UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LETRAS MODERNAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS
LINGUÍSTICOS E LITERÁRIOS EM INGLÊS

JOACYR TUPINAMBÁS DE OLIVEIRA

A Linguística de *Corpus* na formação do
tradutor: compilação e proposta de análise de
um *corpus* paralelo de aprendizes de tradução

Versão Corrigida

São Paulo
2015

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LETRAS MODERNAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS
LINGUÍSTICOS E LITERÁRIOS EM INGLÊS

A Linguística de *Corpus* na formação do
tradutor: compilação e proposta de análise de
um *corpus* paralelo de aprendizes de tradução

Joacyr Tupinambás de Oliveira

Dissertação apresentada à Faculdade
de Filosofia, Letras e Ciências
Humanas, Departamento de Letras
Modernas da Universidade de São
Paulo para a obtenção do título de
Mestre em Letras no Programa de
Estudos Linguísticos e Literários em
Inglês.

Orientadora: Profa. Dra. Stella Esther
Ortweiler Tagnin

São Paulo
2015

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catalogação na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

O481 Oliveira, Joacyr Tupinambás de
 A Linguística de Corpus na formação do tradutor:
 compilação e proposta de análise de um corpus
 paralelo de aprendizes de tradução / Joacyr
 Tupinambás de Oliveira ; orientadora Stella Esther
 Ortweiler Tagnin. - São Paulo, 2015.
 147 f.

Dissertação (Mestrado)- Faculdade de Filosofia,
Letras e Ciências Humanas da Universidade de São
Paulo. Departamento de Letras Modernas. Área de
concentração: Estudos Linguísticos e Literários em
Inglês.

1. Ensino e aprendizagem. 2. Tradução. 3.
Linguística de corpus. I. Tagnin, Stella Esther
Ortweiler, orient. II. Título.

Folha de Aprovação

Joacyr Tupinambás de Oliveira

A Linguística de *Corpus* na Formação do Tradutor: compilação e proposta de análise de um *corpus* paralelo de aprendizes de tradução

Dissertação apresentada à Faculdade de Filosofia, Letras e Ciências Humanas, Departamento de Letras Modernas da Universidade de São Paulo para a obtenção do título de Mestre em Letras no Programa de Estudos Linguísticos e Literários em Inglês.

Aprovada em: ____ / ____ / ____

Banca examinadora

Prof. Dr.: _____

Instituição: _____ Assinatura: _____

Prof. Dr.: _____

Instituição: _____ Assinatura: _____

Prof. Dr.: _____

Instituição: _____ Assinatura: _____

Dedicatória

Aos meus pais, Ary, *in memoriam*, e Lina, por me tornarem o que sou hoje.

Agradecimentos

A meus pais e irmãos, pelo amor e incentivo incondicionais desde o meu primeiro dia na Terra.

Em especial à minha orientadora, Stella Tagnin, pelo exemplo de pesquisadora, por acreditar em mim e por me guiar e iluminar com sua sapiência até mesmo nos caminhos mais caliginosos.

Aos professores Heloisa Cintrão e João Azenha Júnior pelos valiosos comentários e sugestões durante o exame de qualificação.

À professora e pesquisadora Diana Santos, da Universidade de Oslo, pelas sugestões pertinentes, pelas perguntas que me puseram a pensar e, principalmente, por acreditar na minha pesquisa.

A todos os colegas do Projeto COMET, por todas as conversas, discussões; à Malila Prado por sempre estar por perto em momentos de aflição; e especialmente à Rozane Rebechi pela revisão detalhada e inúmeras sugestões.

Aos meus colegas do Departamento de Letras da FMU e à minha coordenadora e amiga Beatriz Santana, que sempre teve palavras de apoio.

Aos meus colegas do curso de tradução da Anhanguera-Unibero e especialmente à minha professora, coordenadora e amiga, Cleide Cerdeira.

A todos os meus alunos que gentilmente cederam suas traduções para que fizessem parte do meu *corpus*.

Ao Tiago Kietzmann, pelo amor, companheirismo e apoio em todos os momentos, principalmente pelos sábados e domingos em meio a livros, artigos e computador.

Ao meu amigo-irmão, Élcio Souza, que sempre acreditou em mim, e sempre me incentivou, mesmo nos momentos em que eu não sabia por onde começar.

Aos meus gatos, Teobaldo e Frederico, por estarem sempre ao meu lado durante horas e horas de pesquisa.

E a todos aqueles que, de uma ou outra maneira, me ajudaram, incentivaram ou apoiaram para que esta fase da pesquisa chegasse ao fim.

*Possible does not mean likely. It just means
it's not impossible.*

Sherlock Holmes, série de TV *Elementary*, temporada
1, episódio 17, Columbia Broadcasting System

Sumário

Índice das figuras	x
Índice dos gráficos	xii
Índice das tabelas	xiii
Lista de abreviaturas.....	xiv
Resumo.....	xv
Abstract.....	xvi
Introdução.....	1
Capítulo 1 – Fundamentação teórica	13
1.1 O que é Linguística de <i>Corpus</i>	14
1.1.1 Tipologia dos <i>corpora</i>	15
1.1.2 Ferramentas para os estudos de <i>corpora</i>	18
1.2 A linguística de <i>corpus</i> nos estudos da tradução.....	26
Capítulo 2 – O ensino da tradução.....	28
2.1 Bacharelado em tradução	29
2.2 Proposta de aula.....	38
Capítulo 3 – Materiais e métodos	48
3.1 Perfil dos participantes.....	49
3.2 A compilação do <i>corpus</i>	54
3.3 <i>Microsoft Excel</i> como uma ferramenta na Linguística de <i>Corpus</i>	57
3.4 Alinhamento e etiquetagem do texto usando o <i>Microsoft Excel</i>	63
3.4.1 Aba ‘Lista de Alunos’	64
3.4.2 A criação de etiquetas	68
3.4.3 Aba ‘Texto xx’	68
3.4.4 Aba ‘Texto xx vert’	78
3.4.5 Aba ‘Texto xx alinhado’	84
3.4 Uso do <i>WordSmith Tools</i> para manipular as traduções	86
Capítulo 4 – Análise das traduções.....	96
4.1 Análise do Texto 3.....	97
4.1.1 Dados estatísticos	98
4.1.2 Análise.....	98
4.2 Atividades para o texto 03.....	109

Considerações finais.....	112
Referências bibliográficas	115
Glossário.....	119
Apêndices	122
Apêndice 1 - Formulário de autorização e questionário para formação do perfil dos participantes.....	123
Apêndice 2 – Modelo da versão digital da atividade de tradução enviada aos alunos por e-mail.....	124
Apêndice 3 - Texto 3 como disposto na planilha do <i>Excel</i>	125
Apêndice 4 - Texto 3 na LP etiquetado.....	126
Apêndice 5 - Texto 3 – Lista das 40 primeiras palavras-chave em inglês, em ordem decrescente de chavicidade	127
Apêndice 6 - Texto 3 – Lista das 80 primeiras palavras-chave em português, em ordem decrescente de chavicidade	128
Anexo	129

Índice das figuras

Figura 1.1 Nível de densidade lexical com base na razão entre <i>types</i> e <i>tokens</i> no <i>corpus</i> em inglês	20
Figura 1.2 Nível de densidade com base na razão entre <i>types</i> e <i>tokens</i> no <i>corpus</i> em português.....	20
Figura 1.3 Comparação de lista de palavras com lista de palavras-chave no <i>corpus</i> de medula óssea.....	22
Figura 1.4 Exemplo de linhas de concordância com a palavra ‘cidades’ em nosso <i>corpus</i> de estudo	23
Figura 1.5 Reorganização (<i>sorting</i>) das linhas de concordância no <i>AntConc</i>	24
Figura 1.6 Reorganização (<i>sorting</i>) das linhas de concordância no <i>WordSmith Tools</i>	24
Figura 1.7 Parâmetros de reorganização das linhas de concordância no <i>AntConc</i>	25
Figura 1.8 Linhas de concordância da palavra ‘efeito’ reorganizadas 1L, 1R, 2L no <i>AntConc</i>	26
Figura 2.1 Lista de palavras geradas a partir do texto na LP.....	39
Figura 2.2 Passos para a conversão de texto em tabela no Word do Microsoft Office	40
Figura 2.3 Resultado após o comando ‘Inserir coluna’	40
Figura 2.4 Inserção de coluna à direita para disponibilizar o espaço onde a tradução será feita	41
Figura 2.5 Planilha do Excel com o texto na LP e as traduções dos alunos	41
Figura 2.6 Lista de palavras do texto na LP e das traduções	42
Figura 2.7 Dez primeiras linhas de concordância da palavra ‘pode’ nas traduções	43
Figura 2.8 Comando ‘Colar’ com o submenu ‘Transpor’	44
Figura 2.9 Ferramenta ‘Classificar e Filtrar’ no Excel	45
Figura 2.10 Resultado do comando ‘Colar especial - transpor’ com a coluna B em ordem alfabética	45
Figura 2.11 Resultado do comando ‘Colar especial - transpor’ com a coluna C em ordem alfabética	46
Figura 3.1 Programa gratuito <i>Easy Duplicate Finder</i> que identifica textos iguais.....	56
Figura 3.2 Ferramenta que busca por textos iguais no <i>WordSmith Tools</i>	57
Figura 3.3 Esquema das inúmeras traduções de um único texto de partida.....	58
Figura 3.4 Identificação das células na planilha eletrônica	61
Figura 3.5 Exemplo de fórmula na planilha eletrônica	61
Figura 3.6 Exemplos de fórmulas aplicadas a texto	62
Figura 3.7 Menu disponível ao clicar na guia da planilha com o botão direito do mouse	64
Figura 3.8 Distribuição das planilhas em um arquivo	64
Figura 3.9 Exemplo da planilha ‘Lista de alunos’ com as respostas dos questionários tabuladas	65
Figura 3.10 Totais da tabulação e fórmulas para contagem automática	67
Figura 3.11 Distribuição das etiquetas nas colunas CMTr.	69
Figura 3.12 Exemplo do cabeçalho do texto de partida na planilha	70

Figura 3.13 Exemplo da camada superior do cabeçalho das traduções	71
Figura 3.14 Fórmula na planilha 'Texto xx' para buscar informações na planilha 'Lista de alunos'	71
Figura 3.15 Exemplo de fórmula que buscará a identificação do aluno na planilha 'Lista de alunos'	72
Figura 3.16 Equivalência numérica de cada coluna e extensão de onde começa e termina a extração dos dados	73
Figura 3.17 Uso da fórmula '=(coluna)(número da linha)' para copiar o conteúdo de uma célula em outra	74
Figura 3.18 Significado de cada elemento da etiqueta de abertura.....	75
Figura 3.19 Significado de cada elemento das etiquetas das linhas	75
Figura 3.20 Primeiras linhas etiquetadas tanto no CMTP quanto no CMTr	76
Figura 3.21 Passos para usar o comando 'colar especial' no <i>Excel</i>	77
Figura 3.22 Uso de código para indicar que o aluno dividiu um segmento em dois períodos na língua de chegada.....	78
Figura 3.23 Exemplo do resultado da aplicação de fórmulas para trazer dados da planilha 'Texto xx'	79
Figura 3.24 Processo de divisão do arquivo 'Textoxx' em traduções individuais	81
Figura 3.25 Processo de renomeação dos arquivos no <i>WordSmith Tools</i>	82
Figura 3.26 Eliminação dos arquivos que não contêm tradução.....	83
Figura 3.27 Exemplo de tradução com cabeçalho e etiquetagem	83
Figura 3.28 Todas as traduções do segmento do texto de partida alinhadas verticalmente	85
Figura 3.29 Janela de configurações no <i>WordSmith Tools</i>	87
Figura 3.30 Lista parcial de palavras gerada no <i>WordSmith Tools</i>	88
Figura 3.31 Lexema organizado na lista de palavras.....	89
Figura 3.32 Aba <i>Statistics</i> mostra, parcialmente, em destaque a razão entre <i>types</i> e <i>tokens</i> do texto 07	90
Figura 3.33 Janela para selecionar apenas determinadas linhas dos textos	91
Figura 3.34 Lista de palavras apenas da linha 3 do texto 07	92
Figura 3.35 Janela do <i>WordSmith Tools</i> onde se pode selecionar os textos a partir das informações no cabeçalho	93
Figura 3.36 Lista de palavras da linha 03 do texto 07 após a aplicação das etiquetas '<years_abroad>t_abr_none' e '<english_background>eng=bas' do cabeçalho usadas como filtros	94
Figura 3.37 Dez primeiras linhas de concordância da palavra 'energia'	94
Figura 4.1 Comparação das listas de palavras em inglês e português de acordo com a frequência.....	100
Figura 4.2 Comparação das listas de palavras-chave em inglês e português	101
Figura 4.3 Distribuição do pronome 'ela' na ferramenta 'Plot' no <i>WordSmith Tools</i> ...	102
Figura 4. 4 Resultado do agrupamento das palavras 'lindo', 'belo' e 'bonito' utilizando a ferramenta <i>Lemma Forms</i>	103
Figura 4.5 Resultado da aba <i>Collocates</i> da palavra 'dia' com 'lindo', 'belo' e 'bonito' 104	
Figura 4.6 Excerto das linhas de concordância da colocação 'belo dia' no <i>Corpus</i> do Português	105
Figura 4.7 Amostra das soluções encontradas para <i>lurching vehicle</i>	107
Figura 4.8 Amostra das soluções encontradas para traduzir <i>approaching landmark</i> .	108

Índice dos gráficos

Gráfico 3.1 Faixa etária dos participantes.....	50
Gráfico 3.2 Sexo dos participantes.....	51
Gráfico 3.3 Nível de proficiência em inglês dos participantes.....	51
Gráfico 3.4 Alunos com vivência em países anglófonos.....	52
Gráfico 3.5 Língua nativa dos participantes.....	53
Gráfico 3.6 Experiência dos participantes em trabalhos de tradução.....	53
Gráfico 3.7 Participantes com outras áreas de formação.....	54

Índice das tabelas

Tabela 3.1 Distribuição de textos traduzidos por cada texto-fonte.....	57
Tabela 4.1 Total de ocorrências do pronome ‘ela’ nas traduções	102
Tabela 4.2 Tabulação dos fraseologismos com a palavra ‘dia’	104
Tabela 4.3 Tabela comparativa das colocações com a palavra ‘dia’ no <i>corpus</i> de estudo e no <i>Corpus</i> do Português	105

Lista de abreviaturas

LP – Língua de partida

LC – Língua de chegada

CMTP – Conjunto matriz do texto de partida

CMTr – Conjunto matriz das traduções

Resumo

Os estudos sobre o ensino da tradução no Brasil ainda oferecem muito espaço para discussões. Valendo-se disso, este trabalho traz como um de seus objetivos uma breve reflexão sobre a sala de aula e sugere um possível método de ensino de tradução baseado na análise de material produzido por tradutores-aprendizes. A intenção é que, por meio da Linguística de *Corpus*, consigamos observar o processo de construção do texto de chegada pela ótica do aluno, nos mesmos moldes que o fazemos ao analisar material produzido por aprendizes de idiomas.

Para tanto, compilamos um *corpus* de aprendizes de tradução, constituído por oito textos originais e cerca de 100 traduções para cada um deles. Alinhar tantas traduções referentes a um original de modo a permitir análises não foi tarefa fácil. A estratégia empregada para superar tal dificuldade foi o desenvolvimento de uma metodologia específica de alinhamento tendo como ferramenta planilhas eletrônicas. Tal metodologia tornou-se o foco central desta pesquisa.

A utilização de fórmulas para a manipulação de dados textuais na planilha eletrônica resultou em um *corpus* alinhado, com todos os textos de partida e suas referidas traduções com cabeçalhos e com todas as linhas etiquetadas. Esse procedimento possibilitou a organização de um *corpus* para ser analisado tanto no editor de planilhas eletrônicas quanto em programas como *AntConc* e *WordSmith Tools*. Além disso, também apresentamos a planilha eletrônica como uma ferramenta didática para ser usada nas aulas de prática de tradução.

Palavras-chave: *corpus* de aprendizes de tradução, Linguística de *Corpus*, alinhamento automático, planilha eletrônica, cabeçalho e etiquetagem

Abstract

Studies on the teaching of translation in Brazil still offer room for discussions. Having that in mind, one of the goals of this research aims at fostering a brief reflection upon the classroom and proposes a teaching method based on the analyses of material produced by translation learners. We show that Corpus Linguistics can be used to analyze student translations in the same way we do when we analyze material produced by language learners.

For that purpose, we compiled a corpus of translations produced by learners, consisting of eight source texts in English and about 800 translations into Portuguese, approximately 100 for each text. Aligning so many translations to their original texts to favor analyses was not a simple task. Such difficulties were overcome by the development of a methodology for alignment, which became the central focus of this research.

By utilizing formulas to deal with textual data in spreadsheets resulted in an aligned corpus containing source texts and their referring translations with headers and all lines tagged. Such procedure allowed us to come up with a corpus to be analyzed in both the spreadsheet editor and in programs such as AntConc and WordSmith Tools. In addition to that, we also introduced the spreadsheets as a didactic tool to be used in translation practice classes.

Keywords: translation learner corpus, corpus linguistics, automatic alignment, spreadsheet, header and tagging

Introdução

A busca por um texto na língua de chegada, isto é, o idioma para o qual se traduz (doravante referida apenas como LC), que ao mesmo tempo transmita os significados presentes na mensagem da língua de partida, ou seja, a língua da qual se traduz (doravante referida apenas como LP), e que seja fluente a ponto de não torná-lo ininteligível ao leitor tem sido, há muito tempo, o objetivo de muitos tradutores. Como atingir tal objetivo tem gerado ardentes discussões entre tradutores e estudiosos, tanto nos meios acadêmicos quanto profissionais. Além disso, no campo da tradução ainda há vários pontos que oferecem oportunidades de ponderações como, por exemplo, as discussões sobre o que é a fidelidade textual; a lealdade textual; a dicotomia ‘fiel-livre’; a linha divisória entre tradução literária e tradução técnica; os parâmetros para avaliar uma tradução; a tecnologia ora como aliada, ora como inimiga do tradutor; etc., isso sem considerar a discussão proposta por Lawrence Venuti (1995 e 1998) sobre a virtude ou o infortúnio de um texto traduzido ser ou não domesticado à cultura da LC, o que, certamente, traz à cena reflexões sobre as questões éticas por trás de cada ato de tradução. E é nesse terreno ainda muito movediço que se insere esta pesquisa, cujo objetivo é, por meio das ferramentas e metodologias oferecidas pela Linguística de *Corpus*, propor uma reflexão sobre alguns processos de ensino da tradução a alunos universitários que aspiram à profissão de tradutor.

A história do ofício da tradução no Brasil desde a época da colonização portuguesa até os dias atuais é descrita por Wyler (2003). Os profissionais inicialmente eram chamados de ‘línguas’, depois de ‘poetas’ e, por fim de ‘bacharéis em tradução’, ou seja, aqueles que ingressam no nível superior de educação com o objetivo de obter formação acadêmica na área. Tanto o processo de instrução quanto o produto (textos traduzidos) desses bacharéis, futuros profissionais da tradução, que denominaremos ‘tradutores-aprendizes’, serão objetos de reflexão em nosso estudo. Nosso norte se mantém em oferecer a pesquisadores e professores de tradução material que possibilite identificar e sistematizar os procedimentos adotados pelos alunos durante as atividades de tradução e seus reflexos no texto traduzido.

Os primeiros cursos superiores de formação de bacharéis em tradução e interpretação no Brasil foram abertos cerca de 30 ou 40 anos atrás. Em muitos

centros universitários da rede privada esses cursos são classificados como uma ramificação da licenciatura em Letras. O INEP – Instituto Nacional de Estudos e Pesquisas Educacionais¹ – não faz qualquer menção à habilitação em tradução ao convocar os alunos de letras para prestarem o ENADE – Exame Nacional de Desempenho de Estudantes –, que avalia os estudantes egressos dos cursos superiores no Brasil. Vale ressaltar que o reconhecimento de habilitações diferentes de um mesmo curso é algo que o INEP já faz. Por exemplo, a área de Comunicação Social oferece duas provas distintas: uma para a habilitação em jornalismo e outra para a habilitação em publicidade e propaganda. O mesmo ocorre com a avaliação do curso de Engenharia, em que há oito modalidades de provas para contemplar suas diferentes habilitações. Entretanto, o ENADE – Letras de 2011² ofereceu apenas uma modalidade de prova aos alunos de letras. Havia uma seção denominada ‘Componente Específico – Bacharelado’, porém em nenhum momento havia alguma menção a algo sobre tradução. A falta de uma prova específica para o bacharelado em tradução talvez seja um reflexo de que a alta gestão de alguns centros universitários não identifique as particularidades de tal curso.

Apesar de ser um ofício antiquíssimo, a aptidão de traduzir é muitas vezes associada, por leigos, à capacidade de falar outro idioma. É fácil encontrar professores de língua estrangeira atuando como tradutores e não menos incomum ver ‘falantes’ de um idioma estrangeiro, sem nenhuma formação específica, apresentando-se como profissionais da tradução. Claro que o tradutor, assim como o professor de idiomas, tem de ter um conhecimento linguístico bem amplo, o que torna as duas profissões muito próximas. Porém, o tradutor necessita desenvolver técnicas bem específicas que lhe permitam navegar entre dois idiomas de modo suave.

Esta pesquisa não vê o mister da tradução como um dom nato de quem fala uma língua estrangeira, mas sim como uma habilidade que pode ser aprendida, desenvolvida, praticada e aprimorada, não sendo apenas uma simples transferência de uma mensagem a outro idioma, como pensam os leigos. Nida (1975: 190) compara o processo tradutório com a transferência de

¹ Disponível em: <http://portal.inep.gov.br/enade>.

² A prova de 2011 está disponível para download em: http://download.inep.gov.br/educacao_superior/enade/provas/2011/LETRAS.pdf. Acesso em março de 2013.

carga de um trem para outro, ficando claro que não importa o quanto de carga cada vagão leve, contanto que toda a carga chegue a seu destino final. Não há como discordar de tal descrição se, metaforicamente, encararmos toda a significação do texto na LP como se fosse uma carga que deve ser levada em sua integridade (se isso for mesmo possível) à LC. Porém, entendemos que a forma como esse ‘carregamento’ será alocado no trem não pode ser aleatória; o tradutor deve decidir a melhor maneira de distribuir e acondicionar a carga nos vagões. Levý (1967) vê a tradução como um processo de tomada de decisões que podem ser estimuladas pelas várias interpretações do texto na LP e pelas diversas possibilidades lexicais na LC.

Acreditamos que um tradutor cumpre o seu papel quando as funções esperadas para o texto na LC são cumpridas e quando o leitor percebe o texto traduzido de modo natural, sem estranhezas que poderiam prejudicar a compreensão. É assim que nos posicionamos quanto ao ato tradutório para esta pesquisa, e, neste momento, não temos espaço para refletirmos sobre as considerações de Venuti (1995), que fala sobre a ilusão de transparência quando o tradutor se esforça para se certificar de que seu texto seja de fácil leitura. Até entendemos que tais estranhezas muitas vezes formam a ‘personalidade’ de uma tradução, mas, como dito, esta pesquisa não pretende advogar nem pôr em discussão se o texto na LC deve ou não ter traços da LP, se o texto deve ou não ser domesticado, se o leitor do texto da LC deve ou não perceber que se trata de um texto traduzido etc.

Nosso foco está em tentar identificar mecanismos de escolha adotados pelos tradutores-aprendizes. Da mesma forma que a distribuição de carga entre diferentes vagões proposta por Nida (1975) não deve ser feita de maneira fortuita, as escolhas sugeridas por Levý (1967), a nosso ver, não deveriam ser baseadas apenas na intuição linguística do tradutor, profissional ou aprendiz. A Linguística de *Corpus* surge como uma ferramenta que pode direcionar as escolhas do profissional, assegurando-lhe, por meio de quantificações estatísticas e estudos qualitativos, um levantamento das opções preferidas pelos falantes nativos da LC.

Ao se ter como respaldo os resultados de pesquisas e análises de levantamentos oriundos de *corpora* específicos, a avaliação qualitativa da tradução começa a deixar o estágio subjetivo, que tem como base o recorte

linguístico do tradutor, pois a Linguística de *Corpus* torna viável ao profissional da tradução a possibilidade de avaliar suas escolhas de modo mais objetivo. Ao analisar o índice de frequência de ocorrência de determinados itens lexicais, ele pode se certificar das convencionalidades e dos idiomatismos característicos de gêneros textuais específicos de um idioma. Por exemplo, ao traduzir um contrato de aluguel residencial, o tradutor depara com três possíveis equivalentes para o termo *landlord* em inglês: Mello (2009: 822) traz em seu dicionário ‘Senhorio; Proprietário’ de terra; Castro (2010: 752) traz apenas ‘locador’. Para escolhermos o equivalente que melhor se adequasse ao gênero jurídico na área contratual, utilizamos um *corpus* compilado a partir de textos extraídos da Internet para uso em nossas aulas. Tal *corpus* é composto de 154 contratos de aluguel residencial escritos originalmente em português, com 212.902 palavras. Em nossas análises, observamos que ‘locador’ aparece 2.114 vezes, ‘proprietário’, 89, ao passo que ‘proprietário de terra’ e ‘senhorio’ não aparecem em nenhum dos contratos. O termo ‘proprietário de terra’ nem seria uma opção por se tratar de um contrato de aluguel residencial; ‘senhorio’, mesmo ocorrendo frequentemente em legendas de filmes, também não é uma escolha adequada para contratos de aluguel, segundo nosso *corpus*. Ao consultar o *corpus*, o tradutor, então, tem a certeza de que a escolha por ‘locador’ é a mais adequada, pois agora não são apenas as ‘duas vozes’ dos dicionários falando, mas as ‘154 vozes’ que formam o *corpus*.

Outro fator a ser levado em conta é a avaliação das traduções. O professor de prática de tradução tem o papel de crítico e de revisor, sendo aquele que julga se uma tradução está ‘boa’, ‘regular’ ou ‘ruim’. Subjetividade na avaliação está quase sempre presente e não é raro encontrarmos dois professores com opiniões divergentes quanto à qualidade do texto na LC. Ao propormos uma análise mais objetiva, a Linguística de *Corpus* pode ser aplicada para identificar e quantificar as escolhas feitas pelos alunos-aprendizes, tornando factível uma sistematização para a avaliação do texto traduzido no ambiente escolar. Mais à frente, Capítulo 4, veremos como as linhas de concordância podem nos ajudar nesse processo.

É nesse contexto acadêmico de formação e preparação de futuros tradutores que se insere nossa pesquisa. A Linguística de *Corpus* passa a ser

utilizada com uma das abordagens de ensino que nos revela as tendências das escolhas que estão por trás da aprendizagem do ofício da tradução.

A justificativa e a motivação para a compilação do *corpus* têm como base o fato de que, com raras exceções, o ensino de tradução nos cursos de graduação das faculdades privadas na cidade de São Paulo é feito sem uma formação pedagógica específica. As aulas de prática de tradução são divididas em gêneros como tradução jornalística, jurídica, literária etc. e geralmente focam no treinamento do futuro profissional por meio de uma quantidade excessiva, e às vezes exaustiva, de exercícios práticos, a fim de expor os alunos a várias problemáticas textuais. Além de selecionar os textos, que são traduzidos em casa e, posteriormente, discutidos durante a aula, o professor atua como revisor e crítico das traduções apresentadas pelos alunos; na maior parte das vezes, suas correções são baseadas em sua intuição linguística e experiência profissional. A abordagem é dogmática e a avaliação seguida é a da ‘tentativa de acerto e erro’, com soluções bem pontuais e específicas do texto que está sendo traduzido. Ou seja, sendo o professor o referente linguístico e cultural, a aula é baseada numa via de mão única que parte de sua experiência como tradutor. Assim, o professor escolhe o texto a ser trabalhado e decide quais os pontos serão discutidos. Uma reflexão inevitável que emerge aqui é se com esse modelo pedagógico estamos realmente atingindo as verdadeiras necessidades dos alunos. Será que o aluno egresso dessa faculdade consegue desenvolver as habilidades que o capacitarão para entrar no mercado profissional de tradução?

Há um bom tempo, o estudo de *corpora* compostos a partir de material produzido por aprendizes de idiomas vem revolucionando o ensino de línguas, pois possibilita ver a aprendizagem pela perspectiva do aluno. A identificação, quantificação e classificação dos erros e acertos permitem ao professor direcionar suas aulas de acordo com as necessidades reais dos alunos. Os estudos de *corpora* de falantes nativos e de aprendizes já se tornaram tão importantes e essenciais ao ensino da língua inglesa que passaram a fazer parte da elaboração dos livros didáticos publicados por editoras como Pearson Longman, Macmillan, OUP, CUP, entre outras. Da mesma forma, acreditamos que um trabalho que envolve a compilação, a organização e a disponibilização de um *corpus* de aprendizes de tradução venha contribuir com as pesquisas

sobre seu ensino e aprendizagem, sendo de grande valia para professores e estudiosos.

Para tanto, compilamos um *corpus* paralelo e alinhado, cuja característica principal é a disposição de sequências formadas por um texto em inglês com as referidas traduções (de 20 a 50 alunos) em português. A utilização desse *corpus* de aprendizes de tradução possibilita a identificação e o mapeamento das tendências de escolhas lexicais e morfossintáticas dos tradutores-aprendizes. Um estudo aprofundado desse *corpus* pode trazer respostas a questionamentos como: a) a estrutura de determinadas unidades fraseológicas no texto da LP gera erros na LC?; b) há ou não padrões e recorrências em erros e acertos?; c) certos itens morfossintáticos e lexicais geram dificuldades?; entre outros. Tais perguntas são o ponto-chave para auxiliar professores de tradução na preparação de cursos e aulas que sejam mais adequadas às necessidades do futuro profissional da tradução.

O acesso a um *corpus* de aprendizes de tradução oferece ao professor a possibilidade de entender melhor o processo cognitivo envolvido no ensino e aprendizagem do aluno. A partir do momento em que o professor se conscientiza de tal processo, passa a ter os subsídios necessários para propor e elaborar um projeto pedagógico para o curso de tradução que contemple a formação do tradutor de modo a suprir as necessidades e eliminar as deficiências do aluno e, ao mesmo tempo, desenvolver suas habilidades tradutórias.

Citando Fillmore (1979), Alves e Tagnin (2009) comparam o tradutor-aprendiz ao falante ingênuo e apontam evidências de que esses alunos tendem a focar em questões microtextuais, dedicando grande parte de seu tempo a resolver problemas pontuais de ordem lexical. Uma sugestão para ajudar o aluno a ter uma visão mais macrotextual seria expô-lo a um *corpus* de aprendizes de tradução, estimulando reflexões, principalmente sobre como as unidades lexicais do texto original foram traduzidas, destacando a variedade de escolhas dos alunos.

Segundo Sinclair (1996 e 2004), as palavras observadas isoladamente não conseguem expressar todos os significados possíveis em um idioma. Para transmitir certos conceitos, o falante combina palavras que formam unidades de significado. Tais unidades de significado são formadas por combinações de

duas ou mais palavras. Essas combinações, também chamadas de fraseologismos, não são aleatórias e ocorrem com certa frequência, podendo ser colocações (palavra lexical + palavra lexical), coligações (palavra lexical + palavra gramatical), expressões idiomáticas etc. O tradutor-aprendiz ingênuo muitas vezes não identifica tais unidades como um bloco e tenta transferir seu significado traduzindo cada palavra individualmente. Por exemplo, em um texto em que uma pessoa aconselha a outra a não se precipitar, e surge a expressão idiomática *Cross that bridge when you come to it*, esse aluno, não compreendendo que se trata de uma única unidade de significado, arriscaria uma tradução literal que o levaria a algo como ‘Cruze aquela ponte quando chegar a ela’, enquanto a expressão com equivalência de significado esperada em português seria ‘Não coloque o carro na frente dos bois’.

Ao preparar suas aulas com base em um *corpus* de aprendizes de tradução, o professor tem a possibilidade de detectar se os alunos, ao lerem o texto de partida, conseguem ou não identificar as unidades de significado. Some-se a isso, também, a intenção do autor ao produzir o texto na LP (Reiss 2000). Com esse procedimento, o docente tem condições de elaborar exercícios de leitura do texto na LP que trabalhem a identificação das fraseologias, trazendo à tona uma reflexão de como tais significados podem ser representados no texto da LC. Para complementar, o aluno pode ser exposto às soluções encontradas por seus colegas de sala e compará-las com suas próprias escolhas.

Uma reflexão proposta por Bernardini (2000 e 2002) é se devemos treinar ou capacitar os futuros tradutores em cursos de formação profissional. Segundo a pesquisadora, enquanto o treinamento forma tradutores para lidarem com determinadas situações, um curso que foca na capacitação faz com que os alunos se tornem conscientes de seus próprios processos de aprendizagem e de seu desenvolvimento profissional, estimulando, assim, a sua autonomia. Nesta pesquisa, acreditamos que tanto a utilização na preparação das aulas quanto a exposição dos alunos a um *corpus* de aprendizes propiciem um curso que priorize a capacitação do futuro profissional da tradução.

