

Exercícios e exemplos do uso do COMPARA para estudos contrastivos

Diana Santos

12 de Dezembro de 2008

Versão 1.0

Este texto foi criado como ajuda para os alunos do curso "Curso avançado de estudos contrastivos usando o COMPARA como ferramenta", que foi leccionado na Segunda EbraLC em São José do Rio Preto. Como segundo objectivo, espera-se que possa constituir uma forma de apoio ao mesmo para futuros alunos ou interessados no COMPARA.

Assim, várias das questões postas no curso são respondidas de forma preliminar, para ajudar o aluno a consolidar os seus conhecimentos. A ideia é que o aluno tente primeiro obter os resultados, e apenas depois confirme se o seu modo de proceder foi o mais adequado. Note-se que a ordem dos exercícios ou perguntas segue a do curso, e não a da dificuldade.

É também conveniente familiaridade com o IMS-CWB e com os outros corpos disponibilizados pela Linguateca, em particular os do projecto AC/DC, e com o PALAVRAS desenvolvido por Eckhard Bick. No fim deste texto encontra-se leitura recomendada para quem queira explorar a fundo a forma de interrogar o COMPARA e outros corpos disponibilizados de forma semelhante. Em anexo junto também uma versão mais actualizada da ajuda do COMPARA, melhorada por ocasião do curso, em Novembro de 2008, portanto.

Exemplo de simples consulta:

Como se escreve (o correspondente a) *publicamente* em inglês: *publicly* ou *publically*?

Escolher a direcção inglês para português

"public.*ly"

Comentário: Obviamente que isto não será uma boa forma de proceder no caso de *actualmente* ou *atualmente*, visto que não são os correspondentes de *actually*. Ou seja, este é um caso típico de sócias enganadores, para os quais os professores deverão mesmo chamar a atenção explicitamente.

Nesses casos, o ideal seria procurar a tradução de *actually*, inserindo simplesmente essa palavra na caixa de procura

E a tradução de

ac*tualmente

(para permitir simultaneamente a grafia com e sem c)

Se ficar ainda com dúvidas sobre a intertradutibilidade dessas duas palavras, experimentar

actually restrição¹ ac*tualmente

e ver a distribuição combinada².

Esta é, aliás, a forma mais garantida de verificar se palavras cognatas são boas traduções uma da outra.

Como é que se traduz *supposed to*?

Escolher a direcção inglês para português

"supposed" "to"

¹ Uso a palavra restrição neste texto para indicar a segunda caixa de procura (identificada por "restrição de alinhamento" na interface do COMPARA).

² Uso indiscriminadamente os termos distribuição cruzada ou combinada para indicar essa funcionalidade do COMPARA, que já mudou várias vezes de nome ao longo da história deste projecto.

Os modais *poder* e *can/could* são usados da mesma maneira?

Escolher

[lema="poder" & pos="V.*"] restrição "can|could"

Distribuição cruzada

Para ver quantos "poderes" haverá, faça a distribuição pela categoria gramatical de

[lema="poder"]

Idem para *can*...

Que verbos são mais usados com o advérbio *home*?

Vamos começar com o pedido de uma visão geral. Se forem os imediatamente anteriores

[pos="V.*"] [word="home" & pos="R.*"]

Distribuição dos lemas

Se for qualquer verbo antecedendo o advérbio *home* (não necessariamente imediatamente)

[pos="V.*"] [pos!="V.*"] [word="home" & pos="R.*"]

Que adjetivos modificam 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

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

Pedido de: distribuição de lemas

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

Tipo de resultados: distribuição de lemas

Pode tb ter interesse verificar qual a distribuição de *pedra* em termos do número, e ver se isto muda quando as pedras estão sozinhas ou qualificadas.

[lema="pedra"]

Distribuição de pessoa ou número

O que é mais azul em português? E em inglês? E cor-de-rosa?

Admitindo que estamos sobretudo interessados em coisas (e portanto substantivos), procuramos

[pos="N.*"] "azu.*"

[pos="N.*"] [lema="azul"]

[pos="N.*"] [cor="Azul.*"]

Distribuição das formas ou dos lemas

As várias alternativas correspondem à palavra *azul* e seus derivados, de uma forma simplesmente gráfica (palavras começadas pelos caracteres *azu*), de uma forma gramatical (pedindo o lema *azul*), ou de uma forma semântica (qualquer palavra marcada como pertencendo ao grupo de cor Azul), isto porque as cores principais foram anotadas como tal em grupos de cor, um dos quais é o Azul.

