

Questões de alinhamento e anotação semântica em um *corpus* de traduções inglês ↔ português: o CorTrad

Stella E. O. Tagnin, Jamilly Brandão Alvino, Camila Barcaro Corrêa*

1. Introdução

Corpora paralelos são fonte inestimável para estudos contrastivos e de tradução (Johansson & Hofland, 1993; Santos, 1996). No entanto, são bastante escassos devido à dificuldade de se encontrar textos para os compor. Um dos mais conhecidos é o Europarl¹, constituído de transcrições das sessões do parlamento europeu. Contempla 21 línguas, dentre elas o português, na variante europeia. Um *corpus* dedicado exclusivamente ao par português-inglês é o COMPARA (Frankenberg-Garcia & Santos, 2002), que contempla extratos de textos literários em três variantes da língua inglesa (americana, britânica e sul-africana) e quatro da língua portuguesa (angolana, brasileira, europeia e moçambicana), permitindo especificar buscas em quaisquer das variantes. Em 2008 foi criado o CorTrad, um dos corpora que compõem o projeto CoMET – *Corpus* Multilíngue para Ensino e Tradução² (Tagnin, 2003; Tagnin, 2008; Teixeira, Santos, & Tagnin, 2012) numa parceria entre o CoMET, a Linguateca³ e o NILC⁴. Enquanto o CoMET se encarrega da coleta e preparação dos textos, a Linguateca é responsável por toda a parte computacional, ou seja, a disponibilização dos corpora on-line e seu elaborado sistema de buscas. Inicialmente o *corpus* estava hospedado no NILC, mas atualmente está na Linguateca, onde também pode ser acessado. O CorTrad é

* Stella E. O. Tagnin é livre-docente aposentada da USP, mas ainda atuante na pós-graduação; Jamilly Brandão Alvino é graduanda em Letras pela USP e foi bolsista do Projeto CoMET até julho de 2016; Camila Barcaro Corrêa é graduanda em Letras pela USP e é bolsista do Projeto CoMET desde setembro de 2016.

¹ <http://www.statmt.org/europarl/>

² <http://www.comet.usp.br>

³ <http://www.linguateca.pt/>

⁴ <http://www.nilc.icmc.usp.br/>

composto de três *corpora* com textos em inglês e português brasileiro⁵: um corpus jornalístico com 180 textos originalmente escritos em português da Revista FAPESP, num total de 19 números dos anos 2001, 2002 e 2003; um corpus técnico-científico composto, por ora, por apenas um livro de culinária também escrito originalmente em português, e um corpus literário com textos completos em inglês, traduzidos para o português. Esse último é constituído de duas partes, uma denominada Traduções e Revisões, com textos que apresentam várias etapas de uma tradução, ou seja, a primeira, uma versão corrigida e a versão publicada. A outra parte, intitulada Múltiplas Traduções, contém traduções de uma mesma obra realizadas por tradutores distintos.

A parte com traduções e revisões constitui-se de 20 contos canadenses e 28 australianos, todos traduzidos por alunos do extinto Curso de Especialização em Tradução da USP e posteriormente publicados (Tagnin, 2002; 2005). A parte com múltiplas traduções é composta pelos livros *Alice in Wonderland* (2011) e *Through the Looking-Glass and What Alice Found There* (1939), ambos de Lewis Carroll, e *Dubliners* (1914 (1996)) de James Joyce.

Como se depreende do acima exposto, além de contemplar especificamente a variante brasileira da língua portuguesa, o CorTrad caracteriza-se por apresentar, no geral, várias traduções do mesmo texto, sejam elas versões preliminares que permitem acompanhar a evolução de uma tradução, sejam traduções por profissionais distintos que ensejam comparações quanto às estratégias tradutórias empregadas, aos equivalentes encontrados, assim como ao estilo dos tradutores. Recentemente Santos e Simões (2015) demonstraram sua relevância para o ensino tanto de tradução quanto de língua por meio da ferramenta ParaEnsinador. Outro diferencial é seu refinado sistema de busca, que permite pesquisas específicas de acordo com cada *corpus*, respeitando, assim, suas particularidades. Finalmente, os textos do *corpus* são etiquetados morfossintaticamente, em inglês pelo CLAWS (Rayson & Garside, 1998) e em português pelo PALAVRAS (Bick, 2000). Os textos em português recebem etiquetas semânticas para cor, vestuário, corpo e sentimentos (Silva

⁵ Cogita-se também incluir versões dos textos em português europeu.

& Santos, 2012; Freitas et al., 2015; Santos et al., 2011; Mota & Santos, 2015), porém os em inglês só tem anotação para cor.

Todo esse material pode ser disponibilizado para pesquisa graças ao DISPARA (Santos, 2002), “uma infraestrutura tecnológica que permite a um usuário com conexão à Internet em qualquer lugar do mundo consultar um *corpus* paralelo”. (Teixeira, Santos, & Tagnin, 2012)

Os textos do CorTrad são alinhados automaticamente pelo programa *EasyAlign*⁶, mas necessitam de uma revisão humana. É dessa revisão, em particular dos textos que compõem a parte das Múltiplas Traduções, que este artigo trata. Na próxima seção oferecemos uma breve apresentação das obras que compõem essa parte e informações sobre como o material chegou às nossas mãos. Na seção 3 descrevemos as várias etapas do processo de alinhamento desse material, bem como os problemas enfrentados e a forma como foram resolvidos. A seção 4 ocupa-se da revisão da anotação semântica e discute, em especial, um impasse que surgiu em relação à concepção cultural de ‘cor’. O artigo encerra-se com as Considerações finais.

2. As obras e a sua disponibilização no CorTrad

O processo de disponibilização de textos implementado no Projeto CoMET foi baseado nos manuais inicialmente estabelecidos para o COMPARA (Frankenberg-Garcia, 2004; 2006a; Frankenberg-Garcia, Santos, & Silva, 2006b) e obedece à seguinte sequência de etapas:

1. Escolher o material que será alinhado
2. Escanear e/ou digitar os textos
3. Revisar os textos e salvar em formato .txt
4. Alinhamento automático pelo programa *EasyAlign*⁷
5. Revisão humana do alinhamento automático:
 - i. Numeração das Unidades de Alinhamento
 - ii. Eventual inserção de etiquetas
 - iii. Leitura final
6. Disponibilização no site do projeto CoMET

⁶ Alinhador do antigo IMS Corpus Workbench, hoje denominado Open CWB. Disponível em: <http://cwb.sourceforge.net/>

⁷ Alinhador integrado ao IMS Corpus Workbench. Disponível em: <http://cwb.sourceforge.net/>

Nesta seção trataremos apenas dos três primeiros itens.

2.1. As obras de Lewis Carroll

A obra *Alice's Adventures in Wonderland* (Carroll, 2011) foi inspirada em uma história oral contada por Lewis Carroll a pedido de Alice, uma das três filhas de Henry Lidell, o então reitor do *Christ Church College*, local onde o autor foi aluno e, posteriormente, professor de matemática. Carroll, então, escreveu a história que havia contado e presenteou Alice com o manuscrito que foi publicado em 1865, após ser ilustrado por John Tenniel.

Composta por 12 capítulos, a obra tornou-se um clássico da literatura mundial, com traduções em mais de 174 idiomas⁸ e inúmeras releituras e adaptações cinematográficas, televisivas e teatrais.

Seguiu-se *Through the Looking-Glass and What Alice Found There* (Carroll, 1939), publicada em 1872, e também composta por 12 capítulos. As obras de Lewis Carroll foram um marco para a literatura infantojuvenil, sendo o autor um dos precursores da literatura *nonsense*. Ambas as obras estão em domínio público e disponíveis, já em formato .txt, no site do Projeto Gutenberg⁹.