O ensino de tradução no Brasil ainda dá seus primeiros passos. Muitos professores da área contam apenas com sua experiência para preparar as

aulas; por isso, acreditamos que o acesso a um *corpus* de aprendizes possa contribuir com o avanço dos estudos sobre o ensino da tradução. Até o final da elaboração desta pesquisa, tínhamos conhecimento de apenas um projeto similar chamado *MeLLANGE Project*³, que foi encerrado em 2007, apesar de ainda estar disponível para acesso e consultas. Os objetivos do projeto consistiam em elaborar uma metodologia para a criação de material para o ensino de tradução com base em *corpus* e de coordenar programas de mestrado dentro da comunidade europeia. Para tanto, uma das possibilidades oferecidas era a de pesquisar um *corpus* paralelo de aprendizes de tradução compilado, pelo projeto, em inglês, catalão, francês, espanhol, alemão e italiano, levando em conta o perfil linguístico dos alunos e a tipificação de erros.

Os objetivos deste trabalho estão divididos em duas fases. Na primeira, fizemos a compilação, organização e alinhamento de um *corpus* de aprendizes de tradução em nível universitário. O *corpus* é formado por um conjunto de textos em inglês alinhados às suas respectivas traduções produzidas pelos alunos. Cada aluno recebe uma identificação numérica e são classificados por sexo, idade, conhecimento da língua nativa e estrangeira e experiência como tradutor. Para efeitos desta pesquisa, classificamos nosso *corpus* como *fechado*, ou seja, não receberá acréscimos no futuro. Entretanto, esperamos que, ao seu término, o produto final seja disponibilizado para pesquisadores nos mesmos moldes do projeto CoMAprend⁴, que permite o acesso aos *corpora* de aprendizes de idiomas, e que outras traduções sejam acrescentadas.

Na segunda fase, aplicamos as ferramentas da Linguística de *Corpus* ao *corpus* compilado e apresentamos um estudo analítico das soluções encontradas pelos alunos em textos selecionados. Por meio de uma metodologia *corpus-driven*, ou seja, orientada pelo *corpus*⁵, estudamos as traduções apresentadas, tentando identificar e entender as opções feitas pelos tradutores-aprendizes. Tais levantamentos servem-nos para identificar quais aspectos presentes no texto na LP motivaram as escolhas lexicais e morfossintáticas dos alunos. Muitas vezes o tradutor-aprendiz iniciante não

³ Disponível em: <http://mellange.eila.jussieu.fr/>. Acesso em: maio de 2014.

⁴ Disponível em: <http://www.fflch.usp.br/dlm/comet/comaprend.html>. Acesso em: dezembro de 2013.

⁵ No capítulo 4 distinguiremos as duas formas de abordar um *corpus*: *corpus-based*, com base no *corpus*, e *corpus-driven*, orientada pelo *corpus*.

consegue identificar certas combinações lexicais como idiomáticas ou como colocações específicas da língua de partida. Por exemplo, o *phrasal verb* ‘*start out*’, como em *The day started out rotten*, tem uma nuance em seu significado sobre algo que começa de uma forma, mas que sofre uma transformação no decorrer da ação. O exemplo foi retirado de um texto em que o dia da personagem começa bem ruim, mas tem uma mudança durante a viagem de volta para casa. Ao compararmos todas as soluções encontradas pelos alunos, podemos usar os dados resultantes dessas análises e propor atividades de *lead-in*, que visam preparar os alunos antes de iniciarem a tarefa da tradução, ou de *follow-up*, que oferecem aos alunos a oportunidade de retomar as escolhas feitas nos textos e julgá-las como adequadas ou não. O objetivo das atividades é levar os alunos a um processo de leitura crítica e reflexão sobre o texto traduzido.

‘Como a compilação e a análise de um *corpus* de aprendizes de tradução podem contribuir na elaboração do projeto pedagógico dos cursos de graduação para a formação de tradutores?’ é a questão que norteia a proposta desta pesquisa científica. Quando comparada a outras áreas do conhecimento, a Teoria da Tradução ganhou sua condição oficial de disciplina há relativamente pouco tempo, algo como pouco mais de três décadas. Mesmo assim, a tentativa de se estabelecerem métodos e procedimentos que sistematizem o processo tradutório sempre esteve presente em sua história. Não queremos questionar as abordagens utilizadas para o processo da tradução, sejam elas mais próximas ao texto de partida ou se privilegiam o texto na LC, ou ainda se o tradutor pode ou não aplicar um processo criativo na elaboração do texto na LC. Apesar de adotarmos uma postura que privilegia uma visão funcionalista para a tradução (Nord, 2000 e 2005), tanto como tradutor quanto professor de prática de tradução, neste momento o que queremos é propor uma discussão sobre como a Linguística de *Corpus* pode se tornar uma ferramenta essencial na elaboração e organização de cursos que visam à formação do futuro tradutor, independentemente do estilo de tradução do professor ou da abordagem teórica adotada.

Outra pergunta a ser considerada é como os professores de tradução podem se beneficiar dos resultados de análises e estudos de um *corpus* de textos procedentes de tradutores-aprendizes. Ao se estabelecer uma

metodologia de análise das traduções feitas pelos tradutores-aprendizes, abre-se a possibilidade para professores e pesquisadores identificarem e compreenderem os processos cognitivos necessários pelos quais os alunos devem passar até atingirem o ponto em que possam ser considerados profissionais formados.

Além de aprender as técnicas e os procedimentos para a tradução, o tradutor-aprendiz ainda precisa adquirir um conhecimento linguístico que lhe permita julgar se suas escolhas lexicais estão ou não de acordo com o gênero textual e o assunto a serem traduzidos. Nas traduções de caráter não literário, a questão da terminologia faz-se necessária para que o texto esteja de acordo com o que se espera em determinada área. Ao comparar a terminologia encontrada no *corpus* de aprendizes com aquela encontrada nos *corpora* específicos, o professor pode auxiliar o aluno a ajustar suas escolhas lexicais de acordo com determinada área técnica. O mesmo vale para as estruturas morfossintáticas mais comuns da área técnica em questão. Por meio da pesquisa proposta neste projeto, pretendemos discutir como a Linguística de *Corpus* pode contribuir com a sistematização de um processo técnico de levantamentos lexicais e morfossintáticos cujos resultados permitam ao tradutor fazer escolhas com mais embasamento, sem necessariamente alterar seu estilo de tradução.

O *corpus* disponibilizado, aliado a um modelo de análise das traduções e à proposta de atividades, pode, a princípio, ser utilizado de duas maneiras pelos pesquisadores: de forma direta, ou seja, o professor baixa o *corpus*, faz os estudos desejados utilizando um programa como o *AntConc*, e, se desejar, utiliza as atividades sugeridas; ou de forma aplicada, usando tanto o *corpus* quanto as análises propostas nesta pesquisa como um modelo de diagnóstico para identificação das características das traduções apresentadas por seus próprios alunos. De uma forma ou de outra, a intenção é oferecer um modelo de pesquisa que possa trazer benefícios para o ensino da tradução.

Para tanto, apresentamos no Capítulo 1 a fundamentação teórica que nos dará respaldo ao longo do trabalho. Discutimos o que é Linguística de *Corpus*, os tipos de *corpora*, que resultados podem ser obtidos e a presença da Linguística de *Corpus* nos estudos da tradução. No Capítulo 2 apresentamos um breve relato sobre o ensino de tradução e explanamos como planilhas

eletrônicas podem ser usadas em aula de modo a facilitar o cotejo de todas as traduções feitas pelos alunos para um texto na LP. No Capítulo 3 descrevemos em detalhe uma grande problemática tecnológica que encontramos para alinhar diversas traduções e como uma planilha eletrônica mostrou-se absolutamente eficaz e prática não só para etiquetar o *corpus* como para alinhar o texto na LP com suas diversas traduções, permitindo uma visão macro do seu conteúdo. Também nesse capítulo descrevemos alguns procedimentos no uso da ferramenta digital *WordSmith Tools* e como o cabeçalho das traduções permite que o programa filtre o *corpus* e selecione apenas os textos que desejamos analisar. No Capítulo 4 apresentamos as análises a partir do *corpus* e sugestões de exercícios para serem aplicados aos alunos. Por fim, nas Considerações Finais retomamos os principais pontos discutidos e apresentamos nossas conclusões. Após as Referências Bibliográficas encontram-se três seções: uma com um glossário de alguns termos recorrentes neste trabalho, uma de apêndices com arquivos que, por sua extensão, não se encaixariam no corpo do texto dos capítulos, e uma de anexo com um gráfico de faixas etárias da população brasileira.

Capítulo 1 – Fundamentação teórica

1.1 O que é Linguística de *Corpus*

A Linguística de *Corpus* será o arcabouço teórico da investigação proposta nesta pesquisa. Uma vez que a Linguística de *Corpus*, como entendida hoje, é associada à tecnologia, seu grande desenvolvimento se dá nos anos 80 do século XX, em decorrência dos avanços na área da computação, principalmente do acesso aos computadores pessoais. Ainda recente no Brasil, a Linguística de *Corpus* começa a tomar fôlego na virada do século. Berber Sardinha e Almeida (2008) destacam os *Encontros de Linguística de Corpus*, que tiveram início na Universidade de São Paulo em 1991, como um meio de intercâmbio e parcerias entre pesquisadores e instituições brasileiras e estrangeiras. Berber Sardinha (2004) define a Linguística de *corpus* como a coleta e a exploração de *corpora* com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. McEnery e Wilson (2001:5) simplificam e a descrevem como o estudo de língua baseado em exemplos do idioma em seu uso ‘na vida real’. McEnery e Hardie (2012) definem Linguística de *Corpus* não como o estudo de um aspecto de uma língua em especial, mas, sim, como uma área que enfoca no conjunto de procedimentos ou métodos para estudar o idioma. Krieger (2003) ressalta que a Linguística de *Corpus* se concentra na descoberta de padrões no uso do idioma e que uma análise baseada em *corpus* não gera teorias, mas evidencia as probabilidades de uma língua. Entretanto não podemos deixar de ressaltar que evidências de padrões e recorrências poderiam, sim, gerar pressuposições e até mesmo teorias linguísticas.

Como o próprio nome diz, a base da Linguística de *Corpus* é, obviamente, o *corpus*, ou seja, uma coleção de textos. Quanto ao verbete ‘*corpus*’, tanto a edição impressa de 2001 quanto a edição digital 3.0 do dicionário da Língua Portuguesa do Antônio Houaiss trazem a seguinte definição:

3 Rubrica: linguística.

conjunto de enunciados numa determinada língua, ger. colhidos de atos reais da fala, que servem de material para análise linguística

As definições ‘atos reais de fala’ e ‘análise linguística’ explicam com precisão o que podemos obter a partir de um *corpus*, ou seja, uma amostra de recortes linguísticos que nos permite supor, confirmar ou rechaçar hipóteses sobre o uso do idioma. Tagnin (2013: 21) explica que um *corpus* é uma coletânea de textos, necessariamente em formato eletrônico, compilados e organizados segundo critérios ditados pelo objetivo de pesquisa a que se destina. O formato eletrônico se faz necessário devido ao fato de que, atualmente, não há como pensar em pesquisa e exploração de *corpora* por outros meios que não os digitais. Para tanto, há vários programas computacionais, entre eles destacamos dois que são citados neste trabalho: *WordSmith Tools* e *AntConc*. Além do uso de ferramentas digitais específicas, os critérios estabelecidos para compilação são importantes, pois um *corpus* não é apenas uma coleção de textos agrupados sem parâmetros. Teóricos como Kilgarriff e Grefenstette (2003) são mais flexíveis e consideram a *Wide-World-Web* como um *corpus* gigantesco que pode ser pesquisado com um foco específico. Já Sinclair (2005), apesar de reconhecer a importância da WWW como fonte de material, não a considera um *corpus*, pois seus textos não foram coletados a partir de uma perspectiva linguística.

1.1.1 Tipologia dos *corpora*

A tipologia dos *corpora* é vasta e, dentre os critérios usados para sua classificação, podemos destacar formato, idiomas envolvidos, dinamismo, conteúdo, origem e disposição dos textos:

- a. o formato pode ser tanto escrito quanto falado. Os textos que formam um *corpus* escrito podem ser oriundos de *sítes* da Internet, de arquivos escaneados a partir de fontes impressas, ou escritos diretamente em editores de texto em computadores ou outros dispositivos digitais. Um *corpus* falado é formado pelas transcrições das falas gravadas e compiladas. Tais transcrições seguem normas já estabelecidas para representarem, de forma escrita, as pausas, hesitações, ruídos etc. comuns no discurso falado;

- b. o número de idiomas presentes determina se o *corpus* é monolíngue – composto por textos em um só idioma –, bilíngue – com textos em dois idiomas –, ou multilíngue – com textos em três ou mais idiomas;
- c. quando um *corpus* é compilado durante certo tempo e depois é considerado como concluído, ou seja, outros textos não são adicionados, trata-se de um ‘*corpus* fechado’. Em contrapartida, quando textos novos são adicionados com frequência e seu tamanho aumenta significativamente, esse *corpus* passa a ser classificado como ‘aberto’ ou ‘monitor’;
- d. o conteúdo é determinado pela característica marcante dos textos que formam o *corpus*. Por exemplo, os textos podem ser de uma área técnica específica ou podem representar uma variante regional;
- e. quanto à autoria, segundo Berber Sardinha (2004: 21) os *corpora* podem conter textos produzidos por nativos ou por aprendizes, quando não são nativos do idioma. A nosso ver, essa diferenciação pode ser renomeada como ‘nativos’ e ‘não nativos’ quando o foco está no estudo do idioma em questão. Da mesma forma que podemos compilar um *corpus* de brasileiros aprendizes de inglês, podemos, por exemplo, compilar um *corpus* de redações escritas por alunos do ensino fundamental, que são nativos e também aprendizes do português. Por isso, além de ‘nativo’ e ‘não nativo’, poderíamos acrescentar aqui outra subcategoria quanto à autoria. Um texto pode ter sido escrito por um especialista em determinado assunto, por um aprendiz de tal assunto ou por um curioso ou leigo no assunto, sendo que todos podem ou não ser falantes nativos do idioma usado. Por exemplo, ao intentar compilar um *corpus* sobre transplantes de medula óssea, podemos estabelecer que os textos coletados sejam apenas aqueles escritos por médicos-cientistas ou por alunos de medicina ou apenas por pessoas que relatam a experiência de terem passado por um transplante. Isso quer dizer que tanto especialistas quanto aprendizes

ou leigos podem ou não ser nativos do idioma usado. Partindo dessa premissa, em nossa pesquisa determinaremos que nosso *corpus* é de aprendizes e é composto por textos (traduções) escritos em sua língua nativa. Claro que, por serem traduções, talvez tenham alguma influência do texto de partida; mesmo assim, não há como negar que esses alunos escrevem em sua língua pátria;

- f. aos estudos dos processos de tradução interessam, principalmente, os *corpora* bilíngues, e em alguns casos, os multilíngues. Com relação à disposição dos textos dentro dos *corpora* bilíngues e multilíngues, com frequência encontramos três possibilidades: os paralelos, os comparáveis e os mistos. Apesar de os nomes parecerem autoexplicáveis, há algumas divergências entre os pesquisadores. Berber Sardinha (2004) estabelece dois tipos de disposição, classificando de ‘paralelos’ os *corpora* em que se pode comparar o original e sua tradução, e ‘alinhados’ quando a tradução aparece logo abaixo da linha do original. De modo geral, quando se fala que um *corpus* é alinhado, entende-se que um período sintático (programas de memória de tradução como o *Wordfast* chamam de ‘segmento’) na língua de partida tem sua tradução correspondente logo abaixo ou logo ao lado. Nesta pesquisa, adotamos as descrições de Bowker e Pearson (2002) e Philip (2009) que consideram que os *corpora* paralelos trazem textos na língua A com suas devidas traduções na língua B. Já os *corpora* comparáveis não apresentam traduções, mas, sim, textos originalmente escritos na língua A e na língua B sob a mesma função comunicativa. Um *corpus* de aprendizes de tradução pode ser monolíngue (quando considerados apenas os textos traduzidos, sem levar em conta o texto de partida), bi- ou multilíngue, comparável ou paralelo (Zanettin, 2012).

Nesta pesquisa, os critérios adotados para a compilação do *corpus* serão os seguintes: todos os textos representam a forma escrita, o *corpus* é bilíngue, sendo que a língua de partida sempre é o inglês e a de chegada, o português. A princípio é

um *corpus* fechado para viabilizar o andamento deste estudo, porém nada nos impede de torná-lo um *corpus* aberto com a contribuição de outros pesquisadores, já que nossa intenção é disponibilizá-lo para acesso público. Os textos não são específicos de determinada área; apresentam um tamanho que varia de 200 a 300 palavras; é de aprendizes de tradução, na maioria falantes nativos de português; é paralelo, sendo que o texto original e respectivas traduções foram alinhadas lado a lado. Outra característica marcante de nosso *corpus* é a quantidade de traduções referentes ao mesmo texto de partida. Para cada texto original há cerca de 100 a 150 traduções. Pelo que nos consta, esta pesquisa é a primeira no Brasil a trabalhar com o alinhamento de múltiplos textos. Tal originalidade é mencionada por Santos (2014). No artigo a pesquisadora relata a análise de um *corpus* bilíngue português-norueguês com múltiplas traduções de aprendizes. A diferença em relação a nossa pesquisa é que Santos trabalha com um número bem limitado de textos, com cerca de oito a dez traduções por original, enquanto em nosso *corpus*, para alguns textos na LP alinhamos cerca de 150 traduções.

1.1.2 Ferramentas para os estudos de corpora

Como já explicado acima, nos dias de hoje não há como pensar em fazer um estudo com base em *corpus* sem ferramentas digitais. Portanto, neste trabalho utilizaremos duas: *Wordsmith Tools* e *AntConc*. O *Wordsmith Tools*, programa com licença de uso paga, foi desenvolvido por Michael Scott e traz três ferramentas fundamentais: um gerador de lista de palavras, um concordanciador e um gerador de lista de palavras-chave. Dentro de cada ferramenta há várias funções que possibilitam diferentes processos de análise do *corpus*. O *AntConc*, com licença de uso gratuita, foi desenvolvido por Laurence Anthony. Um pouco mais simples que o *Wordsmith Tools*, traz as ferramentas essenciais para qualquer pesquisa com *corpora*. Abaixo listamos algumas das funções presentes nos dois programas:

- a. **gerador de lista de palavras:** lista todas as palavras presentes no *corpus*, classificadas por número de ocorrências ou por ordem alfabética. Essa lista faz dois tipos de contagem: uma determina o número de *types*, ou

seja, de palavras distintas, e a outra, o número de *tokens*, ou seja, o número total de palavras presentes. Por exemplo, na frase ‘O tempo perguntou ao tempo quanto tempo o tempo tem’ temos seis *types* e dez *tokens*. Esses valores servem de base para um cálculo estatístico básico, que pode determinar o nível de densidade do *corpus*. Tagnin e Teixeira (2004) propõem uma fórmula de Regra de Três para calcular o *type/token ratio* de um *corpus*:

$$X = \frac{\text{número de types} \times 100}{\text{número de tokens}}$$

O resultado esperado deve estar entre 0 (zero) e 100 (cem). A frase acima, sobre o tempo, usada como exemplo, tem um índice de 60. Quanto mais próximo de 100 estiver o valor final, mais denso será o *corpus*. Quanto mais alto for o valor de *types*, mais variações de vocábulos há no *corpus*, tornando-o mais rico lexicalmente. Tomemos como exemplo um *corpus* cujo tema seja ‘regras’ de modo geral. É de se esperar que o termo ‘regra’ tenha um valor de recorrência bem alto, já que esse é o assunto discutido nos textos que compõem o *corpus*. Entretanto, enquanto *type* a palavra ‘regra’ será contada apenas uma vez. Pode ser que os sinônimos ‘regimento’, ‘regulamentação’ e ‘regulamento’ apareçam em alguns textos e serão contados como *types* diferentes, apesar de se referirem ao mesmo assunto. Isso quer dizer que há uma variedade lexical no *corpus*. Claro que essa é uma análise simplista e não queremos aqui menosprezar todos os outros cálculos estatísticos mais complexos que são usados para determinar as características de um *corpus*. Entretanto, essa conta básica pode dar ao pesquisador uma noção geral de como os termos estão distribuídos nos textos. Ao trabalharmos com *corpora* específicos de um determinado assunto, esse cálculo pode dar ao pesquisador uma noção da riqueza ou não da terminologia presente. Em nossa atividade como docente ministramos um curso cujo objetivo principal é a compilação de um *corpus* de um assunto específico para extrair termos que farão parte

de um glossário. O cálculo *type/token ratio* é usado como parâmetro para sabermos o quão denso o *corpus* é e avaliar se existe ou não material necessário para organizar um glossário com 300 termos. Como exemplo, abaixo seguem os gráficos de um *corpus* sobre transplante de medula óssea. Esse *corpus* foi compilado por um grupo de alunas⁶ do último semestre de um curso de tradução. A Figura 1.1 mostra os dados do *corpus* em inglês:

Figura 1.1 Nível de densidade lexical com base na razão entre *types* e *tokens* no *corpus* em inglês

Trata-se de um *corpus* pequeno e o resultado da razão entre *types* e *tokens* é 2,69. Geralmente os *corpora* em inglês compilados por esses alunos estão na faixa de 2,5 a 3,5. A figura 1.2 abaixo mostra o gráfico do *corpus* em português:

Figura 1.2 Nível de densidade com base na razão entre *types* e *tokens* no *corpus* em português

⁶ Participaram desse grupo as alunas Alessandra Lopes Moreira, Caroline Suite e Cibele Gonçalves. O trabalho foi concluído em junho de 2013. No último semestre, os alunos devem compilar um *corpus* em inglês e português de um assunto específico e extrair 300 termos e suas fraseologias para organizar um glossário bilíngue.

Interessante notar que em português o valor de *tokens* é mais baixo e o de *types*, mais alto, o que, por consequência, gera uma razão mais alta. Os *corpora* em português, mesmo que lematizados, ou seja, uma palavra e suas variações agrupadas e contadas como apenas uma, tendem a ter um número de *types* relativamente mais alto que em inglês⁷. Talvez isso se deva ao fato de o português tentar evitar repetições de palavras;

- b. **gerador de lista de palavras-chave:** para se obter uma lista de palavras-chave é necessário haver dois *corpora*: um é o *corpus* de estudo, aquele que é nosso alvo de pesquisa, e o outro é o *corpus* de referência, via de regra, maior e sem um tema específico, é representativo de uma língua como um todo, e pode ser usado para observações linguísticas gerais (Bowker e Pearson, 2002: 12). Por meio de cálculos estatísticos, o programa identifica as palavras que têm um índice de frequência relativamente baixo no *corpus* de referência, porém alto no *corpus* de estudo. O resultado desse cálculo é chamado de 'índice de chavicidade'. Quanto mais alto o índice de chavicidade, mais relevante para o *corpus* de estudo é a palavra. Na Internet, encontramos alguns *corpora* de referência disponíveis para *download* gratuito. Um exemplo em língua portuguesa é o Lácio-ref, disponibilizado no portal Lácio-Web⁸. Para exemplificar a diferença entre a lista de palavras e a lista de palavras-chave, apresentamos as vinte primeiras ocorrências encontradas em cada uma das listas no *corpus* em português de transplante de medula óssea mencionado acima. O programa *AntConc* foi utilizado para gerar as listas, exibidas na figura 1.3. Na lista de palavras há três colunas. A frequência apresentada na segunda coluna determina a posição (*rank*) das palavras na lista. Como podemos observar, as palavras gramaticais (preposições, artigos, pronomes, conjunções etc.) são as mais frequentes. Entretanto, o

⁷ Viana et al (2008) apresentam uma análise da razão *types/tokens* (formas/itens) de obras literárias brasileiras traduzidas para o inglês em que o valor na tradução é mais baixo que no texto original.

⁸ Para baixar gratuitamente o Lácio-Ref e outros *corpora* basta acessar o portal Lácio-Web e se registrar. Disponível em: <http://www.nilc.icmc.usp.br/lacioweb/index.htm>.

que nos interessa neste momento são as palavras lexicais (substantivos, adjetivos, verbos e advérbios), pois são essas que têm a carga semântica importante para o nosso estudo⁹. A primeira palavra lexical a aparecer é ‘células’. A segunda, ‘pacientes’, e a terceira, ‘transplante’. A lista de palavras-chave é formada por quatro colunas. A terceira coluna (*Keyness*) apresenta o índice de chavicidade, que determinará o posicionamento das palavras. É interessante percebermos que a frequência apresentada na segunda coluna não segue uma ordem decrescente como na ‘lista de palavras’. A posição das palavras na primeira coluna é determinada, agora, pelo índice de chavicidade. As palavras-chave (*keywords*) estão na quarta coluna. Podemos observar que agora há a presença principalmente de substantivos:

Lista de palavras			Lista de palavras chave			
Rank	Freq	Word	Rank	Freq	Keyness	Keyword
1	28163	de	1	4820	12872.541	células
2	15726	a	2	3825	10165.086	pacientes
3	12729	e	3	3116	9113.302	transplante
4	10767	o	4	2883	8529.812	medula
5	7611	do	5	2304	6737.778	óssea
6	7325	em	6	1349	3949.085	doador
7	6892	que	7	1692	3579.983	doença
8	5992	com	8	1413	3552.796	sangue
9	5927	da	9	1095	3319.464	tmo
10	5026	para	10	1665	3171.627	tratamento
11	4820	células	11	1246	2777.893	paciente
12	4451	é	12	922	2716.109	leucemia
13	4208	os	13	874	2329.380	tronco
14	4045	um	14	741	2170.459	sobrevida
15	3825	pacientes	15	695	2106.874	dech
16	3535	no	16	705	2099.907	remissão
17	3387	uma	17	727	2094.969	aguda
18	3116	transplante	18	698	2024.000	doadores
19	3096	ou	19	869	2008.616	t
20	3066	se	20	665	2001.425	linfócitos

Figura 1.3 Comparação de lista de palavras com lista de palavras-chave no *corpus* de medula óssea

Nem sempre o *corpus* de referência é de língua geral. Há momentos em que o pesquisador sente a necessidade de compilar o próprio *corpus* de referência. Isso acontece quando a pesquisa demanda algo específico.

⁹ Não podemos deixar de observar, todavia, que talvez o foco da pesquisa seja exatamente uma palavra gramatical. Um pesquisador pode querer saber, por exemplo, que conjunções adversativas são usadas em determinado gênero.

Digamos que queiramos gerar uma lista de palavras-chave das matérias publicadas na seção policial de um jornal. Nosso *corpus* de estudo é formado pelos artigos da seção policial e nosso *corpus* de referência é formado pelos artigos publicados em todas as outras seções do jornal. Ao cruzar os dois *corpora*, o programa gerará uma lista de palavras-chave formada pelas palavras recorrentes nas matérias policiais, mas raras ou inexistentes nos artigos das outras seções;

- c. **concordanciador**: é uma ferramenta que permite buscar certa palavra dentro do *corpus* e listar todas as suas ocorrências dentro dos co-textos em que ela aparece. Co-texto, aqui, é entendido como um determinado número de caracteres à esquerda e/ou à direita da palavra de busca, que compõe palavras e espaços. Ao analisarmos as linhas de concordância, podemos observar a palavra, ou combinações de palavras, e como essa se comporta ao lado de outras palavras e expressões. As possibilidades oferecidas por essa ferramenta são tão amplas que Sinclair em 2003 publicou *Reading Concordances: an introduction*, dedicado inteiramente a apresentar dezoito tipos de exercícios sobre como explorar as linhas de concordância de um *corpus*. Abaixo (figura 1.4) segue um exemplo de linhas de concordância para a palavra ‘cidades’ encontrada em nosso *corpus* de estudo:

HR	KWIC
1	ânico adverte que no futuro, as temperaturas nas cidades podem ser até 10 graus mais altas que no interior
2	das 'ilhas de calor urbanas' já significa que as cidades estão de 2 a 3 graus mais quentes que as imediaçõ
3	mediações das áreas rurais. A energia usada nas cidades se intensificará com o crescimento da população e
4	itânico informa que no futuro as temperaturas nas cidades podem ser 10° mais altas do que no campo. A 'il
5	. A 'ilha urbana de calor' já quer dizer que as cidades são 2-3°C mais quentes do que a área rural que ci
6	rea rural que circunda. A energia utilizada nas cidades irá se intensificar com o crescimento da populaçã
7	retanha alerta que no futuro as temperaturas nas cidades podem ser 10 graus mais elevadas que no campo. O
8	to dessa onde de calor urbano já significa que as cidades estão dois ou três graus mais quentes que as áreas
9	as áreas rurais próximas. O uso da energia nas cidades vai aumentar assim como as populações vão crescer
10	Bretanha adverte que nas temperaturas futuras, as cidades podem atingir temperaturas 10° maiores do que no
11	urbano do calor de uma ilha já significa que as cidades são 2-3°C mais quentes do que o campo dos arredor
12	dores das zonas rurais. A energia utilizada nas cidades deve intensificar o crescimento de suas populaçõe
13	Bretanha alerta que as temperaturas no futuro nas cidades pode ser 10 graus mais quente do que no campo.
14	o da ilha de de calor urbano "já significa que as cidades são 2-3 ° C mais quente do que as áreas rurais pr
15	s áreas rurais próximas. A Energia utilizada em cidades irá intensificar devido ao crescimento da populaç
16	Bretanha adverte que no futuro, a temperatura nas cidades pode ser 10° mais alta do que na zona rural. O
17	. O efeito da ilha de calor já faz com que as cidades sejam de 2 a 3 graus mais quentes do que as áreas
18	do que as áreas rurais. A energia utilizada nas cidades aumentará conforme a população cresce e o clima t
19	ico da Inglaterra adverte que as temperaturas nas cidades futuramente serão 10° maiores do que na rural.
20	ilhas de calor na área urbana já mostram que as cidades são 2-3° mais quentes que as áreas rurais. A e

Figura 1.4 Exemplo de linhas de concordância com a palavra ‘cidades’ em nosso *corpus* de estudo

Tanto o *AntConc* quanto o *WordSmith Tools* permitem fazer uma reorganização das linhas de concordância tendo como parâmetro as palavras à esquerda ou à direita que acompanham o termo sendo analisado. A figura 1.5 mostra o console do *AntConc*, que chama essa função de 'Kwic Sort'. O item 'level' permite escolher de um a três elementos à esquerda e/ou à direita da palavra de busca. Os valores de cada 'level' vão de 0 a 20 e podem variar da esquerda (L) para a direita (R):

Figura 1.5 Reorganização (sorting) das linhas de concordância no *AntConc*

Já o *WordSmith Tools*, figura 1.6, chama a função de 'Concordance Sort':

Figura 1.6 Reorganização (sorting) das linhas de concordância no *WordSmith Tools*

O que o *AntConc* chama de 'level' o *WordSmith Tools* chama de 'Main Sort', 'Sort 2' e 'Sort 3'. A distância vai de 'centre' a 5, tanto para a esquerda quanto para a direita. Ao reorganizar as linhas de concordância, podemos perceber as tendências de outros elementos aparecerem juntos

com as palavras de busca. A figura 1.7 mostra, no *AntConc*, o '*Kwic Sort*' da palavra 'efeito'. De acordo com os parâmetros que selecionamos, essa função do concordanciador inicialmente considerará as primeiras palavras à esquerda de todas as linhas de concordância do termo pesquisado e as agrupará em ordem alfabética (L1). Em seguida, para cada grupo de linhas de concordância organizadas conforme as letras das palavras à esquerda, a ferramenta colocará em ordem alfabética as linhas de concordância de acordo com as palavras que ocorrem imediatamente à direita (1R) do termo pesquisado. Por fim, dentro desse segundo grupo, a ferramenta organizará as linhas de concordância conforme a ordem alfabética das palavras que se encontram na segunda posição à esquerda (2L) do termo pesquisado:

Figura 1.7 Parâmetros de reorganização das linhas de concordância no *AntConc*

Vale ressaltar que a função *Kwic Sort* pode ser usada de acordo com as necessidades do pesquisador. Por exemplo, quando o *corpus* é em inglês e queremos ver como o termo pesquisado é adjetivado, podemos selecionar a reorganização apenas das palavras à esquerda. Em contrapartida, em português selecionaríamos a organização apenas a partir das palavras à direita. A figura 1.8 mostra o resultado dos parâmetros que estabelecemos acima. Em azul está o termo pesquisado, em vermelho o '*level 1*', em verde o '*level 2*' e em lilás o '*level 3*':

Concordance		Concordance Plot		File View		Clusters/N-Grams		Collocates		Word List		Keyword List	
Concordance Hits 248													
Hit		KWIC										File	
1	do que nas cidades do interior. Devido ao efeito 'ilha de calor', as cidades já estão 2º											text_id=1pr	
2	de branco, isso faria uma mudança climática com efeito de menos 44 bilhões de toneladas de CO2 lit											1pr_info_ar	
3	mais que no interior O calor urbano, com efeito de ilha já indica que as cidades são 2-3º											text_id=1pr	
4	do que no interior. O 'calor urbano com efeito de ilha' já mostra que cidades são de 2											text_id=1pr	
5	lhados pintados de branco, o resultado teria como efeito de 44 bilhões de toneladas a menos de CO2											text_id=1pr	
6	na zona rural. O 'urban heat island effect', (efeito da ilha de calor) já significa que aquelas											text_id=1pr	
7	alta do que na zona rural. A expressão efeito de ilha de calor urbano já significa que											1pr_info_ar	
8	pintassem o telhado de branco, teriam um grande efeito sobre as mudanças no clima, diminuiria em 4											text_id=1pr	
9	eriais mais claros, também haveria um importante efeito positivo. Seria o equivalente a reverter o au											text_id=1pr	
10	uivalente do processo de reversão pelo qual maior efeito é absorver pela Terra porque a (branca) cam											text_id=1pr	
11	a menos na atmosfera porque poderá causar mais efeito a ser refletido de volta no espaço. Casas											text_id=1pr	
12	pintasse o teto de branco, teria o mesmo efeito climático alteraria e liberaria 44 bilhões de											1pr_info_ar	
13	o telhado pintado de branco, teríamos o mesmo efeito climático que lançar menos 44 bilhões de tor											text_id=1pr	
14	pintado ele de branco, isso teria o mesmo efeito climático que emitir 44 bilhões de toneladas											text_id=1pr	
<div><div>< > <</div><div></div><div>> < ></div></div>													
Search Term		☒ Words ☐ Case ☐ Regex		Search Window Size									
efeito		Advanced		100									
Start		Stop		Sort									
Kwic Sort													
☒ Level 1		1L		☒ Level 2		1R		☒ Level 3		2L			
Clone Result													

Figura 1.8 Linhas de concordância da palavra 'efeito' reorganizadas 1L, 1R, 2L no AntConc

1.2 A linguística de *corpus* nos estudos da tradução

Já faz algum tempo que a Linguística de *Corpus* está presente nos estudos da tradução. Laviosa (2004) sugere como marco inicial dos *Estudos da Tradução Baseados em Corpora* (*corpus-based translation studies*) o ano de 1993, quando Mona Baker publicou *Corpus Linguistics and translation studies: implications and applications*. Nesse artigo, Baker advoga contra a abordagem tradicional (e ainda cultivada entre os leigos nos dias de hoje), que vê o texto-fonte como superior ao texto-alvo e defende que o texto traduzido tem o mesmo valor que qualquer outro ato de comunicação. A tradução não é melhor nem pior; é diferente, e, assim sendo, necessita de uma disciplina que a tenha como objeto de estudo: os Estudos da Tradução. A autora, então, sugere que os estudos descritivos da tradução oferecem um campo fértil para os trabalhos feitos com *corpora*.

