

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
DEPARTAMENTO DE LETRAS ESTRANGEIRAS MODERNAS
FACULDADE DE LETRAS**

Ana Cláudia Duarte Corrêa

**CONTRIBUIÇÕES DA LINGUÍSTICA DE *CORPUS*
PARA OS ESTUDOS DA TRADUÇÃO:
um estudo de caso a partir da Tradução Técnica**

Juiz de Fora
2023

Ana Cláudia Duarte Corrêa

**CONTRIBUIÇÕES DA LINGUÍSTICA DE *CORPUS*
PARA OS ESTUDOS DA TRADUÇÃO:
um estudo de caso a partir da Tradução Técnica**

Trabalho de Conclusão de Curso apresentado ao Departamento de Letras Estrangeiras Modernas da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do grau de Bacharel em Letras-Tradução.

Orientador: Profa. Dra. Patrícia Fabiane Amaral da Cunha Lacerda

Juiz de Fora
2023

Ana Cláudia Duarte Corrêa

**CONTRIBUIÇÕES DA LINGUÍSTICA DE *CORPUS*
PARA OS ESTUDOS DA TRADUÇÃO:
um estudo de caso a partir da Tradução Técnica**

Trabalho de Conclusão de Curso
apresentado ao Departamento de Letras
Estrangeiras Modernas da Universidade
Federal de Juiz de Fora como requisito
parcial à obtenção do grau de Bacharel
em Letras-Tradução.

Aprovada em 26 de junho de 2023

BANCA EXAMINADORA

Profa. Dra. Patrícia Fabiane Amaral da Cunha Lacerda - Orientadora
Universidade Federal de Juiz de Fora

Profa. Dra. Mayra Barbosa Guedes
Universidade Federal de Juiz de Fora

Profa. Dra. Sandra Aparecida Faria de Almeida
Universidade Federal de Juiz de Fora

Dedico este trabalho ao meu marido e ao meu filho que sempre me inspiram e me apoiam.

AGRADECIMENTOS

Agradeço a oportunidade de conclusão desta etapa a Deus, a toda a minha família e, em especial, ao meu filho e ao marido pela paciência e pelo apoio.

Agradeço também à minha orientadora, Profa. Pós-Dra. Patrícia Fabiane Amaral da Cunha Lacerda, pela dedicação, pelo apoio, pela compreensão e por todo conhecimento compartilhado.

Agradeço à Profa. Dra. Mayra Barbosa Guedes e à Profa. Dra. Sandra Aparecida Faria de Almeida por aceitarem compor minha banca, pela disponibilidade para a leitura e pelas contribuições a este trabalho.

Agradeço a todo o corpo docente e da Faculdade de Letras pelos ensinamentos durante as aulas.

Agradeço a todos os colaboradores da Universidade Federal de Juiz de Fora pela disponibilidade.

E, enfim, agradeço à Universidade Federal de Juiz de Fora por mais uma oportunidade de adquirir conhecimento e me proporcionar novas oportunidades.

A todos, dedico este trabalho.

RESUMO

Este trabalho tem como objetivo fundamental propor uma interface entre os Estudos da Tradução e a Linguística de *Corpus* por meio de um estudo de caso em particular, pautado na Tradução Técnica e, de modo mais específico, na tradução de resumos acadêmicos da língua portuguesa para a língua inglesa. Nesse sentido, assumindo como objeto de análise resumos e suas respectivas traduções publicados na Revista Veredas – revista publicada periodicamente pelo Programa de Pós-graduação em Linguística da Universidade Federal de Juiz de Fora – nos últimos cinco anos, busca-se demonstrar como a Linguística de *Corpus* pode contribuir para a realização da tradução técnica a partir da utilização do programa *Antconc*. Dessa forma, no que se refere à Linguística de *Corpus*, em conformidade com as proposições teóricas de Cunha Lacerda e Dall'orto (no prelo), Baker (1995), Barreiros (2017), Davies (2015), Laviosa (2002), Novodvorski e Finatto (2014), Salgado, (2013), Sardinha (2000), Tagnin (2003, 2015), Tognini-Bonelli (2001, 2010), entende-se que o trabalho com terminologia, no processo de tradução de textos técnicos, pode possibilitar ao tradutor maior acesso a diferentes conjuntos de textos para adequar ainda mais o texto traduzido, visto que ele será capaz de compreender melhor o funcionamento dos termos e fraseologias no âmbito da Tradução Técnica (AIXELÁ, 2016; ALMEIDA; SIMÕES 2002; AUBERT, 2001; BERNARDINO; VALENTIM, 2016; BEVILACQUA; KILIAN, 2017; BRAGAGNOLLO, 2011; DE CAVACO-CRUZ, 2012; KRIEGGER; SANTIAGO, 2014; RIBEIRO, 2004; TAVARES, 2022; TELES, 2007). Para tal, apresenta-se uma análise pautada no método misto, a partir da coadunação entre a metodologia qualitativa e a metodologia quantitativa, e calcada na aplicação das ferramentas *Wordlist*, *Concordance* e *File View*, que constituem o programa *Antconc*. A partir da realização do trabalho, conclui-se que, de fato, a Linguística de *Corpus* traz contribuições substanciais para os Estudos da Tradução – especialmente no que tange à tradução de resumos acadêmicos –, fomentando uma prática tradutória extremamente reflexiva e fundamentada.

Palavras-chave: Tradução. Tradução Técnica. Tradução de resumos. Linguística de *Corpus*. Programa *Antconc*.

ABSTRACT

The fundamental objective of this work is to propose an interface between Translation Studies and Corpus Linguistics through a particular case study, based on Technical Translation and, more specifically, on the translation of academic abstracts from Portuguese to English. In this sense, assuming as object of analysis abstracts and their respective translations published in *Revista Veredas* – journal that is periodically published by the Graduate Program in Linguistics at the Federal University of Juiz de Fora – in the last five years, this work seeks to demonstrate how Corpus Linguistics can contribute to the achievement of technical translation using a program called *Antconc*. Thus, with regard to Corpus Linguistics, in accordance with the theoretical propositions by Cunha Lacerda and Dall'orto (in press), Baker (1995), Barreiros (2017), Davies (2015), Laviosa (2002), Novodvorski and Finatto (2014), Salgado, (2013), Sardinha (2000), Tagnin (2003, 2015), Tognini-Bonelli (2001, 2010), it is understood that working with terminology, in the process of translation of technical texts, can allow the translator to have a greater access to different sets of texts to further adapt the translated text, since the translator will be able to better understand the functioning of terms and phraseologies within the scope of Technical Translation (AIXELÁ, 2016; ALMEIDA; SIMÕES 2002; AUBERT, 2001; BERNARDINO; VALENTIM, 2016; BEVILACQUA; KILIAN, 2017; BRAGAGNOLLO, 2011; DE CAVACO-CRUZ, 2012; KRIEGGER; SANTIAGO, 2014; RIBEIRO, 2004; TAVARES, 2022; TELES, 2007). To this end, an analysis based on the mixed method is presented, based on the combination between the qualitative research methodology and the quantitative research methodology, and based on the application of the Wordlist tool, the Concordance tool and the File View tool, which constitute the Antconc program. From the completion of the work, it is concluded that, in fact, Corpus Linguistics brings substantial contributions to Translation Studies – especially with regard to the translation of academic abstracts –, fostering an extremely reflective and reasoned translation practice.

Keywords: Translation. Technical translation. Abstract translation. Corpus Linguistics. Antconc program.

LISTA DE FIGURAS

Figura 1 – Tela inicial do programa Antconc 3.5.9.....	34
Figura 2 – Ferramenta Wordlist distribuída pelo critério Sort by Freq.....	35
Figura 3 – Ferramenta Wordlist distribuída pelo critério Sort by Word.....	35
Figura 4 – Ferramenta Wordlist distribuída pelo critério Sort by Word End.....	36
Figura 5 – Tela representativa da ferramenta Concordance.....	37
Figura 6 – Tela representativa da ferramenta File View.....	37
Figura 7 – Tela da ferramenta Concordance Plot.....	38
Figura 8 – Tela da ferramenta Clusters/N-Grams.....	39
Figura 9 – Tela da ferramenta Collocates.....	40
Figura 10 – Tela da ferramenta Keywords.....	41
Figura 11 – Resultados dos dez primeiros nomes nos corpora distribuídos pelo critério Sort by Freq.....	61
Figura 12 – Contextos relacionados aos termos “ensino” e “teaching” por meio da ferramenta Concordance.....	63
Figura 13 – Nomes dentro do texto bruto dos arquivos representados através da ferramenta File View.....	72

LISTA DE QUADROS

Quadro 1 - Principais corpora compilados ou em compilação.....	24
Quadro 2 - Comparação dos termos mais frequentes do português para o inglês... 67	
Quadro 3 - Comparação dos termos mais frequentes do inglês para o português... 70	

LISTA DE TABELAS

Tabela 1 - Os dez nomes mais frequentes no corpus em português..... 60

Tabela 2 - Os dez nomes mais frequentes no corpus em inglês..... 60

SUMÁRIO

INTRODUÇÃO.....	11
CAPÍTULO I - LINGUÍSTICA DE CORPUS: UMA VISÃO GERAL.....	14
1.1. Linguística de Corpus: conceitos fundamentais.....	15
1.2. Linguística de Corpus e tradução.....	29
1.3. O programa Antconc e sua contribuição para os Estudos da Tradução.....	32
1.4. Conclusões.....	42
CAPÍTULO II - TRADUÇÃO TÉCNICA: PRESSUPOSTOS FUNDAMENTAIS.....	43
2.1. Tradução técnica: uma revisão.....	43
2.2. Terminologia e tradução.....	46
2.3. O gênero resumo / abstract.....	50
2.4. Conclusões.....	54
CAPÍTULO III - RESUMO X ABSTRACT: O PAPEL DA LINGUÍSTICA DE CORPUS NA TRADUÇÃO TÉCNICA.....	56
3.1. Procedimentos metodológicos.....	56
3.2. Análise.....	59
3.2.1. Aplicação da ferramenta Wordlist.....	59
3.2.2. Aplicação da ferramenta Concordance.....	62
3.2.3. Aplicação da ferramenta File View.....	71
3.3. Conclusões.....	73
CONSIDERAÇÕES FINAIS.....	75
REFERÊNCIAS.....	77

INTRODUÇÃO

Na atualidade, percebe-se que, nos últimos anos, os estudos da Linguística de *Corpus* têm crescido e exercido bastante influência nas pesquisas na área da tradução. Porém, observa-se que ainda é preciso maior aprofundamento e mais pesquisas sobre o assunto, visto que tais estudos demonstram ser ainda incipientes a respeito disso. Portanto, se fazem necessários trabalhos que busquem, cada vez mais, ratificar a produtividade da interface entre a Linguística de *Corpus* e a tradução.

Nesse sentido, este trabalho busca apresentar uma visão geral da Linguística de *Corpus*, demonstrando a relação presente entre ela e a tradução, ou seja, discutindo a viabilidade da aplicação da Linguística de *Corpus* aos Estudos da Tradução no que tange especialmente à Tradução Técnica. Nesse contexto, a fim de estabelecer uma relação entre a Linguística de *Corpus* (CUNHA LACERDA; DALL'ORTO, no prelo; BAKER, 1995; BARREIROS, 2017; DAVIES, 2015; LAVIOSA 2002; NOVODVORSKI; FINATTO, 2014; SALGADO, 2013; SARDINHA, 2000; TAGNIN, 2003, 2015; TOGNINI-BONELLI, 2001, 2010) e a Tradução Técnica (AIXELÁ, 2016; ALMEIDA; SIMÕES, 2002; AUBERT, 2001; BERNARDINO; VALENTIM, 2016; BEVILACQUA; KILIAN, 2017; BRAGAGNOLLO, 2011; DE CAVACO-CRUZ, 2012; KRIEGGER; SANTIAGO, 2014; RIBEIRO, 2004; TAVARES, 2022; TELES, 2007), assumindo como objeto de análise, de modo particular, a tradução do gênero resumo da língua portuguesa para a língua inglesa. A fim de cumprir o objetivo geral deste trabalho, analisamos como a Linguística de *Corpus* - no que se refere especificamente à aplicação do programa *Antconc*¹- pode contribuir, de modo substancial, para realização e para a compreensão da tradução técnica que se realiza no gênero textual resumo acadêmico.

À vista disso, entende-se que, no momento atual, é necessário trabalhar com as terminologias para traduzir textos técnicos, considerando que o desenvolvimento das atuais tecnologias possibilita ao tradutor maior acesso a diferentes conjuntos de textos, com diferentes gêneros e temáticas, podendo, assim, fazer uso deles para suas necessidades informativas. Em outras palavras, entende-se que a terminologia permite que o tradutor adéque ainda mais o texto ao ponto de vista linguístico e

¹ O programa *Antconc* será devidamente apresentado na seção 1.3.

especializado, visto que este passa a conhecer melhor o funcionamento dos termos e das fraseologias especializadas nos contextos inseridos.

A fim de cumprir o objetivo geral deste trabalho, assume-se como *corpus* de análise os resumos acadêmicos presentes nos últimos cinco anos da Revista Veredas, que constitui um periódico regularmente publicado pelo Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora. Nesse caso, foram analisados os resumos em língua portuguesa – que compreende a língua de partida – e suas respectivas traduções para a língua inglesa – que compreende a língua de chegada. Do ponto de vista metodológico, a análise será calcada na aplicação das ferramentas *Wordlist*, *Concordance* e *File View* que constituem o programa *Antconc*.

À vista disso, este texto está organizado nos seguintes capítulos: no primeiro capítulo, será apresentada uma visão geral da Linguística de *Corpus* e de seus conceitos fundamentais; já no segundo capítulo, serão discutidos os pressupostos fundamentais que constituem a tradução técnica e será também discutido o gênero resumo/*abstract* de modo particular; por fim, no terceiro capítulo, serão destacados os procedimentos metodológicos e será apresentada a análise propriamente dita.

CAPÍTULO I

LINGUÍSTICA DE *CORPUS*: UMA VISÃO GERAL

Neste capítulo, será apresentada uma visão geral da Linguística de *Corpus*, com enfoque no papel que ela desempenha nas pesquisas relacionadas à tradução. Também serão apresentadas as finalidades das ferramentas que constituem o programa *Antconc*, visando a destacar suas contribuições para as pesquisas no âmbito dos Estudos da Tradução.

A fim de cumprir os objetivos propostos, na primeira seção, o capítulo tratará dos conceitos fundamentais da Linguística de *Corpus*, pautando-se nas considerações de Sardinha (2000) no que diz respeito à definição da área. Para responder à problematização de a Linguística de *Corpus* ser uma teoria ou uma metodologia, este capítulo, portanto, se fundamentará nas considerações de Sardinha (2000), Novodvorski e Finatto (2014), Aluísio e Almeida (2006) e Tognini-Bonelli (2010). Para apresentar o entendimento do que seja *corpus*, este trabalho se pautará nas considerações de Sardinha (2000), Silva e Silva (2013), Baker (2010), Novodvorski e Finatto (2014) e Tognini-Bonelli (2001). Já quanto à definição do que se entende por *corpus* e sua tipologia, serão apresentadas como base as afirmações de Davies (2015), Tognini-Bonelli (2010), Baker (1995) e Sardinha (2000). Já no que concerne à convencionalidade, serão consideradas ponderações de Tagnin (2003). E com o propósito de propor um encaminhamento referente a ferramentas para o tratamento de *corpora*, a primeira seção deste capítulo ainda se pautará nas considerações de Baker (1995) e Laviosa (2002).

Em um segundo momento, para tratar das contribuições da Linguística de *Corpus* para a tradução, a segunda seção deste capítulo se pautará nas considerações de Tagnin (2015).

Em um terceiro momento, com a finalidade de apresentar o programa *Antconc* e suas contribuições para os Estudos da Tradução, serão consideradas as afirmações de Cunha Lacerda e Dall'Orto (no prelo) e Salgado (2013).

Por fim, na última seção, serão encaminhadas algumas conclusões a que se chegou a partir das discussões realizadas durante o capítulo.

1.1. Linguística de *Corpus*: conceitos fundamentais

Nos últimos anos, os estudos a respeito da Linguística de *Corpus* têm crescido consideravelmente. De acordo com Sardinha (2000), na atualidade, ela exerce grande influência na pesquisa linguística. No entanto, no Brasil, ela ainda é incipiente. Segundo o autor, a Linguística de *Corpus* vem ganhando espaço além dos centros acadêmicos, sendo utilizada também em aplicações comerciais. Nesse sentido, Sardinha (2000) a define da seguinte forma:

A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram 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 meio de computador (SARDINHA, 2000, p. 325).

Ainda de acordo com o autor, a Linguística de *Corpus* apresenta uma abordagem empirista, uma vez que se pauta no estudo da língua a partir de dados reais:

O empirismo é, em termos bem simples, uma doutrina filosófica segundo a qual o conhecimento se origina da experiência. Na lingüística, o empirismo significa dar primazia aos dados provenientes da observação da linguagem, em geral reunidos sob a forma de um corpus. O empirismo se coloca em oposição ao racionalismo, segundo o qual, em linhas gerais, o conhecimento provém de princípios, estabelecidos *a priori*. (SARDINHA, 2000, pp. 349-350)

A Linguística de *Corpus* também expressa uma visão da língua enquanto sistema probabilístico, que, segundo Sardinha (2000, p. 350), “pressupõe que, embora muitos traços linguísticos sejam possíveis teoricamente, eles não ocorrem com a mesma frequência”.

O autor ainda afirma que as possibilidades da estrutura não se realizam todas na mesma frequência. Para ele, o mais importante acerca da diferença de frequências entre os traços é o fato de essas diferenças não serem aleatórias, pois

há uma correlação entre as características linguísticas e os contextos de uso. Sob essa perspectiva, Sardinha (2000) afirma que:

[...] quando se diz que a variação não é aleatória, na verdade, está se afirmando que a linguagem é padronizada ('patterned'). A padronização se evidencia pela recorrência, isto é, uma colocação, coligação ou estrutura, que se repete significativamente, mostra sinais de ser na verdade um padrão lexical ou léxico-gramatical. A linguagem forma padrões que apresentam regularidade (se mostram estáveis em momentos distintos, isto é, tem freqüência comparável em corpora distintos) e variação sistemática (correlacionam-se com variedades textuais, genéricas, dialetais, etc) (SARDINHA, 2000, p. 351).

Por essa razão, mais do que a “possibilidade teórica de ocorrência, uma teoria da linguagem deve incorporar a probabilidade de ocorrência dos traços”. Recorre-se, então, à introspecção ou à intuição do falante nativo para responder a essa questão. No entanto, o que ele “pode informar é somente se o traço ou a estrutura em questão é intuitivamente provável ou não”, visto que o ser humano não é capaz de isolar conscientemente de forma bem organizada o que é central e típico da linguagem (SARDINHA, 2000, p. 351-352).

O autor afirma, nesse sentido, que é necessária uma observação empírica da frequência do emprego de determinado elemento linguístico em contextos definidos, realizados por diferentes usuários, para saber a probabilidade de um traço ou de uma estrutura. Dessa forma, ele apresenta duas conclusões:

A primeira é a importância primordial de um corpus como fonte de informação, pois ele registra a linguagem natural realmente utilizada por falantes e escritores da língua em situações reais. A segunda é a não-trivialidade da investigação da frequência de ocorrência de traços linguísticos de várias ordens (lexicais, sintáticos, semânticos, discursivos, etc), pois é através do conhecimento da frequência atestada que se pode estimar a probabilidade teórica. (SARDINHA, 2000, p. 352)

Sardinha (2000) ainda busca responder à seguinte questão: a Linguística de *Corpus* é uma disciplina ou uma metodologia? Ele esclarece que, para ele, a Linguística de *Corpus* não é uma disciplina, pois seu objeto de pesquisa não é delimitado como em outras áreas. De acordo com o autor, a Linguística de *Corpus* se dedica a vários fenômenos comumente enfocados em outras áreas. Assim sendo, ele conclui que, para se tratar de uma metodologia ou não, a definição da

Linguística de *Corpus* depende da definição de metodologia usada. Se entendida como *instrumental*, “é possível aplicar-se o instrumental da Linguística de *Corpus* livremente e manter a orientação teórica da disciplina original”. Por outro lado, se entendida como “*modo típico de aplicar um conjunto de pressupostos de caráter teórico*”, passa a ser vista como uma metodologia, visto que traz “algo mais do que simplesmente o instrumental computacional” (SARDINHA, 2000, p.356).

A esse respeito, Novodvorski e Finatto (2014) vão ao encontro de tais considerações e reafirmam que a Linguística de *Corpus* se coloca como uma nova perspectiva para a Linguística, mas não como um novo tipo de Linguística. Segundo os autores, ela se mostra

[...] tanto como uma metodologia quanto como uma abordagem teórica diferenciada dos Estudos da Linguagem. De quem queira se aproximar da LC, apenas por se interessar por seu instrumental ou por seus procedimentos, nada será cobrado em termos de uma filiação teórica – ou epistemológica – ainda que insistamos que LC também é um modo de compreender a língua, que temos nosso modo de defini-la como objeto de estudo: a língua é um sistema probabilístico de combinatórias, no qual uma unidade se define pelas associações que mantém com outras unidades (NOVODVORSKI e FINATTO, 2014, pp. 7-8).

