

# COMPARA – Aula Prática em Português

Ana Frankenberg-Garcia

[última actualização a 26 de Junho de 2008]

<b>1. INTRODUÇÃO RÁPIDA AO CORPUS .....</b>	<b>2</b>
<b>Escolha a sua língua de trabalho* .....</b>	<b>2</b>
<b>Comece a utilizar o COMPARA.....</b>	<b>2</b>
<b>Pesquisas que resultam.....</b>	<b>3</b>
<i>Linguagem das pesquisas .....</i>	<i>3</i>
<i>Pesquisas realistas .....</i>	<i>4</i>
<b>De onde vêm os resultados?.....</b>	<b>5</b>
<b>Pesquisas interessantes .....</b>	<b>7</b>
<b>2. PESQUISA AVANÇADA .....</b>	<b>7</b>
<b>Passo 1 – Selecção da direcção linguística .....</b>	<b>7</b>
<b>Passo 2 – Tipo de pesquisa pretendido .....</b>	<b>8</b>
<i>Restrições de alinhamento .....</i>	<i>8</i>
<i>Notas de tradução .....</i>	<i>9</i>
<i>Títulos.....</i>	<i>9</i>
<i>Palavras estrangeiras .....</i>	<i>9</i>
<i>Ênfase dentro de frases.....</i>	<i>10</i>
<i>Entidades mencionadas .....</i>	<i>10</i>
<i>Alterações na divisão das frases .....</i>	<i>10</i>
<b>Passo 3 – Utilização de apenas uma parte do corpus.....</b>	<b>10</b>
3.1 <i>Selecção de textos segundo as variantes das línguas.....</i>	<i>10</i>
3.2 <i>Selecção de textos segundo as datas de publicação.....</i>	<i>11</i>
3.3 <i>Separação de originais e traduções .....</i>	<i>12</i>
3.4 <i>Selecção de textos específicos .....</i>	<i>13</i>
3.5 <i>Pesquisar em autores específicos .....</i>	<i>13</i>
<b>Passo 4 – Escolha os resultados .....</b>	<b>13</b>
<i>Propriedades de alinhamento .....</i>	<i>14</i>
<i>Esconder notas de tradução .....</i>	<i>14</i>
<i>Mostrar categoria gramatical.....</i>	<i>14</i>
<i>Distribuição das formas .....</i>	<i>15</i>
<i>Distribuição da categoria gramatical .....</i>	<i>15</i>
<i>Distribuição dos lemas .....</i>	<i>15</i>
<i>Distribuição do tempo verbal, pessoa, número e género .....</i>	<i>15</i>
<i>Distribuição de fontes .....</i>	<i>15</i>
<i>Distribuição em texto original e traduzido .....</i>	<i>16</i>
<i>Distribuição em termos de variante do português .....</i>	<i>16</i>
<i>Distribuição em termos de variante do inglês .....</i>	<i>16</i>
<i>Distribuição combinada das expressões pesquisadas em inglês e português.....</i>	<i>16</i>
<i>Distribuição por autor .....</i>	<i>17</i>
<i>Distribuição por cor .....</i>	<i>17</i>
<i>Distribuição semântica .....</i>	<i>17</i>
<b>3. CONCLUSÃO.....</b>	<b>18</b>

## 1. INTRODUÇÃO RÁPIDA AO CORPUS

### Escolha a sua língua de trabalho\*

A interface do COMPARA na internet está disponível em português e em inglês. Vá a <http://www.linguateca.pt/COMPARA/> e escolha a língua com que se sente mais à vontade. Se as especificações do seu *browser* o trouxerem para a interface inglesa e quiser mudar para português, clique em [Esta página em português](#), no canto superior esquerdo.

\*Esta aula prática refere-se apenas ao serviço português. Se preferir utilizar a interface inglesa, siga o [Tutorial](#) em inglês disponível em [Ajuda](#).

### Comece a utilizar o COMPARA

Clique em [Pesquisa simples](#). Escreva *mar* na caixa de pesquisa de português para inglês e clique no botão **Pesquisar (de português para inglês)**.

Os resultados obtidos são concordâncias paralelas. A palavra *mar* aparece a negrito no lado esquerdo do ecrã, inserida num bloco de texto em português. Os equivalentes em inglês aparecem no lado direito. Um lado da concordância é tradução do outro. Para saber qual dos lados corresponde ao texto original e qual corresponde à tradução, veja o código alfanumérico ao lado da concordância. Os códigos começados por P indicam que o original está em português, aqueles que começam por E indicam que o original está em inglês.

→ Quantas ocorrências de *mar* existem?<sup>1</sup>

Percorra os resultados para ver os equivalentes de *mar* em inglês. Como pode verificar, apesar de *mar* ser frequentemente equivalente a *sea*, também encontra palavras como *seas*, *seabed*, *maritime* e *ocean* na parte inglesa da concordância. Também há concordâncias onde não aparece qualquer equivalência. Surpreendente? Nem tanto. O COMPARA tem como base a tradução humana (e não traduções automáticas), em que nem sempre se verifica uma tradução palavra por palavra.

→ Em que língua está o texto original das primeiras três concordâncias?<sup>2</sup>

→ Em que língua está o texto original das últimas três concordâncias?<sup>3</sup>

→ Os textos das concordâncias são todos do mesmo tamanho?<sup>4</sup>

No COMPARA, uma concordância paralela corresponde sempre a uma frase ortográfica completa do texto original e à sua correspondente tradução, seja ela qual for (lembre-se que os tradutores poderão não preservar a estrutura da frase do original). Os critérios utilizados na separação e alinhamento de frases no COMPARA encontram-se detalhados em [Construção do COMPARA](#), a partir de [Documentação específica](#).

Retroceda até à [Pesquisa simples](#), utilizando o botão “Retroceder” do *browser* ou a ligação para [Voltar ao formulário de pesquisa](#) no canto superior direito. Mude o separador do formulário de modo a pesquisar na direcção inversa, de inglês para português, escreva *mar* na caixa de pesquisa correspondente e clique no botão **Pesquisar (de inglês para português)**.

<sup>1</sup> Existem 221 ocorrências de *mar* na parte portuguesa do COMPARA . Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>2</sup> Inglês (o código alfanumérico das primeiras três concordâncias começa por E). A parte portuguesa destas concordâncias são portanto traduções.

<sup>3</sup> Português (o código alfanumérico das últimas três concordâncias começa por P). A parte inglesa destas concordâncias são portanto traduções.

<sup>4</sup> Não. Algumas concordâncias são muito longas, e algumas são muito curtas.

- Quantas ocorrências de *mar* existem?<sup>5</sup>
- Porque é que existem tão poucas?<sup>6</sup>
- Se a sua pesquisa tem como ponto de partida uma palavra portuguesa, qual dos separadores do formulário de pesquisa deve utilizar?<sup>7</sup>
- Se a sua pesquisa tem como ponto de partida uma palavra inglesa, qual dos separadores do formulário de pesquisa deve utilizar?<sup>8</sup>

Retroceda até à [Pesquisa simples](#), escreva *that* na caixa de pesquisa do separador de inglês para português e clique no botão **Pesquisar (de inglês para português)**.

