

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Faculdade de Letras
Programa de Pós-Graduação em Estudos Linguísticos

Leticia Guedes Guimarães

**ESCRITA, TRADUÇÃO AUTOMÁTICA E AUTOPÓS-EDIÇÃO DE RESUMOS
ACADÊMICOS: UMA ANÁLISE TEXTUAL COM BASE EM TAXONOMIA DE
ERROS**

Belo Horizonte
2025

Letícia Guedes Guimarães

**Escrita, tradução automática e autopós-edição de resumos acadêmicos:
uma análise textual com base em taxonomia de erros**

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Estudos Linguísticos.

Área de concentração: Linguística Aplicada
Linha de pesquisa: Estudos da Tradução (3B)

Orientador: Prof. Dr. Leonardo Pereira Nunes

Belo Horizonte
2025

G963e

Guimarães, Letícia Guedes.

Escrita, tradução automática e autopós-edição de resumos acadêmicos [manuscrito] : uma análise textual com base em taxonomia de erros / Letícia Guedes Guimarães. – 2025.

1 recurso online (101 f. : il., tabs., grafs. (algumas color.) : pdf.

Orientador: Leonardo Pereira Nunes.

Área de concentração: Linguística Aplicada.

Linha de Pesquisa: Estudos da Tradução.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 94-99.

Exigências do sistema: Adobe Acrobat Reader.

1. Tradução e interpretação – Teses. 2. Tradução automática – Teses. 3. Redação acadêmica – Teses. I. Nunes, Leonardo Pereira. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD : 418.02



UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Escrita, tradução automática e autopós-edição de resumos acadêmicos: uma análise textual com base em taxonomia de erros

LETÍCIA GUEDES GUIMARÃES

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Mestre em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA APLICADA, linha de pesquisa Estudos da Tradução.

Aprovada em 10 de fevereiro de 2025, pela banca constituída pelos membros:

Prof(a). Leonardo Pereira Nunes - Orientador

UFMG

Prof(a). Deise Prina Dutra

UFMG

Prof(a). Norma Barbosa de Lima Fonseca

CMBH

Belo Horizonte, 10 de fevereiro de 2025.



Documento assinado eletronicamente por **Leonardo Pereira Nunes, Professor do Magistério Superior**, em 12/02/2025, às 12:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Deise Prina Dutra, Professora do Magistério Superior**, em 12/02/2025, às 14:21, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Norma Barbosa de Lima Fonseca, Usuário Externo**, em 12/02/2025, às 15:36, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3965590** e o código CRC **245C4475**.

Referência: Processo nº 23072.209704/2025-64

SEI nº 3965590

AGRADECIMENTOS

Uma pesquisa nunca é mérito de um único pesquisador. Por isso, seria injusto começar de outra forma que não agradecendo a todos que, de forma direta ou indireta, contribuíram para que este trabalho se concretizasse.

Em especial, gostaria de agradecer ao meu orientador, Leonardo Nunes. Mais do que apenas agradecer pelas discussões e pelos conhecimentos compartilhados e construídos, gostaria de reforçar o ser humano incrível que você é. Nenhum saber supera o poder da empatia. Muito obrigada pelo acolhimento, pela paciência, pela disponibilidade. Não sei como esta jornada teria sido se não fossem seus incontáveis “vai ficar tudo bem” nos momentos de maior ansiedade. Foi uma honra poder trilhar o início da minha vida como pesquisadora sob a sua orientação.

Gostaria de agradecer, também, às professoras Adriana Pagano, Deise Dutra e Norma Barbosa. Todas as vezes que fui até vocês em busca de informações, recebi bem mais do que havia pedido. As discussões durante as aulas, as indicações de palestras e cursos, as sugestões de bibliografia, as dicas para apresentações e todas as conversas paralelas no meio disso foram essenciais para que meu trabalho tomasse a forma que tomou.

Não poderia deixar de agradecer à FAPEMIG, ao CNPq e ao Poslin, sem cujo apoio financeiro e logístico esse trabalho jamais poderia ter acontecido.

Muito obrigada, também, a todos os colegas do mestrado que compartilharam esse caminho confuso e incrível comigo. Em especial, agradeço à Isabela Lee, que foi quem me convenceu, no início de tudo, que trocar de linha de pesquisa não era uma loucura, e que tem me acompanhado desde então. Foi incrível aprender com você.

Como nem só de universidade vive a mestrandia, gostaria de agradecer também à minha mãe. O ano passado não foi fácil e, se hoje eu concluo esta pesquisa, é porque nenhuma de nós duas desistiu. Obrigada por não desistir junto comigo. Agradeço de coração, também, a cada amigo que esteve comigo nos meus altos e baixos ao longo desta caminhada. Em especial, obrigada à Maiara e à Nathália, que me conhecem desde bem antes de eu entrar no mestrado; à Joice, ao Lucas e à Larissa, presentes que a faculdade me deu; e à Ariany, que o Pilates transformou em amiga. Não consigo imaginar como teriam sido esses dois últimos anos sem a amizade de vocês.

Por fim, mas não menos importante, muito obrigada à Letícia de 12 anos, que foi quem primeiro cultivou o sonho que, hoje, se concretiza nesta dissertação.

“Pra que cara feia?
Na vida, ninguém paga meia”

(Paulo Leminski)

RESUMO

O inglês, como língua dominante da ciência, representa uma barreira para pesquisadores não anglófonos, que precisam ler, pesquisar e publicar em uma língua estrangeira (Van Dijk, 1994). Dado que publicações em periódicos de alto impacto são essenciais para o avanço na carreira acadêmica, os tradutores automáticos apresentam-se como alternativa acessível (Bowker, Ciro, 2019). Essa ferramenta viabiliza a pós-edição, um processo em que um texto é traduzido automaticamente para outra língua e, em seguida, é editado, de forma a corrigir possíveis erros e gerar um texto de qualidade (Allen, 2003, Mesa-Lao, 2015). Na continuidade do projeto de Nunes (2021), o presente estudo investiga a influência da autopós-edição na qualidade de resumos acadêmicos em inglês, analisando a frequência e os tipos de erros encontrados nesses textos com base nas categorias estabelecidas pela taxonomia de Costa *et al.* (2015). Para tanto, compilamos um corpus paralelo bilíngue, composto por resumos originais escritos em português brasileiro e suas respectivas traduções automáticas e autopós-edições. Os textos foram produzidos por doutorandos de três grandes áreas do conhecimento: Ciências Exatas e da Terra, Ciências Humanas e Ciências Sociais Aplicadas. Para a anotação dos erros, adaptamos a taxonomia de Costa *et al.* (2015), destinada à categorização tanto de erros humanos quanto de sistemas de tradução automática. Realizamos a análise quantitativa dos dados, extraindo-os com o auxílio de *scripts* em Python para a contagem de erros e análise estatística. Os resultados indicam um aumento significativo na frequência de erros nos resumos autopós-editados em comparação com os originais em português, o que pode ser atribuído tanto a limitações na proficiência linguística dos autores quanto a uma pré-edição inadequada. A ordem em que as tarefas foram realizadas também influenciou os resultados, dado que a escrita em inglês anterior à autopós-edição contribuiu para textos de maior qualidade. Finalmente, os estudantes de Ciências Humanas cometeram um número significativamente maior de erros, sugerindo que a familiaridade com a escrita acadêmica em inglês pode variar entre as áreas do conhecimento. Esta pesquisa tem o potencial de contribuir tanto para o ensino e aprendizado de línguas quanto para a compreensão da influência dos sistemas de tradução automática na produção acadêmica de estudantes universitários.

Palavras-chave: escrita acadêmica; tradução automática; autopós-edição; taxonomia de erros.

ABSTRACT

English, as the dominant language of science, poses a barrier for non-native researchers who must read, conduct research, and publish in a foreign language (Van Dijk, 1994). Given that publications in high-impact journals are essential for academic career advancement, machine translation tools emerge as an accessible alternative. These tools enable post-editing, a process in which the machine translated text is revised to correct potential errors and produce better quality outputs (Allen, 2003; Mesa-Lao, 2015). Drawing on Nunes (2021), this study investigates the impact of self-post-editing on the quality of academic abstracts in English by analyzing the frequency and types of errors found in these texts based on the categories established by Costa *et al.*'s taxonomy (2015). A bilingual parallel corpus was compiled, consisting of original abstracts in Brazilian Portuguese, their automatic translations, and their post-edited versions produced by doctoral students from three major fields of knowledge: Earth and Exact Sciences, Humanities, and Applied Social Sciences. For error annotation, an adapted version of Costa *et al.*'s taxonomy (2015), designed for categorizing both human and machine errors, was employed. Quantitative analysis was performed, with error counts and statistical analysis conducted using Python scripts. The results indicate a significant increase in error frequency in self-post-edited abstracts compared to their original Portuguese versions, which can be attributed to both linguistic proficiency limitations of the authors and inadequate pre-editing. Task sequencing also influenced the outcomes, as writing directly in English prior to self-post-editing contributed to higher-quality texts. Finally, students from the Humanities realm made significantly more errors, suggesting that familiarity with academic writing in English may vary across disciplines. This research has the potential to contribute both to language teaching and learning and to understanding the influence of machine translation systems on the academic production of university students.

Keywords: academic writing; machine translation; self post-editing; error taxonomy.

LISTA DE ILUSTRAÇÕES

Figuras

Figura 1 - Estudos Multilíngues como um Espaço Multidimensional de Estudos Linguísticos Interconectados	13
Figura 2 - Mapa de Toury para a concepção de Holmes dos Estudos da Tradução	14
Figura 3 - Ramo aplicado dos Estudos da Tradução	14
Figura 4 - Proposta do estudo de Nunes (2021)	29
Figura 5 - Ordens possíveis no experimento	50
Figura 6 - Planilha contendo textos do corpus e metadados	52
Figura 7 - Interface do UdPipe	53
Figura 8 - Texto anotado gerado pelo UdPipe	54
Figura 9 - Interface de criação da taxonomia no INCEpTION	55
Figura 10 - Processo de anotação realizado no INCEpTION	55
Figura 11 - Trecho do código gerado pelo Chat GPT-4o mini e executado no Google Colab	63
Figura 12 - Organização da pasta de trabalho no Google Sheets	64

Gráficos

Gráfico 1 - Erros por categoria	68
Gráfico 2 - Porcentagem de erros por categoria (EL1)	71
Gráfico 3 - Porcentagem de erros por categoria (TA)	72
Gráfico 4 - Porcentagem de erros por categoria (PE L2)	73
Gráfico 5 - Porcentagem de erros por categoria (CET)	74
Gráfico 6 - Porcentagem de erros por categoria (CH)	75
Gráfico 7 - Porcentagem de erros por categoria (CSA)	76
Gráfico 8 - Porcentagem de erros por categoria (Ordem 1)	77
Gráfico 9 - Porcentagem de erros por categoria (Ordem 2)	77
Gráfico 10 - Porcentagem de erros por categoria (Ordem 3)	79

Tabelas

Tabela 1 - Divisão de resumos por área e modo de produção	51
Tabela 2 - Divisão de resumos por ordem e modo de produção	51
Tabela 3 - Contabilização de types e tokens	66
Tabela 4 - Média, amplitude e desvio padrão de types e tokens	67
Tabela 5 - Erros por etiqueta	69
Tabela 6 - Erros por modo de produção (CET)	80
Tabela 7 - Erros por modo de produção (CH)	82
Tabela 8 - Erros por modo de produção (CSA)	83
Tabela 9 - Erros por modo de produção (Ordem 1)	84
Tabela 10 - Erros por modo de produção (Ordem 2)	86
Tabela 11 - Erros por modo de produção (Ordem 3)	87

Quadros

Quadro 1 - Taxonomia proposta por Flanagan (1994)	37
Quadro 2 - Taxonomia proposta por Elliott <i>et al.</i> (2004)	38
Quadro 3 - Taxonomia proposta por Vilar <i>et al.</i> (2006)	40
Quadro 4 - Taxonomia proposta por Farrús <i>et al.</i> (2010)	41
Quadro 5 - Taxonomia proposta por Kirchhof <i>et al.</i> (2012)	42
Quadro 6 - Taxonomia proposta por Federico <i>et al.</i> (2014) e Girardi <i>et al.</i> (2014)	43
Quadro 7 - Taxonomia proposta por Stymne e Ahrenberg (2012)	44
Quadro 8 - Taxonomia de Lommel <i>et al.</i> (2014) com base na MQM	45
Quadro 9 - Taxonomia proposta por Costa <i>et al.</i> (2015)	46
Quadro 10 - Camadas de anotação	52
Quadro 11 - Etiquetas utilizadas no INCEPTION	59
Quadro 12 - Taxonomia de Costa <i>et al.</i> (2015) com adaptações	60
Quadro 13 - Estrutura da pesquisa	65
Quadro 14 - Síntese da pesquisa com as perguntas respondidas	88
Quadro 15 - Taxonomia de Costa <i>et al.</i> (2015) com exemplos dos autores	100

LISTA DE ABREVIATURAS E SIGLAS

Abreviaturas

Ort.	Ortografia
Léx.	Léxico
Gram.	Gramática
Sem.	Semântica
Disc.	Discurso

Siglas

CET	Ciências Exatas e da Terra
CH	Ciências Humanas
CSA	Ciências Sociais Aplicadas
EL1	Escrita em primeira língua
EL2	Escrita em segunda língua
E/T	Erros/ <i>Token</i>
IFA	Inglês para Fins Acadêmicos
L1	Primeira língua
L2	Segunda língua
PE L2	Autopós-edição em segunda língua
TA	Tradução automática
TTR	Razão <i>type/token</i>
UD	Dependências Universais

SUMÁRIO

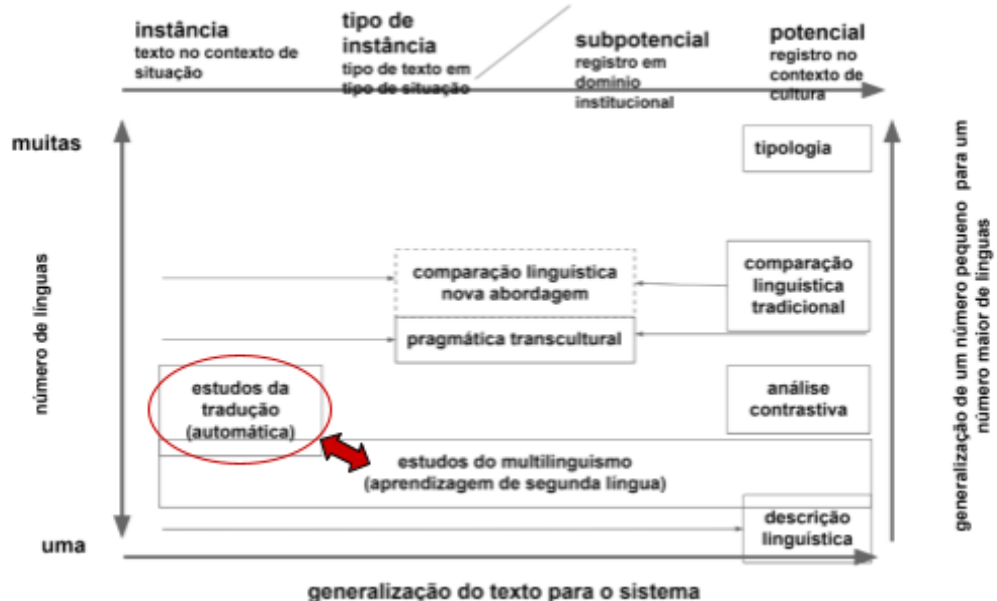
1 INTRODUÇÃO.....	13
1.1 Justificativa e formulação do problema.....	15
1.2 Objetivos.....	17
2 REVISÃO DA LITERATURA.....	18
2.1 Globalização, internacionalização institucional e inglês como língua da ciência.....	18
2.2 Tradução automática, pós-edição e autopós-edição.....	22
2.2.1 Tradução automática.....	22
2.2.2 Métodos de tradução automática.....	24
2.2.3 Pós-edição.....	26
2.2.4 A tradução e a autopós-edição no contexto acadêmico.....	27
2.3 Análise de erros e taxonomias de erros.....	34
2.3.1 Análise de erros humanos.....	34
2.3.2 Análise de erros de sistemas de tradução automática.....	35
2.3.2.1 <i>Taxonomias: classificação manual de erros de tradução automática.....</i>	<i>37</i>
2.3.3 Desafios no desenvolvimento e na aplicação de taxonomias.....	48
3 METODOLOGIA.....	49
3.1 Metodologia de coleta, compilação, armazenamento e anotação do corpus.....	49
3.1.1 O corpus: contexto de coleta e participantes.....	49
3.1.2 O corpus: composição, compilação e armazenamento.....	51
3.1.3 Anotação morfosintática.....	52
3.1.4 Anotação de erros.....	54
3.1.4.1 Adaptação da taxonomia.....	58
3.2 Metodologia de análise do corpus.....	62
3.2.1 Análise quantitativa.....	62
4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS.....	65
4.1 Types, Tokens e TTR.....	66
4.2 Contagem dos erros totais.....	67
4.3 Contagem dos erros por modo de produção.....	70
4.4 Contagem dos erros por área.....	74
4.5 Contagem dos erros por ordem.....	76
4.6 Contagem dos erros por área e modo de produção.....	80
4.7 Contagem dos erros por ordem e modo de produção.....	84
4.8 Síntese.....	88
5 CONSIDERAÇÕES FINAIS.....	90
REFERÊNCIAS.....	94
APÊNDICE A - Taxonomia de Costa et al. (2015) com exemplos.....	100

1 INTRODUÇÃO

A presente dissertação investiga a qualidade linguística, com base em análise de erros, de resumos acadêmicos produzidos por doutorandos de diferentes áreas do conhecimento. Esses resumos foram escritos em português brasileiro, traduzidos automaticamente para o inglês e, posteriormente, pós-editados por seus próprios autores (ou seja, autopós-editados). Assim, a pesquisa incide sobre três grupos de resumos: (1) os escritos em português brasileiro, (2) os traduzidos automaticamente para o inglês e (3) os autopós-editados em inglês. O objetivo é, por meio da investigação, avaliar a influência que o sistema de tradução automática tem na qualidade do texto resultante das autopós-edições, em termos do número e dos tipos de erros presentes.

Por se delinear como um estudo baseado em texto que lida com o contato linguístico entre duas línguas naturais, este trabalho se encontra abrangido pelo escopo dos Estudos Multilíngues (Matthiessen, Teruya e Wu, 2008). Na Figura 1, encontram-se mapeados os diversos fenômenos englobados por esse campo. Essa representação diferencia os estudos em termos: 1) de sua localização no contínuo de instanciação conforme o enfoque que adotam (Halliday, Matthiessen, 2014); e 2) do número de línguas com que operam. Nesse espaço, esta dissertação se caracteriza como um estudo da tradução (automática) que dialoga diretamente com os estudos do multilinguismo (Figura 1). Isso porque trabalha, ao mesmo tempo, com a tradução e com a aquisição do registro acadêmico no contexto de aprendizagem de língua estrangeira para fins acadêmicos.

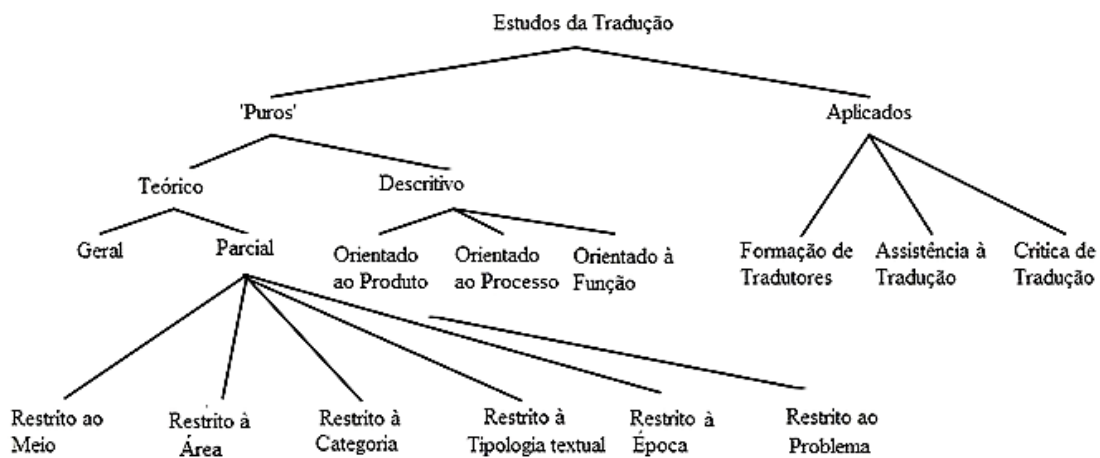
Figura 1 - Estudos Multilíngues como um Espaço Multidimensional de Estudos Linguísticos Interconectados



Fonte: Adaptado de Pagano (2023).

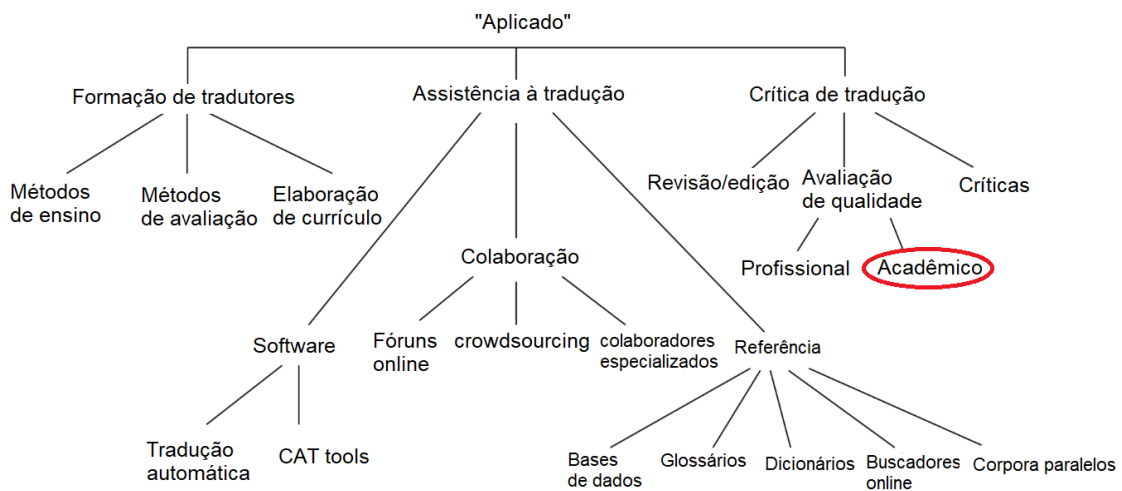
Considerando os Estudos da Tradução conforme seminalmente delineados por Holmes (1988), a presente pesquisa encontra-se abarcada pelo ramo aplicado dos Estudos da Tradução. Mais especificamente, enquadra-se como uma crítica de tradução, na medida em que propõe a avaliação da qualidade dos textos traduzidos e autopós-editados (Figura 2). Na versão expandida do ramo aplicado oferecida por Munday (2016), configura-se como uma crítica realizada enquanto atividade acadêmica (Figura 3), uma vez que ocorre no interior de uma pesquisa de mestrado.

Figura 2 - Mapa de Toury para a concepção de Holmes dos Estudos da Tradução



Fonte: adaptado de Pym (2016, p. 260).

Figura 3 - Ramo aplicado dos Estudos da Tradução



Fonte: Traduzido e adaptado de Munday (2016, p. 20).

A relevância da pesquisa, bem como as questões que guiaram a sua realização, são apresentadas em mais detalhes na próxima seção.