Alúcio e Almeida (2006) também apresentam a Linguística de *Corpus* como uma abordagem. Segundo eles, com o desenvolvimento amplo das pesquisas baseadas em *corpus*, nota-se uma relevância e pertinência nos domínios da Linguística, da Linguística Aplicada e da Linguística Computacional, fazendo surgir uma abordagem para sistematizar procedimentos e dar conta desse novo modo de pesquisar. A essa abordagem eles chamam de Linguística de *Corpus*. Os autores afirmam que tal abordagem é auxiliada pelo desenvolvimento de ferramentas computacionais.

Sardinha (2000) ainda argumenta que o fato de os praticantes produzirem conhecimento novo, como os de caráter contestatório de práticas e preceitos correntes, é uma das razões para não definir a Linguística de *Corpus* como uma metodologia. Assim, o autor apresenta uma terceira possibilidade: a de que a Linguística de *Corpus* não é nem disciplina, nem metodologia, mas sim uma perspectiva, uma maneira de se chegar à língua, ou seja, ela não é apenas um

instrumental, mas sim uma abordagem. Dessa forma, Sardinha (2000) entende-a como uma abordagem baseada em *corpus*.

Tognini-Bonelli (2010) afirma, nesse sentido, que muitos linguistas que trabalham com um *corpus* tendem a concordar que a Linguística de *Corpus* vai muito além de um papel puramente metodológico. Ela diz que os linguistas apontam para uma conexão entre o uso de métodos computacionais, algorítmicos e estatísticos. Ainda, segundo a autora, o que está sendo testemunhando é o fato de que a Linguística de *Corpus* “se tornou um novo empreendimento de pesquisa e uma nova abordagem filosófica da investigação linguística²” (TOGNINI-BONELLI, 2001, p. 1, tradução nossa).

Sob essa perspectiva, Sardinha (2000), ao apresentar definições para o termo *corpus*, chega à conclusão de que, para uma definição completa, é preciso incorporar pontos importantes, tais como: i) a *origem* dos dados, pois esses devem ser autênticos; ii) o *propósito*, visto que sua finalidade deve ser objeto de estudo linguístico; iii) a *composição*, por considerar que seu conteúdo deve ser escolhido de forma criteriosa; iv) a *formatação*, dado que o *corpus* precisa ser inteligível por computador; v) a *representatividade*, em razão da necessidade de o *corpus* ser representativo de uma língua ou variedade; e vi) a *extensão*, pois o *corpus* deve ser vasto para que seja representativo. Partindo disso, o autor resume quatro pré-requisitos para a formação de um *corpus* computadorizado, a saber:

(1) Primeiramente, o corpus deve ser composto de textos autênticos, em linguagem natural. Assim, os textos não podem ter sido produzidos com o propósito de serem alvo de pesquisa linguística. E não podem ter sido criados em linguagem artificial, tais como linguagem de programação de computadores ou notação matemática.

(2) Em segundo lugar, quando se fala em autenticidade dos textos, subentende-se textos escritos por falantes nativos. Tanto assim que, quando este não é o caso, deve-se qualificá-lo, falando-se em corpora ‘de aprendizes’ (‘learner corpora’).

(3) O terceiro pré-requisito é que o conteúdo do corpus seja escolhido criteriosamente. Os princípios da escolha dos textos devem seguir, acima de tudo, as condições de naturalidade e autenticidade. Mas devem também obedecer a um conjunto de regras estabelecidas pelos seus criadores de modo que o corpus coletado corresponda às características que se deseja dele. Ou seja, o conteúdo do corpus deve ser selecionado a fim de garantir que o

² Cf.: “What we are witnessing is the fact that corpus linguistics has become a new research enterprise and a new philosophical approach to linguistic enquiry.” (TOGNINI-BONELLI, 2001, p. 1).

corpus tenha uma certa característica. Por exemplo, se o desejo é construir um corpus de português brasileiro escrito que represente a língua portuguesa, tal qual ela é escrita no Brasil, em sua totalidade, a coleta deve ser guiada por um conjunto de critérios que garanta, entre outras coisas, que o maior número possível de tipos textuais existentes no português brasileiro esteja representado, que haja uma quantidade aceitável de cada tipo de texto e que a seleção dos textos seja aleatória, a fim de que não se contamine a coleta com variáveis indesejáveis.

(4) O quarto pré-requisito é mais problemático: representatividade. Tradicionalmente, tende-se a ver um corpus como um conjunto representativo de uma variedade linguística ou mesmo de um idioma. Mas a questão não pode ser enfocada no vácuo. Cabe se perguntar 'representativo do quê?' e 'representativo para quem?'. A questão da representatividade é discutida abaixo com mais detalhes (SARDINHA, 2000, p. 338).

Segundo Silva e Silva (2013), o conceito de *corpus* tem sido definido por múltiplos posicionamentos. Entre as semelhanças e diferenças indicadas por diferentes autores, o elemento em comum, o *corpus*, é apontado como uma coleção de materiais, sendo conceituado por alguns como "corpo", o que reforça a ideia de elementos em sua constituição. No entanto, os autores destacam que, no âmbito da Linguística moderna, o termo apresenta conotações específicas.

Baker (2010), nesse sentido, destaca que a origem da palavra *corpus* é latina, tendo o termo *corpora* como seu plural. Segundo ele, um *corpus* é, portanto, um "corpo" de linguagem, ou mais especificamente, uma coleção muito grande de linguagem natural, armazenada como arquivos de computador. Além disso, ele afirma que os *corpora* geralmente são cuidadosamente construídos para que possam ser considerados representativos de uma determinada língua ou variedade linguística reafirmando o que foi dito por Sardinha (2000).

Tognini-Bonelli (2001) também vai ao encontro da reflexão de Sardinha (2000) ao afirmar que um *corpus* constitui uma coleção de textos considerados representativos de um determinado idioma, dialeto ou outro subconjunto de idioma a ser usado na análise linguística. Segundo ela, alguns autores consideram o *corpus* como uma coleção de textos de ocorrência natural escolhido para caracterizar um estado ou uma variedade de um idioma, um subconjunto específico construído de acordo com um projeto e propósito específicos, uma coleção de amostras de texto que podem ser falados ou escritos, configurando uma coleção de pedaços de

linguagem selecionados e ordenados de acordo com critérios linguísticos para serem usados como uma amostra linguística. Ainda de acordo com Tognini-Bonelli (2001), muitos pesquisadores concordam que um *corpus* compreende uma coleção de textos linguísticos, embora não sejam necessariamente textos. Assim, alguns estudiosos consideram que se trata de amostras ou pedaços de linguagem, outros, por sua vez, consideram que se trata de uma coleção de textos inteiros. Segundo Tognini-Bonelli (2001), enquanto para muitos um *corpus* deve parecer um conjunto de textos armazenados em um computador, há um consenso de que um *corpus* não apenas fornecerá *insights* sobre seu conteúdo, mas também que os resultados dessas investigações serão reivindicados ou assumidos como típicos da linguagem da qual ele foi selecionado.

De acordo com Novodvorski e Finatto (2014), mesmo que não apresente respostas para todas as indagações, todo *corpus* sempre traz questões novas ou que não se imaginava encontrar, o que torna a sistematização dos dados e as observações cruciais.

Baker (1995) também apresenta uma definição para o termo *corpus*. Segundo a autora,

[...] a palavra *corpus* originalmente significava qualquer coleção de escritos, de forma processada ou não processada, geralmente de um autor específico. Nos últimos anos, e com o crescimento da linguística de corpus, essa definição mudou de três maneiras importantes: (i) *corpus* agora significa principalmente uma coleção de textos mantidos em forma legível por máquina e capazes de serem analisados automática ou semi-automaticamente em uma variedade de maneiras; (ii) um *corpus* não está mais restrito a 'escritos', mas inclui tanto texto falado quanto escrito, e (iii) um *corpus* pode incluir um grande número de textos de uma variedade de fontes, por muitos escritores e falantes e em uma infinidade de tópicos. O importante é que ele seja elaborado para um propósito específico e de acordo com critérios de design explícitos, a fim de garantir que seja representativo da área ou amostra de linguagem que pretende dar conta (BAKER, 1995, p. 224, tradução nossa)³.

³ Cf.: "The word corpus originally meant any collection of writings, in a processed or unprocessed form, usually by a specific author. In recent years, and with the growth of corpus linguistics, this definition has changed in three important ways: (i) corpus now means primarily a collection of texts held in machine-readable form and capable of being analyzed automatically or semi-automatically in a variety of ways; ii) a corpus is no longer restricted to 'writings' but includes spoken as well as written text, and (iii) a corpus may include a large number of texts from a variety of sources, by many writers and speakers and on a multitude of topics. What is important is that it is put together for a particular purpose and according to explicit design criteria in order to ensure that it is representative of the given area or sample of language it aims to account for." (BAKER, 1995, p. 224)

Além de buscar dimensionar uma definição para o termo *corpus* no âmbito da Linguística de *Corpus*, outra questão se torna extremamente relevante: quais são as diferentes tipologias a partir das quais um *corpus* pode se configurar e se apresentar?

A esse respeito, Davies (2015, p. 11) considera que os *corpora* podem se apresentar a partir da seguinte tipologia:

1. *Corpora* pequenos de 1 a 5 milhões de palavras, de primeira geração, como o Brown Corpus (e outros da “família” Brown, como o LOB, Frown e FLOB);
2. *Corpora* de tamanho moderado, de segunda geração e com equilíbrio de gênero, como o British National Corpus de 100 milhões de palavras;
3. *Corpora* maiores e mais atualizados (mas ainda equilibrados por gênero), como o Corpus de 450 milhões de palavras do inglês americano contemporâneo (COCA);
4. Grandes arquivos de texto, como Lexis-Nexis;
5. Arquivos de texto extremamente grandes, como o Google Livros;
6. A Web como *corpus*, vista aqui pelas lentes do Google pesquisas⁴. (DAVIES, 2015, p. 11, tradução nossa).

Tognini-Bonelli (2010) considera alguns tipos de *corpora* mais comuns. Segundo a autora, existem tantos *corpora* para tantos propósitos que fica difícil listá-los. Assim, apenas uma classificação superficial pode ser tentada. Nesse sentido, ela primeiramente apresenta os seguintes *corpora*: i) os *corpora de amostra*, que partem do princípio de que a maioria dos *corpora* são ‘instantâneos’ no tempo e, como tal, são amostras de uma determinada linguagem em um determinado momento, tendo como finalidade apresentar as características linguísticas normais de uma língua ou variedade em aproximadamente as proporções encontradas no uso geral; ii) os *corpora para comparação*, que

⁴ Cf.: “1. Small 1–5-million-word, first-generation corpora like the Brown Corpus (and others in the so-called Brown family, such as LOB, Frown, and FLOB).

2. Moderately sized, second-generation, genre-balanced corpora, such as the 100-million-word British National Corpus (BNC).

3. Larger, more up-to-date (but still genre-balanced) corpora, such as the 450-million-word Corpus of Contemporary American English (COCA).

4. Large text archives, such as Lexis-Nexis.

5. Extremely large text archives, such as Google Books.

6. The Web as Corpus, seen here through the lens of Google-based searches” (DAVIES, 2015, p. 11).

envolvem dois ou mais *corpora* comparáveis quando são construídos com os mesmos critérios e tamanhos; iii) os *corpora geográficos*, que parte do princípio de uma exploração das variedades mundiais de inglês por meio da construção de *corpora* de variedades locais em muitas regiões, usando um *design* baseado nos acervos da pesquisa; iv) os *corpora históricos*, que são projetados para serem comparados ao longo de uma dimensão temporal; v) os *corpora temáticos*, que envolvem textos coletados de acordo com os temas; vi) o *corpus contrastivo*, que é um *corpus* único cujos principais componentes foram escolhidos para facilitar o estudo da variedade analisada. Enquanto os *corpora* de amostra são projetados principalmente para que as várias diferenças entre as variedades se misturem em uma imagem geral dos padrões de uso, um *corpus* contrastivo é projetado para destacar o padrão característico de variedades distinguíveis. Trata-se, portanto, de um *corpus* construído com o propósito específico de estabelecer contrastes dentro de si mesmo (TOGNINI-BONELLI, 2010). Dessa forma, segundo a autora, supõe-se que tais *corpora* constituem um repositório de evidências sobre a língua ou variedade em seu estado natural.

Já de acordo com Baker (1995), os *corpora* são geralmente concebidos com base em vários critérios de seleção., dentre quais os mais importantes são:

- (i) linguagem geral versus domínio restrito
- (ii) linguagem escrita vs. falada
- (iii) sincrônico vs. diacrônico
- (iv) tipicidade em termos de variedade de fontes (escritores/falantes) e gêneros
(por exemplo, editoriais de jornais, entrevistas de rádio, ficção, artigos de jornais, audiências judiciais)
- (v) limites geográficos, por exemplo. Inglês britânico x americano
- (vi) monolíngue vs. bilíngue ou multilíngue⁵ (BAKER, 1995, p. 229, tradução nossa).

⁵ Cf.: "(i) general language vs. restricted domain
(ii) written vs. spoken language
(iii) synchronic vs. diachronic
(iv) typicality in terms of range of sources (writers/speakers) and genres
(e.g. newspaper editorials, radio interviews, fiction, journal articles, court hearings)
(v) geographical limits, e.g. British vs. American English
(vi) monolingual vs. bilingual or multilingual" (BAKER, 1995, p. 229)

Atualmente, existem vários *corpora* disponíveis na internet, que podem ser acessados facilmente. O quadro a seguir, baseado em Sardinha (2000), mostra os principais *corpora* compilados ou em compilação:

Corpus	Lançamento / Referência na literatura	Palavras	Composição
Brown Corpus (Brown University Standard Corpus of Present-Day American English)	1964	1 milhão	Inglês americano, escrito
AHI(American Heritage Intermediate Corpus)	1971	5 milhões	Inglês americano, escrito
LOB (Lancaster-Oslo-Bergen)	1978	1 milhão	inglês britânico, escrito
LLC (London-Land Corpus)	1980	500 mil	inglês britânico, falado
Birmingham Corpus (Birmingham University International Language Database)	1987	20 milhões	Inglês britânico
Kolhapur Corpus (of Indian English)	1988	1 milhão	inglês indiano, escrito
TOSCA Corpus (Tools for Syntactic Corpus Analysis)	1988	1,5 milhão	Inglês britânico, escrito
SEU Corpus(Survey of English Usage)	1989	1 milhão	Inglês britânico, escrito e falado
CHILDES (Child Language Data Exchange)	1990	20 milhões	Inglês infantil, falado
Nijmegen Corpus	1991	132 mil	Inglês britânico, escrito e falado
LLELC (Longman-Lancaster English Language Corpus)	1991*	50 milhões*	inglês de vários tipos, escrito e falado
Map Task Corpus	1991	147 mil	Inglês escocês, falado
LCLE (Longman Corpus of Learner's English)	1992	10 milhões	Inglês escrito por estrangeiros
SEC (Lancaster/IBM Spoken English Corpus)	1992	53 mil	Inglês britânico, falado
Wellington Corpus (of Written New Zealand English)	1993	1 milhão	Inglês neozelandês, escrito
POW (Polytechnic of Wales	1993	65 mil	Inglês infantil, falado

Corpus)			
Wellington Corpus of Spoken New Zealand English	1995	1 milhão	Inglês neozelandês, falado
BNC (British National Corpus)	1995	100 milhões	Inglês britânico, escrito e falado
Corpus of Spoken American English	1991	2 milhões	Inglês americano, falado
ICLE (International Corpus of Learner English)	1997	200 mil**	Inglês escrito por estrangeiros
Bank of English	1997	320 milhões	Inglês britânico
* previsão ** cada variedade nacional			

Quadro 1 - Principais *corpora* compilados ou em compilação
Fonte: Sardinha (2000)

Apesar de poderem ser facilmente acessados, os diferentes *corpora* disponíveis na internet devem ser utilizados de forma consciente. A esse respeito, Tagnin (2003) cita o termo "falante ingênuo", estabelecido por Fillmore (1979) ao se referir a um indivíduo que desconhece as convenções da língua. Segundo ela:

Um falante ingênuo desconhece, entre outras coisas, a ordem preferencial de binômios como cats and dogs, bed and breakfast, knife and fork, cujos correspondentes em português seriam cães e gatos, cama e mesa, garfo e faca. Também não sabe que há determinadas combinações fixas ou semi-fixas conhecidas como colocações (TAGNIN, 2003, p. 192).

Segundo a autora, o falante ingênuo não tem consciência das expressões, das partes pré-fabricadas, das fraseologias e das unidades da língua que não são geradas todas as vezes que são empregadas. Além disso, qualquer pessoa pode ser um falante ingênuo de sua própria língua a depender da situação. Porém, é no confronto das línguas que isso fica mais evidente. De acordo com autora, temos o seguinte:

Basicamente, a ingenuidade do tradutor se configura numa compreensão composicional do significado e numa falta de consciência do quanto uma língua é constituída dessas partes pré-

fabricadas. A ingenuidade do tradutor pode transparecer tanto na sua habilidade de compreensão, quanto na de produção. Em termos de compreensão, ele pode não entender expressões idiomáticas como a *hard nut to crack*, *put one's best foot forward*, ou *cut corners*, por serem não-composicionais, ou seja, uma expressão cujo significado total não corresponde à soma dos significados individuais de seus componentes. Ele pode não compreender muitas fórmulas discursivas por não conhecer as convenções sociais que determinam seu uso na língua alvo. Pode também não compreender referências humorísticas que resultem da manipulação das categorias convencionais da língua (TAGNIN, 2003, p. 193).

Nesse sentido, de acordo com a autora, no que concerne à convencionalidade, seria muito ingênuo se um tradutor acreditasse que o dicionário poderia resolver todos os problemas. Ainda conforme Tagnin (2003), o tradutor considerado profissional, que traduz textos, deve buscar colocações e caminhos para encontrar uma tradução mais adequada. Além disso, ele deve saber observar, nos dicionários, colocações registradas que já podem ter caído em desuso. Ademais, ao fazer uso de *corpora*, ele demonstrará mais destreza, visto que este não apresentará apenas a forma correta, mas também a mais usual da língua.

Nesse sentido, Tagnin (2003) também ressalta a importância da consciência quanto à relevância de *corpora* técnicos para o trabalho do tradutor. Além disso, destaca a relevância do uso de *corpora* na tradução, visto que estes possibilitam a busca pelo termo mais adequado dentro de determinado contexto de uso para que a tradução seja feita de forma natural e fluente.

Novodvorski e Finatto (2014) também afirmam que o uso de *corpus* e de *corpora* tem se expandido atualmente para além da menção a muitas ferramentas e aos princípios da Linguística de *Corpus*, chegando a alcançar áreas inimagináveis e, até mesmo, incompatíveis em um primeiro momento. Segundo os autores,

[...] a alusão às terminologias típicas de LC (como *types*, *tokens* e *concordâncias*) vem se tornando cada vez mais recorrente. Em eventos científicos, em publicações, em nomes de disciplinas, teses e dissertações, a recorrência com que aparecem referências ou vestígios da LC denotam já uma presença marcada no plano acadêmico e servem como um bom termômetro do estado da arte (NOVODVORSKI; FINATTO, 2014, pp. 8-9).

A esse respeito, segundo Baker (2010), há um potencial de uso de *corpora* começando a se concretizar nos Estudos de Tradução. Os *corpora* informatizados estão se tornando cada vez mais populares nas diferentes disciplinas, o que é particularmente verdadeiro para terminologia e tradução automática, em que a ênfase é principalmente, senão exclusivamente, em textos científicos e técnicos.

Nesse sentido, Baker (1995) afirma, em um de seus trabalhos, que a pesquisa baseada em *corpus* tornou-se amplamente aceita a fim de melhorar o desempenho dos sistemas de tradução automática. Ainda, de acordo com a autora, a compilação de terminologia baseada em *corpus* tornou-se uma norma, e não uma exceção. Ela discute, sob essa perspectiva, o uso e o potencial dos *corpora* nos Estudos da Tradução principalmente no que se refere ao uso dos *corpora* informatizados, os quais têm se tornado populares em relação à produção de textos científicos e técnicos.

Firmemente baseada em *corpus*, segundo Baker (1995), a compilação da terminologia se dá pela necessidade de abordar a vida real, já que os termos comumente têm sido extraídos de *corpora* representativos de textos autênticos em formato eletrônico. De acordo com a autora, os *corpora* não são usados apenas para escrever regras para a operação das máquinas, mas também para fornecer diretamente conhecimento para as próprias máquinas. A esse respeito, a autora diz que:

O desenvolvimento de técnicas baseadas em *corpus* em terminologia e tradução automática é encorajador. De certa forma, vai ao encontro da crescente necessidade de uma metodologia descritiva rigorosa na tentativa de aumentar a intersubjetividade das áreas aplicadas dos estudos da tradução, como a formação de tradutores e a crítica de tradução, e, claro, na busca de um conhecimento teórico mais satisfatório⁶ (BAKER, 1995, p. 224, tradução nossa).