As duas obras e suas traduções (Alice no País das Maravilhas, 2000; Alice no País das Maravilhas, 2002; Alice no Reino do Espelho, 1962; Alice no País do Espelho, 1973) nos foram gentilmente cedidas pelo Prof. Guilherme Fromm, da Universidade Federal de Uberlândia e são resultado de trabalhos de algumas de suas alunas. Embora já tenha vindo em formato .txt, o material necessitou de cuidadosa revisão, pois havia trechos que faltavam e alguns erros ortográficos, ambos ocasionados por falha no reconhecimento dos caracteres nos processos de conversão para o formato de texto simples. A revisão, portanto, foi realizada com a comparação entre as versões convertidas para o formato .txt e os arquivos digitalizados, com a exceção de *Alice no*

⁸<https://www.oakknoll.com/pages/books/120410/jon-a-lindseth-general-technical-alan-tannenbaum/alice-in-a-world-of-wonderlands-the-translations-of-lewis-carrolls-masterpiece>

⁹ Projeto Gutenberg: www.gutenberg.org

Reino do Espelho (1962) e *Alice no País do Espelho* (1973), cujas versões em formato .txt foram comparadas com os livros físicos.

2.2. Dubliners, de James Joyce

Joyce começou a escrever, entre 1903 e 1904, o livro que ele chamou de “um capítulo da história moral de meu país”. A obra é constituída por 15 contos que retratam o cotidiano dos habitantes da Irlanda, em especial, de Dublin. Originalmente, o livro abarcava 12 doze contos, e ao longo de 10 anos, enquanto James Joyce tentava, incansavelmente, publicar *Dubliners*, mais 3 contos foram incorporados à obra.

A publicação de *Dubliners* foi recusada inúmeras vezes por diversas editoras por ser um livro de caráter crítico e abarcar conteúdo considerado subversivo para a época, como o conto *Two Gallants* e as referências depreciativas a Eduardo VII – Rei do Reino Unido então.

Em 1914, Joyce aceitou fazer certas mudanças em seu livro, e conseguiu publicá-lo em um período muito próximo ao início da Primeira Guerra Mundial. Por esse motivo a obra não recebeu atenção nem dos críticos, nem do público. Desde então, entretanto, *Dubliners* já conquistou seu lugar como um dos escritos representativos de James Joyce.

Os textos que compõem o *corpus* de *Dubliners*¹⁰ são o texto original, publicado pela primeira vez em 1914 (Joyce, 1914 (1996)) e duas traduções, uma por José Roberto O’Shea (Joyce, 1994) e outra por Hamilton Trevisan (Joyce, 1997). O material recebido já estava previamente formatado em .txt e alinhado pelo *Aligner* do WordSmith Tools, mas assim mesmo foi submetido ao alinhamento pelo EasyAlign.

3. O processo de alinhamento

O alinhamento dos textos é o processo em que as frases das duas traduções são relacionadas numericamente às frases do arquivo original em inglês. Dessa forma, uma busca no CorTrad retornará as três orações alinhadas lado a lado, facilitando o trabalho de comparação e análise do pesquisador.

¹⁰ Esse *corpus* foi gentilmente cedido por Lourdes Bernardes Gonçalves, que o compilou para sua tese de doutorado intitulada *Dubliners sob a lupa da Linguística de Corpus: uma contribuição para a análise e a avaliação da tradução literária* (2006).

Conforme vimos na seção 2 acima, o processo de alinhamento envolve as seguintes etapas:

1. Escolher o material que será alinhado
2. Escanear e/ou digitar os textos
3. Revisar os textos e salvar em formato *.txt*
4. Alinhamento automático pelo programa *EasyAlign*¹¹
5. Revisão humana do alinhamento automático:
 - i. Numeração das Unidades de Alinhamento
 - ii. Eventual inserção de etiquetas
 - iii. Leitura final
6. Disponibilização no *site* do projeto CoMET

As três primeiras já foram tratadas na seção 2. Passaremos agora a abordar as outras três.

Após a escolha, coleta e revisão, os textos devem ser salvos em formato *.txt* para que o programa *EasyAlign* os reconheça e efetue o alinhamento automático (etapas 1, 2, 3 e 4). Ressalta-se que a revisão do material é de extrema importância, pois podem ocorrer erros de digitalização, bem como falta de trechos ou duplicação dos mesmos, o que pode ocasionar futuros erros de alinhamento.

Todo o processo de alinhamento é fundamentado na divisão frásica do material coletado e, segundo Frankenberg-Garcia, Santos e Silva (2006b):

Entende-se por frase uma palavra ou sequência de palavras iniciada por letra maiúscula e terminada em ponto final, reticências, ponto de exclamação ou ponto de interrogação, seguida de uma nova sequência de palavras iniciada por letra maiúscula, ou sem seguimento nenhum, no caso de fim de parágrafo.

¹¹ Alinhador integrado ao IMS Corpus Workbench. Disponível em: <http://cwb.sourceforge.net/>

Dessa forma, o alinhamento automático etiqueta os textos por frases – também denominadas sentenças, donde o símbolo <s> –, porém a revisão humana (etapa 5) é necessária já que o alinhamento automático é um processo sujeito a erros, como se vê na Figura 1 abaixo:

```
And here Alice began to get rather sleepy, and went on saying to
herself, in a dreamy sort of way, « Do cats eat bats? </s>
Do cats eat bats? </s>
» andsometimes « Do bats eat cats? </s> A
» for, you see, as she couldn't answer either question, it didn't
much matter which way she put it . </s>
She felt that she was dozing off, and had just begun to dream that
she was walking hand in hand with Dinah, and was saying to her,
very earnestly, « Now, Dinah, tell me the truth: </s>
did you ever eat a bat? </s>
» when suddenly, thump! </s> B
thump! </s> B
down she came upon a heap of sticks and dry leaves, and the fall
was over . </s>
</p> <p par=99> . </s> C
</p> <p par=100> Alice was not a bit hurt, and she jumped up on to
her feet in a moment: </s>
she looked up, but it was all dark over-head: before her was
another long passage, and the White Rabbit was still in sight,
hurrying down it . </s>
</p> <p par=101> . </s> C
</p> <p par=102> There was not a moment to be lost: </s>
```

Figura 1 – Erros na etiquetagem de sentenças.

De acordo com a imagem, os erros assinalados são:

A – por causa dos pontos de interrogação os trechos da sentença foram considerados como sentenças completas pelo alinhador automático;

B – devido à pontuação, a onomatopeia “thump” foi apontada como sentença, ainda que o ponto de exclamação esteja seguido por uma palavra em letra minúscula;

C – o alinhador automático inseriu um ponto final no espaço entre os parágrafos e, por ser um sinal de pontuação, o considerou como sentença.

A revisão humana do alinhamento automático é composta por quatro subetapas, entre elas a numeração das unidades de alinhamento (subetapa 5, i.). De acordo com Frankenberg-Garcia, Santos & Silva (2006b), “uma unidade de alinhamento é sempre uma frase ortográfica do texto de partida mais o texto correspondente do texto traduzido, seja ele uma frase, mais do que uma frase ou mesmo apenas parte de uma frase”.