Em alguns casos, marcamos mais de uma cor para a mesma unidade. Há portanto diferença entre, por exemplo, os dois pedidos seguintes:

[pos="N.*"] [cor="Verde"]

[pos="N.*"] [cor=".*Verde.*"]

Para obter as correspondentes em inglês, a pergunta seria:

```
[colour="Blue.*"] @[pos=N.*]
```

Reparem que o nome do atributo neste caso, e só neste caso, é diferente (`colour`). E também que se queremos as coisas (e não as cores) temos de marcar com `@` o campo de que pretendemos a distribuição.

E o cor de rosa, como identificar?

Aqui o que é interessante é que há três maneiras, só para cor de rosa, de explicitá-la: com hífens, sem hífens, e sem *cor de...* e há várias maneiras, portanto, de obter o resultado, sendo contudo a melhor a semântica, visto que foi revista exaustivamente pela equipa.

```
[pos="N.*"] [cor=".*Rosa.*"]
```

Que tradução é mais comum para *despedir* ?

A primeira questão a considerar é que certamente há pelo menos dois campos semânticos em inglês para esta palavra: o que tem a ver com emprego, e o que tem a ver com dizer adeus (se em português há uma relação, não entramos por aqui agora). Naturalmente que existem casos fáceis de distingui-los em português (a passiva com *ser*, ou a forma transitiva com pronome *ele despediu-o*), mas nem todos os casos são tão fáceis de decidir automaticamente, veja-se *Ele despediu-se* (do emprego ou dos filhos?)

Aqui a vantagem do COMPARA é que se pode logo pedir três casos separados:

```
[lema="despedir\+se"]
```

```
[lema="despedir\+ele"]
```

```
[lema="despedir" & temcagr="PCP"]
```

Não se lembrava de qual a mnemónica para o particípio passado? Nada mais simples. Ponha um caso no COMPARA de que sabe a análise, e peça a distribuição por `temcagr`.

E porquê o carácter `\`? Porque o `+` significa um ou mais casos, por isso `despedir+se` emparelharia com *despedirse*, *despedirirse*, *despedirrrrrrrrse*, etc., mas não com o carácter `+`.

Para procurar os casos de *despedir de* (e notem que pertencem aos dois campos), poderíamos usar o seguinte:

```
[lema="despedir"] [lema="de(\+.*)*"]
```

O segundo lema junta não só a preposição *de* sozinha como todos os casos de contracção entre *de* e outra coisa. Os lemas das contracções juntam os vários lemas através do carácter `+`, como já referido. Sem o `()*`, exigíamos sempre uma contracção, como podem apreciar se experimentarem

```
[lema="despedir"] [lema="de\+.*"]
```

Vemos, pois, que há muitas despedidas no COMPARA, mas que cada forma parece levar a uma tradução diferente. *Dismiss, fire, sack, let go...* por um lado, e *say farewell, say goodbye, see off, take leave...* por outro.

Aqui poderia ser interessante ver os casos do substantivo também: *despedida* e *despedimento*.

```
[pos="N" & lema="desped.*"]
```

Observar os casos de tradução de *andar a -ar* ou *andar -ndo*

Passando agora para casos mais gramaticais:

```
[lema="andar"] [temcagr="GER"]
```

```
[ lema="andar" ] "a" [ temcagr="INF" ]
```

Este é um caso em que o português é muito mais expressivo do que o inglês, e é interessante ver as várias formas que o inglês usa nos correspondentes (originais ou traduzidos).

Passando agora ao usos do COMPARA em lexicografia bilingue:

Quais as traduções de *bom*? Que não sejam as estafadas, tais como *good, fine, great...*

Poderíamos começar por aqui:

```
[ lema="bom" ] restrição ! "(good|fine|great)"
```

Mas daí víamos que ainda poderíamos retirar casos como *better, best*, ou maiúsculas, o que passa o número de casos de 1087 para 648 (na versão 11.0.0)...

Outra forma poderia ser olhar para as concordâncias paralelas de palavras com frequência, por exemplo, maior do que 15, para servirem de entradas no dicionário.

Poder-se-ia inspeccionar os adjectivos na lista de frequências do COMPARA, ou obtê-las automaticamente através do COMPARA:

Na direcção de inglês para português

```
[ pos="J.*" ] restrição [ lema="bom" ]
```

Distribuição dos lemas

Outra forma seria simplesmente obter directamente a distribuição de algumas palavras que já sabemos que são traduções:

```
(bom|famoso|maravilhoso|apropriado|bem) restrição good
```

Qual a construção mais comum com o verbo *dizer*?