Vários teóricos que descrevem a tradução como um processo de escolhas são citados por Gentzler (2001). Dentre eles, Jirí Levý, que em 1967 escreve o artigo *Translation as a Decision Process*, e James Holmes, que discute dois aspectos 'delicados' na tradução: as 'decisões subjetivas', que são as escolhas morfosintáticas e lexicais feitas de modo consciente pelo tradutor a fim de respeitar

as características da LC, e os ‘acidentes’, quando, raramente, duas línguas têm equivalentes em pontos específicos, tornando a tradução literal a mais frequente. Gentzler (2001:96) ainda observa que, para Holmes, o processo de tradução envolve decisões iniciais que afetarão as decisões futuras. A partir do momento em que consideramos o ato tradutório como um processo de escolhas do tradutor, a Linguística de *Corpus* vem ocupar a função fundamental de auxiliar o profissional em tais escolhas. Com acesso a um *corpus*, o tradutor tem a oportunidade de pesquisar frequências e/ou probabilidades de uso de determinada palavra ou estrutura e fazer uma tradução que estará bem próxima das expectativas do leitor. Quando um tradutor traduz tendo *corpora* como base, o seu texto deixa de ter apenas a sua ‘voz’ e passa a ter todas as ‘vozes’ dos autores dos textos que compõem o *corpus*.

‘Devemos treinar ou capacitar tradutores?’ é uma reflexão importante que Bernardini (2002) traz à tona sobre a formação do profissional da tradução. O treinamento visa dar ao profissional as ‘fórmulas’ do que fazer em determinadas situações, já a capacitação foca a habilidade do tradutor de se adaptar e desenvolver suas capacidades. A nossa experiência de mais de dez anos no magistério em cursos de tradução de nível universitário em instituições privadas nos traz evidências de que, atualmente no Brasil, a maioria das faculdades particulares ‘treina’ tradutores por meio de exaustivos exercícios de tradução, de redação e de revisão, sem faltarem as ‘dicas’ e ‘receitas’ para enfrentarem as possíveis dificuldades do trabalho. Nos últimos anos, o uso da TAC – Tradução Assistida por Computador – passou a fazer parte da grade curricular de várias instituições, sempre com o objetivo de simular em sala de aula a realidade do mercado. Sem dúvida, essa abordagem de ensino tem obtido sucesso ao formar alunos prontos para se inserirem no mercado. Esse mesmo mercado, porém, está em constante transformação e o profissional deve também estar preparado para se adaptar a essas mudanças. A capacitação do profissional é uma abordagem de ensino que se preocupa menos em focar na realidade de um mercado específico e mais na preparação do tradutor para se adaptar a qualquer realidade, promovendo principalmente a sua autonomia. Os procedimentos de pesquisa possibilitados por meio da Linguística de *Corpus* podem oferecer as condições necessárias para a formação e preparação desse tipo de profissional.

Neste capítulo apresentamos os conceitos básicos da Linguística de *Corpus* e descrevemos algumas funções, como lista de palavras, lista de palavras-chave e linhas de concordância, que permitem a investigação dos *corpora*. Essas funções estão presentes em dois programas amplamente usados por pesquisadores: o *AntConc*, de licença livre, e o *WordSmith Tools*, com mais funções e com licença de uso paga. Por fim destacamos a Linguística de *Corpus* como uma ferramenta de pesquisa para tradutores.

Capítulo 2 – O ensino da tradução

Neste capítulo discutiremos as características dos cursos de tradução no Brasil, mais especificamente no Estado de São Paulo. O primeiro ponto é um levantamento sobre a estrutura dos bacharelados em tradução. Em seguida apresentaremos o que alguns teóricos têm discutido sobre a matriz curricular e como o texto na LP pode ser abordado em aula. Na segunda parte detalharemos a nossa sugestão de uma metodologia com planilhas eletrônicas para ser usada em aula após os alunos terem feito a tradução.

2.1 Bacharelado em tradução

Escolher o bacharelado em tradução ainda é algo não tão comum no Brasil e poucas instituições de ensino superior oferecem o curso. Segundo o relatório *Mapa do Ensino Superior no Estado de São Paulo*¹⁰, publicado pelo Sindicato das Entidades Mantenedoras de Estabelecimentos de Ensino Superior no Estado de São Paulo – SEMESP – em 2012, o estado de São Paulo, o mais populoso do país com cerca de 41 milhões de habitantes, contava em 2010 com 571 instituições de ensino superior, sendo 495 particulares e 76 públicas. Entre essas, de acordo com o Guia do Estudante *on-line*¹¹, apenas dez oferecem o curso universitário em tradução: a Unesp e a Unilago, ambas em São José do Rio Preto, a Universidade Sagrado Coração em Bauru, a Unifran em Franca, a UniSantos em Santos, a Universidade Metodista em São Bernardo do Campo, e, na Capital, o Centro Universitário Anhanguera-Unidade Brigadeiro, a FMU – Faculdades Metropolitanas Unidas, a Universidade São Judas e a Uninove. Ou seja, apenas 0,01% das instituições. Esse percentual, que a princípio pode desenhar uma imagem drástica, na verdade é muito bom, pois muitos desses cursos foram abertos há pouco mais de 10 anos, o que demonstra que talvez as instituições de ensino superior estejam começando a considerar o bacharelado em tradução como uma nova opção dentro dos

¹⁰ O relatório está disponível para acesso online em http://www.semesp.org.br/portal/pdfs/publicacoes/mapa_do_ensino_superior_sp_2012.pdf.

¹¹ A página do guia com as informações sobre o curso de tradução está disponível em: <http://guiadoestudante.abril.com.br/profissoes/comunicacao-informacao/traducao-interpretacao-87919.shtml>. Acesso em: julho de 2014.

departamentos de letras. Assim, fazendo uma projeção otimista para o futuro, se essa tendência continuar logo teremos novos cursos de bacharelado em tradução. Nessa linha de raciocínio, justifica-se a discussão do ensino da tradução em faculdades brasileiras.

Como já discutimos, o bacharelado em tradução é considerado como uma alternativa à licenciatura em letras, porém nem todos nas faculdades ou nos órgãos que regimentam os cursos superiores sabem exatamente qual a diferença entre os dois cursos. Tendo linguagem, comunicação e expressão como foco, há, sim, muitos pontos em comum. Tanto para o tradutor quanto para o professor de idiomas é necessário demonstrar conhecimento sólido dos aspectos linguísticos e culturais das línguas envolvidas. Há momentos no ensino do idioma estrangeiro em que o professor utiliza recursos e estratégias empregados na tradução, como equivalências lexicais e estruturais, omissões, explicitação de conceitos inexistentes em uma das línguas, entre outros, durante lições de vocabulário, gramática e compreensão. Da mesma forma, um tradutor também é um pesquisador e estudioso de idiomas, pois faz uso de técnicas de leitura e compreensão de vocabulário típicas das aulas de língua. Ademais, não há como negar que durante o processo de tradução de um texto aprendemos e revisamos, no mínimo, vocabulário e gramática das línguas envolvidas. Tanto o ensino de idiomas quanto a tradução tem o mesmo objeto de estudo; entretanto, deve estar muito claro para as instituições de ensino superior que o objetivo final almejado pelos alunos é diferente. Da mesma forma que um tradutor não é naturalmente um professor de idiomas, o professor não é naturalmente um tradutor.

À vista disso, os métodos adotados em um curso superior de tradução devem ser planejados tendo em mente que aquele alunato necessita de uma formação profissional específica. Assim, espera-se que a universidade questione, por exemplo, se o ensino e a prática da língua portuguesa e da língua estrangeira podem ou não ter a mesma abordagem para futuros professores e futuros tradutores. Será que um curso de tradução não precisa focar mais na produção textual do que nas análises gramaticais? E sobre os estudos literários? Devem-se oferecer aulas de literatura e teoria literária? Se a resposta for afirmativa, qual abordagem traria benefícios para os alunos? Além dessas questões, ainda há outro assunto mais delicado a ser considerado, já que muitos dos cursos superiores

oferecem o programa sob o nome de ‘tradutor e intérprete’. Seria, então, correto alocarmos tradutores-aprendizes e intérpretes-aprendizes na mesma sala de aula? Certamente esses dois profissionais trabalham com o fazer tradutório, mas ao mesmo tempo em que algumas estratégias são semelhantes, as formas como a tradução se materializa são definitivamente distintas.

O objetivo deste trabalho não é buscar respostas para os questionamentos acima; entretanto, acreditamos que essas e muitas outras perguntas precisam sempre estar em pauta ao falarmos sobre ensino dentro dos Estudos da Tradução. Ao menos no Brasil, o cenário não é tão desolador. Vários eventos e congressos dirigidos a tradutores ocorrem no país. A maioria é voltada ao profissional e ao mercado de trabalho, mas entre aqueles que congregam profissionais e estudiosos, damos destaque ao XI Congresso Internacional da ABRAPT – Associação Brasileira de Pesquisadores em Tradução¹², que ocorreu no campus da Universidade Federal de Santa Catarina em 2013. Durante o evento houve sete mesas redondas, sendo uma para discutir o perfil do professor de tradução, e houve 63 simpósios, cinco dos quais congregaram comunicações que tratavam do ensino da tradução e interpretação e da formação do profissional. Apenas para efeito de comparação, na 10ª edição do mesmo congresso em 2009 na cidade de Ouro Preto¹³, houve 14 mesas redondas, sendo uma dedicada ao ensino, e 81 sessões de comunicações, das quais apenas duas foram dedicadas ao ensino. É essencial que discutamos a formação do professor de tradução, pois ele é o agente transformador dentro das instituições de ensino.

Ao analisarmos as matrizes curriculares¹⁴ propostas pelos cursos de tradução citados no início deste capítulo, observamos que não há muita semelhança entre elas. Enquanto algumas instituições propõem uma carga horária bem alta em estudos literários, outras separam as práticas em aulas de tradução, quando o texto é traduzido para a língua nativa do aluno, e de versão, quando a tradução é para a língua estrangeira. Em uma das faculdades, os alunos do bacharelado estudam junto com os alunos da licenciatura durante o primeiro ano, seguindo a mesma

¹² O programa completo do congresso de 2013 está disponível em: <http://abrapt.files.wordpress.com/2013/11/abrapt-2013-resumos.pdf>.

¹³ O programa completo do congresso de 2009 está disponível em: <http://www.nastrilhasdatraducao.ufop.br/programacao.html>.

¹⁴ As matrizes curriculares estão disponíveis nos sites das instituições de ensino superiores citadas.

matriz curricular. A duração do curso também pode variar entre seis e oito semestres, que equivalem respectivamente a 2400 e 3200 horas de instrução. Em algumas há um estágio obrigatório, enquanto em outras os alunos devem escrever um Trabalho de Conclusão de Curso (TCC) no último ano. Devido a essa falta de equivalência entre os programas, a tentativa de traçar o perfil dos alunos egressos do bacharelado em tradução é uma tarefa frustrante, pois cada faculdade decide por si quais matérias farão parte da matriz curricular.

Essas disparidades, a princípio, não são comuns apenas no Brasil. Nord (2000) argumenta que a falta de uniformidade curricular também é um problema nas faculdades europeias. Além de questionar o perfil de quem leciona as aulas de tradução, já que não há instituições que ofereçam cursos específicos de formação a esses profissionais, nesse mesmo artigo a pesquisadora propõe uma matriz curricular que julga ser ideal para a formação do tradutor. Sua sugestão é de um curso com quatro anos de duração que consistiria de três semestres de estudos dos conceitos teóricos e metodológicos de tradução e comunicação interculturais, um semestre de prática de tradução geral e especializada, tanto na direção da língua nativa quanto da língua estrangeira dos alunos, um semestre de estágio ou estudos no exterior, dois semestres de prática e teoria de tradução especializada, terminologia e uso de ferramentas tradicionais e digitais de tradução e um último semestre para o desenvolvimento de um trabalho de pesquisa. Ainda segundo a autora, o ideal seria formar um ‘tradutor funcional’, ou seja, aquele que, de modo geral, sabe que as funções comunicativas do texto na LP nem sempre são as mesmas na LC, está ciente dos aspectos culturais tanto da LP quanto da LC, busca formas de resolver divergências culturais entre as línguas, sabe que estruturas aparentemente equivalentes podem não ser usadas da mesma forma em duas línguas e sabe usar tanto as fontes de pesquisa tradicionais quanto as digitais.

Como disciplina, os estudos da tradução têm proporcionado várias reflexões e as conclusões que têm emergindo em decorrência dos debates e das discussões propostas por estudiosos acabam por influenciar o modo de traduzir de muitos profissionais. Até mesmo o mercado, na *persona* do cliente, hoje está mais ciente das dificuldades, limitações e virtudes da tradução. Entretanto, nossa experiência no ensino da tradução há mais de dez anos tem nos mostrado que ainda há uma distância significativa entre o que é discutido nos congressos e publicado em artigos

e o que é, de fato, praticado na sala de aula. Albir (1999: 10) reconhece que os avanços nas pesquisas da tradução não têm refletido no modo de ensino e propõe que qualquer didática empregada em um curso de tradução deve responder às seguintes perguntas: 1. para quem se ensina (características e necessidades dos alunos)?; 2. o que se ensina (objetivos e conteúdos)?; 3. como se ensina (métodos e meios)?; e 4. com qual resultado (critérios de correção e nivelamento, tipos de avaliações)?¹⁵ ¹⁶.

Ao discutir a didática tradicional de ensino, a autora descreve um cenário que ainda é visto nos dias de hoje, com atividades ‘leia e traduza’, em que apenas as questões lexicais são discutidas e os alunos aguardam pela ‘alternativa correta’ fornecida pelo professor. Infelizmente esse cenário ainda ocorre no Brasil. Um curso universitário deve ter como foco o viés profissionalizante e a faculdade de tradução cumpre o seu papel quando o aluno egresso tem plenas condições de satisfazer as exigências do mercado, já que a maioria dos discentes pretende atuar como tradutor. Sem dúvida, ao elaborar um curso de tradução, é essencial que o professor, além de contemplar os pontos propostos por Albir (1999), também leve em consideração ‘para qual mercado’ estamos preparando os estudantes. Por exemplo, nos últimos anos observamos que muitos alunos, ao buscar estágio em agências de tradução, são contratados para o cargo inicial de revisor e não de tradutor, algo diferente do que acontecia no passado, quando o revisor em tradução era o tradutor experiente.

Levando-se em conta o mercado de trabalho, Ruuskanen (1996) propõe um modelo que foi baseado na experiência profissional de vários tradutores. A autora sugere que tal modelo seja usado em aulas para que os alunos percebam que uma tradução não existe em um vácuo e para que se crie ‘um cenário’ para o texto que será traduzido. Tal modelo é chamado de CATLEGS (*Client, Audience, Time, Lexicon, Genre e Style*) e é baseado nas perguntas que tradutores profissionais fazem antes de aceitar um trabalho. Quando compiladas, representam os pontos a serem considerados antes de realizarmos uma tradução. As respostas obtidas servem de subsídios para construirmos o perfil do texto, cujas características e

¹⁵ Todas as traduções nesta dissertação são de nossa autoria, salvo quando especificado.

¹⁶ No original: “[p]ara quién se enseña (características y necesidades de los estudiantes), qué se enseña (objetivos y contenidos), como se enseña (métodos y medios) y con qué resultado (criterios de corrección y nivelación, tipo de pruebas)”.

expectativas servirão de guia nos processos de escolha e tomadas de decisão durante o ato tradutório. As perguntas propostas são: 1. Quem solicita a tradução? Quem é o autor? – **C**liente; 2. Para quem é o texto? – **A**udiência; 3. Qual é o prazo de entrega? – **T**empo; 4. Qual é o assunto/área do texto? – **L**éxico/registro; 5. Qual é o propósito do texto? – **G**ênero; 6. Em qual formato e onde o texto será publicado? – **eS**tilo. Há uma última pergunta no modelo que, apesar de não entrar no modelo mnêmico, tem um caráter importantíssimo, pois questiona se o cliente concordará com os valores do orçamento passado: 7. O valor do meu orçamento será aceito? A resposta determinará se o trabalho será ou não iniciado.

Como se pode notar, a ordem das perguntas e suas respostas favorecem um processo mnemônico ideal para ser usado em sala de aula e que funciona também em português: **C**liente, **A**udiência, **T**empo, **L**Éxico, **G**ênero, **eS**tilo = CATLEGS. Acreditamos que o uso adequado do modelo CATLEGS seria aquele em que os próprios alunos determinam as respostas às perguntas, pois, ao participarem da construção da atividade fazem com que a aula não seja centrada no professor. As possibilidades de atividades são inúmeras; por exemplo, a turma pode ser dividida em grupos, sendo que cada grupo recebe respostas diferentes para o CATLEGS, e posteriormente, em aula, comparam as diferenças. Outra sugestão seria trabalhar com revisão: os alunos receberiam um texto fonte, sua tradução e uma proposta de respostas ao CATLEGS, que seria usada para revisar e avaliar se a tradução está ou não de acordo com o pedido e, se não estiver, propor uma nova versão.

A utilização do modelo CATLEGS estabelece uma metodologia de abordagem do texto na LP que favorece o processo de escolhas conscientes por parte do tradutor-aprendiz, isto é, o aluno passa a ter elementos para explicar e justificar cada uma dessas escolhas.

Retomando o que já discutimos sobre a tradução ser um processo de escolhas não aleatórias, parece-nos que o modelo funcionalista, proposto por Christiane Nord (2005) a partir dos princípios difundidos por Hans J. Vermeer (2000) e Katharina Reiss (2000)¹⁷, adequa-se perfeitamente como fundamentação teórica dessa metodologia de ensino. Antes de prosseguirmos, deve ficar claro que há outras teorias e abordagens para a tradução e que julgamos ser de fundamental

¹⁷ Cf. Vermeer (1978, 1989 e 1996) e Reiss e Vermeer (1984 e 1991).

importância que o aluno seja exposto a elas. Além disso, também é importante que o professor esteja ciente de que, como outras propostas, o modelo funcionalista também tem seus críticos, principalmente em relação à tradução de textos literários¹⁸.

Em resumo, o modelo funcionalista, segundo Nord (2005), toma a tradução como uma forma de comunicação intercultural. O texto na LP tem um sistema cultural distinto daquele em que se inserirá o texto na LC. Assim, as funções do texto fonte devem ser analisadas uma a uma, levando em conta que a tradução gerada na LC terá contexto e público distintos. Partindo do princípio que o texto é um ato de comunicação, ao considerar quais elementos serão ou não mantidos e quais serão adaptados, o tradutor deve ter em mente que o público alvo da tradução é o leitor da LC.

Ao levar o modelo funcionalista para a sala de aula, o professor torna o texto mais ‘concreto’ para o aluno, orientando-o a transpor as fronteiras linguísticas e a tentar identificar (ou supor) o que motivou o texto e por que o autor o escreveu. Uma atividade que proponha, por exemplo, a leitura do texto na LP de modo a identificar as intenções do autor e os aspectos culturais intrínsecos propicia a discussão entre os alunos de como tais aspectos serão representados na LC e o que fazer quando a intenção do texto de partida é diferente da do texto de chegada. Essa interação com o texto é o primeiro passo para guiar o tradutor-aprendiz em seu processo de escolhas.

Segundo Reiss (2000), após a fase inicial de análise textual, que engloba a identificação da tipologia (narração, descrição, instrução etc.), do gênero (comunicado, notícia, bilhete, manual etc.) e do estilo (formal, informal, emocional etc.), o tradutor deve reconhecer as unidades de tradução¹⁹. Alves (1997 e 2011) sugere aos tradutores-aprendizes que tentem localizar no texto na LP as Unidades de Tradução (UTs) presentes de modo consciente. Essa conscientização é um

¹⁸ Alice Leal discute o modelo funcionalista e a tradução literária no artigo *Funcionalismo e tradução literária: o modelo de Christiane Nord em três contos ingleses contemporâneos*. Disponível em: <https://periodicos.ufsc.br/index.php/scientia/article/view/12916>. Acesso em: maio 2014.

¹⁹ Vale ressaltar que o termo ‘unidade de tradução’ aqui tem um significado distinto daquele tipicamente encontrado em programas de Tradução Assistida por Computador. Tanto no *Wordfast* quanto no *Trados*, ‘unidade de tradução’, muitas vezes abreviado como ‘TU’ (*translation unit*), se refere à combinação de um trecho na LP, chamado de ‘segmento’, que equivale a um período sintático, junto a sua respectiva tradução na LC, dispostos paralelamente na Memória de Tradução (TM – *translation memory*).

procedimento pelo qual, durante a leitura, o aluno identifica combinações específicas de palavras que juntas transmitem um significado distinto. O autor ainda salienta que o objetivo é aumentar o grau de autonomia dos alunos durante o processo de tradução para que tenham controle sobre sua própria produção. O grande desafio aos aprendizes, entretanto, é delimitar as UTs, que podem corresponder a uma palavra, a um ou mais sintagmas, a uma oração ou mesmo a um período, ou seja, não há como especificar do que exatamente se constituem as UTs:

Unidade de Tradução é um segmento do texto de partida, independente de tamanho e forma específicos, para o qual, em um dado momento, se dirige o foco de atenção do tradutor. Trata-se de um segmento em constante transformação que se modifica segundo as necessidades cognitivas e processuais do tradutor. A Unidade de Tradução pode ser considerada como a base cognitiva e o ponto de partida para todo o trabalho processual do tradutor. Suas características individuais de delimitação e sua extrema mutabilidade contribuem fundamentalmente para que os textos de chegada tenham formas individualizadas e diferenciadas. O foco de atenção e consciência é o fator direcionador e delimitador da Unidade de Tradução e é através dele que ela se torna momentaneamente perceptível. (Alves, 2011: 38)

No artigo, o autor apresenta o trecho de um texto com *how* e *many* em que cada um deles pertence a uma UT diferente:

Visitors who want to see how many of Brazil's citizens live can now add the Morro da Providência favela to their tourism itinerary.

Ao perceber que *how* e *many* não formam uma UT, o aluno identifica que o significado aqui não é de ‘quantos’, significado muito mais frequente da combinação *how many*, mas, sim, de ‘como’ e ‘muitos’, motivando a seguinte tradução: ‘...ver como moram muitos dos cidadãos...’ (Alves 2011: 33, grifo nosso).

Na área jurídica encontramos UTs que são constituídas de várias palavras e que, isoladamente, não têm equivalência unívoca com as palavras que formam a unidade em português. Em um curso de tradução jurídica que ministramos, os alunos compilaram um pequeno *corpus* comparável bilíngue de procurações em inglês e português com cerca de 20 mil palavras em cada idioma para auxiliá-los na tradução de uma procuração. O texto de partida continha *Know all persons by these*

presents, uma combinação que deixou os alunos muito intrigados, pois, a princípio, não fazia sentido tendo como base o conhecimento da língua inglesa para fins gerais. Ao suporem que se tratava de uma UT, foram ao *corpus* em inglês e em seis dos vinte e um textos compilados a expressão aparecia. Concluíram, então, que realmente se tratava de uma UT e que sua ocorrência era sempre para introduzir o primeiro parágrafo das procurações. Conscientes desse fato, e que por isso as palavras não poderiam ser traduzidas individualmente, os alunos passaram a analisar o *corpus* em português. A busca foi por uma UT em português que tivesse a mesma função, ou seja, introduzir o primeiro parágrafo de uma procuração. Logo identificaram a expressão ‘Saibam quantos este público instrumento de procuração bastante virem que’ também encontrada em seis textos diferentes e usada para iniciar a procuração. Estar ciente da função da UT na LP e eleger a UT na LC que melhor se adequa ao texto traduzido é uma forma de o aluno ter controle sobre suas escolhas.

Da mesma forma, o reconhecimento de que uma UT tem um significado idiomático na LP viabiliza a busca de uma solução mais adequada no texto da LC. A frase abaixo foi retirada do texto 24²⁰, que faz parte do *corpus* de textos em inglês desta pesquisa. O texto fala sobre um juiz americano que processou a lavanderia de seu bairro por ter perdido suas calças. Ao falar sobre o processo, o autor usa a seguinte frase:

He has decided to take his local dry cleaner's to the cleaners for a cool \$65 million.

Esse trecho gerou muita dificuldade aos alunos que não perceberam que a UT é toda a expressão idiomática e que quer dizer ‘tirar todo o dinheiro possível de alguém’.

Localizar e identificar as UTs em um texto não é tarefa fácil e requer muitos exercícios. Antes de iniciar qualquer tarefa de tradução o professor deve promover junto aos alunos uma leitura ‘consciente’ do texto. Esse exercício de pré-tradução permite que o tradutor-aprendiz reconheça a função do texto, o estilo do autor, pense em um público leitor na LC, identifique as UTs etc. Toda essa conscientização

²⁰ Futuramente pretendemos disponibilizar em formato eletrônico os textos que formam nosso *corpus*.

colocará o texto sob controle do aluno e esse, por sua vez, terá condições de fazer as escolhas para a sua tradução com muito mais segurança.

2.2 Proposta de aula

O que foi apresentado neste capítulo até agora sugere algumas formas de abordar o texto na LP ao apresentá-lo aos alunos. Entretanto, nosso centro de atenção agora se volta para o que acontece durante a aula propriamente dita, principalmente quando levamos em conta a realidade das salas de aula das universidades privadas brasileiras. É impraticável falar sobre *feedback* individualizado a um professor que leciona para sete ou oito turmas com cerca de 40 alunos em cada uma. O que propomos aqui é uma metodologia cujo plano de aula parta das análises de um *corpus* composto pelas traduções dos alunos. Nosso foco está no ‘pós-tradução’ e leva em consideração salas com um número elevado de discentes, tecnologia simples, como quadro de giz, projetor *Datashow* e computador, e um professor com conhecimentos médios de um editor de texto e de um manipulador de planilhas eletrônicas²¹. Também é essencial que o professor conheça os comandos básicos de um programa que pelo menos gere listas de palavras, lista de palavras-chave e linhas de concordância. Neste capítulo mencionaremos apenas o uso do *AntConc*, por se tratar de um programa gratuito.

Após selecionar um texto na LP que atenda os objetivos da ementa do curso, devemos salvá-lo em *txt* e carregá-lo no *AntConc*. A forma mais simples para isso é copiar e colar o conteúdo em um arquivo novo do programa Bloco de Notas. Em seguida, carregamos esse arquivo no *AntConc* e geramos uma lista de palavras, que nos possibilitará construir uma série de expectativas para o texto. Como dissemos, são apenas expectativas, pois somente os resultados obtidos a partir de um *corpus* formado pelas traduções dos alunos nos darão os subsídios para a aula. A figura 2.1 abaixo mostra a lista de palavras gerada a partir de um dos textos que utilizamos nesta pesquisa. O texto é sobre como lidar com o estresse e foi escolhido para ser

²¹ Nesta pesquisa utilizamos o Pacote Office da Microsoft. Caso o professor não tenha acesso a programas pagos, há alternativas gratuitas como o Pacote Office Online no site: www.outlook.com, com acesso após criar uma conta de *e-mail*, ou os programas como o *OpenOffice*, que pode ser baixado de: <http://www.openoffice.org/pt-br/>, e o *LibreOffice*, que pode ser baixado de: <https://pt-br.libreoffice.org/>.

trabalhado em sala, inicialmente, pela repetição da palavra *stress*. O objetivo da aula é discutir quais estratégias adotar uma vez que se recomenda que a boa redação em português evita a repetição de palavras em um mesmo período ou parágrafo:

Word Types: 113		Word Tokens: 223
Rank	Freq	Word
1	13	you
2	12	your
3	9	to
4	8	can
5	7	and
6	7	stress
7	6	a
8	6	it
9	6	the
10	5	time
11	4	avoid
12	4	is
13	4	try

Figura 2.1 Lista de palavras geradas a partir do texto na LP

Ao analisarmos essa lista de palavras podemos prever que os alunos terão de lidar não somente com a repetição de *stress*, mas também de *you*, *can*, *time*, *avoid* e *try*. O verbo modal *can* traz ainda outra dificuldade, pois há variantes em sua tradução com ‘poder’, ‘saber’, ‘conseguir’, podendo até desaparecer por completo em português. Ao preparar a aula referente à fase de pré-tradução do texto com os alunos, devemos chamar atenção para essas repetições e discutir quais estratégias podem ser usadas na tradução.

