPEC_rel, ADV_rel e DET_rel, e que ADV_rel, por exemplo, inclui **como** e **quando**...

Assumindo que todos estes nos interessam, basta procurar o verbo mais próximo, e depois simplesmente fazer uma distribuição dos lemas.

Procura: [pos=".*rel"] [pos!="V.*"]*@[pos="V.*"]

Resultado: Distribuição dos lemas

Com essa expressão, teremos os verbos mais usados em orações relativas. Para possibilitar o aparecimento de qualquer elemento entre o pronome relativo [pos=".*rel"] e o verbo (por exemplo, “uma realidade **que já se impunha**”), mas ao mesmo tempo garantir que esse elemento **não** seja um verbo, usamos [pos!="V.*"]*, que significa qualquer pos diferente (!= significa “diferente”, veja a seção “caracteres especiais”) de V.

l) Com que verbo a palavra **NÃO** é mais frequente?

Admitindo que a palavra **não** antecede sempre o verbo, basta pedir

Procura: "não" @[pos="V.*"]

Resultado: Distribuição dos lemas

Ou, então, aceitar também pronomes clíticos:

Procura: "não" [pos="PERS"]* @[pos="V.*"]

Resultado: Distribuição dos lemas

m) Quando é que a palavra **SEMPRE** é usada? Por exemplo, qual o tempo verbal com que mais co-ocorre?

Basta pedir a distribuição do tempo verbal e/ou do caso pronominal:

Procura: [pos="V.*"] [pos!="V.*"]* "sempre"

Resultado: Distribuição do tempo verbal e/ou do caso pronominal

n) Como saber que preposições estão imediatamente antes e imediatamente depois de "apenas" e de "só"?

Para procurar as preposições, basta fazer

Procura: [pos="PRP"] "só"

e pedir "Distribuição de lema" - se quiser apenas saber qual a preposição que aparece antes de “só” - , ou pedir "Concordância" - se quiser ver o contexto em que a construção aparece.

Para encontrar as preposições que aparecem depois de “só”, é preciso juntar a marca @ para indicar que não é no primeiro elemento que se está interessado...

Procura: "só" @[pos="PRP"]

E, igualmente, pedir "Distribuição de lema" - se quiser apenas saber qual a preposição que aparece depois de "só" - , ou pedir "Concordância" - se quiser ver o contexto em que a construção aparece.

o) Quais os substantivos que aparecem antes de "só"?

Procura: "só" @[pos="N"]

Distribuição de lemas.

Se quiser ver as frases completas, selecione "Concordância".

2.1. MAIS EXEMPLOS DE BUSCA

p) Como frases com verbos no condicional?

Basta incluir na busca [temcagr="COND.*"]

q) Como encontrar frases interrogativas iniciadas por "Como" ?

Procura: [word="Como"] []* "\?" within s

r) Quais palavras ocorrem como complemento do verbo "dar" com a função de objeto direto?

Procura: [lema="dar"] @[func="<ACC"]

Resultado: Distribuição dos lemas

s) Em quais orações o pronome relativo tem a função de objeto direto?

Procura: [pos="SPEC_rel" & func="ACC.*"]

Resultado: Concordância

t) Que palavras foram identificadas como verbos pelo analisador através de heurísticas morfológicas?

Procura: [pos="V.*" & deriv="DER.*"]

u) Como encontrar participios passados de gênero feminino?

Procura: [temcagr="PCP" & gen="F"]

v) Como encontrar formas do imperfeito do conjuntivo/subjuntivo?

Procura: [temcagr="IMPF_SUBJ"]

w) Como encontrar os objetos do verbo MATAR no máximo quatro palavras distantes?

Procura: [lema="matar"] [][0,3] [func="<ACC"]

3. PERGUNTAS GERAIS SOBRE EXPRESSÕES DE BUSCA & FUNCIONAMENTO

Como descobrir / entender as siglas linguísticas para a pesquisa?

Toda a informação está em

<http://www.linguateca.pt/acesso/ anotacao.html>

É possível incluir mais de um elemento na mesma procura? Por exemplo, construções com o verbo MANDAR ou DEIXAR.

Sim. Para isso basta usar o sinal |. Para uma pesquisa com as ocorrências dos verbos mandar ou deixar:

Procura: [lema="mandar|deixar"]

Se fizer uma pesquisa por [word="[Qq]ue"] obtenho as ocorrências de "que" maiúsculo e minúsculo?

Sim, mas sugerimos usar [lema="que"], porque assim são consideradas as formas (word) "Que" e "que" e também "QUE".

Como salvar/guardar os resultados da pesquisa?

Uma das possibilidades é abrir a janela do AC/DC em uma janela separada do browser (por exemplo, selecionando o corpo de pesquisa desejado escolhendo a opção abrir em uma nova janela/em uma nova aba), fazer a pesquisa e em seguida Salvar como/Save as o resultado.

Os verbos estão anotados (marcados) quanto a sua transitividade?

Os verbos não estão classificados nos corpos. O que está anotado é a sua ocorrência. Qualquer verbo (quase) pode ocorrer em contexto transitivo ou intransitivo. A única coisa que está marcada são os argumentos dos verbos nos contextos transitivos.

Como excluir ou evitar a ocorrência de um dado elemento na pesquisa?

O sinal de negação é !=

(veja a seção "Caracteres especiais")

Nos corpos que contêm textos em português de Portugal e do Brasil, como é o exemplo de CHAVE, há alguma maneira de verificar em cada uma das ocorrências se se trata do PT ou do BR?

No caso do CHAVE, basta ver os códigos dos documentos: os que começam por "F" são da Folha de São Paulo, e, portanto, português do Brasil; os que começam por "P" são do jornal Público, e, portanto, português de Portugal.

De qualquer forma, é possível restringir a busca para que apareçam exemplos apenas de uma das variantes. Basta acrescentar informação referente à variante na expressão de busca. Se estiver buscando ocorrências do verbo “calhar” apenas em Portugal, a expressão deve ser

```
[lema="calhar" & variante= "PT"]
```

Se estiver buscando pelo verbo “calhar” apenas no Brasil:

```
[lema="calhar" & variante="BR"]
```

E, se quiser simplesmente ver como se dá a distribuição do verbo calhar com relação às duas variantes:

```
[lema="calhar"]
```

Resultado: Distribuição por variante do português.

Se não houver a opção Distribuição por variante do português nos resultados, é porque o corpo selecionado contém apenas uma das variantes. De qualquer maneira, sugerimos visitar a parte de “Visualização dos resultados” para mais explicações.

Por que às vezes a informação de pos contém .* ? Qual a diferença entre uma busca por, no caso da preposição, [pos="PRP"] e [pos="PRP.*"]?

“.*” indica a possibilidade de aparecer qualquer coisa após o PRP. Assim

[pos="PRP"] retorna apenas preposições, enquanto [pos="PRP.*"] também retorna todos os casos de contração da preposição com artigo ("da") ou com pronome pessoal ("conosco").

Como indico, na minha busca, que entre um elemento e outro podem aparecer “várias palavras”, que eu não sei quais são?

Utilize a expressão []*, que representa "zero ou várias palavras". Para uma palavra apenas, use [].

Posso sempre combinar qualquer expressão de procura com qualquer tipo de resultado (concordância, distribuição por POS, FUNC, etc...)?

Claro! Uma coisa é a seleção das entidades em que se está interessado, outra é o tipo de resultados que deseja (concordância, distribuição) sobre o que selecionou.