A autora também apresenta os tipos de *corpora* que estão sendo usados ou que precisam ser criados especificamente para a pesquisa em tradução. Embora,

⁶ Cf.: “The development of corpus-based techniques in terminology and machine translation is encouraging. It goes some way towards fulfilling the growing need for a rigorous descriptive methodology in an attempt to increase the inter-subjectivity of the applied areas of translation studies, such as translator training and translation criticism, and of course in the pursuit of a more satisfying” (BAKER, 1995, p. 224).

segundo ela, a terminologia para discutir tipos de *corpora* em Estudos da Tradução ainda não tenha sido estabelecida, Baker (1995) propõe três principais tipos, a saber: (i) *corpora* paralelos, (ii) *corpora* multilíngues; e (iii) *corpora* comparáveis.

De acordo com Baker (1995), os *corpora* paralelos geralmente são bilíngues, podendo ser monolíngues e se referem a dois conjuntos de textos. Esses conjuntos correspondem a um conjunto de textos de uma língua específica e a um outro conjunto formado por versões traduzidas dos mesmos textos em outros idiomas. Tais textos são organizados segundo os critérios definidos para a sua utilização e buscam possibilitar a identificação de um determinado padrão em línguas de chegada e em línguas de partida. Já os *corpora* multidirecionais são formados por uma quantidade maior de textos, os quais devem estar ligados entre si, de forma que se utiliza apenas o texto fonte numa única língua e as traduções em diferentes línguas, ainda sendo possível estabelecer uma ligação de várias línguas de partida a várias línguas de chegada, o que torna esses *corpora* mais complexos que os demais. Por fim, os *corpora* comparáveis podem ser monolíngues, bilíngues ou multilíngues. Trata-se, nesse caso, de um conjunto de textos que formam um quadro de amostras comparáveis, sendo representado por um *corpus* monolíngue. Ele é, nesse contexto, composto por um *sub-corpus* de textos traduzidos e por textos não traduzidos. Para isso, é imprescindível que estes sejam compilados de acordo com os mesmos critérios a fim de que possam cobrir um domínio, uma variedade linguística ou um período de tempo.

No tange especificamente ao papel dos *corpora* paralelos nos Estudos da Tradução, Baker (1995, p. 230-231) destaca o seguinte:

Um *corpus* paralelo consiste em textos originais da língua fonte na língua A e suas versões traduzidas na língua B. Este é o tipo de *corpus* que se pensa imediatamente no contexto dos estudos de tradução. Desde que rotinas de software robustas sejam desenvolvidas para alinhar automaticamente trechos de textos-fonte com suas traduções, *corpora* paralelos rapidamente se tornarão indispensáveis na redação de materiais, no treinamento de tradutores auxiliados por computador e na melhoria do desempenho dos sistemas de tradução automática. Sua contribuição mais importante para a disciplina em geral é que eles apoiam uma mudança de ênfase, da prescrição para a descrição. Eles nos permitem estabelecer, objetivamente, como os tradutores superam as dificuldades de tradução na prática, e usar essa evidência para fornecer modelos realistas para tradutores estagiários. Eles também têm um papel importante a desempenhar na exploração de normas

de tradução em contextos sócio-culturais e históricos específicos⁷ (BAKER, 1995, p. 230-231, tradução nossa).

Dando destaque ao *corpus* paralelo, Baker (1995) afirma que este tipo é o mais utilizado nos Estudos da Tradução. A autora acredita que ele é indispensável em diferentes situações, desempenhando um papel importante para a disciplina no que se refere a uma mudança de ênfase.

A esse respeito, Laviosa (2002) destaca que os *corpora* paralelos fornecem evidências empíricas de estratégias e alternativas adotadas por tradutores a fim de solucionar dificuldades e obstáculos durante a tradução, o que é bastante útil para a uma investigação relacionada a específicos contextos históricos e socioculturais.

Dessa forma, após trazer uma reflexão sobre a tipologia de *corpora*, a autora tece a seguinte conclusão acerca dos tipos de *corpora* por ela categorizados:

Tentei dar uma visão geral dos tipos de *corpora* que podem ser usados em estudos de tradução e exemplos do tipo de pesquisa baseada em *corpus* que pode ser realizada na disciplina. Ainda estamos muito longe de alcançar uma metodologia coerente nesta área. A pesquisa baseada em *corpus* oferece um enorme potencial para os estudiosos da tradução, mas o processo de criação dos *corpora* necessários e de elaboração do software relevante é repleto de dificuldades (BAKER, 1995, p. 230-231, tradução nossa⁸).

Nesse sentido, ela confirma a necessidade de que o *corpus* a ser construído deve se ater ao tipo de pesquisa e à função a que visa desempenhar.

⁷ Cf.: "A parallel corpus consists of original, source language-texts in language A and their translated versions in language B. This is the type of corpus that one immediately thinks of in the context of translation studies. Provided that robust software routines are developed for automatically aligning stretches of source texts with their translations, parallel corpora will quickly become indispensable in materials writing, computer-aided translator training and improving the performance of machine translation systems. Their most important contribution to the discipline in general is that they support a shift of emphasis, from prescription to description. They allow us to establish, objectively, how translators overcome difficulties of translation in practice, and to use this evidence to provide realistic models for trainee translators. They also have an important role to play in exploring norms of translating in specific socio-cultural and historical contexts" (BAKER, 1995, p. 230-231).

⁸ Cf.: "I have tried to give an overview of the types of corpora that can be used in translation studies and examples of the kind of corpus-based research that can be carried out in the discipline. We are still a long way from achieving a coherent methodology in this area. Corpus-based research offers enormous potential for translation scholars but the process of setting up the required corpora and of devising the relevant software is fraught with difficulties." (BAKER, 1995, p. 238).

A esse respeito, Laviosa (2002, p. 6, tradução nossa⁹) afirma que “qualquer análise linguística de *corpus* depende tanto da criação de seu objeto de estudo - que é o próprio *corpus* - quanto do desenvolvimento de ferramentas de *software* para observá-lo, analisá-lo e processá-lo”.

Nesse sentido, os parâmetros de construção de um *corpus* devem estar claros e definidos ao longo de sua compilação. Também devem considerar o tipo de pesquisa e a função que o *corpus* desempenha nela para garantir resultados confiáveis e precisos. Além disso, sua forma e sua estrutura devem ser definidas com base nos objetivos a serem alcançados e na metodologia utilizada. Assim, deve-se ater também ao fato de que o *corpus* escolhido deverá ser o mais representativo possível do objeto de pesquisa em análise.

1.2. Linguística de *Corpus* e tradução

Nesta seção, será apresentada a relação entre a Linguística de *Corpus* e a tradução, ou seja, será discutida a viabilidade de se aplicar a Linguística de *Corpus* aos Estudos da Tradução.

Nesse sentido, partindo das considerações de Tagnin (2015), reitera-se que a Linguística de *Corpus* é uma abordagem empírica para o estudo da língua e que, por essa razão, é especialmente útil no estudo da tradução. Essa abordagem parte da observação de uma grande quantidade de textos, reunidos em *corpora* a fim de promover inferências a respeito de como a língua é usada por meio de ferramentas computacionais. Essas ferramentas fornecem, nesse sentido, dados quantitativos, os quais também permitem análises qualitativas.

Tagnin (2015), nesse contexto, define os *corpora* como bancos de textos representativos de linguagem autêntica. Segundo ela, eles são criteriosamente construídos, destinados à pesquisa e facilmente lidos por computador, podendo ser consultados *on-line* ou *off-line* e sendo geralmente especializados e compilados por pesquisadores conforme seus objetivos de estudo. A autora também afirma que:

⁹ Cf.: “Any corpus linguistic analysis is dependent on both the creation of its object of study - that is the corpus itself - and the development of software tools to observe, analyze and process it. (LAVIOSA, 2002, p. 6)

Na Linguística de *Corpus*, enfatiza-se sempre a linguagem autêntica, pois a língua é tida como um sistema probabilístico (HALLIDAY, 1961), ou seja, um sistema em que, embora muitas construções sejam possíveis, algumas delas têm probabilidade maior de ocorrer. Em outras palavras, nem tudo o que é possível ou gramatical numa língua, de fato, ocorre (TAGNIN, 2015, p. 20).

Por essa razão, segundo ela, os *corpora* devem ser criteriosamente construídos. Devem visar, portanto, ao objetivo pretendido para que os textos compilados sejam representativos e tenham procedência de fonte confiável para que os resultados também o sejam. Tagnin (2015) também afirma que os textos devem estar em formato digital para serem legíveis por computadores. Ela também afirma que o primeiro passo para a compilação de um *corpus* é definir seus objetivos. Dessa forma, de acordo com ela, devem ser estabelecidos os parâmetros de sua composição, destacando as seguintes questões:

- a. será um *corpus* estático (i.e. fechado) ou dinâmico (i.e. com atualizações)?
- b. incluirá apenas textos escritos ou também língua falada?
- c. será monolíngue ou multilíngue?
- d. se for multilíngue, será paralelo ou comparável?
- e. que tipos de textos serão incluídos: jornalísticos, manuais, artigos acadêmicos, teses e dissertações?
- f. qual será a proporção entre os diversos tipos?
- g. quais domínios abrangerá? Por exemplo, natação, sustentabilidade, segurança bancária, obras de determinado autor, obras de determinada época etc.
- h. qual será a fonte dos textos a serem compilados: Internet ou material impresso? No caso de material impresso, é preciso lembrar que esse deverá ser escaneado ou, na pior das hipóteses, digitado.
- i. os textos serão completos ou apenas fragmentos? Por exemplo, os textos que compõem o Brown Corpus são constituídos de trechos de 2.000 palavras, enquanto o Bank of English abriga textos completos. Já o COMPARA é constituído, no geral, de 30% das obras contempladas.
- j. como será estabelecido o tamanho do *corpus* pelo número de textos ou pelo número de palavras? Essa decisão, muitas vezes, terá de ser modificada durante o processo de compilação devido à especificidade dos textos. (TAGNIN, 2015, p. 27-28)

A autora destaca que, embora essas questões devam ser estabelecidas antes de começar a coleta dos textos, elas podem sofrer ajustes ao longo da compilação, já que, em determinados casos, poderá ser necessário recorrer a textos

impressos para obter um *corpus* bilíngue com conteúdo equiparável nas duas línguas. Nesse contexto, ela também aborda a importância da representatividade de um *corpus*:

Um fator frequentemente discutido na literatura (e. g. LEECH, 1991, BIBER, 1993, MACENERY et al., 2006) diz respeito à representatividade de um *corpus*, ou seja, o fato de que um *corpus* deve ser representativo daquilo que se pretende investigar, contendo uma amostragem significativa de universo a ser estudado. Quando isso é alcançado, os resultados obtidos podem ser generalizados para todo o universo em estudo (LEECH, 1991). No entanto, o que seria uma amostragem significativa? Segundo Biber (1993, p. 243), a representatividade diz respeito ao quanto uma amostra inclui toda a gama de variabilidade da produção linguística que se pretende investigar. Dessa forma, um *corpus* geral, por exemplo, deve incluir o maior número de gêneros e tipos textuais para assegurar sua representatividade. (TAGNIN, 2015, p. 29).

Assim, após se ater à representatividade dos textos, estes devem ser salvos no formato .txt para serem lidos por ferramentas computacionais e devem ser retirados os elementos que não são de interesse da pesquisa. Por fim, os textos estarão prontos para serem explorados por ferramentas computacionais.

Tagnin (2015) também cita o programa *Antconc* como uma das ferramentas de investigação mais conhecidas, que apresenta três funcionalidades básicas para análise de um *corpus*: a lista de palavras, a lista de palavras-chave e a concordância. Ela apresenta, nesse sentido, as finalidades de cada ferramenta do programa com o objetivo de fornecer um panorama acerca das possibilidades facultadas pelo uso de *corpora* nos Estudos da Tradução.

Com intuito de apresentar o programa *Antconc*, que será aplicado na realização desta pesquisa, na seção seguinte, será discutido como cada uma das funcionalidades que constituem essa ferramenta pode contribuir para a área de tradução.

1.3. O programa Antconc e sua contribuição para os Estudos da Tradução

O propósito desta seção é apresentar, de modo geral, o programa Antconc e sua contribuição para os Estudos da Tradução, considerando sua importância para uma metodologia adequada ao tratamento dos dados a serem analisados no Capítulo III. Assim sendo, são apresentadas, a seguir, algumas reflexões acerca das funcionalidades e das ferramentas que compõem o programa.

Conforme Cunha Lacerda e Dall'Orto (no prelo), o uso do programa pode trazer importantes evidências empíricas tanto de um ponto de vista qualitativo quanto quantitativo. Nesse sentido, as autoras apresentam uma versão mais atualizada do programa – versão 3.5.9. – e ressaltam que ele apresenta duas grandes vantagens, a saber: a) a rapidez durante o *download*, pois apresenta um tamanho pequeno para o arquivo, evitando, assim, ocupar muito espaço de armazenamento; e b) o fato de não necessitar de licenciamento e processo de instalação. Conforme são apresentadas as funcionalidades do programa, é possível perceber que ele apresenta uma interface bastante simples e prática, sendo possível navegar por meio de suas diversas opções facilmente. Ainda segundo as autoras, é "possível navegar por diferentes opções de análise, que permitem descobrir como o elemento pesquisado ocorre, o quanto ocorre e em que contextos ocorre" (CUNHA LACERDA; DALL'ORTO, no prelo), o que torna a utilização do programa ainda mais relevante.

De acordo com Salgado (2013), o programa *Antconc* apresenta diversas vantagens, tendo uma grande contribuição para os Estudos da Tradução. Segundo a autora, além de ser um arquivo leve, ele também dispensa a necessidade de instalação, facilitando seu uso em qualquer computador com diferentes sistemas operacionais (*Windows, Mac e Linux*) e sua utilização por iniciantes no estudo da Linguística de *Corpus*.

Após a testagem das ferramentas do programa, a autora afirma ser possível verificar que o *Antconc* apresenta uma série de vantagens, as quais ela lista da seguinte forma:

- o tamanho pequeno do arquivo, que permite um download rápido (mesmo com conexões à internet de baixa velocidade) e não ocupa muito espaço em disco;
- o fato de não haver necessidade de instalação e licenciamento;
- ser um *freeware* com versões para diferentes sistemas operacionais;
- a facilidade de uso, que permite acessar várias funcionalidades em uma mesma interface, com apenas um clique (SALGADO, 2013, p. 88).

Em conformidade com as considerações de Salgado (2013), Barreiros (2017) também descreve as vantagens do programa *Antconc*. Segundo a autora:

O uso do AntConc mostra-se eficiente porque dimensiona o volume de dados do *corpus* em análise, observando-se as ocorrências e as variações de uso da língua. Ele permite explorar o *corpus* em sua totalidade de maneira rápida e eficiente, o que seria muito difícil manualmente (BARREIROS, 2017, p. 231).

Dessa forma, os autores apresentam as vantagens quanto ao uso do programa *Antconc* e descrevem algumas contribuições que ele pode proporcionar nas pesquisas realizadas no âmbito dos Estudos da Tradução.

Desenvolvido pelo pesquisador Lawrence Anthony, da Faculdade de Ciências e Engenharia da Universidade de Waseda, no Japão, o programa *Antconc* constitui um *freeware*, sendo possível fazer seu *download* gratuitamente¹⁰.

Nesse sentido, a figura, a seguir, representa a tela inicial do programa, a qual pode ser aberta através de duplo clique:

¹⁰ Disponível para *download* por meio do seguinte endereço <http://www.laurenceanthony.net/software/antconc/>.

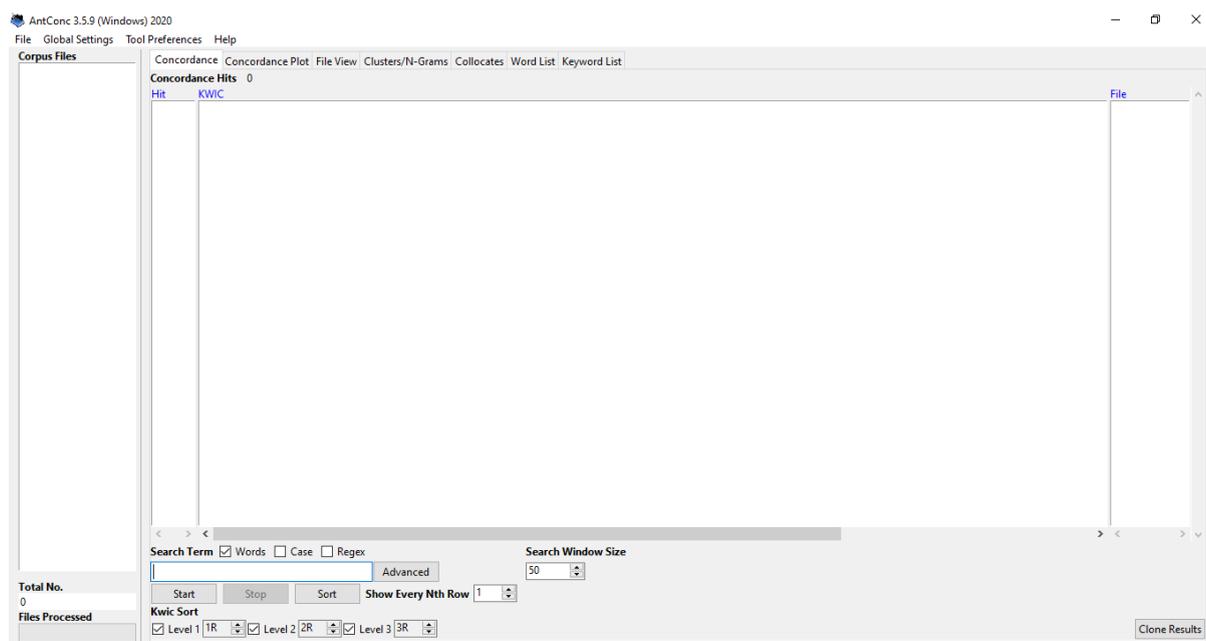


Figura 1 – Tela inicial do programa *Antconc* 3.5.9.
Fonte: imagem retirada do programa *Antconc*

Como destacam Cunha Lacerda e Dall’Orto (no prelo), o programa *Antconc* possui as seguintes ferramentas para o processamento de dados: i) *WordList*; ii) *Concordance*; iii) *Concordance Plot*; iv) *File View*; v) *Clusters/N-Grams*; vi) *Collocates*; e vii) *Keywords*.

As autoras também destacam que, por meio da ferramenta *WordList*, é possível dispor os dados em ordem alfabética ou por frequência. Através dessa ferramenta, é possível encontrar rapidamente termos ou construções com maior frequência ou aqueles com menor ocorrência em um *corpus* por meio do critério *Sort by Freq*, como mostra a figura a seguir.