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

A primeira é

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

A segunda é

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

A terceira é

```
[ lema="dizer" ] [ word="a" & pos="PREP" ] [ pos="N.*" ]
```

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

```
( [ 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. Isto nem sempre, naturalmente, acontece, e noutros casos ter-se-ia portanto de fazer três perguntas distintas ao COMPARA.

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 sintáctica³).

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 SPECrel (as duas categorias mais frequentes para *que*), e simplesmente pedir esses casos:

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

Isto daria para ver que KS indica *que* como conjunção integrante.

```
[word="que" & pos="SPECrel"]
```

E este mostraria que SPECrel era ou parecia ser o que nós queríamos. Poderíamos depois ver se há mais palavras marcadas com esta categoria

```
[pos="SPECrel"]
```

Pedindo distribuição de lemas, o que inesperadamente apenas daria *que* e *quem*. Admitindo que tb gostávamos de encontrar orações com *cujo* e com *qual* etc teríamos de ir 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".

```
"cujo|qual|quanto"
```

Daí veríamos que tb existe a categoria DETrel, e que possivelmente há mais categorias com o sufixo "rel". Fazendo

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

Pedindo então a distribuição por categoria gramatical e por lema, vemos que existem SPECrel ADVrel e DETrel, e que ADVrel 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.

```
[pos="*.rel"] [pos!="V.*"]*@[pos="V.*"]
```

Mais do que resolver alguma questão linguística sobre verbos em orações relativas, esta secção mostrou como procurar n(a parte portuguesa d) o COMPARA e explorar a quantidade de informação nele presente. Não é preciso saber todas as decisões minuciosas feitas no PALAVRAS para conseguir procurar com a ajuda deste.

Com que verbo *não* é mais frequente?

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