De quantas palavras é a distância máxima possível entre duas formas na procura?

O tamanho máximo de palavras encontradas com a expressão `[]*`, que representa "várias palavras", é de 100 unidades.

4. CARACTERES ESPECIAIS

`@` → o `@` só deve ser usado quando o resultado da busca envolve “distribuição de...” e não “Concordância”. Ele indica de qual item se quer fazer a distribuição. Quando não se inclui o `@` na expressão, o resultado da distribuição será sempre o da primeira palavra. Assim, como, por padrão, o elemento considerado é sempre o primeiro, não é preciso usar `@` quando se deseja a distribuição do primeiro elemento na busca. Por exemplo...

`+` → o `+` é um caractere especial (significa "uma ou mais ocorrências"). Para fazer uma pesquisa específica desse caractere é necessário colocar antes um caractere de escape `\"` e é isso o que se faz, por exemplo, quando se procura por pronomes clíticos.

`.` → o `.` significa qualquer caractere.

`*` → Indica opcionalidade. O elemento marcado com `*` pode ou não ser considerado, isto é, indica 0 ou mais.. Por exemplo, na expressão

```
[pos="V.*"]*
```

, a parte `[pos="V.*"]` significa uma palavra com categoria gramatical V, mas está modificada pelo operador `*`, que indica 0 ou mais. Se desejar só uma ocorrência de V, é preciso tirar o asterisco; se desejar uma ou mais ocorrências, basta substituir o `*` por `+`.

`[]` → essa expressão representa uma palavra apenas, qualquer que seja.

`[]*` → essa expressão representa "ZERO ou mais palavras". É muito útil para pesquisas em que não se sabe ou não se precisa saber, o que pode aparecer entre dois elementos da construção. É como um “coringa” que equivale a nenhuma ou mais de uma palavra. O limite máximo de palavras capturadas por `[]*` é de 100 unidades. Além disso, pode utilizar o curinga de maneira negativa (ver *negação* abaixo), ou seja: pode aparecer qualquer coisa, exceto um verbo: `[pos!="V.*"]*`

É importante notar que, muitas vezes, ao usar o `[]*`, acabamos por aceitar elementos demais na busca, que frequentemente pode ultrapassar os limites de uma única frase. Nesse caso, se quiser garantir que a sua expressão de busca se limite a uma frase, deve usar *within s* (ver abaixo) ao final da expressão.

`[].*` → essa expressão representa "UMA ou mais palavras"

within s → delimita a busca a uma única sentença, e é útil principalmente se a expressão de busca contém `[]*`, que indica “uma ou mais palavras”. Deve ser usado no final da expressão de busca. No exemplo abaixo, a expressão indica perguntas iniciadas por “como”:

```
[word="Como"] []* "\?" within s
```

<s> e </s> → marcam o início e o fim de uma sentença, respectivamente. Assim, a expressão delimita a busca a uma sentença inteira. Pode ser usada quando se quiser que toda a sentença apareça em negrito, e não apenas os elementos indicados, e deve ser usada em conjunto com o caracter [], que indica uma ou mais palavras. Por exemplo, com a expressão abaixo, é possível obter todos os adjetivos terminados em –ento, e somente eles aparecem em negrito se o resultado for concordância.

Procura: [pos="ADJ.*" & lema=".*ento"]

Resultado: Concordância

Já com a expressão abaixo, a busca é igualmente por adjetivos terminados em –ento, mas dessa vez a frase inteira aparecerá em negrito.

Procura: <s>[]* [pos="ADJ.*" & lema=".*ento"] []*</s>

Resultado: Concordância

!= → representa a negação. Para uma busca por verbos que não incluam “ser, estar, ter, ir, parecer”:

Procura: [pos="V" & lema!="ser|estar|ter|ir|parecer"]

4.1 EXEMPLOS DE BUSCA COM CARACTERES ESPECIAIS

Como obter as frases em que os verbos de movimento como *ir, vir, entrar, subir, partir, chegar, caminhar, correr, nadar, cambaleiar, tropeçar* etc, nos diversos tempos e modos, co ocorrem com as preposições *a, para, e até* ?

A forma mais simples seria pedir

Procura: [lema="ir|vir|entrar|subir|partir|chegar|caminhar|correr|nadar|cambaleiar|tropeçar"] [lema="a|para|até" & pos="PRP"]

Em que os verbos listados aparecerão em todos os tempos e modos, e serão seguidos **imediatamente** pelas preposições *a, para, e até*

No entanto, como é muito extensa e com verbos muito frequentes, desaconselhamos essa forma de procura. Sugerimos então que se separe cada verbo em um busca diferente. Para o verbo “ir”, portanto, a expressão seria simplesmente

Procura: [lema="ir"] [lema="em|a|para|até" & pos="PRP"]

É possível, ainda, realizar uma busca para cada tipo de preposição. Assim, para procurar apenas o verbo “ir” seguido imediatamente da preposição “para”:

Procura: [lema="ir"] [lema="para" & pos="PRP"]

É possível, também, desejar que a preposição **não** apareça imediatamente após o verbo “ir”. A expressão abaixo significa: encontre quaisquer formas do verbo IR seguidas de um número zero ou mais de palavras que não sejam elas próprias verbos (isto é, cuja POS **não** comece por V), nem preposições (cuja POS **não** comece por P), o que também elimina Pontuação.)

Procura: [lema="ir"] [pos!="[VP].*"]* [lema="para" & pos="PRP"]

Se quiser saber quais são as palavras que aparecem entre o verbo IR e a preposição PARA, basta colocar @ imediatamente antes do elemento da expressão correspondente, e pedir a distribuição dos lemas:

Procura: [lema="ir"] @[pos!="[VP].*"]* [lema="para" & pos="PRP"]

Distribuição dos lemas.

Como garantir que não existe um verbo no infinitivo entre uma determinada palavra e o fim da frase? Por exemplo, não desejo verbos no infinito depois da ocorrência das preposições “com”, “de” ou “a”:

Procura: [lema="com|a|de" & pos="PRP"]? [pos!="INF"]+ </s>

Como buscar palavras não contíguas em um mesmo período?

Para procurar, por exemplo, "com" seguido de "só" com qualquer número de palavras no meio, basta fazer

Procura: "com" []* "só"

No entanto, talvez seja interessante criar alguns filtros ou **restrições**. Por exemplo, para **restringir o número de palavras no meio** (por exemplo entre 1 e 5):

Procura: "com" []{1,5} "só"

Para restringir o número de palavras para menos de 6

Procura:"com" []{0,6} "só"

Para restringir a busca a palavras classificadas de um certo modo, e, por exemplo, **não aceitar verbos** entre as duas palavras:

Procura: "com" [pos != "V.*"]* "só"

5. VISUALIZAÇÃO E APRESENTAÇÃO DOS RESULTADOS

5.1. NÚMERO DE RESPOSTAS

Como limitar o número de respostas de uma busca?

Para limitar o número de ocorrências de uma busca basta digitar, ao lado da expressão de procura, a expressão “cut” imediatamente seguida pela quantidade de ocorrências desejada .

Por exemplo, para limitar a 500 as ocorrências de frases com verbos no condicional:

Procura: [temcagr="COND.*"] cut 500

Como limitar o número de respostas de uma busca e fazer com que os resultados sejam aleatórios? (Ou: Como fazer com que os resultados da busca sejam um subconjunto aleatório?)