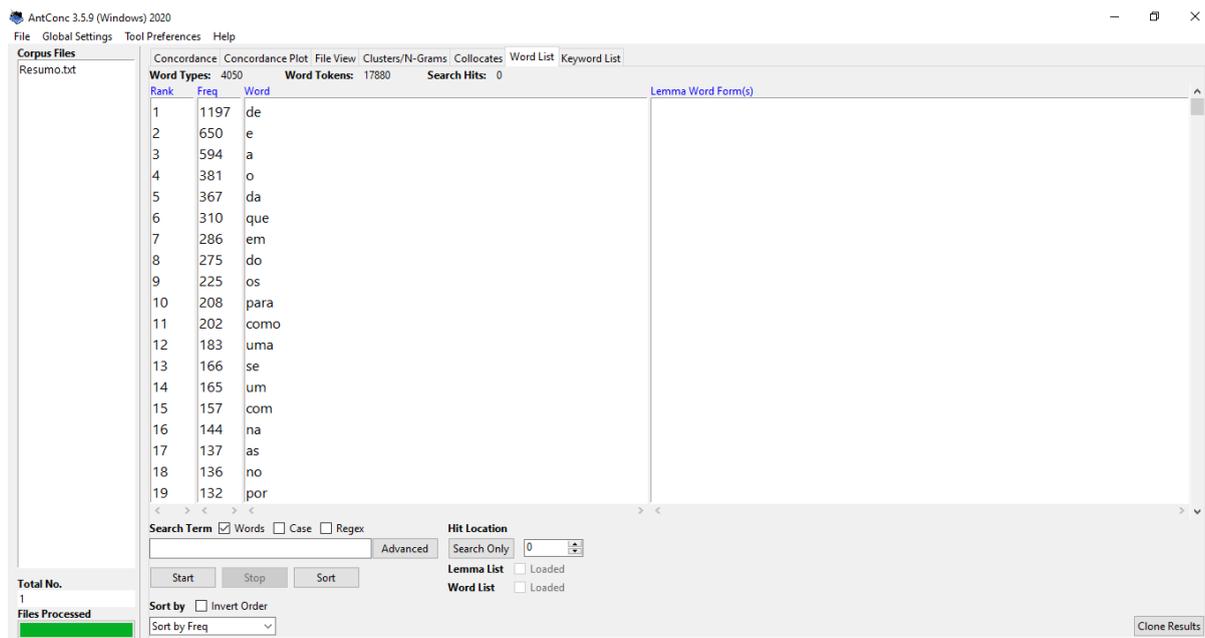


Figura 2 – Ferramenta *Wordlist* distribuída pelo critério *Sort by Freq*
Fonte: imagem retirada do programa *Antconc*

Nessa ferramenta, também é possível se pautar na organização das palavras por ordem alfabética por meio do critério *Sort by Word*:

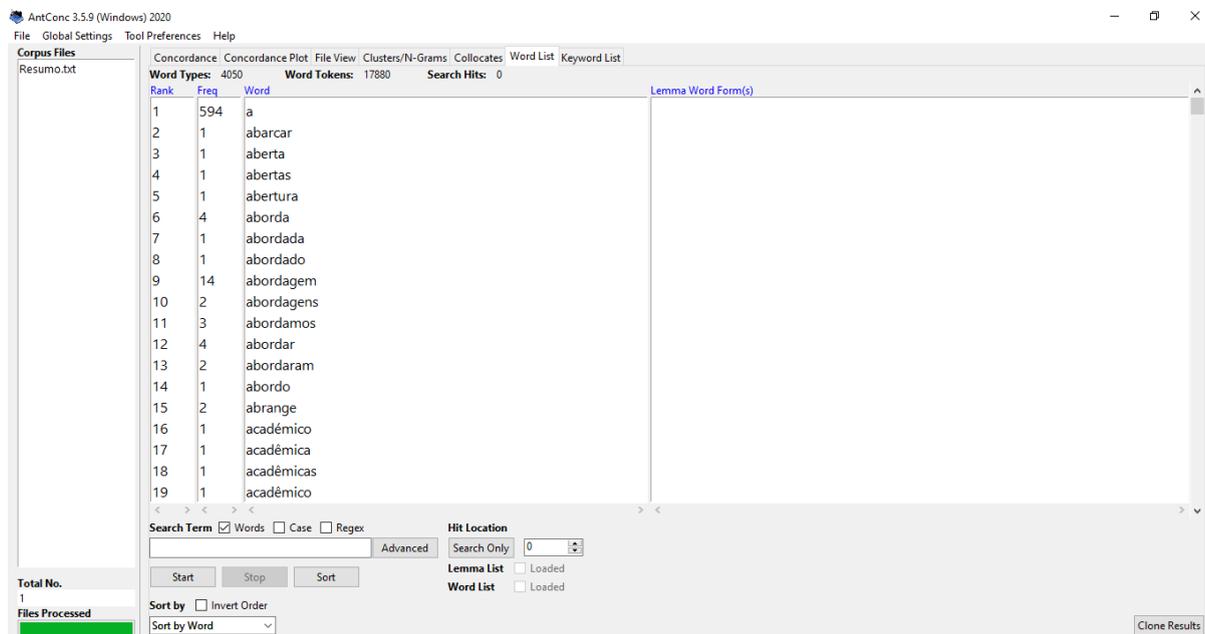


Figura 3 – Ferramenta *Wordlist* distribuída pelo critério *Sort by Word*
Fonte: imagem retirada do programa *Antconc*

E, por fim, é possível pautar a ordenação segundo as terminações:

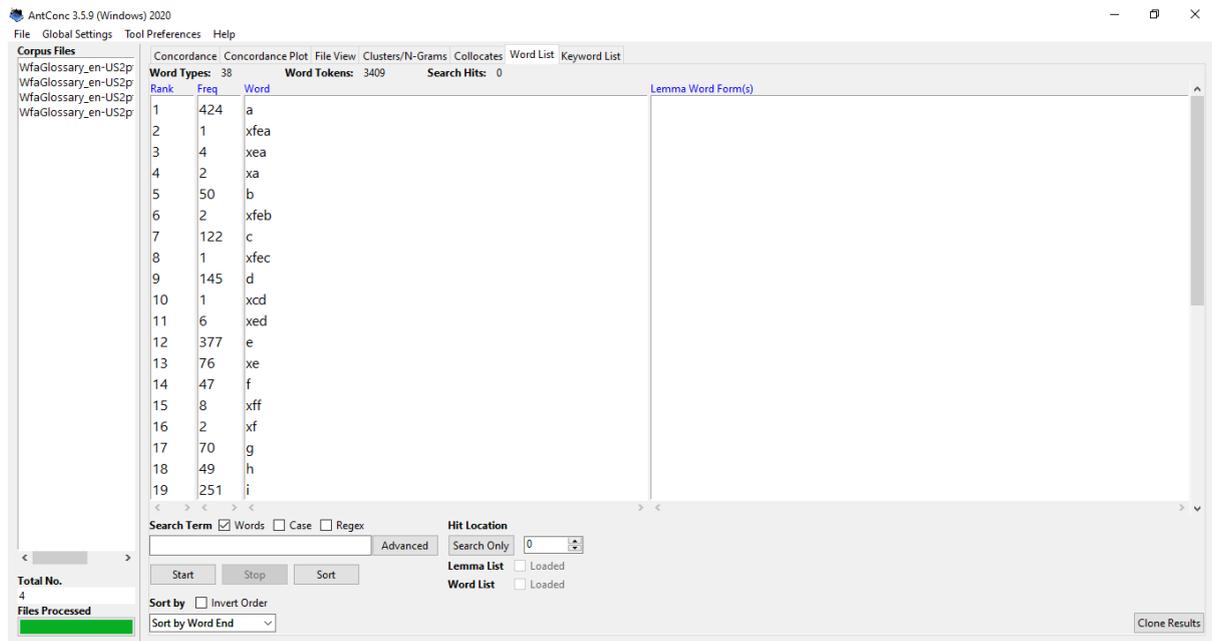


Figura 4 – Ferramenta *Wordlist* distribuída pelo critério *Sort by Word End*
Fonte: imagem retirada do programa *Antconc*

Cunha Lacerda e Dall’Orto (no prelo) descrevem a ferramenta *Concordance* como uma ferramenta capaz de mostrar os resultados no formato 'KWIC' (*KeyWord In Context*), permitindo visualizar como as construções são utilizadas em um *corpus* partindo da distribuição no contexto. Além disso, segundo as autoras, "o programa nos fornece quantos *tokens* representativos de determinado padrão construcional foram verificados e também apresenta o contexto de que faz parte cada ocorrência empiricamente atestada", podendo esta ferramenta ser caracterizada por meio da seguinte tela:

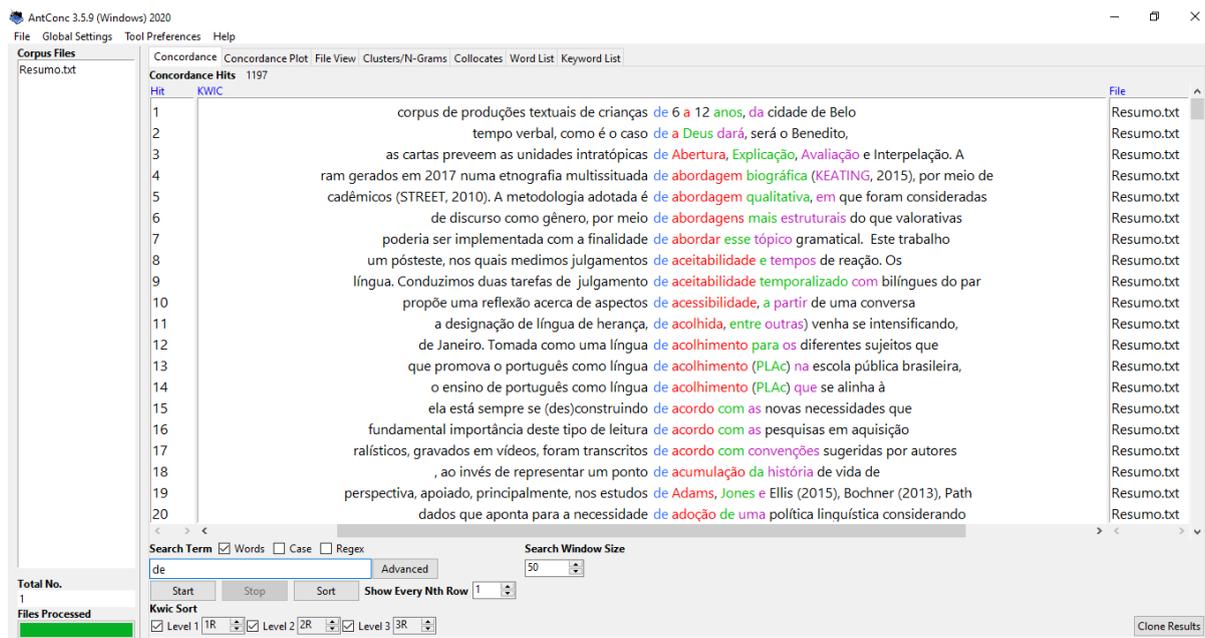


Figura 5 – Tela representativa da ferramenta *Concordance*
 Fonte: imagem retirada do programa *Antconc*

A seguir, apresenta-se a tela da ferramenta *File View*, a qual possibilita a visualização do texto bruto de arquivos individuais:

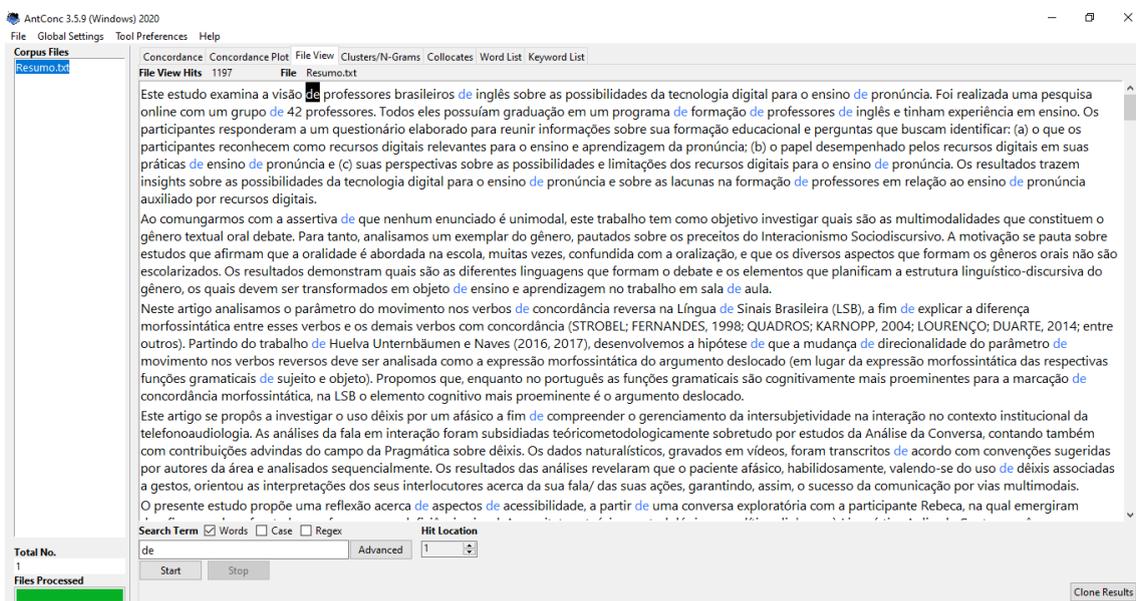


Figura 6 – Tela representativa da ferramenta *File View*
 Fonte: imagem retirada do programa *Antconc*

Por meio dessa ferramenta, é possível visualizar o texto bruto de arquivos individuais, o que permite uma investigação mais detalhada em relação aos resultados dos produtos das ferramentas *Concordance* e *Concordance Plot*.

Concordance Plot constitui uma ferramenta de plotografia de concordância, como destacam Cunha Lacerda e Dall’Orto (no prelo). Ela possibilita a visualização dos resultados da pesquisa em forma de gráfico, mostrando a distribuição da palavra ou construção analisada pelo *corpus*. Abaixo é possível visualizar a tela representativa da ferramenta *Concordance Plot*:

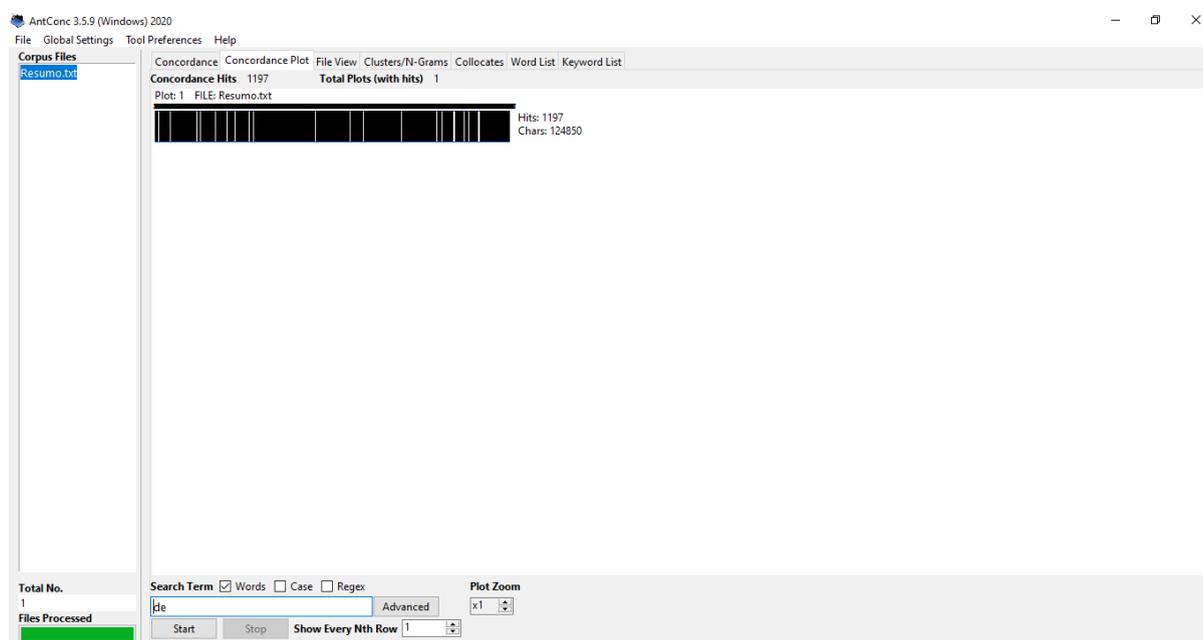


Figura 7 – Tela da ferramenta *Concordance Plot*
Fonte: imagem retirada do programa *Antconc*

De acordo com Cunha Lacerda e Dall’Orto (no prelo), a ferramenta *Clusters/N-Grams* gera uma lista com resultado da pesquisa conforme a ordem alfabética, de frequência, de probabilidade ou de terminações. A ordem dessas listas pode ser invertida por meio da opção *Invert Order*. Ela gera dados a partir dos resultados gerados em *Concordance* ou *Concordance Plot*, o que possibilita verificar quais elementos aparecem mais à esquerda ou à direita do padrão construcional, o que faz com que o olhar do analista, quanto a ele, seja mais refinado. Além disso, esse processo é vantajoso quando se trabalha com o levantamento de lexias

compostas e complexas por possibilitar a localização de expressões comuns em todo o *corpus*, visto que é possível selecionar o número de palavras em cada *cluster* e a frequência mínima de *clusters*. A seguir, é apresentada a tela representativa dessa ferramenta.

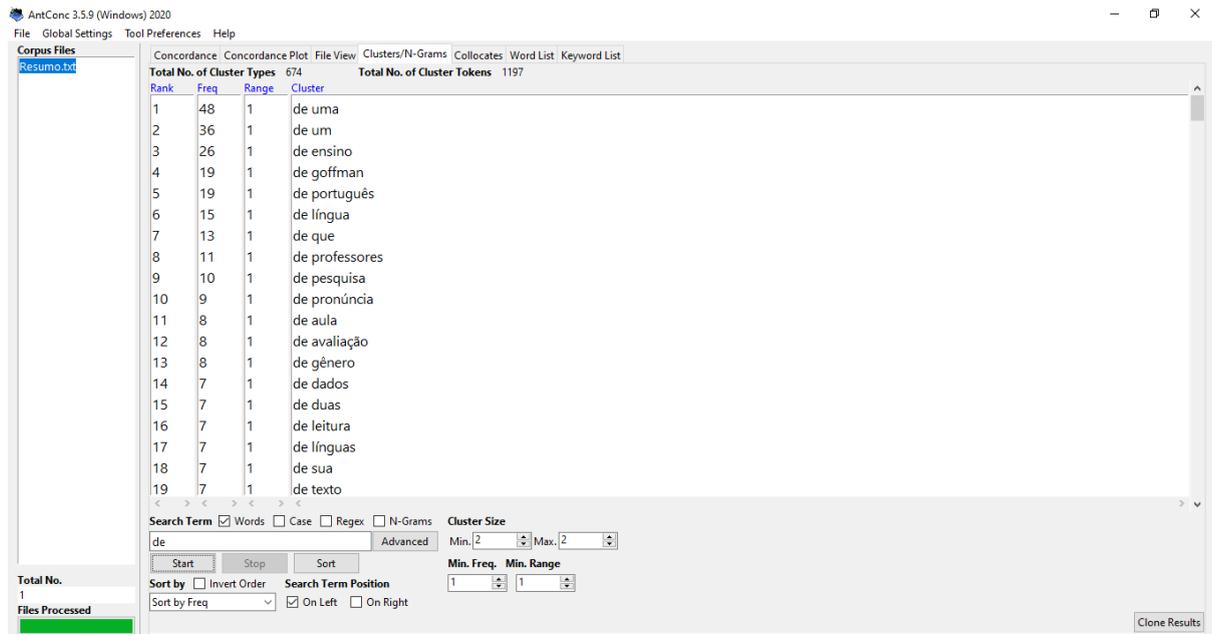


Figura 8 – Tela da ferramenta *Clusters/N-Grams*
 Fonte: imagem retirada do programa *Antconc*

A seguir, é apresentada a imagem representativa da ferramenta *Collocates*, a qual pode contribuir para o refinamento da pesquisa, permitindo uma investigação de padrões não sequenciais na língua. Por meio dela é possível ter uma média estatística, que é gerada automaticamente, possibilitando medir o nível de relação entre o termo pesquisado e o colocado.

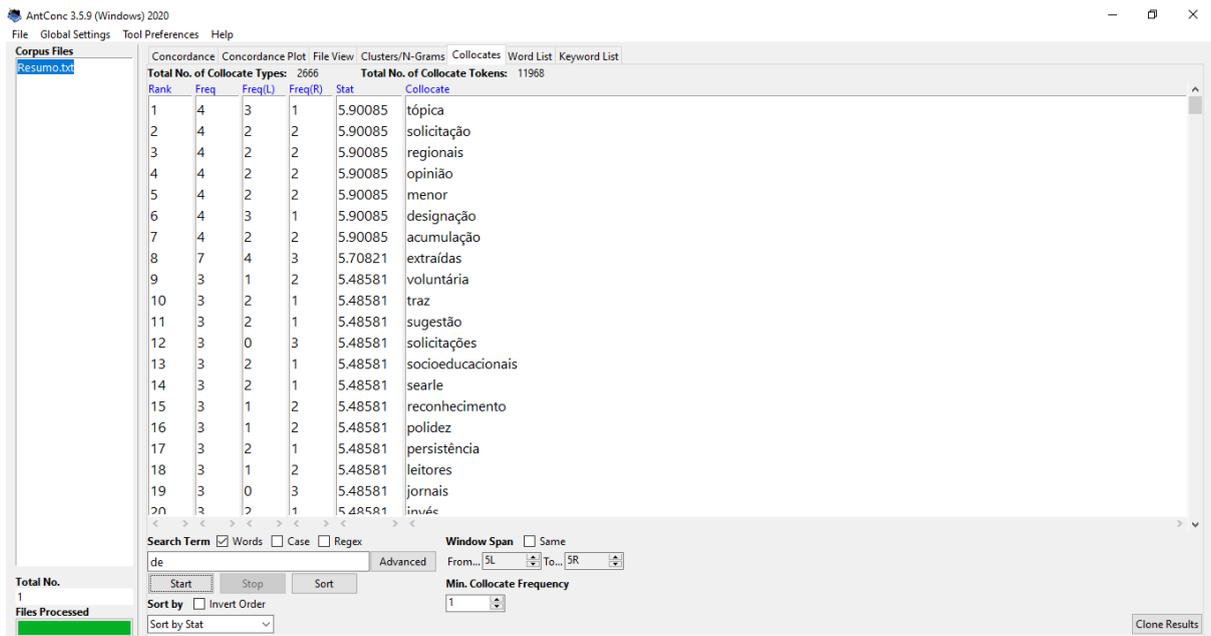


Figura 9 – Tela da ferramenta *Collocates*
 Fonte: imagem retirada do programa *Antconc*

Por fim, abaixo é apresentada a tela representativa da ferramenta *Keywords*, a qual permite identificar palavras características no *corpus* como parte de um gênero ou variedade linguística, segundo Cunha Lacerda e Dall’Orto (no prelo), visto que esta gera uma lista de palavras-chave, comparando a frequência das palavras dos arquivos analisados com a frequência das palavras de um outro *corpus*, pertencente a outro gênero textual.