- Quantas ocorrências de *that* existem no COMPARA?<sup>9</sup>
- Quantas concordâncias de *that* foram mostradas?<sup>10</sup>

### **Pesquisas que resultam**

#### *Linguagem das pesquisas*

Retroceda até à [Pesquisa simples](#) e, utilizando o separador de português para inglês, pesquise *MAR* (com letras maiúsculas) depois *Mar* (só com o *M* maiúsculo).

- Quantas ocorrências de *MAR* encontrou?<sup>11</sup>
- Quais as diferenças entre os resultados obtidos para *Mar* e os resultados para *mar*?<sup>12</sup>

Por norma, o COMPARA distingue sempre entre letras maiúsculas e minúsculas. Se não quiser que se faça essa distinção, seleccione a opção “Não fazer distinção entre maiúsculas e minúsculas”. Assim poderá pesquisar de uma vez só *mar*, *Mar* e *MAR*.

Retroceda até à [Pesquisa simples](#) e clique em [Ajuda](#) para obter mais informações sobre a linguagem de pesquisa do COMPARA. Experimente pesquisar as formas abaixo numa única procura:

- *deus* e *Deus*<sup>13</sup>
- *sit up*<sup>14</sup>

Quando a sua pesquisa envolve uma sequência de palavras, separe cada palavra com aspas: “desta” “forma”. Se a sua pesquisa tiver uma única palavra, as aspas são opcionais.

<sup>5</sup> Só existem 2 ocorrências de *mar* na parte inglesa do COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>6</sup> Porque a parte de baixo do formulário serve para pesquisas que têm como ponto de partida o lado inglês do corpus e o verbo inglês *mar* é muito mais raro do que o nome português *mar*.

<sup>7</sup> Utilize o primeiro separador (1. De português para inglês) para procuras que partem do português.

<sup>8</sup> Utilize o segundo separador (2. De inglês para português) para procuras que partem do inglês.

<sup>9</sup> Existem 17954 ocorrências de *that* na parte inglesa do COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>10</sup> Apesar de existirem 17954 ocorrências, só 1000 concordâncias foram mostradas. As palavras muito comuns geram um número de concordâncias muito elevado. Para proteger os direitos de autor dos textos que fazem parte do corpus, só pode ser exibida uma amostra limitada e aleatória de concordâncias.

<sup>11</sup> Não existem ocorrências de *MAR* na parte portuguesa do COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, pode ser que já exista alguma.

<sup>12</sup> Uma consulta com *mar* retorna nomes comuns, ao passo que uma consulta com *Mar* retorna nomes próprios (ou o nome comum *mar* no início de uma frase).

<sup>13</sup> Escreva “*deus*” na caixa de pesquisa de português para inglês e seleccione a opção “Não fazer distinção entre maiúsculas e minúsculas” ou, sem seleccionar esta opção, escreva “**deus**” %c.

<sup>14</sup> Escreva “*sit*” “*up*” na caixa de pesquisa de inglês para português.

- *aberto* e *aberta*<sup>15</sup>
- *book* e *books*<sup>16</sup>
- *perfeito*, *perfeita*, *perfeitos* e *perfeitas*<sup>17</sup>
- *write*, *writes*, *wrote*, *writing* e *written*<sup>18</sup>
- palavras inglesas começadas por *un*<sup>19</sup>
- palavras inglesas acabadas em *ness*<sup>20</sup>
- *has been* e *have been* com a possibilidade de intercalar uma palavra qualquer entre o auxiliar e o particípio passado<sup>21</sup>
- *emoção* sem utilizar cedilha ou til<sup>22</sup>

A maior parte das pesquisas lexicais no COMPARA pode ser feita com apenas isto. Mas se quiser mais detalhes sobre a linguagem de pesquisa do COMPARA, veja mais informações sobre o nosso software de processamento de corpora, o IMS Corpus Workbench, em:  
<http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPUserManual/HTML/>  
<http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPTutorial/html/>

### *Pesquisas realistas*

Utilize a [Pesquisa simples](#) para pesquisar *At last he had come to his senses* (lembre-se que tem de colocar aspas em torno de cada uma das palavras).

- Quantas ocorrências de *At last he had come to his senses* encontrou?<sup>23</sup>

Provavelmente não obteve qualquer resultado para a expressão anterior porque há uma grande probabilidade de a frase não se encontrar em nenhum dos textos do corpus. Utilizar frases completas ou longas sequências de palavras no COMPARA (ou em qualquer outro corpus) não resulta. Para obter bons resultados com *corpora*, deve trabalhar com sequências de palavras realistas, com elevada probabilidade de constarem nos textos que compõem o corpus.

Pesquise agora as seguintes formas e veja quantas ocorrências existem para cada uma delas:

- "come" "to" "(my|your|his|her|our)" "senses"<sup>24</sup>
- "to" "(my|your|his|her|our)" "senses"<sup>25</sup>
- "(my|your|his|her|our)" "senses"<sup>26</sup>

<sup>15</sup> Escreva "**abert[oa]**" na caixa de pesquisa de português para inglês.

<sup>16</sup> Escreva "**book(s)?**" na caixa de pesquisa de inglês para português.

<sup>17</sup> Escreva "**perfeit[oa](s)?**" na caixa de pesquisa de português para inglês.

<sup>18</sup> Escreva "**(write|writes|writing|written|wrote)**" na caixa de pesquisa de inglês para português. Note que as flexões do verbo devem ser separadas com barras verticais e não oblíquas. No teclado português, deverá usar *shift* + a tecla do canto superior esquerdo; no teclado brasileiro, a barra vertical fica no canto inferior esquerdo, ao lado da tecla *shift*.

<sup>19</sup> Escreva "**un.\***" na caixa de pesquisa de inglês para português.

<sup>20</sup> Escreva "**.\*ness**" na caixa de pesquisa de inglês para português.

<sup>21</sup> Escreva "**(has|have) [l]\* "been" within 3**" na caixa de pesquisa de inglês para português; para intercalar com duas palavras, escreva **within 4**.

<sup>22</sup> Escreva "**emocao**" na caixa de pesquisa de português para inglês e seleccione a opção "Juntar acentos e cedilhas automaticamente" ou, sem seleccionar esta opção, escreva "**emocao" %d**".

<sup>23</sup> Não existem ocorrências de *At last he had come to his senses* no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, pode ser que já exista alguma.

<sup>24</sup> Existem 2 ocorrências de "come" "to" "(my|your|his|her|our)" "senses" no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>25</sup> Existem 8 ocorrências de "to" "(my|your|his|her|our)" "senses" no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

→ "senses"<sup>27</sup>

Caso não obtenha quaisquer ocorrências quando pesquisa uma sequência muito longa de palavras, tente pesquisar expressões mais curtas ou substituir certas palavras por ".\*", a expressão que equivale a qualquer palavra.

Utilize agora a [Pesquisa simples](#) para pesquisar *greenhouse gas emissions* (lembre-se de pôr aspas em torno de cada palavra).

→ Quantas ocorrências encontrou para este termo?<sup>28</sup>

Neste momento, o COMPARA só contém textos literários, e é pouco provável que encontre muitos termos técnicos neste género de textos. Por esta razão, o COMPARA funciona melhor com pesquisas que têm a ver com linguagem geral e linguagem literária.

