

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE LETRAS E LINGUÍSTICA
CURSO DE BACHARELADO EM TRADUÇÃO

STHEFANY KAMILLA ALVES



**SISTEMAS DE MEMÓRIA DE TRADUÇÃO E *CORPORA*:
INTEGRAÇÃO QUE CRIA NOVAS POSSIBILIDADES AO TRADUTOR**

Tradução
Translation

Uberlândia - MG

2018

STHEFANY KAMILLA ALVES



**SISTEMAS DE MEMÓRIA DE TRADUÇÃO E *CORPORA*:
INTEGRAÇÃO QUE CRIA NOVAS POSSIBILIDADES AO TRADUTOR**

Monografia apresentada ao Curso de Tradução do Instituto de Letras e Linguística da Universidade Federal de Uberlândia, como requisito parcial para obtenção do Grau de Bacharel em Tradução. Orientadora: Profa. Dra. Marileide Dias Esqueda.

STHEFANY KAMILLA ALVES

**SISTEMAS DE MEMÓRIA DE TRADUÇÃO E *CORPORA*:
INTEGRAÇÃO QUE CRIA NOVAS POSSIBILIDADES AO TRADUTOR.**

Monografia apresentada ao Curso de Tradução do Instituto de Letras e Linguística da Universidade Federal de Uberlândia, como requisito parcial para obtenção do Grau de Bacharel em Tradução. Orientadora: Profa. Dra. Marileide Dias Esqueda.

Banca examinadora:

Orientadora: Profa. Dra. Marileide Dias Esqueda (UFU)

Orientadora

Examinadora: Profa. Dra. Cynthia Beatrice da Costa (UFU)

Membro

Examinador: Mestrando Thaíke Augusto Narciso Ribeiro (UFU)

Membro

Uberlândia/MG, 26 de julho de 2018.

AGRADECIMENTOS

Aos meus pais, Arnaldo e Edilma, e ao Arthur, meu irmão, por sempre me apoiarem no caminho dos estudos, me incentivarem no conhecimento de outras culturas, e pelas palavras de apoio durante a graduação e, em especial, durante a escrita desta pesquisa.

Ao Wender, que com muita paciência, me apoiou, incentivou e é a pessoa que mais acredita nos meus sonhos.

À Nina, uma grande amiga, que estava comigo nos momentos em que eu mais precisava de apoio e que me alegrou em momentos nos quais pensei que não iria conseguir.

À minha orientadora, Marileide Dias Esqueda, por ser um exemplo profissional no qual todos se espelham, pela seriedade e dedicação com a qual ela levou cada linha desse texto e por ser uma pessoa a quem podemos sempre recorrer.

Às minhas amigas e companheiras de graduação, Maria e Helda, que sempre me estimularam a seguir com a minha pesquisa, com meus inúmeros projetos em andamento, e pela companhia durante todas as idas e voltas desses três anos e meio.

Aos professores do curso, que cada um, à sua maneira, colaborou para meu crescimento pedagógico e pessoal.

À professora Dra. Silvana Maria de Jesus, que me orientou na iniciação científica e foi crucial para as ideias contidas nesse trabalho.

E à Empresa Júnior Babel Traduções, onde ganhei experiência e fiz amigos que levarei para vida toda.

RESUMO

Há 15 anos, o uso de sistemas de memória de tradução (SMT) era uma possibilidade ainda convidativa para os tradutores e muitas pesquisas foram feitas em torno das vantagens e desvantagens da utilização desses sistemas. Com o surgimento da internet e com os avanços tecnológicos, tais sistemas ganharam visibilidade e tornaram-se indispensáveis para a inserção do tradutor no mercado de trabalho. Atualmente, o foco das pesquisas mudou e o que vem sendo discutido é o produto desses sistemas – a memória de tradução (MT), como são construídas, gerenciadas e, principalmente, como o uso dessas memórias interfere na produtividade e qualidade da tradução. Esse trabalho consiste em uma pesquisa empírico-descritiva, que avalia possíveis métodos de criação de memórias de tradução integrando tecnologias da tradução e o estudo com *corpus*, com o objetivo de criar novas possibilidades de construção de memória utilizando um *corpus* paralelo. Constatou-se que a ferramenta de alinhamento do sistema gratuito *Wordfast Anywhere* é capaz de construir uma MT utilizando *corpora* paralelos e que tal MT pode auxiliar na tradução de textos específicos de uma determinada área, aqui foi criada uma MT da Análise Transacional, subárea da psicologia. O trabalho enfatiza que esses sistemas e memórias necessitam ser utilizados sob um olhar crítico do usuário, alertando o tradutor sobre o uso indiscriminado e confiança excessiva em seus resultados. Um subproduto, que descreve de forma detalhada a construção e carregamento de uma MT no *Wordfast Anywhere* também foi criado, contribuindo para a utilização dos *corpora* paralelos na criação de MT de áreas diversas.

Palavras-chave: Tecnologias da tradução. Sistemas de memória de tradução. *Wordfast Anywhere*. *Corpus* paralelo. Memória de tradução. Análise Transacional.

ABSTRACT

Fifteen years ago, the use of translation memory systems (TMS) was still an inviting possibility for translators, and much research was done on the advantages and disadvantages of using such systems. With the emergence of the internet and technological advances, such systems have gained visibility and became essential for the insertion of the translator in the job market. Currently the focus of research has changed and what is being discussed is the product of these systems - the translation memory (TM), how they are constructed, managed, and particularly how the use of these memories intervenes in the productivity and quality of the translation. This work consists of an empirical-descriptive research that evaluates possible methods of creating translation memories integrating translation technologies and the study with *corpus*, in order to create new possibilities for memory construction using a parallel *corpus*. The results show that the free alignment tool of *Wordfast Anywhere Translation System* is able to construct a TM using parallel *corpora* and that such TM can aid in the translation of specific texts of a given area, here was constructed a Transactional Analysis TM, a specific area of psychology. This study also emphasizes that these systems and memories need to be used under the translator's critical eye, warning him or her about the indiscriminate use and excessive confidence in its results. A secondary result of this research shows in detail the construction and use of a TM in *Wordfast Anywhere*, aiming at contributing to the use of parallel *corpora* in the creation of TM of different areas.

Keywords: Translation technologies. Translation Memory Systems. Wordfast Anywhere. Parallel *corpus*. Translation Memory. Transactional Analysis.

LISTA DE FIGURAS

Figura 1: Wordfast Anywhere - versão (2018).....	22
Figura 2: Worfast Anywhere - versão WFP4 (2018)	22
Figura 3: Interface do AntConc 3.5.7 – versão para Macintosh OS X – já carregado com o corpus da pesquisa.....	24
Figura 4: Tela com o resultado do alinhamento do corpus previamente compilado pelos alunos da disciplina de Linguística de Corpus.....	28
Figura 5: Tela com o tamanho do corpus paralelo de AT, aberto na aba Word List do AntConc.....	29
Figura 6: Página inicial do Wordfast AutoAligner (2018)	30
Figura 7: Tela do Wordfast AutoAligner com o resumo dos dados da MT gerada e a opção de download.....	31
Figura 8: Tela do Excel® com o documento de verificação da segmentação feita pelo Wordfast AutoAligner.	32
Figura 9: Quantidade de resgates da MT externa, construída utilizando corpus paralelo da área de AT, ao traduzir o artigo de Steiner.	34
Figura 10: Tela de criação da conta de acesso no Wordfast Anywhere.....	37
Figura 11: Acesso à ferramenta Wordfast AutoAligner pela barra de ferramentas do Wordfast Anywhere.....	39
Figura 12: Seleção dos arquivos do corpus paralelo a serem carregados no Wordfast AutoAligner.	40
Figura 13: Corpus paralelo carregado no Wordfast AutoAligner.....	40
Figura 14: Identificação da língua original e traduzida dos arquivos do corpus paralelo.....	41
Figura 15: Tempo de processamento da ferramenta para gerar o arquivo.	42
Figura 16: Resumo do processo de alinhamento.....	42
Figura 17: Arquivos gerados após o alinhamento no AutoAligner.....	43
Figura 18: Gerenciamento das MT e glossários no Wordfast Anywhere.....	44
Figura 19: Carregamento da MT no Wordfast Anywhere.....	44
Figura 20: Relatório da MT do corpus paralelo feito pelo Wordfast Anywhere.	45
Figura 21: Ativação da MT no Wordfast Anywhere.....	46

LISTA DE ABREVIATURAS E SIGLAS

SMT- Sistema(s) de Memória de Tradução

MT- Memória de tradução

LC- Linguística de Corpus

TA- Tradução Automática

TO- Texto Original

TT- Texto Traduzido

EN- Inglês

PT- Português

EO- Inglês Original

PT- Português Traduzido

AT- Análise Transacional

SUMÁRIO

INTRODUÇÃO	11
CAPÍTULO 1- EMBASAMENTO ACERCA DAS TEORIAS ENVOLVIDAS NA PESQUISA	13
CAPÍTULO 2 –EXPLICITAÇÃO DOS SOFTWARES E MÉTODOS ENVOLVIDOS NA PESQUISA	20
CAPÍTULO 3- PROCESSO DE CRIAÇÃO DA MT UTILIZANDO O <i>CORPUS</i> PARALELO.....	26
3.1.2 Compilação de um <i>Corpus</i> paralelo específico	28
3.2.1 Verificação da qualidade de alinhamento do <i>AutoAligner</i> e da MT construída.....	31
3.3 Ativação e utilização da MT construída no <i>Wordfast Anywhere</i>	32
3.4 Teste da MT externa.....	33
3.5 Avaliação das vantagens e desvantagens do SMT gratuito <i>Wordfast Anywhere</i> na criação de MT utilizando <i>corpus</i> paralelo.	34
CAPÍTULO 4- DESCRIÇÃO DETALHADA DOS PROCEDIMENTOS PARA CRIAÇÃO DA MT A PARTIR DO <i>CORPUS</i> PARALELO NO <i>WORDFAST ANYWHERE</i>	37
4.1 Abertura de uma conta de acesso no <i>Wordfast Anywhere</i>	37
4.2 Escolha do <i>corpus</i> paralelo, geral ou de uma área específica	38
4.3 Acesso ao <i>Wordfast AutoAligner</i>	38
4.4 Inserção do <i>corpus</i> paralelo dentro da ferramenta <i>AutoAligner</i>	39
4.5 Processo de alinhamento na ferramenta <i>AutoAligner</i>	41
4.6 <i>Download</i> e descompactação do arquivo gerado pelo <i>AutoAligner</i>	43
4.7 Utilização da MT do <i>corpus</i> paralelo no <i>Wordfast Anywhere</i>	43
CONSIDERAÇÕES FINAIS.....	47
REFERÊNCIAS.....	48

INTRODUÇÃO

Definir parâmetros para a compilação de um *corpus* paralelo, encontrar textos originais e traduzidos de qualidade e dentro dos critérios estabelecidos, editar *tags*, converter esses arquivos para formato de texto simples (.txt), criar e nomear pastas, e finalmente conseguir construir um *corpus* para análise linguística de textos não são tarefas fáceis. Esse é o processo de compilação de um *corpus* paralelo. Depois de todos esses passos, por que não aproveitar os *corpora* paralelos em seus arquivos e transformá-los em uma memória de tradução, que pode nos auxiliar ainda mais na tradução de textos de uma determinada área específica?

Transformar *corpora* paralelo em memórias de tradução também pode ser uma saída para os tradutores iniciantes, ou profissionais, que desejam se especializar em uma área específica e ainda não têm uma memória de tradução sobre a área em questão. Igualmente, este trabalho contribuirá tanto para linguistas ou tradutores experientes, que já possuem alguns *corpora* compilados e desejam dar um novo uso para esse material, quanto para os tradutores em início de carreira.

É, portanto, sobre essa temática que este trabalho está versado: a criação de memórias de tradução a partir da compilação de *corpora*.

Dentro da grande área dos Estudos da Tradução, estudos relacionados às tecnologias da tradução e às ferramentas da Linguística de *Corpus* vêm crescendo e ganhando visibilidade junto aos tradutores, pesquisadores, alunos e professores.

Nas tecnologias da tradução, o que vem sendo amplamente estudado atualmente são os sistemas de memórias de tradução (doravante SMT), suas funções, o avanço na criação de memórias de tradução (doravante MT) e o quanto esses sistemas são essenciais no cotidiano do tradutor moderno. Stupiello (2008) e Garcia (2015) defendem o uso e a importância dos SMT, afirmando que esses programas atualmente são vitais ao trabalho de tradução e que negar o uso desses sistemas implica estar à margem do mercado atual da tradução.

Com relação à Linguística de *Corpus*, também como uma área que contribui para a formação e atuação profissional do tradutor com suas inúmeras ferramentas tecnológicas, as pesquisas se voltam à descrição linguística, mostrando-se muito útil nos Estudos da Tradução. Os produtos resultantes dos estudos com *corpora* auxiliam o tradutor de várias formas, seja na consulta da frequência das palavras, no estudo de exemplos por meio das linhas de concordância ou na verificação dos termos de uma área de especialidade. Contribuem, ainda, para o

desenvolvimento de sistemas de tradução automática (doravante TA) e para o ensino da tradução.

A união entre os Estudos da Tradução e a Linguística de *Corpus*, portanto, tem produzido trabalhos e investigações acerca do ferramental tecnológico que facilita o processo tradutório; com isso, traduzir se tornou um processo mais eficiente, menos lento e mais dependente das tecnologias. Entretanto, um diálogo teórico e prático maior entre os SMT e as ferramentas da Linguística de *Corpus* pode potencializar e maximizar os avanços concernentes aos tipos de ferramentas disponíveis aos tradutores.

A possibilidade de criar uma memória de tradução utilizando *corpora* paralelos de uma área específica ajudaria não só tradutores iniciantes, que ainda não possuem uma memória de grande volume, como tradutores e pesquisadores de uma área técnica específica que já tenham feito pesquisa com a Linguística de *Corpus* e que tenham *corpora* paralelos disponíveis. Desta forma, torna-se relevante o objetivo principal desta monografia, qual seja, o de criar uma MT utilizando *corpora* paralelos em um SMT.

Com base no exposto, esta monografia está dividida nas seguintes partes: além desta introdução, o capítulo 1 busca descrever no que consiste um SMT e suas atuais funcionalidades, bem como os propósitos e ferramentas utilizadas pela Linguística de *Corpus*, com vistas ao aumento de produtividade e eficácia do processo tradutório. Neste capítulo, serão utilizadas as propostas teóricas daqueles autores que lidam com o tema. O capítulo 2 detalha a metodologia adotada neste trabalho, explicitando as tecnologias usadas na pesquisa. O capítulo 3 apresenta os procedimentos empregados na criação de uma MT utilizando um *corpus* paralelo, os resultados obtidos com a pesquisa e as discussões sobre as vantagens e desvantagens da utilização desses métodos. E o capítulo 4 descreve detalhadamente cada processo de criação dessas MT, oriundas dos *corpora* paralelos, visando a possibilidade de reprodução do método posteriormente.

Esse trabalho justifica-se na contribuição, de maneira relevante, para os pesquisadores dos Estudos da Tradução, da Linguística de *corpus*, da Análise Transacional, subárea da Psicologia utilizada para a compilação do *corpus*, ou de quaisquer áreas de especialidade que possam fazer uso dessas ferramentas para fins de tradução.

CAPÍTULO 1- EMBASAMENTO ACERCA DAS TEORIAS ENVOLVIDAS NA PESQUISA

1.1 Os SMT

Para que se possa definir melhor o que é um SMT, primeiramente busca-se esclarecer a diferença entre tradução automática (TA), ferramentas de tradução e sistemas de memória de tradução (SMT).

Qun e Xiaojun (2015, p.105) definem tradução automática como uma subárea da linguística computacional, que estuda o uso de programas para traduzir textos ou discursos de uma língua para outra, fornecendo uma tradução completa para o idioma de chegada, de uma só vez, utilizando dicionários, cálculos algoritmos, bancos de dados de memórias de tradução e até *corpora* paralelos. A tradução automática é chamada, pela maioria dos teóricos de *machine translation*, equivalente a tradução automática em língua portuguesa. Para evitar confusões com os termos, nessa pesquisa utiliza-se TA, como já mencionado, para designar a tradução automática.