Na interface de busca, basta usar a opção que está no fim das opções de pesquisa (“Amostra aleatória de xxxx linhas”)

5.2. CONTEXTO DAS CONCORDÂNCIAS

Como, nas concordâncias, obter um contexto maior que uma frase?

Embora, na grande maioria das vezes, uma frase seja suficiente para ilustrar/compreender o seu contexto, é possível que, em alguns raros casos, uma frase não seja suficiente, e se queira alargar o contexto.

No AC/DC, é possível pedir um contexto maior, por exemplo duas frases – uma à esquerda da frase-alvo, e outra à direita. Basta escrever, na caixa de procura do AC/DC, essa indicação antes da procura. No exemplo abaixo, a procura é pela palavra "caixa":

```
set c 2s; "caixa";
```

(o “c” vem de *context*; o “s”, de *sentence*)

Se quiser aumentar apenas o contexto à esquerda, basta usar lc (“left context”):

```
set lc 2s; "caixa";
```

Se quiser aumentar apenas o contexto à direita, basta usar rc (“right context”):

```
set rc 2s
```

Atenção: dependendo do corpo utilizado na procura, pode haver restrições adicionais quanto a mostrar mais.

Além disso, é possível também definir o contexto em número de unidades (tokens), ou seja, palavras e/ou pontuação e também atributos estruturais (invisíveis, portanto):

```
set c 40; "mulher";
```

```
ou set lc 50; "mulher"
```

Contudo, não aconselhamos usar este tipo de contexto, porque, às vezes, não produz o que se pretende devido a algumas regras que introduzimos para evitar usos abusivos dos corpos, e que alteram o pedido inicial. Será preferível usar em termos de frases.

Sugerimos, contudo que estas opções sejam primeiro testadas em corpos menores (por exemplo o AmostRA-NILC) antes de interrogar o corpo total.

É possível encontrar o texto na íntegra de onde foi tirado o excerto para uma concordância em contexto?

Em princípio, a nossa interface não se destina a consultar textos completos. No entanto, disponibilizamos textos completos para 3 corpos: CHAVE, FLORESTA e CD HAREM.

No CHAVE, basta seguir a ligação para consultar o artigo de onde o texto provém.

Na FLORESTA, basta ir à página do projeto

<http://www.linguateca.pt/Floresta/>

ou ir diretamente a

<http://www.linguateca.pt/Floresta/levantamento.html>

Na CD HAREM, basta ir à página específica do HAREM:

<http://www.linguateca.pt/HAREM/>

No corpo CETEMPúblico, por sua vez, isso não é possível porque o corpo não contém textos completos; os textos foram divididos em extratos de dois parágrafos. O máximo que se pode obter é esse extrato de dois parágrafos completos, da seguinte maneira. Supondo que esteja interessado na palavra “porém”

```
set c 1 ext; "porém";
```

(a resposta pode ser demorada para palavras muito frequentes).

Para os outros corpos, é possível definir o contexto que se pretende pelo número de frases. Por exemplo, se estiver interessado na palavra “porém”, e quiser duas frases à esquerda e uma à direita, a expressão de busca é:

```
set lc 2 s; set rc 1 s; "porém";
```

Se quiser ver o contexto de uma única frase, é possível inserir uma sequência de palavras dessa frase, cada uma entre aspas, e sem omitir a pontuação: (Como o exemplo abaixo é bem específico, sugerimos usar o corpo Todos juntos)

```
set c 1 ext; "Agora " " " "porém" " " "tudo";
```

5.3. VISUALIZAÇÃO DE ETIQUETAS NAS CONCORDÂNCIAS

É possível visualizar, nas concordâncias, a informação de alguma etiqueta? Por exemplo, da etiqueta de pos? Ou da função sintática

Sim. Para ver a informação de pos, basta acrescentar, **ANTES** da expressão de busca,

```
show +pos ;
```

ou seja, para o exemplo, a expressão seria

Procura: **show +pos; [lema="qual"]**

Para ver as funções sintáticas associadas ao “qual” expressão, basta acrescentar, **ANTES** da expressão de busca,

show +func;

e a expressão de busca seria

Procura: **show +func; [lema="qual"]**

E o procedimento pode ser feito para qualquer tipo de atributo, isto é, é possível show +qualquer tipo de etiqueta;

É possível visualizar, nas concordâncias, mais de dois tipos de etiquetas ao mesmo tempo? Por exemplo, as etiquetas de classe ou função?

Sim!

Basta usar show + ...; "..."

Para ver «func» e «pos» de da palavra livro, basta usar

Basta fazer show +pos; show +func; "livro";

5.4. SELEÇÃO DE RESULTADOS POR CORPO, POR VARIANTES ETC.

Corpos jornalísticos:

Nos resultados da busca, como encontrar informação sobre o semestre e a seção do jornal de que os extratos foram tirados?

Para obter a distribuição por semestre ou por seção, basta selecionar essa opção nos "Resultados". Além disso, junto de cada ocorrência aparece algo que permite identificá-la. No corpo CHAVE, por exemplo, isso é dado por

F940101-008:

que significa Folha de 94, mês 01, dia 01, artigo 008

P951130-045:

que significa Público de 95, mês 10, dia 30, artigo 045

É possível fazer uma busca apenas nos textos jornalísticos do NILC São Carlos? Por exemplo, como encontrar as sentenças interrogativas iniciadas por "Como" nos textos da Folha de São Paulo?

Sim. Os valores relativos ao atributo classe, no corpo NILC São Carlos, estão indicados em <http://www.linguateca.pt/aceso/contabilizacao.php#saocarlos>

E, dali, sabemos que os textos jornalísticos podem ser

JO.* → jornalístico

JOCF → jornalístico só CETENFolha

JO → jornalístico sem CETENFolha

Para selecionar texto jornalístico na busca, basta fazer assim:

```
[word="Como" & classe="JO.*"] [* "\?" within s
```

Se quiser apenas na Folha de São Carlos de 1994 (que é o que está no NILC/São Carlos), e não nos vários jornais, a busca deve ser

```
[word="Como" & classe="JOCF"] [* "\?" within s
```

(E lembramos que essas buscas são apenas para o corpo NILC/São Carlos)

O que significam os códigos que aparecem antes de cada frase no corpo CHAVE?

F940101-008:

significa Folha de 94, mês 01, dia 01, artigo 008

P951130-045:

significa Público de 95, mês 10, dia 30, artigo 045

É possível, usando o corpo Todos, eliminar um ou mais corpos dos resultados?

Sim! Para saber o código de cada corpo, é possível consultar

<http://www.linguateca.pt/acesso/legenda.php>

que contém as legendas de todos os corpos do AC/DC, ou é possível, na própria busca, pedir, nos resultados, a Distribuição por corpo.

Por exemplo, se estiver interessado nas ocorrências do verbo “amarelar”:

Procura: [lema="amarelar"]

Resultado: Distribuição de corpo

Com isso, verá que os corpos são

CHA

VER

CPU

FLOR

NSC

CDIV

DIA

EBR

ENPC

Com essa informação, usando o atributo “corpo” e a negação, é possível filtrar os resultados por corpos.

Assim, caso queira eliminar os resultados do corpo Vercial na busca, consulte

<http://www.linguateca.pt/acesso/legenda.php>

e verá que o código para o Vercial é VER. Para eliminar da busca os resultados do Vercial, basta fazer

Procura: [lema="amarelar" & corpo!="VER"]

Nos corpos que contem textos das variantes de Portugal e do Brasil, é possível selecionar apenas ocorrências de uma variante?