```
"não" @[pos="V.*"]
```

Ou então aceitar também clíticos

```
"não" [pos="PERS"]* @[pos="V.*"]
```

Tipo de resultado: distribuição de lemas.

Quando é que *sempre* é usado?

Por exemplo, qual o tempo verbal que mais co-ocorre? Basta pedir a distribuição do tempo verbal, caso ou grau.

³ Para esta segunda opção, é preciso fazê-lo no AC/DC visto que na interface do COMPARA não dá essa possibilidade, por a equipa ter considerado que era complicado demais.

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

Acabando este conjunto de exemplos por questões que podem não ser evidentes no COMPARA, comecemos pelo lema de substantivos femininos "regulares" ou de advérbios: Dependendo da teoria gramatical, pode haver pessoas que esperem que o lema de *professora* seja *professor*, ou que o lema de *amigavelmente* seja *amigável*. É por isso importante que verifiquem qual a abordagem seguida no COMPARA, bastando para isso procurar algumas dessas palavras e pedir o seu lema (tipo de resultado: distribuição de lemas).

Ainda mais complicada é a análise de participios passados, donde sugiro que observem a categoria gramatical de palavras em *-ado*, e já agora o mesmo para o lado inglês para as palavras em *-ing*.

Outra questão que é importante salientar tem a ver com o facto da nossa escolha em permitir análises vagas, ou seja, mais do que uma análise possível por palavra, quando o contexto não nos permite decidir.

A melhor forma é procurar, por exemplo, a categoria gramatical de *amigo* e continuar a explorar em profundidade. O mais importante, contudo, é compreender que as contas (a descrição quantitativa) se torna mais complicada ao aceitar categorias vagas.

Leitura recomendada

Além do material acessível de <http://www.linguateca.pt/COMPARA>, que, se repetido aqui, aparece a negrito, há muita outra documentação que pode ser útil.

- Acesso a corpora de português: Exemplos. Projecto AC/DC, <http://www.linguateca.pt/acesso/exemplos.html>
- Anotação dos corpora. Projecto AC/DC, <http://www.linguateca.pt/acesso/ anotacao.html>
- Atomização e separação de frases. Projecto AC/DC, <http://www.linguateca.pt/acesso/atomizacao.html>
- Eckhard Bick. *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Dr.phil. thesis. Aarhus University. Aarhus, Denmark: Aarhus University Press. November 2000.
- Eckhard Bick. Portuguese Syntax: Teaching manual, http://visl.sdu.dk/visl/pt/syntaxmanual_0.html
- CLAWS7 tagset. <http://www.comp.lancs.ac.uk/ucrel/claws7tags.html>
- Oliver Christ, Schulze, Bruno M., Hofmann, Anja, & Koenig, Esther. The IMS Corpus Workbench: Corpus Query Processor (CQP): User's manual, Institute for Natural Language Processing, University of Stuttgart, August 16, 1999 (CQP V2.2). <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPUserManual/HTML/>
- Stefan Evert. "The CQP Query Language Tutorial (CWB version 2.2.b90)", University of Stuttgart, 10 July 2005. <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPTutorial/cqp-tutorial.pdf>
- IMS Corpus Workbench: A (very) brief description of the query syntax. <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPSyntax.html>
- IMS Corpus Workbench: Some query examples. <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPExamples.html>
- **Susana Inácio & Diana Santos. "Documentação da anotação morfossintáctica da parte portuguesa do COMPARA". Primeira versão: 9 de Dezembro de 2005. Última versão: Dezembro de 2008.** <http://www.linguateca.pt/COMPARA/DocAnotacaoCorCOMPARA.pdf>
- Diana Santos. "Contrastive Linguistics and NLP", lecture at the *Forskerutdanningsseminar om Midler og Maal i Kontrastiv Linvistikk*, Arts Faculty, University of Oslo, 26 November 2002.
- Diana Santos. "Utilizações de corpora", 7 de Outubro de 2003. http://www.linguateca.pt/Diana/usuarios_corpora.html
- Diana Santos. "Working with Portuguese corpora", apresentação na Universidade de Oslo, 22 de Outubro de 2004, como extensão da apresentação no ISLA, em Lisboa, a 1 de Outubro de 2004. <http://www.linguateca.pt/Diana/download/KRI2004.pdf>
- Diana Santos. "Breves explorações num mar de língua". *Ilha do Desterro* 52 (2007), pp. 127-150. Florianópolis. ISSN: 0101-4846.
- Diana Santos. "Corpos linguísticos da Linguateca: apresentação", *TaLC at TaLC: Teaching and Linguateca's (Portuguese language) Corpora* (ISLA, Lisboa, 3 de Julho de 2008). <http://www.linguateca.pt/Diana/download/SantosWorkshopTaLC2008.pdf>
- **Diana Santos. "Curso avançado de estudos contrastivos usando o COMPARA como ferramenta". Módulo na EBraLC, Segunda Escola Brasileira de Linguística Computacional (Universidade Estadual Paulista - UNESP - Campus de São José do Rio Preto, 3-5 de Novembro de 2008).** <http://www.linguateca.pt/documentos/cursosCOMPARASantosEBRALC2008.pdf>

- 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. <http://www.linguateca.pt/documentos/SantosRanchhodPROPOR99.pdf>