1.1 Justificativa e formulação do problema

Em 2024, Arenas-Castro *et al.* (2024) conduziram uma investigação que avaliou 736 periódicos da área de Ciências Biológicas. Seu objetivo era analisar a inclusão linguística de periódicos internacionais. A avaliação, que envolveu pesquisadores de diversos países, levou em consideração as diretrizes aos autores, questionários respondidos pelos editores-chefe, além de quatro variáveis consideradas indicativas de inclusão: 1) o fator de impacto; 2) a proporção de artigos de acesso livre; 3) a diversidade linguística dos editores; e 4) o fato de o periódico ser de domínio público.

Dentre outros achados, os resultados evidenciaram uma lacuna significativa no suporte oferecido a autores que têm o inglês como segunda língua (L2), posto que, embora 50% dos editores-chefe recomendem o uso de serviços de edição gratuitos, somente 1% dos periódicos disponibiliza-os de fato. Além disso, a análise do fator de impacto demonstrou que periódicos regionais ou publicados em países não-anglófonos tendem a exibir índices inferiores. Essa disparidade sugere que a publicação em inglês confere uma vantagem considerável em termos de visibilidade e alcance, o que não surpreende, uma vez que o inglês é a língua da ciência (Bowker, Ciro, 2019; Dutra *et al.*, 2019). Nesse cenário, a ausência de suporte por parte dos periódicos tem se provado um obstáculo adicional aos pesquisadores não-anglófonos, que já precisam fazer face ao triplo desafio de ler, pesquisar e escrever em uma língua estrangeira (van Dijk, 1994, p. 276).

Um obstáculo não só adicional, mas inescapável. Isso porque, atualmente, há uma estreita associação entre a realização de pesquisas científicas e a divulgação de seus resultados. Desse modo, cria-se uma pressão por produções relevantes, sublinhada pela máxima “publique ou pereça”. Com isso, as publicações em inglês deixam de ser opcionais, tornando-se pré-requisito para que os pesquisadores obtenham títulos e avancem na carreira universitária (Dutra *et al.*, 2019; Arruda *et al.*, 2021). Diante deste cenário, surgem três possibilidades, que podem ser combinadas ou adotadas conforme a necessidade: o aprendizado da língua inglesa, a contratação de serviços profissionais de tradução e/ou revisão e a colaboração com colegas anglófonos (Bowker, Ciro, 2019).

Mais recentemente, os avanços tecnológicos e a popularização dos sistemas de tradução automática possibilitaram a emergência de uma quarta opção: a aplicação dessas ferramentas na produção textual acadêmica (Bowker, Ciro, 2019). Em comparação à escrita em segunda língua (EL2), o processo de traduzir um texto automaticamente e editar a tradução gerada, denominado pós-edição (Allen, 2003), além de reduzir esforço cognitivo realizado (Krings, 2001; Koglin, 2020; Fonseca *et al.*, 2024), oferece a vantagem adicional de

poder ser empregado em um cenário de baixa proficiência na língua estrangeira na qual se precisa escrever (Garcia, 2011; Bowker, 2019).

Essas vantagens, aliadas à crescente demanda por publicações internacionais, têm motivado o desenvolvimento de estudos sobre a implementação da pós-edição nesse contexto. Dentre eles, encontra-se o projeto de pesquisa empírica-experimental de Nunes (2021), que investiga aspectos do processo e do produto envolvidos na escrita e na autopós-edição de resumos acadêmicos realizadas por doutorandos de diferentes áreas do conhecimento. No que tange ao processo, o esforço temporal, cognitivo e emocional já foram escrutinados (Vieira, 2022; Fonseca *et al.*, 2024; Avelar, 2023). Quanto ao produto, embora haja pesquisadores se dedicando ao estudo dos elementos retóricos dos textos, os elementos linguísticos ainda permanecem sem análise.

Assim, a proposta desta dissertação é, precisamente, ocupar essa lacuna. Mais especificamente, tenciona dar um primeiro passo na investigação da qualidade linguística dos resumos, por meio de uma análise comparativa dos erros em três conjuntos de textos: 1) resumos originalmente redigidos em português brasileiro; 2) suas traduções automáticas para o inglês; e 3) a autopós-edição em inglês dos textos gerados pelo tradutor. Os erros serão categorizados com base em uma versão adaptada da taxonomia de erros estabelecida por Costa *et al.* (2015), destinada à classificação de erros de tradutores automáticos e de humanos.

Dessa forma, estabelece-se a seguinte pergunta geral:

a) Qual é a influência da autopós-edição no número e nos tipos de erros em textos de diferentes áreas do conhecimento?

Ela subdivide-se em quatro outras, específicas:

b) Em que medida o número e os tipos de erros encontrados nos textos em português brasileiro influenciaram os erros encontrados nos textos traduzidos automaticamente para o inglês?

c) Em que medida o número e os tipos de erros encontrados nos textos traduzidos automaticamente para o inglês influenciaram os erros encontrados nos textos autopós-editados em inglês?

d) Em que medida o número e os tipos de erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente para o inglês e nos resumos autopós-editados em inglês variam conforme as diferentes áreas do conhecimento?

e) Em que medida o número e os tipos de erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente para o inglês e nos resumos autopós-editados em inglês variam conforme a ordem de execução das tarefas?

A abordagem do produto pela perspectiva da análise de erros interessa pela potencial aplicação de seus resultados nos estudos do multilinguismo e nos estudos da tradução. Isso porque, por um lado, esse tipo de análise é uma importante fonte de informação sobre as estratégias que os aprendizes (formais ou informais) empregam na aquisição de uma língua estrangeira (Dulay, Burt, Krashen, 1982); por outro lado, ela é relevante por possibilitar uma melhor compreensão da influência dos sistemas de tradução automática na qualidade linguística da produção textual acadêmica de estudantes da pós-graduação. Finalmente, dado haver uma carência de trabalhos que lidem com a avaliação de erros em tradução automática — erros em pós-edição inclusos — no par linguístico inglês-português brasileiro (Costa *et al.*, 2021), o presente estudo acaba por preencher, também, essa lacuna.

1.2 Objetivos

Esta pesquisa tem por objetivo geral contribuir para os Estudos do Multilinguismo por meio da investigação, no par linguístico português brasileiro-inglês, da influência da autopós-edição na quantidade e nos tipos de erros encontrados em resumos acadêmicos de distintos domínios do conhecimento.

Para tal, estabelece quatro objetivos específicos, quais sejam:

- 1) Analisar a relação entre os erros observados nos resumos acadêmicos em português e os erros observados nos resumos traduzidos automaticamente para o inglês.
- 2) Analisar a relação entre os erros encontrados nos resumos acadêmicos traduzidos automaticamente e os erros encontrados nos resumos autopós-editados em inglês.
- 3) Analisar a relação entre os erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente para o inglês e nos resumos autopós-editados em inglês para cada uma das áreas do conhecimento.
- 4) Analisar a relação entre os erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente para o inglês e nos resumos autopós-editados em inglês para cada uma das ordens de execução das tarefas.

Além da presente Introdução, este trabalho encontra-se estruturado em mais quatro seções. A segunda revisa a literatura pertinente, fornecendo o embasamento teórico para o estudo. A terceira seção detalha a metodologia empregada na pesquisa. Na sequência, a quarta

apresenta e discute os resultados. Por fim, a quinta seção encerra o trabalho com as considerações finais.

2 REVISÃO DA LITERATURA

2.1 Globalização, internacionalização institucional e inglês como língua da ciência

Ainda que em muitos contextos, os termos ‘globalização’ e ‘internacionalização’ sejam usados de forma indistinta (Knight, 1994), eles não são sinônimos. A globalização guarda estreita relação com o crescente nível de interconexão e interdependência mundial, que, elevado pelos progressos tecnológicos, resultou em transformações econômicas, sociais e políticas significativas (National Geographic Society, 2024). A troca mundial de produtos, costumes e ideias remonta à Rota da Seda, já em 50 a.C., mas, na chamada Era da Informação, ela foi impulsionada a níveis nunca antes vistos.

A conexão entre as diferentes partes do mundo sempre teve seu ritmo ditado pela tecnologia. O aperfeiçoamento dos instrumentos e dos meios de navegação determinou as Grandes Explorações e subsequentes colonizações; as Revoluções Industriais e as duas Grandes Guerras foram marcadas pelas fábricas, ferrovias, carros, aviões e armas nucleares. No entanto, foram os avanços nas tecnologias computacionais e informacionais que cimentaram, de forma definitiva, a interconexão mundial que temos hoje. Assim, teve início a Era da Informação.

Esse cenário traz desvantagens, gerando e aprofundando a desigualdade social, causando danos ao meio ambiente, apagando a pluralidade cultural, acelerando a transmissão de doenças e dando origem a instabilidades políticas. Entretanto, também abre espaço a trocas econômicas importantes, dá origem a alianças políticas, viabiliza a cooperação mundial em cenários de crise; não menos importante, possibilita o trânsito de pessoas e informações e o compartilhamento de conhecimentos — inclusive no cenário acadêmico-científico.

Assim, a globalização pode ser definida como a soma dessas forças econômicas, políticas e sociais que, fazendo parte da realidade do século XXI, são responsáveis por fomentar o engajamento internacional das instituições de ensino superior (Altbach e Knight, 2007). Por sua vez, a internacionalização, enquanto processo, diz respeito à “integração da dimensão internacional nas funções de ensino, pesquisa e extensão de uma instituição de ensino superior” (Knight, 1993, *apud* Arruda *et al.*, 2021, p. 175). Em contrapartida, enquanto produto, “abrange as políticas e práticas adotadas por sistemas e instituições acadêmicas — e

até mesmo por indivíduos — para lidar com o contexto acadêmico global”¹ (Altbach e Knight, 2007, p. 290).

Desse modo, a internacionalização pode ser vista, em última análise, como uma resposta à globalização ou como um de seus produtos (Beck, 2012). Embora ocorra em nível mundial, sua implementação varia conforme as políticas adotadas em cada país². No caso do Brasil, o Plano Nacional de Educação 2014-2024 (PNE) (Brasil, 2014a) desempenhou papel central no reconhecimento da relevância da internacionalização para a educação superior (Arruda *et al.*, 2021). Isso porque, em sua décima quarta meta, voltada à pós-graduação, o PNE traz diretrizes relacionadas diretamente ao processo de internacionalização, abrangendo ambas as suas dimensões: a local e a global (Knight, 2008). Prevê, primeiramente, a consolidação de programas, projetos e ações que tenham esse objetivo; ademais, incentiva o intercâmbio científico e tecnológico, tanto a nível nacional quanto internacional (Brasil, 2014a). Com isso, visa uma maior interconexão entre os pesquisadores de diferentes partes do mundo.

Dentre as iniciativas governamentais com esse propósito, a mais importante empreendida no Brasil foi o programa Ciências sem Fronteiras (CsF). Entre 2011 e 2017 (Brasil, 2011), ele estimulou a mobilidade estudantil por meio da concessão de bolsas a alunos de graduação e pós-graduação em Ciências Exatas interessados em realizar estudos no exterior. Paralelamente, buscou fomentar o intercâmbio científico ao incentivar a vinda de pesquisadores estrangeiros para o país. Em 2017, após cortes orçamentários, o CsF foi encerrado.

Outra iniciativa central foi o programa Idiomas sem fronteiras (IsF), que surgiu em 2012 (Brasil, 2014b) como resposta a uma lacuna identificada por meio do CsF. Dado que muitos estudantes demonstraram uma baixa proficiência em idiomas estrangeiros, o programa visava suprir essa carência. No início, tinha como foco apenas o inglês e, por isso, até 2013, era chamado de Inglês sem Fronteiras; mas, em 2014, expandiu a oferta, passando a abranger, também, outros idiomas — como o italiano, o francês, o espanhol e o alemão — com isso, mudando seu nome para Idioma sem Fronteiras.

Seu objetivo, então, era fornecer aos estudantes do ensino superior as ferramentas linguísticas essenciais para a atuação acadêmica em uma língua estrangeira a nível local (publicação de artigos, participação em eventos internacionais) e global (mobilidade

¹ Nossa tradução para: “Internationalization includes the policies and practices undertaken by academic systems and institutions—and even individuals—to cope with the global academic environment”.

² Uma comparação entre os contextos brasileiro, norte-americano e canadense foi apresentada em Arruda *et al.*, 2021.

internacional). Em 2019, o IsF deixou de ser um programa administrado pelo governo, passando a ser gerido diretamente por cada instituição de ensino superior.

No caso da UFMG, a universidade continua a fazer parte do IsF por meio da adesão à Rede Andifes-IsF. Além disso, localmente, a Faculdade de Letras da universidade, com o apoio do Setor de Proficiência da Diretoria de Relações Internacionais (DRI), mantém a oferta de disciplinas de Inglês para Fins Acadêmicos (IFA)³. Ademais, desenvolve um programa de extensão, o Idiomas para Fins Acadêmico-Profissionais (iUFMG)⁴, que abrange outras línguas, como francês, italiano, alemão e espanhol.

Como a universidade opera na intersecção entre ensino, pesquisa e extensão, vários pesquisadores da UFMG têm se dedicado ao estudo de idiomas para fins de internacionalização; dentre os trabalhos, pode-se citar os de Dutra *et al.* (2019); Orfanò, Oliveira e da Silva (2018); Alves (2019a, 2019b); Terra (2019); e Nunes (2021). Mais recentemente, com o avanço dos sistemas de tradução automática, o uso dessas tecnologias no contexto da leitura e da publicação em inglês passou a ser mais intensamente explorado pelos pesquisadores. Destacam-se os projetos “Cognitive load and potential cost of ELF in academic contexts” (Alves, 2018; Alves *et al.*, 2019a e 2019b), no qual colaboraram oito pesquisadores de três instituições, e “A produção textual acadêmica em distintos domínios: investigando a escrita e a pós-edição em língua inglesa sob a perspectiva linguística e cognitiva” (Nunes, 2021), financiado pelo CNPq, no qual se insere a presente pesquisa de mestrado.

Todas essas pesquisas concentram-se, principalmente, no inglês para fins acadêmicos. Essa predominância se explica pelo domínio dessa língua, que é, atualmente, a língua franca mundial e, por isso, também a língua da ciência (Bowker, Ciro, 2019). As razões que explicam a disseminação do inglês pelo mundo remontam aos anos de domínio do Império Britânico, responsável pela dominação de inúmeros territórios ao redor do mundo e berço das Revoluções Industrial e Científica. No entanto, foi a ascensão dos Estados Unidos que determinou os rumos que o cenário linguístico global tomou (Montgomery, 2013).

Se a Inglaterra foi o berço da primeira revolução industrial, os Estados Unidos foram o berço da segunda, o que lhes gerou enorme riqueza e os consolidou como uma potência global após a Primeira Guerra Mundial, ao lado da Inglaterra e da França (Montgomery, 2013, p. 33). Durante esse período, os EUA também foram responsáveis por várias inovações que

³ Para mais informações sobre o IFA, cf. Dutra *et al.*, 2019.

⁴ Para mais informações sobre o iUFMG, cf. <http://www.lettras.ufmg.br/iufmg/>

transformaram a vida no Ocidente, dentre as quais o telefone, a eletricidade, os automóveis, e a cultura de massa, aprofundando a sua importância (Montgomery, 2013, p. 33). Com as vitórias conquistadas na Segunda Guerra Mundial e, logo em seguida, na Guerra Fria, os EUA firmaram-se como uma potência política, econômica e sociocultural sem paralelo no mundo (Montgomery, 2013).

Não menos relevante, firmaram-se, também, como uma potência científica. Isso porque, com o fim da Segunda Guerra Mundial, enquanto a Europa encontrava-se devastada, os EUA contavam com cidades e universidades intactas, infraestrutura expandida, recursos abundantes e avanço em diversas áreas de pesquisa. Não havia rival para o seu vasto e prestigiado sistema científico, elevado pela guerra a um patamar de grande produtividade (Montgomery, 2013, p. 74). Esse contexto foi determinante para que o inglês se estabelecesse como a língua da ciência.

Comparado às publicações acadêmicas de algumas décadas atrás, logo após a Segunda Guerra Mundial, o cenário da comunicação de pesquisas alterou-se drasticamente. O alemão era o idioma principal para a publicação científica internacional, tendo o francês e o inglês como importantes rivais. Desde então, o inglês assumiu esse papel. (Mauranen; Hynninen; Ranta, 2016, p. 50, *apud* Arruda *et al.*, 2021, p. 178).

Atualmente, o inglês deixou de ser uma língua dependente dos países anglófonos, tornando-se uma língua franca, “uma 'língua de contato' entre pessoas que não compartilham nem uma língua nativa comum nem uma cultura (nacional) comum, e para as quais o inglês é a língua estrangeira escolhida para a comunicação”⁵ (Firth, 1996, p. 240). Embora essa definição tenha a limitação de não incluir os falantes de inglês como primeira língua (L1) (Flowerdew, 2015), capta bem a ideia de que, mais do que se associar a um centro mundial de poder, o inglês atende ao interesse de estabelecimento uma língua comum, que permita a troca de informações e conhecimento entre pessoas de diferentes origens no contexto globalizado atual.

Esse cenário apresenta vantagens e desvantagens. Por um lado, como colocado, a existência de uma língua franca é vantajosa por permitir a comunicação entre pesquisadores de diversas origens, além de se alinhar ao uso global do inglês, exigindo dos pesquisadores, então, o domínio de uma única língua franca para diferentes atividades. Por outro, como apontado na Introdução, representa um desafio a mais àqueles que, por não advirem de países anglófonos, precisam aprender uma língua adicional.

⁵ Nossa tradução para: “a ‘contact language’ between persons who share neither a common native tongue nor a common (national) culture, and for whom English is the chosen foreign language of communication”.

No caso da internacionalização global, torna-se necessário que os acadêmicos invistam tempo e, muitas vezes, dinheiro na aprendizagem do inglês. É nesse contexto que ganham relevância iniciativas como o IsF e o IFA, que oferecem alternativas gratuitas àqueles que desejam progredir no domínio desse idioma.

No caso da internacionalização que ocorre localmente, uma menor proficiência em inglês significa acesso a bibliografia reduzida, posto a maior parte estar redigida nessa língua; publicações com menor alcance, dado os periódicos de maior impacto exigirem textos em inglês (Arenas-Castro *et al.*, 2024); e dificuldade de progressão da carreira, tendo em vista a necessidade dos dois anteriores para que esta ocorra (Bowker, Ciro, 2019). Nesse ponto, conforme sinalizado na introdução, a tradução automática e a autopós-edição se apresentam como alternativas que podem poupar recursos humanos, temporais e financeiros. De onde a importância de estudos como o de Alves (2018), Nunes (2021), e, no esteio destes, o empreendido nesta dissertação.

2.2 Tradução automática, pós-edição e autopós-edição

2.2.1 Tradução automática

A tradução automática pode ser definida como "o emprego de computadores na tarefa de traduzir textos de uma língua natural para outra"⁶ (EAMT, 2024). As primeiras propostas práticas relacionadas à tradução automática surgiram na década de 1930, com Georges Artsrouni, na França e Petr Trojanskii, na Rússia (Hutchings, 2007). Eles patentearam, respectivamente, uma máquina para diversos usos (incluindo a tradução automática), denominada "cérebro mecânico" (*cerveau mécanique*); e uma máquina para automatizar buscas em dicionários para fins de tradução (Hutchings, 2004).

No entanto, foi a publicação do Memorando de Weaver, em 1949, que realmente fez proliferarem as pesquisas em tradução automática. Seu autor, Warren Weaver, não foi o primeiro a desenvolver a ideia de um tradutor automático, mas foi ele que levou a público soluções a problemas linguísticos, envolvidos na concepção desses sistemas, que, até então, não haviam sido abordados por outros estudiosos (Hutchings, 2007). Isso foi o estímulo de que os pesquisadores precisavam.

Em 1952, sob a organização de Yehoshua Bar-Hillel, foi realizada a primeira conferência voltada ao assunto (Hutchings, 2007). Nela, os pesquisadores puderam

⁶ Nossa tradução para: "the application of computers to the task of translating texts from one natural language to another".

compartilhar ideias para solucionar problemas concretos identificados durante suas tentativas de realizar traduções automáticas entre duas línguas naturais. Dentre as sugestões apresentadas, estavam estratégias para a intervenção humana no texto, antes (pré-edição) ou após (pós-edição) a passagem do texto pelo sistema de tradução.

Àquela altura, “já estava claro que a automação completa de traduções de boa qualidade era virtualmente impossível, e que a intervenção humana (...) seria essencial; alguns esperavam que isso fosse apenas uma medida temporária, mas a maioria acreditava que sempre seria necessária”⁷ (Hutchings, 2007, p. 2). Isso, no entanto, não tolheu os esforços para o desenvolvimento de sistemas capazes de traduzir.

Na verdade, em 1954, ano em que foi realizada a primeira demonstração pública de um tradutor automático, os investimentos nas pesquisas cresceram ainda mais, e elas começaram a proliferar pelo mundo. Assim, entre 1956 e 1966, diversos grupos de pesquisas na América, na Europa e na Ásia dedicaram-se ao desenvolvimento e ao aperfeiçoamento desses sistemas de tradução. No entanto, conforme os anos passavam sem que a qualidade das traduções automáticas melhorasse notavelmente, o otimismo esmoreceu. Na época, a infraestrutura computacional era muitas vezes inadequada, demandando um grande esforço para o aprimoramento do *hardware* básico (Hutchings, 2007).

Em 1966, após dois anos de análises solicitadas pelo governo norte-americano, o Comitê Consultivo em Processamento Automático de Linguagem (ALPAC) concluiu que os esforços e o dinheiro empregado não se justificavam. O relatório que divulgaram afirmava não haver “perspectiva imediata ou previsível de tradução automática útil”⁸ (ALPAC, 1996, *apud* Hutchings, 2007, p. 6), que apenas havia se provado mais custosa e menos eficiente do que a tradução humana. Além disso, afirmava que a quantidade de tradutores humanos disponíveis era mais do que suficiente para suprir as demandas por tradução existentes à época. Ainda que o relatório da ALPAC tenha recebido críticas quanto à sua parcialidade e ausência de visão a longo prazo (Hutchings, 2007), seu impacto foi grande o suficiente para manchar a reputação dos tradutores automáticos frente ao público e minar os investimentos na área, tanto nos EUA e como em outras partes do mundo.

A verdadeira retomada ocorreu à medida que a demanda por traduções cresceu a tal ponto que os recursos humanos não eram suficientes para responder a ela. As pesquisas, que nunca haviam sido completamente interrompidas, voltaram a receber atenção pública e

⁷ Nossa tradução para: “It was already clear that full automation of good quality translation was a virtual impossibility, and that human intervention either before or after computer processes (...) would be essential; some hoped this would be only an interim measure, but most expected that it would always be needed.”

⁸ Nossa tradução para: “immediate or predictable prospect of useful machine translation.”

financiamentos. Contribuiu para isso o fato de que, nos anos anteriores, a área da tradução automática tinha visto alguns sucessos, tais quais os sistemas MÉTEO, desenvolvido e adotado no Canadá; e o Systran, de origem estadunidense, utilizado pela Comissão Europeia. Além disso, os tradutores humanos em exercício tornavam-se cada vez mais cientes dos benefícios que os *softwares* de auxílio à tradução tinham em seu trabalho e começavam a fazer cada vez mais uso deles.

A partir de então, a globalização crescente, marcada pela multiplicação e pela disseminação vertiginosa da informação — por que a internet foi, em grande parte, responsável — fez com que o campo apenas crescesse. Conforme as tecnologias progredem, os sistemas se tornam cada vez mais eficientes, respondendo a uma diversidade de cenários que o mundo hiperconectado cria. Em 2014, os avanços na inteligência artificial revolucionaram ainda mais esse cenário, tornando os sistemas mais precisos, rápidos, menos custosos e, por isso mesmo, mais acessíveis. É justamente a respeito dos diferentes tipos de tradutores que foram desenvolvidos ao longo dos anos que passamos a falar a seguir.