Claro! Basta incluir, na expressão de busca, “variante="PT" ou “variante="BR". Se estiver interessado nas ocorrências do verbo “calhar”, apenas na variante de Portugal, basta fazer

Procura: [lema="calhar" & variante="PT"]

Se quiser, por outro lado, saber se o verbo “calhar” é mais usado no Brasil ou em Portugal, basta fazer

Procura: [lema="calhar"] e, no resultado, pedir Distribuição por variante do português.

5.5. NEGRITOS

Como indicar que desejo marcar a frase inteira em que o fenômeno procurado aparece? Por exemplo, procuro pela palavra “caixa”, e desejo selecionar (isto é, desejo que apareça em negrito) não apenas a palavra “caixa”, mas a frase inteira que a contém.

Basta incluir, no início e no fim da expressão de busca, <s>[*] e [*</s>, respectivamente. No exemplo da palavra caixa:

Procura: <s>[*] [lema="caixa"] [*</s>

[*] representa "várias palavras" e <s> e </s> marcam o início e o fim de uma sentença, respectivamente

6. COMO CITAR OS CORPOS DO PROJETO AC/DC?

Faça referência ao(s) corpo(s) do projeto AC/DC que utilizou e indique uma das seguintes referências gerais para o AC/DC: em português, Costa et al. (2009) e, em inglês, Santos (2011). (E, idealmente, também sugerimos referenciar o endereço <http://www.linguateca.pt/ACDC/>).

Por exemplo, se usou o corpo Avante e a referência é em português:

"...corpo AVANTE, do projeto AC/DC (Costa et al.,2009) em <http://www.linguateca.pt/ACDC/>.

As referências mencionadas são:

Santos, Diana. "Linguateca's infrastructure for Portuguese and how it allows the detailed study of language varieties". OSLa 3 (2), 2011, pp. 113-128. [J.B. Johannessen (ed.), Language variation Infrastructure]

Costa, Luís, 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).

Para mais informação, veja os vários artigos associados ao projeto AC/DC, consultando o catálogo de publicações da Linguateca com a marca acdc.

PARTE II – PALAVRA

7. BUSCAS QUE ENVOLVEM FORMAÇÃO DE PALAVRAS

Como encontrar ocorrências do sufixo -ão (ou de outro sufixo)?

Basta procurar ".*ão". No entanto, aconselhamos a restringir a procura para uma determinada categoria gramatical. Por exemplo, se quer adjetivos terminados em -ão:

Procura: [pos="ADJ" & lema=".*ão"]

Resultado: Concordância

Se quer substantivos terminados em -ão:

Procura: [pos="N" & lema=".*ão"]

Recomendamos essas buscas para procuras em contexto. Para encontrar apenas as palavras (sem contexto) terminadas em -ão, basta ir às listas de lemas que disponibilizamos para cada corpo ou pedir o resultado por Distribuição de lemas.

8. EXPRESSÕES MULTIVOCABULARES/MULTIPALAVRAS; LOCUÇÕES

Preposições: Como encontrar expressões complexas da forma Prep Nome Prep, como *em relação a; em termos de; em vez de; a favor de* etc?

Basta exigir que a locução preposicional esteja dentro de uma marca <mwe>. Ou seja, a expressão deve ser:

Procura: <mwe> [pos="PRP"] [pos="N"] [pos="PRP"] </mwe>

Como encontrar variações para a expressão "andar à deriva", com o verbo em todas as suas formas?

Procura: [lema="andar" & pos="V.*"] "à" [lema="deriva"]

ou

Procura: [lema="andar" & pos="V.*"][word="à"][lema="deriva"]

(Caso não encontre nenhuma ocorrência, sugerimos repetir a busca no corpo Todos juntos)

9. TEMPOS VERBAIS

Como encontrar verbos no Futuro do Conjuntivo/Subjuntivo?

Para encontrar as orações que contêm verbos no futuro do conjuntivo/subjuntivo:

Procura: [temcagr="FUT_SUBJ.*"]

Resultado: Concordância

Caso deseje ver quais verbos se encontram nesse tempo, e com que frequência, basta repetir a procura e pedir a opção "Distribuição dos lemas"

Como encontrar verbos no Pretérito Imperfeito e Pretérito Perfeito do Indicativo?

Procura: [temcagr="(PS|IMPF)_.*"]

No entanto, como essa busca retorna um número excessivo de ocorrências, sugerimos experimentar em um corpo bem pequeno ou restringir, por exemplo, às ocorrências da terceira conjugação:

Procura: [lema=".*ir" & temcagr="(PS|IMPF)_.*"]

Para encontrar apenas ocorrências do **Pretérito Perfeito**:

Procura: [temcagr="PS_.*"]

.. e a mesma coisa se aplica ao número excessivo de ocorrências.

Como encontrar apenas ocorrências do Imperfeito?

Para esta busca, caso não se deseje encontrar também ocorrências do conjuntivo/subjuntivo, é importante garantir que o modo é Indicativo:

Procura: [temcagr="IMPF_.*IND.*"]

Como saber se um determinado verbo é mais usado no perfeito ou no imperfeito?

Para saber a distribuição do verbo “gostar”, por exemplo, em termos de uso no perfeito ou no imperfeito:

Procura: [lema="gostar" & temcagr="(PS|IMPF)_.*"]

Distribuição do tempo verbal e/ou do caso pronominal

Como saber quais pessoas do verbo são mais empregues no imperfeito?

A expressão de procura é a mesma, mas deve selecionar

Distribuição de pessoa e/ou número

Por uma opção condicionada por questões técnicas, o verbo seguido de clíticos é considerado como uma unidade (cf. seção Clíticos).

Como encontrar as ocorrências de infinitivo não flexionado e de infinitivo flexionado, de qualquer verbo?

Procura: [temcagr="INF.*"]

Como encontrar frases com verbos no condicional?

Com a expressão

[temcagr="COND.*"]

é possível encontrar todos os verbos no condicional. Para que a busca retorne as frases que contém esses verbos, a expressão é

<s>[* [temcagr="COND.*"] [*</s>

Na qual <s> e </s> marcam o início e o fim de uma sentença, respectivamente, e [* representa “várias palavras”.

Como encontrar os verbos no condicional com clíticos?

[temcagr="COND.*\+.*"]

10. PRONOMES

Como encontrar as formas nominativas do pronome (eu, tu, eles, nós, vós eles/elas)?

[temcagr="NOM"]

Clíticos

Por questões técnicas, o verbo seguido de clíticos é considerado como uma unidade (um átomo, ou *token*). Os valores são concatenados depois do sinal +.

Assim, para obter frases do tipo "mando-o passear", a expressão deve ser

Procura: [lema="mandar(\+.*)*"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Para uma busca SÓ com os clíticos, é preciso usar:

Procura: [lema="mandar\+.*"] [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Se não quiser os verbos com clíticos, poderá fazer a exigência de que NÃO haja um "+" no atributo morfológico (ou no lema)...

Por exemplo, verbos iniciados por "des" no pretérito perfeito sem clíticos:

Procura: [lema="des[^+]+" & temcagr="PS_.*"]

Se quiser, por outro lado, saber qual a distribuição destes últimos, terá de recorrer a

Procura: [temcagr="PS_.*\+.*"]

(verbos no pretérito perfeito com enclíticos)

e seleccionar a distribuição dos lemas.