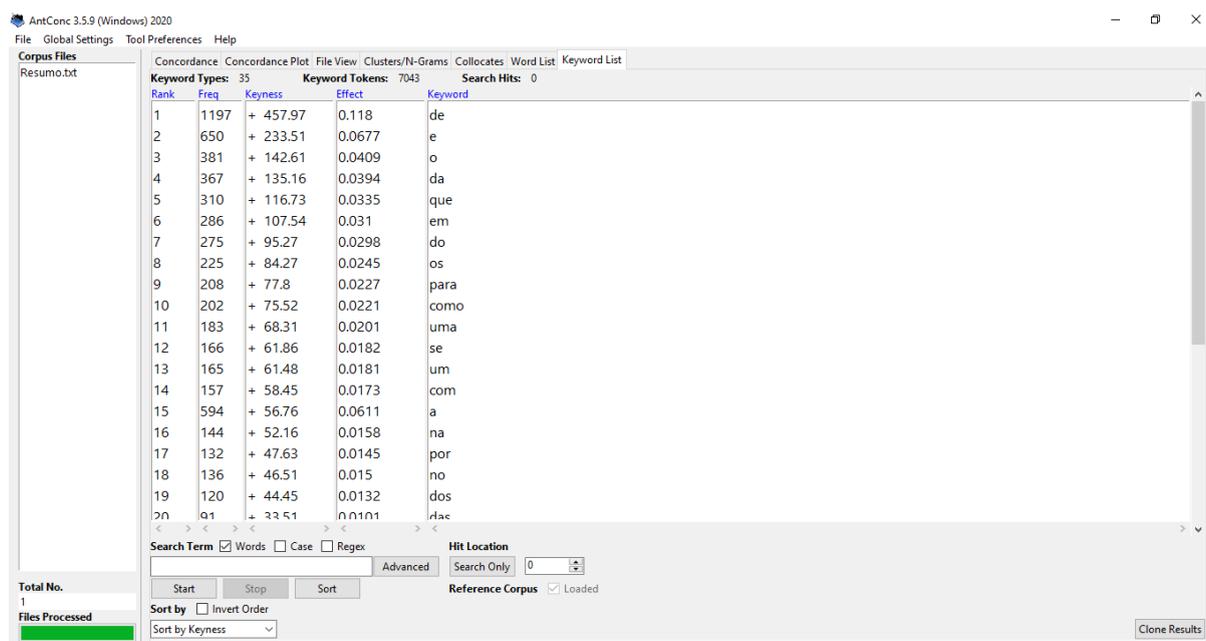


Figura 10 – Tela da ferramenta *Keywords*
 Fonte: imagem retirada do programa *Antconc*

Cunha Lacerda e Dall’Orto (no prelo) concluem, a partir da apresentação dessas ferramentas, que a utilização do programa *Antconc* pode permitir, de modo rápido e preciso, o processamento e a análise de um *corpus* ou de dois ou mais *corpora*. Para as autoras, o uso do programa permite:

[...] refinar o nosso olhar para uma identificação mais apurada dos padrões construcionais com os quais trabalhamos, nos permitindo, de uma forma muito mais rápida e precisa, ter acesso a informações sobre frequência de uso e também acesso a todas as ocorrências empiricamente atestadas que são, de fato, representativas dos padrões construcionais investigados em determinado *corpus* ou em determinados *corpora*. (CUNHA LACERDA; DALL’ORTO, no prelo)

Nesse sentido, reafirma-se aqui assim, como Cunha Lacerda e Dall’Orto (no prelo), as contribuições que a utilização do programa *Antconc* pode proporcionar, visto que, do ponto de vista analítico, o olhar sobre os dados acaba sendo mais refinado.

1.4. Conclusões

Partindo da noção de que a Linguística de *Corpus* coleta e explora *corpora* ou conjunto de dados linguísticos de forma criteriosa a fim de servir a uma pesquisa e de permitir que a coleta e exploração sejam advindas de evidências empíricas extraídas por ferramentas digitais, entende-se que a Linguística de *Corpus* aplicada aos Estudos da Tradução pode trazer importantes contribuições para a área.

Entendida como uma abordagem empírica para o estudo da língua, a Linguística de *Corpus* se torna bastante útil nesse sentido, visto que, a partir da observação de *corpora*, é possível fazer inferências por meio de ferramentas computacionais a respeito de como uma língua é usada. Para isso, é preciso que os textos sejam representativos da língua em pesquisa para serem considerados uma fonte confiável. Além disso, é preciso objetivar previamente os parâmetros de composição para assegurar sua representatividade.

Nesse sentido, as ferramentas computacionais são fundamentais no processamento e coleta de dados. Reiterando o que foi apresentado neste capítulo, o programa *Antconc* apresenta aspectos vantajosos, que são de grande contribuição para os Estudos da Tradução. Por ser de fácil acesso, possuir tamanho compacto, apresentar a facilidade de uso e ser um *freeware* que atende a diferentes sistemas operacionais, o programa ganha destaque por facilitar a busca de informações durante a análise dos dados e por apresentar sete ferramentas com funcionalidades específicas que são de grande valia para o pesquisador, dando a ele a possibilidade de um resultado mais refinado e rápido. Vale ressaltar que, mesmo com diversas vantagens, o olhar do pesquisador ainda é fundamental durante o processo para filtrar dados e se ater ao que é objeto de pesquisa.

Dessa forma, entende-se que a Linguística de *Corpus* desempenha um papel importante para as pesquisas no campo da tradução e que o programa *Antconc* apresenta vantajosas contribuições para a área, podendo ser facilmente manipulado por pesquisadores durante as pesquisas envolvendo *corpora*.

CAPÍTULO II

TRADUÇÃO TÉCNICA: PRESSUPOSTOS FUNDAMENTAIS

A partir da década de 1980, os estudos sobre gêneros textuais começaram a se destacar. De acordo com Bernardino e Valentim (2016, p. 27), é notório que, nas universidades, os gêneros discursivos vêm sendo reconhecidos como importantes espaços para a interação social, para a negociação, para a construção de posicionamentos e avaliações. À luz do que consideram esses e outros estudiosos, tais como Bragagnollo (2011) e Teles (2007), este capítulo busca, nesse sentido, situar a compreensão do funcionamento social do gênero resumo acadêmico, visto que, ao longo deste trabalho, busca-se analisar resumos e *abstracts* das edições da Revista Veredas dos últimos 5 anos – o que torna a apresentação desse gênero e de seus principais aspectos imprescindíveis.

Desse modo, a fim de cumprir este objetivo mais geral, neste capítulo, serão apresentados os pressupostos fundamentais da tradução técnica, a partir das considerações de Aixelá (2016) e de Cavaco-Cruz (2012) a respeito da sua constituição, destacando também os pontos de vista de Tavares (2022) quanto ao trabalho do tradutor nesse processo. Também serão apresentadas a noção de Terminologia e a sua relação com a tradução à luz das reflexões de Aubert (2001), Kriegger e Santiago (2014), Ribeiro (2004), Almeida e Simões (2002) e Bevilacqua e Kilian (2017). Por fim, serão explicitadas as considerações de Bragagnollo (2011), Bernardino e Valentim (2016) e Teles (2007) a respeito da função do gênero resumo e dos processos que o constituem.

2.1. Tradução técnica: uma revisão

A tradução técnica e científica é um ramo da tradução que consiste em envolver a noção de que a literatura engloba a prática linguística de forma criativa, exigindo do tradutor uma reescrita que acompanhe o processo criativo. Para Aixelá (2016, p. 132), a "tradução técnica e científica se constitui tradicionalmente como bedel das discussões teóricas sobre tradução". Em conformidade com esse

raciocínio, Cavaco-Cruz (2012, p.1) considera a tradução técnica uma atividade altamente especializada e afirma que ela “requer do linguista profissional, enquanto tradutor técnico, uma grande polivalência de conhecimentos e um leque alargado de competências que são fulcrais à comunicação empresarial, industrial e tecnológica por todo o mundo”.

Nesse sentido, Aixelá (2016) compara a tradução literária à tradução técnica e afirma que a primeira requer uma maior reflexão, enquanto a segunda necessita apenas de profissionais com um conhecimento terminológico considerável em determinada área.

O autor também afirma que a tradução técnica representa o grosso dos textos traduzidos. Segundo o autor, atualmente, essa questão tem transparecido nos cursos de graduação de tradução, atribuindo à área técnica um peso maior a fim de adaptá-la ao mercado e deslocando a tradução literária para outro plano. Para ele, há "um crescente entendimento de que a tradução técnica merece mais atenção, mesmo que seja porque se trata do principal nicho profissional em que atuarão os estudantes universitários da área" (AIXELÁ, 2016, p. 135).

O autor também reflete a respeito da tendência demonstrada por aspectos referentes à tradução técnica e científica nos Estudos da Tradução e conclui que:

Existe hoje um amplo público leitor diretamente envolvido na área de tradução, geralmente sobre tradução técnica em particular, e acadêmicos estão começando a desenvolver uma nova percepção acerca de uma disciplina que finalmente se encontra protegida sob o manto da dignidade acadêmica (AIXELÁ, 2016, p. 159-160).

Isso reafirma, segundo o autor, que ainda há muito o que ser discutido, visto que há, cada vez mais, espaço para publicações desse tipo. À vista dessa ponderação, Cavaco-Cruz (2012, p. 2) também tece suas considerações:

A tradução técnica está-se [sic] a tornar, de forma pausada mas firme, numa profissão que tem como caminho de sentido único a via da qualidade. Aliada às ferramentas da gestão, como o *benchmarking* e os processos de controlo [sic] de qualidade, e imbuída cada vez mais dos rumos investigativo e empresarial, a tradução técnica apresenta-se internacionalmente como uma das profissões com um dos maiores índices de crescimento e com uma procura formativa de relevo.

Nesse sentido, os conceitos apresentados por Cavaco-Cruz (2012) são tomados como relevantes por Tavares (2022), que os coloca, entretanto, sobre o seguinte ponto de vista:

O trabalho de um tradutor técnico também envolve o processo criativo de reformular a informação do texto de partida de forma clara, descodificando as estruturas frásicas mais complicadas, utilizando estratégias de tradução como por exemplo a adaptação ou a clarificação. Utilizando as palavras de Polchlopek e Aio: “as competências e habilidades” do tradutor não são desenvolvidas só a nível terminológico ou gramatical e estão, na verdade, associadas ao processo de tradução e às: “questões de equivalência textual, fidelidade, funcionalidade do texto e, até mesmo, o próprio conceito de tradução” (TAVARES, 2022 *apud* POLCHLOPEK; AIO, 2009, p.108).

Para a autora, a tradução técnica difere de todos os outros tipos de tradução, tendo em vista a clareza, a concisão e o rigor, que são seus principais atributos. Tavares (2012) também coloca em evidência que:

[...] a tradução técnica requer muito mais do que apenas escrever os equivalentes de palavras numa outra língua e, embora a familiaridade com a língua de partida seja crucial, ainda assim, o conhecimento da língua de partida e a habilidade de escrita na língua de chegada são insuficientes para garantir uma tradução de qualidade (TAVARES, 2022, p. 23).

Nesse sentido, ela destaca que a matéria do texto fonte é essencial para a realização de uma boa tradução, sendo necessário que o tradutor entenda o tópico originalmente apresentado. Somente dessa forma, haverá uma tradução clara, concisa e correta.

Além disso, Tavares (2022) afirma que o objetivo do tradutor é comunicar informação por meio de um texto. Nesse sentido, ele se torna um escritor técnico intercultural quando este propósito se sobrepõe à intenção de transferir o discurso presente no texto original para o texto traduzido. Ela também classifica a tradução técnica “como a tradução de qualquer matéria pertencente a uma área do saber técnica ou tecnológica, que requeira conhecimento especializado” (TAVARES, 2022, p. 24).

Em relação ao uso de ferramentas para a tradução, a autora afirma que atualmente trata-se de uma parte integrante da tradução. Sob essa perspectiva, o

fato de a tradução ser vista como um produto faz com que ela seja considerada "o resultado de um processo de fabricação que inclui a utilização de recursos básicos como é o caso da terminologia, da fraseologia ou de componentes "reciclados" como as memórias de tradução ou segmentos previamente traduzidos" (TAVARES, 2022, p. 94).

Desse modo, é possível afirmar que a tradução técnica está evoluindo e sofrendo uma "industrialização" associada à globalização. Por conseguinte, acredita-se que haverá uma maior mecanização e um maior desenvolvimento das ferramentas envolvidas no processo, como afirma Tavares (2022, p. 94).

Assim sendo, é importante considerar os desenvolvimentos tecnológicos da área para entender parcialmente como a tradução funciona, visto que, apesar de todas as vantagens do uso dessas ferramentas, ainda existem limitações que só podem ser observadas pelo tradutor. Isso reforça a afirmação de que a tradução "baseia-se na aplicação prática do conhecimento científico enquanto a tradução científica está relacionada ao conhecimento teórico" (TAVARES, 2022, p. 95).

Em resumo, para uma boa tradução técnica, é preciso que o tradutor reconheça suas dificuldades, troque suas ferramentas, mude de estratégia, corrija seus erros e estude diferentes temas e áreas para se manter preparado durante o processo, trazendo e levando conceitos de uma cultura para outra.

2.2. Terminologia e tradução

Por muito tempo, acreditou-se que a tradução de textos técnicos fosse mais fácil de ser realizada, mas esse pensamento tem se alterado. Segundo Aubert (2001), a demanda por materiais terminográficos está em expansão. De acordo com o autor, "a relação entre a terminologia e a tradução técnico-científica não é apenas antiga, no sentido de que a necessidade de tradução técnico-científica induziu e induz, até hoje, uma elevada produção de materiais terminológicos" (AUBERT, 2001, p. 42).

Atualmente, sabe-se que é preciso lidar com as terminologias para traduzir textos técnicos. Nesse sentido, os estudiosos têm direcionado um olhar para a Terminologia em direção à tradução, o que tem motivado bastantes estudos sobre os termos técnicos. À vista disso, Kriegger e Santiago (2014) consideram que:

Como não há comunicação especializada sem termos técnico-científicos, explica-se também que os tradutores técnicos busquem na Terminologia, enquanto campo de estudos, fundamentos teóricos capazes de explicar a constituição e o funcionamento dos termos (KRIEGGER; SANTIAGO, 2014, p. 43).

Segundo os autores, o uso da terminologia não é recente, uma vez que povos de civilizações antigas já a utilizavam. Além disso, na metade do século XX, a Terminologia surgiu como área de conhecimento. De acordo com eles, ao longo do tempo, desenvolveram-se teorias que visavam a “descrever os três componentes constitutivos do termo: o aspecto linguístico propriamente dito, sua dimensão conceitual e a dimensão comunicativa que se materializa nos textos que veiculam conhecimentos especializados” (KRIEGGER; SANTIAGO, 2014, p. 45).

Ainda de acordo com Krieger e Santiago (2014), a face linguístico-textual da Terminologia começou a ter vários desdobramentos, tanto no que se refere a pesquisa teórica quanto à aplicada, tornando-se, assim, um campo de investigação de interesse pelos tradutores. Os autores também afirmam que, com o desenvolvimento das atuais tecnologias, o tradutor passou a ter acesso a diversos conjuntos de textos, com diferentes gêneros e temáticas, podendo fazer uso deles para suas necessidades informativas. Para os autores, isso é possível tendo em vista os estudos entre *corpora* paralelos. Além disso, argumentam que o tradutor encontra à sua disposição diversos recursos *on-line* que auxiliam no reconhecimento da frequência de ocorrências lexicais e sintagmáticas, o que também permite observar a língua em funcionamento, impulsionando as descrições comparativas.

Em relação a esses recursos, Ribeiro (2004) diz que, com a chegada da informática na década de 1990, os computadores passaram a desempenhar um papel fundamental no mercado profissional e no meio acadêmico. Tais avanços foram de grande importância para a área da tradução. Isso se deu principalmente com o aumento de ferramentas disponíveis para agilizar o trabalho dos tradutores, como editores de texto, corretores ortográficos e sintáticos e outros programas específicos de tradução. Assim, surgiram diferentes mecanismos de busca na internet, mas, ao mesmo tempo, surgiu o risco da utilização de *corpora* não

critérios, sem controle da qualidade. Nesse sentido, Ribeiro (2004, p. 171) destaca que:

Os recursos tecnológicos vêm mudando a perspectiva a respeito dos estudos da linguagem em geral e da tradução. A quantidade de informações que pode ser processada e analisada e a rapidez com que isso pode ser feito possibilitam observações e estudos que seriam impossíveis manualmente, além de conferirem um caráter de imparcialidade e representatividade à pesquisa.

Para que isso aconteça, é preciso que dados sejam coletados e que haja a criação de *corpus* com critérios rigorosos para evitar resultados equivocados. Atualmente, isso é possível graças à utilização de ferramentas como o programa *Antconc*, que foi apresentado no capítulo anterior.

Relacionando a problemas com a tradução decorrentes do uso das variações linguísticas, Kriegger e Santiago (2014, p.50) argumentam que as investigações comparativas entre textos de idiomas diferentes têm se intensificado, impulsionadas pela Linguística de *Corpus* e pela Retórica Contrastiva. Como a Linguística de *Corpus* tem como objetivo a "descrição empírica da língua em uso, tomando por base grandes conjuntos de textos autênticos, selecionados e organizados com critérios pertinentes à investigação proposta", tal abordagem tornou-se uma espécie de ferramenta indispensável para as investigações durante as traduções.

Nesse sentido e na busca de explicitar o entendimento da Terminologia, Kriegger e Santiago (2014, p.50) afirmam que:

A Terminologia busca, portanto, dar conta da identificação ao definir características e tentar estabelecer limites entre sintagmas terminológicos e unidades fraseológicas especializadas. Por sua vez, o foco maior para a tradução é a escolha da fraseologia especializada equivalente que seja mais adequada a cada uso profissional. Nessa perspectiva, o estudo da fraseologia especializada é importante para a Tradução, pois no trabalho tradutório há transposição de significados de uma língua para outra. Considerando esse fato, o tradutor se depara muitas vezes com dificuldades em traduzir expressões características dos dizeres profissionais, como *baixar o arquivo*, típica da Informática, e *fazer febre*, oriunda da área médica.

Para eles, questões envolvendo a fraseologia especializada se tornam presentes na relação entre as práticas terminológica e tradutória, fazendo com que

o tradutor fique diante de uma questão semântica relevante, visto que as fraseologias nem sempre dependerão do somatório dos elementos que as constituem. Nesse sentido, "evidencia-se que a Tradução, contrariando o que tradicionalmente se julgava, não se resume a um facilitado e mecânico processo de transposição de mensagens" (KRIEGGER; SANTIAGO, 2014, p. 45).

Na verdade, o fato é que há poucos estudos descritivos sobre isso, assim como há pouco material de base para trabalhos sobre a tradução técnica e científica. Assim, é na terminologia que os tradutores têm se valido de recursos informatizados, realizando uma prática reflexiva para o avanço de pesquisas relacionadas aos textos e termos especializados. Portanto, a relação entre Terminologia e Tradução tem apresentado um caminho bastante produtivo.

Neste sentido, Almeida e Simões (2002, p. 18) também afirmam que um conjunto de terminologias adequadas ao documento a ser traduzido é de grande valia para uma boa tradução:

Determinados termos, em algumas áreas, têm traduções que pelo seu uso durante vários anos, e por várias instituições, ficaram cristalizadas e são as traduções oficiais. Nestes casos, os tradutores não devem tentar fazer a sua própria tradução, mesmo que julgue que a sua seria melhor. Como estes termos e respectivas traduções oficiais não aparecem em dicionários, o tradutor deve ter um conjunto de bases terminológicas, para cada uma das áreas em que está a trabalhar.

Segundo os autores, as bases terminológicas podem ser disponibilizadas pelo cliente ou podem ser construídas e mantidas pelos tradutores por meio de documentos disponíveis, como, por exemplo, dicionários, enciclopédias, internet ou, até mesmo, por conversas com técnicos da área em que se vai traduzir.

Nesse contexto, além da utilização das bases terminológicas, Bevilacqua e Kilian (2017) acreditam que a formação em Terminologia contribui para o desenvolvimento da competência tradutória e de suas subcompetências, visto que o texto passará a ser traduzido de forma mais especializada. Segundo as autoras,

[...] a Terminologia auxilia na aquisição da competência tradutória e de suas subcompetências, principalmente no que se refere à construção de obras terminográficas (glossários, dicionários, bases de dados) e de mapas conceituais, bem como a avaliação desse tipo de recursos a serem utilizados na tradução (BEVILACQUA; KILIAN, 2017, p.1708).

Nesse sentido, elas destacam a importância do conhecimento sobre o funcionamento das terminologias e seus princípios de organização durante a solução de problemas terminológicos ao longo do processo tradutório, pois, dominando esse conhecimento, o tradutor será capaz de atuar com mais segurança e rapidez.