2.2.2 Métodos de tradução automática

As abordagens da tradução automática distinguem-se em duas principais: a que se baseia em regras e a que se baseia em corpus. Até o início do século XXI, prevaleceram os tradutores automáticos baseados em regras. Nesse método, as traduções são orientadas por regras lexicais, morfológicas e/ou sintáticas das línguas fonte e alvo. Essas regras, que precisam direcionar: 1) a análise da língua de partida, 2) sua passagem para a de chegada e 3) a geração do texto final na língua alvo (Zhang, Zhong, 2020), devem ser manualmente escritas por especialistas. Isso significa que, para cada língua e para cada par linguístico, é necessária a realização de um mapeamento específico, que coordene todo esse processo tradutório tripartido (Bowker, Ciro, 2019). Por essa exigência, o desenvolvimento de sistemas de tradução automática baseados em regras demanda tempo e trabalho humano intensivo, uma das razões pelas quais os pesquisadores buscaram alternativas a ele (Bowker, Ciro, 2019).

A evolução dos computadores e da Linguística de Corpus fez com que os tradutores baseados em corpora começassem a receber atenção. Esse tipo de tradutor recorre a corpora bilíngues que contêm textos originais e suas respectivas traduções, alinhados sentencialmente. Assim, para traduzir, esses sistemas recorrem não a regras, mas à pesquisa de dados. Três são as abordagens que se orientam dessa forma: a baseada em exemplos, a baseada em estatísticas e a baseada em redes neurais.

Os tradutores que se baseiam em exemplos ganharam proeminência a partir da metade da década de 1980. Seu funcionamento é simples: diante de um texto na língua fonte que precisa ser traduzido para a língua alvo, eles consultam os dados que foram obtidos dos corpora bilíngues em busca de ocorrências semelhantes. Embora resultem em traduções de boa qualidade, esses sistemas exigem corpora bilíngues de grandes dimensões, que ainda assim não são capazes de atender a todos os fenômenos linguísticos (Wang *et al.*, 2021).

Além disso, em situações em que mais de uma tradução é possível para uma dada palavra ou sentença, um tradutor que se baseasse apenas em exemplos teria problemas para encontrar a melhor solução tradutória; nesse contexto, a estatística se apresenta como uma alternativa (Bowker, Ciro, 2019). Assim, a partir de 1990, ganham relevância os sistemas de tradução automática que se utilizam de dados estatísticos para realizar as traduções. Nesse cenário, por meio de cálculos probabilísticos e do acesso a grandes quantidades de dados, os sistemas aprendem a traduzir sozinhos, dispensando a formulação manual de regras linguísticas (Wang *et al.*, 2021). Entretanto, ainda que lide melhor com cenários em que há mais de uma tradução possível, esse tipo de tradutor ainda necessita de intensivo trabalho humano para a criação de modelos de linguagem, tradução e reordenação; e de corpora bilíngues extensos, nem sempre disponíveis, ainda mais no caso de línguas menos utilizadas (Bowker, Ciro, 2019).

Desde o início, os pesquisadores sabiam que o grande salto na área de tradução automática viria do aperfeiçoamento da inteligência artificial (Hutchings, 2007). De fato, em 2014, o avanço do *deep learning* tornou possível o desenvolvimento dos sistemas de tradução automática baseados em redes neurais (Wang *et al.*, 2021). Diferentemente do que ocorre no caso dos tradutores estatísticos, no caso das redes neurais, os pesquisadores não necessariamente precisam dizer aos algoritmos o que procurar, posto que o sistema é capaz de encontrar padrões e tendências sozinho, por mais complexos que sejam, mesmo quando os dados são complicados ou imprecisos (Bowker, Ciro, 2019). Assim, “a ideia básica do NMT [tradutor automático neural] é mapear a língua de origem como uma representação semântica densa e, em seguida, gerar a tradução usando um mecanismo de atenção”⁹ (Wang *et al.*, 2021, p.144).

Uma das grandes vantagens da tradução neural é que ela pode processar pares linguísticos que possuem poucos dados disponíveis para o treinamento, ainda que não produza traduções de alta qualidade. Isso porque, diferente dos sistemas anteriores, ela

⁹ Nossa tradução para: “The basic idea of NMT is to map the source language into a dense semantic representation, and then generate the translation by using an attention mechanism”.

trabalha bem com tradução indireta auxiliada por um idioma pivô, como o inglês (Bowker, Ciro, 2019). Além disso, esse sistema demanda muito menos trabalho manual e tempo para ser desenvolvido. Enquanto o primeiro tradutor estatístico precisou de aproximadamente 16 anos para ser disponibilizado online, o primeiro tradutor neural precisou de apenas um (Wang *et al.*, 2021). Entretanto, assim como seus antecessores, ele ainda demanda uma grande quantidade de dados para produzir traduções acuradas.

Atualmente, os sistemas disponíveis congregam dois ou mais desses métodos (Bowker; Ciro, 2019), uma vez que “não há um método único para obter traduções automáticas de qualidade”¹⁰ (Hutchings, 2007, p. 14). Todos apresentam suas forças e fraquezas e a combinação deles resulta em tradutores cada vez mais eficientes, levando mesmo a alegações de superação da qualidade de traduções humanas. Entretanto, essa é uma assunção precipitada e, embora haja muito otimismo em torno do futuro da tradução automática, ainda há problemas que a impedem de equiparar-se à tradução humana (Bowker, Ciro, 2019; Wang *et al.*, 2021; Koehn, 2024).

Tradutores humanos têm uma tolerância muito maior a erros e a usos particulares da língua, o que facilita lidarem com textos que estejam fora da linguagem padrão e/ou façam uso de gírias, expressões idiomáticas e jargões. Além disso, diferentemente dos sistemas de tradução automática, cujo desempenho depende diretamente da quantidade e da qualidade dos dados com que foram alimentados, seres humanos podem aprender com uma quantidade menor de exemplos (Wang *et al.*, 2021). Ademais, a tradução exige não só conhecimentos linguísticos aprofundados, mas conhecimento de mundo. Dessa forma, “é necessário desenvolver novos métodos que combinem regras simbólicas, conhecimento e redes neurais para melhorar ainda mais a qualidade da tradução [automática]”¹¹ (Wang *et al.*, 2021, p. 150).

Isso, no entanto, não muda o fato de que as tecnologias atuais com frequência produzem traduções de excelente qualidade, cuja intervenção humana precisa ser mínima para que se tornem plenamente adequadas. Assim, cria-se cada vez espaço para o fomento de uma prática já consolidada no mercado tradutor: a pós-edição, da qual trata a próxima seção.

2.2.3 Pós-edição

A pós-edição pode ser definida como a revisão humana de uma tradução gerada por tradutores automáticos com vistas a atingir uma qualidade adequada a um padrão específico.

¹⁰ Nossa tradução para: “there can be no single method for achieving good-quality automatic translation”.

¹¹ Nossa tradução para: “it is necessary to develop new methods that can combine symbolic rules, knowledge, and neural networks to further improve translation quality”.

Nesse processo, corrigem-se erros e realizam-se modificações no texto conforme necessidade. Em geral, o objetivo é realizar o menor número de edições possível (Mesa-Lao, 2013; Allen, 2003). Para realizar pós-edições realmente de qualidade, é necessário um pós-editor capacitado, com profundos conhecimentos linguísticos, de tradução e bom entendimento do processo de tradução automática e de suas falhas (Vasconcellos, León, 1985; O'Brien, 2002; Mesa-Lao, 2013). De um ponto de vista mercadológico, a pós-edição interessa porque possibilita reduzir os custos de produção, publicar mais conteúdo em menos tempo e distribuí-lo em mais línguas.

Do ponto de vista do tradutor, a pós-edição também traz vantagens, já havendo diversos trabalhos que comprovam que essa atividade implica menor carga cognitiva, reduzindo os esforços técnico, temporal e cognitivos empregados quando em comparação com uma tarefa de tradução (Krings, 2001; Koglin, 2020; Fonseca *et al.*, 2024). A redução da carga cognitiva, porém, parece estar condicionada a uma série de fatores, ainda a serem estudados mais detalhadamente (Koponen, 2016). Assim, dependendo da estrutura e da qualidade da tradução automática gerada, por exemplo, a pós-edição pode acabar sendo mais dispendiosa do que a tradução, exigindo mesmo a retradução completa do texto original.

Entretanto, uma vez que, diferente dos erros humanos, os erros gerados pelos tradutores automáticos são previsíveis (Mesa-Lao, 2013), é possível adotar práticas que possibilitem a geração de saídas de traduções automáticas de maior qualidade. Para isso, basta realizar edições na produção textual antes que ela seja submetida ao tradutor automático, ou seja, pré-editar o texto (Mesa-Lao, 2013). Assim como a pós-edição, isso exige competências e técnicas específicas, ligadas ao conhecimento da operação dos tradutores automáticos e de suas limitações. Dentre as mudanças passíveis de serem implementadas, estão a quebra de frases longas em frases mais curtas e a supressão de inversões, substituindo-as por frases na ordem direta (Mesa-Lao, 2013; Bowker, 2019).

Como mencionado na introdução, a tradução e a pós-edição também já vem sendo investigadas no contexto acadêmico-científico. É isso que a seção seguinte discute em mais detalhes.

2.2.4 A tradução e a autopós-edição no contexto acadêmico

A multiplicação e a democratização dos sistemas de tradução automática fizeram deles uma excelente ferramenta, também, para os não-profissionais que necessitassem lidar com textos em idiomas estrangeiros no dia a dia. No contexto acadêmico, tornaram-se uma opção atrativa especialmente para os discentes da pós-graduação, que estão em contato constante

com textos em línguas outras que não a materna, de forma que um tradutor automático pode ser útil tanto no processo de leitura quanto de produção dos textos, ambos essenciais à participação na comunidade acadêmico-científica.

No que tange à produção textual, os sistemas de tradução automática mostram-se uma opção interessante principalmente quando o aprendiz não possui um domínio muito grande do idioma no qual precisa escrever. As possibilidades vão desde a tradução de palavras até a do texto em sua totalidade. Nesse cenário, surge a importância de estudos que se dedicam à autopós-edição, ou seja, à edição de um texto traduzido automaticamente que foi, originalmente, escrito pela própria pessoa que pós-edita (Bowker, Ciro, 2019). Interessando-se especialmente pelo processo, Alves (2018), por exemplo, investigou os esforços cognitivo e emocional envolvidos na autopós-edição de resumos acadêmicos realizada por pesquisadores da área de saúde, comparando-os aos demandados por tarefas de EL1 e EL2. O estudo focou, ainda, nos desafios extras que a autopós-edição pode requerer, especialmente para especialistas e estudantes de tradução.

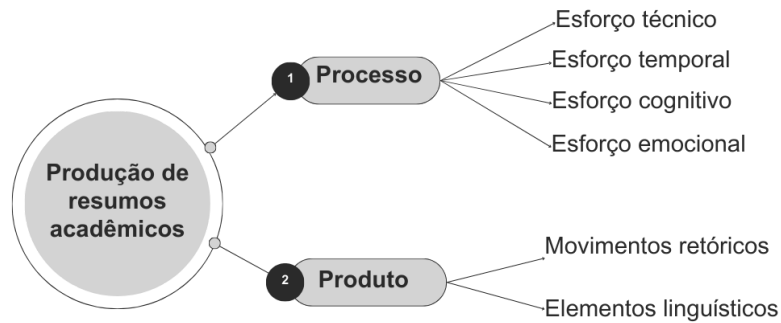
Os participantes envolvidos no estudo eram doutorandos ou pós-doutorandos brasileiros e suíços, os quais tinham como L1 o português, o espanhol ou o alemão. A única L2 envolvida no estudo era o inglês. Para a coleta, aplicou-se a técnica de triangulação, de modo que houve “o uso conjunto de métodos quantitativos e qualitativos dentro de uma perspectiva de complementaridade” (Alves, 2001, p. 70). Assim, foram considerados indicadores de esforço realizadas medidas de diferentes natureza, tais como tempo de execução da tarefa, pausas, número de teclas pressionadas, cliques do mouse, rolagem, fixações, regressões, revisões, dilatação da pupila, frequência cardíaca, autoavaliação, comentários e indicadores linguísticos nos textos alvo.

Na continuidade de Alves (2018), Nunes (2021), propôs um estudo que também compara dados da autopós-edição àqueles da EL1 e EL2. Trabalhando no par linguístico português-brasileiro/inglês, o projeto propôs abordar resumos acadêmicos: 1) escritos em português brasileiro; 2) escritos em inglês; e 3) autopós-editados em língua inglesa. Embora a pesquisa de Nunes (2021) adote um desenho experimental semelhante ao de Alves (2018), expande-o ao ampliar as áreas do conhecimento implicadas e ao analisar em detalhes, além de aspectos do processo, os do produto (Figura 4).

Dentre os dados que advieram do estudo, Vieira (2022) e Fonseca *et al.* (2024) analisaram o esforço temporal, comparando o tempo despendido nas tarefas de EL1, de EL2 e de PEL2. Os resultados indicaram que, dentre as três, a tarefa mais dispendiosa foi a EL2, provavelmente em razão do esforço adicional demandado pelo uso de uma língua estrangeira.

Além disso, a ordem em que a tarefa foi realizada mostrou-se inversamente proporcional ao esforço requerido. A tradução automática, por sua vez, provou-se um elemento facilitador, de forma que a PEL2, seja considerada isoladamente ou em combinação com a EL1, foi menos onerosa do que a EL2.

Figura 4 - Proposta do estudo de Nunes (2021)



Fonte: Elaboração própria, 2025.

A seu turno, Avelar (2023), também lidando com os dados processuais, comparou os esforços cognitivo e emocional realizados durante as três tarefas. Seus achados mostraram que, em comparação à EL1 e à EL2, a PEL2 demandou menos esforço cognitivo. Isso, no entanto, não necessariamente correspondeu a um menor esforço emocional, o qual parece mais condicionado à ordem de realização do que às tarefas propriamente ditas. Conforme apontado em análise de pares, no entanto, os resultados desta pesquisa devem ser considerados com cautela, uma vez que os dados não foram normalizados.

Assim, os resultados sugerem que, de forma semelhante ao que ocorre no caso da pós-edição profissional, a atividade aplicada ao contexto acadêmico também resulta em menor esforço realizado. Contudo, é preciso ressaltar que a autopós-edição em contexto acadêmico envolve a edição de um texto por parte de leigos (tradutores não profissionais). Isso significa alguns desafios adicionais para lidar com o texto, tais quais: dificuldade de encontrar os erros e os pontos problemáticos; falta de conhecimentos especializados a respeito da língua e do processo tradutório; desconhecimento das limitações dos sistemas de tradução automática e dos erros que tipicamente cometem; inexperiência com a pré-edição do texto com o fim de gerar uma tradução automática de maior qualidade. Isso levanta questões a respeito da qualidade do texto final, uma vez que, para que a autopós-edição seja vantajosa, não basta que represente ganhos do ponto de vista do processo; é essencial que o produto final — o texto traduzido e revisado — alcance padrões de qualidade minimamente adequados ao contexto acadêmico.

No contexto da aprendizagem formal de L2, Niño (2008), trabalhando com estudantes avançados, mostrou que a autopós-edição, pelo menos quando executada por esse grupo, não impacta negativamente a qualidade do texto resultante. García (2011), a seu turno, trabalhando com aprendizes iniciantes de inglês, apontou que, comparado à escrita sem assistência do tradutor automático, a atividade de autopós-edição tem o potencial de ajudar esses sujeitos a se comunicar mais, melhor e com menos esforço, resultando em textos com mais palavras e maior qualidade.

Já em se tratando da escrita acadêmica em inglês L2, buscando investigar se os tradutores automáticos são realmente úteis nesse cenário e em que medida impactam na qualidade do texto resultante, O'Brien, Simard e Goulet (2018) conduziram uma pesquisa envolvendo dez acadêmicos: sete doutorandos; um pós-doutorando; um tutor da pós-graduação e um docente efetivo da universidade. Não houve restrições quanto à língua nativa dos sujeitos, de forma que havia participantes cuja L1 era o árabe (1), o chinês (1), o francês (4); o alemão (1); o romeno (1); e o espanhol (2).

Solicitou-se aos participantes que redigissem um resumo de aproximadamente 500 palavras em suas respectivas áreas de estudo, o qual poderia ser referente à pesquisa de doutorado que estavam desenvolvendo; a um artigo em que estivessem trabalhando; ou a um projeto de pesquisa futuro. Dos dez participantes, quatro escolheram a primeira opção, dois a segunda, e quatro a terceira.

A coleta dos dados junto aos participantes ocorreu em duas etapas: a primeira de escrita e a segunda de autopós-edição. Na primeira, o resumo deveria ser redigido em duas partes distintas, cada uma em uma língua — ou a L1, ou o inglês (L2). A parte A deveria conter o título, a contextualização da pesquisa e a sua motivação, enquanto a parte B deveria apresentar a metodologia e os impactos esperados. Para ambas as partes, foi solicitado um número médio de 250 caracteres. Um grupo redigiu a parte A em L1 e a B em L2, enquanto o outro fez o inverso. A ordem das tarefas foi aleatorizada.

Informou-se aos participantes que o objetivo dessa primeira etapa era produzir um rascunho do resumo. Assim, todos eles estavam cientes de que poderiam e iriam revisar o texto posteriormente. Finalmente, foi solicitado aos participantes que registrassem todos os recursos empregados na escrita (fora tradutores automáticos e a ajuda de colegas, que eram opções interditas) e o tempo total gasto em cada parte do resumo.

Para a realização da segunda etapa da coleta, a parte do resumo redigida na L1 dos sujeitos foi traduzida automaticamente para o inglês via Google Tradutor e o texto gerado por esse sistema substituiu a sua contraparte no texto original. Então, o texto, agora integralmente

em inglês, foi devolvido aos participantes para que fosse possível dar continuidade ao trabalho científico.

Na parte seguinte do experimento, voltada à autopós-edição, os participantes foram informados de que a parte originalmente escrita em L1 tinha sido automaticamente traduzida para o inglês e que, agora, deveria ser pós-editada de forma a tornar o resumo resultante um texto coerente e coeso. Essa pós-edição deveria ser realizada em cima de quatro orientações previamente fornecidas: 1) checar a correção gramatical das sentenças; 2) averiguar a tradução da terminologia especializada; 3) verificar a existência de adições ou omissões de informação, bem como de trechos não traduzidos; e 4) seguir as regras básicas de ortografia e pontuação.

Mais uma vez, foi-lhes pedido que registrassem o tempo gasto e os recursos empregados nesta etapa (com exceção dos interditos, que permaneceram os mesmos). Nesse ponto, os participantes estavam cientes de que a versão resultante de sua revisão seria considerada a versão final do texto. Portanto, foi-lhes autorizado revisar também a parte originalmente escrita em inglês, sob a condição de que, também nesse caso, registrassem tempo e recursos utilizados. Finalmente, após a realização dessa segunda etapa, solicitou-se aos participantes que respondessem um questionário a respeito da experiência de escrita pela qual haviam passado.

Os resumos resultantes foram então revisados por um revisor profissional e, também, submetidos ao Antidote, uma ferramenta de correção automática de erros. Assim, O'Brien, Simard e Goulet (2018) teceram suas análises em cima de quatro conjuntos de dados: o perfil dos participantes; o tempo (autorreportado) despendido durante a realização das tarefas e a autopercepção dos participantes sobre as tarefas; a correção realizada pelo revisor profissional e suas percepções sobre os textos; e a correções sugeridas pelo Antidote.

Com relação ao primeiro conjunto, aos dados relativos à L1 e ao grau de instrução dos participantes, mencionados anteriormente, acrescentaram-se informações a respeito: da experiência com a escrita acadêmica em inglês, com 50% tendo experiência substancial e 50%, nenhuma ou pouca; do nível de proficiência em inglês, com todos considerando-se pelo menos usuários intermediários da língua; e da experiência prévia com tradutores automáticos, com 40% dos participantes tendo recorrido à ferramenta em escritas acadêmicas anteriores.

No que diz respeito ao segundo conjunto, quanto ao tempo despendido, o tempo médio gasto na primeira etapa não variou de forma significativa entre as tarefas, mas, durante a segunda, a revisão da autopós-edição demandou mais tempo, o que sugere um maior esforço temporal quando se considera a totalidade do processo. Já sobre a autopercepção sobre as

tarefas, embora os participantes tenham ficado divididos quanto a qual foi a mais difícil — se a redação em L2 ou a autopós-edição —, a qualidade de ambos os textos foi reconhecida por seis dos participantes como equiparável.

Quanto ao terceiro e ao quarto, a correção realizada pelo revisor profissional não evidenciou diferenças de qualidade entre a parte escrita diretamente em inglês e a parte autopós-editada, uma vez que ele afirmou não haver diferença significativa de qualidade em seis dos dez textos, inclusive apontando um sétimo como apresentando uma autopós-edição de qualidade superior à redação em L2. Esse dado é, por sua vez, reforçado pelos resultados alcançados através do Antidote.

Parra Escartín *et al.* (2017) conduziram pesquisa semelhante, agora no domínio específico da medicina. As perguntas de pesquisa giravam em torno, justamente, da qualidade do texto produzido em uma autopós-edição realizada por esses profissionais. Todos os participantes eram falantes de espanhol como L1 e usuários do inglês como língua franca. O desenho experimental previa coletas de dados de diferentes naturezas. A primeira delas foi realizada por meio de um questionário a respeito dos hábitos desses profissionais quanto ao uso de tradutores automáticos.

O questionário recebeu 50 respostas e, do total de respondentes, cinco, com proficiência em inglês que variava do nível B1 ao C2¹², se disponibilizaram a participar da segunda coleta. Solicitou-se a eles, então, que enviassem uma publicação autoral (ou uma seção de uma) de aproximadamente 750 palavras, originalmente escrita em espanhol. Esse texto foi traduzido automaticamente para o inglês e devolvido aos participantes para que eles revisassem e corrigissem o resultado da tradução automática. Finalmente, a última coleta de dados se deu por meio de um questionário pós-tarefa, ao qual os cinco participantes também responderam.

Ao serem devolvidos pelos participantes, os textos autopós-editados foram anotados segundo as edições que haviam recebido, conforme fossem essenciais ou preferenciais. Além disso, erros e edições essenciais não realizadas também foram marcados. Esses textos foram, então, enviados a um tradutor e revisor especializado no domínio em questão para uma última revisão. Quando retornados, foram anotados segundo as edições realizadas pelo revisor profissional, adotando-se, para isso, as mesmas categorias aplicadas na primeira anotação.

Os resultados da pesquisa mostraram que os médicos foram capazes de realizar muitas edições essenciais e preferenciais por conta própria e sem instrução prévia em pós-edição. Apesar disso, eles também acabaram por introduzir novos erros, e o revisor profissional ainda

¹² Conforme Quadro Comum Europeu de Referência para Línguas (CEFR).

precisou realizar edições essenciais, de forma que não é possível afirmar que os textos autopós-editados atingiriam os parâmetros mínimos para a publicação sem passar, antes, por uma correção adicional. Uma vez que os participantes não haviam recebido instrução prévia em pós-edição, também não é possível saber em que medida essa variável teria alterado o cenário final.

Resultados como o de Parra Escartín *et al.* (2017) não invalidam o interesse em seguir estudando esse tópico. Como os próprios autores colocam, eles apenas apontam novos caminhos. Ainda que uma pós-edição de alta qualidade requeira um tradutor profissional, “ciente das palavras cujos diferentes significados dependem do contexto extralinguístico”¹³ (Vasconcellos; León, 1985, p. 2), o principal objetivo é que o uso de sistemas de tradução automática permita aos estudantes de pós-graduação produzir textos que atinjam os padrões mínimos exigidos para aceitação.