Como encontrar apenas o clítico “lhe/lhes”, tanto como enclítico como proclítico?

Para encontrar as ocorrências tendo em vista um determinado verbo, por exemplo, o verbo mandar:

Procura: [word=".*lhes*" & func=".*DAT>"] [lema="mandar"] | [lema="mandar\+.*" & word=".*lhes*"]

Se, por outro lado, não deseja especificar o verbo, e interessam todas as ocorrências de lhe/lhes, basta substituir, na expressão de procura, o lema mandar por .*

Procura: [word=".*lhes*" & func=".*DAT>"] [lema=".*"] | [lema=".*\+.*" & word=".*lhes*"]

Como encontrar os verbos no condicional com clíticos?

[temcagr="COND.*\+.*"]

O carácter "+" é um carácter especial (significando "uma ou mais ocorrências"). Para fazer uma pesquisa específica desse carácter é necessário colocar antes um carácter de escape "\\".

11. LISTAS DE PALAVRAS E FREQUÊNCIA

Existe alguma lista pronta com as palavras dos corpos organizadas por classe gramatical?

O AC/DC tem muitas listas de palavras, já divididas por categoria gramatical, prontas para serem utilizadas e específicas de cada corpo. Os dados, para cada corpo, estão em

<http://www.linguateca.pt/acesso/contabilizacao.php>

Como encontrar informações referentes à frequência das formas e lemas nos corpos?

As informações sobre a frequência estão disponíveis no Ordenador, que permite obter frequências de itens lexicais nos variados corpos :

<http://www.linguateca.pt/acesso/ordenador.php>

PARTE III – FRASES E/OU SENTENÇAS

12. CONCORDÂNCIA VERBAL E NOMINAL

Como encontrar expressões de quantificação que ocupam a posição de sujeito (e verificar a sua (eventual) concordância com o predicado)? Por exemplo, como encontrar frases do tipo “Um terço das trabalhadoras apoiou/apoiaram | foi/foram apoiadas/os...”

Procura: [func="*.SUBJ.*" & pos="NUM.*"] [lema="de(\+.*)*"]

Como encontrar um particípio que se refere a uma coordenação nominal anterior a ele? Por exemplo: “os critérios e as opções selecionadas”. A coordenação pode ser de NPs ou Ns, e o verbo não necessariamente precisa estar imediatamente após a coordenação: “os critérios e as opções foram selecionadas”. E como garantir que o primeiro conjunto nominal está no masculino, o segundo no feminino e o verbo está no feminino também?

Sugerimos usar a seguinte expressão, eventualmente com variações se quiser material entre os nomes e o particípio passado...

Procura: [pos="N"] [pos="KC.*"] [pos="N"] [temcagr="PCP"]

A expressão significa “nome seguido por conjunção seguido por nome seguido por particípio passado”. A expressão abaixo faz o mesmo, mas especifica o gênero, isto é, garante o primeiro conjunto nominal no masculino, o segundo no feminino, e o verbo também no feminino:

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

13. VERBOS COM PRONOMES CLÍTICOS

Como encontrar verbos com clíticos? (tornou-se; tornando-se; tornou-lhe)

O LEMA de um verbo com clíticos é o lema do verbo associado ao lema dos clíticos. Por exemplo, para encontrar as ocorrências de “tornar” + pronomes clíticos:

Procura: [lema="tornar\+.*"]

Se estiver interessado em apenas um tipo específico de pronome, em Resultados selecione a opção “Distribuição da categoria gramatical (PoS)” e veja como os casos foram anotados pelo analisador PALAVRAS. Em seguida, refaça a busca especificando, além do lema, a classe gramatical (pos) desejada.

Como encontrar clíticos antes do verbo principal (proclíticos), em construções do tipo “para não o/lhe mandar”?

Os proclíticos são procurados como palavras normais. Portanto, para incluir as contrações de dois pronomes pessoais, (marcados PERS+PERS), é preciso usar [pos="PERS(\+PERS)*"]

Para encontrar frases com a do exemplo, a expressão será:

```
Procura: [pos="PERS(\+PERS)*"] [pos!="V.*"]{0,3} [lema="mandar"]
```

Essa busca encontrará todos os pronomes pessoais, inclusive os sujeitos. Para encontrar apenas os proclíticos mesmo, a expressão deve ser:

```
Procura: [pos="PERS(\+PERS)*" & temcagr!="NOM"] [pos!="V.*"]{0,3} [lema="mandar"]
```

(ou seja, a expressão acima elimina os casos nominativos)

Se quiser só obter os que foram considerados pelo analisador automático como objetos (diretos ou indiretos) de *mandar*, então seleciona-os forçando o atributo "func"

```
Procura: [pos="PERS(\+PERS)*" & func="ACC>|DAT>|DAT>\+ACC>"] [pos!="V.*"]{0,3} [lema="mandar"]
```

Como encontrar verbos com clíticos na primeira pessoa, por exemplo “sinto-me forte”?

```
Procura: [lema=".*\+eu" & pos="V.*"]
```

Para ver quais são os verbos que aparecem nas construções, peça "Distribuição de lemas".

14. ORDEM E COMBINAÇÃO DE PALAVRAS

Como pesquisar o verbo que aparece depois de uma determinada palavra?

Para pesquisar um verbo que segue uma palavra, basta digitar, na caixa de procura, imaginando que a palavra é "caixa":

```
Procura: "caixa" @[pos="V.*"]
```

Se pedir, no Resultado, a “Distribuição dos lemas”, terá os lemas de todos os verbos que se seguem à palavra "caixa", por ordem de frequência. Se pedir “Concordância”, terá as formas de verbo que seguem caixa, assim como a palavra “caixa”, em negrito.

O caracter @ indica qual a parte da expressão de busca deve ser considerada na “Distribuição dos lemas”. Isto é importante quando não se trata do primeiro elemento da busca.

No entanto, a expressão de procura acima encontra apenas a caixa, e não encontra “caixas”, por exemplo. Para expandir a sua busca, use o atributo “lema”, e a expressão de busca é:

```
Procura: [lema="caixa"] @[pos="V.*"]
```

Experimente retirar o elemento “@” da expressão, e veja o que acontece quando pede “Distribuição dos lemas”.

Como pesquisar os nomes que antecedem uma determinada palavra?

Para procurar os nomes que antecedem uma determinada palavra, por exemplo a palavra "ontem", a expressão de busca é

Procura: [pos="N"] "ontem"

Neste caso, não é preciso usar o caracter @ porque quer a distribuição da primeira palavra da expressão. Em Resultados, pode escolher Distribuição dos lemas, ou Distribuição das formas, ou Concordância...

Como encontrar um verbo seguido de um advérbio de negação (por exemplo: "decidiu não" "resolvi nunca")?

Procura: [pos="V"] [lema="nada|não|ninguém|nunca|nenhum|nem"]

A primeira parte indica que pretendemos um verbo (pos significa Part of Speech). Caso queira limitar a um determinado tempo verbal (por exemplo, para evitar os participios), deve usar o campo "temcagr" (pode encontrar exemplos detalhados na página de anotação - <http://www.linguateca.pt/acesso/anotacao.html>). Por exemplo, para só encontrar verbos no perfeito:

Procura: [temcagr="PS.*"] [lema="nada|não|ninguém|nunca|nenhum|nem"]

Os advérbios de negação não estão explicitamente marcados, pelo que a única maneira é listá-los separados por uma barra vertical. É melhor usar o campo lema porque algumas palavras (como "nenhum") podem ser flexionadas.