O texto na LP deve ser disponibilizado aos alunos digitalmente e em formato *doc*. Para facilitar a transferência das traduções para a planilha eletrônica, temos de inserir uma tabela. Para isso, selecionamos o texto e, na aba ‘Inserir’ do Word, clicamos em ‘Tabela’ e selecionamos ‘Converter texto em tabela’. Em seguida abre-se uma janela onde selecionamos ‘1’ em ‘Número de colunas’ e em ‘Texto separado em’ selecionamos ‘Outro:’ e digitamos ‘.’ no espaço reservado, como mostra a figura 2.2:

Figura 2.2 Passos para a conversão de texto em tabela no Word do Microsoft Office

O resultado será o texto dividido de modo que cada período sintático esteja em uma linha, como mostra a figura 2.3:

Figura 2.3 Resultado após o comando 'Inserir coluna'

Para inserir uma coluna à direita, onde os alunos farão a tradução, basta selecionar o texto na tabela, clicar com o botão esquerdo do mouse e pedir para inserir uma coluna à direita, como demonstrado na figura 2.4:

Figura 2.4 Inserção de coluna à direita para disponibilizar o espaço onde a tradução será feita

Respeitando o prazo estabelecido pelo professor, os alunos enviam por e-mail as traduções. Após receber todos os arquivos, passamos a compilar o *corpus* de traduções do texto. Inicialmente criamos um arquivo no gerenciador de planilhas eletrônicas Excel com o texto na LP na coluna A. Em seguida abrimos os arquivos dos alunos e copiamos apenas a coluna que contém o texto na LC e passamos a colar as traduções a partir da coluna B no Excel. O resultado pode ser visto na figura 2.5 abaixo:

	A	B	C	D	E	F
1	Stress Management - Ways to Avoid Stress	André	Bárbara	Carolina	Eduardo	Marcos
2	Stress is a part of life, and you can't always avoid it.	Gerenciamento de estresse - Maneiras de evitar estresse	Controle de Estresse - Maneiras de Evitar Estresse	Controle de Estresse - Maneiras para evitar o estresse.	Administração do estresse - Maneiras de evitá-lo	Controle do estresse - Maneiras de evitar estresse
3	But you can try to avoid situations that can cause it, and you can control how you respond to it.	Mas você pode tentar evitar situações que possam estressá-lo e pode controlar como reagir a isto.	Mas você pode tentar evitar situações que causam estresse e controlar a suas reações.	Mas você pode tentar evitar situações que podem causá-lo e pode controlar como responder à essas situações.	Mas você pode tentar evitar situações que o causam e pode controlar como responde a ele.	Mas você pode tentar evitar situações que o causam, e controlar a maneira como você enfrenta essas situações.
4	The first step is knowing your own coping strategies.	O primeiro passo é conhecer sua própria estratégia de luta.	O primeiro passo é saber as suas estratégias.	O primeiro passo é conhecer suas estratégias de sobrevivência.	O primeiro passo é saber sua própria estratégia para enfrentá-lo.	O primeiro passo é conhecer suas próprias estratégias para lidar com isso.
5	Try using a stress journal to record stressful events, your response to them, and how you coped.	Tente usar o estresse diário para gravar eventos estressantes, sua reação à eles, e como você lidou com isso.	Tente anotar situações estressantes em um caderno, suas reações e como você lidou com elas.	Tente usar um diário de estresse para marcar eventos estressantes, a sua resposta para eles e como os sobreviveu.	Tente usar um diário para registrar situações estressantes, como respondeu a ele e como o enfrentou.	Tente fazer um registro diário anotando as situações estressantes, como você as enfrenta e lida com elas.

Figura 2.5 Planilha do Excel com o texto na LP e as traduções dos alunos

Na preparação do *corpus*, temos que salvar o conteúdo da planilha em formato *txt* para que possamos explorá-lo no *AntConc*. Há duas formas de fazê-lo. A mais simples é salvar a planilha em formato *txt*, porém, não há como separar o texto da LP. Como os trechos da LP aparecem apenas uma vez, ao gerar a lista de palavras, os termos da LC serão os mais frequentes. A segunda forma é criar uma nova planilha que contenha apenas as traduções e salvá-la como *txt*. Para exemplificar esse procedimento, utilizamos uma planilha com 26 traduções. A figura 2.6 mostra a lista de palavras gerada a partir das traduções ao lado da lista de palavras do texto na LP:

Word Types: 113			Word Tokens: 223		Word Types: 419			Word Tokens: 5957	
Rank	Freq	Word	Rank	Freq	Word	Rank	Freq	Word	Rank
1	13	you	1	245	o	1	245	o	1
2	12	your	2	214	de	2	214	de	2
3	9	to	3	201	você	3	201	você	3
4	8	can	4	198	seu	4	198	seu	4
5	7	and	5	180	e	5	180	e	5
6	7	stress	6	179	que	6	179	que	6
7	6	a	7	159	pode	7	159	pode	7
8	6	it	8	148	estresse	8	148	estresse	8
9	6	the	9	143	vida	9	143	vida	9
10	5	time	10	138	a	10	138	a	10
11	4	avoid	11	124	tempo	11	124	tempo	11
12	4	is	12	104	é	12	104	é	12
13	4	try	13	92	não	13	92	não	13
14	3	are	14	90	como	14	90	como	14
15	3	at	15	88	fazer	15	88	fazer	15
16	3	how	16	82	sua	16	82	sua	16
17	3	life	17	76	mais	17	76	mais	17
18	3	stressful	18	75	evitar	18	75	evitar	18
19	3	t	19	74	tente	19	74	tente	19
20	3	that	20	71	as	20	71	as	20

Figura 2.6 Lista de palavras do texto na LP e das traduções

O estudo comparativo das duas listas possibilita preparar uma aula que vai tratar dos pontos que realmente tenham gerado dificuldade para os alunos. Por exemplo, ao analisar os índices de recorrência da lista de palavras em português na figura 2.6 acima, é possível detectar que deve ter havido alguma dificuldade por parte dos alunos em encontrar opções para o verbo modal *can*. Ao preparar a aula, o esperado seria a forma ‘consegue’. Porém, o que observamos é que ‘pode’ tem um índice de frequência bem alto. Esse dado demanda uma investigação mais detalhada das linhas de concordância a fim de detectarmos como ‘pode’ foi usado. A

figura 2.7 mostra as dez primeiras linhas de concordância do termo ‘pode’ nas traduções:

Concordance		Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits		159					
Hit	KWIC						
1	stress faz parte da vida e você não pode evitá-lo O estresse é uma parte da						
2	uma parte da vida, e você nem sempre pode evita-lo. O estresse faz parte da vida						
3	estresse faz parte da vida e você não pode evita-lo. O estresse faz parte da vida,						
4	estresse faz parte da vida, e você não pode evita-lo sempre. Stress faz parte da vida,						
5	.Stress faz parte da vida, e nem sempre pode ser evitado. O estresse é uma parte da						
6	é uma parte da nossa vida, mas você pode sempre evitar isso. O estresse é uma parte						
7	uma parte da sua vida e você não pode sempre evita-lo. O estresse faz parte da						
8	estresse faz parte da vida e você não pode sempre evita-lo. Estresse é uma parte da						
9	.O stress faz parte da vida, e você pode não pode evitá-lo sempre. Estresse faz parte						
10	faz parte da vida, e você pode não pode evitá-lo sempre. Estresse faz parte da vida,						

Figura 2.7 Dez primeiras linhas de concordância da palavra ‘pode’ nas traduções

Um dos problemas que enfrentamos ao ministrar aulas de prática de tradução em salas com número de alunos elevado é como avaliar individualmente as traduções feitas. Outra questão a ser levada em consideração é que o aluno, ao traduzir, sabe que apenas o professor lerá o seu texto, o que pode, de certa maneira, ser desestimulante. Ao levar todas as traduções para a aula, o professor abre a possibilidade de toda a turma participar na avaliação do texto na LC. Além disso, também há o efeito ‘psicológico’ no aluno, que fará a tarefa de tradução com mais apreço, pois sabe que todos os seus colegas de sala lerão seu texto.

O professor pode usar a mesma planilha que foi usada para gerar o arquivo *txt* (ver figura 2.5 acima) ou pode, a partir dela, criar uma nova planilha, que permitirá reorganizar as células de modo a unir as soluções iguais. Para isso, basta selecionar todas as linhas e colunas com o texto na LP e as traduções, usar o comando ‘copiar’, tanto no atalho ‘ctrl+C’ quando no botão direito do *mouse*, e colar o conteúdo em uma nova planilha. Ao transferir esse conteúdo para a nova tabela, devemos apenas utilizar o botão direito do *mouse* para termos acesso ao menu ‘Opções de Colagem’. Nesse menu, seleciona-se ‘Colar Especial’ e, no submenu, seleciona-se a opção ‘Transpor’, como mostrado na figura 2.8:

Figura 2.8 Comando 'Colar' com o submenu 'Transpor'

Esse procedimento fará com que as linhas que estavam na horizontal sejam dispostas na vertical, permitindo, assim, que a primeira linha contenha cada segmento do texto na LP e, a partir da segunda linha, os nomes dos alunos e suas traduções referentes àquele trecho. Assim, a coluna A contém os nomes dos alunos, a coluna B o primeiro segmento e suas traduções, a coluna C, o segundo segmento e suas traduções e assim por diante. A vantagem desse procedimento é a possibilidade de ordenar as células das colunas verticalmente de modo a agrupar as traduções para aquele trecho específico em ordem alfabética. Para isso, selecionamos as células que contêm a tradução e os nomes dos alunos, deixando de fora apenas a primeira linha, que contém os segmentos do texto na LP. Em seguida, na ferramenta 'Classificar e Filtrar' da planilha, clica-se em 'Personalizar a Classificação'. Ao abrir a janela 'Classificar', em 'Classificar por', escolhemos a coluna que será a âncora da ordenação das linhas e clicamos em 'Ok', como mostra a figura 2.9 abaixo:

	A	B	C	D	E	F
1		Stress Management - Ways to Avoid Stress	Stress is a part of life, and you can't always avoid it.	But you can try to avoid situations that can cause it, and you can control how you respond to it.	The first step is knowing your own coping strategies.	Try using a stress journal to record stressful events, your response to them, and how you coped.
2	André	Gerenciamento de estresse - Maneiras de evitar estresse	Stress faz parte da vida, e você não pode evitá-lo sempre.	Mas você pode tentar evitar situações que o causam, e controlar a maneira como você enfrenta essas situações.	O primeiro passo é conhecer suas próprias estratégias para lidar com isso.	Tente usar o estresse diário para gravar eventos estressantes, sua reação a eles, e como você lidou com isso.
3	Bárbara	Controle de Estresse - Maneiras de Evitar Estresse	Estresse faz parte da vida, e você nem sempre pode evitá-lo.	Mas você pode tentar evitar situações que podem causá-lo e pode controlar como responder a essas situações.	O primeiro passo é saber as suas estratégias de sobrevivência.	Tente anotar situações estressantes em um caderno, suas reações e como você lidou com elas.
4	Carolina	Controle de Estresse - Maneiras para evitar o estresse.	Estresse é uma parte da vida, e não sempre você consegue evitá-lo.	Mas você pode tentar evitar situações que o causam, e controlar a maneira como você enfrenta essas situações.	O primeiro passo é conhecer suas próprias estratégias para lidar com isso.	Tente usar um diário de estresse para marcar eventos estressantes, a sua resposta para eles e como os sobreviveu.
5	Eduardo	Administração de Estresse - Maneiras de evitar estresse	Estresse faz parte da vida, e não é sempre que se pode evitá-lo.	Mas você pode tentar evitar situações que o causam, e controlar a maneira como você enfrenta essas situações.	O primeiro passo é conhecer suas próprias estratégias para lidar com isso.	Tente usar um diário para registrar situações estressantes, como respondeu a ele e como o enfrentou.
6	Marcos	Controle do estresse - Maneiras de evitar estresse	Estresse faz parte da vida, e não é sempre que se pode evitá-lo.	Mas você pode tentar evitar situações que o causam, e controlar a maneira como você enfrenta essas situações.	O primeiro passo é conhecer suas próprias estratégias para lidar com isso.	Tente fazer um registro diário anotando as situações estressantes, como você as enfrenta e lida com elas.

Figura 2.9 Ferramenta 'Classificar e Filtrar' no Excel

Esse comando fará com que as linhas da tabela se reorganizem em ordem alfabética a partir dos conteúdos da coluna selecionada. Por exemplo, ao selecionarmos a coluna B como âncora, o programa ordenará todas as linhas a partir do conteúdo de suas células (figura 2.10). Isso quer dizer que sempre teremos o nome do aluno e sua tradução na mesma linha:

	A	B	C	D	E	F
1		Stress Management - Ways to Avoid Stress	Stress is a part of life, and you can't always avoid it.	But you can try to avoid situations that can cause it, and you can control how you respond to it.	The first step is knowing your own coping strategies.	Try using a stress journal to record stressful events, your response to them, and how you coped.
2	Eduardo	Administração do estresse - Maneiras de evitá-lo	Estresse faz parte da vida e não dá para evitá-lo sempre.	Mas você pode tentar evitar situações que o causam e pode controlar como responde a ele.	O primeiro passo é saber sua própria estratégia para enfrentá-lo.	Tente usar um diário para registrar situações estressantes, como respondeu a ele e como o enfrentou.
3	Bárbara	Controle de Estresse - Maneiras de Evitar Estresse	Estresse faz parte da vida, e você nem sempre pode evitá-lo.	Mas você pode tentar evitar situações que causam estresse e controlar a suas reações.	O primeiro passo é saber as suas estratégias.	Tente anotar situações estressantes em um caderno, suas reações e como você lidou com elas.
4	Carolina	Controle de Estresse - Maneiras para evitar o estresse.	Estresse é uma parte da vida, e não sempre você consegue evitá-lo.	Mas você pode tentar evitar situações que podem causá-lo e pode controlar como responder a essas situações.	O primeiro passo é conhecer suas estratégias de sobrevivência.	Tente usar um diário de estresse para marcar eventos estressantes, a sua resposta para eles e como os sobreviveu.
5	Marcos	Controle do estresse - Maneiras de evitar estresse	Estresse faz parte da vida, e não é sempre que se pode evitá-lo.	Mas você pode tentar evitar situações que o causam, e controlar a maneira como você enfrenta essas situações.	O primeiro passo é conhecer suas próprias estratégias para lidar com isso.	Tente fazer um registro diário anotando as situações estressantes, como você as enfrenta e lida com elas.
6	André	Gerenciamento de estresse - Maneiras de evitar estresse	Stress faz parte da vida, e você não pode evitá-lo a todo instante.	Mas você pode tentar evitar situações que possam causá-lo e pode controlar como reagir a isto.	O primeiro passo é conhecer sua própria estratégia de luta.	Tente usar o estresse diário para gravar eventos estressantes, sua reação a eles, e como você lidou com isso.

Figura 2.10 Resultado do comando 'Colar especial - transpor' com a coluna B em ordem alfabética

Quando o professor passar para a análise do segundo trecho, repete-se o procedimento de 'Personalizar a Classificação'. A figura 2.11 mostra as linhas ordenadas a partir da coluna C:

	A	B	C	D	E	F
		Stress Management - Ways to Avoid Stress	Stress is a part of life, and you can't always avoid it.	But you can try to avoid situations that can cause it, and you can control how you respond to it.	The first step is knowing your own coping strategies.	Try using a stress journal to record stressful events, your response to them, and how you coped.
1	Carolina	Controle de Estresse - Maneiras para evitar o estresse.	Estresse é uma parte da vida, e não sempre você consegue evitá-lo.	Mas você pode tentar evitar situações que podem causá-lo e pode controlar como responder à essas situações.	O primeiro passo é conhecer suas estratégias de sobrevivência.	Tente usar um diário de estresse para marcar eventos estressantes, a sua resposta para eles e como os sobreviveu.
2	Eduardo	Administração do estresse - Maneiras de evitá-lo	Estresse faz parte da vida e não dá para evitá-lo sempre.	Mas você pode tentar evitar situações que o causam e pode controlar como responde a ele.	O primeiro passo é saber sua própria estratégia para enfrentá-lo.	Tente usar um diário para registrar situações estressantes, como respondeu a ele e como o enfrentou.
3	Marcos	Controle do estresse - Maneiras de evitar estresse	Estresse faz parte da vida, e não é sempre que se pode evitá-lo.	Mas você pode tentar evitar situações que o causam, e controlar a maneira como você enfrenta essas situações.	O primeiro passo é conhecer suas próprias estratégias para lidar com isso.	Tente fazer um registro diário anotando as situações estressantes, como você as enfrenta e lida com elas.
4	Bárbara	Controle de Estresse - Maneiras de Evitar Estresse	Estresse faz parte da vida, e você nem sempre pode evitá-lo.	Mas você pode tentar evitar situações que causam estresse e controlar a suas reações.	O primeiro passo é saber as suas estratégias.	Tente anotar situações estressantes em um caderno, suas reações e como você lidou com elas.
5	André	Gerenciamento de estresse - Maneiras de evitar stress	Stress faz parte da vida, e você não pode evitá-lo a todo instante.	Mas você pode tentar evitar situações que possam stressá-lo e pode controlar como reagir a isto.	O primeiro passo é conhecer sua própria estratégia de luta.	Tente usar o estresse diário para gravar eventos estressantes, sua reação à eles, e como você lidou com isso.
6						

Figura 2.11 Resultado do comando 'Colar especial - transpor' com a coluna C em ordem alfabética

Em aula, o professor usa o *Datashow* para projetar a planilha, focando em cada segmento e nas suas traduções. Ao expormos aos alunos todas as possibilidades de tradução para apenas um trecho, deixamos explícito o processo de escolhas. Além disso, quando o aluno compara as diversas possibilidades, abrimos a possibilidade para o desenvolvimento de seu senso crítico em relação à tradução. Com essa metodologia o professor consegue dar um retorno avaliativo para todas as traduções feitas e, ao envolver todos os alunos nas análises de cada trecho, a aula, mais uma vez, deixa de ser centrada no professor, ou seja, no seu conhecimento e experiência, e passa a ter a participação ativa de todos os discentes. Essa metodologia ainda traz mais benefícios para a aula. Primeiro é o abandono do modelo tradicional em que o professor lê o trecho na LP e um ou dois alunos leem suas traduções, enquanto os outros se esforçam para tomar notas, e muitos acabam frustrados por não saberem o quão adequadas são as suas próprias escolhas na LC. Ademais, muitas vezes o aluno não faz a tarefa e fica esperando a 'tradução correta'; segundo, ao saber que sua tradução será exposta e lida por todos os colegas, há uma espécie de pressão extrínseca para que o aluno faça a atividade com mais cuidado e dedicação. Terminado o cotejo das traduções, o professor pode inclusive propor exercícios pós-tradução que podem ser relativos a assuntos inerentes à LP, à LC ou aos processos de tradução.

Neste capítulo revisamos a situação dos cursos de tradução no Brasil, mais especificamente no Estado de São Paulo, e revisitamos brevemente quatro

pesquisadores, Nord, Albir, Ruuskanen e Alves, que apresentam sugestões de abordagens didáticas para lidar com um texto a ser traduzido pelos alunos.

Nesta época de ecletismo em que vivemos, acreditamos que não haja apenas uma abordagem a ser adotada em aula, mas sim uma combinação de sugestões feitas, não só pelos teóricos citados neste capítulo, como por muitos outros que se preocupam com o ensino de tradução. Assim, sugerimos um curso que leve em conta as reflexões propostas por Amparo Albir (1999): quanto ao que se ensina, para quem se ensina e como se ensina. Do mesmo modo, acreditamos que, ao explicitar para o aluno as funções e objetivos do texto, como proposto por Christiane Nord em seu modelo funcionalista, a atividade de tradução torna-se mais ‘palpável’. Também adotamos o modelo CATLEGS de Deborah Ruuskanen (1996), que apresenta objetivos claros e explícitos. Por fim, aderimos à leitura por meio da identificação das Unidades de Tradução (UTs) apresentada por Fábio Alves (2006), que ensina ao aluno que tradução é muito mais do que traduzir palavra por palavra. A prática de uma leitura consciente do texto, em que o tradutor-aprendiz identifica as UTs presentes, viabiliza o processo de escolhas do futuro tradutor durante a produção de sua tradução, favorecendo, assim, a boa qualidade do texto na LC. Nessa linha, pretendemos apresentar como contribuição ao ensino de tradução uma metodologia para ser empregada durante a aula, principalmente em salas com número elevado de alunos. A planilha eletrônica, pouco utilizada como ferramenta pedagógica na área de letras, quando associada aos métodos de compilação e análise de *corpora* da Linguística de *Corpus*, possibilita que as aulas de tradução deixem de ser centradas no professor, abrindo inúmeras possibilidades de participação e envolvimento dos alunos em seu processo de aprendizagem.

Capítulo 3 – Materiais e métodos

Apresentamos neste capítulo o perfil dos alunos que participaram desta pesquisa e gentilmente cederam suas traduções para compor nosso *corpus* de estudo. Os resultados foram obtidos a partir das respostas dadas às perguntas específicas de um questionário. Junto a cada pergunta é apresentado um gráfico que permite formar uma visão global das características dos discentes envolvidos. Em seguida detalhamos como o *Microsoft Excel* tornou-se uma ferramenta fundamental para que esta pesquisa se desenvolvesse, pois sua multifuncionalidade nos permitiu usá-lo para uma série de tarefas, como a tabulação das respostas dos questionários e a transferência automática dos dados tabulados para os cabeçalhos dos textos, para a etiquetagem dos cabeçalhos e das linhas de cada unidade de tradução, e para o alinhamento tanto horizontal quanto vertical das inúmeras traduções com seus textos de partida. A riqueza de detalhes apresentada neste capítulo é intencional, pois acreditamos que os processos descritos aqui possam servir àqueles que queiram replicar a metodologia desta pesquisa.

3.1 Perfil dos participantes

A construção de nosso *corpus* de estudo parte das traduções elaboradas pelos alunos matriculados nos 1º e 2º semestres do curso de Letras, habilitações: tradutor e intérprete das Faculdades Metropolitanas Unidas, Campus Liberdade, São Paulo, SP, onde lecionamos a disciplina de Prática de Tradução. Apesar de compilarmos as traduções feitas por alunos desde 2010, por razões apenas de delimitação temporal, incluímos nesta pesquisa apenas arquivos do 2º semestre de 2011 ao 1º semestre de 2013. Por questões éticas só incluímos os textos dos alunos que nos deram autorização e responderam a um questionário²² informativo que nos permitiu traçar o perfil dos participantes. Tivemos a adesão de 261 alunos.

A primeira pergunta no questionário é sobre a faixa etária. A maior parte, 68%, se encontra na faixa de 17 a 25 anos, um indicativo de que, provavelmente, o curso de tradução é sua primeira faculdade. De modo geral, esses alunos podem ser divididos em dois grupos: aqueles que já conhecem a profissão e aqueles que

²² Um modelo do questionário utilizado encontra-se no Apêndice 1.

gostam de idiomas e veem no curso de tradução uma alternativa à licenciatura. Mais abaixo demonstramos como a sexta e a sétima perguntas do questionário nos auxiliam a chegar a essa conclusão. A segunda faixa etária dominante, 16%, tem mais de 40 anos e a terceira, 12%, entre 31 e 40 anos. Os alunos que estão nessas faixas muitas vezes decidiram fazer um segundo curso universitário pensando em trocar de profissão ou até mesmo ter a tradução como uma alternativa depois da aposentadoria. A menor faixa, 4%, engloba os alunos entre 26 e 30 anos. A divisão dos grupos etários foi feita com base no índice de distribuição da população de acordo com o censo do IBGE de 2010²³. O gráfico 3.1 abaixo demonstra a divisão de faixa etária dos participantes:

Gráfico 3.1 Faixa etária dos participantes

A segunda pergunta é sobre o sexo do participante. 76% dos participantes são do sexo feminino e 24%, do masculino. Praticamente para cada três mulheres há um homem. A identificação do sexo do aluno participante permite que, se for o caso, o pesquisador possa analisar e identificar se há ou não diferenças nos processos de escolhas entre homens e mulheres durante a tradução. Nesse primeiro momento, sem nenhum dado estatístico oficial, percebemos que no curso de tradução há maior presença masculina do que no curso de licenciatura. Logo abaixo, o gráfico 3.2 demonstra as porcentagens:

²³ A maior parte da população brasileira se encontra entre 10 e 24 anos. Como os alunos só podem entrar na faculdade a partir dos 17 anos, criamos a faixa 'de 17 a 25'. O segundo maior grupo populacional vai dos 25 aos 29 anos. Assim criamos a faixa 'de 26 a 30'. Por se tratar de um curso universitário, a princípio acreditávamos que grande parte dos alunos estaria em uma dessas duas faixas. Por essa razão, criamos as faixas 'de 30 a 40' e 'acima de 40'. No Anexo encontra-se uma tabela com a distribuição etária da população brasileira de acordo com o IBGE. Fonte: <http://vamoscontar.ibge.gov.br/atividades/ensino-fundamental-6-ao-9/49-piramide-etaria>.

Gráfico 3.2 Sexo dos participantes

A terceira pergunta é sobre o nível de proficiência em inglês. Ressaltamos aqui que não fizemos um teste de nível com os alunos, mas perguntamos como eles se reconheciam quanto ao seu conhecimento de inglês para fazerem o curso de tradução. As duas maiores faixas registram ‘avançado’, 52%, e intermediário, 24%. Interessante observar que 16% dos alunos consideram seu inglês de nível básico; entretanto, optaram por fazer um curso de tradução, talvez com o objetivo de aprender o idioma. O último grupo, 8%, considera-se ‘pós-avançado’. Esse é o único item do questionário que estipula um parâmetro: para ser pós-avançado, o aluno deve ter um certificado de proficiência como o *Cambridge Esol CPE*, *Michigan ECPE* ou similar. O gráfico 3.3 traz a distribuição por níveis de proficiência em inglês:

Gráfico 3.3 Nível de proficiência em inglês dos participantes

Dando sequência à questão linguística e ao conhecimento da cultura estrangeira, perguntou-se sobre a vivência no exterior. 84% dos alunos afirmam que

nunca viajaram para algum país de língua inglesa, 8% viveram em um país anglófono por até um ano, 4% de um a cinco anos e 4%, mais de cinco anos:

Gráfico 3.4 Alunos com vivência em países anglófonos

Quando comparamos com a pergunta anterior, sobre o nível de inglês, observamos que uma parcela alta dos alunos que consideram seu nível de inglês avançado nunca esteve em um país de língua inglesa. Tal constatação pode se refletir no modo como esses alunos percebem a cultura estrangeira.

Os dados obtidos a partir das respostas das perguntas 3, sobre o nível de proficiência, e 4, sobre a vivência no exterior, podem ser combinados pelo pesquisador para determinar se e o quanto a proficiência na língua estrangeira e a vivência no país onde o idioma estrangeiro é falado influenciam o processo de escolhas durante a tarefa de tradução para o português.

A pergunta seguinte, de número 5, trata a questão da língua nativa. Decidimos incluir essa pergunta pelo alto número de alunos estrangeiros, principalmente chineses, em nossas salas de aula. Todavia, não esperávamos que um fato inusitado ocorresse: nenhum deles quis contribuir com sua tradução para nossos estudos. Como nossa intenção não era, em momento algum, pressionar os alunos a participarem, não perguntamos a razão da recusa. Com isso a participação de alunos estrangeiros praticamente não acrescentou um novo elemento significativo a nossa pesquisa. O gráfico abaixo mostra apenas a proporção de alunos brasileiros e estrangeiros no universo de 261 participantes:

Gráfico 3.5 Língua nativa dos participantes

A sexta pergunta é sobre se os alunos já tiveram ou não alguma experiência como tradutores. A resposta foi dividida em três categorias: nenhuma experiência, 76%; alguns trabalhos amadores, 16%; e se já trabalhou ou trabalha como tradutor profissional, 8%. Essa pergunta abre a possibilidade de, durante as análises, comparar se o fator ‘experiência’ influencia ou não o tipo de escolhas tradutórias que os alunos fazem:

Gráfico 3.6 Experiência dos participantes em trabalhos de tradução

A última pergunta é sobre se os alunos já fizeram ou cursaram ao menos por dois anos algum outro curso universitário. 72% dos alunos responderam que não e 28%, que sim. Dentre os cursos mais mencionados temos os seguintes: jornalismo, comércio exterior, pedagogia, publicidade, editoração, filosofia, direito, gastronomia, administração de empresas, matemática, economia, moda, veterinária, comunicação social, fonoaudiologia, psicologia, informática, fisioterapia, secretariado, entre outros:

Gráfico 3.7 Participantes com outras áreas de formação

Ao compararmos os resultados da última pergunta com os da primeira (faixa etária), podemos concluir que a maioria dos alunos optou pelo curso de tradução como sua primeira faculdade. Não é nosso objetivo aqui, mas um estudo interessante seria tentar identificar como esse aluno percebe a profissão de tradutor.

De modo geral, com base nos gráficos podemos afirmar que a maior parte dos participantes desta pesquisa tem entre 17 e 25 anos, é do sexo feminino, tem nível avançado de inglês, nunca viajou para um país de língua inglesa, é falante nativo do português, nunca fez nenhum trabalho – mesmo que amador – de tradução, e nunca fez outro curso universitário.

Além de propiciar um perfil geral e identificar as características dos alunos, os dados obtidos a partir do questionário são utilizados no cabeçalho de cada tradução utilizada no *corpus*. A função desse cabeçalho é permitir ao pesquisador agrupar as traduções conforme critérios estabelecidos. Por exemplo, por faixa etária, por sexo, por experiência com tradução etc.

3.2 A compilação do *corpus*

Considerando-se as turmas dos turnos matutino e noturno, uma média de 30 a 50 alunos frequentam as aulas e entregam as traduções periodicamente. Na primeira semana de aula, os alunos recebem a versão digital de uma apostila contendo os textos que deverão ser traduzidos durante o semestre²⁴. Em data

²⁴ No Apêndice 2 há um exemplo do arquivo enviado aos alunos.

determinada, todos os alunos devem enviar via *e-mail* o texto traduzido, que deve estar alinhado conforme o modelo recebido. O texto na LP é segmentado em trechos, respeitando os limites dos períodos sintáticos e é disposto em uma página na posição ‘paisagem’ e traz uma tabela dividida em quatro colunas, sendo a primeira para a identificação numérica de cada trecho, a segunda para o texto de partida em inglês, a terceira para o aluno escrever sua tradução em português e a quarta é um espaço para as anotações feitas em aula. Verticalmente, a tabela é dividida conforme o número de períodos sintáticos presentes no texto original, acrescida de uma linha para o título. A segmentação dos textos na LP segue o modelo adotado por ferramentas digitais TAC – Tradução Assistida por Computador – como *Wordfast*, *OmegaT* e *Trados*. Os alunos são orientados a, inicialmente, fazer a tradução por período sintático. Ao terminarem, antes de enviarem ao professor, devem fazer a revisão da tradução e, caso julguem adequado, podem fazer junções de período ou outras alterações. Mais à frente, neste capítulo, na seção 3.4.3, explicamos como lidamos com os casos de omissões e junções. Para garantir que as traduções sejam enviadas, os alunos recebem créditos que serão somados à nota final do semestre.

Nosso *corpus* de estudo é composto por traduções que datam de agosto de 2011 a junho de 2013, e referem-se a oito textos em inglês selecionados a partir de uma coletânea de originais que são traduzidos no primeiro e segundo semestres do curso de tradução. Os textos que formam tal coletânea são simples, com cerca de 200-300 palavras, e cobrem temas como celebridades, aquecimento global, acidentes aéreos, racismo etc. A escolha dos oito originais que entraram para o *corpus* se deu por apresentarem dificuldades e estimularem discussões durante o processo de tradução. Além disso, para garantir não só a quantidade de traduções como a diversidade de alunos, o texto teve de ser traduzido por pelo menos duas turmas.

Como se trata de um *corpus* bilíngue, da mesma forma que os textos na língua de partida foram escolhidos a partir de parâmetros pré-estabelecidos, cada tradução também teve de seguir critérios para ser incluída. Primeiramente, só participaram da pesquisa os alunos que nos deram autorização por escrito para o uso do material; em seguida, cada texto foi individualmente avaliado, a fim de garantir que a tradução tivesse sido realmente feita pelo aluno e não por um tradutor

automático. Os alunos não são proibidos de usar tradutores automáticos, até porque, como os trabalhos são feitos em casa, não há como o professor controlar o seu uso. Entretanto, são enfaticamente desestimulados a utilizar tradução por máquina, já que estão em um curso para aprender a traduzir. Caso isso ocorra, o tradutor-aprendiz é orientado a fazer uma revisão do texto final. Esse cuidado foi importantíssimo para conseguirmos manter o nosso objetivo central, que é identificar os processos cognitivos dos alunos durante a tradução. Além da avaliação individual do professor, usamos para nos auxiliar nessa tarefa o *Machine Translation Detector*, que pode ser acessado *on-line* no site <http://www.translatordetector.com/>. Por fim, identificamos a individualidade do texto, ou seja, verificamos se não houve a ocorrência de dois ou mais alunos nos enviando a mesma tradução. Para isso, há duas possibilidades: uma é usar a ferramenta gratuita *Easy Duplicate Finder*²⁵. Após instalá-la, o seu uso é bem intuitivo. Primeiramente, seleciona-se a pasta que contém os textos para comparação, clica-se em *add* e finalmente em *start*.

Figura 3.1 Programa gratuito *Easy Duplicate Finder* que identifica textos iguais

Outra possibilidade é usar a ferramenta *Duplicates* no *WordSmith Tools*. Para encontrá-la, basta iniciar o programa, clicar em *File Utilities* e na aba *Duplicates*. No

²⁵ Disponível em: <http://www.easyduplicatefinder.com/duplicatefilecleaner.html>.

campo *Folder 1*, localizamos a pasta que contém os textos que devem ser comparados. Para iniciar a busca por textos iguais, clicamos em *search*. A figura 3.2 traz o menu sob *File Utilities*, onde encontramos a ferramenta *Duplicates*²⁶:

Figura 3.2 Ferramenta que busca por textos iguais no *WordSmith Tools*

Após considerarmos concluída a compilação, obtivemos um *corpus* bilíngue composto de oito textos em inglês e 798 traduções, totalizando 1.961 *tokens* e 816 *types* em inglês e 171.345 *tokens* e 4.692 *types* em português. A tabela 3.1 mostra a identificação de cada texto-fonte e o número total de traduções coletadas:

Identificação do texto	03	07	14	17	24	31	32	38
Número total de traduções	132	146	19	138	95	14	103	151

Tabela 3.1 Distribuição de textos traduzidos por cada texto-fonte

3.3 Microsoft Excel como uma ferramenta na Linguística de *Corpus*

A primeira fase da nossa pesquisa englobou dois estágios. O primeiro foi definir a metodologia de coleta das traduções que fariam parte de nosso *corpus*.