- Diana Santos & Luís Sarmiento. "O projecto AC/DC: acesso a corpora/disponibilização de corpora". In Amália Mendes & Tiago Freitas (eds.), *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística (APL 2002)* (Porto, Portugal, 2-4 de Outubro de 2002), Lisboa : APL, pp. 705-717. <http://www.linguateca.pt/documentos/SantosSarmientoAPL2002.pdf>
- **Silva, Rosário, Susana Inácio & Diana Santos. "Documentação da anotação relativa à cor no COMPARA". Primeira versão: 27 de Novembro de 2007. Última versão: Novembro de 2008.**
<http://www.linguateca.pt/COMPARA/DocAnotacaoCorCOMPARA.pdf>
- **Rosário Silva & Ana Frankenberg-Garcia. "Anotação inglesa do COMPARA com o CLAWS C7: critérios de revisão". Novembro de 2008.**
<http://www.linguateca.pt/COMPARA/docum/AnotIngRevisao.pdf>

Também sugiro a visita às páginas dos projectos AC/DC, Floresta Sintá(c)tica, e CorTrad, projectos congéneres do COMPARA e que podem complementar a informação ou a experiência de investigar corpos paralelos (ou monolingues) analisados sintacticamente com o PALAVRAS:

- <http://www.linguateca.pt/ACDC/>
- <http://www.linguateca.pt/Floresta/>
- <http://nilc.icmc.usp.br/dispara/CorTrad/>

Anexo: **Ajuda à pesquisa no COMPARA** (instruções actualizadas no fim de Outubro de 2008, sobre o texto comum criado pela equipa)

Seleccione a direcção de pesquisa

Deixe a opção "de português para inglês" seleccionada se vai escrever uma palavra ou expressão em português de modo a pesquisar a sua tradução em inglês. Para inverter a direcção da pesquisa, seleccione a opção "de inglês para português".

Pesquisa por palavra ou expressão

A caixa de pesquisa é o espaço onde deve introduzir a palavra ou expressão que deseja pesquisar. A sua pesquisa pode ser uma simples palavra, uma combinação de palavras, um prefixo, um sufixo ou muitas outras formas de representar elementos da língua (associados à sua morfologia ou sintaxe).

Como o COMPARA é consituído por textos literários, está mais indicado para pesquisas que envolvam vocabulário genérico, a linguagem própria da literatura e palavras com significados gramaticais (por exemplo, artigos, preposições, conjunções, pronomes, etc.). É pouco provável que encontre termos técnicos no COMPARA, assim como assuntos de interesse político da actualidade.

Também não recomendamos a pesquisa de frases inteiras. Para obter bons resultados com corpos linguísticos, é melhor trabalhar com sequências de palavras mais curtas, cuja probabilidade de já terem sido proferidas seja elevada. Ou seja, será mais bem sucedida se pesquisar palavras isoladas e expressões convencionais como *hoje em dia* do que frases únicas como *Gosto de beber chá antes de me deitar*.

Abaixo encontram-se alguns exemplos de expressões de pesquisa com base apenas na forma do texto. Relembramos que é preciso pôr aspas em torno de cada palavra ou elemento da expressão de pesquisa se esta contiver mais de uma palavra.

O que quer pesquisar	Exemplo do que deve digitar	Possíveis resultados
<i>palavra isolada</i>	assim	assim
<i>sequência de mais de uma palavra</i>	"a" "pensar"	a pensar
<i>duas palavras intercaladas por uma palavra qualquer</i>	"por" ".*" "tempo"	por um tempo, por muito tempo, por algum tempo, por pouco tempo, etc.
<i>duas palavras intercaladas por zero, uma, duas ou três outras quaisquer</i>	"bebeu" []* "vinho" within 3	bebeu vinho, bebeu um vinho, bebeu o meu vinho, bebeu muito vinho...
<i>palavra com grafias diferentes</i>	"a(c)?ção"	acção, ação
	"lo[iu]ça"	loiça, louça
<i>um conjunto de palavras diferentes</i>	"(enfim finalmente)"	enfim, finalmente
<i>palavras com um dado prefixo</i>	"des.*"	desrespeito, desfavorecer, desânimo, destino, desta, destaque...
<i>palavras com uma dada terminação</i>	".*ção"	canção, lição, relação, Conceição, atenção....
<i>sinais de pontuação</i>	"\!"	!
	"\?"	?

O que quer pesquisar	Exemplo do que deve digitar	Possíveis resultados
	"\,"	,
	"\--"	--
<i>travessão e aspas</i>	"(\« »)"	« , »
	"(\` ´)"	` , ´
<i>um adjetivo em todas as suas formas</i>	[lema="gracioso"]	gracioso, graciosa, graciosos, graciosas, graciosíssimo...
<i>um substantivo em todas as suas formas</i>	[lema="carro"]	carro, carros, carrinho, carrão....
<i>um verbo em todas as suas formas</i>	[lema="fazer"]	fazer, fiz, fez, fazia, fará, faça, fazendo....
<i>verbos antes de uma palavra</i>	[pos="V.*"] "silêncio"	fez-se silêncio, pediu silêncio, guardasse silêncio....
<i>adjetivos depois de uma palavra</i>	"amor" @[pos="ADJ.*"]	amor próprio, amor incompreendido, amor materno....
<i>advérbios depois de uma palavra</i>	"disse" @[pos="ADV.*"]	disse devagarinho, disse apenas, disse ainda, disse baixinho, disse lentamente...
<i>substantivos a seguir a um verbo</i>	[lema="cometer"] @[pos="N.*"]	cometerem erros, cometido suicídio, cometeu excessos....
<i>apenas a forma nominal de uma palavra</i>	[word="ajuda" & pos="N.*"]	casos de ajuda só como substantivo, não como forma do verbo ajudar

Para mais informações sobre diferentes maneiras de pesquisar, veja como funciona a sintaxe do IMS Corpus Workbench.