Como encontrar combinações de “verbo + adjetivo”, como *falar alto*, *dançar rápido* etc?

Procura: [pos="V.*"] [pos=".*ADV.*"]* [word="alto|rápido|forte" & pos="ADV" & func="<ADVL"]

No entanto, essa expressão só irá retornar frases em que os adjetivos são “alto”, “rápido” ou “forte”. Isto acontece porque, infelizmente, o analisador sintático PALAVRAS não marca especialmente os advérbios homônimos como um adjetivo, mas limita-se a incluí-los no dicionário como também advérbios. Por isso, a melhor - e talvez única - maneira de obter estes casos é primeiro (nos corpos que a interessam) ver a lista de frequências dos advérbios e selecionar aqueles que considera adjetivos, e depois procurá-los – ou seja, acrescentá-los na expressão acima.

Em <http://www.linguateca.pt/acesso/ordenador.php>

é possível obter frequências de itens lexicais nos variados corpos.

Como obter as frases em que os verbos de movimento como *ir*, *vir*, *entrar*, *subir*, *partir*, *chegar*, *caminhar*, *correr*, *nadar*, *cambaleiar*, *tropeçar* etc, nos diversos tempos e modos, coocorrem com as preposições *a*, *para*, e *até*.

A forma mais simples seria procurar

Procura: [lema="ir|vir|entrar|subir|partir|chegar|caminhar|correr|nadar|cambaleiar|tropeçar"] [lema="a|para|até" & pos="PRP"]

Em que os verbos listados aparecerão em todos os tempos e modos, e serão seguidos **imediatamente** pelas preposições *a*, *para*, e *até*.

No entanto, como é muito extensa e com verbos muito frequentes, desaconselhamos essa forma de procura. Sugerimos então que separe cada verbo em uma busca diferente. Para o verbo “ir”, portanto, a expressão seria simplesmente

Procura: [lema="ir"] [lema="em|a|para|até" & pos="PRP"]

Pode ainda realizar uma busca para cada tipo de preposição. Assim, para procurar apenas o verbo “ir” seguido imediatamente da preposição “para”:

Procura: [lema="ir"] [lema="para" & pos="PRP"]

Pode, ainda, desejar que a preposição não apareça imediatamente após o verbo “ir”. A expressão abaixo significa encontrar quaisquer formas do verbo IR seguidas de um número zero ou mais de palavras que não sejam elas próprias verbos (isto é, cuja POS não comece por V) nem preposições (cuja POS não comece por P) e que também corta PONTUAÇÃO...)

Procura: [lema="ir"] [pos!="[VP].*"]* [lema="para" & pos="PRP"]

Se quiser saber quais são as palavras que aparecem entre o verbo IR e a preposição PARA, basta colocar @ imediatamente antes do elemento da expressão correspondente e pedir a distribuição dos lemas:

Procura: [lema="ir"] @[pos!="[VP].*"]* [lema="para" & pos="PRP"]

Distribuição dos lemas.

Como encontrar sujeitos pronominais entre uma locução verbal (construída com verbo auxiliar)?

(“Para o exercício destas funções era ele assessorado por três funcionários..”; “Seja ela causada pelos deveres de actor, de encenador ou de político”; “Que tinha ele feito, o amigo Artur?”)

Procura: [pos="PERS" & temcagr="NOM"] [func="IMV_#ICL-AUX.*"]

15. ARGUMENTOS E VALÊNCIAS

É possível obter listas de subcategorização verbal? Como encontrar argumentos de verbos? E de nomes?

Não temos listas de subcategorização verbal, mas é possível procurar ocorrências de preposições que foram consideradas, pelo analisador sintático automático, como argumentos dos verbos ou substantivos específicos.

Por exemplo, ao procurar

Procura: [word="a" & func="<PIV"]

aparecem todos os casos da preposição “a” analisada como argumento do verbo.

Se quiser os verbos que aparecem associados a essa preposição, basta fazer

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

Resultado: Distribuição de lemas

Essa expressão busca todos os verbos que, naquele corpo particular, aparecem imediatamente antes do sintagma preposicional.

Para saber quais palavras aparecem depois da preposição “a” (e que, portanto, completam o sintagma preposicional), basta procurar

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

Para ver apenas as palavras que aparecem depois da preposição “a”, basta acrescentar @ imediatamente antes do elemento da expressão correspondente:

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

Resultado: Distribuição dos lemas

Com esse tipo de expressão, é possível, naturalmente, buscar argumentos de verbos específicos. Por exemplo, para os argumentos de “assistir”:

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

SUBSTANTIVOS: Com relação aos substantivos, no caso de “medo”, por exemplo, a expressão será:

Procura: [lema="medo"] [func="N<"]* [func="N<" & pos="PRP"] [func=">N"]* [pos="N.*"]

A diferença nessa busca está na função sintática (func) do segundo elemento da busca: como medo é um nome, os complementos desse nome terão a função “N<”.

Se não quiser especificar o substantivo da busca, a expressão deve ser:

Procura: [pos="N.*"] [func="N<"]* [func="N<" & pos="PRP"] [func=">N"]* [pos="N.*"]

No entanto, essa é uma busca que retornará um grande número de ocorrências, e portanto sugerimos usar um corpo não muito grande. Além disso, é importante notar que a expressão não diferencia entre os complementos (ou argumentos) do nome, e os modificadores do nome, pois essa informação não é fornecida pelo analisador PALAVRAS. Se está especialmente interessado neste tipo de procura, aconselhamos o uso do Bosque e da Selva, partes da Floresta Sintática, e que foram anotados com esse tipo de informação/distinção, representados pelas categorias N<ARG e N<, para argumentos e modificadores do nome, respectivamente.

A Floresta Sintá(c)tica está em <http://www.linguateca.pt/Floresta/>

16. ORAÇÕES COM VERBOS NO GERÚNDIO

Como encontrar orações no gerúndio do tipo “Ela saiu cantando” ; “O navio afundou matando 100 pessoas” ?

Procura: [pos="V.*" & lema!="ser|estar|ir|acabar|continuar|vir|ficar" & temcagr="PR_IND|PS_IND"]
[temcagr="GER" & func="IMV_#ICL.*"]

A concordância mostra as frases, a distribuição mostra as palavras diferentes.

Experimente, primeiro, a concordância. Depois se quiser saber quais os verbos que estão no gerúndio, peça acrescente @ imediatamente antes do elemento da expressão que indica o verbo no gerúndio, isto é:

Procura: [pos="V.*" & lema!="ser|estar|ir|acabar|continuar|vir|ficar" & temcagr="PR_IND|PS_IND"]
@[temcagr="GER" & func="IMV_#ICL.*"]

E peça Distribuição de lemas. O sinal @ indica de qual item se quer fazer a distribuição, se não puser, será sempre da primeira palavra.

Note também que, na expressão acima, tentamos eliminar verbos que costumam funcionar como partes de uma locução (quando fizemos lema != (isto é, lema diferente de) *ser, estar, ir, acabar, continuar, vir, ficar*). Mas note que outros verbos também podem ter este comportamento. Se quiser incluir outros verbos no filtro, basta acrescentá-los à expressão de busca.

Como encontrar orações com gerúndio em posição adnominal, ou seja, nome seguido de gerúndio? E como especificar que esse nome comece com a letra b? E como inserir uma vírgula entre o nome e o gerúndio?