²⁶ O *AntConc* não possui ferramenta com essa função.

Esse processo foi o mais pragmático, pois, como relatado acima, os textos traduzidos foram coletados a partir dos exercícios que nossos alunos fazem semanalmente. De forma habitual, eles nos enviam por e-mail suas traduções já alinhadas em tabelas dispostas em arquivos do Microsoft Office Word em formato *doc*. Entretanto, para que possam ser trabalhadas em programas como o *WordSmith Tools* ou *AntConc*, inicialmente, as traduções devem ser separadas de seus originais e salvas em formato *txt*. Uma cópia dos arquivos em formato *doc* em que se encontra o original e sua tradução alinhados pelos alunos fica salva em uma pasta separada e pode, eventualmente, servir de arquivo *back-up*. O segundo momento foi a busca por uma metodologia de alinhamento do *corpus*. Bowker e Pearson (2002) afirmam que a vantagem principal de se trabalhar com um *corpus* bilíngue é a possibilidade de gerar linhas de concordância bilíngues. Tal fato é de fundamental importância para o nosso estudo uma vez que, para cada trecho do original, devemos analisar as soluções sugeridas por cada um dos tradutores-aprendizes a fim de buscarmos um entendimento de suas escolhas. Diferentemente do CorTrad²⁷, que trabalha com duas traduções de um mesmo texto, estabeleceu-se aquela que seria a característica mais marcante de nosso *corpus* de aprendizes: um original e suas múltiplas traduções. A figura 3.3 mostra o esquema do *corpus*, em que um texto de partida na língua 1, 'L₁', está paralelo '/' às suas traduções na língua 2, 'L₂', sendo que a sequência de letras, 'a', 'b', 'c', indicam os diferentes tradutores-aprendizes:

Texto orig L₁ // Texto trad L₂a + Texto trad L₂b + Textro trad L₂c ...n

Figura 3.3 Esquema das inúmeras traduções de um único texto de partida

Iniciamos, então, a busca por um programa que nos permitisse organizar um *corpus* bilíngue de modo que as linhas de concordância apresentassem o texto original e todas as suas referidas traduções. A princípio havíamos considerado fazer o alinhamento sentencial, ou seja, sentença no texto de partida alinhada à sentença de sua tradução no texto de chegada. Devido ao alto número de traduções, a

²⁷ Disponível em: http://www.fllch.usp.br/dlm/comet/consulta_cortrad.html.

viabilidade desse processo dependia de encontrarmos um método de alinhamento automático.

Nossa primeira tentativa foi com o programa EAlign²⁸. O fato de ser um programa antigo e de não haver um tutorial disponível dificultou a compreensão sobre seu funcionamento, mas superado esse problema inicial, descobrimos que o programa permitia alinhar apenas um original com sua tradução. Partimos então para um segundo programa, o CWB-Corpus Workbench²⁹. Infelizmente não conseguimos nem mesmo instalar o programa por problemas de compatibilidade com o sistema operacional Windows 7. Nossa tentativa seguinte foi o *ParaConc*³⁰, porém a limitação em alinhar no máximo três textos fez com que o descartássemos também.

Uma solução seria ter para cada tradução o seu texto original, porém esse modelo frustraria nosso objetivo de ter todas as traduções lado a lado. Outro problema é que esses programas precisam de uma lista de palavras-âncora, necessária para que consigam fazer o alinhamento automático do original e sua tradução. Uma lista de palavras-âncora contém uma série de palavras e suas possíveis traduções. O programa compara os segmentos na língua A com os segmentos na língua B com base nessa lista. Além da lista, os programas usam sinais de pontuação, nomes próprios e outros itens que não sofrem alteração no processo de tradução para buscar os segmentos equivalentes. Entretanto, como o nosso *corpus* tem como característica inúmeras traduções, a geração de uma lista de palavras-âncora demonstrou-se um tanto trabalhosa, se não impossível, pois para cada palavra âncora no texto da LP há várias possibilidades nos inúmeros textos na LC.

Assim, desistimos de encontrar um programa específico de alinhamento. Entre testes e tentativas que fizemos com esses diversos aplicativos acabamos por concluir que o caminho para encontrarmos uma solução para o alinhamento das traduções estava na codificação (*encoding*) do *corpus*.

A codificação é feita por meio de etiquetas de SGML (Standard Generalized Markup Language, ISO8879) escritas no formato <etiqueta de abertura> e </etiqueta

²⁸ No site *Sourceforge* há um link para o download do programa: <http://sourceforge.net/projects/ealign/>.

²⁹ Acesse <http://cwb.sourceforge.net/> para mais detalhes.

³⁰ Acesse http://athel.com/product_info.php?cPath=30_32&products_id=87 para mais detalhes.

de fechamento> (Berber Sardinha, 2004). Um *corpus* etiquetado permite o uso de cabeçalho com informações extratextuais como nome do tradutor, fonte, data da tradução etc., da mesma forma que também possibilita a etiquetagem de cada segmento do texto original de modo a combiná-lo com o segmento correspondente na sua tradução. Esses códigos desempenham um papel fundamental ao processar os textos no *WordSmith Tools*, pois as etiquetas nos permitem, por exemplo, estabelecer critérios de seleção de partes do arquivo e trabalhar apenas com as linhas que desejamos, facilitando a análise³¹.

Assim, demos nosso passo em busca por um programa que inserisse um cabeçalho em cada tradução e uma etiqueta para identificar cada uma das linhas. Passamos, então, para a fase de teste com o *Corpus Tool 2.8.7*³² e o *Sacodeyl Annotator 3.1*³³. O *Corpus Tool* é bem fácil de ser usado; porém, as etiquetas devem ser inseridas em cada texto individualmente. O *Sacodeyl Annotator* é um pouco mais complexo, mas também trabalha com textos individualmente. Da mesma forma que os programas de alinhamento textual não deram conta do elevado número de traduções, os etiquetadores também se limitavam a trabalhar com arquivos individuais. Devido a tais dificuldades, empenhamo-nos em buscar um método prático que nos permitisse alinhar as traduções, adicionar cabeçalhos e etiquetar cada linha em um sistema de *linha de montagem*. Após vários testes com aplicativos que não são tradicionalmente usados em linguística, a solução para as necessidades de nosso *corpus* foi encontrada no Microsoft Office *Excel*, o editor de planilhas desenvolvido pela Microsoft.

A princípio talvez cause estranheza a utilização de uma planilha eletrônica para manipular texto; contudo, esse tipo de programa manipula dados textuais de modo muito prático e flexível, e tais características se mostraram fundamentais como solução para nossa problemática de alinhamento e etiquetagem.

Uma planilha eletrônica é formada por 'células'. Como se fosse num jogo de batalha naval, a célula é uma caixa de dados onde uma coluna, identificada por letras, encontra-se com uma linha, identificada por números. Ou seja, uma célula

³¹ Mais à frente, na seção 3.4 neste capítulo, detalhamos o uso do *WordSmith Tools*.

³² Disponível em: <http://www.wagsoft.com/CorpusTool/features.html>.

³³ Disponível em: http://wiki.tei-c.org/index.php/Sacodeyl_Annotator.

denominada 'C6' quer dizer que é o encontro da coluna C com a linha 6, conforme ilustrado na figura 3.4:

Figura 3.4 Identificação das células na planilha eletrônica

O uso de fórmulas torna seu manuseio quase intuitivo e permite que o programa selecione, repita, mescle e junte os dados que se deseja manipular de uma célula, ou conjunto de células, em outra. Por exemplo, uma das fórmulas mais simples é a réplica de dados. Digamos que haja um texto na célula A1 e que queiramos repeti-lo nas células A2, A3 e B2. A fórmula utilizada para isso é '=A1'. Ao digitarmos essa fórmula em todas as células onde queremos que o dado se repita, clicamos na tecla 'enter' e a informação se replica, conforme apresentado na figura 3.5:

Figura 3.5 Exemplo de fórmula na planilha eletrônica

É importante observarmos que a planilha trabalha em camadas. Como podemos ver na figura 3.5, o texto está na célula, mas para sabermos qual é a fórmula que está gerando aquele dado temos que ler o campo 'fx'. A figura 3.6 mostra o uso de duas fórmulas para texto. A primeira, aplicada à célula A2, mostra o número de caracteres, inclusive espaços, no texto da célula A1. A segunda, aplicada à célula B3, não só copia o texto da A1 como também o transforma em letras maiúsculas:

Figura 3.6 Exemplos de fórmulas aplicadas a texto

Como podemos observar, as fórmulas dão muito dinamismo às planilhas. O comando '=(letra da coluna, número da linha)' foi uma das primeiras fórmulas que usamos. O seu uso viabilizou etiquetar todos os segmentos das traduções de uma só vez. A outra vantagem dessa fórmula de replicação é que, se for necessário modificar a codificação de alguma etiqueta, basta alterarmos o parâmetro na célula matriz e todos os parâmetros equivalentes nas células dos outros textos se alterarão também, ao mesmo tempo. Outra vantagem é que, ao copiar um texto cujos segmentos estejam dispostos em tabela em um arquivo *doc* ou cujos segmentos estejam cada um em uma linha de arquivo *txt*, a planilha eletrônica automaticamente identifica que da letra maiúscula até o ponto final se trata de um segmento, dispondo-o em uma célula. O processo é semelhante ao usado no *Wordfast* durante a segmentação do texto a ser traduzido. Em nossos testes conseguimos inserir um texto com cerca de 3.000 caracteres em uma só célula. Além da facilidade de inserção de linhas e colunas necessárias para o alinhamento de diversos textos, também há a possibilidade de se ter várias planilhas cujos dados podem ser

cruzados. Como se observa, a planilha eletrônica supriu todas as nossas necessidades.

Várias razões nos levaram a escolher especificamente o *Microsoft Office Excel*. A questão fundamental de usar um programa pago em vez de um programa aberto como o *LibreOffice* ou o *BrOffice* é que o pacote da Microsoft traz recursos de fórmulas que nem sempre estão disponíveis na versões abertas. Outro fator importante é que, por razões de investimento da Microsoft, há vastíssima documentação disponível gratuitamente *on-line*, tanto no *site* da própria Microsoft quanto em *sites* de usuários. Além do mais, a sua popularidade e disponibilidade em vários centros universitários públicos e privados tornam a divulgação e a troca de conhecimento entre pesquisadores muito mais fácil e prática.

3.4 Alinhamento e etiquetagem do texto usando o *Microsoft Excel*

Cada arquivo do *Excel* pode ser formado por inúmeras planilhas. A princípio iríamos utilizar apenas um arquivo com várias planilhas. Mesmo trabalhando em um computador *Intel Core™ i7*, esse arquivo ficou muito grande e se provou muito pesado e lento. Por esse motivo dividimos as traduções de acordo com as turmas dos alunos que as fizeram. O primeiro passo foi criar um número de planilhas no *Excel* de modo que pudéssemos ter onde lançar as informações dos alunos e alinhar os textos traduzidos. Ao abrir um arquivo novo, o programa automaticamente cria três planilhas em branco, denominadas 'Plan1', 'Plan2' e 'Plan3'. Logo depois da 'Plan3' há uma quarta guia que, ao ser clicada, cria mais planilhas. Ao clicar com o botão direito sobre a guia da planilha, abre-se um menu com vários comandos, entre eles 'Renomear' e 'Cor da Guia'. A figura 3.7 mostra a lista de comandos:

Figura 3.7 Menu disponível ao clicar na guia da planilha com o botão direito do mouse

A primeira planilha que criamos foi denominada ‘Lista de alunos’; as que seguem se referem aos textos traduzidos e são sempre em conjuntos de três. Nesse conjunto, a primeira traz o texto na LP alinhado com suas várias traduções horizontalmente; a segunda traz apenas os textos na LC com cabeçalhos e linhas etiquetadas, porém dispostos um logo abaixo do outro. Essa planilha utiliza várias linhas, mas apenas três colunas: uma para as etiquetas de abertura, uma para o texto e outra para a etiqueta de fechamento. Essa planilha serve de fonte para criar arquivos em *txt*. A terceira apresenta o texto original com seus segmentos distribuídos nas células horizontalmente e suas referidas traduções dispostas nas células logo abaixo. A figura 3.8 mostra a distribuição das planilhas. A seguir detalhamos cada uma das abas:

Figura 3.8 Distribuição das planilhas em um arquivo

3.4.1 Aba ‘Lista de Alunos’

A aba ‘Lista de alunos’ contém as informações do questionário respondido pelos alunos na fase das autorizações para uso das traduções. As perguntas sobre idade, sexo, experiência como tradutor, nível de conhecimento de inglês e viagens

para países de língua inglesa foram distribuídas horizontalmente nas colunas, como demonstrado na figura 3.9:

	A	B	C	D	E	F	G	H	I	J
1	Lista Geral dos alunos									
2	Esta lista é a tabulação dos questionários de todos os alunos. Também deve ser usada para controlar o número de identificação de cada aluno.									
3										
4	Nome do Aluno	Número que identifica o aluno	Ano_semestre e qdo a tradução 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. p_adv	Vivência no país de língua inglesa antes de iniciar o primeiro semestre: a. None b. 1yr c. 1-5yrs d. +5_yrs	Língua nativa: a. Português = Port b. Espanhol = Span c. Japonês = Jap d. Chinês = Ch e. Coreano = Cor	Experiência como tradutor antes de iniciar o primeiro semestre: a. None b. Amateur c. Profissional	Você fez outro curso superior (pelo menos 2 anos) antes de iniciar a faculdade de tradução? a. Não b. Sim – Qual?
5	2012_1 - 1o semestre manhã									
6	Ana_Karina	st_id=001	date=2012_1	age=31-40	gender=f	eng=p_adv	t_abr_1-5yrs	Native=Port	exp=amateur	Other_area=bio
7	Ana_Paula	st_id=002	date=2012_1	age=17-25	gender=f	eng=bas	t_abr_none	Native=Port	exp=none	Other_area=none
8	André_Luis	st_id=003	date=2012_1	age=17-25	gender=m	eng=inter	t_abr_none	Native=Port	exp=none	Other_area=none
9	Andre_Nardini	st_id=004	date=2012_1	age=17-25	gender=m	eng=adv	t_abr_none	Native=Port	exp=none	Other_area=none
10	Beatriz	st_id=005	date=2012_1	age=17-25	gender=f	eng=adv	t_abr_none	Native=Port	exp=none	Other_area=none

Figura 3.9 Exemplo da planilha ‘Lista de alunos’ com as respostas dos questionários tabuladas

Para que fosse possível a criação de gráficos e para facilitar os levantamentos estatísticos de cada grupo de alunos, foram estabelecidos parâmetros para a tabulação das respostas. A coluna A é para os nomes dos alunos. Os nomes não entram no cabeçalho das traduções, apenas o número designado na coluna B. Vale ressaltar aqui que apesar de os nomes não aparecerem nas traduções eles terão uma função nas fórmulas dos cabeçalhos; para que funcionem corretamente, os nomes devem estar em ordem alfabética, facilmente obtida por meio do comando ‘Classificar e Filtrar’: basta selecionar as colunas e as linhas que devem ser ordenadas, clicar e expandir ‘Classificar e Filtrar’. Feito isso, devemos clicar em ‘Personalizar Classificação’ e escolher qual coluna deve ser usada como âncora. No nosso caso foi a coluna A que, ao ser organizada em ordem alfabética, levou junto as informações das outras colunas.

O número de identificação alocado ao aluno na coluna B é sempre o mesmo. Assim, se em uma pesquisa quisermos avaliar as traduções feitas por um aluno específico, basta pesquisar pelo número, seguindo o parâmetro ‘st_id=xxx’, em que xxx, nesse momento, começa em 001 e vai até 270. Na parte de descrição do *corpus*, informamos que o total de alunos que aceitaram participar do projeto foi 261. Essa discrepância se dá devido ao fato de alguns alunos terem desistido da autorização do uso de suas traduções. Como a numeração foi dada em uma

sequência e os arquivos já haviam sido salvos, simplesmente retiramos esses alunos e seus números de identificação da sequência.

A coluna C indica o ano e o semestre, 'date=xxxx_1' ou 'date=xxxx_2' em que a tradução foi feita, sendo que 'xxxx' é o ano e os valores 1 e 2 indicam o primeiro e segundo semestres, respectivamente.

A coluna D traz os grupos etários. O código usado é 'age=xx-xx', em que xx é igual à idade inicial e final de cada faixa.

A coluna E tem dois parâmetros: 'gender=f', para indicar que a tradução foi feita por uma aluna, e 'gender=m', para indicar que foi feita por um aluno.

A coluna F informa o nível de inglês em autoavaliação dos alunos e tem os seguintes parâmetros: 'eng=bas' para o nível básico, 'eng=inter' para o intermediário, 'eng=adv' para o avançado e 'eng=p_adv' para o pós-avançado.

A coluna G indica se os alunos tiveram ou não a experiência de morar em um país de língua inglesa. Os parâmetros são 't_abr_none' para quem nunca viveu num país anglófono, 't_abr_1yr' para quem ficou até um ano, 't_abr_1-5yrs' para quem ficou de um a cinco anos, e 't_abr_+5_yrs' para quem morou mais de cinco anos.

A coluna H mostra a língua nativa do aluno. Como explicado acima, acabamos trabalhando praticamente apenas com falantes da língua portuguesa e alguns poucos falantes de espanhol e um de italiano e um de coreano. De qualquer maneira, os parâmetros usados foram 'Native=Port' para os falantes nativos do português, 'Span' para espanhol, 'Jap' para japonês, 'Ch' para chinês, 'Cor' para coreano e 'Ital' para italiano. Como a variedade de idiomas foi muito baixa a ponto de não trazer nenhuma característica representativa para a pesquisa, apesar de registrarmos o idioma, para fins estatísticos consideramos apenas 'língua portuguesa' em comparação com 'língua estrangeira'.

A coluna I registra a experiência do aluno como tradutor. Os parâmetros usados foram 'exp=none' para aqueles que nunca fizeram um trabalho de tradução, 'exp=amateur' para quem fez trabalhos amadores e 'exp=prof' para quem já fez trabalhos pagos como profissionais.

A última coluna é a J e indica se o aluno tem ou não outra área de conhecimento. A princípio íamos registrar quais as áreas presentes, mas percebemos que isso seria irrelevante e a variação, gigantesca; então passamos a marcar apenas se o aluno tem ou não outra formação. Dessa maneira, os

parâmetros usados ficaram 'Other_area=none' ou a abreviação de outras áreas como 'bio' para biologia, 'law' para direito etc.

No final de cada planilha 'Lista de Alunos' que se refere a uma turma há gráficos de cada uma das perguntas a fim de termos uma ideia geral das características daqueles alunos. Os gráficos são similares aos que apresentamos na parte 3.1, que são o resultado dos totais da tabulação do questionário.

A inserção das informações, como idade, conhecimento de inglês etc., obtidas a partir do questionário respondido pelos alunos, é feita de modo manual, uma a uma, porém a contagem dos totais registrados na tabela de tabulação é feito por meio da fórmula '=CONT.SE(D\$2:D\$28; "age=17-25")', como podemos observar na figura 3.10. Essa fórmula conta quantas vezes digitamos os parâmetros de cada pergunta. No exemplo, o programa conta quantas vezes inserimos o parâmetro 'age=17-25' na coluna D, entre as células D2 e D28, que correspondem ao início e final da lista de alunos do grupo de alunos cujo questionário está sendo tabulado. O mesmo processo é feito para os outros parâmetros e perguntas:

					=CONT.SE(D\$2:D\$28;"age=26-30")
Idade	a. De 17 a 25	b. De 26 a 30	c. De 31 a 40	d. Mais de 40.	
	17	1	3	4	
					=CONT.SE(E\$2:E\$28;"gender=f")
Sexo	Total fem	Total masc			
	19	6			
					=CONT.SE(F\$2:F\$28;"eng=inter")
Conh inglês	a. Básico	b. Inter	c. Avançado	d. Pós-Avan	
	4	6	13	2	
Vivencia ext	a. Nenhuma	b. Até 1 ano	c. De 1 a 5 anos	d. Mais de 5 anos	
	21	2	1	1	
Idioma nativo	a. Português	b. Espanhol	c. Japonês	d. Chinês	e. Coreano
	24	1	0	0	0
Exp. Tradut	a. Nenhuma	b. Alguns trab	c. Profissional		
	19	4	2		
Outro curso	a. Não	b. Sim			
	18	7			

Figura 3.10 Totais da tabulação e fórmulas para contagem automática

3.4.2 A criação de etiquetas

Além de alinhar as inúmeras traduções, a planilha eletrônica também foi a solução encontrada para etiquetarmos todos os textos para que pudessem ser manipulados no *WordSmith Tools* e no *AntConc*. Assim, o passo seguinte foi a criação de etiquetas. Bowker e Pearson (2002), Berber Sardinha (2004) e Zanettin (2012) sugerem que, ao criarmos etiquetas, sigamos as recomendações da TEI (*Text Encoding Initiative*)³⁴. Tomando por base a TEI, mas principalmente o seu uso no projeto English-Norwegian Parallel Corpus³⁵, passamos a desenvolver um conjunto de etiquetas que pudessem atender às necessidades de nossa pesquisa. Ressaltamos aqui que as nossas etiquetas são bem simples e se restringem ao uso nos programas citados. Se no futuro esta pesquisa se expandir de modo que o *corpus* passe a ser explorado por uma ferramenta *on-line*, extensões XHTML – *eXtensible HyperText Markup Language*³⁶ – deverão ser usadas. Não o fizemos neste momento porque não há como prever as necessidades que um site requererá para extrair informações do *corpus*.

3.4.3 Aba ‘Texto xx’

A planilha ‘Texto xx’, em que ‘xx’ se refere ao número que identifica o texto original e suas traduções, vem dividida em grupos de três colunas. As três primeiras colunas – A, B e C – chamamos *Conjunto Matriz do Texto de Partida*, que, por simplificação, passaremos a chamar CMTP. A partir de então, cada tradução ocupará um conjunto também de três colunas: D, E e F; G, H e I; J, K e L e assim por diante. O conjunto formado pelas colunas D, E e F, onde se encontra a primeira tradução, chamamos *Conjunto Matriz da Tradução*, que, para simplificar, será referido como CMTr. Essa divisão se faz necessária porque todas as fórmulas usadas na planilha terão tanto o CMTP quanto o CMTr como ponto de partida. A função de cada grupo de três colunas é a seguinte: na coluna da esquerda são inseridas as etiquetas de abertura delimitadas pelos símbolos ‘<’ no início e ‘>’ (sem

³⁴ Disponível em: <http://www.tei-c.org/index.xml>.

³⁵ Disponível em: <http://www.tei-c.org/Activities/Projects/en01.xml>.

³⁶ Disponível em: <http://www.w3.org/>.

as aspas) no final. A segunda coluna é reservada para as informações do cabeçalho e para os segmentos dos textos. A terceira coluna inclui as etiquetas de fechamento delimitadas pelos símbolos ‘</’ no início e ‘>’ no final, em que ‘/’ sempre será o primeiro elemento do parâmetro. Por exemplo, uma etiqueta de fechamento seria </p></ua>, em que há o fechamento do parágrafo e da unidade de alinhamento. A figura 3.11 ilustra como ficam as colunas no CMTr:

<i>f_x</i>		
D	E	F
<ua_id=pt03202><p_par=1>	textotextotextotexto	</p></ua>

Figura 3.11 Distribuição das etiquetas nas colunas CMTr.

As colunas que compreendem o CMTP, da linha 4 até a linha 14, contêm o cabeçalho do texto de partida. As colunas A e C são fixas e, para cada texto original, as informações do cabeçalho são preenchidas manualmente com dados do texto de partida. As etiquetas devem ser dispostas em dois grupos, chamados de ‘níveis’. No primeiro encontram-se as etiquetas que contêm informações do cabeçalho. No segundo nível, encontram-se as etiquetas que identificam os segmentos do texto. Essa divisão é necessária para que programas como o *WordSmith Tools* possam lê-las e interpretá-las independentemente, permitindo uma busca por um dado específico do cabeçalho combinado com determinada linha do texto. No CMTP há as seguintes etiquetas:

```
<header>
  <text_title>(Título do texto)</text_title>
  <text_source>(Origem do texto)</text_source>
  <tokens>(Total de palavras)</tokens>
  <types>(Total de itens)</types>
  <genre>(Gênero do texto)</genre>
  <theme>(Tema geral)</theme>
  <text_id>(Identificação numérica do texto)</text_id>
  <level>(Primeiro ou segundo semestres que formam o
    primeiro ano do curso)</level>
  <comments>(Se há algum comentário para o texto)</comments>
</header>
```

A figura 3.12 abaixo demonstra como fica a distribuição das etiquetas do cabeçalho do texto de partida na planilha eletrônica:

	A	B	C
1			
2			
3			
4	<header>		
5	<text_title>	Stress Management - Ways to Avoid Stress	</text_title>
6	<text_source>	http://www.webmd.com/balance/stress-management/stress-management-avoiding-unnecessary-stress	</text_source>
7	<tokens>	179	</tokens>
8	<types>	117	</types>
9	<genre>	informative	</genre>
10	<theme>	stress control	</theme>
11	<text_id>	32	</text_id>
12	<level>	1	</level>
13	<comments>		</comments>
14	</header>		

Figura 3.12 Exemplo do cabeçalho do texto de partida na planilha

O cabeçalho das traduções é um pouco diferente, pois na planilha do *Microsoft Excel* há camadas: uma superior no estilo WYSIWYG (*What you see is what you get*) e uma inferior, com fórmulas. Essas fórmulas nos permitem contar células e planilhas. Aqui as células entre as etiquetas do cabeçalho da tradução na planilha 'Texto xx' estão conectadas aos dados contabilizados a partir dos questionários respondidos que se encontram na planilha 'Lista de alunos'. A figura 3.13 demonstra como fica a camada superior com as etiquetas. As nove primeiras etiquetas entre <header> e </header> se referem às perguntas dos questionários e as cinco restantes copiam as informações das células do cabeçalho do texto de partida. Essas informações são todas inseridas de forma automática de modo que, se alterarmos alguma informação na planilha 'Lista de alunos', o item é imediatamente atualizado em todas as planilhas 'Texto xx'. Por exemplo, se identificarmos que tabulamos a faixa etária de um aluno de forma errada, basta alterar o dado na 'Lista de alunos' e todos os cabeçalhos das traduções feitas por esse aluno serão corrigidas. A célula-âncora, que permite a ligação com outras

planilhas, está na primeira linha, na coluna que fica entre as etiquetas. Nela inserimos a fórmula ‘=’Lista de alunos’!A2’, sem as aspas duplas, conforme mostra a figura 3.13:

D	E	F
	ALEXA	
<header>		
<student_id>	st_id=209	</student_id>
<date_of_translation>	date=2013_1	</date_of_translation>
<age>	age=31-40	</age>
<gender>	gender=m	</gender>
<english_background>	eng=inter	</english_background>
<years_abroad>	t_abr_none	</years_abroad>
<native_language>	Native=Port	</native_language>
<professional_experience>	exp=none	</professional_experience>
<other_field_background>	Other_area=bus	</other_field_background>
<genre>	informative	</genre>
<theme>	stress control	</theme>
<text_id>	32	</text_id>
<level>	1	</level>
<comments>		</comments>
</header>		

Figura 3.13 Exemplo da camada superior do cabeçalho das traduções

A figura 3.14 mostra a fórmula e os significados de cada elemento:

Figura 3.14 Fórmula na planilha ‘Texto xx’ para buscar informações na planilha ‘Lista de alunos’

Em seguida, nas linhas abaixo da coluna, inserimos outra fórmula que irá buscar as informações na planilha 'Lista de alunos'. A figura 3.15 mostra a fórmula sendo usado ao selecionarmos a célula:

=PROCV(E\$1;'Lista de alunos'!\$A\$2:\$J\$30;2)			
	D	E	F
		ALEX	
	<header>		
	<student_id>	st_id=209	</student_id>
	<date_of_translation>	date=2013_1	</date_of_translation>

Figura 3.15 Exemplo de fórmula que buscará a identificação do aluno na planilha 'Lista de alunos'

A fórmula pode parecer complicada à primeira vista, mas na realidade é simples. O comando 'PROCV' significa 'procurar valor'. Dentro dos parênteses, a letra E indica a coluna na planilha 'Texto xx'. Por exemplo, o próximo será 'H', pois a coluna F contém as etiquetas de fechamento, a coluna G contém as etiquetas de abertura da próxima tradução e a coluna H contém os dados do aluno, e assim por diante. Sendo de três em três, as colunas serão 'K', 'N', 'Q' etc. O símbolo '\$' significa que o próximo elemento não pode ser alterado. O número '1', presente em todas as fórmulas, indica o deslocamento. Como queremos que as informações estejam logo abaixo do nome, na coluna E, o deslocamento é 1: a fórmula está na coluna E e queremos que os dados apareçam na coluna E. Se quiséssemos que aparecessem na coluna F, o número seria 2. O ponto-e-vírgula é usado para separar os parâmetros de cada função. Agora mostramos de onde queremos que o programa retire as informações, colocando o nome da planilha entre aspas simples. O parâmetro 'Lista de alunos'!\$A\$2:\$J\$30 quer dizer que as informações estão na planilha 'Lista de alunos', em uma área de medida quadrada que vai da célula A2 até a célula J30. Por fim, o número após o ponto-e-vírgula indica a coluna, e não a linha, em que a informação será encontrada, por isso se inicia em '2', coluna B. O número '3' indica coluna C e assim por diante. A figura 3.16 abaixo ilustra essa área quadrada e mostra a equivalência numérica e as células de início e final que contêm os dados:

	A	B	C	D	E	F	G	H	I	J
		2	3	4	5	6	7	8	9	10
		Número que identifica o aluno	Ano, semestre 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. None b. Inter c. Adv d. P. Adv	Vivência no país de língua inglesa antes de iniciar o primeiro semestre: a. None b. 1yr c. 1-5yrs d. >5 yrs	Língua nativa: a. Português = Port b. Espanhol = Span c. Japonês = Jap d. Chinesa = Ch e. Coreano = Cor	Experiência como tradutor antes de iniciar o primeiro semestre: a. None b. Amador c. Profissional	Você fez outro curso superior (pelo menos um) antes de iniciar a tradução? a. Não b. Sim – Qual?
1	Nome do Aluno									
2	A	st_id=209	date=2013_1	age=31-40	gender=m	eng=inter	t_abr=None	Native=Port	exp=none	Other_area=Bus
3	BA	st_id=210	date=2013_1	age=17-25	gender=f	eng=bas	t_abr=None	Native=Port	exp=none	Other_area=none
4		st_id=211	date=2013_1	age=17-25	gender=f	eng=adv	t_abr=1yr	Native=Port	exp=none	Other_area=none
5	TI	st_id=212	date=2013_1	age=17-25	gender=f	eng=adv	t_abr=None	Native=Port	exp=none	Other_area=none
6	JINO	st_id=213	date=2013_1	age=17-25	gender=f	eng=adv	t_abr=None	Native=Port	exp=none	Other_area=none
7	RO	st_id=214	date=2013_1	age=26-30	gender=f	eng=bas	t_abr=None	Native=Port	exp=none	Other_area=none
8	RIZ GUTIERRE	st_id=215	date=2013_1	age=17-25	gender=f	eng=adv	t_abr=None	Native=Port	exp=none	Other_area=none
9		st_id=216	date=2013_1	age=17-25	gender=m	eng=adv	t_abr=1yr	Native=Port	exp=none	Other_area=none
10	FEIRA	st_id=217	date=2013_1	age=17-25	gender=f	eng=adv	t_abr=None	Native=Port	exp=none	Other_area=none
11	LARI GIROLDI	st_id=218	date=2013_1	age=17-25	gender=m	eng=adv	t_abr=5yrs	Native=Port	exp=none	Other_area=none
12	SILVA	st_id=219	date=2013_1	age=17-25	gender=f	eng=adv	t_abr=None	Native=Port	exp=prof	Other_area=InstTra
13	IRA	st_id=220	date=2013_1	age=17-25	gender=m	eng=adv	t_abr=None	Native=Port	exp=none	Other_area=none
14		st_id=221	date=2013_1	age=40	gender=f	eng=adv	t_abr=None	Native=Port	exp=amateur	Other_area=Lang
15	NA DE COLAUTTI	st_id=222	date=2013_1	age=40	gender=f	eng=inter	t_abr=None	Native=Span	exp=prof	Other_area=econ
16		st_id=223	date=2013_1	age=17-25	gender=f	eng=bas	t_abr=None	Native=Port	exp=none	Other_area=none
17		st_id=224	date=2013_1	age=31-40	gender=f	eng=bas	t_abr=None	Native=Port	exp=none	Other_area=none
18		st_id=225	date=2013_1	age=40	gender=f	eng=inter	t_abr=1-5yrs	Native=Port	exp=none	Other_area=jour
19	E CAMARGO	st_id=226	date=2013_1	age=17-25	gender=m	eng=adv	t_abr=None	Native=Port	exp=none	Other_area=none
20		st_id=227	date=2013_1	age=17-25	gender=f	eng=adv	t_abr=None	Native=Port	exp=amateur	Other_area=fash
21		st_id=228	date=2013_1	age=31-40	gender=f	eng=inter	t_abr=None	Native=Port	exp=none	Other_area=none
22	A SILVA	st_id=229	date=2013_1	age=17-25	gender=f	eng=inter	t_abr=None	Native=Port	exp=amateur	Other_area=none
23		st_id=230	date=2013_1	age=17-25	gender=f	eng=adv	t_abr=None	Native=Port	exp=none	Other_area=none
24	ALVES CARDOSO	st_id=231	date=2013_1	age=17-25	gender=m	eng=adv	t_abr=None	Native=Port	exp=amateur	Other_area=none
25		st_id=232	date=2013_1	age=17-25	gender=f	eng=adv	t_abr=None	Native=Port	exp=none	Other_area=none
26		st_id=233	date=2013_1	age=40	gender=f	eng=inter	t_abr=None	Native=Port	exp=none	Other_area=law
27										
28										
29										
30										