Dessa forma, conclui-se que o uso da terminologia permite aos tradutores conhecer melhor o funcionamento dos termos e das fraseologias especializadas nos contextos em que estão inseridos, assegurando, assim, que o texto traduzido se adeque ainda mais ao ponto de vista linguístico e especializado.

2.3. O gênero resumo / *abstract*

Resumir textos é algo necessário para o desempenho de várias funções e atividades. O ato de resumir é bastante solicitado no meio acadêmico e, muitas vezes, está atrelado a questões avaliativas como comprovação de compreensão leitora, segundo Bragagnollo (2011). Além disso, Bernardino e Valentim (2016) alegam que os gêneros discursivos vêm sendo reconhecidos nas universidades como “espaços de interação social, de negociação entre pares, de construção de posicionamento e avaliação” (BERNARDINO; VALENTIM, 2016, p. 27 *apud* BERNARDINO, 2007, p. 21).

O resumo acadêmico tem como função a apresentação concisa de pontos relevantes de um texto. Para Bragagnollo (2011), essa função é adquirida quando ele é produzido. Por apresentar objetivos e características diferentes dos resumos produzidos em outros contextos, ele se difere dos demais. Além disso, apresenta um caráter avaliativo e busca, de forma clara, explicitar a compreensão global do texto lido e suas principais considerações.

Por apresentar como característica a fidelidade às ideias expressas no texto-fonte, durante a elaboração de um resumo, quem o escreve deve conhecer as seguintes etapas, segundo Bragagnollo (2011, p. 42 *apud* THEREZO, 2007, p. 117-118):

1. compreender claramente o conteúdo do texto a resumir, de modo a poder fazer escolhas: deixar de lado o acidental (detalhes, explicações, exemplos) e ficar com o essencial (ideias principais);
2. organizar as ideias fundamentais do texto original num discurso seu, coeso e coerente;
3. ser absolutamente fiel às ideias expressas pelo autor, não acrescentando informações subsidiárias;
4. redigir em nível padrão de linguagem, com vocabulário próprio, sem copiar frases ou expressões (a não ser as absolutamente necessárias).

Para ela, este gênero é fundamentalmente parafrástico, visto que é a partir da paráfrase que ele se constitui. Porém, quem o escreve deve estar atento ao fato de que ele não pode se constituir de colagens de trechos do texto original. É preciso, portanto, retextualizar.

De acordo com Teles (2007, p. 24), o processo de retextualização é entendido como a produção de um novo texto a partir do original. Nesse sentido, a autora, destaca que reescrever e retextualizar:

[...] constituem-se atividades distintas e ainda que envolvam operações lingüísticas semelhantes como acréscimo ou supressão de informações, substituição e reordenação tópica, há um distanciamento no que diz respeito às operações discursivo-textuais, já que, do ponto de vista teórico-metodológico, são duas atividades que não se equivalem.

Assim, diferentemente do processo de sumarização, que se dá de forma natural e inconsciente, a retextualização requer mais esforço cognitivo e o uso de estratégias metacognitivas, para que o leitor consiga produzir outro texto a partir do original.

Embora bastante associado à avaliação, o resumo acadêmico pode apresentar propósitos além da comprovação leitora:

[...] com condições de produção, circulação e recepção definidas, pode funcionar como objeto de ensino para o desenvolvimento de capacidades específicas inerentes à compreensão e produção de textos, além de ser um texto autônomo, de uso em diferentes contextos, presente em várias instâncias sociais, no cotidiano das pessoas (BRAGAGNOLLO, 2011, p. 43).

Dessa forma, a autora conclui que o gênero está presente em nossa vivência desde a escola. Durante o ensino, ele vai adquirindo objetivos específicos em situações concretas de comunicação em que o aluno deve considerar a sua compreensão no trabalho de leitura e escrita.

Em consonância com essas reflexões, Teles (2007, p. 20) argumenta que “resumir textos (ou parte de textos) tornou-se hoje uma atividade freqüente - prática que se intensifica ainda mais na mídia digital, dada a forma com que os textos são construídos para atender às expectativas do leitor-navegador da web”.

Ainda de acordo com Bragagnollo (2011), o primeiro passo a ser tomado para a elaboração de um resumo acadêmico é a sumarização, a qual deve ocorrer durante a leitura mentalmente, possibilitando a redução de informações com a eliminação de informações secundárias e conservando as primárias. Segundo Bragagnollo (2011, p. 44):

A sumarização pode acontecer por meio de duas estratégias básicas: o apagamento das informações desnecessárias à compreensão de outras proposições ou de informações redundantes; e a substituição, que seria uma estratégia de construção de novas proposições.

Para tal, é preciso que quem resume tenha conhecimento de quem é autor do texto original, da sua função social, da imagem que o autor tem de seu destinatário, dos locais e/ou veículos onde possivelmente o texto irá circular, do momento possível da produção e do objetivo do autor do texto.

Tudo isso é fundamental no processo de sumarização para avançar para a próxima etapa da produção, a qual deve conter a compreensão global do texto. Para isso, o reconhecimento de alguns elementos do texto original é indispensável, sendo que, a partir da identificação desses elementos, é que quem o escreve estará pronto para compreender as ideias mais importantes do texto para que consiga fazer com que as diferentes relações existentes entre elas sejam percebidas e compreendidas. Dessa forma, quanto maior for o entendimento, maior será a adequação do resumo, visto que a capacidade para produzi-lo está relacionada à capacidade de compreensão do texto, de acordo com Bragagnollo (2011).

Em relação ao processo de sumarização, Teles (2007) argumenta que esse procedimento tem sido confundido com a definição de resumo. Nos termos da autora, temos o seguinte:

Entretanto, estudos revelam que a noção do que seja resumir não tem se mostrado muito clara nem tão pouco homogênea, seja no contexto social, seja no âmbito acadêmico. Machado, A. R. (2005), ao fazer uma análise comparativa entre o conceito de resumo encontrado em um dicionário e os usos que se faz de informações resumidas na mídia (valendo-se particularmente da revista VEJA), constatou que em torno do que se entende por resumo há uma verdadeira confusão terminológica, podendo significar tanto o processo de sumarização quanto o produto, unidade lingüística resultante desse processo (TELES, 2007, p.21).

A autora ressalta que, embora sumarizar seja um processo natural humano, ele não é simples. A sua realização é inconsciente e se dá por meio de aplicações estratégicas cognitivas que são constantes e interiorizadas. Tais estratégias preservam o conteúdo e a materialidade lingüística do texto, excluindo informações secundárias.

Em relação à etapa seguinte, Bragagnollo (2011) ressalta a importância da menção ao autor do texto original, visto que, às vezes, a autoria de ambos os autores não podem ser confundidas. Outra etapa importante é a atribuição de atos do autor do resumo relacionados à sua menção. Para ela, é preciso interpretar esses atos e recorrer a verbos *dicendi*, ou seja, a *verbos do dizer*, como, por exemplo, *dizer*, *afirmar*, *asseverar* etc.. Por fim, ela afirma que a auto-avaliação é a etapa final do resumo. Nessa etapa, quem o escreve deve revisar os procedimentos com a finalidade de verificar se todos os procedimentos foram realizados e com resultados satisfatórios. É nesse momento que entra a reescrita caso seja necessária. Além disso, de acordo com Bragagnollo (2011, p. 46), deve-se observar a estrutura final do texto da seguinte forma:

[...] o resumo deve ser escrito em um bloco só, sem paragrafação; pode ou não ter palavras-chave, dependendo da situação em que é solicitado e apresentado; é redigido na terceira pessoa do singular, no tempo verbal do Presente do Indicativo; é precedido, normalmente, pela referência bibliográfica da obra, objeto do resumo; a linguagem usada deve ser formal, referencial e com alto grau de objetividade.

Em relação à estrutura composicional do resumo, Teles (2007) apresenta os seguintes princípios: (i) o resumo é um texto coeso e coerente; (ii) o resumo é um texto conciso; (iii) o resumo é um texto formalmente diferente do texto-fonte; (iv) o resumo é um texto fiel ao texto-fonte.

Assim, a autora conclui que o resumo é um texto sobre outro texto, e este texto, além de ter que ser claro, também deve esclarecer de quem é a autoria para que o leitor não pense que as ideias são, de fato, de quem resumiu o texto.

2.4. Conclusões

Partindo das considerações apresentadas, conclui-se que a tradução técnica se difere de todos os outros tipos de tradução, levando em consideração a clareza, a concisão e o rigor e indo além de apenas escrever palavras equivalentes em outra língua, pois a familiaridade com a língua de partida é fundamental para que o tradutor desempenhe um bom papel no processo. No entanto, ele pode se valer das vantagens que as ferramentas de tradução oferecem, desde que ele esteja atento às limitações que elas apresentam.

Tais ferramentas também desempenham um papel importante na relação entre a terminologia e a tradução. Tendo em vista que atualmente é preciso lidar com as terminologias para traduzir textos técnicos, é de suma importância que os tradutores busquem, nesse campo de estudo, fundamentos para explicar a constituição e o funcionamento dos termos. Além disso, com o advento das tecnologias, o tradutor passou a ter acesso a diversos recursos, sendo possível fazer uso deles para acessar conjuntos de textos com diferentes gêneros e temáticas para sanar eventuais dúvidas que possam surgir durante o processo tradutório. Dessa forma, essas ferramentas podem ser usadas para agilizar o trabalho dos tradutores. Assim, o uso da terminologia, durante o processo tradutório, permite ao tradutor conhecer melhor o funcionamento dos termos e das fraseologias especializadas em determinados contextos.

Nesse sentido, por ser um gênero bastante presente no dia-a-dia das pessoas, o gênero resumo, o qual é muito requisitado no meio acadêmico e em

diferentes esferas sociais, se fez bastante pertinente neste estudo. Assim, entender os processos que o compõem, como, por exemplo, os processos de retextualização e sumarização, e não confundí-los como sinônimos, é tão imprescindível quanto reconhecer seus aspectos.

CAPÍTULO III

RESUMO X ABSTRACT:

O PAPEL DA LINGUÍSTICA DE *CORPUS* NA TRADUÇÃO TÉCNICA

Este capítulo dedica-se à análise da tradução dos resumos acadêmicos da Revista Veredas, produzidos originalmente na língua portuguesa e traduzidos para a língua inglesa, a partir da aplicação do programa *Antconc* no âmbito da Linguística de *Corpus*. Nesse contexto, serão aplicadas, de modo especial, as seguintes ferramentas que constituem o programa: *Wordlist*, *Concordance* e *File View*.

Em outras palavras, este capítulo visa, por meio das considerações de Cunha Lacerda (2016), de Schiffrin (1987) e de Traugott e Trousdale (2013) e do equacionamento entre a metodologia qualitativa e quantitativa, a evidenciar o uso do programa *Antconc* durante a tradução particular de resumos acadêmicos. Além disso, busca demonstrar a pertinência do uso das ferramentas do programa *Antconc*, visto que elas permitem refinar a tradução, especialmente no que tange à tradução técnica.

A fim de cumprir o objetivo geral deste capítulo, ele se organiza da seguinte maneira: em um primeiro momento, discute-se a metodologia de pesquisa assumida; em um segundo momento, apresenta-se a análise propriamente dita a partir das três ferramentas do programa *Antconc* citadas acima; e, por fim, apresenta-se um encaminhamento que visa a ratificar a pertinência da interface entre a Linguística de *Corpus* e os Estudos da Tradução - principalmente, no que diz respeito à Tradução Técnica.

3.1. Procedimentos metodológicos

Com o objetivo de evidenciar as contribuições do programa *Antconc* durante a tradução técnica, este trabalho tomou como base o método misto de pesquisa, o qual se pauta no equacionamento entre a metodologia qualitativa e quantitativa, se

apoiando nas contribuições de Cunha Lacerda (2016), de Schiffrin (1987) e de Traugott e Trousdale (2013).

Segundo Cunha Lacerda (2016, p. 85), o método misto “pode trazer importantes evidências empíricas no que se refere à ocorrência específica da construcionalização”. Tal método consiste em combinar elementos da abordagem qualitativa a elementos da abordagem quantitativa na busca de aumentar e aprofundar o conhecimento a respeito do objeto de análise. Assim sendo, este trabalho apresenta um levantamento e uma análise dos dados por meio de técnicas estatísticas, de cunho quantitativo, assim como busca uma descrição detalhada do objeto investigado, partindo do contexto em que é instanciado, o que configura a utilização da metodologia qualitativa.

As referidas contribuições de Cunha Lacerda (2016) vão ao encontro das considerações de Schiffrin (1987) na defesa de que o método misto de análise permite que o pesquisador obtenha mais ocorrências e possa analisar, de forma mais adequada, os padrões que são formados, podendo, assim, ganhar novas elaborações e adequações conforme os dados vão sendo observados.

Da mesma forma, Traugott e Trousdale (2013, p. 238), ao tratarem do modo de organização hierárquico de construções da língua, sinalizam que o uso do método misto é pertinente durante a análise de trabalhos no âmbito da mudança linguística. Os autores argumentam que ao considerarem

[...] as abordagens qualitativa e quantitativa como sendo complementares para o trabalho na linguística histórica e prevemos a possibilidade de unir as duas abordagens em estudos de mudança linguística em curso, em que a análise da microvariação no nível individual dos falantes poderia ser combinada com a análise quantitativa da macrovariação no nível do grupo social. (TRAUGOTT; TROUSDALE, 2013, p. 238, *apud* CUNHA LACERDA, 2016, p. 87-88).

Assim, para eles, considerar as abordagens qualitativa e quantitativa como complementares permite uma abordagem mais refinada, estabelecendo uma relação entre frequência e entrenchamento e o grau de abstração a partir da qual os grupos de falantes parecem organizar aspectos de seu conhecimento linguístico.

Nesse sentido, o uso do programa *Antconc* por permitir, durante a tradução, a realização das principais técnicas de Linguística de *Corpus*, como, por exemplo, frequência de palavras, colocações, concordância, *n-grams* e comparação de *corpora* a qualquer tipo de texto, por meio de uma abordagem tanto quantitativa quanto qualitativa, se faz bastante pertinente por admitir uma nova maneira de analisar e descrever os usos da língua, assim como observar fenômenos na área da tradução, o que é exatamente o que busca a Linguística de *Corpus*: se ocupar da coleta e da análise de *corpora* a fim de servirem para pesquisar a língua e suas variedades (SARDINHA, 2000).

Dessa forma, para análise dos dados, textos foram coletados e compilados em dois arquivos com a finalidade de traçar uma comparação entre o *corpus* do português e o *corpus* do inglês. A escolha se deu por textos acadêmicos, mais especificamente, resumos de artigos científicos e seus respectivos *abstracts*. Tais resumos/*abstracts* foram selecionados a partir da Revista Veredas, a qual é um periódico de estudos linguísticos da Universidade Federal de Juiz de Fora, que tem a missão de publicar artigos científicos, ensaios, traduções e resenhas. A escolha da revista se deu pelo fato de os textos publicados em suas edições serem inéditos e reconhecidos por apresentarem qualidade acadêmica na área de Linguística, teórica e aplicada. Após a escolha da revista e a seleção dos resumos/*abstracts* de artigos publicados nos últimos cinco anos, os textos foram compilados em dois arquivos em formato txt. e incorporados ao programa *Antconc* a fim de compor a análise descrita na próxima seção.

Desta forma, com base na análise a seguir e nos procedimentos de análise, partindo da aplicação do programa *Antconc* sob um viés qualitativo e quantitativo, busca-se comprovar empiricamente, por meio da análise da tradução de resumos da Revista Veredas para a língua inglesa, que a Linguística de *Corpus* pode contribuir, de modo substancial, para o refinamento de trabalhos realizados no âmbito da Tradução Técnica. Em outras palavras, busca-se estabelecer, no campo da Tradução Técnica - de modo mais específico, no que tange a um trabalho com resumo/*abstract* -, o papel que a Linguística de *Corpus* pode ter no processo tradutório.

3.2. Análise

A fim de demonstrar a importância da interface entre a Linguística de *Corpus* e a área dos Estudos da Tradução, esta análise está calcada em três das sete ferramentas que compõem o programa *Antconc*, a saber: a ferramenta *Wordlist*, a ferramenta *Concordance* e a ferramenta *File View*.

A escolha da ferramenta *Wordlist* se deu em virtude de ela desempenhar funções como selecionar os nomes mais frequentes no *corpus*, assim como sua posição e número de ocorrências. Já a escolha da ferramenta *Concordance* se deu por ela possibilitar a visualização e a análise de como as construções foram utilizadas no *corpus* original e no *corpus* traduzido a partir de sua distribuição no contexto. Por fim, a escolha da ferramenta *File View* se deu por ela possibilitar a verificação dos nomes dentro do texto bruto dos arquivos, viabilizando, assim, a observação das duas primeiras ocorrências de cada nome e sua análise dentro do texto bruto.

Dessa forma, analisar os dados a partir dessas ferramentas viabiliza investigar as escolhas tradutórias realizadas, ratificando, assim, a interface possível entre a Linguística de *Corpus* e os Estudos da Tradução, no que tange principalmente à tradução como um todo e, mais particularmente, à Tradução Técnica.

3.2.1. Aplicação da ferramenta *Wordlist*

Durante a aplicação da ferramenta *Wordlist*, o critério utilizado foi a busca de como se realiza a tradução técnica dos dez nomes mais frequentes no *corpus* compilado, os quais podem ser observados no quadro a seguir, juntamente com a posição em que se deu e o número de ocorrências.

Dez nomes mais frequentes no corpus em português		
Posição	Nº de ocorrências	Nome
20	106	ensino
25	93	trabalho
26	89	análise
28	85	artigo
29	85	resultados
32	78	língua
35	72	objetivo
36	59	português
37	56	pesquisa
38	53	dados

Tabela 1 - Os dez nomes mais frequentes no *corpus* em português
 Fonte: Elaborada pela autora (2023)

A fim de traçar uma comparação entre o *corpus* do português e o *corpus* do inglês, observou-se também a posição, o número de ocorrências e os dez primeiros nomes presentes do *corpus* do inglês. Durante a análise, foram obtidos os seguintes resultados:

Dez nomes mais frequentes no corpus em inglês		
Posição	Nº de ocorrências	Nome
16	126	language
21	89	results
22	89	analysis
25	83	portuguese
28	68	research
29	68	students
30	65	work
32	61	article
34	60	study
36	59	linguistic

Tabela 2 - Os dez nomes mais frequentes no *corpus* em inglês
 Fonte: Elaborada pela autora (2023)

Tais resultados podem ser comprovados por meio das capturas das telas a seguir:

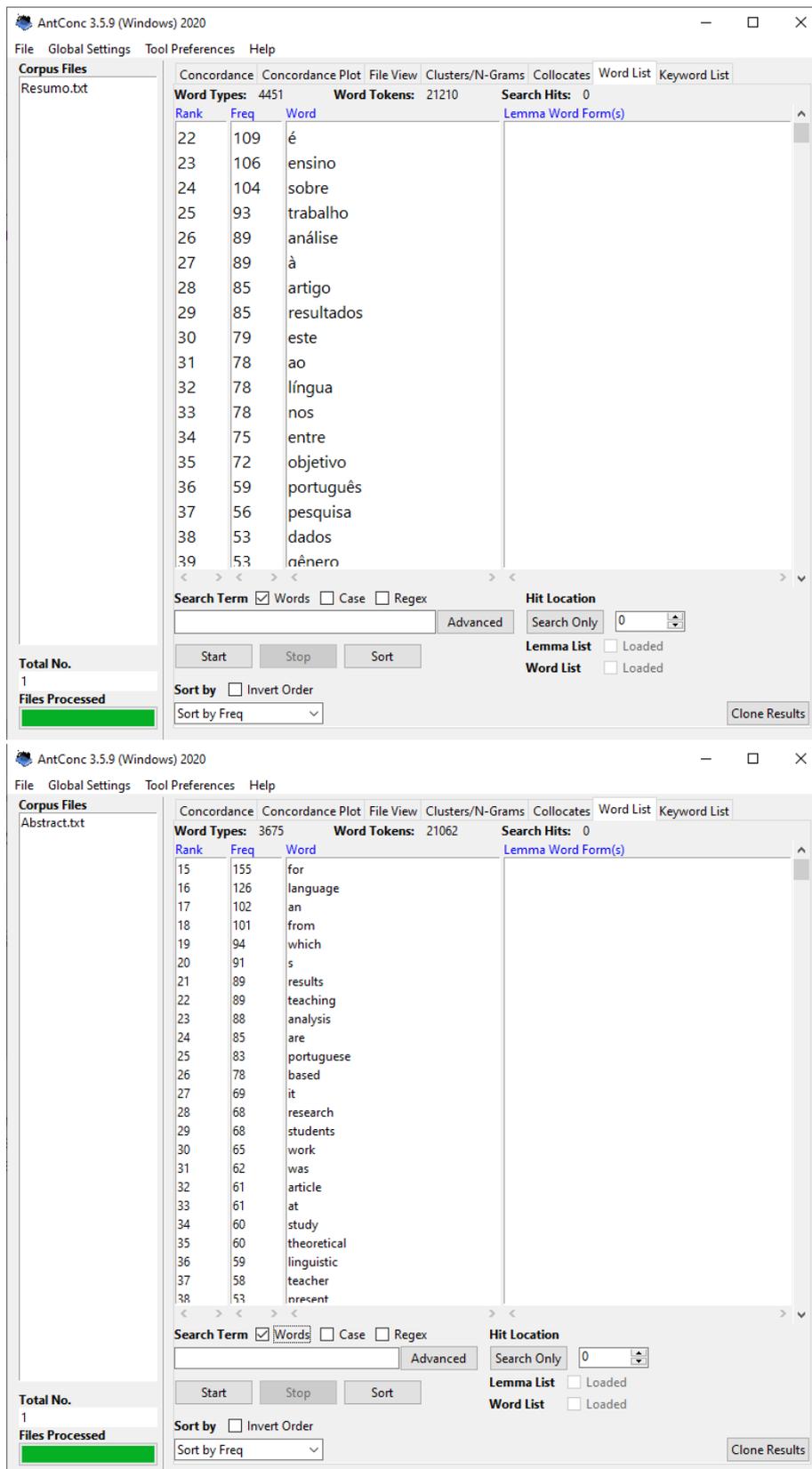


Figura 11 – Resultados dos dez primeiros nomes nos *corpora* distribuídos pelo critério *Sort by Freq*
 Fonte: imagem retirada do programa *Antconc*

Desta forma, o uso da ferramenta *Wordlist* pelo critério *Sort by Freq* possibilitou encontrar rapidamente os nomes mais frequentes, assim como propiciou observar suas posições e o número de ocorrências em ambos os *corpora*. Por meio desses resultados, também foi possível observar que os nomes mais frequentes no *corpus* do português nem sempre aparecem no *corpus* em inglês e vice-versa, ou seja, além de apresentarem posições e número de ocorrências diferentes, alguns termos não se mantêm nos dois *corpora*.