→ Forme dois grupos utilizando os elementos que se encontram em baixo, um com as palavras e termos prováveis, e outro com as palavras e termos pouco prováveis de se encontrar no COMPARA<sup>29</sup>:

What	if	concordances	mecânica dos fluidos
ele(s)	said	heart	heart valve replacement
quando	não	quotient	dia

### De onde vêm os resultados?

Como já sabe, por agora o corpus só contém textos literários. Mas quais são as variantes de português e de inglês desses textos? Quem são os autores? Será que pode ser incluída no corpus uma tradução qualquer? E se a tradução não for boa?

Ao lado de cada concordância existe um código alfanumérico. Se passar o cursor por cima do código, poderá ver quem é o autor e o tradutor da concordância e os títulos das obras em questão. Se clicar no código, obterá a referência bibliográfica dos textos em causa.

Faça uma [Pesquisa simples](#) à sua escolha e percorra a página de resultados até encontrar uma concordância que lhe interesse. Passe o cursor por cima do código alfanumérico que lhe está associado para ver o nome do autor e do tradutor e os títulos das obras em causa. Depois clique no código para obter a referência completa desses textos na página de [Informações bilbiográficas](#).

Pode também aceder à página [Informações bilbiográficas](#) clicando directamente na hiperligação em causa sob a alínea [Textos do COMPARA](#) no menu. Aí encontrará informações sobre todos os textos disponíveis no corpus.

Utilize a informação que se encontra em [Informações bilbiográficas](#) e responda:

→ Quantos textos do autor Julian Barnes existem no corpus?<sup>30</sup>

<sup>26</sup> Existem 23 ocorrências de "(my|your|his|her|our)" "senses" no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>27</sup> Existem 52 ocorrências de *senses* no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>28</sup> Não existem ocorrências de *greenhouse gas emissions* no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, pode ser que já haja alguma.

<sup>29</sup> Palavras e termos prováveis no COMPARA: *What, if, ele(s), said, heart, quando, não, dia*; palavras e termos pouco prováveis: *concordances, mecânica dos fluidos, heart valve replacement, quotient*.

→ Os textos são obras integrais ou são só extractos?<sup>31</sup>

Devido às restrições de direitos de autor, os textos do corpus são geralmente extractos de aproximadamente 30% da obra completa. Estes extractos são seleccionados aleatoriamente do começo, do meio ou do fim da obra.

→ Quantas traduções de *Therapy* de David Lodge existem no corpus?<sup>32</sup>

→ Qual é o código alfanumérico que representa o par de textos com a tradução em português do Brasil de *Therapy* de David Lodge?<sup>33</sup>

Alguns textos no COMPARA estão alinhados com mais do que uma tradução. Sempre que isto acontece, a primeira tradução a ser processada é identificada com T1 no fim do código alfanumérico, a segunda tradução é identificada com T2, e assim sucessivamente.

No menu do lado esquerdo, clique no [Resumo quantitativo](#) dos [Textos do COMPARA](#). Aqui encontra um resumo das informações sobre os textos que formam a base do COMPARA.

→ Quantos textos originais e quantas traduções diferentes existem no corpus?<sup>34</sup>

→ Quantas palavras em inglês e em português existem no corpus?<sup>35</sup>

À data de elaboração desta aula, o COMPARA é o maior corpus paralelo editado do mundo.

→ Quais são as variantes de português e de inglês que aparecem no corpus?<sup>36</sup>

Todas as variantes de português e de inglês são aceites. Apesar de não haver preferência por determinadas variantes, algumas estão mais bem representadas do que outras.

→ Qual a data do texto mais antigo e dos textos mais recentes do corpus?<sup>37</sup>

Podem ser incluídos no corpus textos publicados em qualquer data, mas é necessário obter uma autorização especial para incluir textos com direitos reservados, de autores e tradutores vivos ou que morreram há menos de setenta anos.

→ É possível incluir no corpus um qualquer texto original ou uma qualquer tradução?<sup>38</sup>

<sup>30</sup> Existem 3 textos de Julian Barnes no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>31</sup> Extractos.

<sup>32</sup> Existem duas traduções de *Therapy* de David Lodge no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>33</sup> EBDL1T2.

<sup>34</sup> O COMPARA 10.0.3 contém 72 textos originais e 75 traduções. O número de textos pode já ter aumentado se estiver a utilizar uma versão posterior do corpus.

<sup>35</sup> O COMPARA 10.0.3 contém 1436187 palavras em português e 1544294 palavras em inglês. Estes números podem já ter aumentado se estiver a utilizar uma versão posterior do corpus.

<sup>36</sup> O COMPARA 10.0.3 contém textos em português de Angola, do Brasil, de Moçambique e de Portugal, e textos em inglês da África do Sul, do Reino Unido e dos Estados Unidos da América. Se estiver a utilizar uma versão posterior do corpus, podem já estar incluídas mais variantes.

<sup>37</sup> O texto mais antigo no COMPARA 10.0.3 é *Frankenstein*, de Mary Shelley, publicado em 1818; os textos mais recentes foram publicados em 2002. Se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado.

<sup>38</sup> Não.

Todos os originais e traduções do COMPARA têm de ser textos publicados. Não basta os textos estarem disponíveis em várias cópias. É também necessário que a obra esteja registada num catálogo internacional de identificação de livros. Apesar de isto não garantir só por si a qualidade dos originais e das traduções no COMPARA, normalmente este tipo de texto passa por um processo de revisão e edição e, por algum motivo, mereceu ser publicado. Para além disso, só se incluem no COMPARA originais em português e em inglês, e traduções directamente do inglês e do português. A interferência de outras línguas não é permitida. Também ficam de fora os textos em tradução automática.

### **Pesquisas interessantes**

Depois de já ter uma ideia de como funciona a [Pesquisa simples](#) e de saber qual a origem dos textos do COMPARA, a pergunta que agora se coloca é que tipo de consultas podem ser interessantes. A resposta a esta pergunta não é fácil, uma vez que há inúmeras hipóteses de escolha. Por exemplo:

Um estudante de línguas poderá estar interessado em utilizar o COMPARA como um dicionário bilingue contextualizado e *online*. Pense numa palavra que não saiba dizer em inglês e tente fazer uma pesquisa no COMPARA para descobrir.

Um professor de línguas poderá estar interessado em utilizar o COMPARA para ensinar questões como os falsos amigos, verbos ou preposições que não têm correspondência directa em português e em inglês, etc. Os professores podem copiar os resultados obtidos para um programa de processamento de textos e transformá-los num teste ou numa ficha de exercícios.

Um tradutor ou um lexicógrafo pode estar interessado em estudar as diferentes traduções de uma determinada palavra, quais são as traduções mais utilizadas para essa palavra, quais traduções são usadas em diferentes contextos, e assim por diante.

Um investigador na área da tradução pode querer descobrir se determinadas palavras ou expressões são utilizadas mais frequentemente na língua traduzida ou na língua original, ou se as frases das traduções são mais longas do que as frases dos textos originais, e por aí fora. Para ter acesso a publicações e comunicações orais com base no COMPARA, veja [Publicações](#) no menu de [Documentação específica](#).

Sejam quais forem as suas necessidades, quando achar que a [Pesquisa simples](#) já não é suficiente, estará na altura certa de experimentar a [Pesquisa Avançada](#).