Tanto na pesquisa de O’Brien, Simard e Goulet (2018) quanto na de Parra Escartín *et al.* (2017), a análise dos erros (pós-edições nada mais são do que correções de erros) foi utilizada como meio de aferir a qualidade linguística do texto. Esse não é o único método possível para acessar a qualidade de um texto. É possível, por exemplo, submetê-lo a uma avaliação por especialistas, por meio da sua comparação a um texto de referência ou da atribuição de notas segundo determinados critérios. Pode-se, ainda, recorrer a métricas de avaliação automática, amplamente utilizadas na análise de sistemas de tradução. Mas a análise de erros oferece a vantagem de informar não só se um texto é problemático, mas porque o é; afinal, trata-se do “processo de determinar a incidência, a natureza, a causa e as consequências da linguagem mal-sucedida”¹⁴ (James, 2013, p. 1).

De fato errar é parte integrante de qualquer processo de aprendizagem, seja humano ou de máquina, e aprender com os erros cometidos é um aspecto crucial para que a expertise seja aprimorada (Federico *et al.*, 2014). No caso da análise de erros humanos, examiná-los joga luz sobre a natureza do processo de aprendizagem de uma língua estrangeira, as partes da língua-alvo que os estudantes mais apresentam dificuldades em produzir e os tipos de erro que mais prejudicam a sua comunicação eficaz (Dulay, Burt, Krashen, 1982, p. 138). No tocante à análise de erros em traduções automáticas, seu objetivo, na maioria das vezes, é compreender como esses erros se distribuem entre categorias previamente definidas (Popović, 2018); com

¹³ Nossa tradução para: “only an experienced translator will be aware of the words whose variable meanings are dependent on extra linguistic context”

¹⁴ Nossa tradução para: “is the process of determining the incidence, nature, causes and consequences of unsuccessful language”.

base nesse perfil distribucional, é possível realizar o ajuste (*tuning*) do modelo, aprimorando-o.

Contudo, para que os erros sejam analisados precisam, primeiro, ser classificados. Para isso, são elaboradas taxonomias de erros, entendidas como um conjunto criterioso de categorias de falhas (James, 2013; Popović, 2018). Tais categorias são definidas segundo o conceito de erro adotado e os critérios estabelecidos para a sua classificação, os quais “devem, na medida do possível, refletir fatos objetivos e observáveis sobre as entidades a serem classificadas”¹⁵ (James, 2013, p. 102). Na próxima seção, trataremos em mais detalhes sobre a análise de erros humanos e de sistemas de tradução automática e sobre o desenvolvimento de taxonomias para esses fins.

2.3 Análise de erros e taxonomias de erros

2.3.1 Análise de erros humanos

Apresentando-se como uma alternativa à Análise Contrastiva, a Análise de Erros já foi utilizada, no contexto do ensino e aprendizagem de línguas estrangeiras, para o estudo da competência linguística; dos processos e estratégias de aprendizagem; e da pedagogia das línguas. Nesse contexto, a classificação de erros — e, por consequência, o desenvolvimento de taxonomias — frequentemente se baseou na atribuição de fontes aos erros. Ou seja, os erros eram definidos segundo as suas causas, de forma que a descrição era feita com base em categorias ligadas à explicação.

Segundo Dulay, Burt, Krashen (1982, p. 144), no entanto, “seria mais apropriado utilizar as taxonomias para organizar os erros de acordo com características diretamente observáveis”¹⁶. Das quatro que o autor apresenta, as duas primeiras, que se encontram na base da que adotaremos, são as únicas que, de fato, se atêm à descrição dos erros sem enveredar por suas causas ou sua gravidade (James, 2013). No primeiro caso, trata-se da classificação baseada em categorias linguísticas. No segundo, dos erros classificados com base na superfície do texto produzido.

A primeira categorização localiza os erros segundo o nível em que ocorrem (como fonologia, grafologia ou gramática) e o tipo de estrutura que afetam (estruturas passivas, complementos sentenciais ou semelhantes). Como exemplos, Dulay, Burt, Krashen (1982)

¹⁵ Nossa tradução para: “should as far as possible reflect observable objective facts about the entities to be classified”.

¹⁶ Nossa tradução para: “taxonomies might more appropriately be used to organize errors according to directly observable characteristics”.

mencionam duas taxonomias, que julgam representativas do restante: a de Politzer e Ramirez (1973, *apud* Dulay, Burt, Krashen, 1982) e a de Burt e Kiparsky (1972, *apud* Dulay, Burt, Krashen, 1982). A primeira traz duas grandes categorias, Morfologia e Sintaxe, subespecificadas segundo os diferentes tipos de erros que abrangem, tais como problemas no uso de artigos indefinidos e omissão do verbo. A segunda distingue seis categorias — esqueleto das orações em inglês; o sistema de auxiliares; sentenças passivas; conjunções temporais; complementos sentenciais; e predicados psicológicos —, que também contam com subdivisões.

Quanto ao segundo tipo de taxonomia, baseado na superfície textual, suas categorias se ancoram nos erros identificados nos textos dos alunos quando comparados a um texto ideal hipotético bem elaborado. Por isso mesmo, James (2013) propõe chamar essa classificação de Taxonomia de Modificação de Alvo (Target Modification Taxonomy). Para ela, Dulay, Burt, Krashen (1982) apresentam quatro categorias, posteriormente subespecificadas: omissão; adição (incluindo-se regularizações, marcações duplas e adição simples); formações incorretas (subdivida em arqui-formas, regularizações e formas alternantes) e ordenações incorretas. James (2013), ao apresentar essa taxonomia, propõe uma versão alternativa, que considera mais apropriada, a qual conta com cinco categorias: omissão, excesso de adições (*overinclusion*), seleção incorreta, ordenação incorreta, e mesclas (*blends*).

Embora ambos os tipos de taxonomia — linguística e de modificação de alvo — tenham sido concebidos para aplicação em um contexto de aprendizagem de língua, tanto um quanto o outro podem ser adotados para a análise da produção escrita em geral. No caso do primeiro, por lidar com categorias que dizem respeito à realização linguística dos erros. No segundo, por abordar erros tendo por parâmetro um texto alvo ideal, que, no fim, também é o objetivo da atividade de produção escrita.

Além disso, embora Dulay, Burt, Krashen (1982) apresentem as duas possibilidades de classificação separadamente, James (2013) afirma que a combinação das duas seria bastante vantajosa. De fato, combiná-las é o que fazem Costa *et al.* (2015), autores da taxonomia adotada pelo presente estudo, a qual será descrita na seção 2.3.2.1. Antes, no entanto, abordaremos a análise de erros no contexto da tradução automática.

2.3.2 Análise de erros de sistemas de tradução automática

A análise de erros pode ser realizada com base em diferentes parâmetros. Popović (2018) menciona quatro possibilidades. De acordo com a autora, é possível analisar erros tradutórios: 1) segundo as pós-edições (*edits*) realizadas; 2) de acordo com os erros

cometidos, 3) considerando padrões linguísticos não coincidentes; 4) com base na presença de atributos (*features*) linguísticos específicos. Dentre essas opções, apenas as duas primeiras envolvem a classificação explícita dos erros. Assim, é sobre elas que nos deteremos.

No primeiro caso, trata-se de um método voltado à classificação das ações de pós-edição (*post-editing operations*), ou seja, das correções realizadas no texto de saída do sistema de tradução automática. Isso pode ser feito manualmente, como no caso de Blain *et al.* (2011), ou automaticamente, como em Snover *et al.* (2006). Para que a classificação de edições seja, de fato, uma boa alternativa, é preciso que a pós-edição seja realizada por tradutores qualificados, de forma a reduzir as chances de correções incorretas. Embora essa seja uma tarefa exigente, sempre haverá a necessidade de algum esforço humano (Popović, 2018). Dessa forma, a análise das ações de pós-edição apresenta-se como um método mais confiável do que aqueles que implicam a análise de traduções cruas, uma vez que a tradução automática gerada já terá sido submetida a uma análise e correção de erros.

No caso da análise de erros, ela também pode ser realizada de forma automática ou com base em anotação manual. Quando realizada automaticamente, consiste em um método que visa “estimar a quantidade de diferentes tipos de erros”¹⁷ (Popović, 2018, p. 142). Isso pode ser feito, por exemplo, com o auxílio de medidas como a Word Error Rate (WER) e a Position-independent Word Error Rate (PER), às quais Popović e Ney (2006) recorrem para estimar a quantidade de erros de ordenação (*reordering*) e flexão (*inflectional*). Pode ser feito, também, com recurso a uma taxonomia pré-estabelecida, como ocorre no caso do trabalho de Popović e Ney (2011). Em comparação a uma anotação manual, o escrutínio automático é menos custoso, mais rápido e mais consistente. Entretanto, as ferramentas de análise automática também estão sujeitas a imprecisões, especialmente quando a classificação envolve categorias que representam desafios até mesmo para os anotadores humanos. Além disso, demandam traduções de referência e taxonomias menos detalhadas, posto não serem capazes de operar com um número maior de categorias.

Contudo, para a presente pesquisa, que não envolve tradutores profissionais e demanda uma taxonomia mais complexa, o método mais adequado provou ser a análise com base em anotação manual de erros, que se refere à prática de selecionar as palavras incorretas, etiquetá-las segundo o tipo de erro que apresentam, contabilizar as etiquetas e analisar os resultados. Como os dois anteriores, esse método também apresenta desvantagens: além de demandar bastante recursos — humanos, financeiros, temporais —, pode gerar anotações inconsistentes entre si, a depender do número e da complexidade das categorias componentes

¹⁷ Nossa tradução para: “[...] to estimate the amount of different error types”

da taxonomia, bem como do perfil dos anotadores envolvidos na pesquisa. Contudo, a classificação manual proporciona acesso a uma maior riqueza de detalhes, uma vez que “anotadores humanos podem distinguir um número maior de classes de erro do que as ferramentas automáticas mais avançadas”¹⁸ (Popović, 2018, p. 154) e, por isso, foi escolhida no caso deste estudo. Na próxima seção, detalharemos algumas das taxonomias empregadas nesse tipo de anotação.

2.3.2.1 Taxonomias: classificação manual de erros de tradução automática

Em 1994, Flanagan propôs a primeira taxonomia de erros para avaliar a qualidade de textos gerados por tradutores automáticos, com foco nos pares linguísticos inglês-francês e inglês-alemão. Seu objetivo era criar uma classificação útil para os usuários desses sistemas à época. O Quadro 1 apresenta a taxonomia desenvolvida para o par linguístico inglês-francês.

Quadro 1 - Taxonomia proposta por Flanagan (1994)

Categoria	Descrição
Ortografia	Palavra escrita de forma incorreta
Palavra não encontrada	Palavra que não está no dicionário
Acentuação	Acentuação utilizada de forma incorreta
Maiúsculas e minúsculas	Uso incorreto de letras maiúsculas ou minúsculas
Elisão	Utilização incorreta da elisão
Flexão verbal	Verbo flexionado de forma incorreta
Flexão nominal	Substantivo ou adjetivo flexionado de forma incorreta
Outra flexão	Advérbio ou outras palavras flexionadas incorretamente
Reorganização	Palavras ou frases fora da ordem correta na sentença
Categoria gramatical	Seleção incorreta da classe gramatical da palavra
Pronome	Uso incorreto, ausência ou excesso de pronomes
Artigo	Ausência ou excesso de artigos
Preposição	Uso incorreto, ausência ou excesso de preposições
Negação	Utilização incorreta de palavras negativas ou construção negativa incorreta
Conjunção	Uso incorreto ou falha em identificar as partes de uma frase ligadas por conjunções
Concordância	Falta de concordância entre as palavras (sujeito-verbo, adjetivo-substantivo, etc.)
Limite da oração	Falha na identificação do início e/ou do fim de uma oração
Escolha de palavra	Escolha de palavra inadequada (uma única palavra)
Expressão	Tradução incorreta de expressões multipalavras

Fonte: Traduzido e adaptado de Flanagan (1994, p. 67)

¹⁸ Nossa tradução para: “human annotators can distinguish a larger number of error classes than state-of-the-art automatic tools”.

Para a autora, cada par linguístico demandava um conjunto específico de categorias (Flanagan, 1994). Assim, no caso do par linguístico inglês-alemão, a categoria “elisão” foi descartada e três outras foram acrescentadas: pronome relativo (ausência ou uso incorreto), caso (terminação incorreta) e pontuação (ausente, incorreta ou desnecessária).

Então, em 2004, Elliott *et al.* realizaram uma pesquisa com o objetivo de desenvolver um sistema de detecção automática de erros, capaz de operar eficientemente sem depender de traduções humanas como referência. Para isso, como etapa intermediária, pesquisadores precisaram realizar a anotação manual de erros presentes em traduções automáticas, adotando, para tal, uma taxonomia composta por 58 categorias. Uma vez que não era viável para um sistema automático de detecção operar com uma taxonomia tão detalhada, a finalidade da análise dos erros anotados manualmente era justamente descobrir quais das 58 categorias eram mais relevantes.

Devido ao propósito da pesquisa, o esquema de classificação foi pensado, inicialmente, para abarcar apenas erros de fluência que pudessem ser identificados sem acesso ao texto fonte. No entanto, alguns desses erros se provaram falhas de adequação detectáveis a despeito da ausência do texto de partida. Assim, a categorização final acabou possibilitando a marcação de ambos os tipos. O processo de anotação foi realizado tendo-se em vista as ações de pós-edição de que o texto de saída necessitaria, e um resumo da taxonomia pode ser conferido no Quadro 2.

Quadro 2 - Taxonomia proposta por Elliott *et al.* (2004)

Categoria	Subcategoria (1)	Subcategoria (2)	Subcategoria (3)	Subcategoria (4)
Classe de palavras	Substantivo composto	Cadeia de substantivos ou entidade nomeada	Inadequado	Significado integralmente preciso/impreciso
				Parte do significado preciso/impreciso
			Não traduzido	Integralmente
				Em partes
	Acrônimo	Incorreto		
	Pronome	Inadequado		Anáfora incorreta
				Outro
				Não traduzido
		Não necessário		
		Omitido		Pronome objeto direto
Pronome relativo				
Outro				

Fonte: Traduzido e adaptado de Elliott *et al.* (2004)

Quadro 2 - Taxonomia proposta por Elliott *et al.* (2004)

Categoria	Subcategoria (1)	Subcategoria (2)	Subcategoria (3)	
Classe de palavras	Substantivo, adjetivo, advérbio ou conjunção	Inadequado	Significado preciso/impreciso	
			Extremamente improvável	
		Não traduzido		
		Não necessário		
	Omitido			
	Determinante	Inadequado		
			Não traduzido	
		Não necessário		Artigo definido
				Artigo indefinido
				Outro
		Omitido		Artigo definido
			Artigo indefinido	
			Outro	
	Verbo	Inadequado		Significado preciso/impreciso
				Extremamente improvável
				Verbo composto
		Não traduzido		
		Não necessário		
Omitido				
Tempo ou conjugação	Tempo ou modo			
	Conjugação			
Posição incorreta	Acrônimo, pronome, substantivo comum, adjetivo, advérbio, conjunção, preposição, determinante, verbo, negador ou apêndice na cadeia de substantivos			
	Sequência de substantivos compostos	Cadeia de substantivos ou Entidade nomeada	Ordem das palavras	
			Arranjo	
Outro	Classe de palavra incorreta / estilística inapropriada / expressão incompreensível / erro de ortografia / negação incorreta / número ordinal não traduzido / qualificador não necessário			
	Número	Singular que deveria ser plural ou vice-versa		
	Caixa alta/baixa	Letra maiúscula/ letra minúscula requerida		

Fonte: Traduzido e adaptado de Elliott *et al.* (2004)

Na sequência, com o objetivo de ocupar a lacuna deixada por medidas automáticas de avaliação¹⁹, Vilar *et al.* (2006) desenvolveram a taxonomia que se tornou referência para muitas das subsequentes²⁰. Com base na categorização oferecida por Llitjós, Carbonell, Lavie (2005), estabeleceram cinco grandes categorias — Palavras Ausentes, Ordem das Palavras, Palavras Incorretas, Palavras Desconhecidas e Pontuação —, então subdivididas para maior especificidade. Um resumo está apresentado no Quadro 3.

Quadro 3 - Taxonomia proposta por Vilar *et al.* (2006)

Categoria	Subcategoria (nível 1)	Subcategoria (nível 2)
Palavras ausentes	Palavras de conteúdo	
	Palavras de preenchimento	
Ordem das palavras	Nível da Palavra	Curta extensão
		Longa extensão
	Nível da sequência sintagmática	Curta extensão
		Longa extensão
Palavras incorretas	Sentido	Escolha lexical incorreta
		Desambiguação incorreta
	Forma incorreta	
	Palavras extras	
	Estilística	
Expressões idiomáticas		
Palavras Desconhecidas	Radical desconhecido	
	Forma desconhecida	
Pontuação		

Fonte: Traduzido e adaptado de Popović (2018).

De nome autoexplicativo, a primeira categoria subdivide-se segundo o tipo de palavra ausente: de conteúdo, que são essenciais ao sentido do texto; e de preenchimento, destinadas apenas à possibilitar a formação de uma frase gramaticalmente correta (Vilar *et al.*, 2006). Por sua vez, a segunda categoria, relativa à ordem, diferencia-se conforme o nível — da palavra (uma única palavra no lugar incorreto) ou da sequência sintagmática (um sintagma mal posicionado) — e a extensão — curta, quando envolve uma única sequência sintática (*syntactic chunk*); ou longa, quando compromete mais de uma sequência.

A terceira categoria aplica-se àqueles casos em que o tradutor automático traduz

¹⁹ Os autores mencionam especificamente a Word Error Rate, a Position independent word Error Rate, a BLEU (Papineni *et al.*, 2002, *apud* Vilar *et al.*, 2006) e a NIST (Doddington, 2002, *apud* Vilar *et al.*, 2006).

²⁰ De fato, Bojar (2011) propõe uma versão apenas ligeiramente modificada, na qual remove a categoria das Palavras Desconhecidas e altera a das Palavras Incorretas, excluindo as subcategorias Estilística e Expressões idiomáticas e adicionando a subcategoria de Palavras não traduzidas.

incorretamente uma palavra. Sendo a maior categoria entre as cinco, é também a que apresenta mais subdivisões. A primeira diz respeito aos problemas de sentido — ou a palavra é polissêmica e foi desambiguada incorretamente; ou a palavra presente na tradução não tem relação alguma com a que se encontra no texto fonte. A segunda, referente à forma das palavras, precisa ser subcategorizada segundo o par linguístico envolvido²¹. A terceira concerne às palavras que os sistemas de tradução automática podem adicionar ao texto de saída, ocorrência comum quando se lida com a tradução de textos orais (Vilar *et al.*, 2006). A quarta compreende os problemas estilísticos. Finalmente, a quinta categoria engloba a tradução incorreta de expressões idiomáticas.

A penúltima categoria, relativa às palavras desconhecidas, é subcategorizada conforme envolva uma palavra realmente desconhecida (raiz desconhecida) ou uma forma desconhecida de uma palavra conhecida (forma desconhecida). Finalmente, a última categoria é a dos erros de pontuação que, devido ao baixo impacto que tinha para a pesquisa de Vilar *et al.* (2006), não recebeu maiores especificações.

Em 2010, no contexto de um trabalho que busca correlacionar diferentes métodos de avaliação da qualidade de traduções automáticas (dentre os quais a classificação e análise de erros²²), Farrús *et al.* (2010)²³ desenvolvem um conjunto de cinco categorias de erros linguisticamente motivadas, objetivas e específicas. Com isso, visavam minimizar o impacto do julgamento humano e obter informações direcionadas a respeito dos erros linguísticos encontrados nos textos de saída dos sistemas de tradução automática. Para reforçar tal propósito, além da própria lista de categorias, forneceram ao anotador responsável orientações o mais detalhadas possível. Um resumo da classificação de erros pode ser conferido no Quadro 4.

Quadro 4 - Taxonomia proposta por Farrús *et al.* (2010)

Categoria
Erros morfológicos
Erros lexicais
Erros ortográficos
Erros sintáticos
Erros semânticos

Fonte: Traduzido e adaptado de Popović (2018)

²¹ Vilar *et al.* (2006) citam o caso do espanhol, no qual subcategorias relevantes poderiam ser as de concordância e de tempo verbal.

²² Os demais métodos são as métricas automáticas (BLEU e TER); a avaliação perceptual humana, via comparação de traduções geradas por diferentes sistemas de tradução.

²³ Comelles *et al.* (2012) propõem uma taxonomia similar à de Farrús *et al.* (2010), mas seu propósito é a avaliação automática de erros.

Como pode ser observado, a taxonomia conta com categorias que independem da língua, relativas ao nível linguístico em que o erro se encontra. Para os autores, apenas as subcategorias devem depender do par linguístico envolvido — no caso, catalão-espanhol. Antes de prosseguir para a próxima taxonomia, é interessante destacar um dos resultados obtidos nessa pesquisa, que dialoga com os que obtivemos para a pesquisa atual. Ainda que não tenha sido possível para os autores estabelecer uma correlação entre os três métodos de análise de erros envolvidos, eles puderam observar que erros lexicais e semânticos parecem ser aqueles que mais influenciam a percepção humana da qualidade da tradução automática gerada.

Por sua vez, Kirchhoff *et al.* (2012) desenvolveram uma taxonomia baseada na de Vilar *et al.* (2006), até mesmo por estarem trabalhando no mesmo par linguístico²⁴. A anotação dos erros foi realizada no contexto de uma pesquisa que, por meio de análise conjunta, buscou investigar as preferências dos usuários em relação aos tipos de erros de tradução encontrados em textos específicos. Os erros foram categorizados em doze grupos, cinco deles com subdivisões, como mostra o Quadro 5.

Quadro 5 - Taxonomia proposta por Kirchhof *et al.* (2012)

Nível 1	Nível 2
Palavras ausentes	Palavras lexicais
	Palavras gramaticais
Palavras extras	Palavras lexicais
	Palavras gramaticais
Ordem das palavras	Curta extensão
	Longa extensão
Morfologia	Verbal
	Nominal
Erro de sentido da palavra	
Pontuação	
Ortografia	
Uso de maiúsculas	
Não traduzido	Termo médico
	Nome próprio
Pragmática	
Diacríticos	
Outros	

Fonte: Traduzido e adaptado de Popović, 2018.

²⁴ Ambos os trabalhos operam com o par inglês-espanhol.

Diferentemente das taxonomias anteriores, esta inclui a categoria “Outros” para abarcar erros não contemplados pelas demais. Embora pouco informativa, sua inclusão evita a necessidade de expandir a tipologia indefinidamente. Como o foco da pesquisa era a preferência dos usuários e não a diversidade dos erros, essa adição não comprometeu os resultados.

A pesquisa revelou que a preferência dos usuários por uma tradução não está necessariamente relacionada à quantidade de erros presentes, pois alguns tipos de erro têm maior impacto na escolha do que outros (Kirchhoff *et al.*, 2012, p. 125) — constatação que dialoga com o achado de Farrús *et al.* (2010). A título de exemplo, entre os erros analisados, “Ordenação das palavras” foi o menos aceito pelos usuários, apesar de ser o menos frequente, enquanto o tipo mais comum, “Erros morfológicos”, ocupou apenas a terceira posição em termos de rejeição.

Com propósito similar ao de Kirchhoff *et al.* (2012), Federico *et al.* (2014, p. 1634) propuseram “um modelo estatístico para analisar o impacto dos diferentes tipos de erros na qualidade do texto de saída do sistema de tradução automática”²⁵. Para validá-lo, realizaram testes com três línguas-alvo — chinês, árabe e russo — tendo o inglês como língua de partida. As categorias segundo as quais a anotação dos erros foi realizada remontam, em parte, às de Vilar *et al.* (2006): erros morfológicos; erros no léxico (escolhas lexicais incorretas e palavras extras); palavras ausentes e erros na ordenação. Em Girardi *et al.* (2014), há menção ainda a outras categorias, tais como “capitalização e pontuação” e “erros excessivos”²⁶.

