

A LINGUÍSTICA DE *CORPUS* NA FORMAÇÃO DE TRADUTORES: COMPILAÇÃO E ANÁLISE DE UM *CORPUS* DE APRENDIZES DE TRADUÇÃO

Joacyr Oliveira

Com pouquíssimas exceções, os escassos cursos superiores de formação de tradutores e intérpretes no Brasil são relativamente novos e, em muitos centros universitários da rede privada, são classificados como uma ramificação da licenciatura em Letras. O número ainda bem irrisório de estudos e pesquisas sobre a atual situação do ensino e aprendizagem de tradução em faculdades no Brasil é um forte reflexo dessa realidade e gera uma série de reflexões, entre elas: Será que os currículos dos cursos de tradução europeus podem ser implementados no Brasil, ou ainda, serviriam de modelo para nossas necessidades? A infraestrutura das faculdades particulares brasileiras é comparável à europeia ou norte-americana? Nossos professores têm a mesma formação e capacitação daqueles que atuam no velho continente? Os alunos brasileiros poderiam ser equiparados em sua formação nos ensinamentos fundamental e médio aos europeus? Será que a mesma abordagem usada para ensinar tradução no par francês-inglês ou alemão-inglês poderia ser aplicada ao par português-inglês? Sem respostas convincentes, nosso estudo tem como objetivo mais abrangente conhecer melhor o aluno brasileiro de tradução. Dentro desse universo, propomos um recorte no nível linguístico. Como nossos alunos percebem o texto de partida? São capazes de identificar os idiomatismos e as convencionalidades presentes na língua estrangeira ou acabam por se focar em questões mais microtextuais como salientado por Alves e Tagnin (2009). Considerando que o objeto final do ato tradutório é o resultado de um processo de decisões (Levy 1966 [2000]) torna-se necessário buscar uma forma de conhecermos melhor esses processos nos alunos. Justifica-se, assim, a necessidade da compilação de um corpus de aprendizes de tradução, pois, desse modo, poderemos identificar as tendências nas escolhas lexicais e morfossintáticas dos tradutores-aprendizes, e, a partir da perspectiva do aluno, observar o desenrolar dos processos e das técnicas de aprendizagem da tradução.

Este artigo tem por objetivo apresentar um recorte de uma pesquisa maior que trata da formação de tradutores. Aqui apenas discutiremos a solução encontrada para o alinhamento de várias versões de uma mesma tradução.

Nosso corpus, ainda em processo de etiquetagem, foi compilado a partir das traduções feitas por alunos do primeiro e segundo semestres de uma faculdade particular na cidade de São Paulo. A coleta ocorreu entre 2010 e 2012 em quatro turmas por semestre, cada uma com uma média de 30 alunos, que trabalharam com cerca de oito textos originais, gerando uma média de 900 traduções por semestre. Assim, estimamos ter em mãos um corpus com aproximadamente cinco mil textos. Após responderem a um questionário (que denominamos *Questionário de Autorização*) com informações sobre idade, sexo, conhecimento de inglês, vivência no exterior e experiência em trabalhos de tradução, cerca de 400 alunos assinaram um termo de autorização de uso de suas traduções para a compilação do corpus. Os textos originais foram segmentados em períodos sintáticos e enviados no formato DOC aos alunos para serem traduzidos. O resultado final, enviado por e-mail ao professor, foi apresentado em unidades de tradução, ou seja, segmento original seguido paralelamente por sua tradução. (Ver figura 1. Todas as figuras estão no anexo após as referências.)

Com os textos em mãos, a primeira questão a ser resolvida era como dispor o corpus para iniciar as análises, uma vez que temos por objetivo identificar as tendências na produção das traduções. Para termos uma visão macro, decidimos que as traduções deveriam estar alinhadas trecho a trecho com seu texto original. Ao buscarmos um programa de computador que pudesse fazer isso, iniciamos com as ferramentas tradicionalmente usadas para alinhamento como *ParaConc* (BARLOW2003), *Aligner* do WordSmith Tools (SCOTT 2012), *EAlign*, *CWB-Corpus Workbench* (HEID 2010), assim como algumas *on-line* como *CasualConc* (IMAO 2010) e *ABBYY Aligner Online* (2012) entre outras. No entanto, deparamos com um problema que nenhum dos programas mencionados pode resolver: o número elevado de traduções referentes a um mesmo original. O *ParaConc*, assim como os outros programas, por exemplo, alinha até três textos no máximo; já a versão 6.0 do WordSmith Tools gera erros ao alinhar um número elevado de textos. A segunda questão era encontrar uma forma prática de inserir cabeçalho em cada uma das inúmeras traduções, além de etiquetar todas as suas linhas. O uso do cabeçalho é de fundamental importância, pois permite ao pesquisador

segmentar as traduções por faixa etária, conhecimento de inglês, experiência em tradução etc. A etiquetagem das linhas, por sua vez, viabiliza a segmentação de partes específicas do texto.