Para mais informações sobre as opções que envolveram a anotação gramatical do português, veja a Documentação da anotação morfossintáctica da parte portuguesa do COMPARA, assim como a informação variada sobre a anotação no projecto AC/DC.

Para mais informações sobre pesquisas que envolvem a gramática do inglês, veja o CLAWS tagset.

Para informações referentes à anotação semântica do COMPARA, no que respeita à cor, veja-se a Documentação da anotação relativa à cor no COMPARA.

Restrição de alinhamento

Ao pesquisar uma palavra ou expressão numa das línguas do corpo, a opção de restrição de alinhamento permite-lhe especificar só os resultados que contenham ou excluam as palavras ou expressões na outra língua do corpus que inserir na caixa de restrição de alinhamento. Por exemplo, se escrever *grande* na caixa de pesquisa do lado esquerdo e *big* na caixa de restrição de alinhamento, só serão mostrados os resultados contendo *grande* no lado português e *big* no lado inglês da concordância.

Para pesquisar, por outro lado, *grande* no lado português mas excluindo as traduções que incluam *big* do lado inglês, deve digitar *grande* na caixa de pesquisa do lado esquerdo e *!big*, ou seja, *big* precedido de um ponto de exclamação, na caixa de restrição de alinhamento.

Eis alguns resultados exemplificativos de *grande* alinhado com *big*:

EBDL3T2(798): Deve ser uma cidade **grande**.

That must be a **big** city.

EBDL4(1205):	A família católica, grande , calorosa e feliz.	The big , warm, happy Catholic family.
PPJS1(2767):	-- Lisboa é muito grande .	`Lisbon's too big for that;
Em contrapartida, eis alguns resultados de <i>grande</i> sem alinhamento com <i>big</i> :		
EBDL1T1(436):	Respondi que há anos não escrevia um trecho em prosa tão grande .	I said it was the longest piece of continuous prose I'd written in years.
EBDL3T1(797):	«Vê-se ali uma grande mancha negra	«There's a great dark smudge over there.
PPJSA1(23):	Fica aqui muito perto, seria um grande favor que me faziam.	It's close by and you could do me no greater favour.

É importante notar que o COMPARA não se encontra ainda alinhado palavra a palavra. O sistema limita-se, portanto, a procurar concordâncias paralelas com *grande* no lado português e com (ou sem) *big* no lado inglês. É importante lembrar que poderá haver casos de concordâncias paralelas com *grande* no lado português e *big* no lado inglês sem que uma palavra seja a tradução da outra.

Intimamente relacionada com a restrição de alinhamento está a escolha do tipo de resultado **Distribuição combinada dos resultados em português e inglês**, que permite uma visão resumida da distribuição dos resultados em ambas as línguas.

Mais opções de pesquisa

O COMPARA permite-lhe ainda pesquisar uma série de outros atributos que pode associar (ou não) a uma palavra ou expressão de pesquisa:

A opção **notas de tradução** permite encontrar as unidades de alinhamento que contêm notas de tradução, ou seja, que o tradutor achou necessário comentar por qualquer motivo.

A opção **títulos** procura os casos de títulos de livros, jornais, filmes, peças de teatro, canções, programas de televisão (etc.) citados nos textos do corpus (sejam eles verdadeiros ou fictícios), desde que o autor, tradutor ou editor os tenha assinalado com tipografia diferente (normalmente, itálico, mas também o pode ter feito em maiúsculas). Note que esta opção serve apenas para pesquisar os títulos *citados* nos textos incluídos no COMPARA; se estiver à procura dos títulos das obras que fazem parte do corpus, procure antes na descrição do Conteúdo do COMPARA. Se estiver interessada nos casos de títulos de secções das obras, deve usar o comando <chaptitle> []* </chaptitle>.

A opção **palavras ou expressões estrangeiras** obtém as palavras e expressões numa língua diferente da língua principal do texto que o autor, tradutor ou editor salientou tipograficamente (normalmente, em itálico).

A opção **entidades mencionadas** recupera os nomes próprios utilizados para designar marcas, estabelecimentos comerciais, hotéis, empresas, produtos, etc. que estejam tipograficamente salientes na edição impressa.

Finalmente, a opção **palavras ou expressões com ênfase** permite encontrar as palavras ou expressões que o autor ou o tradutor tiver salientado tipograficamente e que não cabem em nenhuma das categorias anteriores.

Devido à maneira como os textos do COMPARA foram alinhados, também pode pesquisar as frases que foram unidas, separadas, adicionadas, omitidas e reordenadas na tradução. Mais informação sobre os critérios de separação de frases adoptados encontra-se disponível na página relativa à atomização no projecto AC/DC.

Escolher o tipo de resultados

Os resultados do COMPARA são por omissão apresentados em forma de concordâncias paralelas, mas existem muitos outros tipos de resultados que fazem sentido.

Concordâncias