Para encontrar um nome seguido de gerúndio:

Procura: [pos="N"] [temcagr="GER"]

Para especificar que o nome começa com b, basta acrescentar word="b.*" `a expressão:

Procura: [pos="N" & word="b.*"] [temcagr="GER"]

Se quiser o infinitivo das palavras que estão no gerúndio, junte um @, ou seja,

Procura: [pos="N"] @[temcagr="GER"]

e peça Distribuição dos lemas.

Se quiser saber quais os nomes, basta pedir distribuição dos lemas, e não é preciso acrescentar @, pois o nome é o primeiro elemento da procura. No entanto, conforme o corpo utilizado na busca, haverá um número demasiado grande de ocorrências de nome - mais de 300 milhões - e o sistema pode demorar muito tempo e não devolver resultado nenhum. A solução é pedir um número fixo (veja a seção visualização dos resultados), ou escolher um determinado nome ou verbo.

Para inserir uma vírgula, basta inseri-la, entre aspas, entre as duas palavras:

[pos="N" & word="b.*"] "," [temcagr="GER"]

17. ORAÇÕES RELATIVAS/ ORAÇÕES ADJETIVAS

Como encontrar orações relativas/orações adjetivas?

Para encontrar orações relativas, basta buscar pelo pronome relativo:

Procura: [pos="SPEC_rel"]

Para saber quais são os pronomes relativos usados com mais frequência basta pedir

Distribuição dos lemas

Se desejar ampliar a busca para relativas introduzidas também por “o qual”

Procura: [pos=".*_rel" & lema="qual|que|quem"]

Para saber as classes atribuídas aos pronomes na busca acima

Resultados: Distribuição da categoria gramatical (PoS)

Como encontrar pronomes relativos com função de objeto direto?

Procura: [pos="SPEC_rel" & func="ACC.*"]

Para as demais funções que podem ser exercidas pelos pronomes relativos, ver a página referente à anotação dos corpos:

<http://www.linguateca.pt/acesso/ anotacao.html>

Como encontrar estruturas do tipo "as coisas de que precisava", "equipe em que Ayrton Senna foi tricampeão", em que a preposição está anteposta ao verbo que a exige?

Procura: [lema="que" & pos="SPEC_rel" & func="P<"]

Como saber quais são os verbos que participam nas estruturas acima?

Procura: [lema="que" & pos="SPEC_rel" & func="P<"] [pos!="V.*"]* @[pos="V.*"]

Resultado: Distribuição dos lemas

Como saber quais os nomes que participam nas estruturas acima? (isto é, os nomes que antecedem as preposições):

Procura: [pos="N.*"] [pos!="V.*"]*[lema="que" & pos="SPEC_rel" & func="P<"] [pos!="V.*"]* [pos="V.*"]
within s

Resultado: Distribuição dos lemas

18. ORAÇÕES SUBSTANTIVAS / COMPLEMENTO ORACIONAL

Como encontrar orações com um determinado verbo (por exemplo, “mandar”) cujo complemento é uma oração introduzida pela conjunção integrante “que”. Ou seja, frases como “Bateram na testemunha e mandaram que os três entrassem”; “A regra manda que as melhores instalações ficam com quem chegar primeiro.”

Procura: [lema="mandar"] [word="que" & pos="KS"]

Como encontrar uma oração finita com a função sintática de sujeito, predicativo do sujeito, complemento direto ou indireto, e que seja introduzida por uma conjunção subordinante?

A conjunção subordinante / subordinativa é expressa como “KS”.

Estas são as possíveis funções sintáticas:

SUB_#FS-<ACC : complemento direto

SUB_#FS-<SUBJ : sujeito

SUB_#FS-<SC : nome predicativo do sujeito

SUB_#FS-P< : complemento de preposição

Para as conjunções subordinantes com a função de objeto direto, por exemplo, a expressão de busca será:

Procura: [pos="KS" & func="SUB_#FS\-<ACC"]

Que retornará frases como

*Apesar da iniciativa, Rossi reluta em admitir **que** montará um programa de governo.*

Para garantir que o verbo na oração subordinada/subordinante seja finito (no exemplo abaixo o tempo está ou no indicativo ou no conjuntivo/subjuntivo)

Procura: [pos="KS" & func="SUB_#FS\-<ACC"] [pos!="V.*"]* [pos="V.*" & temcagr=".*(IND|SUBJ)"]

Como encontrar frases em que o complemento do verbo é uma oração no conjuntivo/subjuntivo? (“Eu te disse que espero mesmo a sério que, no caso de as coisas continuarem como estão, tu consigas o que queres.”) Como encontrar a estrutura “verbo (por ex, “esperar”) + qualquer coisa + complemento do verbo no conjuntivo/subjuntivo”?

Uma forma aproximada para a busca é

Procura: [lema="esperar"] [* @ [word="que" & pos="KS.*" & func="SUB_#FS-<ACC"] [pos!="V.*"]* [temcagr=".*SUBJ.*"] within s;

(Vale lembrar que, como o PALAVRAS não acerta 100%, é possível que muitos dos casos de KS não seja, de fato, KS. A Florest Sintá(c)tica, sendo revista por pessoas, é naturalmente muito mais confiável.)

Como encontrar frases com um determinado verbo (por exemplo, “mandar”) cujo complemento é uma oração no infinitivo? Ou seja, frases como “Mande ele entrar em contato conosco”; “Eu mandei também pintar uma faixa de boas vindas a ele”; “O governo ditatorial brasileiro detestou a idéia e mandou a embaixada vetar os filmes .”

Procura: [lema="mandar"] [pos!="V.*"]+ [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Como encontrar orações com um determinado verbo (por exemplo, “deixar”) + pronome reflexivo “se” + verbo no infinitivo? (“Capitu deixou-se fitar e examinar”; “Estava fumando à janela, e fumando deixou-se ficar”; “E deixou-se cair para trás, na rede, a estrebuchar.”)

Procura: [lema="deixar\+se"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Para encontrar outras ocorrências de “deixar” + pronomes (sem precisar garantir que sejam reflexivos, como “deixar-me”; “deixar-te”; “deixar-se”)+ verbos no infinitivo:

Procura: [lema="deixar\+(eu|tu|se|nós|vós)"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Se quiser garantir que os pronomes são reflexivos, é preciso uma expressão de procura para cada pronome:

[lema="deixar\+eu" & pessnum="1S\+1S"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"].

[lema="deixar\+tu" & pessnum="2S\+2S"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"].

[lema="deixar\+nós" & pessnum="1P\+1P"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"].

[lema="deixar\+vós" & pessnum="2P\+2P"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"].

Para procurar os pronomes em posição proclítica (anteriores ao verbo), como em “me deixei ficar”, é preciso garantir que estão na mesma pessoa que o verbo.

Para pronomes na primeira pessoa do singular:

Procura: [pos="PERS_refl" & temcagr!="ACC" & pessnum="1S"] [lema="deixar" & pessnum="1S"]
[pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Para pronomes na terceira pessoa do singular:

Procura: [pos="PERS_refl" & temcagr!="ACC" & pessnum="3S/P"] [lema="fazer" & pessnum="3S"]
[pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Como encontrar pronomes do caso nominativo que apareçam entre determinados verbos (por exemplo, “mandar”) e um verbo no infinitivo? “A Maria mandou eles entrarem.” “Ele mandou eu colocar o cinto”; “Quem mandou ele se expor tanto?”; “Mandei você fazer isso”.