Já os termos *Computer-assisted Translation Tools* ou *CAT tools* e *translation memory systems* (TMS) – o primeiro usado por Garcia (2015) e Bowker (2005), e o último utilizado por Pym e Gil (2006), Christensen e Schjoldager (2011) e LeBlanc (2013) –, significam sistemas ou programas de memória de tradução, observando-se que essas nomenclaturas variam de autor para autor. Atualmente, no Brasil, autoras da área de tecnologias da tradução, como Esqueda e Stupiello (2017), referem-se a tal tecnologia como sistemas de memória de tradução (SMT), em português e *translation memory systems* (TMS), quando publicam artigos sobre o tema em inglês (ESQUEDA; STUPIELLO, 2016). Portanto, essa pesquisa utilizou a nomenclatura e a definição usada em língua portuguesa por Esqueda e Stupiello (2017), que ressaltam:

Os SMT são ferramentas tecnológicas que têm como propósito facilitar o trabalho do tradutor, por meio de um mecanismo que armazena em uma memória eletrônica os textos traduzidos, separando-os em segmentos que obedecem a critérios, até certo ponto configuráveis, para posterior recuperação, com o propósito de reduzir o tempo gasto com o processo de tradução de textos ou segmentos que possuem algum nível de correspondência. (ESQUEDA e STUPIELLO, 2017, p. 1765)

Entende-se que os SMT são muito diferentes dos sistemas de TA, uma vez que possuem ferramentas e recursos disponíveis para auxiliar o tradutor no processo tradutório. Os SMT oferecem recursos que aumentam a ergonomia e a produtividade tradutória, tais como: visualização do texto original (doravante TO) e do texto traduzido (doravante TT), dispostos na mesma tela; prevenção de saltos e/ou trechos sem tradução; preservação da diagramação do

TO; leitura de vários formatos de texto; preservação de números e nomes próprios; interface própria ou similar ao do processador de texto do Word, bastante familiar aos tradutores; uso ativo de glossários; pesquisa ativa e imediata; uso de diversos glossários ao mesmo tempo; coleta de termos e preparação dos glossários; buscas difusas. Todas estas funções estão à disposição do usuário, na maioria dos SMT, que podem funcionar com e sem memória ativa. Entretanto, a utilização das memórias aumenta a produtividade do tradutor.

Um SMT gera um projeto de tradução a partir da inserção do texto original; segmenta o texto e à medida que o tradutor traduz um segmento e avança para o próximo, alimenta-se uma memória de tradução (MT), que nada mais é do que o segmento no idioma X e o segmento traduzido para o idioma Y salvos paralelamente. Desse modo, quando armazenado, o segmento fica disponível para recuperação no banco de dados (*e.g.*, na memória), que é ativado assim que aparecer outra estrutura linguística semelhante, podendo o tradutor utilizar ou não esses segmentos recuperados oportunamente. Sempre que faz a recuperação de um segmento da MT, o sistema informa ao tradutor o grau de similaridade entre esse segmento recuperado e o segmento que está sendo traduzido, podendo essas correspondências linguísticas serem denominadas como parciais (*fuzzy matches*) ou totais (*full matches*), pois, quando aparece uma recuperação total, o tradutor pode inserir a tradução resgatada na nova tradução com um simples comando. Isso polpa tempo e otimiza o trabalho do tradutor (ESQUEDA E STUPIELLO, 2017).

Alves (2004) explica que a alimentação do sistema pode ser feita de duas maneiras, interna ou externamente. O que aqui será chamado de MT interna é a memória alimentada pelo tradutor, através da interação durante a tradução, com o SMT. Enquanto ele traduz, os segmentos bilíngues vão sendo arquivados na MT interna do projeto criado para traduzir um determinado texto. A segunda forma de construção de MT é utilizando fontes textuais bilíngues de um assunto específico, que podem ser alinhadas, formando uma MT externa.

Um SMT pode ser alimentado de duas maneiras. Na primeira delas o tradutor interage com o sistema e o insumo linguístico vai sendo armazenado à medida que o tradutor avança no trabalho de tradução. Desta forma, assim que um determinado segmento é armazenado, este se torna disponível para fins de recuperação caso seja identificado como correspondente a um insumo linguístico subsequente. A segunda forma de armazenamento de um SMT consiste em alimentá-lo com fontes textuais disponíveis sobre um assunto específico. Antes que tais recursos possam ser utilizados pelo tradutor, é necessário alinhá-los com os segmentos correspondentes de um dado texto de partida. (ALVES, 2004, p. 199)

O alinhamento é uma maneira de recuperar textos que não foram traduzidos nos SMT, pareando o TO e o TT, segmentando e codificando-os apropriadamente para que sejam transformados em uma MT, principalmente quando o tradutor já realizou muitas traduções fora

dos sistemas. A segmentação desses textos paralelos é feita seguindo as regras de segmentação utilizadas pelos sistemas de memória, dessa maneira a reutilização desses segmentos já traduzidos são maximizados, pois, os SMT enxergam os segmentos gerados a partir de textos alinhados exatamente como enxergam aqueles criados no editor, através da inserção da tradução pelo tradutor (Garcia, 2015).

O trabalho de alinhamento normalmente é uma tarefa que exige muitas etapas e uma quantidade considerável de tempo, pois é um trabalho que deve ser feito individual e manualmente. Stupiello (2014) explicita essa afirmação ao falar sobre as possibilidades que os tradutores têm de reutilizarem seus textos previamente traduzidos para a criação de uma MT e como o aproveitamento dessas traduções pode ser uma possibilidade para melhorar o desempenho tradutório.

O trabalho de digitalização e alinhamento de textos de origem e suas respectivas traduções exige tempo, porém, dependendo do grau de semelhança de um novo trabalho com esses textos, o investimento de esforços pode ser útil não só para reaproveitamento no trabalho atual, mas também para alimentar o banco de dados para posterior utilização em pesquisas terminológicas e fraseológicas. Textos de origem e suas traduções, sempre digitalizados, podem ser alinhados pelos recursos próprios de cada sistema e, então, armazenados no banco de dados, podendo vir a melhorar o desempenho da ferramenta ao aumentar as chances de localização e reutilização desses conteúdos. (STUPIELLO, 2014, p.79)

Os *corpora* paralelos podem ser utilizados como fontes para a construção destas MT externas e tornam esse processo mais rápido e produtivo.

A próxima seção apresenta alguns pressupostos da Linguística de *Corpus*.

1.2 A Linguística de *corpus* e os Estudos da Tradução

Berber Sardinha (2004) define a Linguística de *Corpus* (LC) como uma:

abordagem que se ocupa da coleta e da exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por computador. (Berber Sardinha, 2004, p.3).

Para Sinclair (2005), pesquisador de *corpus* e responsável pela criação do primeiro *corpus* computadorizado, a definição de *corpus* é:

uma coleção de textos em formato eletrônico, compilado a partir de critérios externos pré-definidos, para representar, da melhor forma possível, a linguagem ou uma variação

de linguagem, sendo uma fonte de dados para pesquisas linguísticas. (SINCLAR, 2005, p.16) [tradução nossa].¹

A LC permite investigar a linguagem de uma maneira rápida e objetiva, pois proporciona ao analista uma quantidade enorme de textos e de informação, que antes dos avanços tecnológicos, era praticamente impossível.

Berber Sardinha (2004, p.2) afirma, em uma entrevista à revista ReVEL, que muitos programas de computador, tradutores *online* e aplicativos poderiam ser criados a partir da contribuição da LC. Desde 2004, a área de tecnologia voltada para a Linguística teve um grande desenvolvimento e Sardinha provou que tinha razão. A LC contribui e continua contribuindo para a criação e desenvolvimento de vários *softwares* de tradução automática (e.g *Google Tradutor, WorldLingo, Yahoo! Babel Fish*) e, além disso, é essencial para certas análises de traduções, tais como: avaliar estruturas textuais e discursivas, auxiliar em análises terminológicas e na criação de glossários, possibilitar análises de colocação, prosódia, estilo do autor e estratégias e soluções tradutórias, entre outros.

As pesquisas que utilizam *corpora* com textos traduzidos para estudar a tradução usam como referência a tipologia proposta por Baker (1995), que define basicamente três tipos de *corpora*:

- (i) *Corpora* comparáveis – formados por duas coleção de textos distintos, uma de textos originalmente escritos em uma língua X e outra, com outros textos traduzidos naquela mesma língua X. Os *corpora* comparáveis são monolíngues, pois são formados basicamente de textos originais e textos traduzidos que não possuem nenhuma relação além do idioma que estão escritos.
- (ii) *Corpora* paralelos – formados por textos originais escritos na língua X e suas respectivas traduções para língua Y.
- (iii) *corpora* multilíngues – formados por dois *subcorpora* ou mais, contendo textos que correspondem a critérios pré-definidos, como por exemplo, gênero do discurso ou tipo de texto. Cada *subcorpora* contém textos em uma língua diferente, podendo ser formado por duas ou mais línguas.

¹A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research. (SINCLAIR, 2005, p.16)

Jesus (2014) aponta que a tipologia utilizada por Baker (1995) serviu como base para outros autores incorporarem diversos outros tipos de *corpora* mais complexos e que podem surgir nos Estudos da Tradução, entretanto, aqui foi adotada a terminologia e a visão simplificada apresentada por Baker para os *corpora* paralelos, que é a matéria prima dessa pesquisa.

Normalmente, os *corpora* paralelos são bilíngues, entretanto, também podem ser compostos de traduções do texto original em mais de dois idiomas, esses são chamados de *corpora* paralelos multilíngues. Os Estudos da Tradução têm explorado os *corpora* paralelos e comparáveis em diversas frentes, como, por exemplo, em pesquisas terminológicas, na formação do tradutor, na investigação das propriedades de textos traduzidos e como base de dados de tradução.

Recentemente, tem-se buscado investigar a relação entre *corpora* e tecnologias da tradução, que pode ser feita não somente para suprir as necessidades da TA, mas servir como base de dados de recuperação instantânea, dentro de um SMT. Lars Ahrenberg (2015) destaca essa interface ao falar sobre alinhamento.

Historicamente, a noção de alinhamento na tecnologia da tradução está intimamente ligada ao interesse em *corpora* paralelos, ou bitextos, que surgiu na última metade da década de 80, para suprir as deficiências das tecnologias da tradução automática e os programas de auxílio à tradução. (AHRENBORG, 2015, p.395) [tradução nossa]²

Assim, a presente monografia pretende explorar as perspectivas desses pesquisadores, e utilizar a Linguística de *Corpus* como uma nova maneira de auxiliar o tradutor, criando uma memória de tradução confiável, de tamanho considerável e voltada para a área de especialidade a ser traduzida, fazendo uso de um STM gratuito, o *Wordfast Anywhere*, que não gera custos aos pesquisadores ou tradutores. Para criar MT externas é preciso reutilizar ou compilar *corpora* paralelos – que visto sob a perspectiva da área de tecnologias da tradução, nada mais é do que um texto original e texto traduzido que não foi feito dentro de um SMT, mas que pode ser alinhado – e transformar esses *corpora* em MT, utilizando a ferramenta de alinhamento do SMT.

²Historically, the notion of alignment in translation technology is intimately bound up with the interest in parallel corpora, or bitexts, that emerged in the last half of the 1980s, as a way to deal with the shortcomings of the then existing technologies for machine translation and translation aids. (AHRENBORG, 2015, p.395)

1.3 Olhar crítico sobre a reutilização dos segmentos resgatados da MT

Darin (2001) afirmou que o uso de *softwares* de tradução, por tradutores profissionais, era uma realidade iminente e que cresceria cada vez mais nos anos e décadas que estariam por vir. Stupiello (2008) confirmou a informação de Darin dizendo que traduzir já envolvia estar munido de tecnologias que auxiliassem o tradutor, como dicionários eletrônicos, base de dados e SMT. LeBlanc (2013) ressaltou que os SMT eram as ferramentas de auxílio ao tradutor mais conhecidas e utilizadas pelos tradutores e apontou, em sua pesquisa, dados empíricos de como os SMT podem melhorar o processo tradutório aumentando a produtividade, aprimorando a consistência textual das traduções, diminuindo o tempo gasto na elaboração de traduções, e também mencionou que as memórias podem servir como *corpus* paralelo – o caminho inverso do que essa pesquisa propõe. Com isso, os SMT já provaram que conseguem otimizar o trabalho do tradutor e conseguiram grande popularidade, tanto na área acadêmica quanto na área profissional da tradução. Entretanto, surgiram novos questionamentos relacionados ao uso desses sistemas e das recuperações das MT.

Ideias levantadas por Bowker (2005), vem sendo discutidas por Christensen e Schjoldager (2011), Esqueda e Stupiello (2016; 2017), Esqueda *et alli.* (2017) e também LeBlanc (2013), e dizem respeito à maneira como o tradutor percebe a segmentação e a utilização das memórias, ao uso indiscriminado desses sistemas e das traduções automáticas, à excessiva confiança neles depositados, e à propagação de erros contidos nas memórias.

LeBlanc (2013), apesar de citar as inúmeras vantagens desses sistemas, destaca que a confiança exacerbada nas memórias, principalmente por tradutores iniciantes – que não tem experiência e discernimento para entender a complexidade dos sistemas, olhar crítico e conhecimento linguístico para rejeitar certas propostas tradutórias oferecidas pela MT – leva à propagação de erros com a reutilização de segmentos não confiáveis em conjunto com traduções automáticas mal solucionadas. À medida que os tradutores passam a confiar cada dia mais nas memórias, a propagação de erros se torna um problema a ser considerado.

Observando que existe uma tendência de tradutores inexperientes aceitarem, sem um olhar crítico, as recuperações dos segmentos das MT, professores-pesquisadores dos Estudos da Tradução, em seus estudos sobre as novas tecnologias, como Esqueda *et alli.* (2017) mostraram, através de pesquisas empíricas com alunos de tradução, como os textos com alta grau de repetição, a pressão de tempo, a falta de experiência e extrema confiança nas recuperações da MT culminaram em erros tradutórios graves. Assim, perceberam que

precisariam desenvolver em seus alunos a capacidade de metarreflexão em relação as MT e seu uso.

Como as pesquisas mostram que as MT vêm sendo utilizadas pelos tradutores, muitas vezes, sem razoabilidade, principalmente àqueles aos quais essa pesquisa visa contribuir – os tradutores iniciantes –, surge a necessidade de enfatizar que nenhuma tecnologia é capaz de traduzir um texto com qualidade, sem o olhar atento e sagaz do tradutor. As MT, independentemente de quão confiáveis e bem administradas sejam, não devem ser utilizadas sem que esse profissional tenha consciência e olhar crítico de que essa tecnologia apenas *auxilia* seu trabalho, mas que *cabe ao tradutor* administrar e ponderar seu uso, para que a tecnologia continue contribuindo positivamente para o ofício e não prejudicando-o.

CAPÍTULO 2 –EXPLICITAÇÃO DOS SOFTWARES E MÉTODOS ENVOLVIDOS NA PESQUISA

Para que houvesse uma compreensão dos conceitos envolvidos na pesquisa, tais como: sistema de memória de tradução, memória de tradução, alinhamento, e *corpora* paralelos, foi feito um levantamento bibliográfico de textos referentes a esses assuntos, conforme buscou-se mostrar no Capítulo 1.

Além de refletir sobre tais conceitos, também foram realizadas pesquisas sobre o uso e funcionamento dos SMT e como eles podem auxiliar o processo tradutório através de suas múltiplas funcionalidades. A escolha dos textos citados nessa pesquisa baseou-se nas indicações de leitura dos professores ao longo do curso, juntamente com pesquisas feitas na *Internet*³ e as intervenções da orientadora.