As concordâncias paralelas mostram a sua palavra ou expressão de pesquisa a negrito no lado esquerdo do ecrã, com o texto correspondente (na outra língua) no lado direito.

Não conte encontrar os originais sempre do lado esquerdo e as traduções sempre do lado direito. As colunas de resultados servem apenas para separar o português do inglês, podendo haver originais e traduções em ambas. Se quiser diferenciar originais de traduções, terá de escolher cada uma das partes (original e traduzido) do COMPARA separadamente.

Para ver de onde vem cada concordância, clique no código para obter a referência bibliográfica completa do par de textos em questão. O código, associado ao número da unidade de alinhamento, permite referir univocamente a ocorrência em questão.

Nas concordâncias, a expressão de pesquisa aparece inserida no contexto de uma unidade de alinhamento. No COMPARA, isso equivale a uma frase completa do texto original e o texto correspondente na tradução (que poderá não ser uma frase inteira, pois os tradutores nem sempre mantêm a separação frásica do original).

Se escolher a opção **mostrar propriedades de alinhamento**, poderá ver se as frases do original foram mantidas (1:1), divididas em duas (1:2), omitidas (1:0), anexadas a outras (1:1/2), e assim por diante.

Se procurou explicitamente por notas de tradução, a opção **esconder notas de tradução** serve para omitir o próprio texto das notas. Por omissão, ao procurar pelas notas esse texto é também mostrado. Se não procurou pelas notas de tradução, estas não são mostradas nas concordâncias.

Distribuição combinada das expressões de pesquisa em português e inglês (*em versões anteriores do COMPARA, esta distribuição era designada por resumo quantitativo*)

Esta opção diz respeito unicamente a pesquisas que contêm restrições de alinhamento. Por exemplo, poderá querer ver todas as vezes que a palavra *yes* aparece na mesma unidade de alinhamento que a palavra *sim*. A distribuição combinada diz-lhe quantas vezes foi encontrada a palavra *yes*, quantas vezes foi encontrada a palavra *sim* no outro lado do corpo, e quantas vezes as ocorrências de *yes* coincidem com ocorrências de *sim*.

Se a sua procura tiver um elemento negativo (na restrição de alinhamento), por exemplo *no* e não *não*, ou seja *!não*, esta opção indica quantos casos ocorrem com *no* em inglês, e em quantos casos não há correspondência entre os dois.

A cada unidade do COMPARA encontram-se associados vários campos relacionados com a sua análise sintáctica, que não só podem ser interrogados separadamente ou em conjunto, como as suas distribuições obtidas:

- **Distribuição das formas** ([word="anti.*"])
- **Distribuição dos lemas** [lema=".*mente"]

Note-se que o lema de uma palavra pode corresponder a várias categorias gramaticais, por exemplo a distribuição do lema *fino* inclui *fina*, *fino*, *finas*, *finos*, *fininho* e *finíssimo* quando se refere ao adjectivo, e apenas *fino*, *finos* quando se refere ao substantivo. Por essa razão aconselha-se procuras mais especificadas, tal como [lema="fino" & pos="ADJ.*"]

- **Distribuição da categoria gramatical** ([pos="NUM.*"])

Veja-se a informação presente no atributo pos para o elenco dos valores possíveis.

- **Distribuição do tempo verbal, do caso ou do grau** ([temcagr="INF.*"])

Veja-se a informação morfológica presente no atributo temcagr para o elenco dos valores possíveis. Repare-se que a informação neste campo é diferente conforme a categoria gramatical: tempo verbal se for um verbo, caso se for um pronome, e grau se for um adjectivo ou substantivo.

- **Distribuição da pessoa ou do número** ([pessnum="S"])

Veja-se a informação morfológica presente no atributo `persnum` para o elenco dos valores possíveis. Mais uma vez, conforme estivermos em presença de um verbo ou de um pronome pessoal, por um lado, ou de um substantivo ou adjectivo, por outro, este atributo contém pessoa ou número (singular/plural), respectivamente.

- **Distribuição do género** ([`gen="M.*"`])
- **Distribuição da função sintáctica**

Veja-se a informação sintáctica presente no atributo `func` para mais informação. Até agora, esta informação não se encontra revista, e em muitos casos tem apenas o valor "mudei" que corresponde à mudança radical da categoria gramatical.

Distribuição da semântica

([`sema="cor.*"`])

No campo "sema" foram incluídos até agora apenas informação associada a cor e campos relacionados, como cor:raça ou não maduro.

Distribuição do grupo de cor

([`cor="Branco"`] ou [`colour="White"`])

Todas as palavras marcadas como cor no COMPARA foram além disso subcategorizadas de acordo com o grupo a que pertenciam. O nome desse atributo é diferente conforme a língua seja português ou inglês.

Distribuição das obras

Esta opção serve para ver quantas vezes a sua expressão de pesquisa aparece em cada texto do corpus.

Distribuição em texto original e texto traduzido