A partir dessas observações, fez-se necessário analisar como as construções foram utilizadas no *corpus* a contar da distribuição no contexto, conforme descrevem Cunha Lacerda e Dall’Orto (no prelo). Tal análise será apresentada na próxima seção, considerando o uso da ferramenta *Concordance*.

3.2.2. Aplicação da ferramenta Concordance

Considerando o elevado número de ocorrências para cada nome, nesta seção, serão observadas as duas primeiras ocorrências de cada nome em particular no *corpus*. Assim sendo, apresentam-se, a seguir, imagens da tela com a aplicação da ferramenta *Concordance* a fim de demonstrar como se deu a análise, assim como apresentar um quadro com os nomes referidos e as suas duas primeiras ocorrências no *corpus* do português e no *corpus* do inglês. Inclusive, essas ocorrências serão detidamente analisadas mais adiante a partir da apresentação do quadro 2, o qual evidenciará os nomes dentro do texto bruto dos arquivos representados por meio da ferramenta *File View*.

Após clicar no termo “ensino” no *corpus* do português e em “teaching” no *corpus* em inglês por meio da ferramenta *WordList*, o programa foi direcionado para os resultados da ferramenta *Concordance*, como é possível observar na figura a seguir:

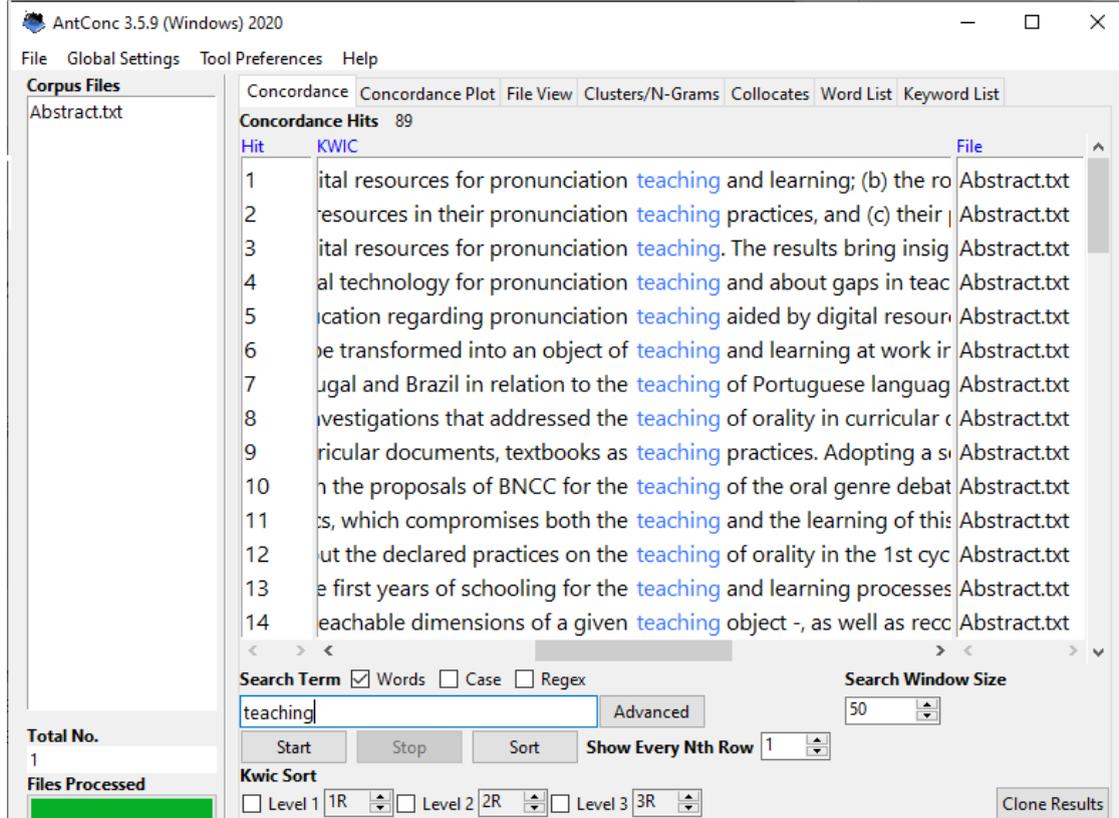
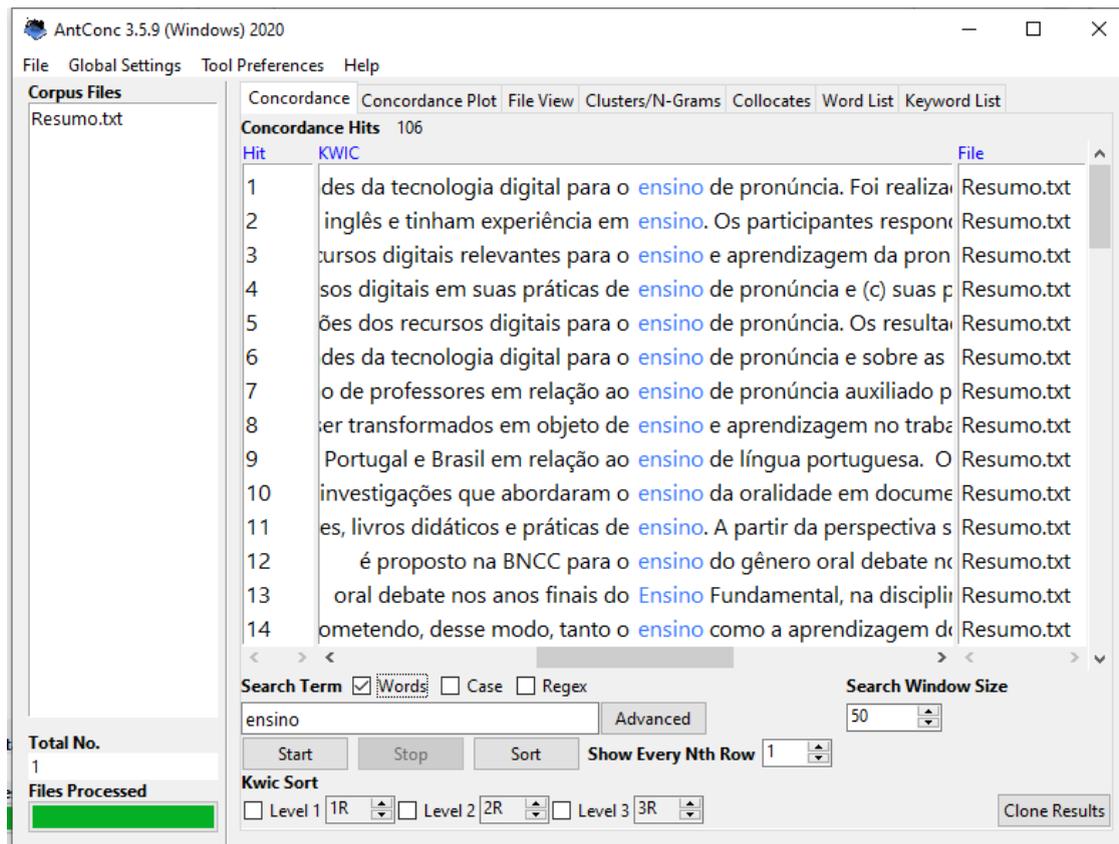


Figura 12 – Contextos relacionados aos termos “ensino” e “teaching” por meio da ferramenta Concordance

Fonte: imagem retirada do programa Antconc

Após observados os resultados, foram selecionados os dois primeiros contextos de ambos os nomes a fim de analisar e comparar seus usos dentro do contexto. Na sequência, o programa abriu as telas da ferramenta *File View*, as quais serão apresentadas na próxima seção por meio da figura 13.

Com o direcionamento da tela para a ferramenta *File View*, foi possível verificar os nomes dentro do texto bruto dos arquivos. Após execução desse passo a passo com todos os dez nomes selecionados em cada *corpus*, foram elaborados os seguintes quadros para análise:

Nome	Contexto		
	Nº da ocorrência	Português	Inglês
ensino	106	Este estudo examina a visão de professores brasileiros de inglês sobre as possibilidades da tecnologia digital para o ensino de pronúncia.	This study examines the view of Brazilian teachers of English on the affordances of digital technology to teach pronunciation.
		Todos eles possuíam graduação em um programa de formação de professores de inglês e tinham experiência em ensino .	All of them held an undergraduate degree from an English Teacher education program.
trabalho	93	Ao comungarmos com a assertiva de que nenhum enunciado é unimodal, este trabalho tem como objetivo investigar quais são as multimodalidades que constituem o gênero textual oral debate.	By sharing the assertive that no statement is unimodal, this work aims to investigate which are the multimodalities that constitute the oral textual genre debate.
		Os resultados demonstram quais são as diferentes linguagens que formam o debate e os elementos que planificam a estrutura linguístico-discursiva do gênero, os quais devem ser transformados em objeto de ensino e aprendizagem no trabalho em sala de aula.	The results demonstrate the different languages that form the debate and the elements that plan the linguistic-discursive structure of the genre, which must be transformed into an object of teaching and learning at work in the classroom.
análise	89	As análises da fala em interação foram subsidiadas teóricometodologicamente sobretudo por estudos da Análise da Conversa, contando também com contribuições advindas do campo da Pragmática sobre dêixis.	The analysis of talk in interaction was theoretically and methodologically supported mainly by studies on Conversation Analysis , and it also counted on contributions from the field of Pragmatics on deixis.

		Inspiradas pelo paradigma qualitativo de pesquisa, usamos os recursos do Subsistema da Atitude para o desenvolvimento de uma análise macro e micro discursiva acerca dos relatos avaliativos da participante.	Inspired by the qualitative research paradigm, we used the resources of the Attitude Subsystem to develop a macro and micro discursive analysis of the participant's evaluative reports.
artigo	85	Neste artigo analisamos o parâmetro do movimento nos verbos de concordância reversa na Língua de Sinais Brasileira (LSB), a fim de explicar a diferença morfossintática entre esses verbos e os demais verbos com concordância (STROBEL; FERNANDES, 1998; QUADROS; KARNOPP, 2004; LOURENÇO; DUARTE, 2014; entre outros).	In this article , we analyze the movement parameter of the so-called backward verbs in Brazilian Sign Language (BSL), in order to explain the morphosyntactic difference between these verbs and the other agreement verbs. (cf. STROBEL; FERNANDES, 1998; QUADROS; KARNOPP, 2004; LOURENÇO; DUARTE, 2014, among others).
		Este artigo se propôs a investigar o uso dêixis por um afásico a fim de compreender o gerenciamento da intersubjetividade na interação no contexto institucional da telefonaudiologia.	This article aims to investigate the use of deixis by an aphasic patient in order to understand the management of intersubjectivity in interaction in the institutional context of speech-language teletherapy.
resultados	85	Os resultados trazem insights sobre as possibilidades da tecnologia digital para o ensino de pronúncia e sobre as lacunas na formação de professores em relação ao ensino de pronúncia auxiliado por recursos digitais.	The results bring insights about the affordances of digital technology for pronunciation teaching and about gaps in teacher education regarding pronunciation teaching aided by digital resources.
		Os resultados demonstram quais são as diferentes linguagens que formam o debate e os elementos que planejam a estrutura linguístico-discursiva do gênero, os quais devem ser transformados em objeto de ensino e aprendizagem no trabalho em sala de aula.	The results demonstrate the different languages that form the debate and the elements that plan the linguistic-discursive structure of the genre, which must be transformed into an object of teaching and learning at work in the classroom.
língua	78	Neste artigo analisamos o parâmetro do movimento nos verbos de concordância reversa na Língua de Sinais Brasileira (LSB), a fim de explicar a diferença morfossintática entre esses verbos e os demais verbos com concordância (STROBEL; FERNANDES, 1998; QUADROS; KARNOPP, 2004; LOURENÇO; DUARTE, 2014; entre outros).	In this article, we analyze the movement parameter of the so-called backward verbs in Brazilian Sign Language (BSL), in order to explain the morphosyntactic difference between these verbs and the other agreement verbs. (cf. STROBEL; FERNANDES, 1998; QUADROS; KARNOPP, 2004; LOURENÇO; DUARTE, 2014, among others).

		Considerações históricas a respeito do contexto situacional são importantes, pois, segundo a Teoria Semi linguística, há íntima relação entre o uso da língua , o contrato comunicativo, os dados situacionais e a performance argumentativa dos enunciadores no discurso.	Some historical reflections about the context of the sermons are considered important, because to the Semi linguistics theory there is a relation among language , the “contract” of communication, the situation and the performance of the subjects in discourse.
português	59	Propomos que, enquanto no português as funções gramaticais são cognitivamente mais proeminentes para a marcação de concordância morfossintática, na LSB o elemento cognitivo mais proeminente é o argumento deslocado.	Our proposal is that, while in Portuguese grammatical functions are cognitively more prominent for the purposes of morphosyntactic agreement marking, in BSL the most prominent cognitive element is the displaced argument.
		A partir disso, pretendeu-se compreender como a inserção de Timor-Leste na Comunidade Lusófona colaborou para o enraizamento linguístico do português no território timorense.	From this, it was intended to understand how the insertion of Timor-Leste in the Lusophone Community contributed to the linguistic rooting of Portuguese in the Timorese territory.
pesquisa	56	Foi realizada uma pesquisa online com um grupo de 42 professores.	An online survey was conducted with a group of 42 in-service teachers.
		Inspiradas pelo paradigma qualitativo de pesquisa , usamos os recursos do Subsistema da Atitude para o desenvolvimento de uma análise macro e micro discursiva acerca dos relatos avaliativos da participante.	Inspired by the qualitative research paradigm, we used the resources of the Attitude Subsystem to develop a macro and micro discursive analysis of the participant's evaluative reports.
dados	53	Os dados naturalísticos, gravados em vídeos, foram transcritos de acordo com convenções sugeridas por autores da área e analisados sequencialmente.	The naturalistic data , recorded on videos, were transcribed according to the conventions suggested by authors in the field and analysed sequentially.
		Para tanto, além de uma revisão das características fonéticas de CCV, o estudo traz dados de fala de 71 crianças entre 2;0-5;11 anos, comparando a produção de líquidas e oclusivas em CCV e em contexto-controle CV.	We conducted a review on the phonetic characteristics of CCV and an experimental study with 71 children aged 2;0-5;11 years. The study compared the production of liquids and obstruents in CCV and in context-control CV.
objetivo	72	Ao comungarmos com a assertiva de que nenhum enunciado é unimodal, este trabalho tem como objetivo investigar quais são as multimodalidades que constituem o gênero textual oral debate.	By sharing the assertive that no statement is unimodal, this work aims to investigate which are the multimodalities that constitute the oral textual genre debate.

		Para alcançar o objetivo pretendido, adotou-se a pesquisa bibliográfica, respaldando-se em materiais tais como artigos científicos e livros, a fim de catalogar bases teóricas que pudessem sustentar as questões levantadas.	To achieve the intended objective , bibliographic research was adopted, supported by materials, such as scientific articles and books, in order to catalog theoretical bases that could support the issues raised.
--	--	--	---

Quadro 2 - Comparação dos termos mais frequentes do português para o inglês
Fonte: Elaborada pela autora (2023)

Nome	Contexto		
	Nº da ocorrência	Português	Inglês
teaching	89	(a) o que os participantes reconhecem como recursos digitais relevantes para o ensino e aprendizagem da pronúncia;	(a) what the participants recognize as relevant digital resources for pronunciation teaching and learning;
		(b) o papel desempenhado pelos recursos digitais em suas práticas de ensino de pronúncia,	(b) the role played by digital resources in their pronunciation teaching practices,
work	65	Ao comungarmos com a assertiva de que nenhum enunciado é unimodal, este trabalho tem como objetivo investigar quais são as multimodalidades que constituem o gênero textual oral debate.	By sharing the assertive that no statement is unimodal, this work aims to investigate which are the multimodalities that constitute the oral textual genre debate.
		Os resultados demonstram quais são as diferentes linguagens que formam o debate e os elementos que planificam a estrutura linguístico-discursiva do gênero, os quais devem ser transformados em objeto de ensino e aprendizagem no trabalho em sala de aula.	The results demonstrate the different languages that form the debate and the elements that plan the linguistic-discursive structure of the genre, which must be transformed into an object of teaching and learning at work in the classroom.
analysis	88	As análises da fala em interação foram subsidiadas teóricometodologicamente sobretudo por estudos da Análise da Conversa, contando também com contribuições advindas do campo da Pragmática sobre dêixis.	The analysis of talk in interaction was theoretically and methodologically supported mainly by studies on Conversation Analysis, and it also counted on contributions from the field of Pragmatics on deixis.

		As análises da fala em interação foram subsidiadas teóricometodologicamente sobretudo por estudos da Análise da Conversa, contando também com contribuições advindas do campo da Pragmática sobre dêixis.	The analysis of talk in interaction was theoretically and methodologically supported mainly by studies on Conversation Analysis , and it also counted on contributions from the field of Pragmatics on deixis.
article	61	Neste artigo analisamos o parâmetro do movimento nos verbos de concordância reversa na Língua de Sinais Brasileira (LSB), a fim de explicar a diferença morfossintática entre esses verbos e os demais verbos com concordância (STROBEL; FERNANDES, 1998; QUADROS; KARNOPP, 2004; LOURENÇO; DUARTE, 2014; entre outros).	In this article , we analyze the movement parameter of the so-called backward verbs in Brazilian Sign Language (BSL), in order to explain the morphosyntactic difference between these verbs and the other agreement verbs. (cf. STROBEL; FERNANDES, 1998; QUADROS; KARNOPP, 2004; LOURENÇO; DUARTE, 2014, among others).
		Este artigo se propôs a investigar o uso dêixis por um afásico a fim de compreender o gerenciamento da intersubjetividade na interação no contexto institucional da telefonaudiologia.	This article aims to investigate the use of deixis by an aphasic patient in order to understand the management of intersubjectivity in interaction in the institutional context of speech-language teletherapy.
results	89	Os resultados trazem insights sobre as possibilidades da tecnologia digital para o ensino de pronúncia e sobre as lacunas na formação de professores em relação ao ensino de pronúncia auxiliado por recursos digitais.	The results bring insights about the affordances of digital technology for pronunciation teaching and about gaps in teacher education regarding pronunciation teaching aided by digital resources.
		Os resultados demonstram quais são as diferentes linguagens que formam o debate e os elementos que planificam a estrutura linguístico-discursiva do gênero, os quais devem ser transformados em objeto de ensino e aprendizagem no trabalho em sala de aula.	The results demonstrate the different languages that form the debate and the elements that plan the linguistic-discursive structure of the genre, which must be transformed into an object of teaching and learning at work in the classroom.
language	126	Neste artigo analisamos o parâmetro do movimento nos verbos de concordância reversa na Língua de Sinais Brasileira (LSB), a fim de explicar a diferença morfossintática entre esses verbos e os demais verbos com concordância (STROBEL; FERNANDES, 1998; QUADROS; KARNOPP, 2004; LOURENÇO; DUARTE, 2014; entre outros).	In this article, we analyze the movement parameter of the so-called backward verbs in Brazilian Sign Language (BSL), in order to explain the morphosyntactic difference between these verbs and the other agreement verbs. (cf. STROBEL; FERNANDES, 1998; QUADROS; KARNOPP, 2004; LOURENÇO; DUARTE, 2014, among others).