## **2. PESQUISA AVANÇADA**

Clique na hiperligação da [Pesquisa Avançada](#). O formulário de pesquisa está dividido em quatro passos. Para navegar de um passo para outro, utilize os separadores de cada um ou, em alternativa, utilize a versão expandida do formulário, para navegar sem os separadores.

A versão expandida é semelhante à versão anterior do formulário de pesquisa avançada. Recomenda-se o seu uso para aumentar o tamanho do texto e melhorar a acessibilidade, ou para pesquisas repetitivas.

### **Passo 1 – Selecção da direcção linguística**

Quando utilizar a [Pesquisa Avançada](#), a primeira coisa que deve fazer é seleccionar a direcção linguística da pesquisa.

- Qual é a direcção linguística que deve utilizar para pesquisar "saudade(s)?"?<sup>39</sup>
- Qual é a direcção linguística que deve utilizar para pesquisar "eventually,"?<sup>40</sup>

Por defeito, a direcção linguística do COMPARA é de português para inglês. Se a sua pesquisa tem como ponto de partida uma palavra ou expressão em inglês, não se esqueça de mudar a direcção linguística para de inglês para português.

## **Passo 2 – Tipo de pesquisa pretendido**

A caixa de pesquisa da esquerda serve para escrever a expressão a pesquisar, exactamente como na [Pesquisa simples](#). Serve tanto para palavras e expressões em português como em inglês, uma vez que já determinou, no passo 1, qual a direcção linguística a utilizar.

### *Restrições de alinhamento*

Ao efectuar uma pesquisa numa das línguas do corpus, a caixa das **restrições de alinhamento** do lado direito permite que pesquise palavras e expressões nas duas línguas do corpus simultaneamente. Por exemplo, se quiser saber se existem correspondências de *eventually* no lado inglês do COMPARA e *eventualmente* no lado português, deve fazer o seguinte:

Seleccione **de inglês para português** no passo 1. Escreva "*eventually*" na caixa de pesquisa da esquerda e seleccione a opção "Não fazer distinção entre maiúsculas e minúsculas". Escreva "*eventualmente*" na caixa das restrições de alinhamento do lado direito e seleccione a opção "Não fazer distinção entre maiúsculas e minúsculas". Clique em **Procurar**.

- Quantas concordâncias com *eventually* no lado inglês e *eventualmente* no lado português encontrou?<sup>41</sup>
- Em que língua está o texto original dessas concordâncias, e quais são as conclusões que daí retira?<sup>42</sup>

Se agora quiser ver se existem concordâncias com *eventually* no lado inglês do corpus mas **sem *eventualmente*** no lado português, retroceda até à [Pesquisa Avançada](#) e faça o seguinte:

Seleccione **de inglês para português** no passo 1. Escreva "*eventually*" e "Não fazer distinção entre maiúsculas e minúsculas" na caixa de pesquisa da esquerda. Escreva *!"eventualmente"* (com um ponto de exclamação antes do termo a pesquisar) e "Não fazer distinção entre maiúsculas e minúsculas" na caixa de pesquisa da direita. Clique em **Procurar**.

- Quantas concordâncias encontrou e quais são as conclusões que daí retira?<sup>43</sup>
- Consegue listar algumas das formas de se dizer *eventually* em português?<sup>44</sup>

<sup>39</sup> De português para inglês. Utilize esta opção quando a pesquisa parte do português.

<sup>40</sup> De inglês para português. Utilize esta opção quando a pesquisa parte do inglês.

<sup>41</sup> Existem apenas 3 concordâncias com *eventually* no lado inglês e *eventualmente* no lado português do COMPARA 10.0.3. (se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado).

<sup>42</sup> O texto original está em inglês (o código alfanumérico começa por E), o que quer dizer que *eventually* foi traduzido por *eventualmente*, e não o contrário.

<sup>43</sup> Existem 82 concordâncias com *eventually* no lado inglês e **sem *eventualmente*** no lado português do COMPARA 10.0.3. (se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado). *Eventually* é equivalente a outra coisa que não seja *eventualmente* em 96% das suas ocorrências num corpus com quase três milhões de palavras. Conclui-se que *eventualmente* não é uma boa tradução da palavra *eventually*.

Quando a direcção linguística é **de inglês para português**, utilize a caixa de pesquisa da esquerda para uma expressão em inglês e a da direita para uma restrição de alinhamento em português. Se mudar a direcção linguística no passo 1 para **de português para inglês**, utilize a caixa de pesquisa da esquerda para escrever a expressão a pesquisar em português e a da direita para escrever uma restrição de alinhamento em inglês. Não é obrigatório utilizar as restrições de alinhamento. Se não está interessado em pesquisar ao mesmo tempo em português e em inglês, deixe a caixa das restrições de alinhamento em branco.

Retroceda até à [Pesquisa Avançada](#) e limpe o formulário. Clique no separador 2.2 de mais opções de pesquisa, que lhe permite pesquisar alguns elementos adicionais. Veja como:

### *Notas de tradução*

Limpe o formulário. Seleccione a caixa que diz **notas de tradução** e clique em **Procurar**.

- Quantas são e em que língua estão as notas de tradução encontradas?<sup>45</sup>
- O que é que tem de fazer para pesquisar notas de tradução na outra língua?<sup>46</sup>

### *Títulos*

Retroceda até à [Pesquisa Avançada](#), limpe o formulário e experimente pesquisar títulos no lado inglês e no lado português do corpus.

- Os títulos encontrados são os títulos das obras que constituem o corpus? Ou seja, os mesmos que se encontram listados em [Informações bibliográficas](#)?<sup>47</sup>

Na [Pesquisa Avançada](#), a opção **títulos** não fornece os títulos das obras que compõem o corpus, e sim os títulos (reais ou fictícios) de livros, jornais, revistas, filmes, peças de teatro, programas de televisão, canções (etc.) *citados* nestes textos. Lembre-se que para encontrar os títulos das obras que constituem o corpus, deve antes clicar no código alfanumérico que se encontra junto das concordâncias ou em [Informações bibliográficas](#) no menu de [Textos do COMPARA](#).

### *Palavras estrangeiras*

Retroceda até à [Pesquisa Avançada](#), limpe o formulário e experimente pesquisar palavras e expressões estrangeiras no lado português do corpus.

- Quantas palavras ou expressões estrangeiras existem no lado português do corpus? Liste quatro.<sup>48</sup>

Retroceda até à [Pesquisa Avançada](#), mude a direcção linguística e experimente pesquisar palavras e expressões estrangeiras no lado inglês do corpus.

- Quantas palavras ou expressões estrangeiras existem no lado inglês do corpus? Liste quatro.<sup>49</sup>

---

<sup>44</sup> *Acabar por, afinal de contas, por fim, finalmente, até, etc.*

<sup>45</sup> Existem 98 notas de tradução portuguesas no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado.

<sup>46</sup> Mudar a direcção linguística no passo 1.

<sup>47</sup> Não.

<sup>48</sup> O COMPARA 10.0.3 contém 2217 palavras e expressões estrangeiras no lado português do corpus. Quatro exemplos disso são: Knees R'Us, Benson and Hedges, chignon, snob. Se estiver a utilizar uma versão posterior do corpus, podem já existir mais palavras e expressões estrangeiras.