Também foi feita uma pesquisa sobre o funcionamento do SMT *Wordfast Anywhere*, destacando a análise do manual ofertado pelo *site* da empresa, observando principalmente a criação e uso das MT. A partir desses manuais, foram testados todos os processos de carregamento, criação e utilização das MT dentro do *Wordfast Anywhere* e do *AutoAligner*. Para tais processos, foram utilizados *corpus* paralelos previamente criados pela VI turma do Curso de Bacharelado em Tradução (ILEEL-UFU), de 2015, e artigos originais e traduzidos do *site Institute for Integrative Psychotherapy*⁴.

Por se tratar do uso da versão *online* do *Wordfast Anywhere*, não são necessárias as especificações técnicas do computador ou sistema utilizado na pesquisa, entretanto, o que pesou na escolha desse *software* foi a facilidade de execução em qualquer sistema operacional, como por exemplo, *Windows*, *Linux* ou *MacOS*. O computador deve somente estar conectado à rede, e preferencialmente aberto no navegador *Google Chrome*. Imagens, por captura de tela, do funcionamento do SMT foram coletadas durante todo o processo de teste da criação da MT utilizando *corpus* e a partir desses testes de processo deu-se início à redação de descrição de cada processo para a criação da MT.

Dentro dos Estudos da Tradução, Willians e Chesterman (2002) definem 12 áreas de pesquisa. Pautada nesses autores, esse trabalho consiste em uma pesquisa empírico-descritiva, que avalia possíveis métodos de criação de memórias de tradução utilizando as tecnologias da tradução e o estudo com *corpus* para realizar esse processo.

³ Rede mundial de computadores. Segundo o dicionário Caldas Aulete, a palavra registra-se, muitas vezes, com inicial maiúscula e os sinônimos net e rede podem ser utilizados.

⁴Disponível em <<http://www.integrativetherapy.com>>. Acesso em nov. 2017.

2.1 O SMT livre: *Wordfast Anywhere*

A escolha de explorar o uso de *corpora* paralelos em um sistema de memória de tradução gratuito e *online* justifica-se pelo fato de que muitos tradutores, principalmente os que estão no início de carreira, não podem fazer um grande investimento para adquirir um *software* proprietário. A popularização dos sistemas pagos é nítida, além da eficácia e alta tecnologia contida nesses SMT. No entanto, as funcionalidades dos sistemas gratuitos, que se assemelham a dos sistemas pagos, podem igualmente oferecer os mesmos recursos de qualidade.

Neste contexto, este trabalho visa descrever algumas possibilidades de uso do SMT gratuito, o *Wordfast Anywhere* e apresentar suas vantagens e desvantagens.

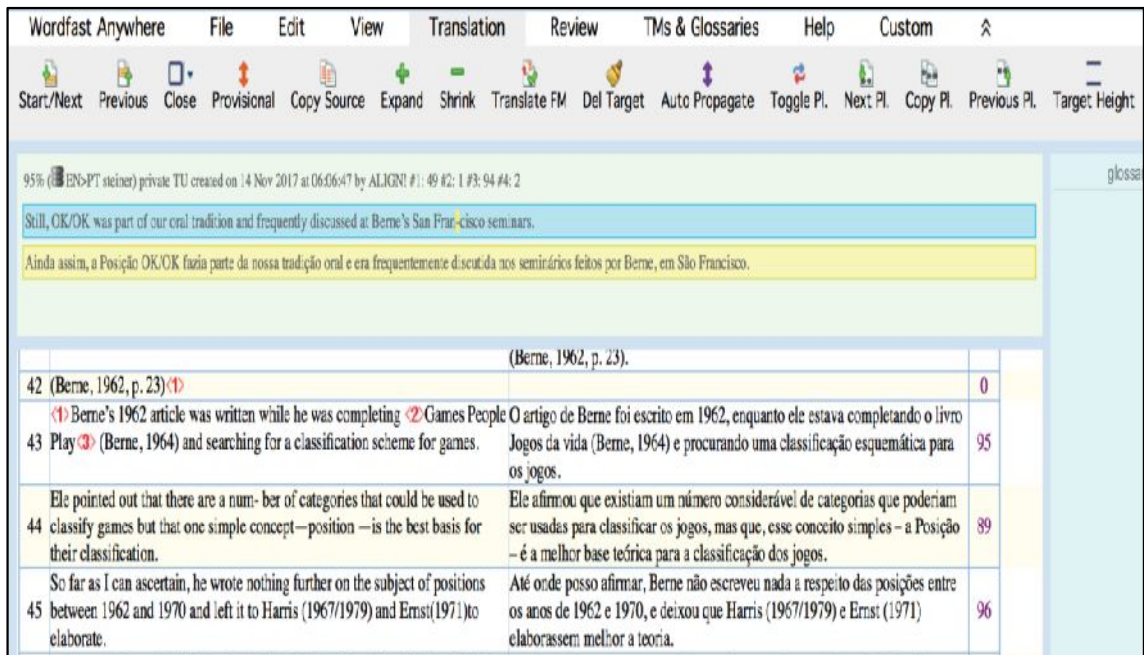
O *Wordfast Anywhere* é um produto gratuito, disponível *online* desenvolvido pela companhia *Wordfast LLC*. Em 1999, o francês Yves Champollion criou a primeira versão de um SMT chamado *Wordfast*, desenvolvido para ser um concorrente do *SDL Trados*. Com a popularização da ferramenta, a companhia lançou produtos como o *Wordfast Classic*, em 2000, um SMT executado através de conjunto de macros do *Microsoft Word*, o *Wordfast Professional*, um SMT independente e de interface própria e que também compete com outros SMT do mercado. Essas últimos três produtos desenvolvidos pela companhia não são gratuitos, mas possuem versões disponíveis para *download* no site da empresa, que podem ser testadas pelo tradutor durante um prazo de 30 dias.

O *Wordfast Anywhere* é a versão *online* e gratuita, como já mencionado, com características muitas próximas do sistema de interface própria da empresa, o *Wordfast Professional*. Foi criado em maio de 2010 e, segundo o site⁵ da companhia e seus desenvolvedores, a versão *online* foi desenvolvida para libertar os tradutores de suas casas e escritórios, permitindo que eles trabalhem de qualquer lugar que tenha conexão com a *Internet*. Os usuários podem carregar suas MT e glossários para uma conta *online*, protegida por senha, e acessá-las virtualmente, de qualquer computador, celular ou *tablet*, com acesso à rede.

As Figuras 1 e 2 mostram os dois layouts exibidos ao tradutor durante o processo tradutório, no SMT *Wordfast Anywhere*. A Figura 1 mostra o *layout* do *Wordfast Anywhere* na seção de tradução, na tela inicial do sistema *online*.

⁵Ver a aba de descrição do *Wordfast Anywhere* no site <<https://www.wordfast.net/?go=anywhere>>. Acesso em fev. 2018.

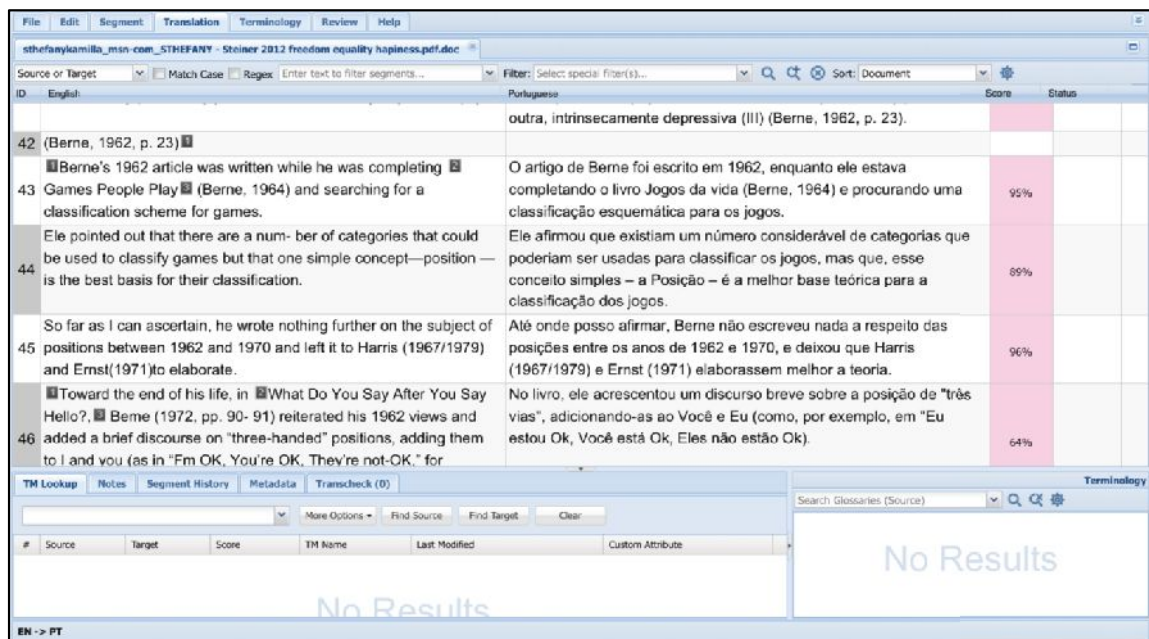
Figura 1: Interface do *Wordfast Anywhere* com um arquivo de tradução - versão (2018)



Fonte: *Wordfast Anywhere* (2018)

Em 2017, o *Wordfast Anywhere* habilitou a função WFP4 – *Wordfast Professional 4* – uma versão demo, *online* do *software* pago. Por essa aba, o usuário consegue ter a experiência de usar a opção paga dentro da versão livre do SMT.

Figura 2: Interface do *Wordfast Anywhere* com um arquivo de tradução - versão WFP4 (2018)



Fonte: *Wordfast Anywhere* (2018)

Sobre os SMT gratuitos, Machado (2014) afirma que “[...]o *software* livre e de código aberto está se tornando, cada vez mais, uma boa opção. [...] os tradutores podem optar por ser independentes e controlar totalmente as ferramentas da sua profissão”. Com essa pesquisa, deu-se visibilidade para um *software online* e gratuito que pode ser utilizado de qualquer computador e lugar com acesso à rede. Assim, esta é uma opção para tradutores que não podem fazer um investimento inicial, além de ser uma maneira alternativa e útil de criar uma memória de tradução.

2.2 *AntConc*

Outra ferramenta utilizada nesta pesquisa para verificar o tamanho do *corpus* e que pode ser usado por qualquer tradutor é o *AntConc*, um concordanciador executável desenvolvido por Lawrence Anthony, Doutor em Linguística Aplicada pela Universidade de Birmingham e, atualmente, docente da *Faculty of Science and Engineering*, em *Waseda*, no Japão. Esse programa reúne ferramentas que permitem buscas e cálculos estatísticos sobre as palavras que compõem um *corpus* escrito. Assim como o *Wordfast Anywhere*, ele é gratuito, pode ser acessado através dos sistemas operacionais *Windows*, *Linux* e *Macintosh OS X*, e está disponível para *download* no site oficial do desenvolvedor⁶.

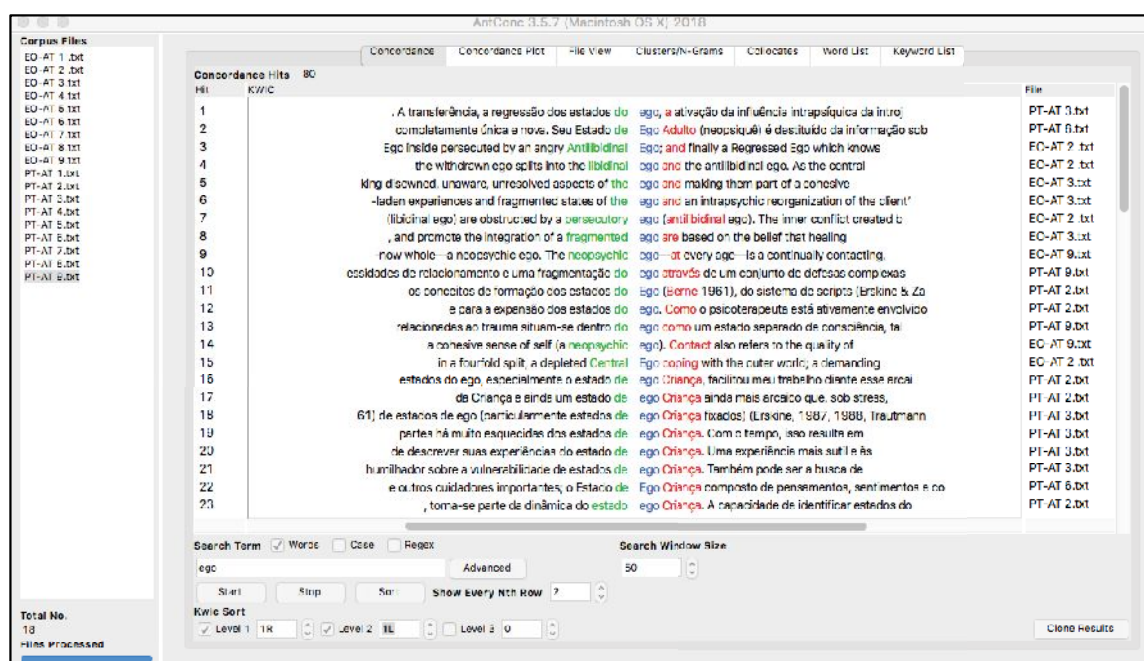
Além do *AntConc* ser gratuito, ele possui uma interface simples, de fácil manuseio e usabilidade de suas funções. Não é necessário fazer a instalação, pois se trata de um arquivo executável, que pode ser acessado com apenas dois cliques de *mouse*. Sua interface também é simplificada, e, em uma única janela, é possível navegar por todas as opções de análise linguística que ele oferece (figura 3), tais como:

- (i) *Concordance* (concordanciador) – fornece as linhas de concordância de um determinado termo pesquisado, aparecendo excertos retirados dos arquivos que compõem o *corpus*.
- (ii) *Concordance Plot* (diagrama das concordâncias) – mostra, através de um gráfico em barras, a incidência e a distribuição do termo pesquisado nos arquivos que compõem o *corpus*.
- (iii) *File View* (visualização no arquivo) – possibilita a visão geral do arquivo/texto que contém o termo pesquisado.

⁶ Versões disponíveis para *download* em: <http://www.laurenceanthony.net/software/antconc/>

- (iv) *Clusters* (colocações) – faz uma busca do termo pesquisado combinado com outras palavras e mostra a frequência com que essa combinação ocorre dentro dos arquivos que compõe o *corpus*.
- (v) *Collocates* (colocados) – gera uma lista dos colocados, ou seja, das palavras que frequentemente antecedem ou sucedem o termo que está sendo pesquisado. A quantidade de palavras antes e/ou depois do termo podem ser ajustadas no painel de controle.
- (vi) *Word List* (lista de palavras) – fornece uma lista com todas as palavras que compõe os arquivos do *corpus* e a frequência que elas aparecem.
- (vii) *Keyword List* (lista de palavras-chave) – gera uma lista de palavras-chave que comparam a frequência das palavras de um *corpus* de referência com o *corpus* que está sendo estudado.

Figura 3: Interface do *AntConc* 3.5.7 – versão para *Macintosh OS X* – já carregado com o *corpus* da pesquisa.



Fonte: *AntConc* (2018)

Na Linguística de *Corpus*, os *corpora* são construídos para investigar questões centrais da linguagem, e o *AntConc* é utilizado como ferramenta para analisar esses arquivos e fornecer dados estatísticos das pesquisas. Já no caso dos Estudos da Tradução, a construção de um *corpus* e o uso do *AntConc* pode ser bastante útil para a verificação de termos específicos de uma área e elaboração de glossários – baseado na frequência que essas palavras aparecem na *Keyword List* quando comparadas com a frequência que aparecem em um *corpus* de referência – ou, como base de dados e *corpus* de consulta, para tradução de colocações (combinação de

palavras que desenvolveram uma relação semântica baseada na coocorrência frequente em uma dada língua) e colocados (palavras que acompanham a palavra base – de maior sentido, dessas colocações).