A ferramenta que nos ofereceu as condições necessárias para os alinhamentos, inserção de cabeçalhos e etiquetagem das linhas em todos os textos de maneira mais prática foi o *Microsoft Excel* do Pacote Microsoft Office 2010. A escolha da Planilha Excel pode gerar alguma estranheza inicial a quem ainda não a viu em estudos de corpora, porém, por ser um programa manipulador de dados, além de cálculos, oferece os recursos necessários para alinhar e etiquetar os textos de nosso corpus.

Para cada turma de alunos, preparamos uma pasta (arquivo em Excel que contém uma ou mais planilhas) com as seguintes planilhas: 1. Tabulação das respostas dadas ao *Questionário de Autorização*; 2. Inserção de cabeçalho e etiquetagem das linhas; 3. Traduções alinhadas verticalmente. Para os próximos textos da mesma turma de alunos repetiram-se as planilhas 2 e 3. A planilha 1 dispõe na linha 1 as perguntas do *Questionário de Autorização*. Na coluna A são lançados os nomes dos alunos e nas interseções das linhas com as colunas adicionamos as respostas dadas. Abaixo da tabela com as respostas há gráficos estatísticos que nos permitem ter uma visão geral do perfil da turma de alunos. (Ver figura 2.)

Ao recebermos os arquivos dos alunos em formato DOC, transferimos manualmente apenas a tradução para a planilha 2, que foi previamente formatada para receber um original em uma coluna e todas as traduções do referido texto nas colunas à direita. Tanto o original como cada uma de suas traduções ocupam três colunas da planilha: à esquerda as etiquetas de abertura, no meio o cabeçalho e o texto e à direita as etiquetas de fechamento. Os cabeçalhos dos textos de partida contêm informações como gênero, número de *types* e *tokens* etc; os cabeçalhos das traduções trazem as informações dos questionários respondidos pelos alunos. (Ver figura 3.)

Logo abaixo do cabeçalho estão os textos (original e traduções). A primeira coluna contém as etiquetas de abertura, por exemplo <ua_id=en00301> (“ua_id” = *unidade de alinhamento – identificação*, “en” = *inglês*, “003” = *número que identifica o texto original*, “01” = *número que identifica a linha*); as traduções também seguem o mesmo código, sendo os dígitos antecidos por “pt”; a segunda coluna traz o texto (original ou tradução) dispondo cada segmento (período sintático) em uma linha; e a

coluna 3 apresenta a etiqueta de fechamento, por exemplo: </ua>. Para cada tradução repetimos as três colunas. (Ver figura 4.) As etiquetas utilizadas foram baseadas nos documentos da TEI – *Text Encoding Initiative*. (BURNARD, L. e BAUMAN, S. 2009)

A planilha 2 também é usada para salvar cada uma das traduções com seu cabeçalho e etiquetas em formato de texto simples (txt), permitindo posteriormente seu processamento no WordSmith Tools. (Ver figura 6.)

A planilha 3 traz na linha 1 o texto de partida, sendo que cada segmento corresponde a uma coluna. Logo abaixo, verticalmente, as linhas de cada coluna trazem as diversas traduções. Essa planilha permite a macroanálise das traduções. (Ver figura 5.)

O fator mais importante em nossa decisão para incorporarmos a planilha eletrônica em nossa pesquisa foi a sua grande praticidade, pois uma vez montada a planilha matriz, todo o processamento é feito de modo automático por meio do uso de fórmulas (Cf. tutoriais PROCH e PROCV no site da Microsoft). Ao inserirmos as informações dos alunos na planilha 1, o cabeçalho da planilha 2 atualiza-se automaticamente. Na planilha 2, ao alteramos algum dado nas etiquetas da primeira tradução, todas as outras etiquetas também serão alteradas. À medida que inserimos as traduções na planilha 2, a planilha 3 realinha-se com os novos textos.