Figura 3.16 Equivalência numérica de cada coluna e extensão de onde começa e termina a extração dos dados

As fórmulas do cabeçalho ficam assim:

```
=PROCV(H$1;'Lista de alunos'!$A$2:$J$30;2)
=PROCV(H$1;'Lista de alunos'!$A$2:$J$30;3)
=PROCV(H$1;'Lista de alunos'!$A$2:$J$30;4)
=PROCV(H$1;'Lista de alunos'!$A$2:$J$30;5)
=PROCV(H$1;'Lista de alunos'!$A$2:$J$30;6)
=PROCV(H$1;'Lista de alunos'!$A$2:$J$30;7)
=PROCV(H$1;'Lista de alunos'!$A$2:$J$30;8)
=PROCV(H$1;'Lista de alunos'!$A$2:$J$30;9)
=PROCV(H$1;'Lista de alunos'!$A$2:$J$30;10)
```

O resultado na camada superior da planilha é vista como aquela exemplificada na figura 3.13 acima. Para finalizar o cabeçalho da tradução, precisamos dos dados que serão inseridos entre as etiquetas <genre></genre>, <theme></theme>, <text_id></text_id> e <level></level>. O conteúdo de cada etiqueta será a simples cópia das informações incluídas no cabeçalho da CMTP. Para isso, utilizamos a fórmula que copia o conteúdo de outras células. Na nossa planilha utilizamos o símbolo '\$' para que os valores não sejam acidentalmente alterados. As fórmulas da parte final do cabeçalho ficam assim:

=B\$9

=B\$10

=B\$11

=B\$12

É a partir do CMTr que o programa vai replicar as etiquetas que abrem e fecham os dados dos alunos extraídos da planilha ‘Lista de alunos’. Da coluna G por diante cada grupo de três colunas é chamado de ‘réplica de etiquetas’. Isso quer dizer que para cada tradução as etiquetas são as mesmas, variando apenas os dados dos alunos. Cada etiqueta de abertura e de fechamento é gerada a partir de uma fórmula que copia os dados inseridos nas colunas D, etiquetas de abertura, e F, etiquetas de fechamento, no CMTr³⁷. A praticidade que essa fórmula oferece é que o processo de réplica não tem limite. Além disso, como já dissemos, se num dado momento precisarmos alterar alguma etiqueta, ao modificá-la no CMTr, todas as outras etiquetas das traduções também se modificarão. A figura 3.17 mostra a camada das fórmulas e a camada em que podemos ler o resultado gerado a partir de tais fórmulas:

D	E	F			G	H	I
	CMTr		Fórmulas			Resultado das fórmulas	
<header>			=D3			<header>	
<student_id>	st_id=209	</student_id>	=D4	st_id=210	=F4	<student_id>	st_id=210
<date_of_translation>	date=2013_1	</date_of_translation>	=D5	date=2013_1	=F5	<date_of_translation>	date=2013_1
<age>	age=31-40	</age>	=D6	age=17-25	=F6	<age>	age=17-25
<gender>	gender=m	</gender>	=D7	gender=f	=F7	<gender>	gender=f
<english_background>	eng=inter	</english_background>	=D8	eng=bas	=F8	<english_background>	eng=bas
<years_abroad>	t_abr_none	</years_abroad>	=D9	t_abr_none	=F9	<years_abroad>	t_abr_none
<native_language>	Native=Port	</native_language>	=D10	Native=Port	=F10	<native_language>	Native=Port
ce	exp=none	</professional_experience>	=D11	exp=none	=F11	<professional_experience>	exp=none
d	Other_area=bus	</other_field_background>	=D12	Other_area=none	=F12	<other_field_background>	Other_area=none
<genre>	informative	</genre>	=D13	informative	=F13	<genre>	informative
<theme>	stress control	</theme>	=D14	stress control	=F14	<theme>	stress control
<text_id>	32	</text_id>	=D15	32	=F15	<text_id>	32
<level>	1	</level>	=D16	1	=F16	<level>	1
<comments>		</comments>	=D17	0	=F17	<comments>	0
</header>			=D18			</header>	

Figura 3.17 Uso da fórmula ‘=(coluna)(número da linha)’ para copiar o conteúdo de uma célula em outra

O cabeçalho ocupa da linha 3 à 18. Na linha 19, coluna A no CMTP e coluna D no CMTr, temos a etiqueta que abre o primeiro nível dos códigos que marcam os segmentos do texto. Essa etiqueta é usada como base para nomear os arquivos no

³⁷ Nas planilhas do *Excel*, as células mostram apenas o resultado gerado a partir das fórmulas. Para conhecer a fórmula por trás de um determinado resultado, temos que clicar na célula e verificar os dados exibidos na ‘Barra de Fórmulas’.

processo final de geração dos arquivos no formato *txt*. A figura 3.18 abaixo explica os significados de cada parte do primeiro parâmetro do texto. Mais à frente veremos que esses elementos são usados para nomear os arquivos *txt*:

Figura 3.18 Significado de cada elemento da etiqueta de abertura

A partir da célula A20 inserimos as etiquetas que nomeiam cada linha (os segmentos) do texto de partida. A figura 3.19 explica o significado do código que criamos para formar o parâmetro:

Figura 3.19 Significado de cada elemento das etiquetas das linhas

Nesse parâmetro apenas o número da linha varia. Junto a esse código podem aparecer mais duas etiquetas:

`<t>` indica que se trata de um título ou subtítulo do texto

`<p_par=1>` indica o início e o número do parágrafo

Essas etiquetas permitem, por exemplo, que o pesquisador selecione apenas os títulos ou um parágrafo inteiro. Mais à frente neste capítulo, explicamos como o

WordSmith Tools pode ser usado para isso. Todos os segmentos são fechados por `</ua>`. Para fechar esse segundo nível, utilizamos a etiqueta `</texto>`.

As traduções se iniciam com uma etiqueta similar à do texto de partida, porém com as informações do aluno: `<text_id=1pr_infor_stress_pt032_s209.txt>`. Aqui 'pt' quer dizer que o texto foi traduzido para o português, em seguida aparece o número de identificação do texto e, por fim, o 's' indica qual aluno fez aquela tradução, nesse caso, o aluno 209. As etiquetas que nomeiam as linhas das traduções seguem o mesmo modelo do texto de partida, apenas alterando 'en' para 'pt'. A figura 3.20 abaixo demonstra as primeiras linhas do CMTP e CMTr com as respectivas etiquetas:

19	<code><text_id=1pr_infor_stress_en032.txt></code>		<code><text_id=1pr_infor_stress_pt032_s209.txt></code>	
20	<code><ua_id=en03201><t></code> Stress Management - Ways to Avoid Stress <code></t></ua></code>		<code><ua_id=pt03201><t></code> Controlando o Stress - Maneiras de evitar o stress <code></t></ua></code>	
21	<code><ua_id=en03202><p_par=1></code> Stress is a part of life, and you can't always avoid it. <code></ua></code>		<code><ua_id=pt03202><p_par=1></code> O stress faz parte da vida e você não pode evitá-lo <code></ua></code>	
22	<code><ua_id=en03203></code> But you can try to avoid situations that can cause it, and you can control how you respond to it. <code></ua></code>		<code><ua_id=pt03203></code> Mas você pode tentar evitar situações que podem causá-lo e controlar como irá enfrentá-lo. <code></ua></code>	
23	<code><ua_id=en03204></code> The first step is knowing your own coping strategies. <code></ua></code>		<code><ua_id=pt03204></code> O primeiro passo é descobrir sua própria estratégia para enfrentá-lo <code></ua></code>	
	<code><ua_id=en03205></code> Try using a stress journal to record stressful events, your response to them, and how you coped. <code></ua></code>		<code><ua_id=pt03205></code> Tente anotar as situações de stress pelas quais você passou e como você as resolveu. <code></ua></code>	

Figura 3.20 Primeiras linhas etiquetadas tanto no CMTP quanto no CMTr

Réplicas da CMTr são coladas em grupos de três colunas. Como trabalhamos com salas grandes, já deixamos nosso modelo pronto para receber até 70 traduções. Com as colunas preparadas, passamos a inserir as traduções de cada aluno nas células apropriadas. Para esse processo, precisamos retomar os arquivos em formato *doc* que os alunos nos enviam por e-mail. Primeiramente selecionamos

e copiamos apenas a coluna que contém a tradução. Em seguida, acessamos a planilha do *Excel* correspondente à turma do aluno e, posicionando o mouse na célula imediatamente após a primeira etiqueta na linha 20, usamos o comando ‘colar’ >> ‘colar especial’ >> ‘colar texto sem formatação’. A figura 3.21 mostra a sequência dos passos:

Figura 3.21 Passos para usar o comando ‘colar especial’ no *Excel*

O mesmo processo é repetido para cada tradução que será copiada para as réplicas do CMTr. No final, ao observar todas as traduções juntas podemos facilmente identificar aquelas que ficaram díspares em seu alinhamento e podemos tratar cada uma localmente e identificar se o tradutor-aprendiz uniu ou separou períodos. Para alguns casos especiais que podem gerar alguma dificuldade no alinhamento dos textos, criamos códigos que indicam o que aconteceu com a tradução. Abaixo segue a lista desses códigos e a figura 3.22 exemplifica o uso do código que mostra que o aluno escolheu dividir o segmento do texto de partida em duas partes.

<! - - absolutamente nada - -> Quando por alguma razão o aluno não traduziu o segmento.

<! - - incompleto - -> Quando o segmento ficou incompleto. O código é usado exatamente no lugar em que falta a tradução, seja no começo, meio ou fim do segmento.

<! - - com o anterior- -> Quando o aluno uniu na tradução dois segmentos separados no texto original. O código é inserido na célula que se refere ao segundo segmento.

<! - - novo período - -> Quando o aluno dividiu um segmento do texto de partida em dois ou mais períodos sintáticos. Para não haver problemas no alinhamento, o texto dividido fica alocado em uma só célula e o código é inserido entre os dois períodos.

<ua_id=pt00705><p_par=3>	No entanto, o Prof. Steven Chu, Secretário de Energia do Governo Obama, está apoiando uma solução de baixa tecnologia. <! - - novo período - -> Ele sugere o uso de tinta branca.	</p></ua>
--------------------------	---	-----------

Figura 3.22 Uso de código para indicar que o aluno dividiu um segmento em dois períodos na língua de chegada

3.4.4 Aba 'Texto xx vert'

Essa segunda aba das traduções é gerada totalmente por meio de fórmulas. Sua função é colocar apenas as traduções etiquetadas em posição vertical, utilizando apenas três colunas: uma para a etiqueta de abertura, uma para os dados do cabeçalho e texto e uma para as etiquetas de fechamento. A função dessa planilha é facilitar a geração dos arquivos no formato *txt*.

Nessa planilha usamos apenas as colunas A, B e C. Todos os dados vêm da planilha 'Texto xx', que, no exemplo, é a planilha do texto 32. A fórmula usada é a deslocamento de dado: '=DESLOC('Texto 32'!D1;2;0;1;1)'. A única variante aqui é o elemento que indica a coluna de origem do dado.

Nesse caso, a coluna D se refere às etiquetas de abertura, 'E' se refere aos dados do questionário ou à tradução, e 'F', às etiquetas de fechamento. Para obtermos a próxima tradução, utilizamos as colunas G, H e I e assim por diante. O seguinte esquema de fórmulas

=DESLOC('Texto 32'!D1;2;0;1;1)		
=DESLOC('Texto 32'!D2;2;0;1;1)	=DESLOC('Texto 32'!E2;2;0;1;1)	=DESLOC('Texto 32'!F2;2;0;1;1)
=DESLOC('Texto 32'!D3;2;0;1;1)	=DESLOC('Texto 32'!E3;2;0;1;1)	=DESLOC('Texto 32'!F3;2;0;1;1)
=DESLOC('Texto 32'!D4;2;0;1;1)	=DESLOC('Texto 32'!E4;2;0;1;1)	=DESLOC('Texto 32'!F4;2;0;1;1)
=DESLOC('Texto 32'!D5;2;0;1;1)	=DESLOC('Texto 32'!E5;2;0;1;1)	=DESLOC('Texto 32'!F5;2;0;1;1)

etc.

resulta nos seguintes dados mostrados na figura 3.23:

	A	B	C
1	<header>		
2	<student_id>	st_id=209	</student_id>
3	<date_of_translat	date=2013_1	</date_of_translation>
4	<age>	age=31-40	</age>
5	<gender>	gender=m	</gender>

Figura 3.23 Exemplo do resultado da aplicação de fórmulas para trazer dados da planilha 'Texto xx'

No caso do texto 32, cada tradução com cabeçalho ocupa 38 linhas, que vão de <header> a </header> e depois de <text_id...> a </text>. O total de linhas ocupado por cada texto varia, pois tudo depende do número de segmentos. Esse conjunto de linhas se repete para cada uma das traduções. Como o número que identifica o aluno é fixo, e não foi em todas as ocasiões que todos os alunos entregaram a totalidade das traduções, há momentos em que células ficam vazias. Quando isso ocorre, o próprio programa adiciona o numeral '0' para indicar que a célula na aba 'texto xx' não contém nenhum dado³⁸.

O próximo passo foi salvar os dados da planilha 'Texto xx vert' para transformá-los em arquivos no formato *txt*. Há duas formas de se fazer isso. A mais

³⁸ Justifica-se assim o uso dos códigos <! -- absolutamente nada --> ou <!-- com o anterior --> a fim de evitar a adição automática do numeral 0.

simples é usar o recurso ‘Salvar como’ do programa e, no campo ‘Tipo:’, selecionar ‘Texto (separado por tabulações) (txt)’. Apesar de prático em um primeiro momento, há dois resultados que podem atrapalhar. Primeiro é que o nome da aba dentro do *Excel* se alterará para ‘txt’; segundo, como as planilhas são modelos que serão usados com todas as turmas, há um número muito maior de conjuntos de linhas que compreende cada texto do que há traduções, já que a média de alunos varia muito por sala. Assim, ao salvar a planilha como *txt*, estamos salvando todas as células preparadas para 70 traduções. Numa sala com 20 alunos, teremos 50 conjuntos de etiquetas de cabeçalho e textos vazios, que terão de ser descartados mais à frente.

Outra forma de se fazer esse trabalho, a nosso ver mais prática, é selecionando da célula A1 até a célula da coluna C na linha que contém a etiqueta `</text>` da última tradução. Feito isso, usamos o comando ‘copiar’ e colamos o conteúdo em um arquivo novo no programa Bloco de Notas. Esse arquivo, que contém todas as traduções, é salvo como ‘textox.txt’. O próximo passo engloba separar as traduções em arquivos individuais. Para isso usamos o programa *WordSmith Tools*. Na interface de abertura do programa, à esquerda, há um índice com as várias ferramentas existentes. Aqui clicamos em *Utilities*, que nos leva a outro índice de ferramentas, onde selecionamos a função *File Utilities*. Ao abrir a janela de trabalho, clicamos na aba *Splitter*. Nessa janela inserimos os comandos que automaticamente separarão o arquivo ‘textox.txt’ em arquivos individuais. Como todas as traduções terminam com `</text>`, usaremos esse elemento como separador. Para que funcione é necessário selecionar o comando *end of text*. Em seguida localizamos a pasta de destino em *Destination folder*: e selecionamos o arquivo que deverá ser dividido em ‘File(s) to Split:’ Feito isso, é só clicar em *Split now*. A figura 3.24 abaixo ilustra esse processo:

Figura 3.24 Processo de divisão do arquivo 'Textox' em traduções individuais

O programa cria uma pasta com o nome do texto no diretório que escolhemos dentro de *Destination folder*. e nela salva os arquivos individualmente, nomeando cada um na sequência 0001, 0002, 0003 etc. de forma automática. Entretanto, precisamos que os arquivos sejam nomeados de forma que possamos identificá-los facilmente. Para isso usamos outra ferramenta do *File Utilities*: a aba *Rename*. Nessa aba localizamos a pasta que contém os arquivos a serem renomeados no campo *Folder*. Abaixo já estará inserida a terminação '*.txt'. O asterisco significa que ele buscará na pasta arquivos com qualquer nome que tenha a extensão *txt*. Agora basta clicar no botão *Find Files*. Feito isso aparecerá a lista de arquivos a serem renomeados. No campo *Mask for new name* inserimos o novo nome. Como temos todas as traduções em ordem numérica, utilizamos o símbolo '#' para indicar que ali deverá constar o número do texto. Um exemplo de nome é '1pr_infor_stress_pt032_s###', em que o programa numerará cada arquivo usando três dígitos a partir do número que inserimos no campo *Start number*. Na figura 3.25 abaixo podemos observar que o número inicial é 234, pois essa é a identificação numérica do primeiro aluno da turma. Justifica-se assim a razão para manter todos os nomes dos alunos, mesmo daqueles que não fizeram a tradução. Nessa turma, por exemplo, os alunos vão de 'st_id=234' a 'st_id=270'. Nem todos entregaram a

tradução do texto 032, mas nesse momento o que nos interessa é que o programa siga a ordem numérica. Agora é só clicar no botão ‘Rename’, conforme apresentado na figura 3.25:

Figura 3.25 Processo de renomeação dos arquivos no *WordSmith Tools*

Agora temos que eliminar os arquivos que não contêm tradução. Essa tarefa é rapidamente feita com auxílio do File Explorer. Abrimos a pasta que agora possui as traduções individuais e renomeadas e vamos visualizando uma a uma, com o comando ‘View’ >> ‘Preview Pane’ acionado. Se depararmos com um arquivo que no lugar de texto tiver o dígito 0, simplesmente o apagamos. Na imagem abaixo (figura 3.26), observamos que os arquivos 258 e 264 já foram eliminados e agora será a vez do arquivo 266:

Figura 3.26 Eliminação dos arquivos que não contêm tradução

Terminada essa tarefa, a pasta está pronta para ser adicionada ao *corpus* e os arquivos já estão prontos para serem manipulados pelo *WordSmith Tools* ou *AntConc*. O resultado final são arquivos individuais com cabeçalho e com cada linha etiquetada como mostra a figura 3.27 abaixo:

Figura 3.27 Exemplo de tradução com cabeçalho e etiquetagem

3.4.5 Aba 'Texto xx alinhado'

Uma das questões cruciais que precisavam ser resolvidas nesta pesquisa era uma maneira de alinhar o segmento do texto de partida com todas as suas traduções. O *WordSmith Tools* oferece uma ferramenta de alinhamento, 'Viewer & Aligner', que se mostra muito prática para alinhar um número limitado de textos. No nosso caso, tal ferramenta não foi tão útil, pois, primeiramente, cada arquivo tem de ser salvo individualmente em uma extensão específica: 'vwr'. Em seguida, os textos devem ser juntados por meio da ferramenta 'Merge with', e o resultado será um arquivo com uma nova extensão: 'ali'. Além desses vários passos com cada um dos textos, a ferramenta faz o alinhamento com base na pontuação. Dessa forma, quando há traduções em que o aluno deixou de traduzir algum segmento ou juntou ou separou os períodos sintáticos, ocorre um 'desalinhamento', o que acaba por atrapalhar a análise.

Mais uma vez o *Microsoft Excel* serviu para nossos propósitos. Nossa necessidade era ter em uma linha o texto de partida e nas linhas de baixo todas as traduções, de modo a termos uma macrovisão de cada segmento. A solução para essa questão está na aba 'Texto xx alinhado'. Utilizando a fórmula '=DESLOC('Texto 32'!\$B\$20;0;A1)', a planilha exibe na primeira linha o segmento do texto de partida e nas linhas abaixo as traduções que os alunos da turma fizeram. A fórmula 'DESLOC' é usada justamente para buscar dados em outra planilha. Essa mesma fórmula já foi usada na aba 'Textoxx vert', que organiza as traduções verticalmente. O resultado é mostrado na figura 3.28. *Stress management – ways to avoid stress* é o título do texto:

B
Stress Management - Ways to Avoid Stress
Controlando o Stress - Maneiras de evitar o stress
Administração do estresse - maneiras de evitar o estresse
Gestão de Estresse - Maneiras de evitar o estresse
0
Gerenciando o estresse - maneiras de evitar o estresse
Administração do estresse - Caminhos para evita-lo
Stress X Tempo - Como evitar o Stress.
Manegamento do Estresse-Uma forma de evitar o estresse
0
0
<! - - absolutamente nada - ->
Tratamento de estresse - Modos de evitar o estresse
Controle de estresse - Maneiras de evitar o estresse
Gerenciamento do estresse - Maneiras de evitá-lo
Gerenciamento do estresse; maneiras de evitar o estresse
Dominando o estresse-Como evitá-lo
Gerenciamento do Estresse - Maneiras de evitá-lo
Controlando o estresse - Maneiras de evitar o estresse
Gerenciamento de estresse - Maneiras de se evitar o estresse
Gestão de estresse - Maneiras de evitar o estresse.
Estresse - Maneiras de evitá-lo
Administração do estresse - Modos de evitar estresse
Administando o estresse - maneiras de se evitar o estresse
Administração do Estresse - Meios de se evitar o estresse
Gerência do Estresse - Maneiras de Evitar o Estresse

Figura 3.28 Todas as traduções do segmento do texto de partida alinhadas verticalmente

Nessa planilha conseguimos facilmente analisar todas as sugestões de tradução, assim como identificar aquela que um aluno simplesmente não traduziu. Essa disposição é ideal para uma análise *corpus-driven*, ou seja, orientada pelo *corpus*. Na sessão 3.4 explicamos como fazemos a análise das linhas no *WordSmith Tools*.

Ressaltamos que apenas o primeiro conjunto de planilhas dá um pouco mais de trabalho devido à inserção das fórmulas. Em seguida, tal conjunto serve de modelo para a etiquetagem e alinhamento dos outros textos, necessitando apenas ajustar o número de segmentos presentes no texto original e nas traduções. Para cada turma basta inserir os dados dos questionários na aba 'Lista de alunos' e o nome dos alunos na célula logo acima do cabeçalho na aba 'Texto xx' para que os dados sejam atualizados.

Abaixo apresentamos um esquema que resume a nossa metodologia de preparação das traduções para análises:

3.4 Uso do *WordSmith Tools* para manipular as traduções

O *WordSmith Tools* é um pacote de aplicativos pago desenvolvido pelo linguista inglês Mike Scott da Universidade de Liverpool, no Reino Unido. A versão utilizada neste trabalho é a de número 6; com as atualizações de 22 de dezembro de 2013 chegamos ao *release* 6.0.0.168. Ao iniciar o *WordSmith Tools*, há na parte superior um *link* para três ferramentas de manipulação de *corpus*: 'Concord', 'Keywords' e 'Wordlist'. Sempre iniciamos pela função 'Wordlist'. Como o nosso *corpus* tem cabeçalho e as linhas são etiquetadas, antes de gerar a lista de palavras precisamos estabelecer alguns parâmetros de seleção para que esses elementos extratextuais não interfiram no resultado almejado. Ao abrir essa funcionalidade, selecionamos na barra de tarefas 'Settings' e depois 'tags'. A janela 'Advanced Settings' se abrirá. O campo 'Mark-up to ignore' deverá ser '<*>' (sem as aspas) em que o asterisco significa qualquer conteúdo entre o sinal de menor e maior. Esse comando evita que o conteúdo das etiquetas seja contabilizado na contagem de palavras. Outra questão a ser resolvida é a não inclusão das informações dos cabeçalhos nas listas de palavras. Para isso vamos até o campo 'Document header

end:’ e inserimos a etiqueta que fecha o nosso cabeçalho – `</header>`³⁹. A figura 3.29 ilustra os ajustes realizados:

Figura 3.29 Janela de configurações no *WordSmith Tools*

Em seguida, clicamos em ‘Close’, voltamos à barra de tarefas e clicamos em *file* e depois em *new*. Abre-se uma pequena janela e, na aba *Main*, clicamos em *Choose Texts Now*, que nos levará para o gerenciador de arquivos do programa, onde localizamos manualmente a pasta que contém os arquivos que serão processados, lembrando que somente os arquivos em formato *txt* podem ser adicionados. Feito isso, clicamos em *ok* e depois em *Make a Word list now*. Esse processo gerará uma lista de palavras utilizando apenas os elementos textuais, descartando todo o cabeçalho e as etiquetas. As onze palavras mais frequentes das traduções que estão sendo analisadas podem ser observadas na figura 3.30:

³⁹ A etiqueta `<teiHeader>` já vem como default; entretanto, não é necessário que seja em inglês. O obrigatório aqui é que o comando se refira exatamente à etiqueta usada no texto. Por exemplo, se foi `</Header>` com a primeira letra em maiúscula, isso deverá ser respeitado. O mesmo se a etiqueta estiver em português `</cabeçalho>`.

N	Word	Freq.	%	Texts	% Lemmas	Set
1	DE	1.922	6,42	146	100,00	
2	O	1.284	4,29	146	100,00	
3	MAIS	822	2,75	146	100,00	
4	QUE	750	2,51	146	100,00	
5	#	724	2,42	146	100,00	
6	DO	611	2,04	142	97,26	
7	A	605	2,02	142	97,26	
8	E	567	1,89	146	100,00	
9	AS	493	1,65	145	99,32	
10	CIDADES	421	1,41	146	100,00	
11	NO	415	1,39	144	98,63	

frequency alphabetical statistics filenames notes

1.137 entries Row 1 DE

Figura 3.30 Lista parcial de palavras gerada no *WordSmith Tools*

Na figura 3.30 vemos as dez primeiras palavras (o elemento ‘#’ indica que se trata de um numeral), em ordem decrescente de frequência, de uma lista extraída de 146 textos. A lista mostra 1.137 *types* (tipos) de um total de 29.926 *tokens* (formas). A palavra mais frequente é ‘de’, com 1.922 ocorrências, que representa 6,42% de todas as palavras. Como aparece nos 146 textos, tem uma presença em 100% dos textos.

Caso queiramos fazer uma contagem por lemas, ou seja, lematizar uma determinada palavra, o que temos de fazer é clicar no(s) elemento(s) e arrastá-lo até uma palavra que será o núcleo do lexema e soltar o botão esquerdo do mouse⁴⁰. A figura 3.31 demonstra como juntamos à preposição ‘de’ as variáveis ‘da’, ‘das’, ‘do’ e ‘dos’. A aba *Alphabetical* facilita a localização dos lemas por trazê-las em ordem alfabética:

⁴⁰ Se os lemas estiverem muito longe um do outro, clicamos no elemento que queremos unir e, segurando o botão esquerdo do mouse, utilizamos a roda central de rolagem para chegarmos até o ponto da lista onde está a palavra âncora, para então fazer a junção liberando o botão com o cursor bem em cima dela.

338	DAS	96	0,32	74	50,68	
340	DE	2.894	9,67	0		de[1922] do[611] da[201] das[96] dos[64]
341	DECORRÊNCIA	2		2	1,37	
342	DEFENDE	5	0,02	5	3,42	
343	DEFENDENDO	3	0,01	3	2,05	
344	DEIXANDO	1		1	0,68	
345	DEIXAR	1		1	0,68	
346	DEIXARIA	1		1	0,68	
347	DEIXARIAM	1		1	0,68	

Lemma Forms

DE	1.922
DO	611
DA	201
DAS	96
DOS	64

Figura 3.31 Lexema organizado na lista de palavras

Para desfazer a lista de lemas, basta clicar sobre a palavra que está sendo usada como núcleo do lexema e em seguida clicar em Ctrl+F4.

A aba *Statistics* traz informações gerais do *corpus* e particulares de cada texto que foi usado para análise. Como já mencionamos anteriormente, quanto mais baixo o valor da média entre *types* e *tokens*, mais repetidas são as palavras no texto. O resultado da conta deve ser um número entre 0 e 100, porém os valores de referência são bem relativos. Quanto maior for o *corpus*, mais baixo será o valor de referência. Um *corpus* com cerca de 300 textos e um milhão de palavras pode ter a razão entre *types* e *tokens* de 2 a 6. Dependendo do que se deseja pesquisar esse valor não tem muita importância. Em nossa análise tal resultado é relevante, pois fazemos a comparação individual entre as traduções. A figura 3.32 mostra a aba *Statistics* do texto 07:

WordList

File	Edit	View	Compute	Settings	Windows	Help	
N	text file	file size	tokens (running words) in	tokens used for word list	sum of	types (distinct)	type/token ratio
1	Overall	318.972	29.946	29.222	1.137	3,89	
2	text_id=1pr_info_art_pt007_s025.txt	2.139	190	187	139	74,33	
3	1pr_info_art_pt007_s153.txt	2.155	198	193	138	71,50	
4	text_id=1pr_info_art_pt007_s130.txt	2.076	186	184	131	71,20	
5	text_id=1pr_info_art_pt007_s065.txt	2.060	182	177	126	71,19	
6	text_id=1pr_info_art_pt007_s076.txt	2.205	199	194	137	70,62	
7	text_id=1pr_info_art_pt007_s121.txt	2.010	177	173	122	70,52	
8	text_id=1pr_info_art_pt007_s010.txt	2.137	192	187	131	70,05	
9	1pr_info_art_pt007_s154.txt	2.004	174	169	118	69,82	
10	text_id=1pr_info_art_pt007_s050.txt	2.287	210	205	143	69,76	
11	text_id=1pr_info_art_pt007_s037.txt	2.148	196	191	133	69,63	
12	text_id=1pr_info_art_pt007_s024.txt	2.202	201	196	136	69,39	
13	text_id=1pr_info_art_pt007_s048.txt	2.181	204	199	138	69,35	
14	text_id=1pr_info_art_pt007_s045.txt	2.189	207	202	140	69,31	
15	text_id=1pr_info_art_pt007_s141.txt	2.206	207	202	140	69,31	
16	1pr_info_art_pt007_s157.txt	2.181	200	195	135	69,23	
17	text_id=1pr_info_art_pt007_s052.txt	2.150	200	195	135	69,23	
18	text_id=1pr_info_art_pt007_s096.txt	2.258	215	210	145	69,05	
19	text_id=1pr_info_art_pt007_s085.txt	2.212	208	203	140	68,97	
20	text_id=1pr_info_art_pt007_s047.txt	2.219	210	206	142	68,93	

Figura 3.32 Aba *Statistics* mostra, parcialmente, em destaque a razão entre *types* e *tokens* do texto 07

Temos um total de 147 traduções e queremos saber o quão criativos ou repetitivos os alunos são. A coluna em destaque na figura mostra a razão entre *types* e *tokens* de cada texto. Podemos notar que o valor de referência é outro, ficando entre 59,22 e 74,33. Um valor alto indica que o aluno usou uma variedade lexical maior e que talvez não tenha feito repetições desnecessárias. Poderíamos considerar que esse índice indica o nível de criatividade do aluno além de estimular outros tipos de estatísticas. Por meio dos cabeçalhos das traduções, poderíamos tentar identificar, por exemplo, o sexo e a faixa etária dos alunos com os índices da razão entre *types* e *tokens* mais altos.