Portanto, assim como o *Wordfast Anywhere*, o *AntConc* também pode servir como ferramenta de auxílio ao tradutor. E os *corpora* paralelos, compilados por quaisquer objetivos, podem ajudar no processo tradutório, tanto como base de dados dentro do *AntConc* quanto como MT, quando construídos através de alinhamento.

CAPÍTULO 3- PROCESSO DE CRIAÇÃO DA MT UTILIZANDO O *CORPUS* PARALELO

Como dito anteriormente, os *corpora* são criados por pesquisadores, linguistas ou tradutores, com a intenção de analisar determinadas ocorrências da língua, tais como: frequência das palavras mais comuns da língua; a frequência das classes gramaticais; comprovação de colocações; a regência dos verbos preposicionados, servir como base de dados para traduções, auxiliar na pesquisa terminológica, entre outras coisas. Essa versatilidade do *corpus* faz com que a cada dia novos *corpora* sejam compilados para alimentar as pesquisas para os fins descritos acima. Contudo, percebe-se que depois dessas pesquisas, os *corpora* se transformam em base de dados e nem todo seu potencial é explorado.

Assim, nesta pesquisa, busca-se estudar a relação entre *corpora* paralelos e as MT – o primeiro um conjunto de textos no original e tradução, selecionados com um propósito específico de análise da língua e o segundo “um sistema de arquivos reunindo trabalhos anteriores de tradução que, organizados com seus textos de origem, possibilitam a consulta terminológica e fraseológica e, em muitos casos, a recuperação de opções anteriores de tradução salvas e mantidas em bancos de dados.

3.1 Escolha do *Corpus* Paralelo

A ideia inicial deste trabalho baseava-se somente na utilização de *corpora* paralelos já compilados, seja por linguistas, por mim ou outros alunos do Curso de Bacharelado em Tradução, com o intuito principal de aproveitar os *corpora* previamente construídos para a criação das MT, reciclando-os e criando uma oportunidade rápida e fácil para tradutores utilizarem esse material como MT, seja ela ativa ou para consulta.

Entretanto, para satisfazer uma necessidade gerada no projeto de Iniciação Científica do qual a proponente desta monografia era também aluna-pesquisadora, que teve como título “A terminologia de textos técnico-científicos a partir da interface entre *corpora* e novas tecnologias da tradução”, delineou-se como objetivo criar uma MT, de tamanho considerável, sobre Análise Transacional (AT) para auxiliar na tradução de alguns artigos de teóricos da área da Psicologia, e devido a escassez de *corpus* previamente compilados nessa área, atestada por professores do Curso de Bacharelado em Tradução do ILEEL-UFU. Após a criação e estudo da terminologia nessa área, mostrou-se oportuno compilar um *corpus* paralelo da área de

Psicologia para suprir ambas as necessidades, gerar uma MT para a iniciação científica e utilizá-la como base para a presente monografia.

Portanto, o processo de escolha do *corpus* a ser utilizado na criação da MT será descrito a seguir.

3.1.1 Seleção de um Corpus já compilado anteriormente

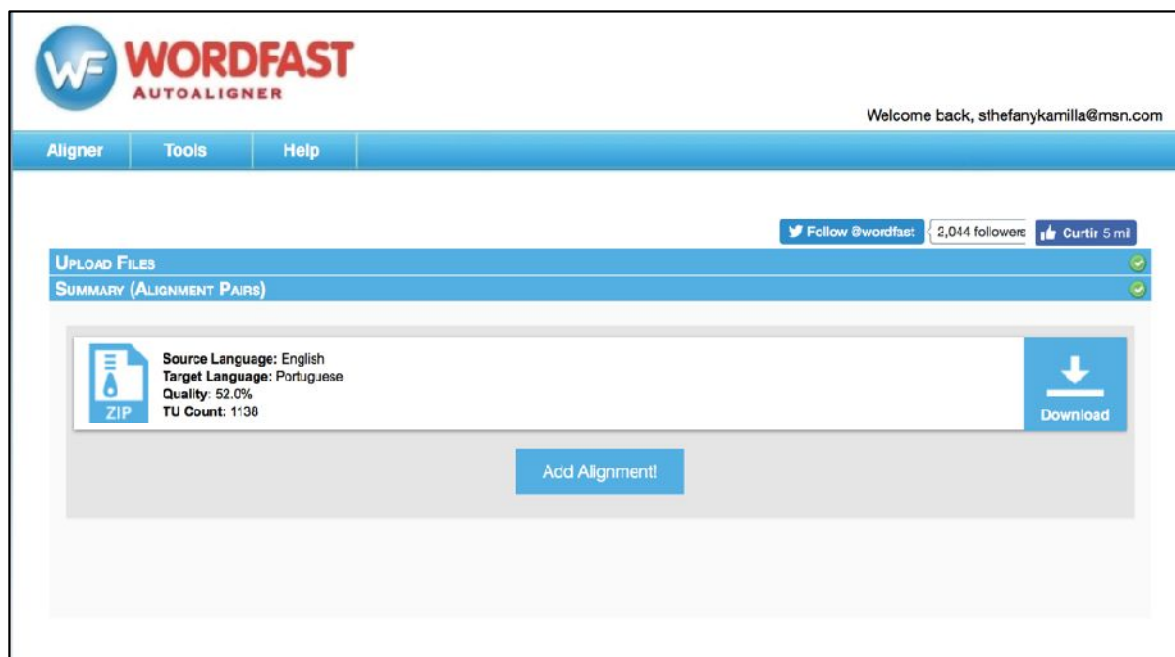
Na fase de testes de criação das MT utilizando *corpus* paralelo e o *Wordfast Anywhere*, foi utilizado um *corpus* paralelo produzido pelos alunos da VI turma de Bacharelado em Tradução da Universidade Federal de Uberlândia, do ano de 2015, que no quinto período cursavam a disciplina de Língua Inglesa: Estudos Descritivos e Linguística de *Corpus*. A turma compilou um *corpus* paralelo do livro da área de Psicologia sobre Análise Transacional, intitulado *Games People Play* (1964) e sua tradução para o português *Os Jogos da Vida* (1974), de Eric Berne, como atividade de compilação e análise de *corpus* da disciplina. O *corpus* paralelo criado pelos alunos foi composto de 36 arquivos em .txt, referentes a cada capítulo do livro e suas respectivas traduções.

Após a compilação, percebeu-se que a tradução *O Jogo da Vida* (op. cit) não estava completamente alinhada com o texto original, *Games People Play* (op. cit), sendo que alguns parágrafos e capítulos foram suprimidos, algumas vezes anexados a outros capítulos do livro traduzido, e que a maior parte da obra foi adaptada. Apesar de os alunos terem compilado o *corpus* acompanhando as modificações, ainda assim o TO e o TT não resultaram em um alinhamento perfeito, muitos segmentos ficaram em branco ou com sentenças que não eram correspondentes ao segmento original.

Tendo essas informações, o *corpus* foi transformado em MT a partir da ferramenta de alinhamento *AutoAligner*⁷ do *Wordfast Anywhere*. Essa foi uma maneira de testar a qualidade de alinhamento e de verificar, através desse primeiro teste, que o *AutoAligner* fornece um resumo do alinhamento automático feito e que nesse resumo é possível ver a qualidade da MT criada. No caso do *corpus* compilado previamente pelos alunos pode-se verificar, na figura 4, que a qualidade de alinhamento da MT foi de 52,0%, número considerado baixo, pois 48% dos segmentos não estavam alinhados corretamente.

⁷ O processo de criação da MT será descrito no item 3.2.

Figura 4: Tela com o resultado do alinhamento do *corpus* previamente compilado pelos alunos da disciplina de Linguística de *Corpus*, em 2015.



Fonte: a autora

Nesse primeiro teste, percebeu-se que nem todo *corpus* paralelo pode ser transformado em uma MT de boa qualidade – entende-se como boa qualidade segmentos de TT correspondentes ao TO, que não estejam em branco ou desconfigurados. E o *corpus* previamente produzido não pôde ser utilizado no decorrer da pesquisa da presente monografia, devido ao seu baixo índice de paridade entre TO e TT.

3.1.2 Compilação de um *Corpus* paralelo específico

Devido à baixa qualidade do alinhamento no *corpus* paralelo previamente compilado na disciplina de Língua Inglesa: Estudos Descritivos e Linguística de *Corpus*, ponto de partida dessa pesquisa, e que a Psicologia, mais precisamente a área de AT, necessitava, tanto de um *corpus* paralelo composto de artigos originais e traduzidos da área, quanto de uma MT para contribuir com os tradutores e auxiliar nas traduções de textos sobre o assunto, compilou-se um *corpus* paralelo da AT.

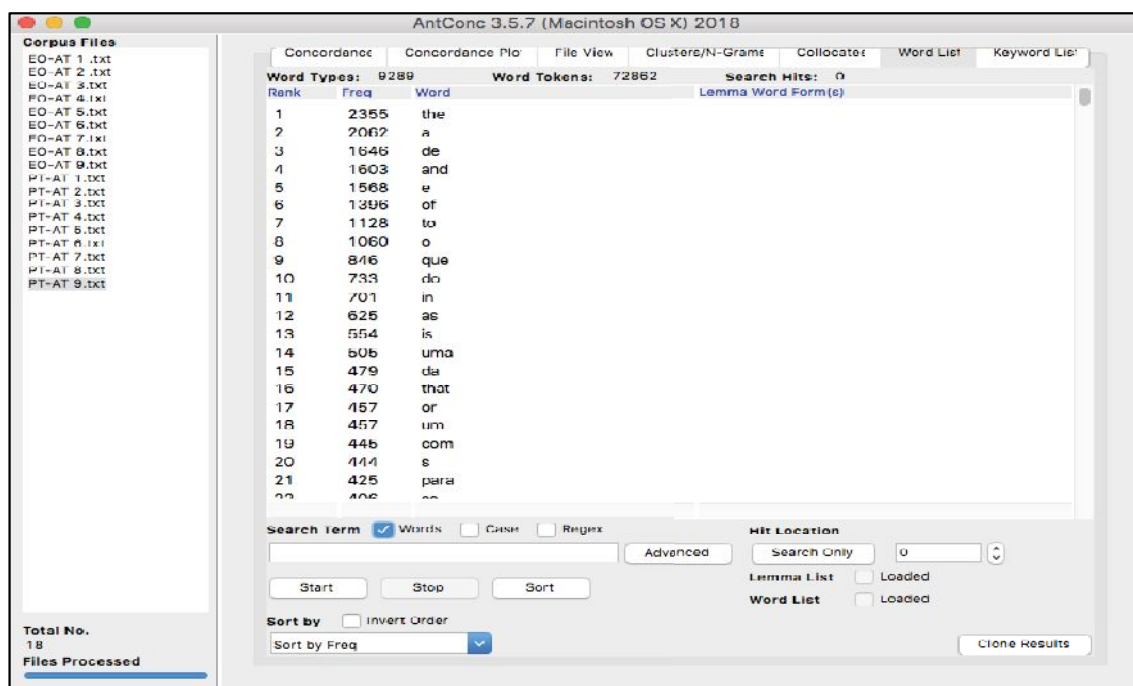
Artigos originais e traduzidos da área de Psicologia, mais especificamente da Análise Transacional, foram retirados do site *Institute for Integrative Psychotherapy*⁸. Com esses textos, um *corpus* foi compilado – os textos originais e traduzidos foram coletados, convertidos para o

⁸Disponível em <<http://www.integrativetherapy.com>>. Acesso em nov. 2017.

formato .txt, renomeados e salvos em uma pasta eletrônica devidamente identificada, gerando, assim, um *corpus* paralelo com 18 textos, entre originais e traduções, com um total de 72 862 *tokens* (número total de palavras do *corpus*). Esse *corpus* é considerado, de acordo com Berber Sardinha (2002), um *corpus* pequeno, pois é inferior a 80 mil palavras, mas que cumpre o requisito de representatividade, por essa pesquisa ser um estudo-piloto com o intuito de testar as funcionalidades dos SMT na construção de MT utilizando *corpora* paralelos.

Na figura 5, na captura da tela do programa *AntConc* carregado com o *corpus* paralelo compilado, pode-se observar a quantidade de arquivos que o compõem (*corpus files*), a quantidade de palavras contidas, excluindo as repetições (*Word Types*) e a quantidade total de palavras ou tamanho do *corpus* (*Word Tokens*).

Figura 5: Tela com o tamanho do *corpus* paralelo de AT, aberto na aba *Word List* do *AntConc*.



Fonte: a autora

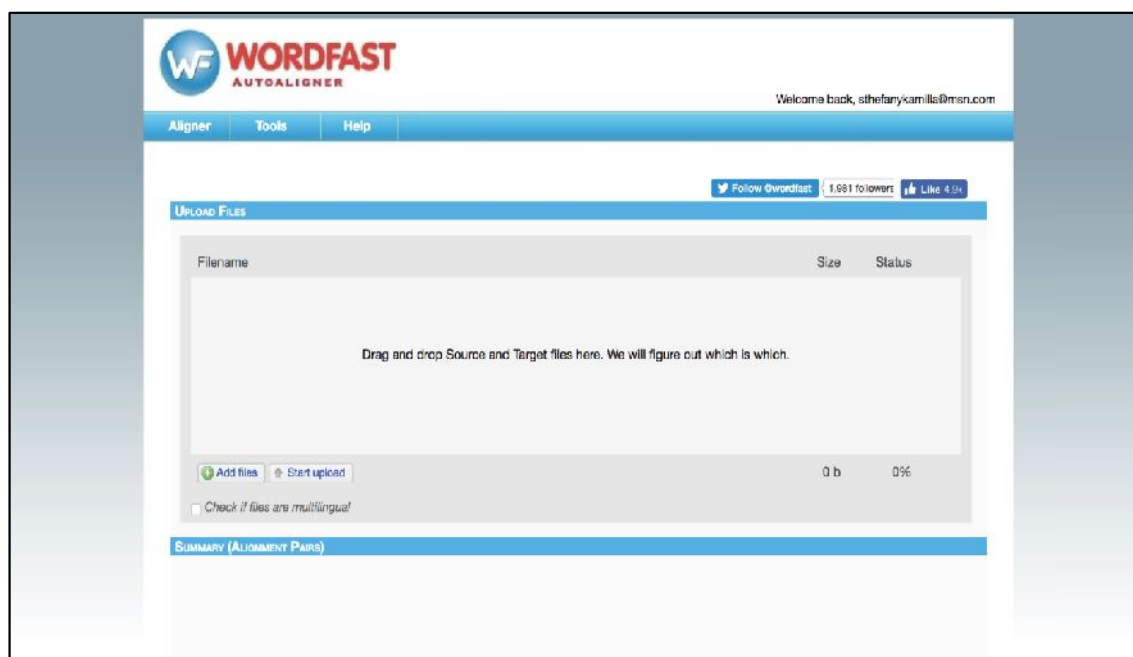
Com a compilação desse *corpus*, foi possível compreender que a compilação de um *corpus* exclusivamente para a criação de uma MT é uma tarefa complexa e que exige certo esforço. Entretanto, se o tradutor trabalha frequentemente com determinada área, vale construir um *corpus* paralelo que supra ambos os objetivos, uma consulta linguística através do *AntConc* com possibilidades de análise que o *corpus* paralelo oferece, e, além disso, criar uma MT que possa ser utilizada ativamente na tradução ou simplesmente para consulta.

3.2 Criação da MT utilizado o *Corpus* paralelo no *Wordfast Anywhere AutoAligner*

Com o *corpus* paralelo da AT compilado, o SMT *Wordfast Anywhere* foi escolhido para ser o sistema possibilitador da construção da MT, como já mencionado. O *Wordfast Anywhere* disponibiliza uma ferramenta de alinhamento *online* chamada *Wordfast AutoAligner*, na qual possibilita que TO e TT sejam alinhados automaticamente, disponibilizando o arquivo em .tmx (*translation memory exchange, i.e.*, formato internacional e intercambiável de arquivos das MT) para *download*.