		Considerações históricas a respeito do contexto situacional são importantes, pois, segundo a Teoria Semiollingüística, há íntima relação entre o uso da língua , o contrato comunicativo, os dados situacionais e a performance argumentativa dos enunciadores no discurso.	Some historical reflections about the context of the sermons are considered important, because to the Semiollingüistic theory there is a relation among language , the “contract” of communication, the situation and the performance of the subjects in discourse.
portuguese	83	Propomos que, enquanto no português as funções gramaticais são cognitivamente mais proeminentes para a marcação de concordância morfossintática, na LSB o elemento cognitivo mais proeminente é o argumento deslocado.	Our proposal is that, while in Portuguese grammatical functions are cognitively more prominent for the purposes of morphosyntactic agreement marking, in BSL the most prominent cognitive element is the displaced argument.
		Este artigo, desenvolvido a partir de um panorama histórico da língua portuguesa no território timorense, realiza uma breve retomada do antes e do depois da inserção de Timor na Comunidade Lusófona.	This article, developed under a historical overview of the Portuguese language in the Timorese territory, makes a brief review of the before and after the insertion of Timor as part of the Lusophone Community.
research	68	Inspiradas pelo paradigma qualitativo de pesquisa , usamos os recursos do Subsistema da Atitude para o desenvolvimento de uma análise macro e micro discursiva acerca dos relatos avaliativos da participante.	Inspired by the qualitative research paradigm, we used the resources of the Attitude Subsystem to develop a macro and micro discursive analysis of the participant’s evaluative reports.
		Para alcançar o objetivo pretendido, adotou-se a pesquisa bibliográfica, respaldando-se em materiais tais como artigos científicos e livros, a fim de catalogar bases teóricas que pudessem sustentar as questões levantadas.	To achieve the intended objective, bibliographic research was adopted, supported by materials, such as scientific articles and books, in order to catalog theoretical bases that could support the issues raised.
study	60	Este estudo examina a visão de professores brasileiros de inglês sobre as possibilidades da tecnologia digital para o ensino de pronúncia.	This study examines the view of Brazilian teachers of English on the affordances of digital technology to teach pronunciation.
		O presente estudo propõe uma reflexão acerca de aspectos de acessibilidade, a partir de uma conversa exploratória com a participante Rebeca, na qual emergiram desafios por ela enfrentados professora com deficiência visual.	The present study proposes a reflection on aspects of accessibility, based on an exploratory conversation with the participant Rebeca, in which challenges faced by a teacher with vision impairment emerged.

linguistic	59	Os resultados demonstram quais são as diferentes linguagens que formam o debate e os elementos que planificam a estrutura linguístico -discursiva do gênero, os quais devem ser transformados em objeto de ensino e aprendizagem no trabalho em sala de aula.	The results demonstrate the different languages that form the debate and the elements that plan the linguistic -discursive structure of the genre, which must be transformed into an object of teaching and learning at work in the classroom.
		Neste artigo, considera-se a argumentação como uma atividade linguístico -discursiva realizada em determinada situação, com finalidade persuasiva do sujeito enunciator sobre o outro na interação.	This article considers argumentation as a linguistic -discursive activity in a specific situation, that has the objective of getting a persuasive force upon the subject of the interaction.

Quadro 3 - Comparação dos termos mais frequentes do inglês para o português
Fonte: Elaborada pela autora (2023)

Ao comparar os quadros 2 e 3, é possível perceber que a maioria dos 10 primeiros nomes do português apresentam seus correspondentes em inglês dentre os 10 primeiros nomes do inglês, exceto pelos termos "dados" e "objetivo", cujas correspondências perderam posição para os termos "study" e "linguistic". Também é possível observar que os termos apresentam posições de ocorrência diferentes, assim como a frequência de termos apresentados, como demonstrado na seção anterior. No quadro 2, representativo do português, ao observar o termo "ensino", por exemplo, percebe-se que o *hit*¹¹ da palavra, dentro da ferramenta *Concordance*, é de 1 para a primeira ocorrência e 2 para a segunda ocorrência. Já na tradução em inglês, o *hit* correspondente à primeira ocorrência em português se deu em 1, ou seja, consistiu na primeira ocorrência do termo "teach". No entanto, durante a tradução da segunda ocorrência, observa-se que não foi utilizado um termo correspondente, havendo uma reestruturação da frase, o que trouxe a omissão de alguns termos, dentre eles, o termo correspondente a "ensino", presente no original. Semelhantemente a esta ocorrência, neste mesmo quadro, houve a omissão do termo "dados", visto que, durante a primeira ocorrência, a tradução se deu com o uso do termo "data" e, na segunda ocorrência, o termo foi omitido. Casos de omissão não foram observados no quadro 3, representativo do inglês.

¹¹ Neste trabalho, nos termos da Linguística de *Corpus*, compreende-se o termo *hit* como sendo representativo do número de ocorrências identificado.

Ainda no quadro 2, notamos que termos correspondentes também compartilham posições semelhantes no *hit*, como é o caso dos nomes "trabalho" e "work", "artigo" e "article", "resultados" e "results", "língua" e "language" e "português" e "Portuguese". O mesmo ocorre com os termos "work" e "trabalho", "article" e "artigo", "results" e "resultados", "Portuguese" e "português", "study" e "estudo" e "linguistic" e "linguístico" no quadro 3, representativo do inglês.

Em ambos os quadros, observa-se também que os tradutores optaram por termos diferentes ao traduzirem uma determinada palavra. Este é o caso do nome "pesquisa", o qual foi traduzido para o inglês de duas formas: "survey", com o total de 5 *hits*, e "research", com o total de 68 *hits*. O mesmo ocorreu com o termo "objetivo", já que, nas traduções para o inglês, constam os termos "aims to" e "objective" como correspondentes, com 40 e 20 *hits*, respectivamente.

No quadro 3, representativo do inglês, não foi observado nenhum uso de termo diferente durante a escolha tradutória dos 10 nomes com maior ocorrência no *corpus* do português. No entanto, houve uma variação maior no que se refere ao *hit* de alguns nomes, como, por exemplo, o termo "teaching", que, no inglês, ocorreu com *hits* 1 e 2, enquanto, no português, tais ocorrências apresentaram *hits* 3 e 4 para o termo "ensino".

Dessa forma, tal análise reforça a importância do uso da ferramenta *Antconc* para o tradutor durante a busca de termos especializados, tendo em vista a rapidez durante a coleta e análise de dados e a observação das ocorrências a fim de optar pela melhor escolha durante a tradução.

3.2.3. Aplicação da ferramenta *File View*

Conforme observado na seção anterior, o uso da ferramenta *File View* se deu em concomitância com a ferramenta *Concordance*, visto que, após observar as duas primeiras ocorrências de cada nome, foi preciso analisá-los dentro do contexto do texto bruto. Tal análise foi possibilitada por meio da ferramenta *File View*. Assim, após clicar nos termos selecionados na ferramenta *Concordance*, o programa abriu as seguintes telas na ferramenta *File View*:

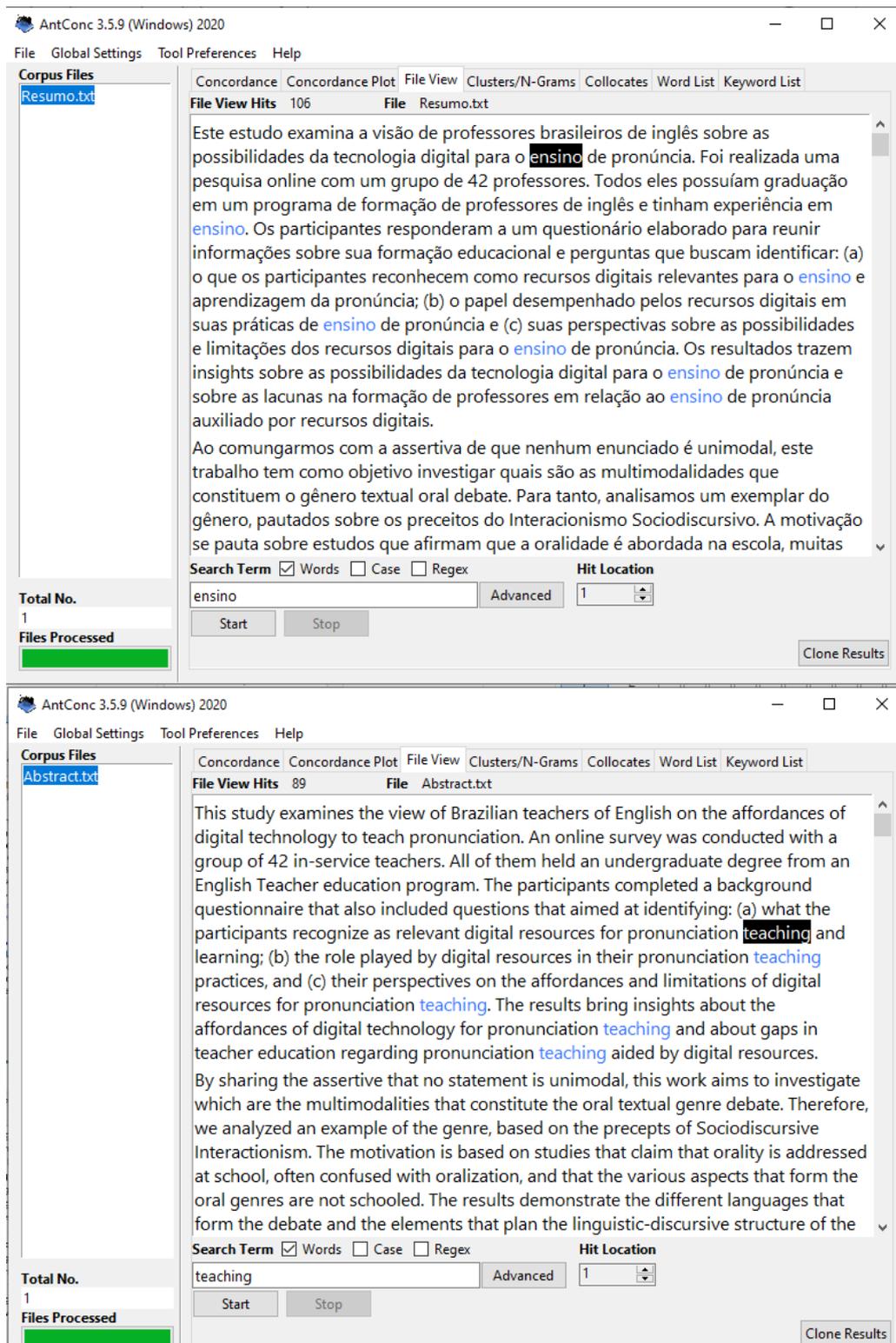


Figura 13 – Nomes dentro do texto bruto dos arquivos representados através da ferramenta *File View*

Fonte: imagem retirada do programa *Antconc*

Como observado, com o direcionamento da tela para a ferramenta *File View*, foi possível verificar os nomes dentro do texto bruto dos arquivos. Nesse sentido,

com a utilização da ferramenta, foi possível visualizar, no texto bruto de arquivos individuais, como se deu o uso de cada nome. Tal visualização favoreceu uma investigação mais detalhada em relação aos resultados dos produtos da ferramenta *Concordance*.

Dessa forma, é possível concluir que, ao visualizar de forma mais detalhada o contexto em que cada nome está inserido, o tradutor poderá ser capaz de compreender melhor seu uso e poderá traçar escolhas mais adequadas durante a tradução, possibilitando que o texto seja traduzido de forma mais especializada – o que confirma ainda mais a contribuição do trabalho com terminologias para traduzir textos técnicos.

3.3. Conclusões

Partindo das análises apresentadas, é possível concluir que o uso do programa *Antconc* durante a tradução técnica pode trazer importantes evidências empíricas a fim de servir como facilitador durante a escolha de termos durante a tradução.

Com base na análise dos dados coletados por meio dos textos compilados em dois arquivos, os quais foram apresentados nas seções anteriores e a partir dos quais foi traçada uma comparação entre o *corpus* do português e o *corpus* do inglês, é possível afirmar que a Linguística de *Corpus* detém um papel importante durante o processo tradutório. Tal afirmação se confirma a partir das análises apresentadas por meio da análise das ferramentas *Wordlist*, *Concordance* e *File View*.

Nesse sentido, observou-se que a ferramenta *Wordlist* possibilitou a seleção dos 10 nomes mais frequentes tanto no *corpus* em português como no *corpus* em inglês. Além disso, ela facilitou a observação da posição e do número de ocorrência de cada termo. Já a ferramenta *Concordance* possibilitou a visualização e a análise de como ocorreram as construções do *corpus* original. Não obstante, a ferramenta *File View* possibilitou a visualização e análise dos nomes selecionados dentro do contexto do texto bruto.

Assim sendo, conclui-se que, quando analisados por meio da ferramenta *Antconc*, os dados coletados permitiram observar a viabilização das escolhas tradutórias que foram realizadas. Tal conclusão reafirma o fato de a Linguística de *Corpus* exercer um papel significativo durante a tradução, além de ratificar a possibilidade da interface entre ela e os Estudos da Tradução, no que se refere não somente à tradução como um todo, mas também, de forma significativa, à Tradução Técnica. Além disso, conclui-se que o uso da ferramenta, durante a tradução, permite ao tradutor observar padrões e usos específicos dos termos dentro de contextos a fim de fazer uso deles de forma mais adequada durante a tradução, tornando-a, assim, mais especializada.

CONSIDERAÇÕES FINAIS

A fim de demonstrar a relação presente entre a Linguística de *Corpus* e os Estudos da Tradução, estabelecendo uma relação entre ela e a Tradução Técnica, este trabalho assumiu como objeto de análise a tradução de textos do português para o inglês a partir da seleção de resumos acadêmicos. Tais resumos foram selecionados da *Revista Veredas*, nos últimos 5 anos, a fim de analisar como a Linguística de *Corpus* pode contribuir para a compreensão da tradução técnica, sendo este o objetivo do trabalho. Para atingir tal objetivo, a aplicação se deu por meio do programa *Antconc*, o qual contribuiu substancialmente para a realização e a compreensão da tradução técnica realizada nos resumos/abstracts.

Dessa forma, compreende-se que a aplicação das ferramentas *Wordlist*, *Concordance* e *File View* possibilitou a seleção dos dez nomes mais frequentes em ambos os *corpora*, facilitando a observação da posição e do número de ocorrência de cada termo e permitindo a visualização e a análise de como ocorreram os nomes selecionados no *corpus* original e no *corpus* traduzido a partir do contexto de uso.

Nesse sentido, confirma-se a importância de trabalhar com as terminologias durante a tradução de textos técnicos, tendo em vista que tais ferramentas possibilitam maior acesso a conjuntos de textos, permitindo, a partir deles, fazer uso dos dados e sanar necessidades informativas e adequando ainda mais o texto traduzido. Em outras palavras, a partir dessas ferramentas, é possível conhecer melhor o uso e o funcionamento dos termos e das fraseologias dentro dos contextos de uso, o que possibilita uma escolha mais especializada durante a tradução.

Assim sendo, após a análise e comparação dos dados obtidos, reforça-se a interface entre a Linguística de *Corpus* e a tradução – e, mais especificamente, entre a Linguística de *Corpus* e a Tradução Técnica. Confirma-se também a noção de que a Linguística de *Corpus* coleta e explora *corpora* ou conjunto de dados linguísticos de forma criteriosa a fim de servir a uma pesquisa, permitindo que a coleta e exploração sejam advindas de evidências empíricas extraídas por ferramentas digitais. Dessa forma, a Linguística de *Corpus* se faz bastante útil, tendo em vista a observação dos *corpora* aqui analisados. Nesse sentido, ressalta-se que o programa *Antconc* apresenta aspectos vantajosos e de grande

contribuição para a pesquisa. Dessa maneira, conclui-se que há ainda uma profícua agenda de trabalho, a ser realizada por pesquisas futuras, no que tange à aplicação da Linguística de *Corpus* à área da Tradução Técnica.

REFERÊNCIAS

AUBERT, F. H.. Tradução técnico-científica e terminologia: um ensaio exploratório de uma via de mão dupla. *TradTerm*, 2001, p. 41-52.

ALMEIDA, J. J. D.; SIMÕES, A. M. B.. Ferramentas de Tradução e Terminologia. [Texto base de disciplina]. Departamento de Informática. Universidade do Minho, 2002.

ALUÍSIO, S. M.; ALMEIDA, G. M. de B.. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa lingüística. *Calidoscópico*, São Leopoldo, v. 4, n. 3, p. 156-178, 2006.

BAKER, M. *Corpora in Translation Studies: An Overview and Some Suggestions for Future Research*. London: John Benjamins Publishing Company, 1995.

BAKER, P. *Sociolinguistics and Corpus Linguistics*. Edinburgh: University Press, 2010.

BARREIROS, L. L. S.. O uso de ferramentas computacionais na elaboração do Vocabulário de Eulálio Motta: AntConc e FLEEx. *Feira de Santana*, v. 18, n. 2, p. 216-241, maio-agosto 2017.

BERBER SARDINHA, T. Linguística de Corpus: Histórico e problemática. *DELTA*. Documentação de Estudos em Linguística Teórica e Aplicada, São Paulo, v. 16, n. 2, p. 323-367, 2000.

BERNARDINO, C. G.; VALENTIM, D. L.. Uma breve análise comparativa entre exemplares do gênero textual “resumo acadêmico”. *Entre palavras*, Fortaleza, v. 6, p. 25-45, jan./jun. 2016.

BEVILACQUA, C. R.; KILIAN, C. K. Tradução e Terminologia: relações necessárias e a formação do tradutor. *Domínios de Lingu@gem*, Uberlândia, v. 11, n. 5, p. 1707–1726, 2017. DOI: 10.14393/DL32-v11n5a2017-17. Disponível em: <https://seer.ufu.br/index.php/dominiosdelinguagem/article/view/37409>. Acesso em: 30 nov. 2022.

CAVACO-CRUZ, L.. *Manual Prático e Fundamental de Tradução Técnica*. Independence: Arkonte. 2012.

CUNHA LACERDA, P. F. A. da. O papel do método misto na análise de processos de mudança em uma abordagem construcional: reflexões e propostas. *Revista Linguística / Revista do Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro*. Volume Especial, dez de 2016, p. 83-101.

CUNHA LACERDA, P. F. A.; DALL’ORTO, L. F. M.. Uso do software Antconc na análise de dados do uso (no prelo).

DAVIES, M.. Corpora: An introduction. In: BIBER, D.; REPPEN, R. (eds.). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge University Press: Cambridge, 2015, p. 11–31.

KRIEGER, M. DA G.; SANTIAGO, M. S. *Estudos de terminologia para a tradução técnica*. Revista de Letras, v. 2, n. 33, 16 out. 2017.

LAVIOSA, S. *Corpus-based Translation Studies: theory, findings, applications*. Amsterdam: Rodopi, 2002. p. 23-31

NOVODVORSKI, A.; FINATTO, M. J. B.. Linguística de Corpus no Brasil: uma aventura mais do que adequada. Linguística de Corpus: abordagem e metodologia em pesquisas linguísticas de base empírica. *Letras & Letras*. Documentação de Estudos em Linguística Teórica e Aplicada, São Paulo, v. 30, n. 2, pp. 8-9, 2014.

SALGADO, A. R. Estudo de ferramenta computacional de análise de *corpora* aplicada à terminologia: Antconc. *Revista Moinhos*. vol. III. n. 03, 2013.

SCHIFFRIN, D. (1987) *Discourse markers*. Cambridge: Cambridge University Press.

SILVA, T. D. L. da; SILVA, E. M. da. *Mas o que é mesmo Corpus? – Alguns Apontamentos sobre a Construção de Corpo de Pesquisa nos Estudos em Administração*. In: Encontro da ANPAD, XXXVII, 2013, Rio de Janeiro. Artigo, Rio de Janeiro, EnANPAD, 2013.

TAGNIN, S. E. O.. *Os Corpora: instrumentos de auto-ajuda para o tradutor*. Cadernos de Tradução (UFSC), Florianópolis, v. 9, n. 2002/1, p. 191-213, 2003.

TAGNIN, S. E. A Linguística de Corpus na e para a tradução. In: VIANA, V.; TAGNIN, S. E. O. *Corpora na tradução*. São Paulo: Hub Editorial, 2015.

TELES, M. E. da S.. *Um estudo do resumo acadêmico em curso de graduação*. Dissertação (Mestrado em Língua Portuguesa). São Paulo: Pontifícia Universidade Católica de São Paulo, 2007.

TOGNINI-BONELLI; E.. *Theoretical overview of the evolution of corpus linguistics*. In: The Routledge Handbook of Corpus Linguistics. London and New York: Routledge, 2010, pp. 13-27.

_____, E.. *Corpus linguistics at work*. Amsterdam; Philadelphia: J. Benjamins, 2001.

TRAUGOTT, E. C.; TROUSDALE, G. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press, 2013.