As unidades de alinhamento são indicadas pelo número à esquerda que acompanha o trecho em questão, como exemplificado na Figura 2 abaixo:

Texto original	104. <s> <p par=31> « What a curious feeling! » said Alice « I must be shutting up like a telescope! » </p> </s> 105. <s> <p par=32> And so it was indeed: she was now only ten inches high, and her face brightened up at the thought that she was now the right size for going through the little door into that lovely garden. </s>
Tradução 1	104. <s> <p par=26> « Que sensação estranha! » exclamou Alice, « devo estar encolhendo como uma luneta! » </p> </s> 105. <s> <p par=27> E de fato estava: tinha agora não mais que trinta centímetros de altura. Seu rosto iluminou-se ao pensar que atingira o tamanho certo para passar pela pequena porta que dava naquele belo jardim. </s>
Tradução 2	104. <s> <p par=27> « Que sensação estranha », disse Alice. « Eu devo estar encolhendo como um telescópio! » </p> </s> 105. <s> <p par=28> E daí era fato, ela estava agora com apenas 25 centímetros de altura, e seu rosto resplandeceu ao pensar que aquele era o tamanho exato para atravessar a portinha em direção ao adorável jardim. </s>

Figura 2 – Revisão do alinhamento: as unidades de alinhamento (subetapa i)

Quanto à inserção e revisão de etiquetas (subetapa 5, ii.), as principais são:

- <t> título </t>
- <s> sentença </s>
- <p par=n^o> parágrafo </p>
- <emph> palavra ou trecho enfatizado </emph>

Quando uma unidade de alinhamento não tem tradução, insere-se a etiqueta <s> sem qualquer complementação, o que indica que a frase é vazia. (ver Fig. 3).

```

427. <s> <p par=225> Este começou imediatamente: </p> </s>
428. <s> <p par=226> <emph> O sol brilhava sobre o mar... </emph> </p> </s>
429. <s>
430. <s>
431. <s>
432. <s>
433. <s>

```

Figura 3 – Revisão do alinhamento: as etiquetas (subetapa ii)

Além disso, o alinhador automático não adiciona etiquetas de sentença nos textos traduzidos, de modo que isso foi feito subsequentemente.

Entretanto, essa tarefa só foi realizada para as obras de Lewis Carroll. O procedimento foi alterado para as de James Joyce, dispensando as etiquetas <s> nas traduções sem afetar o alinhamento.

Em casos específicos, duas ou mais sentenças no texto original equivalem a uma sentença na tradução, ou seja, houve uma junção na tradução, por isso a etiqueta que indica sentença <s> deve conter a quantidade de sentenças do texto original que foram unidas, como em <s2>, que indica que o trecho do texto traduzido se refere a **duas** sentenças do texto original </s>. Além disso, o trecho traduzido deve ser replicado de acordo com essa quantidade para que as unidades de alinhamento sejam formadas corretamente (ver Fig. 4):

Original	365. <s> <p par=200> « What do you call yourself? » the Fawn said at last. </s> 366. <s> Such a soft sweet voice it had! </p> </s>
Tradução 1	365. <s2> <p par=190> -- Como se chama você? perguntou ele por fim numa linda voz de veado. </s> 366. <s2> -- Como se chama você? perguntou ele por fim numa linda voz de veado. </p> </s>

Figura 4 - Revisão do alinhamento: etiquetagem de sentenças replicadas na tradução (subetapa iii)

Quando há sentenças adicionadas pelo tradutor, sem equivalentes no texto original, é utilizada a etiqueta <add>, que indica adição (ver Fig. 5):

Original	1281. <s> <p par=760> « That <emph> proves </emph> his guilt, » said the Queen. </p> </s> 1282. <s> <p par=761> « It proves anything of the sort! » said Alice. </s>
Tradução 1	1281. <s> <p par=782> « Isso prova a sua culpa, é claro », disse a Rainha: <add> « Logo, cortem-lhe... » </add> </p> </s> 1282. <s> <p par=783> « Isso não prova nada! » interveio Alice. </s>

Figura 5 - Revisão do alinhamento: etiquetagem de sentenças adicionais (subetapa iv)

Cumpridas todas as etapas, o texto alinhado é disponibilizado no *site* da Linguateca e pode ser acessado gratuitamente tanto pela página da Linguateca quanto pela página do Projeto CoMET. O *site* conta com sistema de busca próprio, composto por diversas ferramentas que auxiliam o pesquisador, como ilustram as imagens abaixo (Fig. 6 e 7), que ilustram uma busca pela palavra ‘chapeleiro’ na primeira tradução:

CorTrad

Início	Jornalístico	Literário	Técnico-científico
--------	--------------	-----------	--------------------

CorTrad literário várias traduções

O CorTrad é um corpus aberto, sujeito a alterações. Veja [dados quantitativos](#) para informações atualizadas sobre o conteúdo do corpus.

O **CorTrad literário múltiplas traduções** consiste atualmente dos dois livros infanto-juvenis de Lewis Carroll traduzidos cada um por dois tradutores diferentes, e dos contos "Dubliners" de James Joyce, também traduzidos por dois tradutores distintos. Os textos de Lewis Carroll em forma digital foram gentilmente cedidos por Guilherme Fromm, da Universidade Federal de Uberlândia.

A disponibilização do CorTrad na rede é um [projeto conjunto entre o COMET e a Linguateca](#), usando o sistema [DISPARA](#).

Pesquisar no corpus

Original	<input type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver
Primeira tradução	<input checked="" type="radio"/> principal	chapeleiro	<input checked="" type="checkbox"/> ver
Segunda tradução	<input type="radio"/> principal	<input type="text"/>	<input checked="" type="checkbox"/> ver

☒ Ignorar maiúsculas/minúsculas

Procurar por tipo de alinhamento:

- ☐ 1-0
- ☐ 1-N
- ☐ N-1
- ☐ N-M

Pesquisar

Resultado	
<input checked="" type="radio"/> Concordância	<input type="radio"/> Distribuição das formas
<input type="radio"/> Distribuição dos lemas	<input type="radio"/> Distribuição da categoria gramatical (PoS)
<input type="radio"/> Distribuição do tempo verbal e/ou do caso pronominal	<input type="radio"/> Distribuição de pessoa e/ou número
<input type="radio"/> Distribuição do gênero morfológico	<input type="radio"/> Distribuição da função sintática
<input type="radio"/> Distribuição por obra	<input type="radio"/> Distribuição pelos títulos
<input type="radio"/> Distribuição por autor	<input type="radio"/> Distribuição por variante do original inglês (ou tradução)
<input type="radio"/> Distribuição por data de publicação	<input type="radio"/> Distribuição por campo semântico
<input type="radio"/> Distribuição por grupo (de cor, de vestuário, etc.)	
Opções	
<input type="checkbox"/> Resultados por ordem alfabética (só distribuições)	
<input type="checkbox"/> Resultados numerados (só concordância)	

Pesquisar

Figura 6 - Sistema de busca: CorTrad, Projeto CoMET.

CorTrad

Expressão de busca: "chapeleiro" %c
 Resultado escolhido: **concordância em contexto**
 Corpus pesquisado: **primeira tradução** (versão 1.1)

55 ocorrências.

[Voltar](#) [Nova pesquisa](#)

Original	Primeira tradução	Segunda tradução
«In <i>that</i> direction», the Cat said, waving its right paw round, «lives a Hatter: and in <i>that</i> direction», waving the other paw, «lives a March Hare. Visit either you like: both mad.»	«Naquela direção», disse o Gato, ondulando sua pata direita, " mora um Chapeleiro naquela outra ", agitando a outra pata «mora uma Lebre de Março. Visite ou um ou outro: ambos são loucos. »	«Naquela direção», o Gato disse, apontando sua pata direita em círculo, " vive o Chapeleiro, e naquela, apontando a outra pata, «vive a Lebre de Março. Visite qualquer um que você queira, os dois são malucos. »
It was so large a house, that she did not like to go nearer till she had nibbled some more of the lefthand bit of mushroom, and raised herself to about two feet high: even then she walked up towards it rather timidly, saying to herself «Suppose it should be raving mad after all! I almost wish I'd gone to see the Hatter instead!»	Era uma casa tão grande que ela não quis aproximar-se sem antes morder mais um pedacinho do cogumelo da mão esquerda, até alcançar setenta centímetros de altura. Mesmo assim, caminhou em direção à casa com muita timidez, dizendo a si mesma: «Imagine se ela estiver delirando! Eu quase chego a desejar que tivesse ido visitar o Chapeleiro »	A casa era tão grande que Alice não queria se aproximar até morder um pedaço da mão esquerda do cogumelo, e crescer para mais ou menos 70 centímetros: mesmo depois disso ela caminhou em sua direção timidamente, dizendo para si mesma: «Suponhamos que ela esteja delirante afinal! Eu quase desejo que tivesse ido ver o Chapeleiro!»