A Figura 6 mostra a página inicial da ferramenta de alinhamento do SMT *Wordfast Anywhere, o AutoAligner*; basta fazer o carregamento dos textos TO e TT, e o sistema se encarrega de identificar as línguas dos TO e TT, de segmentar o(s) texto(s)-fonte (de acordo com a segmentação padrão do *Wordfast*) e fazer o alinhamento das traduções correspondentes a cada segmento.

Figura 6: Página inicial do *Wordfast AutoAligner* (2018)



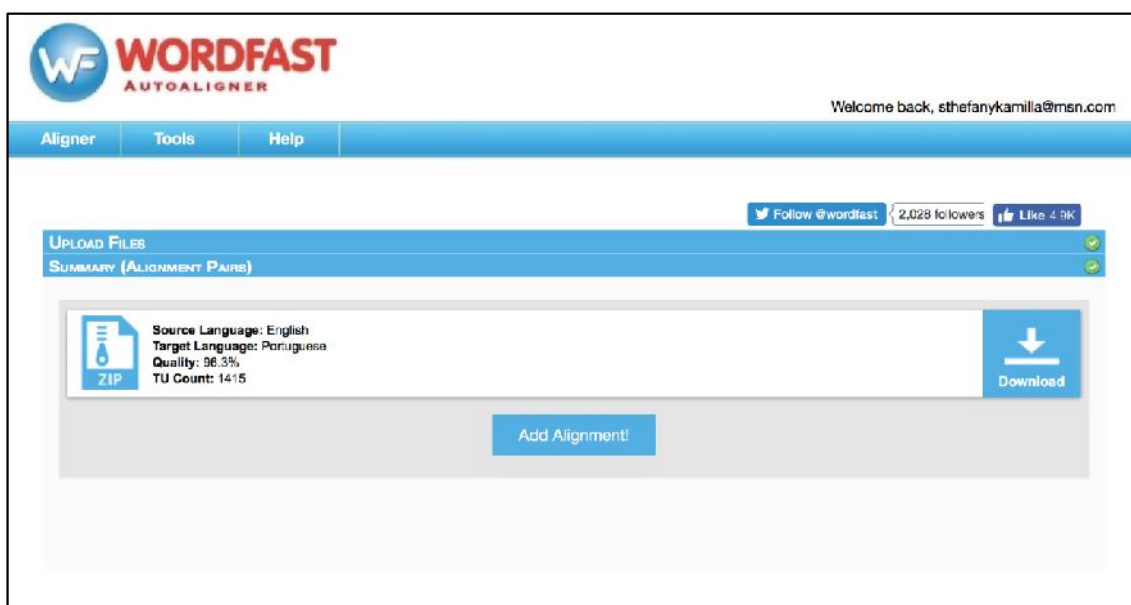
Fonte: *Wordfast AutoAligner* (2018)

Portanto, o processo de criação da MT dessa pesquisa se deu, primeiramente, com a seleção e o carregamento dos 18 arquivos do *corpus* paralelo (TO e TT). Como o *corpus* de pesquisa era paralelo bilíngue, não foi necessário realizar o acréscimo de nenhuma informação.

Os arquivos foram carregados e o programa analisou automaticamente as línguas envolvidas e a direção que elas apareciam, no caso, inglês original para português traduzido.

Depois desses processos, o programa oferece a opção, através de caixas de seleção, para que os marcadores de formatação e segmentação (conhecidas como *tags*) sejam mantidos ou não e que o arquivo gerado seja enviado por *e-mail*. Na pesquisa, as *tags* não foram mantidas, pois o *corpus* é composto de textos em formato .txt e não carrega formatações complexas. A opção de envio por *e-mail* do arquivo gerado foi selecionada e o deu-se início ao alinhamento. O *Wordfast AutoAligner* levou aproximadamente dez minutos para concluir o processo de alinhamento, ou seja, todos os 18 arquivos foram alinhados e transformados em uma única MT.

Figura 7: Tela do *Wordfast AutoAligner* com o resumo dos dados da MT gerada e a opção de download.



Fonte: a autora

O resumo gerado pelo *AutoAligner*, demonstrado pela figura 7 comprova que o *corpus* paralelo compilado para essa pesquisa-piloto gerou uma MT qualidade alta, 96,3%, e que é formada por 1415 segmentos, na direção inglês original para português traduzido. Deve-se entender que o que o programa considera como avaliação de qualidade não é necessariamente o conteúdo semântico de equivalência tradutória entre TF e TT, mas, sim, se o segmento tem um outro segmento que correspondente e se esses pares estão alinhados.

3.2.1 Verificação da qualidade de alinhamento do *AutoAligner* e da MT construída

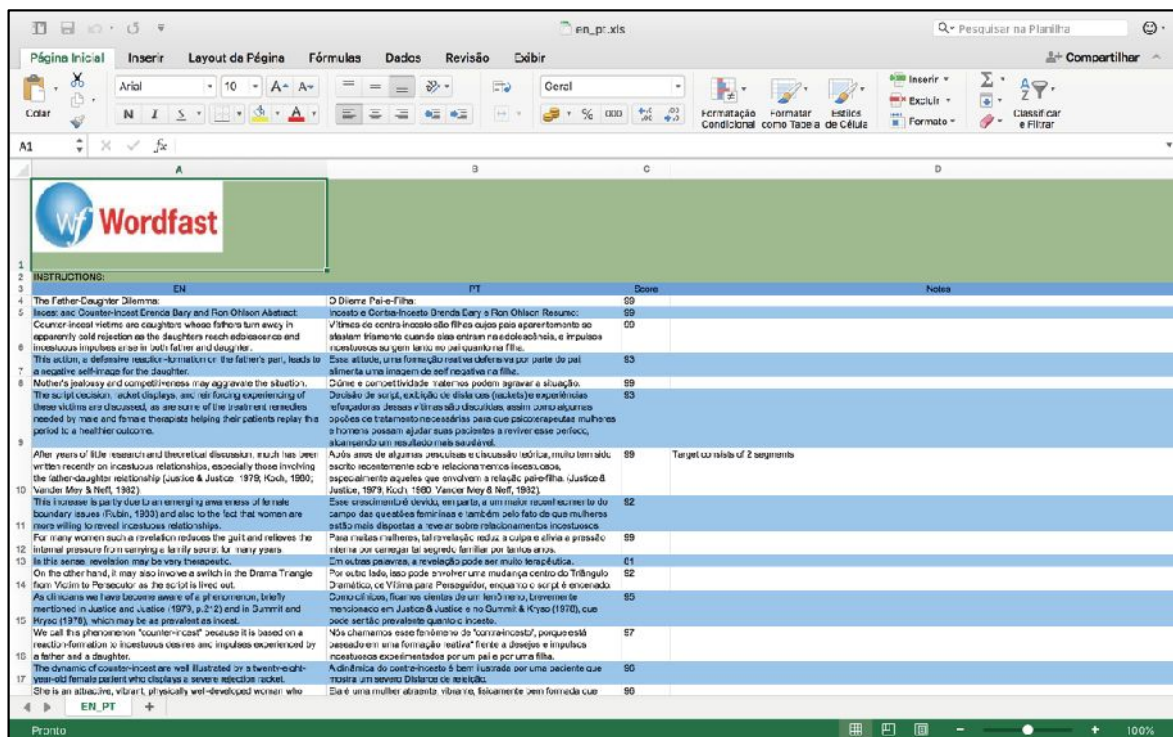
Como dito anteriormente nos tópicos 3.1.1 e 3.2, o *Wordfast AutoAligner* forneceu um resumo do arquivo produzido, e nele foi possível verificar a qualidade da MT. Ou seja, se o alinhamento feito pareou o segmento original com a sua respectiva tradução. Esse resultado é

obtido através da média total de segmentos e quantos foram os que não se alinharam corretamente.

Além disso, o *AutoAligner* forneceu, dentro da pasta de arquivos compactados para *download*, um arquivo em tabela para verificação dos segmentos em formato .xls, que pode ser aberto no *Excel*® ou *Numbers*®, por exemplo. Nesses arquivos, estão contidos todos os segmentos e suas traduções equivalentes, além de conter a porcentagem de paridade entre um e outro.

A figura 8 é uma captura de tela desse arquivo em .xls, composto pelo número do segmento, as línguas-fonte e alvo do segmento (EN e PT), a quantidade de pontos que aquele segmento obteve na paridade (*Score*), e observações sobre o alinhamento. É possível verificar, observando a figura 8, que o segmento traduzido número 10, por exemplo, foi composto por dois segmentos-fonte (*Notes*).

Figura 8: Tela do *Excel*® com o documento de verificação da segmentação feita pelo *Wordfast AutoAligner*.



	EN	PT	Score	Nota
1				
2				
3	INSTRUCTIONS:			
4	The Father-Daughter Dilemma:	O Dilema Pai-e-Filha:	59	
5	Incest and Counter-incest Ennata Bary and Ron Ohlson Abstract	Incesto e Contra-Incesto Ennata Bary e Ron Ohlson Resumo:	59	
6	Counter-incest victims are daughters whose fathers turn away in	Vítimas do contra-incesto são filhas cujos pais aparentemente se	59	
7	apparently self-rejection as the daughters reach adolescence and	distanciam quando elas entram na adolescência, e impulsiona		
8	incessuous impulses arise in both father and daughter.	numerosas su germes tanto no pai quanto na filha.	53	
9	This action, a defensive reaction to rejection or the father's part, leads to	Essa atitude, uma formação reativa defensiva por parte do pai,		
10	a sexual self-love for the daughter.	alimenta uma imagem de self-love na filha.	53	
11	Mother's jealousy and competitiveness may aggravate the situation.	Criar e competitividade maternas podem agravar a situação.	59	
12	The suicidal wishes, self displays, and self-facing experiencing of	Desejos de suicídio, exibição de si mesma (suicídio) experiências	59	
13	these victims are discussed, as are some of the treatment resources	relacionadas. Incesto e Incesto discutidos assim, como algumas		
14	needed by men and female therapists helping their patients replay this	opções de tratamento necessárias para que psicoterapeutas mulheres		
15	periodic a healthier outcome.	e homens possam ajudar suas pacientes a reviver esse período,		
16	After years of film research and theoretical discussion, such has been	abarcando um resultado mais saudável.	59	Target consists of 2 segments
17	written recently on incestuous relationships, especially those involving the	Após anos de algumas pesquisas e discussões teóricas, muito trabalho	59	
18	father-daughter relationship (Justice & Justice 1979; Koch, 1990;	sacrio recentemente sobre relacionamentos incestuosos,		
19	Van der Mey & Neff, 1992)	especialmente aqueles que envolvem a relação pai-filha (Justice &	52	
20	This course is partly due to an interest in areas of the male	Essa abordagem levou, em parte, a um maior interesse no do		
21	boundary issues (Rubin, 1993) and also to the fact that women are	campo das questões femininas e também pelo fato de que mulheres		
22	more willing to reveal incestuous relationships.	estão mais dispostas a revelar sobre relacionamentos incestuosos.	59	
23	For many women such a revelation reduces the guilt and relieves the	Para muitas mulheres, tal revelação reduz a culpa e alivia a pressão		
24	internal pressure from carrying a heavy secret for many years.	interna por carregar um segredo familiar por tantos anos.	59	
25	In this sense, revelation may be very therapeutic.	Em outros casos, a revelação pode ser muito terapêutica.	61	
26	On the other hand, it may also involve a switch in the Drama Triangle	Por outro lado, isso pode envolver uma mudança centro do Triângulo	52	
27	from Victim to Persecutor as the secret is freed out.	Dramático, de Vítima para Perseguidor, enquanto o segredo é liberado.	55	
28	An incestuous man may become aware of a phenomenon, briefly	Um incestuoso, ficar no centro de um fenômeno brevemente		
29	mentioned in Justice and Justice (1979, p.2'2) and in Dummlil and	menção em Justice & Justice e no Dummlil & Kyoa (1970), que		
30	Myrao (1978), which may be as prevalent as incest.	pois seria prevalente quanto a incesto.	57	
31	We call this phenomenon "counter-incest" because it is based on a	Nós chamamos esse fenômeno de "contra-incesto", porque está		
32	reaction formation to incestuous desires and impulses experienced by	baseado em uma formação reativa forte a desejos e impulsos		
33	a father and a daughter.	incestuosos excessivamente por um pai e por uma filha.	50	
34	The dynamic of counter-incest are well illustrated by a twenty-eight-	A dinâmica do contra-incesto é bem ilustrada por uma paciente que		
35	year-old female patient who disclosed a severe psychotic attack.	mostrou um severo ataque psicótico.	50	
36	She is an attractive, vibrant, typically well-developed woman who	Ela é uma mulher atraente, vibrante, tipicamente bem formada que	50	

Fonte: a autora

3.3 Ativação e utilização da MT construída no *Wordfast Anywhere*

Após todos os processos descritos, a MT construída, utilizando um *corpus* paralelo, foi carregada na aba de gerenciamento de MT e glossários do *Wordfast Anywhere*, renomeada

como: EN>PT ATMONO e selecionada como MT ativa. As alterações foram salvas e a aba fechada. A partir desse momento a MT estava disponível para ser usada em qualquer tradução.

Vale observar que a MT construída não precisa necessariamente estar ativa, pode também ser utilizada apenas como MT de consulta, enquanto o tradutor alimenta outra MT ativa.

3.4 Teste da MT externa

Para testar a MT externa utilizando o *corpus* paralelo da AT foi traduzido um artigo acadêmico da área de AT, intitulado *The Ok Position: Freedom, Equality and the Pursuit of Happiness* escrito por Claude M. Steiner e publicado no *Transactional Analysis Journal*⁹, volume 42, de 2012.

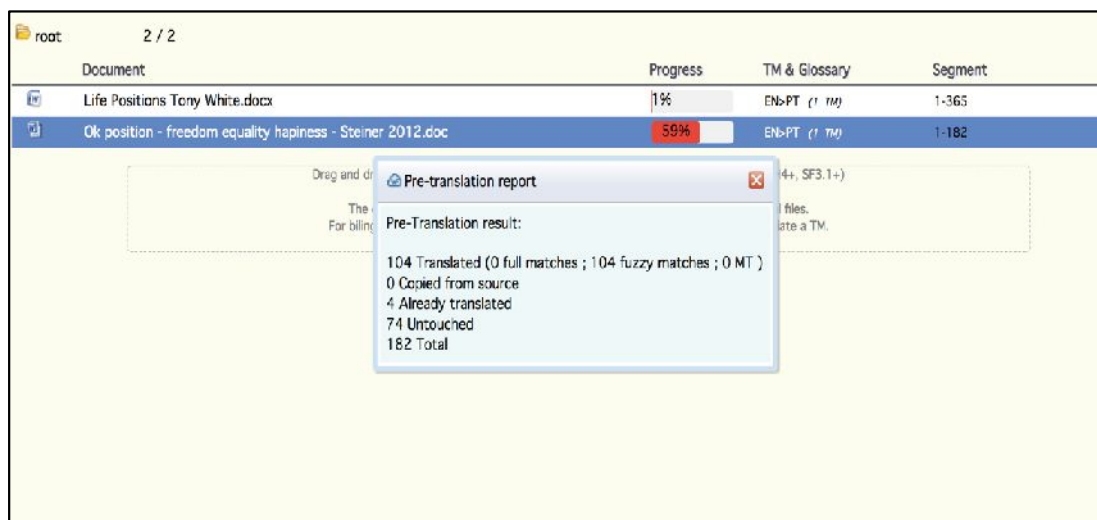
Claude Steiner¹⁰ é um dos principais pesquisadores da linha de pensamento de Eric Berne (quem iniciou os estudos da Análise Transacional), e o pesquisador que mais escreve sobre as teorias de Berne e sobre AT. A revista especializada *Transactional Analysis Journal*, em todas as suas edições, possuem mais de 300 textos de Steiner. O artigo *The Ok Position: Freedom, Equality and the Pursuit of Happiness*, especificamente, trata sobre as posições existenciais descritas por Berne em 1964, no livro *Games people play* e que são muito discutidas por psicólogos. Portanto, selecionando um artigo de um autor que ocupa uma posição de destaque e um tema muito recorrente, pode-se pressupor que a MT auxilie na tradução de vários textos da área.