Quadro 6 - Taxonomia proposta por Federico *et al.* (2014) e Girardi *et al.* (2014)

Categoria
Erros morfológicos
Escolha lexical
Adições
Omissões
Uso de maiúsculas e de pontuação
Erros de reordenação
Muitos erros

Fonte: Traduzido e adaptado de Popović (2018).

²⁵ Nossa tradução para: “robust statistical framework for analysing the impact of different error types on machine translation (MT) output quality.”

²⁶ Castilho *et al.* (2017a; 2017b) adotam taxonomia similar, excluindo duas categorias: “punctuation” e “too many errors” (*apud* Popović, 2018)

Dentre os achados da pesquisa, está o fato de que “a percepção humana de qualidade não depende necessariamente da frequência de um determinado tipo de erro”²⁷ (Popović, 2018, p. 134), reforçando aquilo que Farrús *et al.* (2010) e Kirchoff *et al.* (2012) já haviam apresentado.

No contexto das investigações voltadas para a concordância entre anotadores, destaca-se a de Stymne e Ahrenberg (2012). Com o objetivo de superar as limitações de uma avaliação anterior, marcada por baixa concordância entre os sistemas, os autores desenvolveram e testaram novas estratégias para aumentar a consistência dos resultados. Discutiram o impacto de fatores como o nível de detalhamento da taxonomia, a restrição da análise a sentenças curtas e o treinamento dos anotadores. Propuseram, então, uma taxonomia baseada em diferentes medidas e tipologias, como as de Flanagan (1994), Vilar *et al.* (2006) e Farrús *et al.* (2010). Um resumo das categorias está disponível no Quadro 7.

Quadro 7 - Taxonomia proposta por Stymne e Ahrenberg (2012)

Nível 1	Nível 2
Taxa de erros	Palavras faltantes
	Palavras extras
	Palavra errada
	Ordem das palavras
Linguístico	Ortografia
	Semântica
	Sintaxe
GF	Palavras gramaticais
	Palavras funcionais
Forma	Categorias morfológicas
Classe de palavras e outros	Classe gramatical
	Pontuação
FA	Fluência
	Adequação
	Nem um nem outro
	Ambos
Reo (causa da reordenação)	
Index (posição do erro)	
Outros (outras categorias)	
Ser (seriedade do erro)	

Fonte: Traduzido e adaptado de Popović (2018).

²⁷ Nossa tradução para: “the human perception of quality does not necessarily depend on the frequency of a given error type”.

O estudo mostrou que o número de categorias e o treinamento são fundamentais para aumentar a concordância, enquanto a localização exata dos erros continua sendo um ponto de divergência frequente (Popović, 2018).

A seu turno, Lommel *et al.* (2014) exploram a viabilidade da Multidimensional Quality Metric (MQM) para a análise de textos de saída de tradutores automáticos. Desenvolvida pelo QTLaunchPad project (Dombeck, 2025c), a MQM oferece um conjunto amplo de categorias que possibilite a seleção de subconjuntos adequados a tarefas particulares (Lommel *et al.*, 2014; Popović, 2018). Para o seu estudo, os autores selecionaram um subconjunto dentre as mais de 80 categorias disponibilizadas pelo projeto. A taxonomia efetivamente empregada pode ser encontrada no Quadro 8. Como se pode ver, ela é construída do mais amplo ao mais específico. As duas categorias mais gerais dizem respeito à recepção dos textos enquanto traduções (acurácia) e enquanto construto linguístico (fluência); no primeiro caso, a acurácia se subdivide em categorias mais específicas ligadas às operações tradutórias; no segundo, as subcategorias ligam-se à gramática, ao discurso e à ortografia.

Quadro 8 - Taxonomia de Lommel *et al.* (2014) com base na MQM

Nível 1	Nível 2	Nível 3	Nível 4
Acurácia	Tradução incorreta		
	Terminologia incorreta		
	Omissão de informação		
	Adição de informação		
	Parte não traduzida		
Fluência	Erros gramaticais	Erros morfológicos (forma)	
		Erros em palavras funcionais	Ausência
			Uso excessivo
	Uso incorreto		
	Erro de registro/estilística	Uso incorreto da letra maiúscula	
	Erro de ortografia		
	Erro de formatação	Erro de pontuação	
	Texto incompreensível		

Fonte: Traduzido e adaptado de Popović (2018).

Finalmente, com base nas taxonomias de Vilar *et al.* (2006), Bojar (2011) e Dulay, Burt, Krashen (1982), Costa *et al.* (2015, p. 127) propõem uma classificação linguisticamente motivada e informada, que busca expandir as que a precederam. Trabalhando no par linguístico inglês-português europeu, os autores têm como objetivo propor categorias que atendam às particularidades das línguas românicas, visto que “todas as taxonomias são

influenciadas pelas idiosincrasias das línguas nas quais operam”²⁸ (Costa *et al.*, 2015, p. 130). Além disso, buscam criar uma taxonomia passível de expansão e aplicável tanto a erros de máquina quanto humanos. Essas características fazem dessa taxonomia a mais adequada à presente pesquisa e, por isso, os erros do *corpus* de resumos acadêmicos serão anotados segundo as suas categorias. Um resumo da taxonomia pode ser visto no Quadro 9²⁹

Quadro 9 - Taxonomia proposta por Costa *et al.* (2015)

Categoria	Subcategoria	
Ortografia	Pontuação	
	Letra maiúscula	
	Digitação	
Léxico	Omissão	Palavra lexical
		Palavra gramatical
	Adição	Palavra lexical
		Palavra gramatical
	Não traduzido	
Gramática	Verbos	Classe de palavras
		Tempo/Modo
		Pessoa/Número
	Concordância (nome)	Ambos (blend)
		Gênero
		Número
		Pessoa
	Escolha incorreta	Mais de um (blend)
		Contração
	Ordem incorreta	
Semântica	Sentido incorreto	
	Palavra incorreta	
	Colocação incorreta	
	Expressões idiomáticas	
Discurso	Estilística	
	Dialeto	
	Não deveria traduzir	

Fonte: Traduzido e adaptado de Costa *et al.* (2015)

²⁸ Nossa tradução para: “All taxonomies are influenced by the idiosyncrasies of the languages with which they are working”.

²⁹ Para uma versão expandida com exemplos, ver Apêndice A.

A taxonomia parte de cinco categorias gerais, correspondentes ao componente linguístico afetado pelo erro: Ortografia; Léxico; Gramática; Semântica e Discurso. Elas recebem, então, de um a cinco níveis de subdivisão. Assim, na categoria Ortografia (nível dos caracteres), os autores distinguem entre os erros de pontuação, de uso da letra maiúscula e de grafia. Ainda que o uso incorreto da letra maiúscula seja um erro de grafia, os autores optaram por mantê-los como subcategorias separadas porque, se por um lado erros de capitalização são comuns em traduções automáticas, erros de grafia são mais comuns de serem cometidos por humanos. Assim, essa subdivisão permite capturar essa diferença de performance.

No que tange ao Léxico, apontam, primeiro, as omissões ou adições de palavras lexicais e gramaticais; sinalizam, também, a não tradução de algumas palavras. A manutenção de palavras na língua original é um erro mais comum de ser cometido por tradutores automáticos do que por humanos. No caso destes, a tendência é, na verdade, traduzir o que não necessariamente deveria ser traduzido. Esse erro também é contemplado pela taxonomia, mas localiza-se no nível discursivo.

O componente da Gramática é aquele que conta com mais subdivisões, sendo, portanto, a classe mais detalhada. Dentro dela, os autores destacam duas categorias mais amplas. A primeira, “Escolha incorreta” (*misselection*), reúne a escolha inadequada de classe de palavras; os problemas de concordância verbal (tempo e/ou pessoa verbal) e nominal (gênero, número e/ou pessoa); a incorreção na realização de contrações; e a ordenação incorreta das palavras. Diferente de Vilar *et al.* (2006), não há distinção quanto ao nível e ao alcance.

Em relação à Semântica, sinalizam as possibilidade de sentido incorreto (ou desambiguação incorreta, conforme a taxonomia de Vilar *et al.* (2006); palavra incorreta (ou escolha lexical incorreta, conforme a taxonomia de Vilar *et al.* (2006); colocação incorreta (tradução incorreta de blocos de palavras que geralmente ocorrem juntos na língua fonte); e expressões idiomáticas erroneamente traduzidas.

Finalmente, no nível do Discurso, focalizam os problemas estilísticos (*style*), ligados a inadequações como repetições de palavras; as questões relacionadas à variação linguística (tais como as diferenças dialetais entre o português brasileiro e o europeu); e, como mencionado anteriormente, a tradução de termos que não deveriam ser traduzidos, tais como títulos de livros, nomes de lugares.

Antes de prosseguir ao próximo tópico, no qual serão apresentados alguns desafios implicados na concepção de taxonomias, é interessante destacar algumas das contribuições da pesquisa de Costa *et al.* (2015) no que tange à elucidação da relação entre os erros e a

qualidade de uma tradução. De acordo com os autores, os erros lexicais e os semânticos são aqueles que têm maior impacto no nível da sentença, dialogando com os resultados de Farrús *et al.* (2010). Além disso, sentenças melhor avaliadas apresentam um número inferior de erros gramaticais, mas a exata relação entre sentenças mal avaliadas e os erros gramaticais ainda necessita de maiores esclarecimentos (Popović, 2018).

2.3.3 Desafios no desenvolvimento e na aplicação de taxonomias

Embora de indiscutível relevância, o desenvolvimento de taxonomias não é uma tarefa simples (Popović, 2018). Isso porque implica uma série de decisões que, invariavelmente, resultam em perdas. Uma delas, por exemplo, envolve o número de categorias que compõem a tipologia de erros. O paradoxo reside no fato de quanto mais categorias, maior detalhe de informações, mas menor consistência entre diferentes anotadores e mais possibilidade de que ferramentas de anotação automática cometam falhas — o que acaba prejudicando as informações obtidas. Além disso, no caso de anotações manuais, quanto mais numerosas as possibilidades, maior o esforço cognitivo demandado dos anotadores.

No entanto, conforme sublinhado por Popović (2018, p. 139), “não apenas o número, mas a exata definição das categorias é muito importante”³⁰. Sem que cada uma esteja bem definida, não há meio de se manter a consistência nas anotações. Assim, é preciso que se tenha exata noção de a que tipo de erro cada categoria atende, de forma a não desenvolver uma taxonomia insuficiente ou redundante. Esse ponto coloca em questão o conflito entre a importância de uma taxonomia individualizada e a necessidade de uma taxonomia suficientemente una e geral. Uma solução para isso é estabelecer “um conjunto pequeno e generalizado de amplas classes de erros como ponto de partida, permitindo sua expansão em diferentes direções e profundidades, dependendo da tarefa/par de línguas/domínio”³¹ (Popović, 2018, p. 140), estratégia adotada por algumas das taxonomias apresentadas na seção anterior, inclusive a de Costa *et al.* (2015), adotada neste estudo.

Contudo, ainda assim, restam problemas ligados à categorização de palavras ou trechos nos quais ocorre mais de um erro, o que não é incomum, uma vez que um erro pode dar origem a outro (Flanagan, 1994; Vilar *et al.*, 2006). É controversa, também, a determinação da posição exata dos erros, ou seja, quais palavras ou trechos eles implicam. Isso porque “os erros às vezes envolvem apenas palavras isoladas, mas com mais frequência

³⁰ Nossa tradução para: “not only the number, but also the exact definition of error categories is very important”.

³¹ Nossa tradução para: “generalised small set of broad error classes as a starting point and enabling its expansion in distinct directions and depths depending on the task/language pair/domain.”

envolvem frases, expressões descontínuas, ordem de palavras ou relações que atravessam os limites das sentença”³² (Flanagan, 1994, p. 65). Isso explica, por exemplo, a baixa concordância entre diferentes anotadores.

Não obstante as dificuldades, a análise de erros segue sendo uma atividade de suma importância, uma vez que fornece informações importantes a respeito das dificuldades, inerentes a processos de aprendizagem, enfrentadas pela máquina e por humanos. Assim, pesquisas como a aqui empreendida trazem resultados que podem ser aplicados tanto no ensino e aprendizagem de línguas quanto, potencialmente, no ajuste dos tradutores automáticos. Antes de apresentar os resultados obtidos, no entanto, é necessário descrever a metodologia empregada no estudo, à qual se dedica a próxima seção.

3 METODOLOGIA

3.1 Metodologia de coleta, compilação, armazenamento e anotação do corpus

3.1.1 O corpus: contexto de coleta e participantes

Os resumos que integram o corpus desta pesquisa foram produzidos no contexto do projeto de Nunes (2021), previamente aprovado pelo Comitê de Ética³³. Todos os doze sujeitos envolvidos na produção dos textos eram, na época da coleta, pós-graduandos de diferentes cursos da UFMG que conduziam pesquisas em uma das três grandes áreas do CNPq: Ciências Exatas e da Terra (CET), Ciências Sociais Aplicadas (CSA) ou Ciências Humanas (CH). Para serem incluídos, precisavam atender a cinco critérios: ser brasileiro(a); estar pelo menos no segundo ano do doutorado; desenvolver a pesquisa dentro de uma das três áreas mencionadas; ter o português brasileiro como L1; autodeclarar um nível intermediário de leitura e escrita em inglês (L2)³⁴.

Durante a coleta, três tarefas foram solicitadas a cada participante: escrita de um resumo em português brasileiro; escrita de um resumo em inglês; e pós-edição do resumo em português brasileiro após a sua tradução automática para o inglês. A ordem das tarefas foi aleatorizada (Figura 5), de forma a evitar o elemento facilitador, a analisar o impacto da ordem e a garantir a validade ecológica do experimento. O tradutor automático utilizado foi o

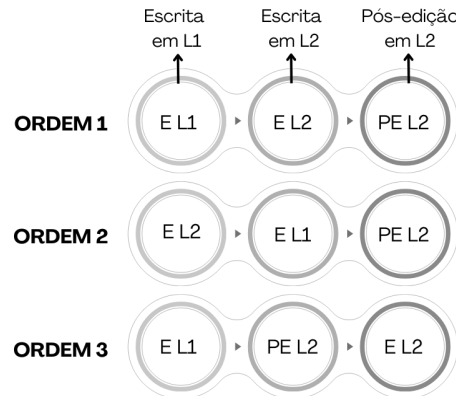
³² Nossa tradução para: “errors sometimes involve only single words, but more often involve phrases, discontinuous expressions, word order or relationships across sentence boundaries”

³³ Número do Certificado de Apresentação de Apreciação Ética (CAAE): 40708520.1.0000.5149

³⁴ Embora o ideal fosse a realização de um teste para verificação do nível de proficiência em inglês, esta se mostrou uma alternativa inviável, devido à dificuldade de recrutamento dos participantes.

DeepL, e a autopós-edição foi realizada diretamente no Translog-II³⁵. A fim de autopós-editar os textos, os participantes puderam recorrer a dicionários e ferramentas de busca, sendo vetado apenas o uso de tradutores automáticos.

Figura 5 - Ordens possíveis no experimento



Fonte: Elaboração própria, 2025.

As coletas aconteceram no Laboratório de Tradução (LETRA) da FALE-UFMG entre novembro de 2022 e junho de 2024. Durante elas, foi solicitada aos participantes a escrita de um resumo acadêmico, no âmbito de suas áreas, relatando uma pesquisa de autoria própria em andamento ou concluída. Como insumo, foram fornecidas chamadas fictícias para um congresso nacional e para um internacional, as quais estabeleciam como normas de submissão:

- a) a língua, sendo ela o português brasileiro ou o inglês, conforme a chamada;
- b) o tamanho do texto, entre 100 e 250 palavras;
- c) a ausência de elementos gráficos ou palavras-chave;
- d) a estruturação do resumo em objetivos, métodos, resultados e possíveis contribuições para a área.

O desenho experimental de Nunes (2021) prevê 10 participantes por domínio do conhecimento. Até o momento de conclusão desta dissertação, havia coletas de quatro doutorandos de CET, cinco doutorandos das CH e três das CSA, somando 12 participantes ao todo.

³⁵ Programa que possibilita a gravação do processo de escrita e leitura de textos, sendo, por isso, amplamente utilizado nos estudos processuais da tradução (Jakobsen, 2006).

3.1.2 O corpus: composição, compilação e armazenamento

O *corpus* desta pesquisa compreende trinta e seis resumos acadêmicos. Constitui, assim, um pequeno³⁶ corpus bilíngue paralelo e comparável, conforme mostram as Tabelas 1 e 2. Para esta pesquisa, ele foi dividido em três subcorpora, segundo o modo de produção (PT, TA ou PE); a grande área (CET, CH ou CSA); e a ordem das tarefas (1, 2 ou 3) — os quais foram analisados separadamente e, também, considerando as suas intersecções.

Tabela 1 - Divisão de resumos por área e modo de produção

Área	EL1	TA	PE L2	Total
CET	4	4	4	12
CH	5	5	5	15
CSA	3	3	3	9
Textos	12	12	12	36
Palavras	2445	2365	2433	7243

Fonte: Elaboração própria, 2025.

Tabela 2 - Divisão de resumos por ordem e modo de produção

Ordem	EL1	TA	PE L2	Total
1	4	4	4	12
2	4	4	4	12
3	4	4	4	12
Textos	12	12	12	36
Palavras	2445	2365	2433	7243

Fonte: Elaboração própria, 2025.

Os textos finais foram extraídos dos arquivos XML gerados pelo Translog-II, a partir do monitor físico do rastreador ocular Tobii T60. Todos os resumos foram, inicialmente, inseridos em uma planilha do Planilhas Google³⁷ para armazenamento e controle, conforme ilustra a Figura 6. As grandes áreas do CNPq foram separadas por cores, para facilitar a visualização. Os resumos se encontram elencados cronologicamente, seguindo as datas de realização das coletas.

³⁶ Classificação de extensão de corpora com menos de 80 mil palavras (Berber-Sardinha, 2004, p. 26)

³⁷ Gerenciador de planilhas semelhante ao Excel.

Figura 6 - Planilha contendo textos do corpus e metadados

ORDEM	CÓD.	ÁREA	EL1	TA	PE L2
ORDEM 1	BPC0551	CET	O estudo cinético da de	The kinetic study of th	The kinetic study of the
ORDEM 3	ESS0146	CET	Um dos grandes desafi	One of the great chall	Nowadays, one of the r
ORDEM 3	GBJ0654	CET	A análise de fontes de c	The analysis of conta	The analysis of contam
ORDEM 2	GBO0456	CET	Deficiências do microni	Deficiencies of the mi	Selenium deficiencies a
ORDEM 2	DRS1017	CH	Este trabalho aborda a	This paper discusses	This paper discusses th
ORDEM 1	EJT0905	CH	Esta pesquisa de douto	This doctoral research	This PHD research in e
ORDEM 1	FBI0418	CH	A proposta deste texto	The purpose of this te	The purpose of this tex
ORDEM 2	LSJ0446	CH	Grupos de interesse sã	Interest groups are a	Interest groups are a de
ORDEM 3	MBV1263	CH	A presente pesquisa ap	This research present	This research presents
ORDEM 1	CBA0267	CSA	O presente estudo bus	This study seeks to u	This study aims to shov
ORDEM 3	LBH0861	CSA	A crise sanitária da par	The health crisis of th	The health crisis of Cov
ORDEM 2	GBS0257	CSA	Devido as mudanças cl	Due to climate chang	Due to climate change

Fonte: captura de tela realizada pela pesquisadora

Em seguida, visando à extração de informações via script em python e ao processamento posterior por anotadores automáticos, cada um dos resumos foi salvo separadamente no formato de texto (.txt), com seus caracteres codificados em UTF-8. Os arquivos foram nomeados seguindo-se um código particular: a) ordem das tarefas (1, 2 ou 3); b) modo de produção do texto (EL1, TA ou PE L2; c) sigla da grande área do conhecimento (CET, CH ou CSA); e d) código alfanumérico para representar o participante. Assim, todos os nomes de arquivos eram semelhantes ao nome: “1 - EL1- CET - BPC0551”. A adoção deste código teve por fim facilitar a posterior análise dos erros por modo de produção, por ordem e por grande área do conhecimento.

Com os arquivos devidamente salvos e preparados para os passos subsequentes do estudo, passamos à anotação do corpus. Com efeito, em sua versão final, o corpus conta com duas camadas de anotação, uma automática e uma manual, discriminadas no Quadro 10, que serão detalhadas em mais detalhes nas duas seções subsequentes.

Quadro 10 - Camadas de anotação

Conteúdo da anotação	Modo de anotação	Ferramenta de anotação
Morfossintática	Automática	UDPipe
Erros	Manual	INCEpTION

Fonte: Elaboração própria, 2025.

3.1.3 Anotação morfossintática

Para a anotação morfossintática, optou-se pelo *framework* das dependências universais (UD). Essa escolha teve por respaldo o fato de esse ser um *framework* concebido para a anotação das diferentes línguas naturais, não de uma única língua específica³⁸. Assim, um

³⁸ Para mais informações, cf. <https://universaldependencies.org>.

corpus anotado segundo as UD possibilita a comparação linguística com base em categorias adequadas a estudos comparativos (Coneglian, 2023). Embora essa anotação não tenha sido utilizada na presente pesquisa, sua relevância se justifica pelo fato de que o corpus será posteriormente analisado por outros integrantes do projeto de Nunes (2021). Assim, toda a sua preparação tem por fim viabilizar, também, as atuais e futuras investigações conduzidas por esses pesquisadores.


Essa etapa de anotação foi realizada automaticamente utilizando-se o UDPipe, pipeline que tokeniza, lematiza e realiza a sua anotação com base nas UD. Para os fins desta pesquisa, optamos por utilizar a versão mais atualizada da ferramenta — à época, a 2. Com relação aos modelos utilizados para a anotação dos resumos, optamos por aqueles referentes à atualização 2.12. No caso do português brasileiro, o modelo foi treinado com a *trebank* Petrogold ('portuguese-petrogold-ud-2.12-230717') e, no caso do inglês, com a *trebank* Gum ('english-gum-ud-2.12-230717'). A escolha baseou-se tanto na avaliação das *trebanks* oferecidas no site oficial das UD, quanto no desempenho dos modelos reportado no próprio site do UDPipe.

Por fim, optamos por solicitar a normalização dos espaços e a pré-segmentação dos textos. Como entrada, fornecemos ao UDPipe os arquivos .txt, previamente salvos. Uma vez que a ferramenta realiza a conversão automática de txt para CONLL-U, a saída gerada são arquivos nesse formato. Nas figuras a seguir, é possível observar a interface disponível no site, as opções escolhidas para esta pesquisa e a saída resultante.

Figura 7 - Interface do UDPipe

The service is freely available for testing. Respect the [CC BY-NC-SA](#) licence of the models – explicit written permission of the authors is required for any commercial exploitation of the system. If you use the service, you agree that data obtained by us during such use can be used for further improvements of the systems at UFAL. All comments and reactions are welcome.

Model: UD 2.12 (description) UD 2.10 (description) UD 2.6 (description) PDT-C 1.0 (description) Evalatin (24/20)

 portuguese-petrogold-ud-2.12-230717

Actions: Tag and Lemmatize Parse

▼ Advanced Options

UDPipe version: UDPipe 2 UDPipe 1

Input [?]: Tokenize plain text CoNLL-U Horizontal Vertical

Tokenizer [?]: Normalize spaces Presegmented input Save token ranges

A Input Text Input File

PT - B.txt (1.4kb loaded) Load File ...