Figura 7 – Resultado de busca pela expressão “chapeleiro” em concordância em contexto, na primeira tradução

3.1. As obras de Lewis Carroll: revisão do alinhamento: considerações e soluções

O alinhamento foi realizado com três versões¹² de cada obra, a versão original em inglês e duas outras traduzidas para o português:

- *Alice in Wonderland* (1991), disponibilizada *on-line* pelo Projeto Gutenberg¹³;
- *Alice no País das Maravilhas* (2002), tradução de Clélia Regina Ramos¹⁴;
- *Alice no País das Maravilhas* (2000), tradução de Isabel de Lorenzo e Nelson Ascher¹⁵;
- *Through the Looking-Glass, and What Alice Found There* (1939)¹⁶;
- *Alice no País do Espelho* (1973), tradução de Monteiro Lobato¹⁷;
- *Alice no Reino do Espelho* (1962), tradução de Maria Thereza Giácomo¹⁸.

¹² Ressalta-se que todas as versões foram cedidas em formato de texto simples ao Projeto CoMET pelo Prof. Dr. Guilherme Fromm, da Universidade Federal de Uberlândia.

¹³ Disponível em: < <https://www.gutenberg.org/> >.

¹⁴ Disponível em: < <http://www.ebooksbrasil.org/eLibris/alicep.html> >

¹⁵ Disponível em: http://200.136.76.125/colégio/livros/download/alice_no_pais_das_maravilhas.pdf

¹⁶ Disponível em: < http://www.gas.org/refbib/Carroll_Works.pdf >.

¹⁷ Livro impresso. Ver Referências bibliográficas: Carroll 1973.

¹⁸ Livro impresso. Ver Referências bibliográficas: Carroll 1962.

Após a revisão do material digitalizado, os *corpora* foram enviados à Linguatca para o alinhamento automático. Entretanto, na etapa de revisão humana desse alinhamento sugeriram alguns impasses, dentre eles, (1) o que deveria ser considerado unidade de alinhamento e (2) como alinhar a grande quantidade de poemas que permeiam as obras de Carroll já que, até então, nenhum *corpus* com poemas havia sido incluído no *site* do Projeto CoMET.

Concluiu-se que a melhor solução seria manter os poemas em suas formas originais, respeitando as rimas e a sonoridade pretendidas pelo autor. Assim, cada estrofe foi considerada um parágrafo (<p>) e cada verso, uma sentença (<s>), como no excerto:

1. <s> <t> Alice's Adventures in Wonderland </t> </s>
2. <s> <p par=1> All in the golden afternoon </s>
3. <s> Full leisurely we glide; </s>
4. <s> For both our oars, with little skill, </s>
5. <s> By little arms are plied, </s>
6. <s> While little hands make vain pretence </s>
7. <s> Our wanderings to guide. </p> </s>
8. <s> <p par=2> Ah, cruel Three! In such an hour, </s>
9. <s> Beneath such dreamy weather, </s>
10. <s> To beg a tale of breath too weak </s>
11. <s> To stir the tiniest feather! </s>
12. <s> Yet what can one poor voice avail </s>
13. <s> Against three tongues together? </p> </s>

De acordo com os manuais (Frankenberg-Garcia, 2006; Frankenberg-Garcia, Santos, & Silva, 2006), a separação das unidades de alinhamento é indicada por pontos finais (.), pontos de interrogação (?) e pontos de exclamação (!). Contudo, em alguns trechos do texto, a divisão de sentenças de acordo com esses sinais ocasionaria certa perda de sentido, como no trecho a seguir:

«Do you know, I was so angry, Kitty,» Alice went on, as soon as they were comfortably settled again, «when I saw all the mischief you had been doing, I was very nearly opening the window, and putting you out into the snow! And you'd have deserved it, you little mischievous darling! What have you got to say for yourself? Now don't interrupt me!» she went on, holding up one finger.

Caso o trecho em questão fosse dividido de acordo com os sinais de pontuação, algumas sentenças não seriam acompanhadas de aspas e, dessa forma, o pesquisador não saberia se a sentença é uma fala da personagem ou apenas uma observação do narrador; além disso, em muitas sentenças, o enunciatário seria omitido.

Portanto, decidiu-se que cada fala completa das personagens seria considerada como uma unidade de alinhamento, sem que houvesse divisões de acordo com os sinais de pontuação, mantendo o sentido e a forma de escrita particular do autor (ver Fig. 8):

Original	Primeira tradução	Segunda tradução
«Do you know, I was so angry, Kitty.» Alice went on, as soon as they were comfortably settled again. «when I saw all the mischief you had been doing. I was very nearly opening the window, and putting you out into the snow! And you'd have deserved it, you little mischievous darling! What have you got to say for yourself? Now don't interrupt me!» she went on, holding up one finger.	Sem erguer-se da poltrona Alice começou a ralar com o maroto. -- -- Eu devia ficar zangada, sabe? Você é o tipo do atrapalhador de serviço. Que faria se eu abrisse a janela e o atirasse na neve, hein? Olhe, nem mais um -- -- <i>pio</i> -- --, está ouvindo? -- E com o dedo indicador erguido: --	-- Fiquei tão zangada com você, Pretinha, -- disse Alice após acomodar-se novamente -- que quase abri a janela para jogá-la para fora. Bem que você merecia, sua malvadinha! E não adianta inventar desculpas! Não me interrompa! -- -- E, ameaçando com o dedinho, Alice continuou a ralar: --

Figura 8 – Resultado de busca pela expressão “mischievous” em concordância em contexto, no texto original

Tal como no COMPARA, foi preciso rever cuidadosamente as digitalizações iniciais.

3.2. Dubliners: revisão do alinhamento: considerações e soluções

A exemplo dos textos de Lewis Carroll, o alinhamento do *corpus* de *Dubliners* no CorTrad foi fundamentado nos manuais de alinhamento de Frankenberg-Garcia (2004; 2006a; 2007; Frankenberg-Garcia, Santos, & Silva, 2006b) com algumas alterações, que serão aqui descritas.

O trabalho de alinhamento é feito no Microsoft Word, a fim de seguir as instruções de alinhamento por sentença já implementadas no COMPARA (Frankenberg-Garcia, Santos, & Silva, 2006).

Quando uma frase do texto original não é traduzida, deve-se alinhar a sentença em inglês com “ABSOLUTAMENTE NADA!!!” (como se vê na Figura 9 abaixo) ou com “SEM TRADUÇÃO” ou até com uma etiqueta <s> vazia.:

12. </p><p par=5> And the sub-matron and two of the Board ladies had heard the compliment. </s>	senhoras do Comitê escutaram o elogio.	12. </p><p par=5> A subchefe e duas senhoras do Comitê tinham escutado o elogio.
13. And Ginger Mooney was always saying what she wouldn't do to the dummy who had charge of the irons if it wasn't for Maria. </s>	13. E Ginger Mooney vivia dizendo que não sabia o que teria sido da surd-muda que cuidava dos ferros de engomar, se Maria não as tivesse separado.	13. E Ginger Mooney vivia dizendo que não saberia o que fazer com a muda que cuidava dos ferros de engomar se não fosse Maria.
14. Everyone was so fond of Maria. </s>	14. ABSOLUTAMENTE NADA !!!	14. Todos queriam muito bem a Maria.
		15. </p><p par=6> As mulheres tomariam o chá às seis horas da

Figura 9 – Alinhamento de frase sem tradução

Ao terminar o alinhamento, os três textos devem ser convertidos de volta para o formato de arquivo de texto.