Tanto a tradução quanto a MT foram feitas utilizando o SMT gratuito *Wordfast Anywhere*, assim, foi possível verificar que o sistema possui ferramentas capazes de construir uma MT externa e, para além disso, que a MT construída com o *corpus* paralelo da área auxilia na tradução de um artigo dessa mesma área. O teste comprovou que a criação de uma memória de tradução utilizando *corpora* paralelos é válida.

⁹ O *Transactional Analysis Journal* é uma revista sobre a teoria da Análise Transacional, princípios e aplicações da teoria em vários campos. É publicada trimestralmente desde 1971 e oferece artigos acadêmicos de todas as perspectivas teóricas e campos de aplicação, incluindo pesquisas quantitativas e qualitativas, estudos de caso, pesquisas de literatura, resenhas de livros e ensaios reflexivos.

¹⁰ O website de Eric Berne contém uma bibliografia sobre Claude Steiner. <http://www.ericberne.com/claude-steiner-biography/>

Figura 9: Quantidade de resgates da MT externa, construída utilizando *corpus* paralelo da área de AT, ao traduzir o artigo de Steiner.



Fonte: a autora

Na figura 9 é possível perceber que muitos segmentos são resgatados da memória de tradução e que todos eles possuem correspondências linguísticas parciais, com os segmentos contidos no texto (*fuzzy matches*). Cabe ao tradutor perceber que essas indicações da MT devem ser avaliadas, lidas com cautela e aceitas somente se corresponderem à tradução daquele segmento, mesmo para as recuperações de memória que apareçam com uma porcentagem de correspondência elevada (maior que 90% ou *golden matches*). Os SMT e as MT são recursos que aumentam consideravelmente o rendimento do trabalho tradutório, entretanto, seus benefícios só podem ser aproveitados se o usuário tiver um olhar crítico sobre cada segmento recuperado, independentemente da qualidade da MT ou da similaridade linguística. Um erro tradutório por aceitação de indicações da MT pode comprometer o trabalho de forma considerável e isso coloca sob questão todos os benefícios oferecidos pelos SMT, conforme explicam Esqueda e Stupiello (2016) e Esqueda *et alii.* (2017).

3.5 Avaliação das vantagens e desvantagens do SMT gratuito *Wordfast Anywhere* na criação de MT utilizando *corpus* paralelo.

Pode-se avaliar o SMT *Wordfast Anywhere* como eficaz na construção de uma MT utilizando um *corpus* paralelo, pois a plataforma gratuita e *online* consegue alinhar esse conjunto de TO e TT de forma integrada, transformando-os em uma única memória de tradução. Isso é o que podemos considerar como principal vantagem do *Wordfast Anywhere/ AutoAligner*, pois a ferramenta desse SMT executa a tarefa dada de forma satisfatória. Outra grande

vantagem é a rapidez e facilidade com que o processo de alinhamento é feito no *AutoAligner*. Os textos são carregados na ferramenta e alinhados automaticamente, como o nome da ferramenta já explicita.

Ao fazer o alinhamento e transformar o *corpus* em MT, o sistema oferece um resumo dos dados da memória construída, informando quantidade de segmentos, qualidade de alinhamento e direção das línguas envolvidas no processo. Essas informações são importantes, já que o processo é feito automaticamente. A automação poderia ser considerada uma desvantagem se o usuário não pudesse verificar se o alinhamento foi executado de maneira correta, e se o segmento de TO realmente corresponde ao segmento de TT. Entretanto, ao fazer o *download* da MT, o sistema fornece, além do arquivo em *.tmx*, uma tabela em *.xls* que possibilita a visualização de cada segmento gerado e a percentagem de paridade semântica entre TO e TT (figura 8), possibilitando ao usuário a confirmação da qualidade de cada segmento.

O método proposto não funciona apenas para os usuários que utilizam o *Wordfast Anywhere*, pois as memórias construídas com a ferramenta *AutoAligner* podem ser executadas/inseridas em qualquer SMT do mercado, já que a MT é fornecida em *.tmx*, que, como já dito, é um formato internacional e intercambiável de MT. Essa é uma vantagem considerável dos sistemas gratuitos e/ou de código aberto, a liberdade que o usuário tem de utilizá-los sem se prender a eles. Eles utilizam e geram formatos compatíveis com todos os tipos de sistema operacional e SMT, e, no caso do *Wordfast Anywhere*, por ser uma ferramenta *online*, pode ser acessada de qualquer computador, em qualquer lugar, assim o usuário não depende de um único computador para utilizar o sistema.

O *Wordfast Anywhere*, bem como a ferramenta integrada *AutoAligner* são executados através de conexão com a *Internet*, precisam ser abertos em um navegador e estarem vinculados a uma conta de usuário, o que pode gerar algumas desvantagens. Uma conexão instável com a rede faz com que o SMT perca a conexão com o servidor e volte para a página de *login* e o usuário tenha que entrar com a sua conta e executar todo o processo de alinhamento novamente, ou, em outros casos, causa demora na execução de tarefas, como, por exemplo, no carregamento dos arquivos ou no processo de alinhamento e construção da MT. O navegador utilizado para acessar o *Wordfast Anywhere* também pode influenciar no funcionamento do sistema, que responde de maneira inesperada quando acessado através de alguns navegadores (*Safari* e *Mozilla Firefox*). Recomenda-se, através de testes durante essa pesquisa, que o navegador *Google Chrome* seja utilizado, evitando que a conexão do SMT com o servidor se interrompa com facilidade e o usuário perca o trabalho realizado anteriormente.

O produto produzido pelo alinhamento no *AutoAligner* não fica vinculado automaticamente ao SMT. Processos como o *download* dos arquivos, a descompactação (executada por programas adicionais) e verificação dos segmentos são feitos externamente e necessitam de um certo domínio de funções básicas de informática. E depois desses processos, ainda deve ser feita a adição da MT no SMT *Wordfast Anywhere* ou em qualquer outro SMT, o que demanda mais alguns procedimentos. Essa quantidade de ações necessárias para utilizar a MT externa, construída pelo *AutoAligner* também pode ser considerada uma desvantagem. Uma vantagem da formação acadêmica dos alunos do Curso de Tradução da UFU é poder estudar e praticar tanto a compilação de *corpus* quanto as formas de utilização nas disciplinas do curso.

Portanto, podemos citar como vantagens proeminentes, a gratuidade do *Wordfast Anywhere*, a facilidade e a rapidez na construção das MT, a unificação de todos os arquivos do *corpus* paralelo em uma única MT, a compatibilidade do *Wordfast Anywhere* com todos sistemas operacionais, e a liberdade que o usuário tem de utilizar o produto. E como desvantagens, a necessidade da conexão com a *Internet*, alguns desajustes no funcionamento do sistema, ao utilizar outros navegadores, e muitos processos para extrair o produto gerado pelo *AutoAligner* devem ser destacadas.

CAPÍTULO 4- DESCRIÇÃO DETALHADA DOS PROCEDIMENTOS PARA CRIAÇÃO DA MT A PARTIR DO *CORPUS* PARALELO NO *WORDFAST ANYWHERE*.

Considerando a importância das MT no processo tradutório e que muitos tradutores têm textos traduzidos ou *corpus* de consulta e não possuem, necessariamente, MT sobre eles, considerou-se a importância de descrever, em detalhes, o processo de criação de memórias utilizando, nesse caso, um *corpus* paralelo bilíngue, o *Wordfast Anywhere* e *AutoAligner*. Dessa forma, este capítulo tornou-se um subproduto da pesquisa, capaz de contribuir para que outros profissionais tenham apoio teórico para darem mais um passo na utilização dos *corpora* paralelos e criarem MT de áreas diversas.

4.1 Abertura de uma conta de acesso no *Wordfast Anywhere*

Para que o acesso ao *Wordfast Anywhere* seja liberado, é necessário criar uma conta gratuita usando endereço de *e-mail* válido e definir uma senha de acesso. A plataforma pede a ativação da conta através de um *e-mail* enviado ao endereço fornecido. Assim que o usuário confirma os dados, a plataforma libera o acesso imediato, basta inserir o *e-mail* e a senha de acesso, e clicar em *Start*. O *Wordfast Anywhere*, como dito anteriormente, é um SMT *online*, todos os arquivos e dados ficam salvos na conta do usuário.

Figura 10: Tela de criação da conta de acesso no *Wordfast Anywhere*.

Wordfast Anywhere

Configurações

Seguro | <https://www.freetm.com/jsp/create.jsp>

W- ANYWHERE

Get started with your free account.
Type an existing email address and a secure password.

tutorialmonografia@gmail.com

.....

A sua palavra-passe deveria ter pelo menos 8 caracteres e conter pelo menos uma letra e um algarismo.

✓ Não sou um robô

By clicking on the Start button, you agree to our [Termos e condições](#).
Your email will be checked first of all.

Start

Fonte: *Wordfast Anywhere*

4.2 Escolha do *corpus* paralelo, geral ou de uma área específica

A escolha do *corpus* paralelo é um processo importante na criação da MT, é ele que fornecerá todo o conteúdo linguístico da MT. Portanto, o *corpus* deve ter sido compilado seguindo critérios importantes para esse fim, tais como, área de estudo, gênero e direção linguística da tradução. Aqui, este processo foi listado logo a seguir da escolha do SMT. Entretanto, deve-se considerar que o *corpus* paralelo já esteja devidamente compilado, pois a compilação do *corpus* não será descrita detalhadamente nesse tópico.

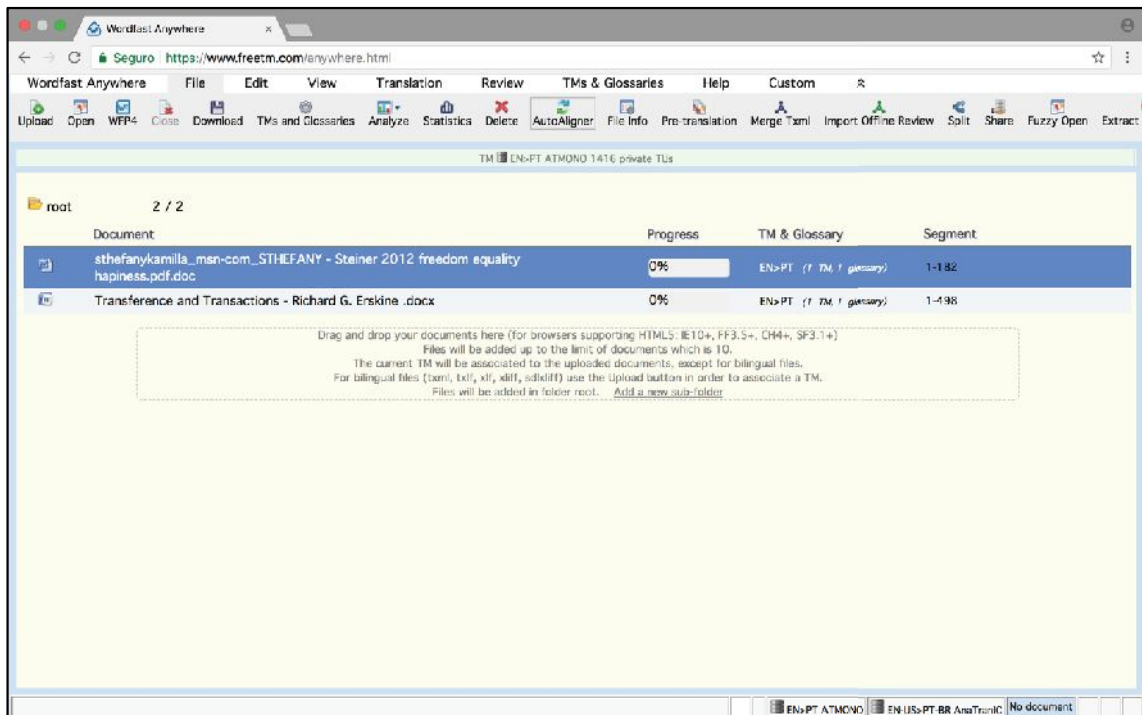
É recomendável que o *corpus* seja composto de TT que reproduzam uma fidelidade estrutural com o TO, do contrário, a ferramenta de alinhamento não conseguirá parear todos os segmentos e suas respectivas traduções, podendo criar uma MT com uma baixa qualidade de alinhamento, cheia de segmentos em branco e/ou com disparidade entre TO e TT.

4.3 Acesso ao *Wordfast AutoAligner*

Após o login e o acesso à página inicial do *Wordfast Anywhere*, é possível visualizar, na barra de opções superior, várias opções dadas pelo SMT. Entre elas, a opção *file* (arquivo), na qual deve ser selecionada para gerar uma segunda barra de ferramentas relacionadas com os arquivos, entre as opções, deve-se clicar em *AutoAligner* para ter acesso à ferramenta de alinhamento fornecida pelo *Wordfast Anywhere*.

Uma nova guia será aberta no navegador, com a função de alinhamento do *Wordfast Anywhere* – o *Wordfast Anywhere AutoAligner*. Esse processo deve ser feito inteiramente dentro do SMT e depende da conta de acesso da plataforma para liberar o uso da ferramenta.

Figura 11: Acesso à ferramenta *Wordfast AutoAligner* pela barra de ferramentas do *Wordfast Anywhere*.

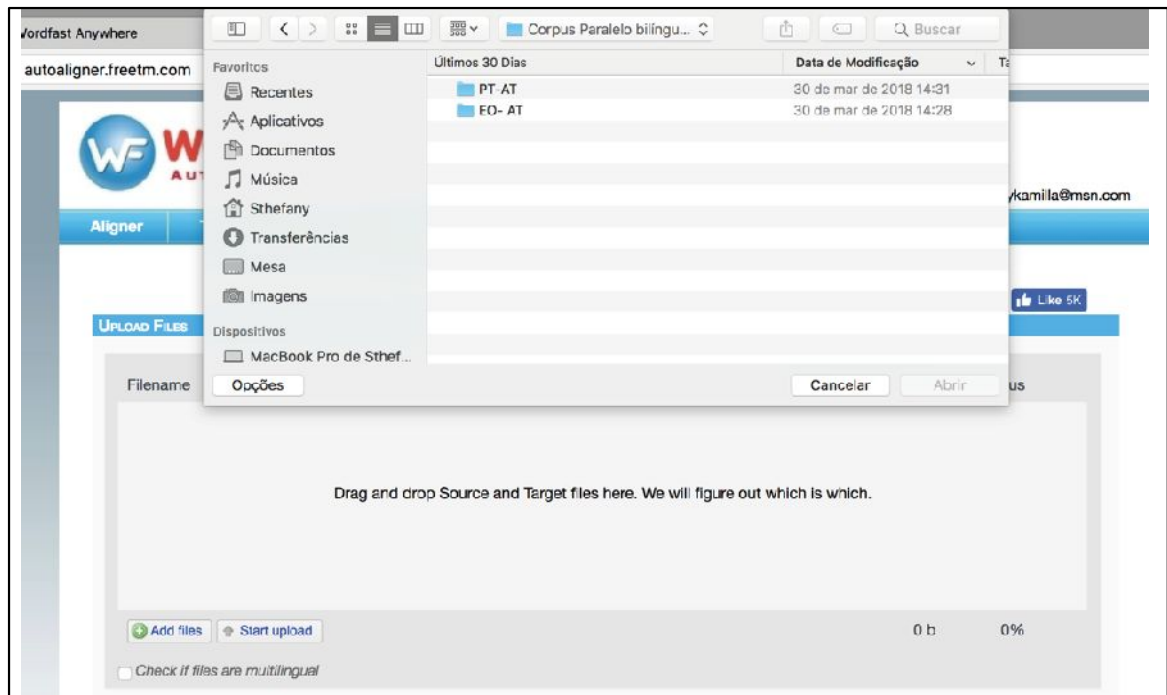


Fonte: *Wordfast Anywhere* (2018)

4.4 Inserção do *corpus* paralelo dentro da ferramenta *AutoAligner*

Com a ferramenta de alinhamento aberta na nova aba, é necessário adicionar os arquivos que deverão ser alinhados. Na opção *add files* (adicionar arquivos), deve-se carregar todos os arquivos (originais e traduções) do *corpus* paralelo, que estarão salvos em uma pasta devidamente nomeada, seguindo as regras estabelecidas pela LC para a compilação de um *corpus* paralelo.