Neste artigo demonstramos como uma planilha eletrônica, como a Microsoft Excel, que não foi projetada para pesquisas linguísticas, mostrou-se eficaz para resolver um problema que os programas criados para alinhamento textual não conseguiam solucionar. Características como a flexibilidade em manipular dados e o cruzamento de informações por meio de fórmulas colocam-na como uma excelente candidata a se tornar uma ferramenta alternativa nas pesquisas em linguística de corpus.

Referências

ABBYY Aligner online (2012). Abbyy Products. Disponível para uso online em:

<<http://aligner.abbyyonline.com/en>>

ALVES, F. e TAGNIN, S. E. O. (2010). “Corpora e ensino de tradução: O papel do automonitoramento e da conscientização cognitivo-discursiva no processo de aprendizagem de tradutores novatos”, in VIANA, V. & TAGNIN, S.(orgs) *Corpora no Ensino de Línguas Estrangeiras*, São Paulo: Hub Editorial, 189-204.

BARLOW, M. (2003). *ParaConc 2.6.0*. Huston: Athelstan. CD-ROM

BURNARD, L. e BAUMAN, S. (2009). *TEI P5: Guidelines for Electronic Text Encoding and Interchange 1.3.0*, Oxford: TEI Consortium. Acesso em: março de 2012.

Disponível em: <<http://www.tei-c.org/Guidelines/P5/>>.

HEID, U. (2010). *Corpus Workbench 3.0*. Stuttgart: University of Stuttgart, Institute for Natural Language Processing. Disponível para download em <<http://cwb.sourceforge.net/>>

IMAO, Yasu. (2010). *CasualConc 1.0.5: CasualMultiPConc Text Aligner*. Osaka, Osaka University, Acesso em 3 de mar. de 2012. Disponível para download em <<https://sites.google.com/site/casualconc/Home>>

AUTOR(es). Título, número da versão (exceto a primeira). Cidade de publicação : Publicador, ano. Descrição física.

LEVÝ, J.(1966[2000]). “Translation as a decision process”, in: VENITI, L. *The translation studies reader*. Londres: Routledge, 148-59.

"PROCH (Função PROCH)." PROCH (Função PROCH). Microsoft, 2012. Acesso em 19 de janeiro de 2012. Disponível em: <<http://office.microsoft.com/pt-br/excel-help/proch-funcao-proch-HP010342579.aspx>>.

"PROCV (Função PROCV)." PROCV (Função PROCV). Microsoft, 2012. Acesso em 19 de janeiro de 2012. Disponível em: <<http://office.microsoft.com/pt-br/excel-help/procv-funcao-procv-HP010343011.aspx>>

SCOTT, M. (2012). *WordSmith Tools*. Liverpool: Lexical Analysis Software.

Anexo – Figuras

	Texto 1 - Língua de partida	Língua de chegada
<ua id=en1201> <p><t>	<u>Beautiful Day, Isn't It?</u>	Lindo dia, não é?
<ua id=en1202><p>	The day started out rotten.	O dia começou podre.
<ua id=en1203>	Mary overslept and was late for work.	Mary dormiu demais e estava atrasada para o trabalho.

Fig. 1: Excerto do texto em tabela onde os alunos fazem a tradução.

	A	B	C	D	E	F	G	H
1	Nome do Aluno	Número que identifica o aluno	Ano_se mestre qdo a trad foi feita	Idade quando fez o curso: a. age=17-25 b. age=26-30 c. age=31-40 d. +40.	Sexo a. Masculino b. Feminino	Conhecimento de inglês ao iniciar o primeiro semestre: a. Bas b. Inter c. Adv d. post-adv	Vivência no país de língua inglesa antes de iniciar o primeiro semestre: a. None b. Up_to_1 yr c. 1_to_5 yrs d. +5 yrs	Língua nativa: a. Português = Port b. Espanhol = Span c. Japonês = Jap d. Chinês = Ch e. Coreano = Cor
2	Ana_██████	1	2012_1	age=31-40	f	Post-Adv	1 to 5 yrs	Native=Port
3	Ana_██████	2	2012_1	age=17-25	f	bas	None	Native=Port
4	André_██████	3	2012_1	age=17-25	m	Inter	None	Native=Port
5	André_██████	4	2012_1	age=17-25	m	Adv	None	Native=Port
6	Beatriz_██████	5	2012_1	age=17-25	f	Adv	None	Native=Port

Fig. 2: Planilha 1 - Excerto da planilha que contém a tabulação dos questionários.