Durante o processo de alinhamento surgiu apenas um obstáculo técnico, que exigiu adaptações do manual inicial. Alguns contos de *Dubliners* possuem canções e poemas – gêneros que podem sofrer grandes modificações quando traduzidos, tanto na parte lexical quanto na ordem de suas sentenças. Em princípio, as estrofes são consideradas parágrafos e os versos, sentenças, como foi feito nos textos de Carroll. No entanto, quando a canção ou o poema havia passado por muitas mudanças ao ser traduzido, foi necessário agrupar várias sentenças, que representam versos, em uma só, com a finalidade de conservar o sentido original e manter certa ordem no alinhamento.

120. </p><p par=28> I had riches too great to count; could boast </s> Of a high ancestral name, </s> But I also dreamt, which pleased me most, </s> That you loved me still the same. </s>	120. </p><p par=28> As riquezas que eu tinha, impossível contar </s> E de um nome ilustre podia me orgulhar </s> Mas também sonhei, o que mais me agradou, </s> Que teu amor por mim nunca mudou.	120. </p><p par=28> As minhas riquezas, não podia contar, </s> De um nome ilustre podia me gabar; </s> Mas também sonhei, e foi do que mais gostei, </s> Que tu sempre irias me amar.
121. </p><p par=29> But no one tried to show her her mistake; and when she had ended her song Joe was very much moved. </s>	121. </p><p par=29> Mas ninguém procurou mostrar-lhe o engano. Quando terminou, Joe estava muito comovido.	121. </p><p par=29> Mas ninguém se dispôs a mostrar-lhe o engano; e quando ela terminou a canção Joe estava bastante emocionado.
122. He said that there was no time like the long ago and no music for him like poor old Balfe, whatever other people might say; and his eyes filled up so much with tears that he could not find what he	122. Afirmava que para ele nada se comparava aos velhos tempos e que música alguma igualava-se à do velho e pobre Balfe, não importava o que dissessem. Seus olhos ficaram tão cheios de lágrimas, que não	122. Afirmou que para ele nada se comparava aos velhos tempos e

Figura 10 – Agrupamento de sentenças dentro de uma canção.

Na imagem acima, todas as sentenças do parágrafo 28 foram agrupadas em uma única unidade de alinhamento (a unidade 120), e assim foi criado um parágrafo com múltiplas sentenças para os três textos.

Para demonstrar a importância da revisão manual do alinhamento automático, segue abaixo a comparação entre o número de frases em cada arquivo alinhado antes e depois de serem revistos.

Texto alinhado	Número de frases arquivo original (antes da revisão)	Número de frases tradução 01 (antes da revisão)	Número de frases tradução 02 (antes da revisão)	Número de frases nos três arquivos (depois da revisão)
<i>Araby</i>	151	97	93	151
<i>Clay</i>	257	258	238	122
<i>A Little Cloud</i>	646	646	579	410
<i>Counterparts</i>	456	456	426	291
<i>The Dead</i>	1951	1951	1905	1036
<i>An Encounter</i>	314	314	303	192
<i>Eveline</i>	204	203	198	132
<i>Two Gallants</i>	496	496	429	299
<i>Grace</i>	988	988	970	639
<i>The Boarding House</i>	284	284	285	171
<i>Ivy Day in The Committee Room</i>	774	774	698	515
<i>A Mother</i>	454	454	459	298
<i>A Painful Case</i>	367	367	359	212
<i>After the Race</i>	243	242	241	143
<i>The Sisters</i>	334	334	334	204

Tabela 1: Dados quantitativos após revisão por frase

No geral, o número de frases tornou-se menor após a revisão devido à grande quantidade de etiquetas (<s> e <p>) vazias, assim como a um erro do programa de alinhamento por adicionar ao arquivo mais frases “ABSOLUTAMENTE NADA !!!” do que seria correto.

4. A revisão da anotação semântica

Neste artigo focamos apenas na revisão da anotação semântica feita no novo subcorpus de múltiplas traduções, composto das duas obras de Lewis Carroll e de *Dubliners*. Essa revisão é composta por uma única etapa nos três

textos: o texto original, a primeira tradução e a segunda tradução. Os procedimentos de busca das anotações semânticas a serem revisadas são feitos no próprio site:

CorTrad

The screenshot shows the CorTrad website interface. At the top, there is a navigation bar with four tabs: 'Início', 'Jornalístico', 'Literário', and 'Técnico-científico'. Below this, the header reads 'CorTrad literário várias traduções'. A paragraph explains that CorTrad is an open corpus subject to changes, with links to 'dados quantitativos' and 'COMET e a Linguateca'. Another paragraph describes the 'CorTrad literário múltiplas traduções' project, mentioning Lewis Carroll and James Joyce. A third paragraph states that the corpus is available through a project between COMET and Linguateca, using the DISPARA system.

The 'Pesquisar no corpus' section contains a search form. It has three radio buttons labeled 'Original', 'Primeira tradução', and 'Segunda tradução'. The 'Original' radio button is selected. To the right of the radio buttons is a text input field containing '[sema="cor"]'. Below the radio buttons are three checkboxes, all of which are checked: 'ver', 'ver', and 'ver'. A 'Pesquisar' button is located to the right of the search form.

The 'Resultado' section shows a list of filters. The 'Concordância' filter is selected. Below the filters, there is a section for 'Opções' with two checkboxes: 'Resultados por ordem alfabética (só distribuições)' and 'Resultados numerados (só concordância)'. The 'Resultados numerados (só concordância)' checkbox is checked.

Figura 11 – Procedimentos de busca de anotações semânticas.

- A. Entrar na área de traduções múltiplas do CorTrad;
- B. Optar por um texto e assinalá-lo como “principal”;
- C. Escolher qual etiqueta será utilizada (como a etiqueta utilizada no primeiro exemplo, [sema=“cor”], que seleciona somente ocorrências com palavras que expressam cores);
- D. Selecionar Resultado → Concordância;
- E. Opções → Resultados numerados (só concordância) – essa fase ajuda a estratégia visual da revisão.

Em seguida, é necessário analisar ocorrência por ocorrência para encontrar os casos em que a anotação semântica está equivocada. Na imagem a seguir encontra-se assinalado um desses casos:

	whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her. __UNDEF__	entorpecida) se o prazer de fazer uma guirlanda de margaridas valeria o esforço de levantar-se e colher as margaridas, quando de repente um coelho branco com olhos rosados passou correndo perto dela. --	fazer um colar de margaridas era mais forte do que o esforço de ter de levantar e colher as margaridas, quando subitamente um Coelho Branco com olhos cor-de-rosa passou correndo perto dela. --
3	So she was considering, in her own mind(as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her. __UNDEF__	Assim, meditava com seus botões (tanto quanto podia, porque o calor daquele dia era tal que ela se sentia sonolenta e entorpecida) se o prazer de fazer uma guirlanda de margaridas valeria o esforço de levantar-se e colher as margaridas, quando de repente um coelho branco com olhos rosados passou correndo perto dela. --	Então, ela pensava consigo mesma (tão bem quanto era possível naquele dia quente que a deixava sonolenta e estúpida) se o prazer de fazer um colar de margaridas era mais forte do que o esforço de ter de levantar e colher as margaridas, quando subitamente um Coelho Branco com olhos cor-de-rosa passou correndo perto dela. --
4	She took down a jar from one of the shelves as she passed; it was labelled « orange marmalade » but to her great disappointment it was empty: she did not like to drop the jar, for fear of killing somebody, so managed to put it into one of the cupboards as she fell past it. __UNDEF__	Enquanto passava, pegou de uma das prateleiras um pote: tinha o rótulo «-- -- geléia de laranja -- -- », mas para seu desapontamento estava vazio: não quis jogar fora o pote, com medo de acertar mortalmente alguém lá embaixo, então, esforçou-se por colocá-lo de volta em uma das prateleiras enquanto passava. --	Alice apanhou um pote de uma das prateleiras ao passar: estava etiquetado «-- -- geléia de laranja -- -- », mas para seu grande desapontamento estava vazio: ela não jogou o pote fora por medo de machucar alguém que estivesse embaixo e por isso precisou fazer algumas manobras para recolocá-lo em uma das prateleiras. --
5	Alice was not a bit hurt, and she jumped up on to her feet in a moment: she looked up, but it was all dark overhead; before her was another long passage, and the White Rabbit was still in sight, hurrying down it.	Alice não estava nem um pouco machucada, por isso levantou-se num instante: olhou para cima, mas estava tudo escuro. Diante dela havia outro longo corredor, e o Coelho Branco ainda estava à vista, correndo apressado.	Alice não estava nem um pouco machucada, e pôde saltar sobre os pés num instante: olhou para cima mas estava tudo escuro sobre sua cabeça, diante dela havia outro grande túnel e o Coelho Branco ainda estava à vista, apressado. --
6	Suddenly she came upon a little three-legged table, all made of solid glass; there was nothing on it but a tiny golden key, and Alice's first thought was that it might belong to one of the doors of the hall; but, alas! either the locks were too	De repente, topou com uma mesa de três pés, toda feita de vidro: não havia nada em cima, a não ser uma chavezinha dourada. A primeira ideia de Alice foi que a chave deveria ser de uma das portas, mas -- que nada! -- ou as	De repente, encontrou uma pequena mesa de três pés, toda feita em vidro sólido: não havia nada sobre ela senão uma minúscula chave dourada e a primeira ideia de Alice foi de que ela deveria pertencer a uma das portas da sala: «mas, ai de mim!, ou as