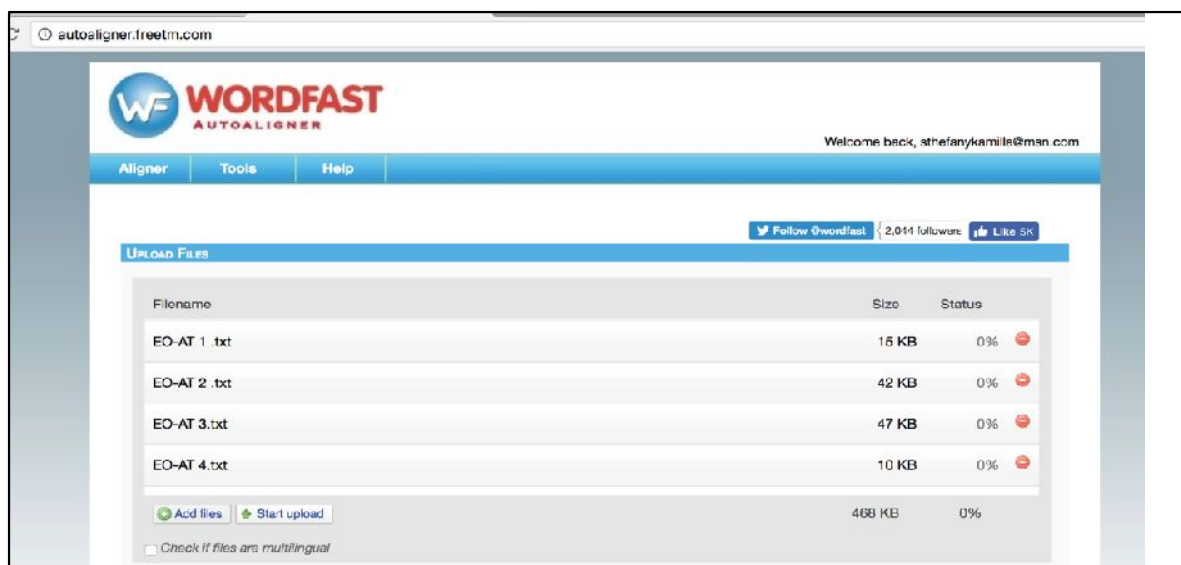
Figura 12: Seleção dos arquivos do *corpus* paralelo a serem carregados no *Wordfast AutoAligner*.



Fonte: a autora

Se o *corpus* paralelo for multilíngue, é necessário marcar a opção para arquivos multilíngues, do contrário, deve-se somente iniciar o carregamento dos arquivos na opção *Start upload*.

Figura 13: *Corpus* paralelo carregado no *Wordfast AutoAligner*.

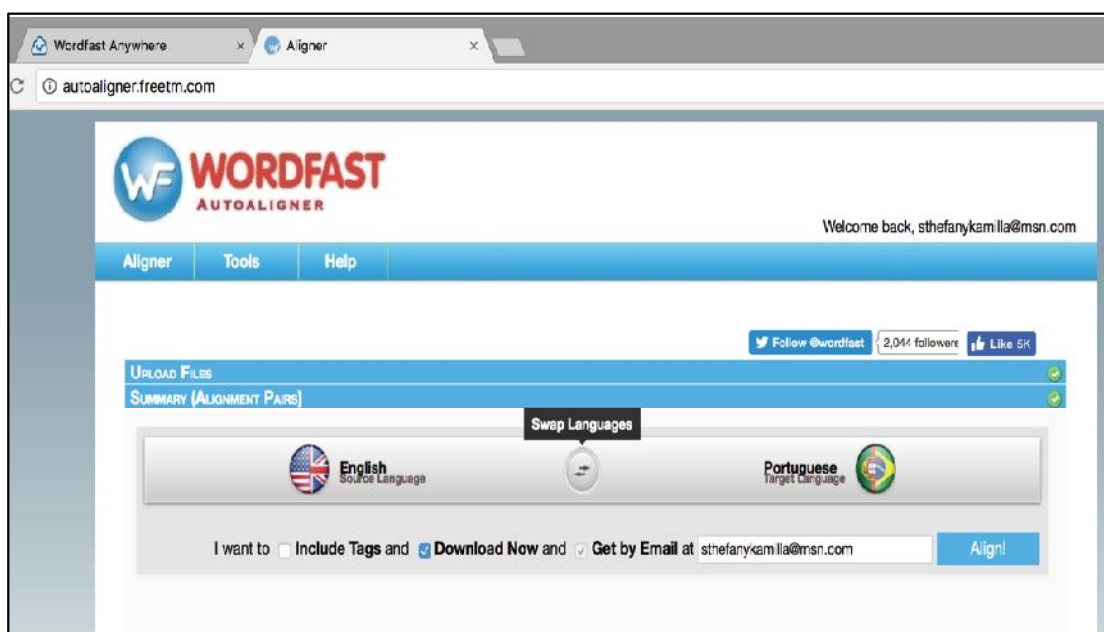


Fonte: a autora

4.5 Processo de alinhamento na ferramenta *AutoAligner*

A ferramenta irá identificar automaticamente a língua dos TO e a língua dos TT, se a direção da tradução estiver em ordem inversa, basta clicar na opção *Swap Languages* (Trocar par linguístico). A ferramenta fornece as opções de receber o arquivo de alinhamento (MT) por *e-mail* e/ou fazer o *download* automaticamente, é necessário selecionar as caixas com as opções desejadas, adicionar um *e-mail* para o envio dos arquivos – não é necessário que o *e-mail* esteja vinculado à conta de acesso do *Wordfast Anywhere*, a MT pode ser compartilhada com qualquer pessoa, usuário ou não – e autorizar o alinhamento clicando em *Align!*.

Figura 14: Identificação da língua original e traduzida dos arquivos do *corpus* paralelo.

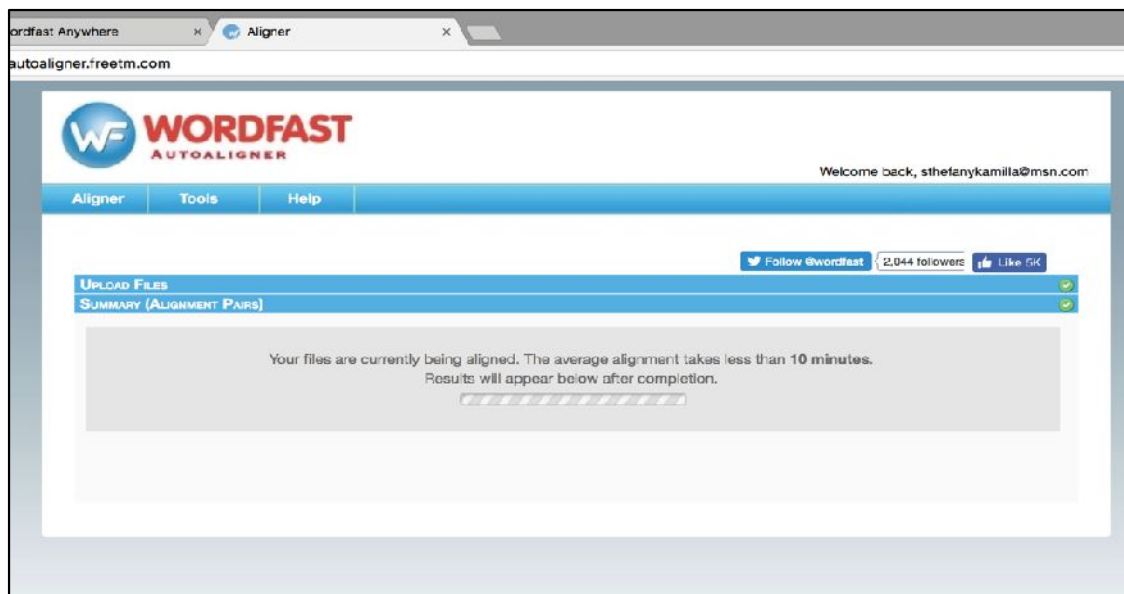


Fonte: *Wordfast AutoAligner* (2018)

Após o carregamento dos arquivos e definição das línguas dos TO e TT, a ferramenta *AutoAligner* executa o processo de alinhamento automaticamente. Aparecerá na tela uma mensagem com os dizeres: *Your files are currently being aligned. The average alignment takes less than xx minutes. Results will appear below after completion*¹¹.

¹¹ Tradução nossa: *Seus arquivos estão sendo alinhados. O tempo máximo de alinhamento será inferior a xx minutos. Depois de completo, os resultados aparecerão abaixo*

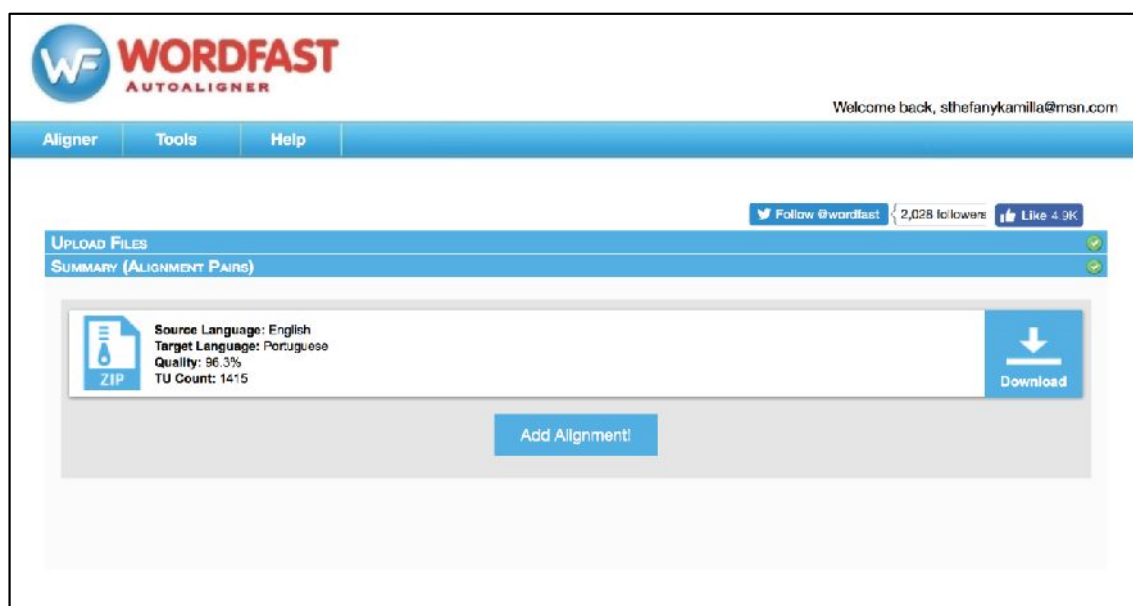
Figura 15: Tempo de processamento da ferramenta para gerar o arquivo.



Fonte: *Wordfast AutoAligner* (2018)

Após o prazo de processamento da ferramenta, o arquivo é enviado para o *e-mail* escolhido e também fica disponível para *download* manual. Nesse momento, aparece na tela um resumo do processo de alinhamento feito pelo *AutoAligner*, que informa a quantidade de segmentos alinhados, as línguas presentes nos segmentos originais e traduzidos, e, o mais importante, a qualidade do arquivo alinhado, ou seja, a qualidade de alinhamento da memória de tradução.

Figura 16: Resumo do processo de alinhamento.

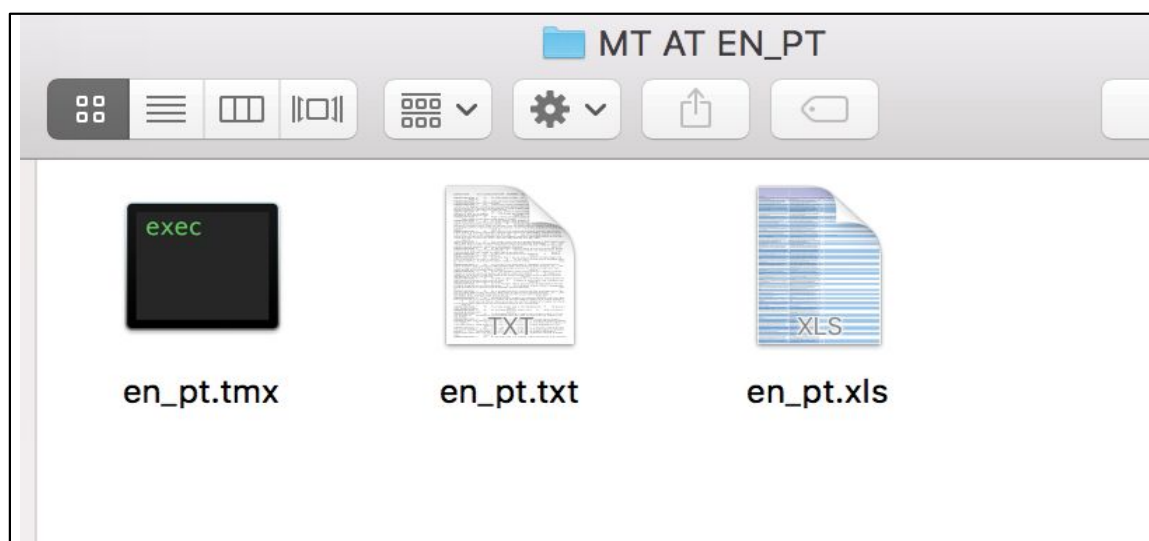


Fonte: a autora

4.6 Download e descompactação do arquivo gerado pelo *AutoAligner*

O *AutoAligner* disponibiliza para *download*, uma pasta compactada contendo os arquivos resultantes do alinhamento. Essa pasta deve ser descompactada utilizando qualquer descompactador de arquivos (como por exemplo, o WinRAR). Nela contém três arquivos, um no formato .tmx (formato internacional e intercambiável do arquivo das MT, compatível com todos os SMT disponíveis no mercado), um arquivo de texto com os segmentos alinhados em .txt (mais utilizado por programas de análise de *corpus*, como o *AntConc*) e uma tabela em .xls (compatível com editores de planilhas, como o *Microsoft Excel*®).

Figura 17: Arquivos gerados após o alinhamento no *AutoAligner*.