Fonte: captura de tela realizada pela pesquisadora

Figura 8 - Texto anotado gerado pelo UdPipe

```
# generator = UDPipe 2, https://lindat.mff.cuni.cz/services/udpipe
# udpipe_model = portuguese-petrogold-ud-2.12-230717
# udpipe_model_licence = CC BY-NC-SA
# newdoc
# newpar
# sent_id = 1
# text = O estudo cinético da decomposição térmica de fármacos e polímeros rotineiramente é feito através do uso de equações lineares que têm como objetivo tratar os dados experimentais obtidos através de analisadores térmicos. Porém, a prática tradicional desse método produz erros de grande ordem de grandeza, fazendo com que os resultados não sejam acurados. Tradicionalmente, os dados são tratados de forma empírica, através de tentativa e erro e, os modelos de decomposição térmica cujas equações descrevem os dados experimentais de maneira mais fidedigna são aceitos como corretos. No entanto, é possível realizar tais estudos através de novos modelos cinéticos, baseados nos tradicionais e uma das formas de conceber esses modelos é através de redes neurais artificiais. O objetivo deste trabalho é construir redes neurais artificiais em linguagem MATLAB que produzam modelos cinéticos de decomposição térmica acurados e com erro residuais de ordem de grandeza desprezível (menores que 10-4), para análise térmica e cinética da decomposição térmica de fármacos e polímeros. As redes arquitetadas até o momento produziram resultados com erro até um milhão de vezes menores do que os métodos tradicionais. O presente estudo possui contribuições diretas na indústria automotiva e farmacêutica, predizendo com grande precisão o tempo de decomposição de polímeros automotivos em condições extremas e o tempo de prateleira de fármacos, respectivamente.
1 O o DET_ Definite=Def|Gender=Masc|Number=Sing|PronType=Art 2 det _ TokenRange=0:1
2 estudo estudo NOUN _ Gender=Masc|Number=Sing 14 nsubj:pass _ TokenRange=2:8
3 cinético cinético ADJ Gender=Masc|Number=Sing 2 amod TokenRange=9:17
```

Fonte: captura de tela realizada pela pesquisadora

3.1.4 Anotação de erros

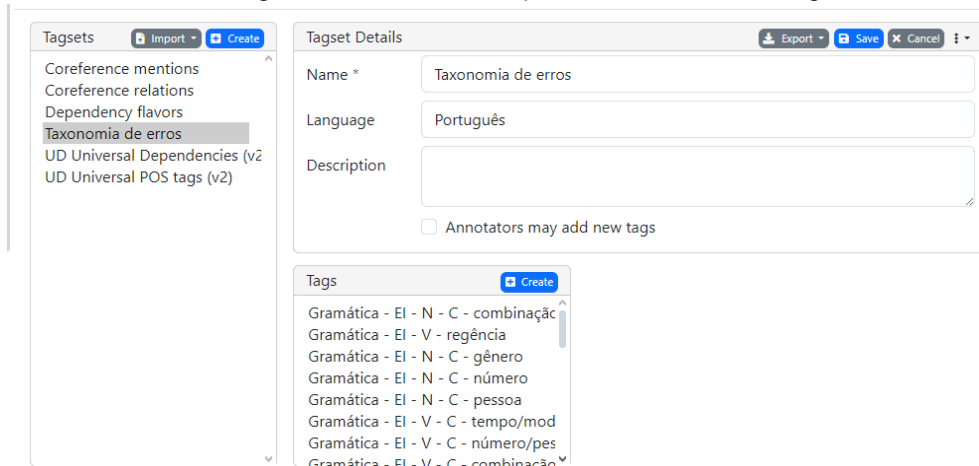
Assim que todos os 36 textos receberam a primeira camada de anotação, teve início a etapa manual, relativa aos erros. Ela foi realizada no INCEPTION, uma plataforma gratuita baseada em Java, disponibilizada pela Technische Universität Darmstadt (Alemanha), na qual é possível realizar anotações de outras naturezas que não apenas a morfosintática.

O primeiro passo foi a criação do projeto na plataforma e a importação dos arquivos CONLL-U. Em seguida, adicionou-se o conjunto de etiquetas (*tagset*), adotando para tal a taxonomia estabelecida por Costa *et al.* (2015), uma vez que, como mencionado na revisão da literatura, 1) é adaptada às línguas românicas, atendendo ao português brasileiro; 2) baseia-se em categorias mais amplas, facilitando a sua expansão e adaptação; 3) foi concebida para ser aplicada tanto a erros de sistemas de tradução automática quanto a erros humanos. Cada subcategoria foi concebida como uma etiqueta individual, de forma a facilitar a sua contabilização posterior. Finalmente, como última etapa antes do início da anotação, criou-se a camada referente aos erros, cujas configurações foram:

- 1) tipo: *span* (para delimitar a parte do texto a ser anotada)
- 2) granularidade: nível do *token* (estabelecida como a unidade mínima para a anotação do erro)
- 3) sobreposição: permitida (para casos em que um *token* estivesse envolvido em mais de um erro)
- 4) além do nível da sentença: permitido (para casos em que o erro envolvesse *tokens* em sentenças diferentes)

Além disso, foi adicionada uma única *feature* primitiva, do tipo *string* (por se tratar de uma sequência de caracteres), relativa aos erros, à qual foi associado o conjunto de etiquetas previamente criado.

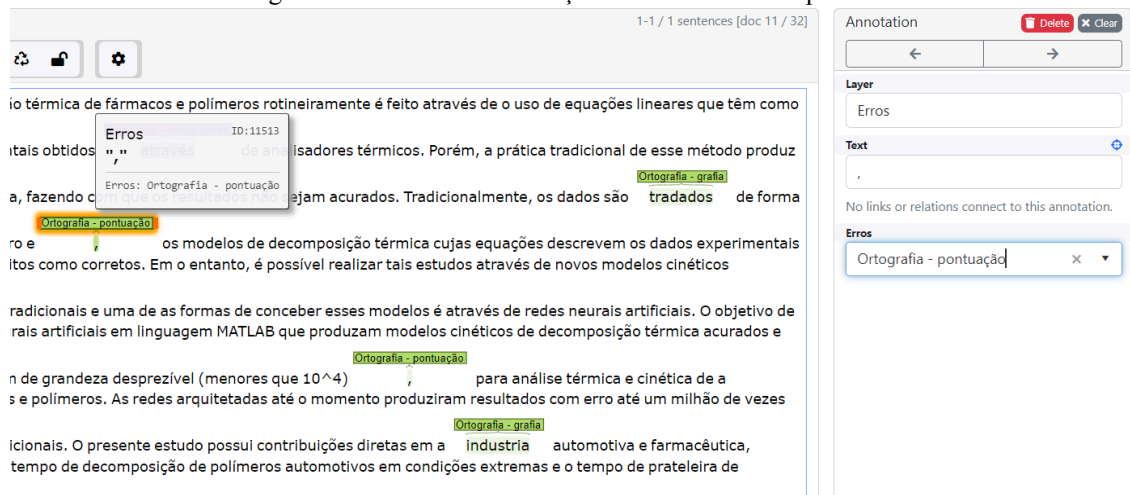
Figura 9 - Interface de criação da taxonomia no INCEPTION



Fonte: captura de tela realizada pela pesquisadora

A interface do INCEPTION é bastante amigável ao usuário. Para a anotação dos erros, foi necessário apenas selecionar o(s) *token(s)* referentes, clicar sobre a etiqueta criada e escolher, dentre os disponíveis, o tipo de erro correspondente.

Figura 10 - Processo de anotação realizado no INCEPTION



Fonte: captura de tela realizada pela pesquisadora

A anotação dos erros foi feita com base no inglês padrão, posto ser o padrão adotado pela maioria dos periódicos. No entanto, casos em que havia desvio da norma, mas os usos se encontravam validados em corpus (Google Scholar), especialmente em artigos de grande relevância, não foram apontados como erros. Ao todo, o processo de anotação contou com cinco etapas: uma anotação inicial realizada com auxílio do Chat GPT; duas revisões

realizadas pelo Chat GPT e validadas pela pesquisadora; uma terceira revisão validada apenas pela pesquisadora; e uma quarta revisão realizada por pares. Isso se deu da seguinte forma:

- 1) A anotação inicial foi realizada pelo Chat GPT-3.5³⁹, entre maio e junho de 2024. Para isso, forneceu-se ao modelo o resumo a ser anotado junto à seguinte orientação: “identificar erros de gramática, ortografia, semântica. Quero uma lista dos erros em inglês, não o texto corrigido”. Em resposta, o Chat forneceu uma lista dos erros, junto a explicações de por que constituíam incorreções. Com base nisso, os erros foram anotados no INCEpTION, sem passar por revisão humana.
- 2) A primeira revisão foi realizada com o auxílio do Chat GPT-4o mini, em julho. Para tal, as orientações foram adaptadas, de forma a que o modelo considerasse, também, que se tratava de uma escrita acadêmica. Assim, solicitou-se: “identificar erros de gramática, ortografia, semântica. Quero uma lista dos erros em inglês, não o texto corrigido. Considere que se trata de um texto em [língua em análise] acadêmico”. Com base na lista e nas explicações fornecidas pelo modelo, os textos anotados foram revisados. Dessa vez, a pesquisadora interveio na anotação, porém seguindo apenas a própria intuição de falante experto do português brasileiro e os conhecimentos adquiridos no estudo formal do português e do inglês, sem validar as decisões em corpora, em gramáticas ou em dicionários.
- 3) A segunda revisão foi realizada da mesma forma que a primeira, a diferença ocorreu apenas na validação humana. Em vez de recorrer somente à própria intuição, a pesquisadora recorreu, também, a corpus. Nesse caso, optou-se pelo Google Scholar, por ele apresentar atualização constante; possibilitar a organização dos resultados por relevância do texto; facilitar o processo de checagem das ocorrências; e informar sobre o número de vezes que o texto no qual a ocorrência está presente foi citado. Recorreu-se, também, a gramáticas — *Longman Grammar of Spoken and Written English* (Biber et al., 2003) e *A Gramática de usos do Português Brasileiro* (Neves, 2000) — e a dicionários gerais — *Oxford Language* (disponível online) e *Minidicionário Houaiss* (Houaiss, 2019) — e de regência (Luft, 2010; Luft, 2013).

³⁹ Para mais informações, ver: <https://openai.com/index/chatgpt/>

- 4) A terceira revisão foi realizada apenas pela pesquisadora, sem recurso ao chatbot, revalidando as escolhas com base no Google Scholar, nos manuais e nos dicionários anteriormente mencionados.
- 5) A quarta revisão, por fim, foi realizada por dois pares: um doutor em linguística e professor de Inglês para Fins Acadêmico-Profissionais; e uma mestranda em Linguística e professora em inglês para fins gerais. Para essa revisão, os arquivos anotados foram exportados em formato TSV (.tsv) e salvos em uma pasta no Google Drive, à qual ambos os revisores tiveram acesso, sem que um, no entanto, tivesse acesso à avaliação do outro.

A opção pelo auxílio do chatbot nas três primeiras etapas se deu por dois motivos. O primeiro foi de ordem prática. O recurso ao Chat GPT (tanto ao 3.5 quanto ao 4o mini) tornou o processo muito mais ágil. Além disso, uma vez que a atenção humana está sujeita a falhas, a assistência do modelo foi uma garantia de que não era apenas o julgamento humano envolvido na anotação e revisão dos erros, ainda que a validação final tenha sido realizada por humanos.

O segundo foi de interesse didático. Com o advento dos grandes modelos de linguagem, o Chat GPT, em suas diversas versões, passou a fazer parte do cotidiano de muitas pessoas (Altman, 2022; Hu, 2023). Ademais, ele se tornou mais uma ferramenta de auxílio à escrita acadêmica, com potencial para superar os tradutores automáticos, na medida em que reúne em um mesmo lugar, não só a tarefa de tradução, mas de revisão, ambas com um alto grau de personalização possível graças ao fornecimentos de instruções (*prompts*). Entretanto, por mais otimistas que pareçam os prognósticos, é necessário proceder com cautela — não apenas por questões éticas, mas por questões ligadas à limitação da própria capacidade do modelo.

Assim, ao unir as sugestões do modelo à validação humana, o objetivo era investigar, de forma preliminar, o potencial dessa ferramenta no auxílio à edição. Embora isso não tenha sido feito de forma sistemática, até por não ser o foco da dissertação, foi possível, através dessa opção metodológica, observar forças e fraquezas mais evidentes do modelo. Assim, enquanto seu ponto forte parece ser a identificação de erros ortográficos e gramaticais, sua principal fraqueza são as questões semânticas — não só deixa de corrigi-las como, por vezes, peca por hipercorreção.

No que tange à revisão das anotações por dois pares, seu propósito foi mitigar o prejuízo das dificuldades relacionadas à localização e à categorização dos erros, das quais trata a seção 2.3.4. A revisão permitiu a discussão dos *tokens* implicados nos erros e da

clareza nas definições das categorias e em sua aplicação. Com isso, assegurou-se uma anotação mais fiável.

Finalmente, antes de apresentar as adaptações realizadas na taxonomia, são necessárias duas observações a respeito da exportação dos arquivos anotados. Primeiro, embora a plataforma ofereça a possibilidade de exportar todos os arquivos conjuntamente em formato CONLL-U e, adicionalmente, em algum outro formato de escolha do usuário⁴⁰, optou-se por baixar cada arquivo individualmente, ao final da última revisão. Segundo, o formato escolhido foi o WebAnno TSV v3.3 (.tsv), que é tabular, permitindo a abertura em planilha quando necessário.

3.1.4.1 Adaptação da taxonomia

Durante as três primeiras etapas da anotação, ficou evidente que o conjunto de etiquetas, conforme originalmente proposto por Costa *et al.* (2015), precisava ser expandido. Como mencionado na Revisão da Literatura, esses autores conceberam sua taxonomia com a intenção de que pudesse ser adaptada em pesquisas futuras, fundamentando-na em categorias mais amplas, dentro das quais é possível incluir categorias mais específicas sem maiores dificuldades.

Para este trabalho, dentro de Gramática, os erros de Concordância foram diferenciados daqueles de Regência. Já em Semântica, foram adicionadas as subcategorias Ambiguidade (sem distinção da natureza) e Coerência/Coesão (problemas de paralelismo; dificuldade/impossibilidade de retomar antecedentes). Por fim, dentro de Discurso, Estilística foi subdividida em problemas de registro (inadequações do texto ao registro acadêmico) e problemas de repetição (repetição de palavras; redundâncias).

Conforme apontado durante a revisão da literatura, Costa *et al.* (2015) não fornecem mais detalhes a respeito da aplicação da categoria Estilística, além de mencionar que atende a problemas como repetições. Para fins de adaptação dessa etiqueta, recorreremos, assim, a bibliografia adicional — no caso, ao texto de O'Brien, Simard e Goulet (2018), que menciona a presença e uso da categoria no corretor gramatical Antidote, disponível online em duas versões, uma gratuita e outra paga⁴¹. O conjunto de etiquetas final se encontra discriminado no Quadro 11.

⁴⁰ Os arquivos podem ser exportados em diversos formatos: CONLL (2000 a 2012); CONLL CoreNLP; CONLL-U (formato padrão); Inline XML; NLP Interchange Format; Plain text; TEI; UIMA CAS JSON 0.4.0; UIMA CAS XMI (tanto XML 1.0 quanto XML 1.1); UIMA binary CAS; WebAnno TSV v3.3 e WebLicht TCF.

⁴¹ Para mais informações, cf. <https://www.antidote.info/en/>.

Houve outros casos em que foi suficiente expandir as categorias em seu escopo, sem necessariamente acrescentar novas etiquetas. Assim, no nível do Léxico, as categorias de adição e omissão de palavras (lexicais ou gramaticais) foram usadas, também, em contextos em que a palavra (não) existia no texto em português, mas era (des)necessária em inglês.

Quadro 11 - Etiquetas utilizadas no INCEpTION

Etiquetas	
Ortografia - grafia	Semântica - coerência/coesão
Ortografia - pontuação	Semântica - colocação incorreta
Ortografia - letra maiúscula	Semântica - ambiguidade
Léxico - omissão PG	Semântica - palavra incorreta
Léxico - adição PG	Semântica - sentido incorreto
Léxico - omissão PL	Semântica - expressão idiomática
Léxico - adição PL	Discurso - estilística - repetição
Léxico - palavra não traduzida	Discurso - estilística - registro
Gramática - EI - V - C - número/pessoa	Discurso - não deveria ter sido traduzido
Gramática - EI - V - C - combinação	Discurso - dialeto
Gramática - EI - V - C - tempo/modo	
Gramática - EI - V - regência	
Gramática - EI - N - C - gênero	
Gramática - EI - N - C - pessoa	
Gramática - EI - N - C - número	
Gramática - EI - N - C - combinação	
Gramática - EI - classe de palavra	
Gramática - EI - N - regência	
Gramática - EI - Contração	
Gramática - OI	

Fonte: Elaboração própria, 2025.

No nível Semântico, por sua vez, erros de terminologia foram classificados no interior da categoria “palavra incorreta”, quando envolviam um único *token*, ou “colocação incorreta”, quanto implicavam mais de um *token*. Quanto à categoria das colocações, ela foi aplicada a qualquer sequência de palavras que ocorre com uma frequência maior do que esperado, sendo ou não contíguas (Frankenberg-Garcia, 2018). Foram considerados erros de colocação casos em que o sistema de tradução automática ou o participante produziu uma sequência de palavras cuja co-ocorrência é muito baixa (menos de 100 ocorrências no Google Scholar). Já os casos de sequências com mais de 100 ocorrências, mas menos de 1000, foram classificados como problemas de registro e, portanto, localizados no nível discursivo. Por meio dessas

estratégias de adaptação, foi possível a continuidade do trabalho de Costa *et al.* (2015), de fato expandindo a taxonomia proposta. A taxonomia adaptada e com exemplos do corpus pode ser conferida no Quadro 12. Etiquetas sem ocorrências não constam no quadro.

Quadro 12 - Taxonomia de Costa *et al.* (2015) com adaptações

Categoria Subcategorias		Exemplos retirados do corpus			Detalhes	
		Grupo	Incorreto	Correto		
Ort.	Pontuação	EL1	"e, os modelos de decomposição térmica"	"e os modelos de decomposição térmica"		
	Letra maiúscula	EL1	"Os Cenários de simulação"	"Os cenários de simulação"		
	Grafia	EL1	"os dados são tradados"	"os dados são tratados"		
Léx.	Omissão	Palavra lexical	TA	"some cancers"	"some types of cancer"	No original em pt-br: "alguns tipos de câncer"
		Palavra gramatical	PE L2	"complete degradation of eight target CECs and 3 log inactivation of bacteria"	"complete degradation of eight target CECs and a 3 log inactivation of bacteria"	
	Adição	Palavra lexical	PE L2	"the contaminants contaminantes of emerging concern (CEC)"	"the contaminants of emerging concern (CEC)"	
		Palavra gramatical	TA	"The objective of this work is to biofortify flaxseed with selenium in a hydroponic culture by measuring (...) the concentration""	"The objective of this work is to biofortify flaxseed with selenium in a hydroponic culture, measuring (...) the concentration""	No original em pt-br: "O objetivo deste trabalho é biofortificar a linhaça com selênio em um cultivo hidropônico, medindo (...) a concentração"
Escolha incorreta	Classe de palavras	TA/ PE L2	"As specific objectives, the thesis presented a description of the educational policies"	"The specific objectives of the dissertation were to describe the educational policies(...)"	No original em pt-br: "Como objetivos específicos a tese apresentou descrever as políticas educacionais"	

Fonte: Elaboração própria, 2025.

Quadro 12 - Taxonomia de Costa *et al.* (2015) com adaptações

Categoria Subcategorias				Exemplos retirados do corpus			Detalhes		
				Grupo	Incorreto	Correto			
Gram.	Escolha incorreta	Verbos	Concordância	Tempo/ Modo	PE L2	"to improve your (sic) life and developed skills and habilities(sic)"	"to improve your (sic) life and to develop skills and habilities(sic)"		
				Número/ Pessoa	PE L2	"a piece for day that relate"	"a piece for day that relates"		
			Ambos	PE L2	"samples of hospital wastewater were sampling"	"samples of hospital wastewater were sampled"			
		Regência		PE L2	"The analysis of contamination sources(...) has currently been studied for several disasters that have happened."	"The analysis of contamination sources(...) has currently been studied due to several disasters that have happened."	No original em pt-br: "A análise de fontes de contaminação (...) atualmente tem sido muito estudada devido a vários desastres que aconteceram"		
		Nome	Concordância	Gênero	EL1	"torna necessário uma regulação democrática do sistema "	"torna necessária uma regulação democrática do sistema "		
				Número	EL1	"com erro residuais"	"com erros residuais"		
			Regência	EL1	"Devido as mudanças climáticas"	"Devido às mudanças climáticas"			
		Sem..	Sentido incorreto			TA	"much researched object"	"much-researched topic"	
			Palavra incorreta			PE L2	"A philosophical experimentation, sometimes"	"Therefore, a philosophical experimentation"	
			Colocação incorreta			TA/ PE L2	"they can be read in the equipment"	"they can be read by the equipment"	

Fonte: Elaboração própria, 2025.

Quadro 12 - Taxonomia de Costa *et al.* (2015) com exemplos do corpus

Categoria Subcategorias		Exemplos retirados do corpus			Detalhes	
		Grupo	Incorreto	Correto		
Sem.	Sentido incorreto	TA	"much researched object"	"much-researched topic"		
	Palavra incorreta	PE L2	"A philosophical experimentation, sometimes"	"Therefore, a philosophical experimentation"		
	Colocação incorreta	TA/ PE L2	"they can be read in the equipment"	"they can be read by the equipment"		
	Coerência/Coesão	EL1	"os equipamentos [...] tem (sic) crescido e ficaram mais sensíveis."	"os equipamentos tornaram-se mais numerosos e sensíveis"		
	Ambiguidade	EL1	"tratar os dados experimentais obtidos através de analisadores térmicos"	"tratar, através de analisadores térmicos, os dados experimentais obtidos"		
Disc.	Estilística	Registro	TA/ PE L2	"making the results inaccurate"	"leading to inaccurate results"	
		Repetição	PE L2	"kinetic models with residual errors with acetable[sic] order"	"kinetic models with negligible order of magnitude [...] residual errors"	
	Não deveria traduzir		TA	"through the project Leisure Street"	"through the project Rua do Lazer"	

Fonte: Elaboração própria, 2025.

3.2 Metodologia de análise do corpus

3.2.1 Análise quantitativa

Para obter um panorama do corpus com base nos arquivos salvos em .txt, foi solicitado ao Chat GPT-4o mini⁴² que gerasse dois códigos: 1) um para contabilizar os *types* de cada arquivo do corpus e 2) um voltado à contagem de *tokens* dos arquivos. Como ambiente computacional para execução dos códigos, escolheu-se o Google Colab, plataforma baseada em nuvem de propriedade da Google. Ela foi escolhida devido à sua facilidade de acesso e uso, pois, além de ser gratuita, pode ser acessada diretamente no navegador. Para o processamento, o Google Colab foi conectado ao Google Drive para acesso ao corpus e, na sequência, os códigos foram executados.

Por meio dessa etapa, tornou-se possível obter informações gerais a respeito dos textos, caras a pesquisas direcionadas por corpus (*corpus driven*) (Heine; Narrog; Biber,

⁴² Ainda que a ideia inicial fosse realizar a contagem das etiquetas com o auxílio do Planilhas Google, isso se provou uma opção pouco produtiva quando comparada à análise automatizada por um código em Python.

2015). Embora um panorama mais completo exigisse a extração de outros dados, como o número médio de palavras por sentença e a lista de termos mais frequentes, optamos por extrair apenas os *types* e os *tokens*. Isso porque, além de nos esclarecer a extensão dos corpus e a razão *type/token* (TTR), esses dados nos permitiriam calcular a proporção de erros por modo de produção do texto, por área e por ordem, afora as proporções das combinações entre esses três.