Figura 12 – Erro na anotação semântica de “cor” em Traduções Múltiplas.

Na figura acima, o item 4 apresenta a palavra *orange* em negrito, por ter sido considerada cor-de-laranja. Porém, neste caso, o termo representa a fruta ‘laranja’, e por isso deve ser apontado como marcação incorreta.

Outro exemplo trata da revisão da anotação semântica para partes do corpo (Fig. 13).

First came ten soldiers carrying clubs: these were all shaped like the three gardeners, oblong and flat, with their hands and feet at the corners: next the ten courtiers; these were ornamented all over with diamonds, and walked two and two, as the soldiers did.	Primeiro surgiram dez soldados armados com maças. Eram todos iguais aos jardineiros, retangulares e achatados, com as pernas e os braços nos quatro ângulos. Em seguida vieram dez cortesãos, paramentados com diamantes em forma de losangos. Caminhavam de dois em dois, assim como os soldados.	Em primeiro lugar chegaram dez soldados carregando clavas: eles eram todos da mesma forma que os jardineiros, retangulares e achatados, com as mãos e os pés saindo dos quatro cantos; depois vinham dez cortesãos, que eram ornamentados com diamantes e caminhavam de dois em dois, como os soldados.
After these came the royal children; there were ten of them, and the little dears came jumping merrily along, hand in hand, in couples: they were all ornamented with hearts.	Depois vieram os infantes reais, também em número de dez, saltitando alegremente de mãos dadas, em pares, todos enfeitados com corações .	Depois desses vinham as crianças reais, dez delas, e as gracinhas iam saltitando alegremente de mãos dadas, em duplas também.
Alice was rather doubtful whether she ought not to lie down on her face like the three gardeners, but she could not	Alice ficou em dúvida se devia ou não atirar-se ao chão como os jardineiros, mas não se recordava de ter ouvido falar	Alice estava em dúvida se deveria ou não atirar-se ao chão de bruços como os três jardineiros, mas não conseguia se lembrar se

Figura 13 – Anotação inadequada de corpo para ‘corações’

Nota-se, na primeira tradução (segunda coluna), a palavra em negrito ‘corações’, termo que, automaticamente, foi marcado com a etiqueta de “corpo”. Porém, nesse caso, ‘corações’ não faz menção ao órgão torácico que é o maior responsável pela circulação do sangue em um corpo, mas sim, à forma estilizada desse órgão, que simboliza o amor e o naipe ‘copas’ do baralho.

Assim, essa etiqueta foi corrigida durante o processo de revisão.

4.1 O impasse da concepção cultural de “cor”

Keesing (1961, p. 551) afirma que “a linguagem é uma forma do comportamento aprendido e, assim, é reconhecida pelos antropólogos como um aspecto da cultura”, e como cada cultura carrega manifestações e hábitos próprios, isso enseja pontos de vista característicos de cada sociedade sobre diversos contextos do cotidiano. Assim, a linguagem será um aspecto singular de cada cultura, uma vez que é ela que exprime os pensamentos, ela é “o registro do pensar conceitual articulado” (Eble, 1982, p. 79).

Dessa forma, a noção de cor também é compreendida e transmitida de forma típica. Ao pesquisar ocorrências por meio do sistema de busca do CorTrad, com o texto original como principal, alguns casos demandaram mais atenção e reflexão por apresentarem certa perspectiva cultural própria, como na Figura 14 a seguir:

34	She was walking by the White Rabbit, who was peeping anxiously into her face. UNDEF	Alice estava caminhando ao lado do Coelho Branco, que a observava com ar muito inquieto. --	Ela estava caminhando bem ao lado do Coelho Branco, que ficava olhando o tempo todo para ela. --
35	«They're done with blacking . I believe.»	Alice olhou para baixo antes de responder: «o sapateiro. " --	«Acho que são lustrados com uma escova, eu acho. São escovados. " --
36	So Alice began telling them her adventures from the time when she first saw the White Rabbit.	Então Alice começou a contar-lhes as suas aventuras desde quando vira pela primeira vez o Coelho Branco.	Então Alice começou a contar suas aventuras desde a primeira vez que viu o Coelho Branco.

Figura 14 – Pesquisa do sema ‘cor’ na anotação semântica do CorTrad com o texto original como principal.

Na unidade de alinhamento 35, nota-se a ocorrência de *blacking* na coluna do texto original. Tal termo, traduzido literalmente para o português, seria *empretecer*, mas, na realidade, significa *lustrar* ou *engraxar os sapatos*.

Assim, questionou-se se a palavra *blacking* deveria continuar marcada como um sema de cor ou não. Ao considerar que a graxa é preta e, ao ser passada no sapato ela o empretece, a conclusão a que se chegou foi que o termo deveria continuar estabelecido como ‘cor’, já que, aparentemente, para os falantes do inglês em questão, engraxar é o mesmo que “tornar o sapato preto”.

Outro exemplo refere-se à palavra *gold*, que pode representar tanto a cor ‘dourado’, em português, quanto o metal ‘ouro’ (Fig. 15).

The warmth, fragrance and colour of her body appealed to his senses.	O calor, a fragrância e o colorido daquele corpo jovem provocavam-lhe os sentidos.	O calor, o aroma do corpo da mulher estimulavam-lhe os sentidos.
Mr Holohan became very red and excited	Holohan ficou mais vermelho e agitado.	Mr Holohan ficou vermelho e nervoso.
The first tenor bent his head and began to count the links of the gold chain which was extended across his waist, smiling and humming random notes to observe the effect on the frontal sinus.	O primeiro tenor baixou a cabeça e começou a contar os elos da corrente de ouro, que lhe cruzava o peito, sorrindo e murmurando algumas notas para observar o efeito na cavidade frontal.	O primeiro tenor abaixou a cabeça e começou a contar os elos da corrente dourada que trazia à cintura, sorrindo e cantarolando notas aleatórias para checar o efeito da voz na cavidade frontal.
Her face was inundated with an angry colour and she looked as if she would attack someone with her hands.	Seu rosto estava ameaçadoramente rubro e ela parecia pronta a agredir alguém.	Seu rosto estava tomado por um rubor irado e parecia prestes a agredir alguém.