Esta opção permite ver quantas vezes a sua expressão de pesquisa aparece nos originais e quantas vezes aparece nas traduções. Poderá ser útil para investigar diferenças entre a língua dos originais e a língua das traduções.

Distribuição em termos da variante do português

Esta opção permite ver com que frequência a sua expressão de pesquisa aparece em diferentes variantes do português.

Distribuição em termos da variante do inglês

Esta opção permite ver com que frequência a sua expressão de pesquisa aparece em diferentes variantes do inglês.

Distribuição por autor

Esta opção serve para examinar quantas vezes um determinado autor utilizou uma palavra ou expressão.

Seleccionar partes específicas do COMPARA

Embora possa restringir a sua procura a partes do COMPARA; note que nesse caso a base textual sobre a qual pesquisa é mais reduzida, em alguns casos significativamente.

Especificar variantes do português e do inglês

O COMPARA comporta várias variantes do português e do inglês. Caso queira trabalhar apenas com determinadas variantes, é aqui que as pode seleccionar. Teoricamente, qualquer combinação é possível. Por exemplo, pode seleccionar todas as variantes do português e apenas inglês britânico, ou só português do Brasil e inglês americano. No entanto, algumas combinações de variantes não se encontram disponíveis. Se escolher inglês sul-africano traduzido para português de Angola, não obterá nenhum resultado, porque, de momento, não há nenhum par de textos que satisfaça estes critérios.

Especificar datas de publicação

O COMPARA contém tanto textos contemporâneos como obras mais antigas. Se na sua pesquisa quiser utilizar apenas as publicações recentes, ou apenas as que forem anteriores a uma determinada época, poderá seleccionar textos consoante o ano de publicação dos

mesmos. A escolha é feita com base no ano da primeira edição de cada obra, ainda que o excerto incluído no COMPARA possa ter sido digitalizado a partir de uma edição posterior.

Diferenciar originais de traduções

O COMPARA separa o português do inglês, mas não é automática a diferença entre o português original e o português das traduções e entre o inglês original e o inglês das traduções. Se na sua pesquisa é importante separar a língua dos originais da língua utilizada nas traduções, terá de seleccionar uma das opções abaixo.

- **Pesquisar só de originais para traduções** : Quando a direcção de pesquisa seleccionada no topo do formulário for de português para inglês, esta opção faz com que as pesquisas mostrem apenas de originais em língua portuguesa alinhados com as suas traduções inglesas. Em contrapartida, quando a direcção de pesquisa seleccionada for de inglês para português, as pesquisas mostrarão apenas originais ingleses com as suas traduções para português.
- **Pesquisar só de traduções para originais** : Quando a direcção de pesquisa for de português para inglês, esta opção permite ver como as traduções para português se relacionam de volta com os seus respectivos textos de partida em inglês. Em contrapartida, ao pesquisar de inglês para português, esta opção mostra traduções inglesas alinhadas com os seus respectivos textos de origem em português.

Pesquisar autores específicos

Esta funcionalidade é útil para pesquisar apenas nos textos de autores específicos. Poderá utilizá-la para ver como os autores representados no COMPARA utilizam certas palavras e expressões. Se seleccionar Eça de Queirós, por exemplo, a sua pesquisa irá incidir apenas sobre os textos de Eça de Queirós presentes no corpus.

Pesquisar apenas em textos específicos

Esta opção serve para seleccionar individualmente os textos que deseja pesquisar, podendo, desta forma, criar um sub-corpus à sua medida. Por exemplo, poderá escolher apenas os textos de um autor ou tradutor em particular. A selecção de textos específicos desactiva todas as outras opções de restrição do corpus. Nesse caso, por exemplo, se escolher um texto da autora sul-africana Nadine Gordimer, não poderá remover do corpus a variante sul-africana da língua inglesa.

Cada par de textos possui um código com ligação à página de Informações bibliográficas dos textos do COMPARA, onde poderá obter a referência bibliográfica completa do original e da tradução, bem como informações sobre a extensão e a variante de cada texto. Os textos cuja identificação começa pela letra P representam textos escritos originalmente em português, enquanto que os começados pela letra E identificam os textos cujos originais são em inglês. Normalmente utilizamos excertos de 30% do começo, meio ou fim de uma obra.

Agradecimento

Este texto foi produzido no âmbito da Linguateca, contrato número 339/1.3/C/NAC, financiado pelo governo português e pela União Europeia. O COMPARA é um projecto liderado por mim e pela Ana Frankenberg-Garcia, tendo como restantes membros da equipa Rosário Silva, Susana Inácio e Pedro Martins Sousa. Agradeço à Cláudia Zavaglia pelo convite para leccionar na Segunda EBraLC e a todos os alunos da mesma pelas dúvidas e sugestões durante o curso.