A segunda parte da análise quantitativa envolveu a extração da quantidade de erros. Para tanto, solicitou-se ao Chat GPT-4o mini que gerasse um código que realizasse a contagem das etiquetas presentes nos arquivos .tsv com base no conjunto de etiquetas fornecido, o qual foi exportado do próprio INCEpTION em formato JSON.

O código ficou responsável por, então:

1. acessar a pasta do Drive na qual os arquivos .tsv estavam;
2. ler o conjunto de etiquetas em formato JSON;
3. remover números e colchetes, inseridos pelo INCEpTION em caso de ocorrências sucessivas de um mesmo tipo de etiqueta dentro de um único arquivo (por exemplo: Ortografia - Pontuação [1], na primeira ocorrência; Ortografia - Pontuação [2]; na segunda ocorrência);
4. realizar a contagem das etiquetas;
5. imprimir o resultado.

Figura 11 - Trecho do código gerado pelo Chat GPT-4o mini e executado no Google Colab

```
import pandas as pd
import re
import os
import json
from collections import defaultdict

# Diretório onde os arquivos TSV estão localizados
folder_path = 'caminho do diretório de arquivos'

# JSON com as tags
tags_json = ""
{
  "name": "Taxonomia de erros",
  "description": null,
  "language": "Português",
  "tags": [
    {"tag name": "Gramática - EI - N - C - combinação".
```

Fonte: captura de tela realizada pela pesquisadora

Adicionalmente, solicitamos três versões alteradas do código: uma que fornecesse a listagem da quantidade de erros por arquivo; e duas outras que pesquisassem as etiquetas

apenas de arquivos que contivessem uma *string* específica (ou mais de uma) em seu nome . Assim, foi possível realizar a análise quantitativa por modo de produção, por área e por ordem, bem como considerando as intersecções entre os três subcorpora.

Para facilitar a visualização dos dados, foi gerado um código para a criação de tabelas com base nos dados quantitativos resultantes da execução dos códigos anteriores. Esse código exigia a inserção não do diretório para a localização dos arquivos, mas dos próprios dados quantitativos resultantes. Eles eram então processados para a geração de tabelas, que eram exportadas diretamente no Google Drive.

Essas tabelas contêm:

- a) o número total de erros;
- b) o número total de erros por etiquetas
- c) o número total de erros por categorias
- d) a discriminação das etiquetas com 0 ocorrências.

Isso foi solicitado para a totalidade do corpus; para cada modo de produção, cada área do conhecimento e cada ordem das tarefas; além de para as combinações deles (área/modo de produção, área/ordem; modo de produção/ordem; área/ordem/modo de produção). Com base nas tabelas geradas, os resultados foram separados em planilhas discretas, uma para cada contagem, as quais foram reunidas em uma única pasta de trabalho. Nesta pasta, constam também os resultados das contagens iniciais realizadas no corpus, conforme ilustra a Figura 12. Por fim, todos os dados foram verificados duas vezes e, após constatar sua justeza, foram normalizados.

Figura 12 - Organização da pasta de trabalho no Google Sheets

	A	B	C	D	E	F	G	H	I	J	K
	Arquivo	Tags	Tokens	E/T	Total	Tokens	E/T		Tag	Contagem	Total
1	1-CET-PT-BPC0551.tsv	34	214	15,89%	1539	7243	21,25%		Ortografia - grafia	81	1539
2	1-CET-TA-BPC0551.tsv	27	199	13,57%					Ortografia - pontuação		83
3	1-CET-PE-BPC0551.tsv	30	198	15,15%					Ortografia - letra maiúscula		3
4	2-CET-PT-GBO0456.tsv	12	235	5,11%					Léxico - omissão PG		41
5	2-CET-TA-GBO0456.tsv	19	220	8,64%					Léxico - adição PG		9
6	2-CET-PE-GBO0456.tsv	18	220	8,18%					Léxico - omissão PL		9
7	3-CET-PT-ESS0146.tsv	13	223	5,83%					Léxico - adição PL		1
8	3-CET-TA-ESS0146.tsv	27	211	12,80%					Gramática - EI - V - C - número/pessoa		16
9	3-CET-PE-ESS0146.tsv	54	197	27,41%					Gramática - EI - V - C - combinação		15
10	3-CET-PT-GBJ0654.tsv	35	189	18,52%					Gramática - EI - N - C - número		14
11	3-CET-TA-GBJ0654.tsv	32	180	17,78%	Total CET	Tokens	E/T		Gramática - EI - V - C - tempo/modo		10
12	3-CET-PE-GBJ0654.tsv	52	174	29,89%	353	2460	14,35%		Gramática - EI - V - regência		11

Fonte: captura de tela realizada pela pesquisadora

Para consulta do corpus de modo a checar as ocorrências quando necessário, recorreremos à planilha de gerenciamento do corpus e aos arquivos reunidos dentro do INCEPTION (recorrendo-se à sua interface amigável ao usuário). Uma vez que todos os passos metodológicos foram descritos, segue-se para a próxima seção, na qual serão apresentados e discutidos os resultados obtidos.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Para uma melhor compreensão da análise e da interpretação dos resultados, o Quadro 13 apresenta uma síntese da condução da pesquisa: perguntas, objetivos e metodologia. A pergunta e o objetivo gerais se encontram em negrito, enquanto os específicos são apresentados sem destaque. As perguntas serão respondidas ao longo da discussão empreendida na seção, a qual se encontra estruturada em torno dos passos metodológicos da análise quantitativa.

Quadro 13 - Estrutura da pesquisa

Perguntas	Objetivos	Metodologia
Qual é a influência da autopós-edição no número e nos tipos de erros de tradução em textos de diferentes áreas do conhecimento?	Investigar, no par português brasileiro-inglês, a influência da autopós-edição na quantidade e nos tipos de erros de tradução encontrados em resumos acadêmicos de distintos domínios do conhecimento	Compilação e preparação do corpus 1) compilação e preparação do corpus 2) anotação dos erros
Em que medida o número e os tipos de erros encontrados nos textos em português brasileiro influenciaram nos erros encontrados nos textos traduzidos automaticamente para o inglês?	Analisar a relação entre os erros observados nos resumos acadêmicos em português e os erros observados nos resumos traduzidos automaticamente para o inglês.	Análise quantitativa 1) contabilização dos <i>types</i> e <i>tokens</i> 2) contabilização dos erros totais
Em que medida o número e os tipos de erros encontrados nos textos traduzidos automaticamente para o inglês influenciaram nos erros encontrados nos textos autopós-editados em inglês?	Analisar a relação entre os erros encontrados nos resumos acadêmicos traduzidos automaticamente pelo tradutor e os erros encontrados nos resumos autopós-editados em inglês.	3) contabilização dos erros por: - modo de produção - área - ordem - modo de produção e área - modo de produção e ordem 4) normalização dos dados 5) interpretação dos dados

Fonte: Elaboração própria, 2025.

Quadro 13 - Estrutura da pesquisa

Perguntas	Objetivos	Metodologia
Em que medida o número e os tipos de erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente para o inglês e nos resumos autopós-editados em inglês variam entre as diferentes áreas do conhecimento?	Analisar a relação entre os erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente pelo sistema para o inglês e nos resumos autopós-editados em inglês de cada uma das áreas do conhecimento.	
Em que medida o número e os tipos de erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente para o inglês e nos resumos autopós-editados em inglês variam conforme a ordem de execução das tarefas?	Analisar a relação entre os erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente pelo sistema para o inglês e nos resumos autopós-editados em inglês para cada uma das ordens de execução das tarefas.	

Fonte: Elaboração própria, 2025.

4.1 *Types, Tokens e TTR*

O primeiro passo da análise quantitativa foi a contabilização dos *types* e *tokens*. Em ambos os casos, eles foram contabilizados por arquivos e, depois, somados de acordo com os subcorpora. Os resultados podem ser conferidos na Tabela 3. Ao todo, o corpus conta com 4382 *types* e 7243 *tokens*. Portanto, uma TTR de 0.60. Todos os subcorpora não apresentaram grandes variações no que tange ao TTR, o que sugere densidades lexicais bastante semelhantes. Com relação ao subcorpus relativo aos modos de produção do texto, é interessante observar que, na passagem do português para a autopós-edição, não houve perda significativa da densidade lexical, que variou em apenas 0.03.

Tabela 3 - Contabilização de *types* e *tokens*

SUBCORPUS	TOKENS	TYPES	TTR
EL1	2.445	1.520	0,62
TA	2.365	1.417	0,60
PEL2	2.433	1.445	0,59
Ordem 1	2.197	1.277	0,58
Ordem 2	2.600	1.588	0,61
Ordem 3	2.446	1.517	0,62
CET	2.460	1.465	0,60
CH	3.255	1.938	0,60
CSA	1.528	979	0,64

Fonte: Elaboração própria, 2025.

Na passagem do português brasileiro ao inglês, o tradutor automático reduziu o número de *tokens*. Isso não surpreende, uma vez que o inglês tende a ser uma língua mais concisa do que o português (Vieira, 1982). No entanto, observa-se que, durante a autopós-edição, o número de *tokens* voltou a aumentar, o que pode ser um reflexo da L1. Isso, no entanto, não resultou em um aumento da densidade lexical — ou seja, os participantes adicionaram palavras, mas não palavras diferentes.

No que se refere à quantidade de *tokens*, a média do corpus é de 201 *tokens* por participante, com um desvio padrão de aproximadamente 41 e uma amplitude de 155. Quanto aos *types*, a média é de 122 por texto, com um desvio padrão de 23 e uma amplitude de 99. Já para os subcorpora, as medidas são as mostradas na Tabela 4.

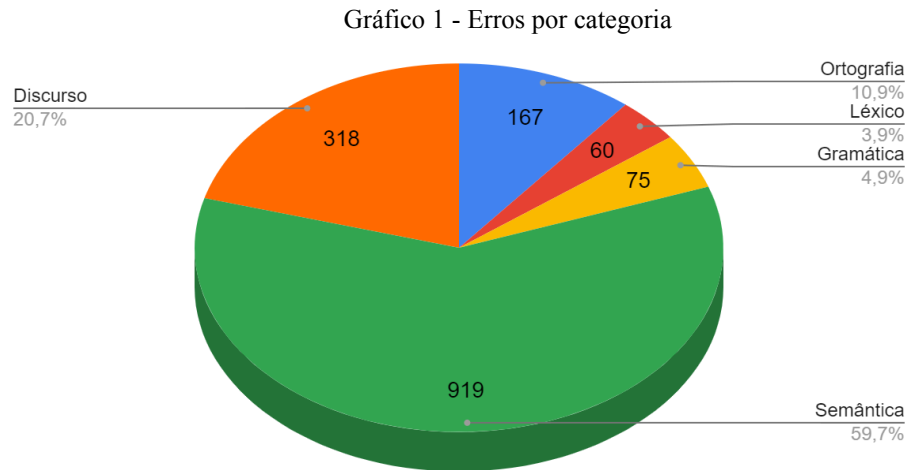
Tabela 4 - Média, amplitude e desvio padrão de *types* e *tokens*

SUBCORPUS		Tokens			Types		
		Média	Amp	DP	Média	Amp	DP
Modo de produção	EL1	203,75	136,00	40,70	126,67	93,00	23,96
	TA	197,08	151,00	41,19	118,08	88,00	22,96
	PEL2	202,75	149,00	44,46	120,42	90,00	23,74
Área	CET	205,00	61,00	18,54	122,08	30,00	10,24
	CH	217,00	81,00	29,23	129,20	46,00	14,34
	CSA	169,78	154,00	61,63	108,78	99,00	39,38
Ordem	ORDEM 1	183,08	144,00	49,14	106,42	73,00	26,55
	ORDEM 2	216,67	105,00	37,92	132,33	50,00	15,92
	ORDEM 3	203,83	80,00	29,63	126,42	53,00	18,63

Fonte: Elaboração própria, 2025.

4.2 Contagem dos erros totais

Em seguida, o segundo passo na análise quantitativa foi a contabilização dos erros totais. Isso foi feito em três níveis de granularidade: sem distinção de categorias, com distinção das categorias e com contabilização das etiquetas individualmente. Ao todo, o corpus apresentou 1539 erros, com uma porcentagem de 21,25% erro/*token* (E/T). Eles se encontram distribuídos nas categorias gerais conforme mostra o Gráfico 1. Os erros mais frequentes foram os de Semântica, representando 59,7% das 1539 ocorrências, enquanto os menos frequentes foram os Lexicais, responsáveis pela fatia de 3,9%.



Fonte: Elaboração própria, 2025.

A Tabela 5 discrimina os erros por etiqueta. Com isso, é possível descobrir quais foram decisivas para os resultados exibidos no gráfico precedente. Começando pelos Lexicais, que foram os menos frequentes, "omissão de palavras gramaticais" foi a ocorrência mais identificada. É importante lembrar que as etiquetas de omissões e adições foram utilizadas para anotar, também, casos em que uma palavra deveria (ou, ao contrário, não deveria) estar presente em um contexto. Essa adaptação foi feita para dar conta dos erros humanos. Assim, grande parte das ocorrências podem ser atribuídas aos participantes, não ao tradutor automático, que responde por apenas seis delas.

Com relação aos erros gramaticais, que foram os segundo menos frequentes, as etiquetas com mais ocorrências diziam respeito à concordância e à regência verbais. No caso da concordância nominal, apenas a de número se repetiu mais vezes. Nenhuma etiqueta particular se distanciou extremamente das demais, no entanto. No caso da Ortografia, os problemas de grafia e os de pontuação apareceram em medidas semelhantes. Dadas as limitações do cenário da produção textual, realizada em um contexto empírico-experimental, é difícil precisar em que medida esses são erros que, de fato, seriam cometidos pelos doutorandos, ainda mais se considerarmos que, em geral, as produções textuais são realizadas com mais tempo para revisão e em editores de textos, que possuem corretores ortográficos embutidos bastante eficientes.

No tocante aos erros discursivos, representativos de 20,7% das ocorrências, as questões ligadas à estilística foram as mais relevantes, respondendo por 316 das marcações. A repetição ficou responsável por mais de 50%, com 182 registros. Desse total, menos de 10% referem-se a digitações duplicadas. É importante lembrar que a etiqueta foi aplicada nos casos de repetição de palavras e de ideias (redundância). Assim, a recorrência da etiqueta é

indicativa, por um lado, de uma baixa variabilidade vocabular e, por outro, de uma proximidade na expressão das ideias. Quanto aos problemas ligados ao registro acadêmico, 122 se encontram nos textos em inglês, tanto os traduzidos pelo tradutor automático quanto aqueles autopós-editados. Isso se explica pela falta de adaptação do *DeepL* ao registro acadêmico, bem como, possivelmente, pela menor familiaridade dos participantes com esse registro em inglês.

Tabela 5 - Erros por etiqueta

TIPO	TOTAL	ETIQUETA	TOTAL (A)	TOTAL (R)
Ortografia	167	Ortografia - grafia	81	48,50%
		Ortografia - pontuação	83	49,70%
		Ortografia - letra maiúscula	3	1,80%
Léxico	60	Léxico - omissão PG	41	68,33%
		Léxico - adição PG	9	15,00%
		Léxico - omissão PL	9	15,00%
		Léxico - adição PL	1	1,67%
Gramática	75	Gramática - EI - V - C - número/pessoa	16	21,33%
		Gramática - EI - V - C - combinação	15	20,00%
		Gramática - EI - N - C - número	14	18,67%
		Gramática - EI - V - C - tempo/modo	10	13,33%
		Gramática - EI - V - regência	11	14,67%
		Gramática - EI - N - C - gênero	2	2,67%
		Gramática - EI - classe de palavra	2	2,67%
		Gramática - EI - N - regência	5	6,67%
Semântica	919	Semântica - coerência/coesão	603	65,61%
		Semântica - colocação incorreta	180	19,59%
		Semântica - ambiguidade	102	11,10%
		Semântica - palavra incorreta	18	1,96%
		Semântica - sentido incorreto	16	1,74%
Discurso	318	Discurso - estilística - repetição	182	57,23%
		Discurso - estilística - registro	134	42,14%
		Discurso - não deveria ter sido traduzido	2	0,63%

Fonte: Elaboração própria, 2025.

Finalmente, a respeito dos problemas de Semântica, coerência/coesão responde por quase 66% das ocorrências. Assim, percebe-se que os textos apresentam, sobretudo,

problemas ligados à construção e à conexão dos significados do texto. Essa distribuição é uniforme independente do modo de produção, o que sugere que já eram um problema em português, o *DeepL* não os resolveu e tampouco a autopós-edição.

Apenas seis etiquetas não tiveram qualquer ocorrência. No nível do Léxico, não houve palavras não traduzidas. No nível da Gramática, não foram registradas algumas escolhas incorretas (concordância nominal de pessoa, concordância nominal envolvendo mais de um problema e contração) ou qualquer ordenação incorreta. No nível da Semântica, não ocorreram expressões idiomáticas. Finalmente, no nível do Discurso, não foram identificados erros dialetais.

Sobre a questão das contrações, sua não ocorrência no português pode ser explicada por essa ser a L1 dos participantes e, portanto, aquela na qual eles estão mais habituados a escreverem na universidade. Em contrapartida, no caso do inglês, o fato é justificável se considerarmos o nível de formalidade dos resumos, uma vez que as contrações são associadas, tipicamente, à linguagem coloquial.

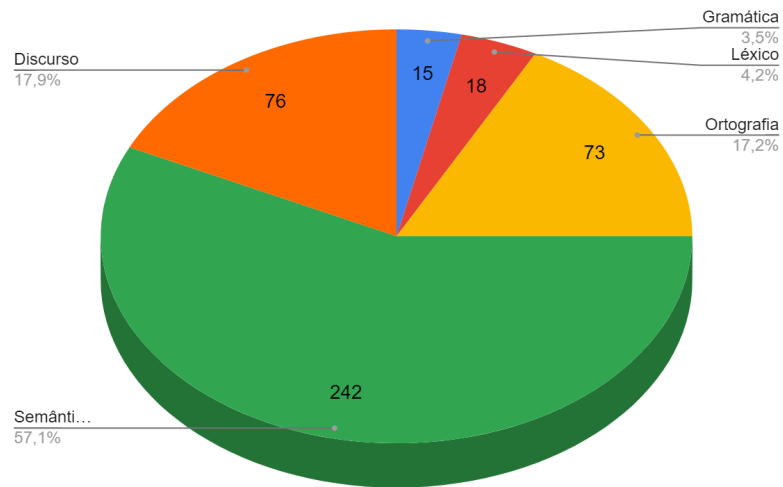
Com relação a ausência de expressões idiomáticas, isso se explica pela própria natureza dos textos, uma vez que a utilização desse recurso linguístico é incomum nos registros acadêmicos. Já a ausência de problemas ligados ao dialeto justifica-se pelo fato de que os textos em português foram redigidos em um contexto brasileiro. Caso o português fosse a língua alvo, não a língua fonte, é possível que houvesse ocorrências. No entanto, não era esse o cenário.

Por fim, é interessante observar que, embora os problemas de ordenação incorreta fossem típicos dos tradutores mais antigos (Bowker, Ciro, 2019), isso não parece persistir, uma vez que não há nenhuma ocorrência desse erro em qualquer um dos textos, mesmo naqueles que foram automaticamente traduzidos.

4.3 Contagem dos erros por modo de produção

O passo seguinte na análise foi a contagem por modo de produção, ou seja, EL1, TA ou PE L2. No caso da EL1, ao todo houve 424 erros, que respondem por uma taxa de cerca de 17 erros a cada 100 *tokens*. Essa é a menor taxa entre os três subcorpora, o que faz sentido, dado tratar-se de uma produção humana em L1. Olhando para as categorias, a mais frequente foi Semântica, contabilizando 57,1% das ocorrências. A menos frequente, a seu turno, é Gramática, com 3,5% das ocorrências. Assim, diferente do que ocorre no caso dos erros totais, Léxico é apenas a segunda categoria menos frequente. As ocorrências discriminadas por categoria podem ser conferidas no Gráfico 2.

Gráfico 2 - Porcentagem de erros por categoria (EL1)



Fonte: Elaboração própria, 2025.

Em Semântica, os maiores problemas foram aqueles de Coerência/Coesão, que respondem por 202 das 242 ocorrências. Quanto ao Discurso, a questão principal parece ter sido a repetição, que se fez presente 64 vezes, em comparação com apenas 12 ocorrências de registro. No caso da Ortografia, categoria que recorreu quase tanto quanto a sua antecessora na hierarquia, a etiqueta mais frequente foi a de pontuação, seguida pela de grafia, esta com 20 marcações a menos. Gramática e Lógico, por sua vez, não apresentaram nenhuma etiqueta especialmente recorrente.

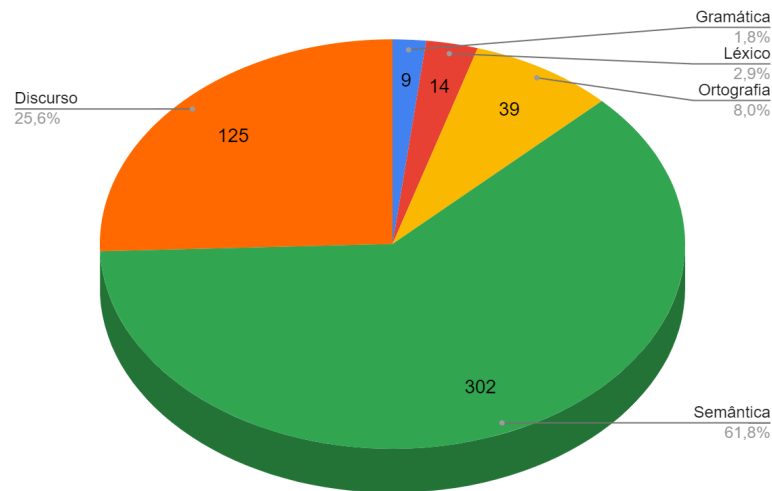
Na passagem para a tradução automática, o número de *tokens* diminuiu, mas os erros aumentaram tanto em termos absolutos quanto relativos. Ao todo, houve 486 erros, correspondendo a uma frequência de 20,68 erros a cada 100 *tokens*. As duas categorias responsáveis pelo aumento dos erros foram Semântica e Discurso, enquanto as outras três sofreram ligeiras reduções. As porcentagens podem ser conferidas em detalhes no Gráfico 3.

Com relação às reduções, a queda foi especialmente expressiva na Ortografia, que passou de 17,2% das ocorrências para apenas 8%. Isso se deveu, sobretudo, à etiqueta relativa aos problemas de pontuação, que apareceu bem menos vezes nos textos traduzidos automaticamente. Assim, é possível perceber que, no que tange a esse tipo de erro, bem mais ligado a regras do que ao uso da língua, o tradutor automático é satisfatoriamente eficaz em sua correção.

Referente aos aumentos, dentro de Discurso, a grande responsável pelo crescimento de 8,1% das ocorrências foi a etiqueta “Estilística - registro”. Isso reforça o que foi dito anteriormente da baixa adaptação do DeepL ao registro específico que estava traduzindo —

afinal, houve a ocorrência de 42 novos erros. Já no caso de Semântica, o aumento da taxa se deveu sobretudo aos erros de colocação, que subiram de 8 ocorrências para 70.

Gráfico 3 - Porcentagem de erros por categoria (TA)



Fonte: Elaboração própria, 2025.