→ As palavras e expressões estrangeiras são só nomes comuns?<sup>50</sup>

Os nomes próprios que incluem nomes comuns foram considerados **palavras estrangeiras**. Por exemplo, *Bouvard et Pécuchet* não é considerado português ou inglês devido à conjunção francesa *et*. Note-se que só aparecerão nos resultados as palavras e expressões estrangeiras salientadas tipograficamente (normalmente em itálico) pelo autor ou tradutor da obra. Pode haver outras palavras ou expressões no corpus numa língua diferente da língua original do texto, mas que não aparecerão automaticamente nos resultados porque o autor ou o tradutor não as diferenciou tipograficamente.

### *Ênfase dentro de frases*

Retroceda até à [Pesquisa Avançada](#), limpe o formulário e experimente procurar palavras ou expressões com ênfase nas duas direcções linguísticas.

→ Qual é a língua do corpus que utiliza mais palavras com ênfase?<sup>51</sup>

### *Entidades mencionadas*

Retroceda até à [Pesquisa Avançada](#), limpe o formulário e pesquise entidades mencionadas nas duas direcções linguísticas.

Entidades mencionadas são palavras que identificam nomes próprios, marcas, lojas, hotéis, empresas, produtos, doutrinas, etc. Só aparecem automaticamente nos resultados do COMPARA se tiverem sido salientadas tipograficamente pelo autor ou pelo tradutor.

### *Alterações na divisão das frases*

O COMPARA também permite que veja as frases que os tradutores dividiram, uniram, adicionaram, apagaram e reordenaram na tradução. Seleccione a caixa apropriada se quiser pesquisar estas mudanças.

## **Passo 3 – Utilização de apenas uma parte do corpus**

A [Pesquisa simples](#) do COMPARA utiliza por defeito todos os textos do corpus. Mas se for à página [Textos do COMPARA](#) e vir quais são esses textos, pode querer não os utilizar todos. O Passo 3 da [Pesquisa Avançada](#) permite-lhe usar só os textos que seleccionar.

### *3.1 Selecção de textos segundo as variantes das línguas*

Vá a [Pesquisa Avançada](#), limpe o formulário e procure "it" com "Não fazer distinção entre maiúsculas e minúsculas" na direcção de inglês para português.

→ Quantas ocorrências de "it" sem distinção entre maiúsculas e minúsculas existem?<sup>52</sup>

Agora, retroceda até à [Pesquisa Avançada](#), **não** limpe o formulário e seleccione a caixa que diz inglês da África do Sul.

→ Quantas ocorrências de "it" sem distinção entre maiúsculas e minúsculas existem agora?<sup>53</sup>

<sup>49</sup> O COMPARA 10.0.3 contém 1181 palavras e expressões estrangeiras no lado inglês do corpus. Quatro exemplos disso são: Genson, cojones, n'est ce pas, paraphema. Se estiver a utilizar uma versão posterior do corpus, podem já existir mais palavras e expressões estrangeiras.

<sup>50</sup> Não.

<sup>51</sup> No COMPARA 10.0.3, existem mais palavras e expressões com ênfase em inglês. Esta informação pode já ter mudado se estiver a utilizar uma versão posterior do corpus.

<sup>52</sup> Existem 16487 ocorrências de *it* e *It* em todos os textos em inglês do COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

Faça o mesmo separadamente para inglês do Reino Unido e dos Estados Unidos.

→ Quantas ocorrências encontrou?<sup>54</sup>

Os resultados obtidos para as diferentes variantes de inglês não indicam que se utiliza mais frequentemente a palavra *it* em certas variantes de inglês do que noutras. Em primeiro lugar, não pode tirar conclusões sobre o inglês da África do Sul, do Reino Unido e dos Estados Unidos tendo só como base os textos do COMPARA, que apenas representam uma quantidade relativamente pequena de textos literários nestas variantes. Em segundo lugar, os resultados não são proporcionais, porque a parte sul-africana do corpus é muito menor do que as partes do Reino Unido e dos Estados Unidos (ver [Textos do COMPARA](#) e [Resumo quantitativo](#)).

Repita as mesmas pesquisas e ao mesmo tempo preste atenção ao topo da página de resultados que diz **Descrição do corpus usado nesta procura** para descobrir quantas palavras inglesas da África do Sul, do Reino Unido e dos Estados Unidos foram usadas nas suas consultas.

→ Utilize estas informações para completar o quadro que se encontra em baixo<sup>55</sup>:

Variante do inglês	Ocorrências <i>it/It</i>	Número total de palavras no sub-corpus	% de <i>it/It</i>
Estados Unidos			
Reino Unido			
África do Sul			

→ Repita agora a pesquisa de *it* e *It*, restringindo o corpus ao inglês da África do Sul e ao português de Angola. Quantas ocorrências existem?<sup>56</sup>

Algumas combinações de variantes de língua, como o inglês da África do Sul e o português de Angola, não estão disponíveis no COMPARA. Para saber que combinações de variantes existem no corpus, consulte a página [Resumo quantitativo](#).

### 3.2 Selecção de textos segundo as datas de publicação

Pode também limitar as suas pesquisas a textos que foram publicados antes ou depois de determinadas datas.

<sup>53</sup> Existem 901 ocorrências de *it* e *It* nos textos em inglês da África do Sul do COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>54</sup> Existem 9647 ocorrências de *it* e *It* nos textos em inglês do Reino Unido e 5939 ocorrências nos textos em inglês dos Estados Unidos do COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, este número pode já ter aumentado.

<sup>55</sup> O quadro abaixo foi preenchido com valores retirados do COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado:

Variante do inglês	Ocorrências <i>it/It</i>	Número total de palavras	% de <i>it/It</i>
Estados Unidos	5939	673046	0,9
Reino Unido	9647	1021729	0,9
África do Sul	901	101962	0,9

<sup>56</sup> Não há nenhuma ocorrência, porque não existem originais em inglês da África do Sul traduzidos para português de Angola nem originais em português de Angola traduzidos para inglês da África do Sul no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, isto pode ter mudado.

- Limpe o formulário e pesquise a palavra *computador* em originais publicados antes de 1975. Quantas ocorrências há?<sup>57</sup>
- Repita a pesquisa com originais publicados depois de 1975. Quantas ocorrências há?<sup>58</sup>
- Limpe o formulário e faça uma pesquisa em branco, ou seja, deixe a caixa de pesquisa vazia, seleccionando a opção de textos traduzidos publicados antes de 1900. De que textos provêm as concordâncias?<sup>59</sup>

Na versão 10.0.3 do COMPARA, o texto mais antigo data de 1818 e os mais recentes foram publicados em 2002.