3				<header>				
4	<header>			<student_id>	1	</student_id>		
5	<text_title>	Beautiful day	</text_title>	<date_of_translation>	2012_1	</date_of_translation>		
6	<text_source>	//www.mdsupport.org/revelations2	</text_source>	<age>	age=31-40	</age>		
7	<tokens>	179	</tokens>	<gender>	f	</gender>		
8	<types>	117	</types>	english_background>	Post-Adv	</english_background>		
9	<genre>	literature	</genre>	<years_abroad>	1 to 5 yrs	</years_abroad>		
10	<theme>	chronicle	</theme>	<native_language>	Native=Port	</native_language>		

Fig. 3: Planilha 2 - Excerto do cabeçalho do texto original e da tradução.

20	<ua_id=en00301><p><t>	Beautiful Day, Isn't It?	</t></p></ua>	<ua_id=pt00301><p><t>	Lindo dia, não é?	</t></ua>
21	<ua_id=en00302><p>	The day started out rotten.	</ua>	<ua_id=pt00302><p>	O dia começou horrível.	</ua>
22	<ua_id=en00303>	Mary overslept and was late for work.	</ua>	<ua_id=pt00303>	Mary perdeu a hora e estava atrasada para o trabalho.	</ua>
23	<ua_id=en00304>	Everything that happened at the office contributed to her nervous frenzy.	</ua>	<ua_id=pt00304>	Tudo o que aconteceu no trabalho contribuiu para seu estado de nervos.	</ua>
24	<ua_id=en00305>	By the time she reached the bus stop for her homeward trip, her stomach was one big knot.	</p></ua>	<ua_id=pt00305>	No momento que ela chegou ao ponto de ônibus para sua jornada à casa, ela tinha um nó no estomago.	</ua>

Fig. 4: Planilha 2 - Excerto do texto de partida e uma tradução.

B	C	D	E
Beautiful Day, Isn't It?	The day started out rotten.	Mary overslept and was late for work.	Everything that happened at the office contributed to her nervous frenzy.
Lindo dia, não é?	O dia começou horrível.	Mary perdeu a hora e estava atrasada para o trabalho.	Tudo o que aconteceu no trabalho contribuiu para seu estado de nervos.
Lindo dia, não é?	O dia começou péssimo.	Mary perdeu a hora e estava atrasada para o trabalho.	Tudo o que aconteceu no escritório contribuiu para seu nervoso incontrolável.
Lindo dia, não?	O dia começou ruim.	Mary dormiu demais e estava atrasada para o trabalho.	Tudo o que aconteceu no escritório contribuiu para o seu nervosismo.
Lindo dia, não é?	O dia já começou ruim.	Mary dormiu demais e se atrasou para o trabalho.	Tudo o que aconteceu no escritório contribuiu para seu estado de nervos.
Lindo dia, não?	O dia começou mal.	Mary dormiu demais e se atrasou para o trabalho.	Tudo o que aconteceu no escritório contribuiu para seu nervosismo.
Lindo dia, não?	O dia começou ruim.	Mary dormiu demais e se atrasou para o trabalho.	Tudo o que aconteceu em seu serviço contribuiu para seu nervosismo.
Lindo dia, não é?	O dia tinha começado podre.	Mary dormiu demais e estava atrasada para o trabalho.	Tudo que acontecia no escritório contribuiu para o seu estado nervoso.

Fig. 5: Planilha 3 - Exemplo das traduções alinhadas verticalmente. (A coluna A não é mostrada aqui por conter apenas números das linhas.)

```

Sem título - Bloco de notas
Arquivo Editar Formatar Exibir Ajuda
<header>
<student_id> 1 </student_id>
<date_of_translation> 2012_1 </date_of_translation>
<age> age=31-40 </age>
<gender> f </gender>
<english_background> Post-Adv </english_background>
<years_abroad> 1 to 5 yrs </years_abroad>
<native_language> Native=Port </native_language>
<professional_experience> exp=amateur
</professional_experience>
<other_field_background> other_area=bio
</other_field_background>
<genre> literature </genre>
<theme> chronicle </theme>
<text_id> 1 </text_id>
<level> 1 </level>
<comments> </comments>
</header>
<text_id=1pr_lit_chron_pt003_s001.txt>
<ua_id=pt00301><p><t> Lindo dia, não é? </t></ua>
<ua_id=pt00302><p> O dia começou horrível. </ua>
<ua_id=pt00303> Mary perdeu a hora e estava atrasada para o trabalho.
</ua>

```

Fig. 6: Exemplo de arquivo em texto simples.