Procura: [lema="mandar"] [temcagr="NOM"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Para saber quais os verbos que mais frequentemente antecedem os pronomes nesse tipo de construção, a busca deve ser

Procura: [pos="V.*"] [temcagr="NOM"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Resultado:: Distribuição dos lemas

Para saber quais os verbos que mais frequentemente aparecem depois dos pronomes nesse tipo de construção, basta colocar @ imediatamente antes do elemento da expressão de busca que corresponde ao verbo no infinitivo (e pedir a distribuição dos lemas)

Procura: [pos="V.*"] [temcagr="NOM"] [pos!="V.*"]* @[temcagr="INF" & func=".*#ICL\-<ACC.*"]

Resultado: Distribuição dos lemas

Como encontrar frases com os infinitivos pessoais? (*“Três feirantes acusados de participarem de grupos de extermínio”; “Então vi os motores pegarem fogo”; “a possibilidade de termos um novo padrão”; “tentando convencer os índios a adotarem a religião cristã.”*)

Tanto os infinitivos pessoais como os infinitivos impessoais estão marcados INF. Para buscar só infinitivos pessoais, é preciso de juntar a restrição ao atributo pessnum:

Procura: [temcagr="INF" & pessnum="(*P)|(2S)"]

Como encontrar o clítico "lhe/lhes" em construções em que o verbo principal é “mandar” ou “fazer” e o outro verbo está no infinitivo? (*“Eu pensei que ele lhe mandasse dar alguns banhos”; “mandou-lhe dizer que viesse”; “que eu lhe mandaria fazer de propósito”.*)

A busca abaixo é apenas para o verbo “mandar”:

Procura: [word=".*lhes*" & func=".*DAT>"] [lema="mandar"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"] | [lema="mandar\+.*" & word=".*lhes*"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"]

Para incluir, na mesma busca, o verbo fazer, basta acrescentar |*fazer* ao lado de mandar na expressão:

Procura: [word=".*lhes*" & func=".*DAT>"] [lema="mandar|fazer"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"] | [lema="mandar|fazer\+.*" & word=".*lhes*"] [pos!="V.*"]* [temcagr="INF" & func=".*#ICL\-<ACC.*"]

19. ORAÇÕES ADVERBIAIS

Como encontrar ocorrências de verbos no futuro do conjuntivo/subjuntivo que seguem a conjunção "quando"?

Procura: [lema="quando"] [pos!="V.*"]* [temcagr="FUT_SUBJ"]

Como encontrar orações introduzidas pela conjunção “se” + verbo no imperfeito do conjuntivo/subjuntivo + verbo no condicional (por ex, verbo “tornar”)? (“Se dependesse unicamente da boa vontade dos paulistanos, São Paulo se tornaria de um dia para o outro uma cidade melhor” ; “Se a bancarrota fosse declarada, Tapie se tornaria inelegível”)

Procura: [lema="se" & pos="KS"] []* [temcagr="IMPF_SUBJ.*"] []* [lema="tornar(\+.*)*" & temcagr="COND.*"] within s

O "within s" é para forçar que a ocorrência se encontre dentro da mesma frase

O lema="tornar(\+.*)*" significa: ou só “tornar”, ou “tornar” seguido de + com qualquer coisa.

Lembramos que como é uma procura bastante específica, é possível que não encontre ocorrências nos corpos menores.

Se estiver especialmente interessado em orações adverbiais, sugerimos consultar também o Bosque e a Selva, partes da Floresta Sintá(c)tica

<http://www.linguateca.pt/Floresta/>

que contêm etiquetas especiais para as orações adverbiais (além de terem sido revistos por linguistas). As informações sobre as orações na Floresta estão em

<http://www.linguateca.pt/Floresta/BibliaFlorestal/sec14.html>

(uma seção do manual da Floresta Sintá(c)tica, a Bíblia Florestal, que está em

<http://www.linguateca.pt/Floresta/BibliaFlorestal>

Para encontrar as frases na Floresta, pode usar a interface de busca Milhafre

<http://www.linguateca.pt/Floresta/milhafre/>

e, no campo Procurável, digitar a etiqueta correspondente à oração desejada. As etiquetas estão em

<http://www.linguateca.pt/Floresta/BibliaFlorestal/anexo1.html#s3>

PARTE IV – Não encontro o que procuro

20. QUANDO NÃO SEI EM QUE SEÇÃO DAS PJR PROCURAR

Nem sempre (aliás, quase nunca) temos uma classificação linguística única para um tipo de fenômeno – isto é, naturalmente, um mesmo fenômeno pode ser abordado por diferentes perspectivas. Por exemplo, o pedido

Como encontrar pronomes do caso nominativo que apareçam entre determinados verbos (por exemplo, “mandar”) e um verbo no infinitivo? “A Maria mandou eles entrarem.” “Ele mandou eu colocar o cinto”; “Quem mandou ele se expor tanto?”; “Mandei você fazer isso”.

Pode ser visto tanto como uma procura que envolve *orações substantivas/complemento oracional* como uma procura que lida com *colocação pronominal*, pelo menos, o que representa um problema do ponto de vista prático da classificação das buscas possíveis no AC/DC. (E o exemplo acima integra a seção “*orações substantivas/complemento oracional*”) Por isso, sugerimos que, caso não encontre o que procura na seção que seria esperada, não desanime e veja se não há nada que interesse em outras seções relacionadas.

21. QUANDO FAÇO UMA BUSCA E OBTENHO ZERO RESULTADOS

Não encontro resultados, mesmo usando uma expressão idêntica à do manual!

Tente mudar o corpo. Provavelmente não está conseguindo resultados pois a expressão é muito restrita - e talvez pouco frequente nos corpos do AC/DC. Nesse caso, experimente a busca em com o corpo Todos juntos. Em seguida, se quiser saber em qual corpo a expressão é mais frequente, para usar somente esse corpo, peça, nos Resultados, a Distribuição por corpo. Em

<http://www.linguateca.pt/acesso/legenda.php>

está uma lista com a explicação das siglas de cada corpo.

Quando tento fazer uma busca de uma expressão (por exemplo, “rir no”), sempre obtenho a indicação de erro.

Isto acontece porque o sistema de procuras do AC/DC é apenas para frequências de palavras (ou grupos de palavras, como todas as palavras que possuem o com o mesmo prefixo, ou o mesmo sufixo, etc), na sua forma flexionada ou como lemas. Assim, espaços na "palavra" fazem com que o sistema não encontre nada. Para encontrar “ir no” deve fazer – caso esteja interessado nessa expressão exatamente:

Procura: [word="ir"] [word="no"]

Para obter não apenas “ir no”, mas também “foi no”; “foram no”; “fui no” etc, basta usar a informação de lema

Procura: [lema="ir"] [word="no"]

(mas lembre-se de que essa busca deve encontrar muitas ocorrências, e por isso desaconselhamos seu uso em corpos muito grandes, que às vezes podem demorar a retornar resultados.)

Por que não obtenho resultados para sequências como "pode-me ajudar" no português de Portugal?

O mais importante quando se está a trabalhar com corpus é perceber de onde vêm os textos. No AC/DC, corpus como CETEMPúblico, CHAVE , NILC/São Carlos, CONDIV e DiaCLAV, contêm principalmente textos jornalísticos, que raramente são escritos na primeira pessoa...