### *3.3 Separação de originais e traduções*

- Limpe o formulário da [Pesquisa Avançada](#) e pesquise *already* (sem distinção entre maiúsculas e minúsculas) na direcção linguística de inglês para português. Quantas ocorrências existem?<sup>60</sup>
- O código alfanumérico das concordâncias obtidas começa sempre com a mesma letra?<sup>61</sup>
- Agora, retroceda até à [Pesquisa Avançada](#), **não** limpe o formulário, seleccione, no passo 3.3, a caixa que diz que as pesquisas devem ser conduzidas apenas de originais para traduções e clique em **Procurar**. Quantas ocorrências existem?<sup>62</sup>
- O código alfanumérico das concordâncias que obteve como resultado começa sempre com a mesma letra?<sup>63</sup>
- Retroceda até à [Pesquisa Avançada](#), **não** limpe o formulário, mas desta vez, no passo 3.3, não seleccione a caixa que diz que as pesquisas devem apenas ser de originais para traduções e seleccione a outra caixa, a que diz que as pesquisas devem apenas ser de traduções para originais. Clique em **Procurar**. Quantas ocorrências existem?<sup>64</sup>
- O código alfanumérico das concordâncias que obteve como resultado começa sempre com a mesma letra?<sup>65</sup>
- A palavra *already* é mais frequente em inglês original ou em inglês traduzido?<sup>66</sup>

---

<sup>57</sup> Só existe uma ocorrência da palavra *computador* em originais publicados antes de 1975. Trata-se de um texto de 1974, da autora sul-africana Nadine Gordimer. Se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado.

<sup>58</sup> Na versão 10.0.3 do COMPARA, *computador* figura 40 vezes nos textos publicados após 1975.

<sup>59</sup> Todas as concordâncias vêm do mesmo texto, pois só existe uma tradução publicada antes de 1900 no COMPARA 10.0.3: *Iracema* de José de Alencar, traduzido para inglês por Lady Isabel Burton e publicado em 1886. Se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado.

<sup>60</sup> Existem 916 ocorrências de *already* no COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado.

<sup>61</sup> Não. Os códigos das primeiras concordâncias que aparecem nos resultados começam com a letra E, o que quer dizer que os originais são ingleses e que o que está em português são traduções. Os códigos das últimas concordâncias que aparecem como resultado começam por P, o que quer dizer que os originais estão em português e que o que aparece em inglês são traduções.

<sup>62</sup> Existem 332 ocorrências de *already* nos originais em inglês do COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado.

<sup>63</sup> Sim. Todas as concordâncias começam por E, o que quer dizer que só foram utilizados textos originais em inglês nesta pesquisa. O lado português, portanto, contém apenas traduções.

<sup>64</sup> Existem 584 ocorrência de *already* nos textos traduzidos para inglês do COMPARA 10.0.3. Se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado.

<sup>65</sup> Sim. Todas as concordâncias começam por P, o que quer dizer que só foram utilizados textos em inglês traduzido do português nesta pesquisa.

<sup>66</sup> No COMPARA 10.0.3, *already* aparece 332 vezes em 819660 palavras de inglês original, e 584 vezes em 724220 palavras de inglês traduzido do português (pode consultar o número total de palavras em inglês original e traduzido na página [Resumo quantitativo](#), ou onde diz *Descrição do corpus usado nesta procura* no início da página de resultados). *Already* é utilizado duas vezes mais em inglês traduzido!

A maioria dos investigadores na área da tradução concorda que a linguagem das traduções é diferente daquela usada nos originais. A opção que permite separar originais de traduções é importante sempre que quiser identificar e comprovar diferenças existentes entre o português original e o português das traduções, ou o inglês original e o inglês das traduções. Esta opção também é relevante se utilizar o COMPARA para ensinar línguas, quando pretende proteger o aprendente de certas características da linguagem da tradução, ou quando, pelo contrário, quer mostrar ao aprendente a tradução de determinadas palavras ou expressões.

### 3.4 Selecção de textos específicos

Cada par de textos, original e tradução, no COMPARA está representado por um código alfanumérico único. Se passar o cursor por cima do código, poderá ver o original e tradução correspondente. Os códigos também estão ligados à página [Informações bibliográficas](#), onde pode obter referências completas e informações acerca da extensão dos textos e das variantes da língua utilizadas. Na secção 3.4 da [Pesquisa Avançada](#), deixe-se orientar por estes códigos para seleccionar os pares de textos que quer utilizar nas pesquisas, sejam eles quais forem.

→ Na secção 3.4 da [Pesquisa Avançada](#) clique no código alfanumérico EBDL1T1, retroceda, e depois clique no EBDL1T2. Que textos são estes?<sup>67</sup>

Limpe o formulário da [Pesquisa Avançada](#), seleccione as caixas EBDL1T1 e EBDL1T2, e faça uma pesquisa para *suit* na direcção de inglês para português.

→ Quais são as traduções de *suit* apresentadas na tradução para português europeu e na tradução para português do Brasil?<sup>68</sup>

→ Repare nos números entre parênteses que se encontram debaixo dos códigos que identificam os textos. Tem alguma ideia por que razão os números 555, 689 e 1339 estão repetidos?<sup>69</sup>

Os números entre parênteses debaixo dos códigos alfanuméricos servem para dar uma identidade única a cada frase dos textos originais do corpus. A palavra *suit* aparece nas frases número 555, 689 e 1339 do extracto do livro *Therapy* de David Lodge usado no COMPARA.

### 3.5 Pesquisar em autores específicos

O passo 3.5 permite-lhe seleccionar autores específicos e pesquisar somente dentro dos textos destes autores.

→ Utilizando o passo 3.5, pesquise *black* apenas em textos de Nadine Gordimer e depois apenas em textos de Henry James. Que autor usa mais esta palavra?<sup>70</sup>

### Passo 4 – Escolha os resultados

Os resultados de todas as pesquisas que efectuou até ao momento, tanto na [Pesquisa simples](#) como na [Pesquisa Avançada](#), foram apresentados sob a forma de concordâncias paralelas, que é o formato padrão de apresentação de resultados do COMPARA. No entanto, o formulário da [Pesquisa Avançada](#) também permite que obtenha outros tipos de resultados.

<sup>67</sup> *Therapy* de David Lodge traduzido para português europeu (EBDL1T1) e para português do Brasil (EBDL1T2).

<sup>68</sup> Em português europeu, *suit* foi traduzido como *fato* e *é bom*; em português do Brasil, *suit* foi traduzido como *agasalho*, *terno* e *não seria mau*.

<sup>69</sup> Os números estão repetidos porque as frases do texto original estão repetidas.

<sup>70</sup> Nadine Gordimer.

### *Propriedades de alinhamento*

Se clicar na caixa que diz *mostrar propriedades de alinhamento*, vai conseguir ver se a frase do texto original na qual cada concordância se baseia foi dividida, unida, apagada, adicionada ou reordenada na tradução.

Limpe o formulário. No passo 2.2, experimente algumas pesquisas com adição, supressão e junção de frases na tradução, ao mesmo tempo que, no passo 3.3, selecciona **conduzir pesquisas apenas de originais para traduções** e, no passo 4, selecciona **mostrar propriedades de alinhamento**. Clique em *procurar* e observe os números que aparecem a seguir ao número entre parênteses que identifica cada frase.

→ O que acha que querem dizer 1-0, 1-1/2, 1-2, e assim sucessivamente?<sup>71</sup>

### *Esconder notas de tradução*

Retroceda até à [Pesquisa Avançada](#). Limpe o formulário. No passo 1, seleccione a direcção **de inglês para português**. No passo 2.2, seleccione **notas de tradução**. Clique em *enter* para ver as notas de tradução dos tradutores ingleses. Retroceda até ao formulário da [Pesquisa Avançada](#), não o limpe e seleccione **esconder notas de tradução** no passo 4. Clique em *enter*.