O *WordSmith Tools* é um programa que oferece várias possibilidades de análise de um *corpus*. Além da análise dos textos de forma integral, como mencionamos anteriormente, com textos etiquetados há a possibilidade de selecionarmos apenas a linha que queremos analisar. Para isso, basta estabelecer os parâmetros de seleção. Ao iniciar o programa, abrimos a ferramenta 'Wordlist' na

barra de tarefas e selecionamos 'Settings'. No menu que se abrirá, clicamos em 'tags', que abrirá uma janela chamada 'Advanced Settings'. As configurações feitas anteriormente devem ainda estar ativas. Por exemplo, quando geramos a nossa primeira lista de palavras do texto 07, configuramos o programa para desconsiderar o cabeçalho. Agora clicamos em 'Only Part of File' e uma nova janela se abrirá (figura 3.33):

Figura 3.33 Janela para selecionar apenas determinadas linhas dos textos

No campo *Sections to keep* adicionamos a linha que queremos analisar. Digamos que quiséssemos analisar a terceira linha do texto 07. No primeiro campo inserimos `<ud_id=pt00703>` e, no segundo campo, a etiqueta de fechamento `</ua>`. A lista de palavras resultante conterá apenas aquelas encontradas na linha 3 das traduções do texto 07 (figura 3.34):

N	Word	Freq.	%	Texts	% Lemmas
1	QUE	285	8,53	141	96,58
2	#	283	8,47	142	97,26
3	AS	201	6,02	122	83,56
4	DE	182	5,45	129	88,36
5	MAIS	139	4,16	138	94,52
6	CIDADES	138	4,13	137	93,84
7	RURAIS	134	4,01	134	91,78
8	JÁ	128	3,83	126	86,30
9	CALOR	127	3,80	127	86,99
10	O	123	3,68	120	82,19

Figura 3.34 Lista de palavras apenas da linha 3 do texto 07

A figura 3.34 mostra as 10 primeiras palavras de uma lista feita apenas com a linha 3 do texto 07.

Além de selecionar uma linha específica, outro tipo de filtro que o *WordSmith Tools* nos permite empregar é a seleção dos textos a partir das informações que estão no cabeçalho de cada texto. Digamos que, por exemplo, dentro da linha 03 do texto 07 quiséssemos analisar as soluções encontradas nas traduções feitas apenas por alunos que nunca tenham viajado para um país de língua inglesa e cujo nível de inglês seja básico. As etiquetas que queremos são

```
<years_abroad>t_abr_none
<english_background>eng=bas
```

sem a necessidade das etiquetas de fechamento. Sempre que mudarmos os parâmetros de pesquisa, precisamos gerar uma nova lista de palavras. Antes de gerar a lista de palavras, na barra de tarefas, selecionamos 'Settings' e depois *Tags*. A janela *Advanced Settings* se abrirá. Após verificar que os parâmetros de *Mark up to ignore* e *Document heading ends*: estejam configurados conforme explicado, clicamos em *Only If Containing* e uma nova janela se abrirá, chamada *Select Only text files containing....* Nessa janela podemos adicionar até quatro filtros (ou seja, as etiquetas e seus valores, como indicado acima), unidos pelo comando *plus*. O

programa também permite o uso de alternativas. Digamos que nossa intenção seja analisar a produção apenas dos alunos que nunca tenham viajado ou que tenham vivido por menos de um ano em um país de língua inglesa. Para isso, basta inserirmos no campo imediatamente ao lado a etiqueta e o seu valor correspondente. Os campos abaixo de *but not* servem para a exclusão de textos. A figura 3.35 abaixo demonstra o uso dos parâmetros do cabeçalho como filtros:

Figura 3.35 Janela do *WordSmith Tools* onde se pode selecionar os textos a partir das informações no cabeçalho

Caso queiramos analisar apenas a linha 03 das traduções desses alunos basta aplicar o filtro *Only Part of File*, como demonstrado acima. A figura 3.36 mostra as 10 primeiras palavras da lista resultante da seleção com os filtros propostos aplicados:

WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%	Texts	% Lemmas	Set
1	#	28	8,64	14	9,59	
2	QUE	27	8,33	14	9,59	
3	AS	22	6,79	13	8,90	
4	DE	18	5,56	13	8,90	
5	O	15	4,63	14	9,59	
6	MAIS	14	4,32	14	9,59	
7	RURAI	14	4,32	14	9,59	
8	CALOR	13	4,01	13	8,90	
9	CIDADES	13	4,01	13	8,90	
10	JÁ	13	4,01	13	8,90	
11	ÁREAS	12	3,70	12	8,22	

frequency alphabetical statistics filenames notes

59 entries Row 1 #

Figura 3.36 Lista de palavras da linha 03 do texto 07 após a aplicação das etiquetas '<years_abroad>t_abr_none' e '<english_background>eng=bas' do cabeçalho usadas como filtros

Para analisar a palavra em seu contexto, basta clicar com o botão direito do mouse e selecionar *Concordance* e a janela dessa ferramenta se abrirá (figura 3.37):

Concordance

File Edit View Compute Settings Windows Help

N	Concordance	Set
1	, Prof. Steven Chu, secretário de Energia de Obama, está apoiando	
2	, Prof. Steven Chu, secretario de energia de Obama, está apoiando uma	
3	, o Prof. Steven Chu, secretário de energia de Obama, está criando uma	
4	, apoia a solução de baixo uso de energia para alguns dos nossos	
5	o professor Steven Chu, secretário de energia de Obama, está apoiando uma	
6	o professor Steven Chu, Secretário de Energia da administração de (Obama o	
7	Professor Steven Chu, secretário de energia do presidente Obama, tem	
8	geladeiras. Entretanto, o secretário de energia de Obama, Prof. Steven Chu,	
9	, o Prof. Steven Chu, secretário de energia do presidente Obama, apóia	
10	que estas áreas. O consumo de energia fica mais intenso a medida	

Figura 3.37 Dez primeiras linhas de concordância da palavra 'energia'

A construção de um *corpus* não é uma tarefa simples e menos ainda é sua análise. O nosso *corpus* de aprendizes de tradução envolveu uma série de passos que dependeu dos alunos e de sua vontade de participar do projeto, aplicação e

tabulação de questionários e a manipulação de cada arquivo incluído. Ao mesmo tempo em que a necessidade de alinhar múltiplos textos tornou a nossa pesquisa original, isso também nos trouxe uma grande dificuldade: que tipo de *software* usar para resolver os alinhamentos. Os programas desenvolvidos para esse propósito não supriram nossas necessidades e tivemos que buscar alternativas não tradicionais. A solução foi encontrada no *Microsoft Excel*. Pouco usado pelos pesquisadores de Linguística de *Corpus*, a sua flexibilidade no uso de inúmeras linhas e colunas e principalmente pela capacidade de usar fórmulas que extraem e manipulam dados foi o programa ideal para nos auxiliar a inserir cabeçalhos e a etiquetar cada uma das linhas das 798 traduções que compõem nosso *corpus* de estudo. Além disso, foram as planilhas do *Excel* que também resolveram a grande problemática que tínhamos: como alinhar um original e cerca de 150 traduções de modo a termos em uma linha o texto da língua de partida e nas linhas abaixo seus equivalentes na linha de chegada. O cabeçalho e a etiquetagem das linhas dos textos que compõem o *Corpus* tornaram o uso do *WordSmith Tools* mais eficiente, pois nos foi possível selecionar linhas específicas e gerar sua lista de palavras e suas linhas de concordância.

No próximo capítulo observaremos como a metodologia apresentada aqui pode ser aplicada para analisar um *corpus* compilado a partir das traduções dos alunos.

Capítulo 4 – Análise das traduções

Este capítulo propõe um modelo de análise a partir do *corpus* de aprendizes de tradução. Primeiramente apresentamos o texto de partida seguido de alguns dados estatísticos. A seguir processamos tanto o texto de partida quanto as traduções nos programas *WordSmith Tools* e *AntConc*. Os resultados obtidos a partir desses processamentos servem de subsídios a uma abordagem *corpus-driven*. Tognini-Bonelli (2001: 84) explica que numa abordagem *corpus-driven* os exemplos não são ajustados a categorias pré-definidas pelo pesquisador e que a recorrência de padrões é o que fornece as evidências da análise. Já numa abordagem *corpus-based* o *corpus* é usado para explicar, testar e exemplificar teorias pré-concebidas (Tognini-Bonelli, 2001: 65). Nesta pesquisa, tentamos sempre iniciar as análises por meio da abordagem *corpus-driven*, principalmente por almejarmos ‘enxergar’ os processos na elaboração de uma tradução com os olhos dos alunos. Assim que se torna possível criarmos alguma hipótese em relação a uma escolha lexical ou morfossintática específica, passamos para a abordagem *corpus-based* a fim de examinarmos os detalhes de tal hipótese e suas variantes, se houver alguma.

Ao final propomos atividades para serem feitas com os alunos, a fim de estimular uma reflexão mais detalhada sobre suas traduções.

4.1 Análise do Texto 3

O texto 3 é o primeiro texto que os alunos traduzem no curso. Quando usado para o ensino de inglês é indicado para o segundo nível do básico. Inicialmente esse texto foi escolhido devido às suas expressões idiomáticas. Após analisarmos as traduções enviadas pelos alunos, percebemos que as questões adjetivais foram bem mais relevantes.

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. By the time she reached the bus stop for her homeward trip, her stomach was one big knot.

As usual, the bus was late - and jammed. She had to stand in the aisle. As the lurching vehicle pulled her in all directions, her gloom deepened.

Then she heard a voice from up front boom, "Beautiful day, isn't it?" Because of the crowd, she could not see the man, but she heard him as he continued to comment on the spring scenery, calling attention to each approaching landmark.

"This church."

"That park."

"This cemetery."

"That firehouse."

Soon all the passengers were gazing out the windows. The man's enthusiasm was so contagious she found herself smiling for the first time that day.

They reached her stop. Maneuvering toward the door, she got a look at their 'guide' an older gentleman with a beard, wearing dark glasses and carrying a thin, white cane.

Fonte: http://datateca.unad.edu.co/contenidos/90121/Archivos%20Habilitacion%202013_1/reading_3.html

4.1.1 Dados estatísticos

Língua de Partida		Total de traduções: 132	
<i>Types</i>	117	Variação de <i>types</i>	De 107 a 142
<i>Tokens</i>	179	Variação de <i>tokens</i>	De 151 a 216

4.1.2 Análise

O primeiro passo é comparar as listas de palavras em inglês e português. Utilizando o programa *AntConc*, geramos uma lista de palavras-chave tanto do texto de partida quanto das traduções, partindo da hipótese que deveria haver uma semelhança lexical entre a lista em inglês e a lista em português. Ao compará-las, tentamos encontrar as preferências de tradução dos alunos. Para gerar as listas de palavras-chave usamos como *corpus* de referência, tanto em inglês quanto em português, *corpora* da área de biologia e saúde nos dois idiomas. Como se trata de um texto em inglês geral, aplicamos um *corpus* com textos mais técnicos, pois a intenção era destacar as palavras e as expressões idiomáticas encontradas em uma crônica.

Abaixo está a lista de palavras na LP e nas traduções, após aplicarmos uma *stop list* contendo os seguintes itens: inglês – *The, the, a, A, an, An, of, in; with, to, To, and, And, for*; português: *a, as, A, As, e, o, os, O, Os, de, De, da, das, Da, Das, do, dos, Do, Dos, para, com, em, na, nas, Na, Nas, no, nos, No, Nos, que, Que, uma, Uma, umas, Umas, um, Um, Uns, uns*. As *stop lists* são arquivos no formato *txt* que, quando carregadas, quer seja no *AntConc* ou no *WordSmith Tools*, excluem de qualquer processamento do *corpus* as palavras nelas contidas.

O resultado da comparação da lista de palavras de um texto original e de sua tradução pode revelar dados importantes sobre as escolhas do tradutor. Entretanto, no estudo do Texto 3 há 132 traduções para apenas um original; assim, temos de levar em conta que os resultados podem ser discrepantes, pois para determinada palavra em inglês os alunos podem ter feito escolhas tradutórias diferentes em português. Por exemplo, *overslept* ora foi traduzido por ‘perdeu a hora’, ora por ‘dormiu demais’. Mesmo assim, a comparação das listas do texto de partida e suas traduções pode revelar características importantes para a análise. A figura 4.1 mostra que há uma coincidência entre dois termos no grupo das três palavras mais frequentes nos dois idiomas: *she* e *day*, e ‘ela’ e ‘dia’. A palavra ‘não’ aparece no topo da lista em português devido ao título (Belo dia, não?). Em inglês, como *not* foi contraído na forma *isn’t* (*Beautiful day, isn’t it?*), sequer aparece na lista. O método que utilizamos aqui é de detectar uma palavra no original e buscar suas possíveis traduções em português, conforme explicado na figura 4.1:

Rank	Freq	Word	Rank	Freq	Word
1	7	she	1	898	ela
2	6	her	2	524	dia
3	4	day	3	388	não
4	4	that	4	358	se
5	4	was	5	318	seu
6	3	as	6	312	ônibus
7	2	all	7	260	homem
8	2	at	8	260	ponto
9	2	beautiful	9	255	estava
10	2	bus	10	227	sua
11	2	heard	11	196	aquele
12	2	isn	12	186	ouviu
13	2	it	13	172	lindo
14	2	late	14	170	como
15	2	man	15	164	é
16	2	out	16	162	pela
17	2	reached	17	145	todos
18	2	stop	18	138	ao
19	2	this	19	137	era
20	2	time	20	132	igreja

Figura 4.1 Comparação das listas de palavras em inglês e português de acordo com a frequência

O próximo passo é gerar uma lista de palavras-chave do texto de partida e das traduções. O resultado a partir do texto original nos mostra palavras que se destacam e que, por isso, podem ser possíveis casos de análise. Já a lista gerada a partir das traduções vai nos mostrar a tendência de soluções dos alunos. O *AntConc* gerou uma lista com 41 palavras-chave em inglês e 143 palavras-chave em português. Novamente há uma coincidência entre a palavra com maior chavacidade em inglês e português: tanto *she* quanto 'ela' são as palavras com maior chavacidade, como mostra a figura 4.2. Outra coincidência é *beautiful* e 'lindo' estarem em quarto lugar em chavacidade. Nesse momento essas coincidências servirão de guia para nossas análises. Os Anexos 4 e 5 trazem as duas listas completas.

Rank	Freq	Keyness	Keyword	Rank	Freq	Keyness	Keyword
1	7	57.607	she	1	893	5493.866	ela
2	6	46.618	her	2	312	2913.930	ônibus
3	4	26.769	day	3	524	2892.912	dia
4	2	24.475	beautiful	4	172	1935.499	lindo
5	2	23.957	bus	5	186	1843.137	ouviu
6	1	21.198	firehouse	6	260	1561.903	homem
7	1	21.198	overslept	7	255	1467.416	estava
8	2	20.658	isn	8	196	1320.600	aquele
9	2	18.436	heard	9	130	1309.916	cemitério
10	1	18.426	homeward	10	127	1308.091	mary
11	1	18.426	lurching	11	125	1306.190	atrasado
12	2	17.555	reached	12	260	1297.788	ponto
13	2	16.890	stop	13	113	1294.975	dormiu
14	1	16.700	deepened	14	128	1278.160	óculos
15	1	16.700	maneuvering	15	117	1251.945	escuros
16	1	16.195	gazing	16	130	1238.120	passageiros
17	1	15.792	rotten	17	111	1237.327	bengala
18	1	15.457	frenzy	18	112	1221.689	sorrindo
19	1	15.457	gloom	19	122	1198.772	corredor
20	1	15.170	scenery	20	119	1174.626	chamando
< >	< >	< >	< >	< >	< >	< >	< >

Figura 4.2 Comparação das listas de palavras-chave em inglês e português

Seguem alguns pontos que nos guiarão em nossa análise *corpus-driven*.

a. O pronome *she*.

De modo geral, prega-se que durante a análise de uma lista de palavras o pesquisador deve se ater às palavras lexicais, ou seja, àquelas que pertencem às categorias dos substantivos, adjetivos, verbos e advérbios. As outras categorias são consideradas gramaticais e pouco contribuem para o significado geral do texto. Entretanto, observamos aqui que a classe dos pronomes do caso reto deve ser estudada com cuidado, já que seu uso pode revelar algumas tendências marcantes no ato tradutório dos alunos. No texto 03, na LP, o pronome *she* é usado sete vezes para narrar, em terceira pessoa, um dia na vida de uma mulher, daí a recorrência do pronome. Entretanto, algo chama atenção nas traduções, que têm o pronome ‘ela’ também com alto índice de frequência e de chavicidade: se a língua portuguesa permite a omissão do pronome, por que o índice de frequência de ‘ela’ é tão alto? O *WordSmith Tools* possui uma ferramenta chamada *Plot* (no *AntConc* a ferramenta chama-se ‘Concordance Plot’), que traça um mapa do

texto e identifica a posição do termo pesquisado. A figura 4.3 mostra a distribuição do pronome ‘ela’ nas catorze primeiras traduções do texto 3. A coluna ‘Words’ traz o total de palavras na tradução e a coluna ‘Hits’, quantas vezes ‘ela’ apareceu no texto. A coluna *Plot* divide o texto em 8 partes iguais e cada barra mostra a localização do termo de busca dentro da tradução:

Figura 4.3 Distribuição do pronome ‘ela’ na ferramenta ‘Plot’ no *WordSmith Tools*

Após analisarmos todas as traduções na aba ‘Plot’ pudemos fazer um levantamento de como se deu o uso do pronome ‘ela’ nas 132 traduções. Com o resultado, montamos a tabela 4.1:

Núm. de ocorrências do pronome “ela”	Núm. de traduções com a frequência do pronome indicada na primeira coluna.
10	2
9	5
8	16
7	70
6	23
5	14

Tabela 4.1 Total de ocorrências do pronome ‘ela’ nas traduções

O fato de que mais de 50% das traduções apresentam a ocorrência do pronome ‘ela’ com a mesma frequência do inglês demonstra que os tradutores-aprendizes não consideraram a possibilidade de omissão

pronominal em alguns momentos. O mais preocupante aqui é que em 17% das traduções o pronome do caso reto foi usado com frequência maior do que no texto original em inglês. Provavelmente esses alunos estão fazendo uso exagerado dos pronomes do caso reto, ou, ainda pior, utilizaram-no com função de objeto, como no exemplo extraído do *corpus*: ‘Na medida em que o veículo balançava ela para todas as direções’.

b. Tradução da palavra *beautiful*

A tradução da palavra *beautiful*, na fraseologia *beautiful day*, traz um aspecto interessante. Tanto a lista de palavras quanto a lista de palavras-chave traz a palavra ‘lindo’ com maior frequência e chavacidade. Ao verificar a palavra no dicionário bilíngue impresso do Houaiss (2000) e no Michaelis *on-line*, verificamos que ‘lindo’ não foi uma tradução sugerida:

beautiful a. belo, harmonioso, formoso; bonito
(Houaiss, 2000: 64)

Para descobriremos as opções de tradução da palavra, utilizamos os filtros de seleção disponibilizados no comando ‘Settings’ do *WordSmith Tools* para gerar um lista de palavras apenas com as linhas 1, do título, e 9, que se referem às traduções da frase *Beautiful day, isn’t it* no texto de partida. Com a lista de palavras gerada, utilizamos a ferramenta *Lemma Forms* para agruparmos as palavras que pretendemos analisar (Figura 4.4):

3	LINDO	260	11,74	0	lindo[172] belo[46] bonit
4	É	151	6,82	79	59,85
5	VOZ	129	5,82	129	97,73
6	ELA	127	5,73	127	96,21
7	UMA	125	5,64	124	93,94

Lemma Forms	
LINDO	172
BELO	46
BONITO	42

Figura 4. 4 Resultado do agrupamento das palavras ‘lindo’, ‘belo’ e ‘bonito’ utilizando a ferramenta *Lemma Forms*

⁴¹ Disponível em: <http://michaelis.uol.com.br/moderno/ingles/index.php?lingua=ingles-portugues&palavra=beautiful>. Acesso em 15/01/2014.

O resultado foi ‘lindo’ com 172 ocorrências, ‘belo’ com 46 ocorrências e ‘bonito’ com 42 ocorrências. A vantagem se de usar essa ferramenta é que podemos gerar linhas de concordância apenas com essas três palavras. Uma rápida análise nos leva a detectar que ora o adjetivo vem pré-posto, ora pós-posto. Para obtermos informações estatísticas, utilizamos a aba *collocates*, presente na ferramenta *Concordance*, que nos permite estudar como as palavras co-ocorrem. A figura 4.5 mostra como os três adjetivos se colocam com ‘dia’ em números:

N	Word	With	Relation	Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4
23	DIA	lindo	0,000	89	172	24	148				4	20		148			
24	DIA	belo	0,000	25	46	1	45					1		45			
25	DIA	bonito	0,000	24	42	23	19					23		19			

Figura 4.5 Resultado da aba *Collocates* da palavra ‘dia’ com ‘lindo’, ‘belo’ e ‘bonito’

O levantamento estatístico nos permite tabular as preferências dos tradutores-aprendizes não só na escolha quanto na posição do adjetivo em relação ao substantivo ‘dia’, como exposto na tabela 4.2 abaixo:

Colocação	Freq.	Colocação	Freq.
lindo dia	148	dia lindo	24
belo dia	45	dia belo	1
bonito dia	19	dia bonito	23
		O dia está lindo	4

Tabela 4.2 Tabulação dos fraseologismos com a palavra ‘dia’

Em um primeiro momento, essas informações poderiam apenas dizer qual é a preferência dos alunos. Entretanto, há a possibilidade de o professor de tradução levar para a sala uma reflexão sobre colocações e sobre o uso da língua. Dicionários de sinônimos apontam que ‘lindo’ tem o mesmo sentido de ‘belo’, mas será que os falantes do português usam o fraseologismo ‘lindo dia’ exatamente da mesma forma que ‘belo dia’? Não questionamos aqui o significado, mas, sim, o uso. Ao fazermos o levantamento de ocorrências dos

fraseologismos em estudo aqui no *Corpus* do Português⁴², surpreendemo-nos com o fato de que ‘belo dia’ tem 162 ocorrências contra apenas nove de ‘lindo dia’. A tabela 4.3 mostra a frequência das colocações no *corpus* dos aprendizes ao lado da frequência no *Corpus* do Português:

Fraseologia	Ocorrências no <i>corpus</i> de estudo	Ocorrências no <i>Corpus</i> do Português
lindo dia	148	9
belo dia	45	162
dia bonito	23	10
dia lindo	20	14
bonito dia	19	1
dia (alguma palavra) lindo	4	4
dia belo	1	1

Tabela 4.3 Tabela comparativa das colocações com a palavra ‘dia’ no *corpus* de estudo e no *Corpus* do Português

Uma análise quantitativa sugeriria que, então, optássemos pela primeira. Porém, um estudo qualitativo das linhas de concordância nos revela algo muito interessante. Apesar de ‘belo dia’ ser usado, claro, para dizer que o dia estava bonito, na maior parte das ocorrências o fraseologismo é usado no sentido de ‘em certo dia’, como mostra a figura 4.6:

Como era muito criticado por amigos, mudei para Montparnasse. Um belo dia o concierge do modesto prédio, Gérard, subiu a n
eu imaginava. O incêndio foi debelado. A sexta-feira foi um belo dia . Estado - E o início da semana? Lopes - Tenso. Na
da, porque ela tem começo, meio e fim. Um belo dia acaba. A última pessoa que entrou no governo vai se aposentar daqui a
problema aqui durante quase uns três meses - até um belo dia que surgiu o ovo de Colombo - é simplesmente a gente não tinha
sio Oswaldo Cruz - - de calça comprida - uhn uhn e um belo dia eu (-) " QUE é isso?você queimou todo o seu o seu
nsa que é seu amigo, conversa, fica próximo e num belo dia a pessoa tá te crocodilando, querendo puxar seu tapete. Ah, não
entes. Tanto caso ele criou com os gastos da mulher que chegou o belo dia de indispor-se com um filho. Logo o Astolfo Bento,
De tudo aproveitou, tanto e tão completamente que chegou o belo dia de agradecer a Sofia a decisão bem tomada: - Nada fari
s, auxiliares diretos do Presidente. Este, um temperamental, um belo dia cansou de tantos estudos, cálculos, tabelas, orçament
e eu era o último a saber. Foi ela quem, um belo dia , quando eu já não morava mais em P. e estava de passagem
stante. A chuva da véspera limpou a atmosfera e fazia um belo dia de primavera quando ele chegou ao prédio de Bia. Ela o re
sperei uma oportunidade para tirar a limpo a minha dúvida e um belo dia , ao saber que o doutor se encontrava na casa do pad

Figura 4.6 Excerto das linhas de concordância da colocação ‘belo dia’ no *Corpus* do Português

A análise do significado do fraseologismo oferece ao professor de tradução a oportunidade de trazer para a sala de aula reflexões que vão além do texto ao discutir com os alunos, por exemplo, se ‘belo dia’ é realmente a

⁴² Disponível em: <http://www.corpusdoportugues.org/>.

melhor escolha como equivalente de *beautiful day*. Em relação ao significado não haveria problema algum. Porém, ao levarmos em conta o modo como falantes da língua portuguesa usam a colocação, talvez fosse interessante sugerir que os 45 alunos que optaram por ‘belo dia’ pensassem se querem ou não manter a sua tradução. É possível traduzir *beautiful day* por ‘belo dia’ no sentido de que o dia está bonito? Sim, mas o *Corpus* do Português nos mostra que a probabilidade de os falantes nativos do português usarem o fraseologismo com esse significado é baixa. O próprio *corpus* de aprendizes aqui mostra uma tendência maior por ‘lindo’ que por ‘belo’. Dando continuidade à discussão, o professor pode ainda levar os alunos a refletir sobre a posposição do adjetivo ‘lindo’, pois no *Corpus* do Português o fraseologismo ‘dia lindo’ registra ocorrência um pouco superior a ‘lindo dia’. Já os alunos que optaram por ‘bonito dia’ possivelmente poderiam repensar sua escolha com base no fato que no *Corpus* do Português há apenas uma ocorrência dessa combinação.

c. Tradução dos fraseologismos *lurching vehicle* e *approaching landmarks*

As questões adjetivais em inglês geralmente podem causar alguma dificuldade no processo de tradução. No texto analisado, há uma estrutura muito comum em inglês formada por um gerúndio na função de adjetivo seguido por um substantivo que constantemente gera certa dificuldade na tradução. Observemos os dois casos que ocorrem na língua de partida:

*As the lurching vehicle pulled her in all directions, her gloom deepened
...as he continued to comment on the spring scenery, calling attention to each
approaching landmark.*

Durante a aula em que o texto foi trabalhado, vários alunos comentaram sobre a dificuldade que tiveram em encontrar em português uma equivalência para essa estrutura. Mais uma vez, se saíram bem aqueles que conseguiram ter uma visão mais macrotextual da língua de partida e de chegada. A aba ‘Texto xx alinhado’ nos permite ver as soluções em um contexto mais

completo. A figura 4.7 mostra algumas das soluções encontradas para traduzir *lurching vehicle*:

I
As the lurching vehicle pulled her in all directions, her gloom deepened.
A medida que o ônibus chacoalhava, seu desânimo aumentava.
A guinada brusca do ônibus a arrastou em todas direções, aumentando a irritação dela.
A maneira que o veículo balançante a puxava em todas as direções, sua melancolia se aprofundava.
A medida que o veículo a balançava em todas as direções, sua agonia piorava.
A medida que o veículo chacoalhava jogando-a em todas as direções, sua chateação aumentava.
A medida que o veículo desenfreado a chacoalhava, sua tristeza se aprofundava.
Com o balance do ônibus, ela se mexeu para todos os sentidos e sua emoção agravava.
Com o balanço do ônibus jogando-a em todas as direções, sua tristeza se
Com o balanço do veículo a jogando para todas as direções, sua tristeza se aprofundou.
Como um carro cambaleando por toda direção, sua tristeza foi aumentando.
Conforme o chacoalhar do ônibus a jogava em todas as direções, seu humor
Enquanto o turbulento veículo a empurrava para todos os lados, sua melancolia foi crescendo.
Enquanto o veículo a arrastava para todos os lados, sua melancolia se intensificava.
Enquanto o veículo balançante a puxava em todas as direções, sua tristeza se aprofundava.
Enquanto o veículo em movimento, jogava ela em todas as direções, seu pessimismo aumentou.
Enquanto o veículo que balançava muito e arrastava-a para todas as direções, sua tristeza aumentou.
O balanço do veículo a empurrou em todas as direções, sua melancolia se aprofundou.
O veículo sacudiu tanto que a jogou para todos os lados, ela ficou extremamente deprimida.
Quando o veículo guinou e arrastou ela para todas as direções, sua tristeza piorou.
Sua tristeza aumentava com o veículo balançando e a jogando para todos os lados.

Figura 4.7 Amostra das soluções encontradas para *lurching vehicle*

Ao analisarmos os resultados mostrados na figura acima conseguimos detectar pelo menos três situações. Há alunos que buscaram uma solução mais macrotextual como ‘À medida que o veículo a balançava em todas as direções’ e que conseguiram passar para a língua de chegada o significado do fraseologismo; há alunos que até tentaram resolver o problema do adjetivo, mas acabaram optando por soluções inapropriadas como ‘A guinada brusca do ônibus a arrastou em todas direções’ mas perderam o sentido na

língua de chegada; e há aqueles que buscaram soluções microtextuais apresentando traduções literais como em ‘Como o chacoalhante veículo a puxava em todas as direções’.

A análise seguinte foi referente às escolhas para a tradução do fraseologismo *approaching landmarks*. Como demonstrado na figura 4.8, observamos que os resultados não foram tão ‘desastrosos’ como no caso de *lurching vehicle*:

K
Because of the crowd, she could not see the man, but she heard him as he continued to comment on the spring scenery, calling attention to each approaching landmark.
Como o ônibus estava bem cheio, ela não podia ver o homem, mas o ouviu falar sobre a paisagem primaveril e chamar a atenção para cada ponto de referência que se aproximava.
Como o ônibus estava cheio, ela não podia ver o homem, mas ela ouvia como ele continuava comentando no cenário primaveril, chamando atenção para cada ponto conhecido que se aproximava.
Como o ônibus estava lotado, ela não conseguiu ver o homem, mas ela continuou o ouvindo, enquanto ele comentava sobre a primavera e chamando atenção para cada lugar onde passavam.
Devido à lotação, ela não conseguia ver o homem, mas podia ouvi-lo conforme ele continuava a comentar a paisagem primaveril, chamando atenção para cada detalhe que se aproximava na paisagem.
Ela não conseguia ver o homem com tanta gente na frente, mas ela o ouviu enquanto ele comentava os cenários e pontos que se aproximavam conforme o ônibus andava.
Por causa da aglomeração, ela não pôde ver o homem, mas ela o ouvia e ele continuou falando da paisagem que surgia, chamando a atenção para os importantes prédios e lugares que se aproximavam.
Por causa da multidão, ela não conseguia ver o homem que falava, mas ouvia enquanto ele continuava a comentar sobre o cenário de primavera, chamando a atenção para cada lugar que aparecia.
Por causa da multidão, ela não conseguia ver o homem, mas ela o ouviu quando continuou a comentar sobre a paisagem de primavera, chamando atenção para cada ponto de referência que se aproximava.
Por causa da multidão, ela não conseguia ver o homem, no entanto, ouvia-lhe enquanto comentava sobre o cenário da primavera, chamando atenção para cada ponto de referência que se aproximava.
Por causa da multidão, ela não conseguiu ver o homem, mas ela o escutou enquanto ele continuava a comentar sobre o cenário de primavera, chamava a atenção para cada ponto que se aproximava.
Por causa da multidão, ela não conseguiu ver o homem, mas ela o ouviu enquanto ele continuava a comentar sobre o cenário de primavera, chamando a atenção para cada ponto de referência que se
Por causa da multidão, ela não conseguiu ver o homem, mas ela o ouviu enquanto ele continuava a comentar o cenário primaveril, chamando atenção para cada ponto de referência que se aproximava.
Por causa da multidão, ela não pode ver o homem, mas ela o ouvia enquanto ele continuava a comentar sobre o cenário da primavera, apontando cada ponto de referencia que via.
Por causa da multidão, ela não pôde ver o homem, mas ela o ouviu continuar comentando sobre o cenário da Primavera, chamando atenção para cada ponto que se encontrasse referência.
Por causa da multidão, ela não podia ver o homem, mas ela o ouviu enquanto ele continuou a comentar da paisagem geral, chamando atenção para cada ponto de referencia que ia se aproximando.
Por causa da multidão, ela não conseguia ver o homem, mas ela escutava ele e ele continuava a comentar o cenário da primavera, chamando a atenção em cada ponto de referencia.
Por causa da multidão, ela não conseguia ver o homem, mas ouviu quando ele continuou comentando daquela vista de primavera, chamando atenção para cada vista que se aproximava.
Por causa das pessoas, ela não conseguiu ver o homem, mas ela o ouviu comentando a respeito da paisagem de primavera, chamando a atenção toda a vez que se aproximavam de um ponto turístico.
Por conta da multidão, ela não conseguia ver o homem, mas ela ouviu enquanto ele continuou a comentar sobre o cenário da primavera, chamando atenção para cada ponto de referência que se aproximava.