Figura 15: Traduções distintas para *gold*

O primeiro tradutor traduziu *gold chain* como ‘corrente de ouro’, identificando *gold* como um metal, enquanto o segundo o entendeu como uma cor, ‘corrente dourada’. Além das implicações de valorização do objeto – certamente uma corrente de ouro tem mais valor do que uma corrente dourada – surge a questão de como anotar *gold* no original: como metal ou como cor? Por outro lado, essa dúvida surgiu ao se analisar as traduções. A anotação semântica deveria levar em consideração as traduções?

Um último exemplo é *brown bread*, costumeiramente traduzido como ‘pão preto’. A pergunta que cabe é: *brown* é cor neste caso ou *brown bread* é uma colocação à qual corresponde, em português, outra colocação ‘pão preto’, mas com uma cor diferente? E o ‘pão preto’, na maioria das vezes, não é exatamente preto, variando em tons de marrom, daí a denominação em inglês. Por outro lado, como se vê no primeiro exemplo da Figura 16 abaixo, *brown bread* opõe-se a *white bread*, esse sim sempre branco.

«Then I suppose they'll soon bring the white bread and the brown ?» Alice ventured to remark. -- UNDEF --	-- Nesse caso o povo não tarda a trazer-lhes o pão branco e o pão preto, não é assim? observou Alice. --	-- Então está quase na hora de comerem o pão? -- indagou Alice, curiosa. --
Haigha and Hatta set to work at once, carrying round trays of white and brown bread.	Ciro e Cairo foram buscar -- -- <i>refrescos</i> -- --, isto é, bandejas de pão branco e preto.	Leal e Feliz imediatamente começaram a servir fatias de pão branco e preto, em bandejas. --
Haigha and Hatta set to work at once, carrying round trays of white and brown bread.	Ciro e Cairo foram buscar -- -- <i>refrescos</i> -- --, isto é, bandejas de pão branco e preto.	Leal e Feliz imediatamente começaram a servir fatias de pão branco e preto, em bandejas. --
«None of your brown bread for me!»	Não quero saber de pão preto. --	-- Para mim nada de pão! --
However, there was the great dish still lying at her feet, on which she had tried to cut the plum-cake. «So I wasn't dreaming, after all.» she said to herself, «unless -- unless we're all part of the	Vendo-o, Alice refletiu que não podia ter sido sonho. E se tivesse sido um sonho do Rei Branco e não dela? Desconfiada disso, Alice resolveu acordá-lo para ver o que acontecia. --	Só restava o prato vazio... -- -- Mas, se restava o prato, não fora sonho. -- -- " Só fizemos todos parte de um mesmo sonho pensou ela. «Se ao menos for sonho meu não sonho do Rei Vermelho! Acho horrível

Figura 16: A tradução de *brown bread*

Fica então a pergunta: *brown* é cor e como tal deve ser anotado, ou é parte de uma colocação, devendo ser anotado como “cor:original”, o que indica um uso arbitrário, idiomático da cor? Na dúvida, talvez o mais conveniente seja anotá-lo como ‘vago’.

Outros aspectos relativos à anotação de cor já foram discutidos em Santos et al. (2012).

4.2. Revisão de “cor” e “corpo”: dados quantitativos

Para uma visualização mais clara do número de alterações feitas nas anotações de “cor” e de “corpo” durante a revisão das etiquetas, apresentamos abaixo os dados quantitativos de cada versão (original, tradução 1 e tradução 2) disponível em Múltiplas Traduções, no CorTrad, tanto das obras de Lewis Carroll (ver Tabela 2) e de James Joyce (ver Tabela 3).

Anotação semântica – Lewis Carroll	Revisão	Número inicial no <u>CorTrad</u>	Número de erros encontrados	Número final no <u>CorTrad</u>
Original	Cor	83	1	82
Tradução 1	Cor	57	1	56
Tradução 1	Corpo	762	16	746
Tradução 2	Cor	54	1	53
Tradução 2	Corpo	662	15	647

Tabela 2: Dados quantitativos após revisão semântica – Lewis Carroll

Como se pode observar, a maior discrepância ocorreu nas etiquetas para “corpo” nas obras de Lewis Carroll, com 16 erros encontrados na Tradução 1 e 15 na Tradução 2. A grande maioria advém de três termos: “rabo”, “cara” e “caracol”. A confusão com a palavra “rabo” ocorreu devido a diferenças linguísticas entre o português europeu e o português brasileiro, pois em Portugal “rabo” refere-se a uma parte do corpo humano – sinônima de “traseiro”, por exemplo – enquanto no Brasil a mesma palavra é somente usada denominar uma parte do corpo de animais. O segundo termo, “cara”, foi, na verdade, utilizado muitas vezes como sinônimo de “querida”, a exemplo de “Como está se sentindo agora, minha cara?”. O terceiro termo, “caracol”, destacado na busca por “corpo” como um estilo de penteado, era, na realidade, um dos personagens animais da história, ou seja, um molusco, e não um formato de cabelo. Por essa razão, foi computado como erro.

Anotação semântica – James Joyce	Revisão	Número inicial no <u>CorTrad</u>	Número de erros encontrados	Número final no <u>CorTrad</u>
Original	Cor	229	13	216
Tradução 1	Cor	206	2	204
Tradução 1	Corpo	845	1	844
Tradução 2	Cor	218	4	214
Tradução 2	Corpo	947	1	946

Tabela 3: Dados quantitativos após revisão semântica - James Joyce

Os casos de maior discrepância no texto original da obra de James Joyce ocorreram na revisão de “cor”, que apresentou 13 erros. Os principais problemas aconteceram devido ao nome da rua *Stephens’s Green*, em que a palavra “Green” foi considerada “cor” em todas as suas ocorrências. Por outro lado, o programa considerou as palavras *gold* e *silver*, respectivamente, como “dourado” e “prateado” quando de fato referiam-se ao metal, tanto

assim que os tradutores optaram corretamente por “de ouro” e “de prata”, como se vê a seguir:

A small gold coin shone in the palm.	Uma pequena moeda de ouro brilhava sobre ela.	Ali brilhava uma pequena moeda de ouro .
While she spoke she turned a silver bracelet round and round her wrist.	-- -- Enquanto falava, ela fazia girar um bracelete de prata .	-- -- Enquanto falava ela girava continuamente uma pulseira de prata que usava.

5. Considerações finais

Ainda que o alinhamento de *corpora* paralelos deva seguir um padrão geral, por vezes são necessárias adaptações e escolhas sobre o modo de alinhar para que as particularidades da escrita de cada autor sejam respeitadas. Cada *corpus* é único e, portanto, não deve ser tratado apenas como um texto coletado adicionalmente. Por vezes, uma nova questão poderá fazer com que todas as decisões feitas anteriormente sejam revistas, ou mesmo que sejam criadas novas categorias.

Os problemas que surgem durante os processos tanto de alinhamento quanto de revisão da anotação semântica ensejam reflexões sobre as estratégias de tradução assim como ampliam o conhecimento da linguagem de uma cultura alheia. Tarefas aparentemente repetitivas podem proporcionar lições e aprendizados significativos.

Quanto ao CorTrad, parte integrante do Projeto CoMET, a crescente importância das pesquisas linguísticas o torna cada vez mais relevante, não apenas para a Tradução, mas também para seus outros segmentos constituintes, a Terminologia e o Ensino e Aprendizagem de Línguas. A disponibilização de *corpora* para pesquisa *on-line* colabora, de maneira efetiva, para o estudo das línguas por meio dos princípios da Linguística de *Corpus*.

Referências

BICK, Eckhard. *The parsing system "Palavras": automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus: Aarhus University Press, 2000.

CARROLL, Lewis. Through the looking-glass, and what Alice found there. In: L. Carroll, *The Complete Works of Lewis Carroll*. London: Collins, 1939. Disponível em:

http://www.gasl.org/refbib/Carroll_Works.pdf

CARROLL, Lewis. *Alice no Reino do Espelho*. (M. T. Giácomo, Trad.) São Paulo: Melhoramentos, 1962 (2a. ed.).