→ O que acontece?<sup>72</sup>

### *Mostrar categoria gramatical*

Volte para a [Pesquisa Avançada](#), limpe o formulário, seleccione a direcção português para inglês, escreva *gosto* no passo 2 e, no passo 4, seleccione **concordâncias e mostrar categoria grammatical**. Faça *enter*.

Em vez do texto normal das concordâncias, verá que aparecem barras oblíquas, letras e códigos a seguir a cada palavra. Os códigos e letras são etiquetas que identificam a categoria grammatical de cada palavra.

→ Que códigos aparecem depois de *gosto*?<sup>73</sup>

Todas as pesquisas desta aula são puramente lexicais, ou seja, não contêm informações grammaticais. Há pouco tempo introduzimos anotação grammatical no COMPARA, para que os utilizadores possam usar informação grammatical nas suas pesquisas. Por exemplo, com anotação grammatical passa a ser possível recuperar concordâncias para a palavra *gosto* como nome separadamente das concordâncias em que *gosto* é um verbo. A anotação grammatical do português foi feita automaticamente com o analisador PALAVRAS e o inglês foi anotado com o CLAWS. De momento, estamos a rever ambos de modo a aumentar a precisão dos resultados.

Embora a anotação grammatical automática ainda não esteja completamente revista, já é possível conduzir pesquisas grammaticais no COMPARA, desde que se leve em conta uma pequena margem de erro nos resultados. Estamos a elaborar um manual separado para ensinar a fazer pesquisas com informação grammatical. Por enquanto, pode ir aprendendo como se fazem pesquisas grammaticais através dos exemplos listados na ajuda a pesquisar. Os últimos exemplos da tabela de expressões de pesquisa mostram pesquisas grammaticais. Para mais informações, veja [Anotação grammatical](#) no menu de [Documentação específica](#).

<sup>71</sup> 1-0 identifica as frases do texto original que não foram traduzidas. 1-1/2 identifica as frases do texto original que são equivalentes a meias frases na tradução. 1-2 identifica as frases do texto original que foram divididas em duas frases separadas na tradução, etc.

<sup>72</sup> Deve conseguir ver onde é que as notas de tradução foram inseridas sem ter de ver as próprias notas.

<sup>73</sup> Aparecem **gosto/N**, **gosto/V\_fmc** e **gosto/V**. O primeiro identifica *gosto* como sendo um nome ou substantivo; os dois segundos identificam *gosto* como sendo tipos de verbos.

### *Distribuição das formas*

Retroceda até à [Pesquisa Avançada](#). Limpe o formulário. No passo 1, seleccione a direcção **de inglês para português**. No passo 2, escreva ".\*ly" na caixa de pesquisa. Salte o passo 3. No passo 4, retire a selecção de **concordâncias** e seleccione **distribuição de formas**. Clique em **Pesquisar**.

→ O que acontece?<sup>74</sup>

### *Distribuição da categoria gramatical*

Volte para a [Pesquisa Avançada](#), limpe o formulário, seleccione a direcção português para inglês, escreva *gosto* (sem fazer distinção entre maiúsculas e minúsculas) no passo 2 e, no passo 4, seleccione distribuição da categoria gramatical. Clique em **Pesquisar**.

→ A palavra *gosto* aparece mais frequentemente como verbo ou como nome?<sup>75</sup>

### *Distribuição dos lemas*

Volte para a [Pesquisa Avançada](#), limpe o formulário, seleccione a direcção português para inglês, escreva *caminho* no passo 2 e, no passo 4, seleccione **distribuição dos lemas**.

→ Quantas vezes aparece a palavra *caminho* associada à primeira pessoa do singular do verbo *caminhar* e quantas vezes aparece a palavra como um nome sinónimo de trilho/estrada/rumo/etc.?<sup>76</sup>

### *Distribuição do tempo verbal, pessoa, número e género*

Estas distribuições só se encontram disponíveis para português.

Ao pesquisar [lema="trabalhar"] e seleccionar uma **distribuição de tempo verbal**, irá ver que tempos verbais deste verbo estão representados no corpus. Que tempo verbal é mais frequente para o verbo *trabalhar*?<sup>77</sup>

Ao pesquisar [lema="sugerir"] e seleccionar uma **distribuição de pessoa e número**, obterá um resumo das flexões de pessoa e número deste verbo no corpus. Qual é a mais frequente?<sup>78</sup>

Ao pesquisar [lema="querido"] e seleccionar uma **distribuição de género**, poderá ver se esta palavra é mais frequente no feminino ou no masculino.<sup>79</sup>

### *Distribuição de fontes*

Retroceda até à [Pesquisa Avançada](#). Limpe o formulário. No passo 1, seleccione a direcção **de inglês para português**. No passo 2, escreva *love* na caixa de pesquisa. Salte o passo 3. No passo 4, não seleccione a caixa das concordâncias e seleccione a caixa que diz **distribuição de fontes**. Clique em *enter*.

---

<sup>74</sup> Aparecem as palavras terminadas em *ly* na parte inglesa do corpus. A mais frequente no COMPARA 10.0.3 é *only*, com 2424 ocorrências. Este número pode já ter aumentado se estiver a utilizar uma versão posterior do corpus.

<sup>75</sup> No COMPARA 10.0.3, há 152 ocorrências de *gosto* como nome e 157 (109+48) ocorrências de *gosto* como verbo. Se estiver a utilizar uma versão posterior do corpus, esta informação pode já ter mudado.

<sup>76</sup> No COMPARA 10.0.3, há apenas sete ocorrências de *caminho* como conjugação do verbo *caminhar*. Nas outras 495 vezes, a palavra é um nome. Estes números podem já ter aumentado se estiver a utilizar uma versão posterior do corpus

<sup>77</sup> No COMPARA 10.0.3, o tempo verbal mais frequente para *trabalhar* é o infinitivo.

<sup>78</sup> No COMPARA 10.0.3, *sugerir* aparece mais vezes na terceira pessoa do singular.

<sup>79</sup> No COMPARA 10.0.3, *querido* aparece mais vezes na forma masculina.

→ Qual é o texto no corpus que menciona mais frequentemente a palavra *love*?<sup>80</sup>

#### *Distribuição em texto original e traduzido*

Retroceda até à [Pesquisa Avançada](#). Limpe o formulário. No passo 1, seleccione a direcção **de inglês para português**. No passo 2, escreva *love* na caixa de pesquisa. Salte o passo 3. No passo 4, **não** seleccione a caixa das concordâncias e seleccione a caixa que diz **distribuição em texto original e traduzido**. Clique em *enter*.

→ A palavra *love* aparece com maior frequência em inglês original ou em inglês traduzido?<sup>81</sup>

#### *Distribuição em termos de variante do português*

Retroceda até à [Pesquisa Avançada](#). Limpe o formulário. No passo 1, seleccione a direcção **de português para inglês**. No passo 2, escreva *fato* na caixa de pesquisa. Salte o passo 3. No passo 4, seleccione a caixa das concordâncias e a caixa que diz **distribuição em termos de variante do português**. Clique em *enter*.