Figura 4.8 Amostra das soluções encontradas para traduzir *approaching landmark*

Entretanto constatamos que *landmark* causou, sim, certa dificuldade na tradução desse segmento e que *approching* desapareceu do texto de alguns tradutores-aprendizes.

Os resultados das análises mostram que as questões adjetivais causam dificuldades no processo de tradução dos alunos e por isso acreditamos que se faz necessária a preparação de exercícios que permitam o estudo e a discussão de soluções para tais casos. Tendo como base os resultados das análises apresentadas acima sugerimos duas propostas de exercícios para a prática de tradução de tais estruturas adjetivais. Os objetivos são principalmente oferecer aos alunos atividades que desenvolvam a percepção macrotextual tanto do original quanto da tradução. Esperamos que dessa forma estejamos *capacitando*, e não apenas treinando, o aluno de tradução. Essa abordagem é defendida por Bernardini (2002: 174):

Um curso de tradução com objetivo na capacitação [...] seria mais focado no crescimento de cada um dos alunos, oferecendo-lhes oportunidades para o desenvolvimento de competências e capacidades de muito maior escopo e aplicação do que aquelas para as quais poderão, e certamente serão, ser treinados assim que ingressarem no mercado de trabalho.⁴³

4.2 Atividades para o texto 03

1. Uso dos pronomes do caso reto. Releia a sua tradução com atenção especial aos pronomes do caso reto. Primeiro circule todos os pronomes do caso reto. Em seguida, sugira quais poderiam ser omitidos sem prejudicar o significado do texto.

Nota ao professor: Se a atividade for feita em aula, os alunos podem trabalhar em pares. Em vez de ler o próprio texto, o aluno pode ler o texto do colega e fazer as sugestões de omissão do pronome.

⁴³ Tradução nossa para: *An educationally-oriented translation course [...] would be more focused on the growth of each individual learner, providing her with room for the development of competences and capacities of much wider scope and application than those they can, and indeed will, be trained into once they enter the job market.*

2. **Posição do adjetivo.** Com um colega, estude os trechos abaixo retirados do *Corpus* do Português. Qual é a diferença de significado quando o adjetivo vem antes e quando vem depois do substantivo?

...do portal as armas dos Mascarenhas. Não pode haver a <u>mínima dúvida</u> . Pois por esse portal vi eu entrar Clara...	Do pouco que é ensinado, uma <u>parcela mínima</u> é aprendida. Os dados educacionais projetam a pobreza do futuro. Pouco adianta festejar o...
Passou um <u>bom tempo</u> convencida de que o sogro de Amaury tinha morrido. (Inimigas íntimas, Cavalcante, J)	A chuva aumenta o problema do arroz, mas estamos preocupados com quem não estava plantando nem com <u>tempo bom</u> .
De mais a mais, veja, o país está em <u>franco desenvolvimento</u> , estarecendo o mundo com uma taxa de crescimento do PIB nunca antes experimentada.	Uma senhora de <u>sorriso franco</u> , atenciosa com as centenas de pessoas que queriam fotos de recordação. Também tirei uma, que mostro sempre aos amigos.
Um momento, Eminência! Este <u>pobre homem</u> foi envenenado! Observe o vômito de vinho!	Gabriel que se queixasse da sorte, que havia feito de Melo um <u>homem pobre</u> .

3. **Posição do adjetivo.** Traduza as frases abaixo para o português.

- In this article, we focus on the following question: What factors are associated with **household adoption** of **clean energy** in **poor countries**?
- I don't think people realize that even a **small amount** of money can make a **huge difference** for someone just starting out.
- NATL. SECURITY ADVISER: I have just a brief statement. Last year when I was in the archives -- reviewing documents, I made an **honest mistake**. It is one that I deeply regret.
- Well, my **dear man**, one might think of death and mortality all the time, or from time to time, before that accident and after it. I painted a picture after the accident called The Persistence of Electrical Nymphs in Space 1985.

4. **Transformação de orações adjetivas em locuções adjetivas.** Reescreva as orações abaixo transformando as locuções em orações. Em seguida proponha uma tradução para o português. Observe o exemplo:

Ex.:

Everybody looked at the approaching landmarks.
Everybody looked at the landmarks that were approaching
Todos olhavam para os pontos turísticos que se aproximavam.

- a. So, we're in for a very, very long fight in Syria in the coming months and perhaps even years.
- b. But I want people to understand, my goal is not to make rich people poor. It's to make poor people rich.
- c. It boasts a queen-size berth with a nice six-inch-thick mattress and a hanging locker. (Motor boating)
- d. Though rooftop gardens go back at least to the Hanging Gardens of Babylon, the modern green roof movement has made its way here from Europe, where for years government policies have encouraged or required green roofs.
- e. Angie dropped Russell off at his apartment, which was walking distance from her own, and then took Anita home.
- f. The Eastland never turned a profit in her Cleveland years. Winters were long, the sailing season short, and the ship burned all that fuel. Her politician owners didn't help. They gave away as many tickets as they sold -- my mother, her three brothers, and all their classmates were frequent nonpaying passengers -- and concerns about safety persisted. (The Eastland é um barco.)

Como dissemos na Introdução desta pesquisa, acreditamos que traduzir é uma habilidade que deve ser desenvolvida e que tem de ser treinada. Entender que a Linguística de *Corpus* não é apenas sobre números, mas uma metodologia que nos permite descobrir tendências nas línguas viabiliza a possibilidade de o professor de tradução introduzir em suas aulas discussões e reflexões que vão além do texto de partida e de chegada. O tipo de análise que apresentamos aqui é um exercício que propicia uma visão mais macrotextual, possibilitando que o aluno deixe de ser 'ingênuo', como sugerido por Alves e Tagnin (2009), citados também na Introdução, e passe a ver que cada uma de suas escolhas tem uma consequência para o texto.

Considerações finais

Nesta pesquisa buscamos responder de que formas a compilação e a análise de um *corpus* de aprendizes de tradução podem contribuir na elaboração da didática adotada no ensino de prática de tradução. A característica fundamental que emerge é a mudança no foco da aula, que passa a ser preparada com base nos pontos levantados a partir das traduções elaboradas pelos alunos. O tradutor-aprendiz assume um papel determinante nas escolhas pedagógicas do professor, o que aumenta sensivelmente as possibilidades das necessidades dos alunos serem atingidas. Desde o início, a compilação do *corpus* de aprendizes de tradução trouxe a questão tecnológica como um fator que demandou muita atenção, pois tivemos que buscar uma forma de alinhar um texto original e suas várias traduções. Programas comumente usados para alinhar textos não possuíam ferramentas capazes de lidar com as necessidades de nosso *corpus*, que tem, para cada texto na LP, cerca de 100 a 150 traduções. Assim, chegamos à conclusão que, para viabilizar esta pesquisa, seria necessário, primeiro, adicionar um cabeçalho em cada uma das cerca de 800 traduções. O cabeçalho mostra-se importantíssimo, pois amplia as possibilidades de pesquisa e permite que o pesquisador estabeleça critérios de busca e selecione apenas os textos dos alunos com os quais deseja trabalhar. Por exemplo, a pesquisa pode envolver apenas traduções feitas por alunos que estejam na faixa etária de 18 a 25 e que nunca tenham morado fora. Além disso, para que conseguíssemos analisar as traduções produzidas para cada trecho do texto na LP, precisaríamos etiquetar todos os segmentos dos textos na LC. O grande problema enfrentado foi buscar uma solução para inserção dos cabeçalhos e etiquetagem dos segmentos de forma automática.

Um programa pouco usado por linguistas tornou essas duas tarefas exequíveis: a planilha de manipulação de dados *Microsoft Excel*. Por meio de fórmulas conseguimos não só adicionar cabeçalhos individualizados com os dados coletados a partir dos questionários respondidos pelos alunos como também etiquetar cada um dos segmentos das traduções de forma totalmente automática. A planilha *Excel* também facilitou a transformação das traduções em formato *txt*, essencial para investigação com programas de manipulação de *corpora*. O resultado final foi um *corpus* bilíngue formado por oito textos em inglês com cabeçalho e segmentos etiquetados que serviram de originais para cerca de 800 traduções, todas também com cabeçalhos e segmentos etiquetados. Além das análises das

traduções alinhadas no Microsoft Excel, os textos etiquetados possibilitaram inúmeras outras análises em dois programas usados na manipulação e exploração de *corpora*: *AntConc* e *WordSmith Tools*.

Visando a avaliação do produto, usamos os resultados obtidos a partir de todas as análises para sugerir um modelo de aula para os cursos de tradução. Esperamos que este trabalho tenha sido apenas o passo inicial para incentivar discussões de como a Linguística de *Corpus* pode ser uma aliada na formação de futuros tradutores. Esperamos também que outros *corpora* de aprendizes sejam construídos de forma a possibilitar aulas menos centradas no professor e mais nos alunos e suas traduções.

Referências bibliográficas

ALBIR, A. H. **Enseñar a traducir**: metodología en la formación de traductores e intérpretes - teoría y fichas prácticas. Madrid: Edelsa, 1999.

ALVES, F. A formação de tradutores a partir de uma abordagem cognitiva: reflexões de um projeto de ensino. **TradTerm**, São Paulo, v. 4, n. 2, p. 19-40, 1997.

ALVES, F. Unidades de tradução: o que são e como operá-las. In: ALVES, F.; MAGALHÃES, C.; PAGANO, A. (Eds.). **Traduzir com autonomia**: estratégias para o tradutor em formação. 4ª. ed. São Paulo: Contexto, 2011. p. 29-30.

ALVES, F.; TAGNIN, S. E. O. *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. (Eds.) **Corpora no Ensino de Línguas Estrangeiras**. São Paulo: Hub Editorial, 2010. p. 189-24.

ANTHONY, L. AntConc Version 3.2.4. Tokyo: Waseda University, Tokyo, 2011. Disponível em: <<http://www.antlab.sci.waseda.ac.jp/>>.

ARROJO, R. **Oficina de tradução**: a teoria na prática. 4. ed. São Paulo: Ática, 2002.

BAKER, M.; SALDANHA, G. **Routledge Encyclopedia of Translation Studies**. London: Routledge, 2009.

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri: Manole, 2004.

BERBER SARDINHA, T.; ALMEIDA, G. M. D. B. A Linguística de *Corpus* no Brasil. In: TAGNIN, S. E. O.; VALE, O. A. (Eds.) **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p. 17-40.

BERNARDINI, S. **I corpora nella didattica della traduzione**. Seminario di studi internazionale, 14-15 novembre 1997. Bertinoro: Clueb. 2000.

BERNARDINI, S. Educating translators for the challenges of the new millennium: The potential of parallel bi-directional corpora. In: HALLER, J.; MAIA, B.; ULRICH, M. (Eds.) **Training the language services provider for the new millennium**. Porto: Universidade do Porto, 2002. p. 173-186.

BIBER, D. A corpus-driven approach to formulaic language in English Multi-word patterns in speech and writing. **International Journal of corpus Linguistics**, v. 14, n. 3, p. 275-311, 2009.

BOWKER, L.; PEARSON, J. **Working with specialized language**: a practical guide to using corpora. London: Routledge, 2002.

CAPELATO, R. **Mapa do Ensino Superior no Estado de São Paulo**. Relatório Semesp - Sindicato das Entidades Mantenedoras de Estabelecimentos de Ensino Superior no Estado de São Paulo. São Paulo. 2012. Disponível em: http://www.semesp.org.br/portal/pdfs/publicacoes/mapa_do_ensino_superior_sp_2012.pdf.

CASTRO, M. M. D. **Dicionário de Direito, Economia e Contabilidade**: português-inglês / inglês-português. 3ª. ed. Rio de Janeiro: Forense, 2010.

GENTZLER, E. **Contemporary Translation Theories**. Toronto: Multilingual Matters, 2001.

HOUAISS, A. **Dicionário Inglês-português**. Rio de Janeiro: Editora Record, 2000.

HOUAISS, A.; VILLAR, M. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2009.

IZWAINI, S. **Building specialised corpora for translation studies**. Proceedings of the Workshop on Multilingual Corpora: Linguistic Requirements and Technical Perspectives, Corpus Linguistics 2003, 27 March 2003. [S.l.]: University of Lancaster, 2003. p. 17-25.

KILGARRIFF, A.; GREFFENSTETTE, G. Web As Corpus. **A Digital Archive of Research Papers in Computational Linguistics**, 2003. Disponível em: <<http://www.aclweb.org/anthology/J03-3001>>. Acesso em: 1 outubro 2013.

KRIEGER, D. Corpus Linguistics: What It Is and How It Can Be Applied to Teaching. **The Internet TESL Journal**, v. 9, n. 3, March 2003. Disponível em <http://iteslj.org/Articles/Krieger-Corpus.html>.

LAVIOSA, S. Corpus-based Translation Studies: Where Does It Come From? Where Is It Going? **TradTerm**, v. 10, p. 29-57, 2004.

LEVÝ, J. Translation as a Decision Process. In: VENUTI, L. (Ed.) **The Translation Studies Reader**. London: Routledge, 2000. p. 148-59.

LONGMAN dictionary of contemporary English. Harlow: Pearson/Longman, 2009.

MCENERY, T.; HARDIE, A. **Corpus Linguistics: Method, Theory and Practice**. Cambridge: Cambridge University Press, 2012.

MCENERY, T.; WILSON, A. **Corpus Linguistics: An Introduction**. 2nd. ed. Edinburgh: Edinburgh University Press, 2001.

MELLO, M. C. **Dicionário Jurídico: português-inglês - inglês-português**. 9ª ed. rev. e amp. ed. São Paulo: Método, 2009.

NESSELHAUF, N. **Collocations in a Learner Corpus**. Amsterdam: John Benjamins, 2004.

NIDA, E. **Language structure and translation**. California: Stanford University Press, 1975.

NORD, C. Training functional translators. **Cadernos de Tradução**. Florianópolis: Núcleo de tradução, v. 5, p. 27-46, 2000.

NORD, C. **Text Analysis in Translation: theory, methodology, and didactic application of a model for translation-oriented text analysis**. Tradução de Christiane Nord e Penelope Sparrow. 2nd. ed. Amsterdam: Rodopi, 2005.

OLOHAN, M. Corpus Linguistics and Translation Studies: Interaction and Reaction. **Linguistica Antverpiensia**, p. 419-429, 2002. Disponível em: <<http://www.lans-tts.be/img/NS1/P419-429.PDF>>.

PHILIP, G. Arriving at equivalence: making a case for comparable general reference corpora in translation studies. In: BEEBY, A.; INÉS, P. R.; SÁNCHEZ-GIJÓN, P. (Eds.) **Corpus Use and Translating: Corpus use for learning to translate and learning corpus use to translate**. [S.l.]: John Benjamins, 2009. p. 59-73.

REISS, K. Type, Kind and Individuality of Text: Decision Making in Translation. In: VENUTI, L. (Ed.) **The Translation Studies Reader**. Tradução de Susan Kitron. London: Routledge, 2000. p. 160-171.

RUUSKANEN, D. D. K. Creating the 'Other': a pragmatic translation tool. In: DOLLERUP, C.; APPEL, V. (Eds.) **Teaching Translation and Interpreting 3: New Horizons**. Papers from the Third Language International Conference, Elsinore, Denmark, 1995. Amsterdam: John Benjamins, 1996. p. 233-242.

SANTOS, D. Ser, estar, ficar, haver e ter contra ha, bli e være: quem disse que era fácil traduzir sentimentos e sensações? In: EBELING, S., et al. (Eds.) **Corpus-based Studies in Contrastive Linguistics**. Oslo: OSLa, Oslo Studies in Language, 2014. <http://www.linguateca.pt/Diana/download/KK.pdf>.

SCOTT, M. **WordSmith Tools (Version 6.0.0.168)**. Liverpool: Lexical Analysis Software, 2013. Disponível em: <http://www.lexically.net/wordsmith/index.html>.

SINCLAIR, M. **Reading Concordances: An Introduction**. London: Pearson Longman, 2003.

SINCLAIR, J. M. The Search for Units of Meaning. In: CABRÉ, M. T.; LORENTE, M. (Eds.) **Cicle De Conferències 94-95, Lèxic, Corpus i Dictionaries**. Barcelona: IULA, 1996. p. 97-108.

SINCLAIR, J. M. **Trust the Text: Language, Corpus and Discourse**. London: Routledge, 2004.

SINCLAIR, J. M. Corpus and Text - Basic Principles. In: WYNNE, M. (Ed.) **Developing Linguistic Corpora: a Guide to Good Practice**. Oxford: Oxbow Books, 2005. Disponível em. p. 1-16.

SINCLAIR, M. **Reading Concordances: an Introduction**. London: Pearson-Longman, 2003.

TAGNIN, S. E. O.; TEIXEIRA, D. Linguística de *corpus* e tradução técnica - relato da montagem de um *corpus* multivarietal de culinária. **TradTerm**, São Paulo, v. 10, p. 313-58, 2004.

TAGNIN, S. E. O. **O jeito que a gente diz - combinações consagradas em inglês e português**. Barueri: Disal Editora, 2013.

TOGNINI-BONELLI, E. **Corpus linguistics at work**. Amsterdam: John Benjamins, 2001.

UZAR, R. S. Corpus Methodology for Analysing Translation. **Cadernos de Tradução**. Florianópolis: Universidade Federal de Santa Catarina, v. 1, n. 9, p. 237-65, 2002. Disponível em: <http://www.periodicos.ufsc.br/index.php/traducao/issue/view/432>.

VENUTI, L. **The Translator's Invisibility: A History of Translation**. London: Routledge, 1995.

VENUTI, L. **The Scandals of Translation: Towards an Ethics of Difference**. London: Routledge, 1998.

VERMEER, H. J. Skopos and Commission an Translational Action. Tradução de Andrew Chesterman. London: Routledge, 2000. p. 221-231.

WYLER, A. **Línguas, Poetas e Bacharéis: Uma Crônica Da Tradução No Brasil**. Rio de Janeiro: Rocco, 2003.

ZANETTIN, F. **Translation-driven corpora: corpus resources for descriptive and applied translation studies**. Manchester: St. Jerome, 2012.

Glossário

alinhado	<i>corpus</i> em que o original e as traduções de seus segmentos estão lado a lado
AntConc	programa de computador desenvolvido por Lawrence Anthony para analisar <i>corpora</i>
cabeçalho	informações adicionadas no topo de um texto
célula	cada subdivisão de uma tabela da planilha eletrônica localizada no encontro de uma coluna com uma linha
chavicidade	termo usado na listas de palavras-chave que determina o grau de relevância da palavra no <i>corpus</i> de estudo
comparável	<i>corpus</i> em que os textos da língua A são sobre o mesmo assunto da língua B, não sendo um tradução do outro
concordância, linhas de	segmentos de textos retirados do <i>corpus</i> que contém a palavra de busca que se está pesquisando
<i>corpus</i> , sing. - <i>corpora</i> , pl.	coleção de textos compilados conforme critérios pré-estabelecidos
etiquetas	informações extratextuais sempre colocadas entre os símbolos '<' e '>'. Para que programas como o <i>AntConc</i> e <i>WordSmith Tools</i> entendam a função de cada etiqueta, essas devem sempre vir em pares, uma que abre '< >' e uma que fecha '< />'.
Excel	planilha eletrônica desenvolvida pela Microsoft
fórmula	programação simples inserida na célula de uma planilha eletrônica
fraseologismo	combinação de duas ou mais palavras cuja recorrência não é aleatória
linguística de <i>corpus</i>	área da linguística que busca estudar a língua a partir de instâncias autênticas de uso
planilha, planilha eletrônica	programa de computador que usa tabelas para manipular dados numéricos ou textuais
segmento	divisão do texto a ser traduzido. Essa divisão é feita com base no período sintático que, de forma muito simplista, pode ser entendido como um conjunto de uma ou mais orações e que é iniciado por uma palavra cuja primeira letra é grafada em maiúscula e que termina por ponto final, ponto de exclamação, ponto de interrogação ou reticências.
tokens	termo em inglês usado para indicar o número de itens (total de palavras) em um <i>corpus</i>

types	termo em inglês usado para indicar o número de formas distintas em um <i>corpus</i>
WordSmith Tools	programa de computador desenvolvido por Mike Scott para analisar <i>corpora</i>

Apêndices

Apêndice 1 - Formulário de autorização e questionário para formação do perfil dos participantes.

Autorização de pesquisa científica

Eu, _____, RG: _____
declaro que estive regularmente matriculado no curso de Letras – tradutor e intérprete da FMU – Faculdades Metropolitanas Unidas, Campus Liberdade, enquanto aluno do professor Joacyr Oliveira, e venho por meio desta autorizá-lo a utilizar, para fins específicos de pesquisa científica, as traduções que elaborei como parte das atividades didáticas propostas no curso Prática de Tradução. Estou ciente de que meu nome não será divulgado em nenhum momento.

Assinatura

Perfil do aluno

1. Idade quando fez o curso:
☒ a. De 17 a 25 ☐ b. De 26 a 30 ☐ c. De 31 a 40 ☐ d. Mais de 40.
2. Sexo:
☒ a. Masculino ☐ b. Feminino
3. Conhecimento de inglês ao iniciar o primeiro semestre:
☒ a. Básico ☐ b. Intermediário ☐ c. Avançado ☒ d. Pós avançado
(com certificação internacional.)
4. Vivência no país de língua inglesa antes de iniciar o primeiro semestre:
☒ a. Nenhuma ☐ b. Até 1 ano ☐ c. De 1 a 5 anos ☐ d. Mais de 5 anos
5. Língua nativa:
☒ a. Português ☐ b. Espanhol ☐ c. Japonês ☐ d. Chinês ☐ e. Coreano
☒ f. Outra – Qual? _____
6. Experiência como tradutor antes de iniciar o primeiro semestre:
☒ a. Nenhuma ☐ b. Alguns trabalhos amadores ☐ c. Profissional
7. Você fez outro curso superior (pelo menos 2 anos) antes de iniciar a faculdade de tradução?
☒ a. Não ☐ b. Sim – Qual? _____

Apêndice 2 – Modelo da versão digital da atividade de tradução enviada aos alunos por e-mail

Texto 04 – ATENÇÃO: Antes de iniciar, leia as instruções nesta página! O não cumprimento das normas invalidará o crédito.

a) Traduza o texto abaixo. Lembre-se dos blocos significativos. Registre suas dificuldades e as soluções encontradas.

Entregar via e-mail: dia: __/__/13__, até as 23:59.

		Tradução nesta coluna. Nome na célula abaixo	Anotações feitas em sala para o portfólio
	Língua de partida		
<ua_id=en04401><p>	What Is an Urban Heat Island?	O que é uma ilha de Aquecimento Urbano?	
<ua_id=en04402><p_p8(=1>	As urban areas develop, changes occur in their landscape.	Conforme as áreas urbanas se desenvolvem, ocorrem mudanças em sua paisagem.	
<ua_id=en04403>	Buildings, roads, and other infrastructure replace open land and vegetation.	Prédios, estradas, e outras infraestruturas substituem o terreno aberto e a vegetação.	
<ua_id=en04404>	Surfaces that were once permeable and moist become impermeable and dry.	Lugares que antes eram permeáveis e úmidos se tornaram impermeáveis e secos.	
<ua_id=en04405>	These changes cause urban regions to become warmer than their rural surroundings, forming an "island" of higher temperatures in the landscape.	Essas mudanças deixaram as regiões urbanas mais quentes do que os arredores rurais, formando uma "ilha" de temperaturas elevadas na paisagem.	
<ua_id=en04406><p_p8(=2>	Heat islands occur on the surface and in the atmosphere.	As ilhas de calor ocorrem na superfície e na atmosfera.	

Apêndice 3 - Texto 3 como disposto na planilha do Excel

<header>		
<text_title>	Beautiful day	</text_title>
<text_source>	http://www.mdsupport.org/revelations2.html	</text_source>
<tokens>	179	</tokens>
<types>	117	</types>
<genre>	literature	</genre>
<theme>	chronicle	</theme>
<text_id>	3	</text_id>
<level>	1	</level>
<comments>	This is the first text students translate.	</comments>
</header>		
<text_id=1pr_lit_chron_en003.txt>		
<ua_id=en00301><t>	Beautiful Day, Isn't It?	</t></ua>
<ua_id=en00302><p_par=1>	The day started out rotten.	</ua>
<ua_id=en00303>	Mary overslept and was late for work.	</ua>
<ua_id=en00304>	Everything that happened at the office contributed to her nervous frenzy.	</ua>
<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=en00306><p_par=2>	As usual, the bus was late - and jammed.	</ua>
<ua_id=en00307>	She had to stand in the aisle.	</ua>
<ua_id=en00308>	As the lurching vehicle pulled her in all directions, her gloom deepened.	</p></ua>
<ua_id=en00309><p_par=3>	Then she heard a voice from up front boom, 'Beautiful day, isn't it?'	</ua>
<ua_id=en00310>	Because of the crowd, she could not see the man, but she heard him as he continued to comment on the spring scenery, calling attention to each approaching landmark.	</p></ua>
<ua_id=en00311><p_par=4>	'This church.'	</p></ua>
<ua_id=en00312><p_par=5>	'That park.'	</p></ua>
<ua_id=en00313><p_par=6>	'This cemetery.'	</p></ua>
<ua_id=en00314><p_par=7>	'That firehouse.'	</p></ua>
<ua_id=en00315><p_par=8>	Soon all the passengers were gazing out the windows.	</ua>
<ua_id=en00316>	The man's enthusiasm was so contagious she found herself smiling for the first time that day.	</p></ua>
<ua_id=en00317><p_par=9>	They reached her stop.	</ua>
<ua_id=en00318>	Maneuvering toward the door, she got a look at their 'guide' an older gentleman with a beard, wearing dark glasses and carrying a thin, white cane.	</p></ua>
</text>		

Apêndice 4 - Texto 3 na LP etiquetado

```
<header>
  <text_title>Beautiful day</text_title>
  <text_source>http://www.mdsupport.org/revelations2.html</text_source>
  <tokens>179</tokens>
  <types>117</types>
  <genre>literature</genre>
  <theme>chronicle</theme>
  <text_id>3</text_id>
  <level>1</level>
  <comments>This is the first text students translate </comments>
</header>

<text_id=1pr_lit_chron_en003.txt>
  <ua_id=en00301><t>Beautiful Day, Isn't It?</t></ua>
  <ua_id=en00302><p_par=1>The day started out rotten.</ua>
  <ua_id=en00303>Mary overslept and was late for work.</ua>
  <ua_id=en00304>Everything that happened at the office contributed to her nervous
  frenzy.</ua>
  <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=en00306><p_par=2>As usual, the bus was late - and jammed.</ua>
  <ua_id=en00307>She had to stand in the aisle.</ua>
  <ua_id=en00308> As the lurching vehicle pulled her in all directions, her gloom
  deepened.</p></ua>
  <ua_id=en00309><p_par=3>Then she heard a voice from up front boom,
  'Beautiful day, isn't it?'</ua>
  <ua_id=en00310>Because of the crowd, she could not see the man, but she heard
  him as he continued to comment on the spring scenery, calling attention to each
  approaching landmark.</p></ua>
  <ua_id=en00311><p_par=4>'This church.' </p></ua>
  <ua_id=en00312><p_par=5>'That park.'</p></ua>
  <ua_id=en00313><p_par=6>'This cemetery.'</p></ua>
  <ua_id=en00314><p_par=7>'That firehouse.'</p></ua>
  <ua_id=en00315><p_par=8>Soon all the passengers were gazing out the
  windows.</ua>
  <ua_id=en00316>The man's enthusiasm was so contagious she found herself smiling
  for the first time that day.</p></ua>
  <ua_id=en00317><p_par=9>They reached her stop.</ua>
  <ua_id=en00318> Maneuvering toward the door, she got a look at their 'guide' an
  older gentleman with a beard, wearing dark glasses and carrying a thin, white cane.
  </p></ua>
</text>
```

Apêndice 5 - Texto 3 – Lista das 40 primeiras palavras-chave em inglês, em ordem decrescente de chavicidade

Concordance	Concordance Plot	File View	Clusters/N-Gr	Concordance	Concordance Plot	File View	Clusters/N-Gr
Types Before Cut: 127 Types After Cut: 118				Types Before Cut: 127 Types After Cut: 118			
Rank	Freq	Keyness	Keyword	Rank	Freq	Keyness	Keyword
1	7	53.464	she	21	1	14.324	aisle
2	6	43.074	her	22	1	14.324	contagious
3	3	37.311	beautiful	23	1	14.324	jammed
4	5	32.720	day	24	1	14.101	cemetery
5	2	22.766	bus	25	1	13.901	smiling
6	1	20.602	firehouse	26	2	13.720	late
7	1	20.602	overslept	27	1	13.552	cane
8	2	19.469	isn	28	2	13.526	man
9	1	17.830	homeward	29	1	12.768	beard
10	1	17.830	lurching	30	1	12.377	knot
11	2	17.249	heard	31	1	12.290	glasses
12	2	16.370	reached	32	1	12.207	gentleman
13	1	16.104	deepened	33	1	12.127	chronicle
14	1	16.104	maneuvering	34	1	12.127	revelations
15	2	15.705	stop	35	1	11.905	landmark
16	1	15.599	gazing	36	1	11.582	crowd
17	1	15.196	rotten	37	1	11.202	pulled
18	1	14.861	frenzy	38	1	11.202	wearing
19	1	14.861	gloom	39	1	11.154	approaching
20	1	14.574	scenery	40	1	10.723	enthusiasm

Apêndice 6 - Texto 3 – Lista das 80 primeiras palavras-chave em português, em ordem decrescente de chavicidade

Concordance	Concordance Plot	File View	Clusters/N-Gr	Concordance	Concordance Plot	File View	Clusters/N-Gr
Types Before Cut: 976 Types After Cut: 796				Types Before Cut: 974 Types After Cut: 796			
Rank	Freq	Keyness	Keyword	Rank	Freq	Keyness	Keyword
1	893	4435.091	ela	21	112	880.292	barba
2	312	2122.788	ônibus	22	123	874.156	entusiasmo
3	524	1984.535	dia	23	132	845.843	igreja
4	172	1408.006	lindo	24	113	842.904	multidão
5	186	1406.805	ouviu	25	104	831.338	carregando
6	260	1267.674	homem	26	104	814.376	direções
7	260	1213.839	ponto	27	124	804.527	escritório
8	255	1171.035	estava	28	106	798.303	bombeiros
9	196	1165.154	aquele	29	94	787.948	none
10	125	985.009	atrasado	30	123	780.360	guia
11	130	954.519	cemitério	31	96	771.246	nó
12	127	953.749	mary	32	109	760.719	contribuiu
13	128	949.793	óculos	33	128	760.157	porta
14	117	924.728	escuros	34	100	741.775	final
15	122	917.410	corredor	35	114	717.117	branca
16	113	917.143	dormiu	36	87	710.195	contagante
17	130	914.557	passageiros	37	91	697.879	estômago
18	119	901.588	chamando	38	125	697.046	parque
19	111	900.525	bengala	39	318	695.357	seu
20	112	889.289	sorrindo	40	88	682.825	lotado

Concordance	Concordance Plot	File View	Clusters/N-Gr	Concordance	Concordance Plot	File View	Clusters/N-Gr
Types Before Cut: 974 Types After Cut: 796				Types Before Cut: 972 Types After Cut: 795			
Rank	Freq	Keyness	Keyword	Rank	Freq	Keyness	Keyword
41	93	671.286	olhando	61	72	479.218	janelas
42	130	665.243	voz	62	64	478.833	continuava
43	91	658.700	primavera	63	61	476.702	aproximava
44	129	622.061	chegou	64	93	471.320	senhor
45	80	603.831	costume	65	64	440.005	inter
46	106	591.210	usando	66	105	429.671	causa
47	110	566.063	aconteceu	67	124	420.394	tão
48	123	555.164	ficar	68	61	414.043	tristeza
49	68	551.902	atrasou	69	116	406.345	teve
50	93	543.153	chegaram	70	84	404.026	cenário
51	78	540.427	comentar	71	78	396.711	velho
52	113	537.095	estavam	72	94	394.238	direção
53	121	524.943	atenção	73	99	391.289	demais
54	69	522.872	olhada	74	128	385.848	casa
55	126	519.024	começou	75	715	383.309	para
56	112	514.825	logo	76	58	374.767	janela
57	91	514.631	veículo	77	75	373.254	dela
58	129	503.720	ver	78	55	370.140	parada
59	64	500.752	atrasada	79	66	365.064	viu
60	89	492.763	naquele	80	41	344.107	manobrando

Anexo

Gráfico de distribuição dos grupos etários da população brasileira segundo o Censo do IBGE de 2010.