CARROLL, Lewis. *Alice no País do Espelho* (2a. ed.). (M. Lobato, Trad.) São Paulo: Brasiliense, 1973.

CARROLL, Lewis. *Alice no País das Maravilhas* (2a. ed.). (I. d. Ascher, Trad.) São Paulo, 2000. Disponível em:

http://200.136.76.125/colegio/livros/download/alice_no_pais_das_maravilhas.pdf

CARROLL, Lewis. *Alice no País das Maravilhas*. (C. R. Ramos, Trad.) Rio de Janeiro: Arara Azul, 2002. Disponível em:

<http://www.ebooksbrasil.org/eLibris/alicep.html>

CARROLL, Lewis. *Alice's adventures in Wonderland* (3rd ed.). The Millennium Fulcrum, 2011. Disponível em:

<https://www.gutenberg.org/>

EBLE, Alroino. B. "Cultura e linguagem". *Revista de Ciências Humanas* v. 1, n. 2, 1982, p. 78-82.

FRANKENBERG-GARCIA, Ana. "Verificação da etiquetagem automática". 2004. Disponível em:

<http://www.linguateca.pt/COMPARA/docum/VerifEtiquetagem.pdf>

FRANKENBERG-GARCIA, Ana. "Instruções de alinhamento por parágrafo". 2006. Disponível em:

<http://www.linguateca.pt/COMPARA/docum/AlinhamentoPar.pdf>

FRANKENBERG-GARCIA, Ana. "COMPARA - Instruções para a revisão do Reconhecimento Óptico de Caracteres (ROC) e etiquetagem preliminar". 2007. Disponível em:

<http://www.linguateca.pt/COMPARA/docum/ROC.pdf>

FRANKENBERG-GARCIA, Ana, & SANTOS, Diana. "COMPARA, um corpus português-inglês na Web". *Cadernos de Tradução* 9, 2002, p. 61-79.

FRANKENBERG-GARCIA, Ana; SANTOS, Diana & SILVA, Rosário. "COMPARA - Revisão do alinhamento por frase". 2006. Disponível em:

<http://www.linguateca.pt/COMPARA/docum/AlinhamentoFrase.pdf>

FREITAS, Cláudia; SANTOS, Diana; CARRIÇO, Bruno; MOTA, Cristina & JANSEN, Heidi. "O léxico do corpo e anotação de sentidos em grandes corpora - o projeto Esqueleto". *Revista de Estudos da Linguagem* 23, 3, 2015, p. 641-680.

JOHANSSON, Stig, & HOFLAND, Knut. Towards an English-Norwegian parallel corpus. In: U. Fries, G. Tottie, & P. Schneider, *Creating and using*

English Language corpora. Papers from the fourteenth International Conference on English Language Research on Computerized Corpora. Amsterdam-Atlanta: Rodopi, 1993, p. 25-37.

JOYCE, James. *Dubliners*. London: Penguin Popular Classics. 1996 [1914]

JOYCE, James. *Dublinenses*. (J. R. O'Shea, Trad.) São Paulo: Siciliano, 1994 (2a. ed.).

JOYCE, James. *Dublinenses*. Rio de Janeiro: Civilização Brasileira, 1997 (5a. ed.).

KEESING, Roger M. (1961). *Antropologia cultural - vol. 2*. Rio de Janeiro: Fundo de Cultura.

MOTA, Cristina, & SANTOS, Diana. "Emotions in natural language: a broad-coverage perspective". Janeiro de 2015. Disponível em:

<http://www.linguateca.pt/acesso/EmotionsBC.pdf>

RAYSON, Paul, & GARSIDE, Roger. "The CLAWS Web Tagger". *ICAME Journal* 22, 1998, p. 121-123.

SANTOS, Diana. "On the use of parallel texts in the comparison of languages". In: M. F. Nascimento, M. G. Rodrigues, & J. Bettencourt, *Actas do XI Encontro da Associação Portuguesa de Linguística* (Vol. I.). Lisboa: Associação Portuguesa de Linguística, 1996, p. 217-239.

SANTOS, Diana. "DISPARA, a system for distributing parallel corpora on the Web". In: E. Ranchod, & N. J. Mamede, *Advances in Natural Language Processing (Third International Conference, PorTAL 2002)*. Berlin: Springer, 2002, p. 209-218.

SANTOS, Diana, & SIMÕES, Alberto. "Ensinador paralelo: alicerces para uma pedagogia nova". In: Alberto Simões, Anabela Barreiro, Diana Santos, Rui Sousa-Silva, & Stella E. Tagnin (Eds.), *Linguística, Informática e Tradução: Mundos que se cruzam*. Oslo: University of Oslo, 2015, p. 235-252.

SANTOS, Diana; SILVA, Augusto Soares da, & MOTA, Cristina. "Guardafatos: notas sobre a anotação do campo semântico do vestuário em português". 21 de fevereiro de 2011. Disponível em:

<http://www.linguateca.pt/acesso/GuardaFatos.pdf>

SANTOS, Diana; TAGNIN, Stella E., & TEIXEIRA, Elisa Duarte. "CorTrad and Portuguese-English translation studies: investigating colours". *Studies in Variation*, v. 12, 2012, pp. 1-15.

SILVA, Rosário, & SANTOS, Diana. *Arco-íris: notas sobre a anotação do campo semântico da cor em português*. 2012. Disponível em:

<http://www.linguateca.pt/acesso/ArcoIris.pdf>

TAGNIN, Stella E. O. (org.). *Lá do Canadá*. São Paulo: Olavobrás. 2002

TAGNIN, Stella E. O. "COMET - a multilingual corpus for teaching and translation". In: *PALC 2001: Practical Applications in Language Corpora*, Lodz,



Poland. Frankfurt am Main: Peter Lang - Europäischer Verlag der Wissenschaften, 2003, p. 535-540.

TAGNIN, Stella E. O. (org.). *Lá da Austrália*. São Paulo: Fólio, 2005.

TAGNIN, Stella E. O. "Disponibilização de corpora on-line: os avanços do Projeto COMET". In: S. E. Tagnin, & O. A. Vale, *Avanços da Linguística de Corpus no Brasil*. São Paulo: Humanitas, 2008, p. 95-115.

TEIXEIRA, Elisa Duarte. *A Linguística de Corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual*. São Paulo: Universidade de São Paulo, 2008.

TEIXEIRA, Elisa Duarte; SANTOS, Diana & TAGNIN, Stella E. O. "CorTrad: um corpus paralelo multiversão para o par de línguas português-inglês". In: Tania M. Shepherd; Tony Berber Sardinha, & Marcia V. Pinto, *Caminhos da Linguística de Corpus*. Campinas: Mercado de Letras, 2012, pp. 151-175.

Resumo

O CorTrad, parte do Projeto CoMET, é um corpus bidirecional inglês-português com textos jornalísticos, técnico-científicos e literários. Este artigo trata de recentes acréscimos à parte literária: duas obras de Lewis Carroll - *Alice's Adventures in Wonderland* e *Through the Looking-Glass and What Alice Found There* e o livro de contos *Dubliners*, de James Joyce. Cada uma das obras conta com duas traduções por tradutores distintos. Por suas peculiaridades estilísticas e linguísticas, surgiram problemas de alinhamento e de anotação semântica, que são descritos e discutidos, em detalhe.

Abstract

CorTrad, part of the CoMET Project, is a bidirectional English-Portuguese corpus with journalistic, technical-scientific and literary texts. This article is about recent additions to the literary section: two works by Lewis Carroll - *Alice's Adventures in Wonderland* and *Through the Looking-Glass and What Alice Found There* and James Joyce's book of short stories *Dubliners*. Each of the works has two translations by different translators. Due to their stylistic and linguistic peculiarities, problems of alignment and semantic annotation have arisen which are described and discussed in detail.