→ Em que variante do português a palavra *fato* aparece com maior frequência? Analise as concordâncias e veja se consegue encontrar uma explicação para tal.<sup>82</sup>

#### *Distribuição em termos de variante do inglês*

Retroceda até à [Pesquisa Avançada](#). Limpe o formulário. No passo 1, seleccione a direcção **de inglês para português**. No passo 2, escreva *colour* na caixa de pesquisa. Salte o passo 3. No passo 4, **não** seleccione a caixa das concordâncias e seleccione a caixa que diz **distribuição em termos de variante do inglês**. Clique em *enter*.

→ Em que variante do inglês a palavra *colour* aparece com maior frequência?<sup>83</sup>

#### *Distribuição combinada das expressões pesquisadas em inglês e português*

Retroceda até à [Pesquisa Avançada](#). Limpe o formulário. No passo 1, seleccione a direcção **de inglês para português**. No passo 2, escreva *actually* na caixa de pesquisa e *a(c)tualmente* na caixa de restrições de alinhamento. Em ambos os casos, seleccione “não fazer distinção entre maiúsculas e minúsculas”. Salte o passo 3. No passo 4, **não** seleccione a caixa das concordâncias e seleccione a caixa que diz **distribuição combinada das expressões pesquisadas em inglês e português**. Clique em *enter*.

---

<sup>80</sup> No PBMA5, a palavra *love* é referida 38 vezes em 22203 palavras, ou seja, 17 vezes em cada 10 mil palavras. Compare com EUEP1, onde a palavra *love* só aparece uma vez em 24411 palavras, ou seja, só 0,4 vezes em cada 10 mil palavras.

<sup>81</sup> No COMPARA 10.0.3, a palavra *love* é referida 451 vezes em 724220 palavras inglesas que foram traduzidas do português, ou seja, 62 vezes em cada cem mil palavras, e 296 vezes em 819660 palavras de textos originais ingleses, ou seja, 36 vezes em cada cem mil palavras. Parece que há mais amor em inglês traduzido do português! Se estiver a utilizar uma versão posterior do corpus, estes números podem já ter mudado.

<sup>82</sup> Proporcionalmente, a palavra *fato* é muito mais frequente em português do Brasil (na versão 10.0.3 do COMPARA, são 68 ocorrências em 557712 palavras, ou seja, 12 vezes em cada 100 mil palavras). As concordâncias mostram que *fato* tem significados diferentes em português do Brasil e em português de Portugal e de Moçambique, o que pode repercutir numa maior frequência da palavra em português do Brasil.

<sup>83</sup> No COMPARA 10.0.3, *colour* é proporcionalmente mais frequente em inglês sul-africano (15 vezes em 100 mil palavras), seguido de inglês britânico (10 vezes em 100 mil palavras) e, em inglês americano, só aparece 0,6 vezes em cada 100 mil palavras. Se repetir a pesquisa com *color*, o inglês americano passa a predominar.

→ Quantas ocorrências existem para *actually* e *Actually*; quantas ocorrências existem para *actualmente*, *atualmente*, *Actualmente* ou *Atualmente*? Quantas vezes é que as duas coincidem numa única concordância?<sup>84</sup>

Se olhar para a **Descrição do corpus usado nesta pesquisa**, também poderá concluir que *actually* aparece 15 vezes em cada 100 mil palavras inglesas, enquanto *actualmente* só aparece 2 vezes em cada 100 mil palavras portuguesas. A palavra *actually* é proporcionalmente mais de sete vezes mais frequente do que *actualmente*!

#### *Distribuição por autor*

Retroceda ao formulário de [Pesquisa Avançada](#). Limpe o formulário. Selecione a direcção **de inglês para português** no passo 1. No passo 2, escreva *gun*. Salte o passo 3. No passo 4, deixe a opção concordâncias vazia e peça uma **distribuição por autor**. Faça o mesmo para a palavra *killer*.

→ Que autor no COMPARA escreve mais sobre revólveres e assassinos?<sup>85</sup>

#### *Distribuição por cor*

Retroceda ao formulário de [Pesquisa Avançada](#). Limpe o formulário. Selecione a direcção **de inglês para português** no passo 1. No passo 2, escreva `[sem="colour.*"]` na caixa de pesquisa. No passo 3.5, seleccione Kazuo Ishiguro. No passo 4, deixe a opção concordâncias vazia e peça uma **distribuição por grupo da cor**.

→ Que cor o autor Kazuo Ishiguro utiliza mais?<sup>86</sup>

Para procurar palavras classificadas com o atributo cor em inglês, use `[sem="colour.*"]`. Para português, utilize a seguinte expressão na caixa de pesquisa: `[sem="cor.*"]`

#### *Distribuição semântica*

Retroceda ao formulário de [Pesquisa Avançada](#). Limpe o formulário. Selecione a direcção **de português para inglês** no passo 1. No passo 2, escreva `[sem="cor.*"]` na caixa de pesquisa. Ignore o passo 3. No passo 4, peça concordâncias e uma **distribuição semântica**.

→ Quantas palavras relacionadas com cor existem no COMPARA?<sup>87</sup>

→ Verifique que tipo de palavras estão anotadas com o atributo "cor".<sup>88</sup>

Por ora, a única categoria semântica disponível para pesquisa no COMPARA é a cor. Se estiver interessado em utilizar o COMPARA para estudar algum outro aspecto da semântica, por favor entre em contacto connosco.

<sup>84</sup> No COMPARA 10.0.3, existem 232 concordâncias para *actually* e *Actually* no lado inglês do corpus, e 31 concordâncias para *actualmente*, *atualmente*, *Actualmente* ou *Atualmente* no lado português do corpus. No entanto, só numa concordância é que os termos coincidiram em inglês e em português, o que indica que não são bons equivalentes linguísticos. Se estiver a utilizar uma versão posterior do corpus, estes números podem já ter mudado.

<sup>85</sup> No COMPARA 10.0.3, Patrícia Mello utiliza as palavras *gun* e *killer* com maior frequência.

<sup>86</sup> No COMPARA 10.0.3, Kazuo Ishiguro usa principalmente a cor *black*.

<sup>87</sup> No COMPARA 10.0.3, há 3436 palavras relacionadas com cor na parte portuguesa do corpus.

<sup>88</sup> Alguns exemplos: *branco*, *acastanhada*, *vermelhos*, *azul-pálido*, *vermelho-framboesa*, *colorida*, etc.

### **3. CONCLUSÃO**

Agora que terminou a nossa aula prática, tente fazer as suas próprias pesquisas e avise-nos se tiver questões ou problemas que ainda não foram tratados na nossa página de [Ajuda](#) ou de [Dúvidas dos utilizadores](#). Quaisquer sugestões que tenha também são bem-vindas. Entre em contacto connosco através do formulário disponível em [Enviar questões, comentários e sugestões](#).

Em <http://www.linguateca.pt/COMPARA/ComparaBibliografia.html>, temos ainda uma página cujo objectivo é fornecer uma bibliografia com descrições do COMPARA e estudos baseados no COMPARA. Se usar o COMPARA em trabalhos publicados ou apresentados em conferências, agradecemos que nos indique, para que os mesmos possam ser referenciados nesta página.

*O COMPARA é (parcialmente) financiado pela Fundação para a Ciência e Tecnologia, co-financiada pelo POSI, através do projecto POSI/PLP/43931/2001, e pelo POSC através do projecto POSC 339/1.3/C/NAC, ambos executados pela FCCN.*