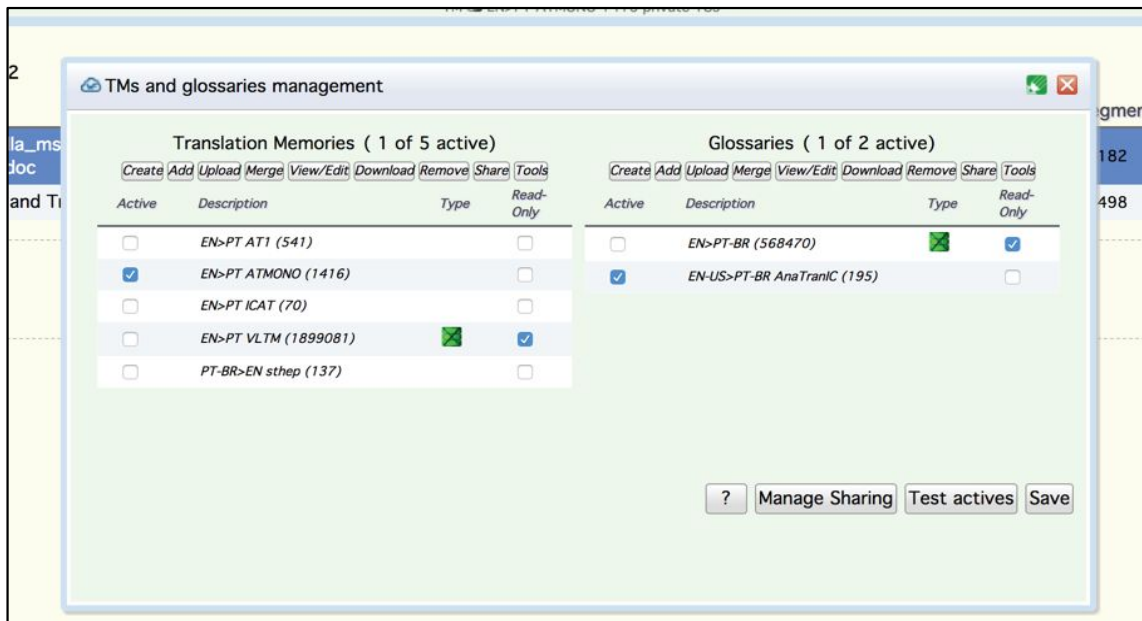
Fonte: a autora

4.7 Utilização da MT do *corpus* paralelo no *Wordfast Anywhere*

Após a criação e download da MT, o próximo e mais importante processo é a utilização da MT no SMT, e, para isso, é necessário adicionar essa memória no *Wordfast Anywhere*. Na página inicial, em *file* (barra de ferramentas superior) contém ferramenta de gerenciamento das MT e dos glossários: *TMs and glossaries*; é necessário clicar nessa opção para abrir a função de gerenciamento e adicionar uma nova MT.

A função gerenciamento é dividido em *Translation Memories* e *Glossaries*, como mostra a figura 17. A opção *upload* permite que uma nova MT seja carregada no SMT.

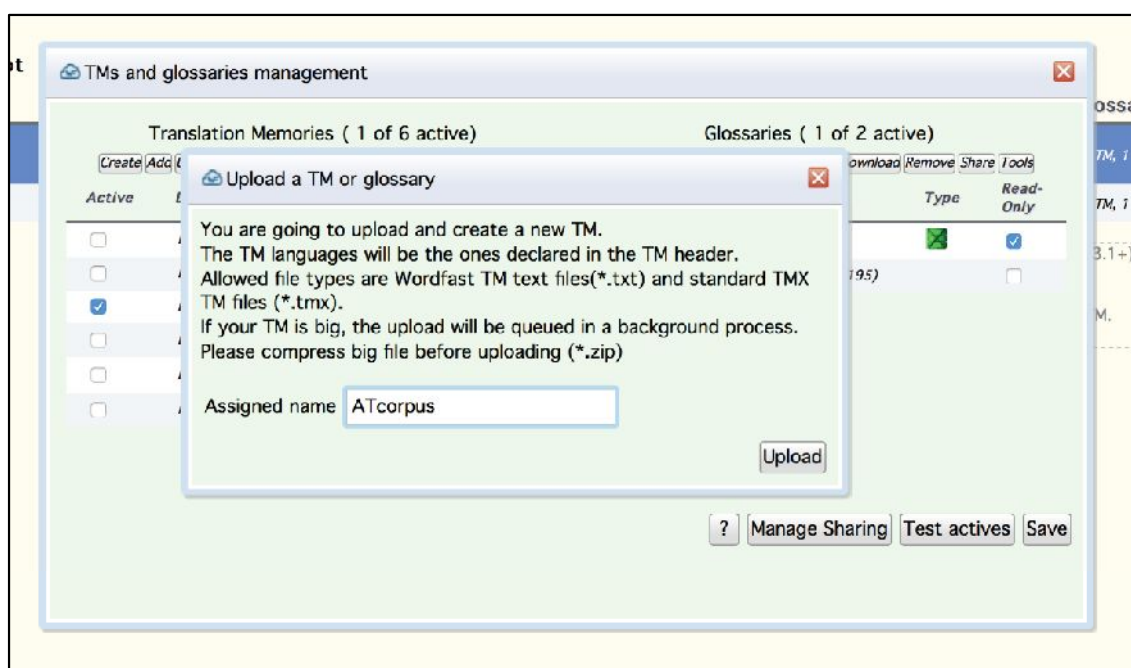
Figura 18: Gerenciamento das MT e glossários no *Wordfast Anywhere*.



Fonte: a autora

O SMT pede que a nova MT seja nomeada e depois carregada para o servidor. É recomendável nomear a MT de uma maneira que remeta ao conteúdo presente nela e que, posteriormente, seja fácil identificar qual estará sendo utilizada pelo SMT. No momento do carregamento, basta localizar a MT dentro na pasta descompactada, selecionando o arquivo .tmx e autorizar o carregamento clicando em *upload*.

Figura 19: Carregamento da MT no *Wordfast Anywhere*.



Fonte: a autora

Após o carregamento do arquivo em .tmx, obter-se-á a MT externa utilizando o *corpus* paralelo, carregada no servidor *Wordfast Anywhere*. O SMT verifica a quantidade de segmentos contidos na memória, se algum segmento possui outras línguas além das duas inicialmente configuradas, a quantidade de segmentos inválidos ou com erros, a quantidade de segmentos duplicados e a quantidade de segmentos válidos carregados no servidor, fornecendo um relatório no final da análise.

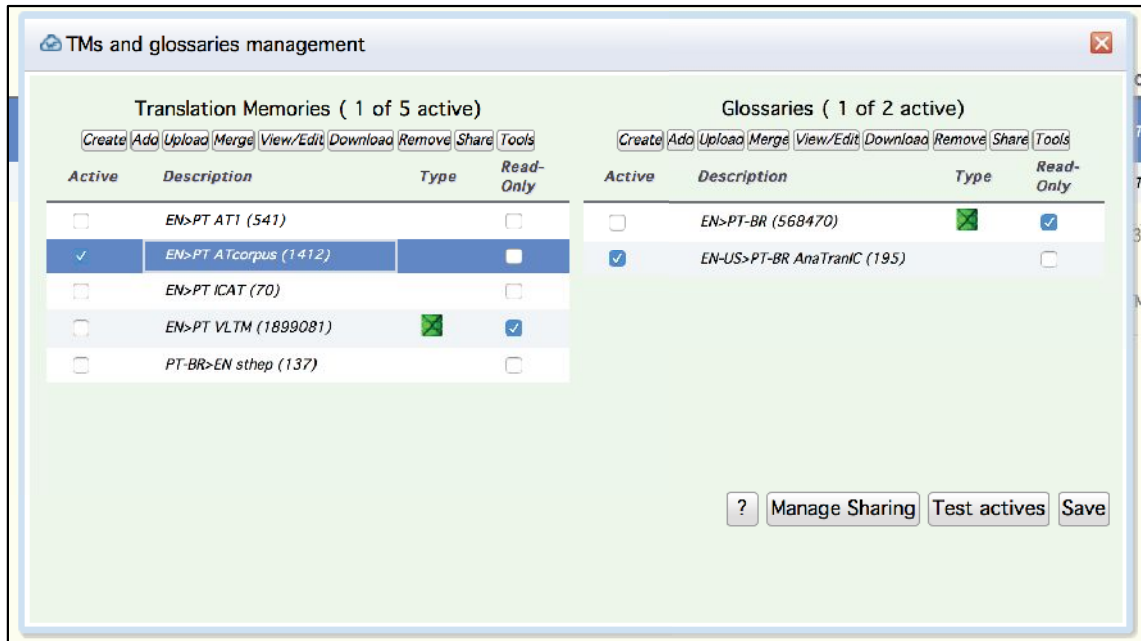
Figura 20: Relatório da MT do *corpus* paralelo feito pelo *Wordfast Anywhere*.

The screenshot shows the 'TMs and glossaries management' window. It contains two main sections: 'Translation Memories (1 of 7 active)' and 'Glossaries (1 of 2 active)'. The 'Translation Memories' section has a table with columns: Active, Description, Type, and Read-Only. The 'EN>PT ATcorpus (1412)' entry is selected. The 'Glossaries' section has a table with columns: Active, Description, Type, and Read-Only. The 'EN-US>PT-BR AnaTranC (195)' entry is selected. A 'TM upload report' dialog box is open, showing the following data:

TM upload report	
Total submitted TUs	1412
TUs with different languages in file	0
Invalid TUs in file	0
	EN>PT 1412
Duplicated TUs in file	0
Added TUs on server	1412

Fonte: a autora

Para definir MT criada através do *corpus* paralelo como a memória ativa, é fundamental que a caixa de opção *Active* em *Translation Memories* esteja selecionada e, após a seleção, que a opção seja salva. Após esse procedimento, a memória ativa é a MT externa.

Figura 21: Ativação da MT no *Wordfast Anywhere*.

Fonte: a autora

CONSIDERAÇÕES FINAIS

Esta pesquisa buscou descrever o uso de *corpus* paralelo na construção de memórias de tradução no *Wordfast Anywhere*, para que os tradutores iniciantes tivessem mais uma alternativa para construírem suas memórias de tradução e, mais do que isso, possam conhecer melhor algumas ferramentas gratuitas e suas funções.

Inicialmente, o objetivo era utilizar *corpora* paralelos previamente compilados, contudo, devido aos problemas de alinhamento do *corpus* disponibilizado e a necessidade de se compilar um *corpus* paralelo para a Análise Transacional, uma subárea da Psicologia, esse trabalho pôde expandir seus objetivos e contribuir para tal subárea, compilando um *corpus* paralelo e construindo uma MT específica dessa área nessa temática.

Pode-se considerar que a pesquisa e sua metodologia foram eficientes, visto que a proposta conseguiu proporcionar um diálogo teórico e prático entre os SMT e as ferramentas da Linguística de *Corpus*, potencializando e maximizando as ferramentas disponíveis e oferecendo novas possibilidades aos tradutores.

Concluiu-se que um SMT *online* e gratuito é capaz de construir MT utilizando *corpora* paralelos, e que tais MT podem auxiliar na tradução de textos específicos para uma determinada área de domínio. E, sobretudo, enfatizar que esses sistemas e memórias devem ser utilizados sob um olhar crítico do usuário, que jamais pode confiar cegamente em qualquer ferramenta tradutória.

Destaca-se que esse trabalho abre perspectiva para estudos exploratórios futuros, que poderão ser feitos utilizando outros *softwares*, proprietários ou não, no intuito de verificar se possuem os mesmos recursos disponíveis para a construção de uma memória utilizando *corpora* paralelos que o sistema gratuito ora investigado. Mesmo assim, acredita-se que o que foi apresentado aqui, principalmente o subproduto dessa pesquisa, é suficiente para auxiliar nas subáreas abordadas aqui, Tecnologias da Tradução e Linguística de *Corpus*, podendo servir como apoio teórico para tradutores em formação e aos interessados no processo de criação de MT utilizando *corpus* paralelos.

REFERÊNCIAS

- ANTHONY, Laurence . **AntConc**. Version 3.5.7 (Macintosh OS X). 2018. Disponível em: <<http://www.laurenceanthony.net/software/antconc/>>. Acesso em: 06 jun. 2018.
- ALVES, Fabio. Tradução, cognição e tecnologia: investigando a interface entre o desempenho do tradutor e a tradução assistida por computador. **Cadernos de Tradução**, v.2, n.14, p. 185-209, 2004.
- AHRENBERG, Lars. Alignment. IN: SIN-WAI, Chan (Ed.). **The Routledge Encyclopedia of Translation Technology**. Londres/Nova Iorque: Routledge, 2015. Cap. 24, p. 295-408.
- BAKER, Mona. *Corpora* in translation studies: An overview and some suggestions for future research. **Target. International Journal of Translation Studies**, v. 7, n. 2, p. 223-243, 1995.
- BOWKER, Lynne. Productivity vs. quality. **A pilot study on the impact of translation memory systems. Localisation Focus**, v. 4, n. 1, p. 13-20, 2005.
- BERNE, Eric. **Games People Play: The psychology of human relationships**. New York: Ballantine Books, 1964.
- BERNE, Eric. **Os Jogos da Vida : a Psicologia Transacional e o Relacionamento entre as Pessoas** [Adapt. Júlio de Moraes; trad. E. Artens]. Rio de Janeiro: Artenova, 1974.
- CHRISTENSEN, Tina Paulsen; SCHJOLDAGER, Anne. Translation-memory (TM) research: what do we know and how do we know it?. **HERMES-Journal of Language and Communication in Business**, v. 23, n. 44, p. 89-101, 2010.
- DARIN, Leila Cristina De Mello. O anseio da língua-mãe nas entrelinhas da tradução automática. **Galáxia. Revista do Programa de Pós-Graduação em Comunicação e Semiótica**, n. 2, p. 149- 167, 2001.
- ESQUEDA, Marileide; DA SILVA, Igor A. Lourenço; STUPIELLO, Érika Nogueira de Andrade. Examinando o uso dos sistemas de memória de tradução na sala de aula de tradução. **Cadernos de Tradução**, v. 37, n. 3, p. 160-184, 2017.
- ESQUEDA, Marileide Dias; STUPIELLO, Érika Nogueira de Andrade. A critical approach proposal for training translation memory systems in translator education programs. **Forum**, [s.l.], v. 14, n. 1, p.79-99, 2016.
- ESQUEDA, Marileide Dias; STUPIELLO, Érika Nogueira de Andrade;. Tecnologias e formação de tradutores. **Domínios de Lingu@gem**, [s.l.], v. 11, n. 5, p.1764-1781, 21 dez. 2017. EDUFU - Editora da Universidade Federal de Uberlândia.

- GARCIA, Ignacio. Computer-aided translation: Systems In: SIN-WAI, Chan (Ed.). **The Routledge Encyclopedia of Translation Technology**. Londres/Nova Iorque: Routledge, 2015. Cap. 3, p. 68-87.
- JESUS, Silvana Maria de. O *corpus* combinado e a pesquisa nos Estudos da tradução baseados em *corpora*. **Letras & Letras**, v. 30, n. 2, p. 291-321, 2014.
- LeBLANC, Matthieu. Translators on translation memory (TM): results of an ethnographic study in three translation services and agencies. **The International Journal of Translation & Interpreting**, v. 5, n. 2, 2013, pp. 1-13.
- MACHADO, Maria José Bellino. *Software* livre e de código aberto — um bom amigo do tradutor. **A folha**, n. 45, p.1-9, 2014.
- PYM, Anthony; BIAU GIL, José Ramón. Technology and translation (pedagogical overview). In: PYM, A.; PEREKRESTENKO, A.; STARINK, B. (Ed.). **Translation technology and its teaching**. Tarragona: Universitat Rovira I Virgili and Intercultural Studies Group, 2006, pp. 89-96.
- SARDINHA, Tony Berber. Tamanho de *corpus*. **The ESpecialist**, vol. 23, n. 2, p. 103-122, 2002.
- SARDINHA , Tony Berber. **Linguística de Corpus**. Editora Manole Ltda, 2004.
- SINCLAIR, J. *Corpus* and Text - Basic Principles. In: M. WYNNE (Ed.), **Developing linguistic corpora: a guide to good practice**. Oxford, Oxbow Books, p. 1-16, 2005.
- STEINER, Claude M. The OK position: Freedom, equality, and the pursuit of happiness. **Transactional Analysis Journal**, v. 42, n. 4, p. 294-297, 2012.
- STUPIELLO, Érica Nogueira de Andrade. A aplicação de sistemas de memória de tradução como ferramentas de produtividade para o tradutor. In: **Ética profissional na tradução assistida por sistemas de memórias** [online]. São Paulo: Editora UNESP, p. 75-139, 2014.
- STUPIELLO, Érica Nogueira de Andrade. O impacto das novas tecnologias no tempo e na qualidade da produção tradutória. **Estudos Linguísticos**, São Paulo, n.37, p. 45-154, 2008.
- WILLIAMS, Jenny; CHESTERMAN, Andrew. **The map: a beginner's guide to doing research in translation studies**. Routledge, 2014.
- WORDFAST Anywhere**. Version online. Nova Iorque, EUA. 2018. Disponível em: <<http://www.FreeTM.com>>. Acesso em: 06 jun. 2018.