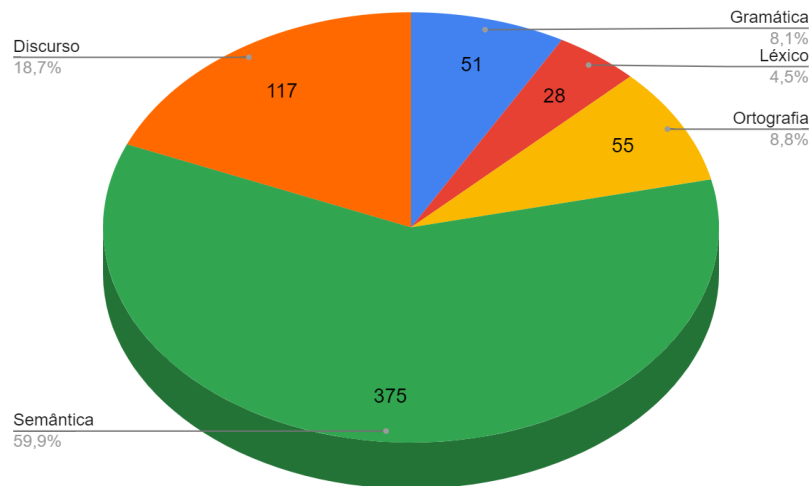
É provável que isso guarde estreita relação com o aumento dos problemas de registro, uma vez que grande parte das colocações presentes nos textos eram do domínio acadêmico. Considerando que o *DeepL* está traduzindo do português brasileiro para o inglês, é possível que sequências de palavras que constituem colocações naquela língua tenham sido literalmente traduzidas para esta, na qual não necessariamente operam da mesma forma. Para que isso fosse confirmado, no entanto, seria necessário uma investigação comparativa das ocorrências individuais.

Por fim, a autopós-edição dos textos traduzidos automaticamente também introduziu novos erros, elevando as ocorrências para 626, correspondente a uma taxa de erros de 25,73%. Como se vê no Gráfico 4, todas as categorias, com exceção de Discurso, sofreram aumentos em números absolutos. Ainda assim, em termos relativos, a representatividade dos erros semânticos também caiu em relação ao todo, devido ao aumento dos erros ortográficos, lexicais e gramaticais. Inclusive, o aumento dos erros gramaticais em relação ao todo foi tão expressivo que os fez ultrapassar os erros lexicais. Nesse caso, o maior problema dos participantes parece ter sido a concordância verbal, que cresceu de 5 para 37 ocorrências.

Assim, em suma, na passagem do português (EL1) para o inglês via tradução automática (TA), houve um crescimento de 3,34% na frequência dos erros. Isso porque, embora o *DeepL* tenha sido bem sucedida em reduzir os erros ortográficos, lexicais e gramaticais, acabou por introduzir outros ligados à Semântica e ao Discurso. Isso condiz com o que a literatura relata a respeito das limitações dos tradutores automáticos (Bowker, Ciro,

2019; Wang *et al.*, 2021), uma vez que essas duas categorias estão muito mais condicionadas ao uso que se faz da língua, que não é exato e varia de forma constante, tornando-se difícil ao entendimento do sistema de tradução automática. Também sugere uma baixa proficiência em autopós-edição por parte dos participantes.

Gráfico 4 - Porcentagem de erros por categoria (PE L2)



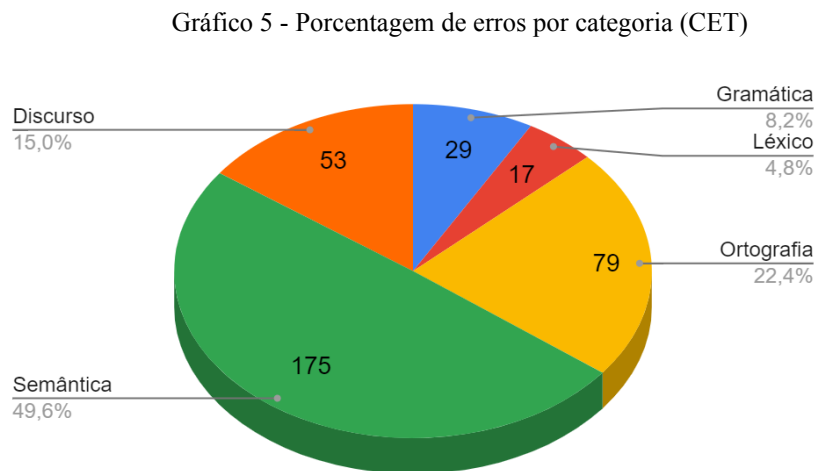
Fonte: Elaboração própria, 2025.

Durante a autopós-edição, os participantes acabaram por introduzir erros que o tradutor havia sido bem sucedido em evitar, especialmente os gramaticais. O principal problema parece estar nos verbos, que, por isso, precisariam ser apontados como um ponto a se ficar atento durante uma autopós-edição. Além disso, houve um aumento nos erros semânticos totais. No caso dos problemas de coerência e coesão, eles já se mostravam frequentes na EL1, portanto, não parecem relacionados à PE L2. Quanto ao crescimento de ocorrências ligadas às colocações, isso ajuda a sinalizá-las como outro aspecto a se ficar atento ao pós-editar um texto.

No caso da autopós-edição, no entanto, além da hipercorreção, houve um fator adicional, que foi a introdução de novos sintagmas e, mesmo, sentenças que não constavam nos dois textos anteriores. Em um estágio futuro do projeto, em que se disponha de mais dados para comparação, seria interessante analisar a relação entre essa inclusão de porções textuais durante a autopós-edição e a ordem em que os participantes estavam executando as tarefas, para averiguar uma possível influência da escrita em L2 nessa realização de acréscimos.

4.4 Contagem dos erros por área

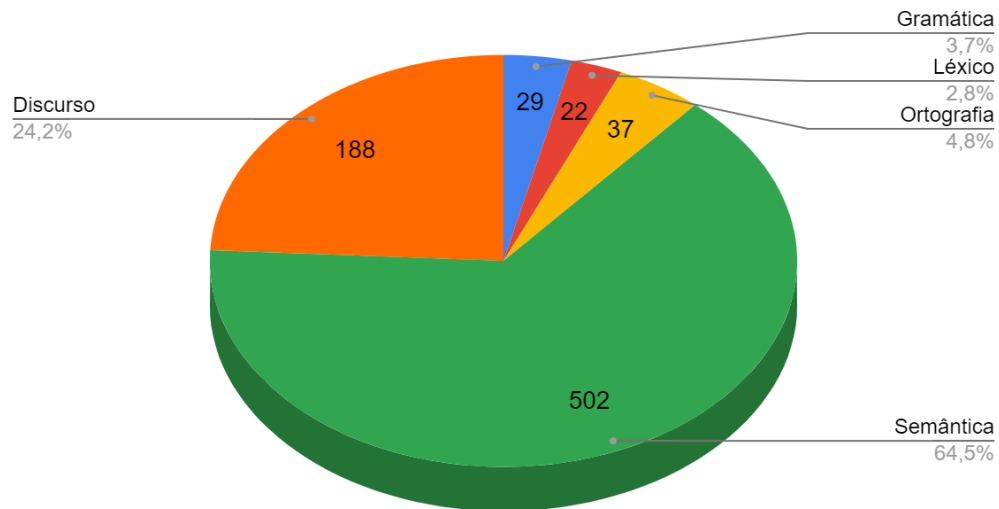
Em seguida, os erros foram contabilizados de acordo com as grandes áreas. Dentre todas as áreas, CET foi a que contabilizou menos erros, com uma taxa de 14,35% E/T. Em número absolutos, esse grupo somou 353 erros, com quase 50% deles pertencendo à categoria Semântica. Diferentemente do que ocorre no caso de CH e CSA, os problemas de ambiguidade foram mais recorrentes do que os de colocação incorreta. Isso, somado à alta frequência dos erros de coerência/coesão (Semântica), e à baixa presença dos erros de registro (Discurso), especialmente em comparação aos erros de repetição (Discurso), sugere que o maior problema dessa área não é com a adequação ao registro acadêmico, mas com a construção e o encadeamento dos significados ao longo do texto. Ademais, diferentemente do que ocorre com CH e CSA, os erros ortográficos foram mais frequentes do que os discursivos. De fato, CET apresentou bem mais erros de pontuação (2,24 a cada 100 *tokens*), o que reforça a tese do prejuízo da construção textual. Dentre todos, os erros lexicais foram os menos numerosos, com 17 ocorrências totais. As quantidades em sua relação com o todo podem ser conferidas no Gráfico 5.



Fonte: Elaboração própria, 2025.

CH, por sua vez, foi a área com a segunda maior taxa de erros: foram 778 em 3255 *tokens*, correspondendo a 23,95% E/T. Assim como no caso anterior, os erros mais frequentes foram os de Semântica, que, dessa vez, respondem por mais da metade do total (64,5%); enquanto isso, os lexicais contaram apenas 2,8% das ocorrências. Desse modo, dentre as três áreas, CH é a que apresenta a maior discrepância entre a categoria com maior número de erros (502) e a com menor (22), conforme mostra o Gráfico 6.

Gráfico 6 - Porcentagem de erros por categoria (CH)



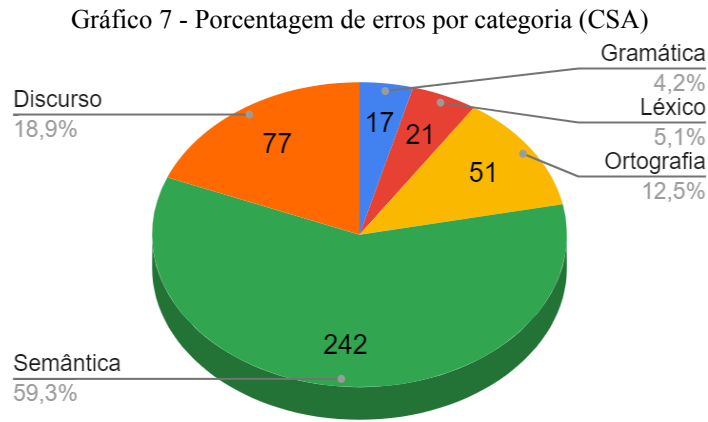
Fonte: Elaboração própria, 2025.

Entre as três áreas, CH apresentou o maior número de problemas discursivos. Diferentemente das outras, teve mais erros de registro do que de repetição. Na verdade, enquanto em CET e CSA os erros de repetição foram mais que o dobro dos de registro, em CH os números foram próximos, com os erros de registro superando os de repetição por apenas duas ocorrências. Isso sugere que essa grande área enfrenta mais dificuldades para adequar-se ao registro acadêmico, possivelmente devido à menor publicação de artigos científicos, resultando em menos experiência acumulada. A isso, somam-se as 92 ocorrências de erros de repetição, que contribuem para que CH tenha a maior e mais influente taxa de erros discursivos. Além disso, a área também apresenta bastantes dificuldades na estruturação de um texto coeso e coerente, uma vez que a taxa desse tipo de erro é de 11% E/T — a maior entre as três grandes áreas.

Por fim, CSA foi a primeira colocada no que diz respeito à porcentagem de E/T (26,7%). Novamente, Semântica responde por mais da metade dos erros cometidos, mas, diferentemente das duas áreas anteriores, nesse caso os erros gramaticais respondem pela menor fatia. No entanto, em todos os três casos, ambas as categorias — Léxico e Gramática — tiveram números de ocorrência bastante próximos, especialmente em comparação às demais categorias. Um resumo dos dados pode ser visto no Gráfico 7.

Dentro de Semântica, as categorias mais problemáticas, são, de forma similar ao que ocorre com CH, as de colocações incorretas e de coerência/coesão. Na verdade, no que tange à dificuldade de estruturar um texto coeso/coerente e que se adeque ao registro acadêmico, CSA é a segunda área com mais dificuldade. Demonstra, também, ter tido problemas durante

a revisão dos textos produzidos, dada a alta ocorrência de erros de grafia: 2,55% E/T, quase o triplo das ocorrências de CET, segunda colocada nesse quesito.



Fonte: Elaboração própria, 2025.

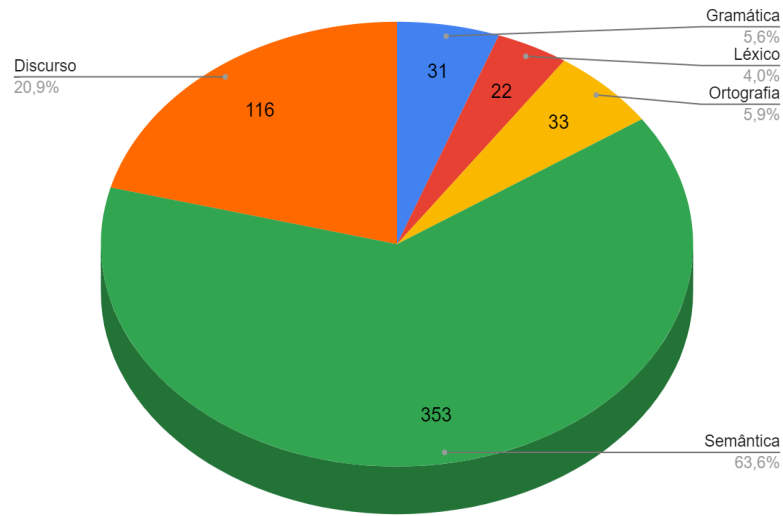
Então, em resumo, todas as três áreas apresentaram problemas na construção e no encadeamento dos significados, registrando bastantes ocorrências de problemas ligados à coerência/coesão e à repetição de palavras. CET, ainda que apresente menos problemas de repetição, traz ainda uma maior taxa de erros de pontuação, que pode sugerir prejuízo da estrutura textual caso sejam, por exemplo, ocorrências de vírgulas incorretas. A confirmação disso, no entanto, demandaria uma análise caso a caso.

CH e CSA destacaram-se ainda pelas dificuldades em adequar-se ao registro acadêmico, com uma maior ocorrência de problemas ligados ao registro. A alta taxa de problemas colocacionais pode indicar ainda, uma menor proficiência em língua inglesa, que, embora também esteja presente em CET, o faz com menor intensidade. A verificação disso, no entanto, depende dos resultados da análise combinada entre área e modo de produção, a qual será reportada mais a frente, em seção posterior.

4.5 Contagem dos erros por ordem

Na etapa seguinte, os erros foram contabilizados por ordem de execução das tarefas. Na Ordem 1, os participantes começaram pela EL1, depois redigiram o resumo em L2 e, por fim, autopós-editaram o resumo em português brasileiro traduzido automaticamente para o inglês. Ou seja, a sequência foi: EL1 → EL2 → PE L2. Ao todo, o grupo de doutorandos que obedeceu a essa ordem incorreu em 555 erros, correspondentes a uma taxa de 25,26% E/T — a maior dentre os três grupos. A distribuição das ocorrências em relação às categorias se encontra ilustrada no Gráfico 8.

Gráfico 8 - Porcentagem de erros por categoria (Ordem 1)

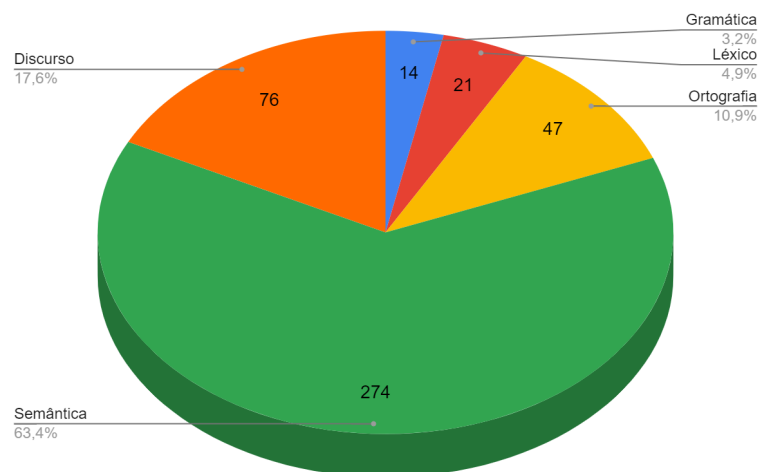


Fonte: Elaboração própria, 2025.

Desse total, 84,5% está concentrado apenas nas categorias relativas à Semântica e ao Discurso, demonstrando que os principais desafios residem na construção e no encadeamento dos significados, bem como na adequação ao registro acadêmico. Isso se confirma quando olhamos as etiquetas com mais ocorrências: no caso da Semântica, colocação incorreta (3,69% E/T), coerência/coesão (9,42% E/T) e ambiguidade (2,37% E/T); no caso de Discurso, estilística - registro (3,10% E/T) e repetição (2,10% E/T). A recorrência dos problemas de colocação pode sugerir, ainda, dificuldades com a língua inglesa, as quais serão averiguadas, assim como no caso da área, na análise combinada com o modo de produção.

Na Ordem 2, os participantes foram orientados a, primeiro, escrever em L2, depois em L1 e, por fim, realizar a pós-edição em L2, conforme o Gráfico 9.

Gráfico 9 - Porcentagem de erros por categoria (Ordem 2)



Fonte: Elaboração própria, 2025.

Assim, o arranjo foi: EL2 → EL1 → PE L2. O grupo que seguiu essa ordem contabilizou, em sua totalidade, 432 erros, resultando em uma taxa de 16,62% E/T — a menor entre as três ordens. Discriminados por categoria, os erros dão origem ao Gráfico 9.

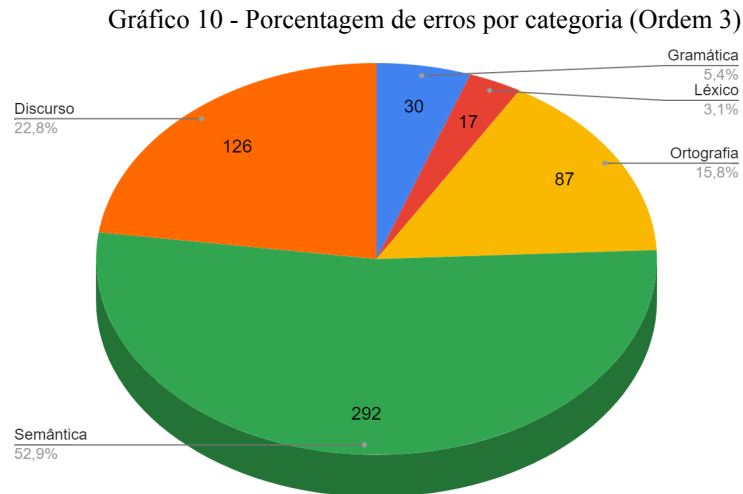
Semântica segue sendo a categoria mais frequente, representando 63,4% das ocorrências, total bem próximo ao encontrado na Ordem 1. No entanto, ainda comparando ambas as ordens, há uma queda nos erros discursivos de 3,3%, bem como um aumento de 5% nos ortográficos. Além disso, na relação com o todo, dessa vez, os erros lexicais mostram-se mais significativos do que os gramaticais, embora tanto a Ordem 1 quanto a Ordem 2 registrem números baixos de ocorrências desses tipos.

Observando-se as principais etiquetas a influenciarem nos resultados, a frequência mais elevada de erros ortográficos deveu-se, sobretudo, à maior presença dos problemas de pontuação (1,27% E/T). A menor taxa de erros gramaticais, por sua vez, adveio do fato de os participantes terem apresentado menos problemas em relação ao uso dos verbos, registrando apenas 8 ocorrências do tipo.

Quanto aos erros discursivos, a queda na comparação com a Ordem 1 ocorreu, em grande parte, devido ao fato de os participantes da Ordem 2 terem incorrido em apenas três inadequações de registro, indicando uma maior familiaridade com o registro acadêmico em ambas as línguas. A queda só não foi maior porque houve um aumento de 0,55% nas ocorrências da etiqueta “estilística - repetição”, que passaram a apresentar uma frequência de 2,73% E/T.

Finalmente, no que tange aos erros semânticos, o maior problema foi coerência/coesão, que registrou uma frequência de 7,85% E/T. Ainda que seja uma taxa inferior àquela presente na Ordem 1, indica que os doutorandos que seguiram a Ordem 2 também tiveram problemas na elaboração e conexão dos significados, o que é reforçado pela recorrência de repetições e pela elevação no número das pontuações ausentes ou inadequadamente empregadas. Entretanto, o menor número de erros de colocação sugere uma maior proficiência em inglês, a ser confirmada na análise da ordem combinada ao modo de produção.

Na ordem 3, por sua vez, os doutorandos escreveram o resumo em português, na sequência o pós-editaram e, só então, produziram o resumo em inglês (EL1 → PE L2 → EL2). Nessa sequência, houve um total de 552 erros em 2446 *tokens*, equivalentes a uma taxa de 22,57% — portanto, a segunda colocada entre as ordens. No Gráfico 10, é possível ver as ocorrências distribuídas por categorias.



Fonte: Elaboração própria, 2025.

Mais uma vez, Semântica encabeça a lista de erros mais frequentes. Entretanto, diferentemente das ordens anteriores, na Ordem 3, os erros semânticos constituem pouco mais de 50% das ocorrências totais. Enquanto isso, amplia-se a relevância dos erros discursivos e ortográficos. Ademais, assim como na Ordem 1, os problemas gramaticais são mais recorrentes do que os lexicais, registrando quase o dobro de ocorrências.

Analisando por etiqueta, vê-se que, de forma distinta ao que ocorre nas demais ordens, o número de ocorrências gramaticais deve-se, em grande parte, à frequência dos erros de regência verbal. Quanto à Ortografia, os casos ligados à grafia são muito mais recorrentes do que nos grupos anteriores, resultando em uma taxa de 2% E/T. Quanto à Semântica e ao Discurso, há registros relevantes de todas as etiquetas ligadas à construção e ao encadeamento dos significados — coerência/coesão, ambiguidade, repetição — e, também, à adequação ao registro acadêmico e à língua — colocações e registro —, sugerindo que ambos os pontos representam problemas aos doutorandos que seguiram a Ordem 3. Novamente, a relação disso com a proficiência linguística só será esclarecida em momento posterior da análise.

Como ressaltado anteriormente, para entender melhor como todos esses números — por modo de produção, por área e por ordem — interagem, é preciso analisá-los em combinação. Por isso, é a isso que as próximas duas seções se voltam. No entanto, antes de prosseguir, é interessante observar que a ordem que apresentou a menor taxa de erros (e, sobretudo, a menor taxa de erros relacionada à adequação ao registro acadêmico) é a única na qual a escrita em inglês precede todas as demais tarefas. Assim, a EL2 ao início parece facilitar a pós-edição nessa língua. Por outro lado, quando a EL1 é a primeira das tarefas, o número de erros tende a ser superior, independente de a pós-edição vir imediatamente na

sequência ou não. Cabe questionarmo-nos, então, em que medida escrever diretamente na L2 influencia na EL1 e/ou na forma como modificamos o texto traduzido automaticamente.

4.6 Contagem dos erros por área e modo de produção

Começando por CET, que foi a área a cometer menos erros, é possível notar que, contrário à tendência geral do corpus e ao que houve nas outras áreas, aqui ocorre uma diminuição constante no número de *tokens*, de forma que os textos autopós-editados somam 72 a menos do que os resumos em português. Considerando ser o inglês uma língua mais sintética, é natural que isso ocorra durante a tradução automática, e o fato de que os participantes mantêm a tendência mesmo ao pós-editarem pode sugerir algum nível de consciência dessa diferença entre as línguas.

O número de erros, no entanto, varia de forma inversamente proporcional, crescendo conforme os *tokens* diminuem. Assim, na passagem para a TA, há uma elevação de 2,4% na taxa de E/T, que cresce ainda em mais 6,56% durante a pós-edição, de modo que há um crescimento total de 8,6% na taxa de erros. As informações completas podem ser conferidas na Tabela 6.

Tabela 6 - Erros por modo de produção (CET)

CET											
EL1			TA			PE L2					
Total	94		Total	105		Total	154				
Tokens	861		Tokens	810		Tokens	789				
E/T	10,92%		E/T	12,96%		E/T	19,52%				
Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T			
Ortografia	31	32,98%	3,60%	Ortografia	22	20,95%	2,72%	Ortografia	26	16,88%	3,30%
Léxico	3	3,19%	0,35%	Léxico	5	4,76%	0,62%	Léxico	9	5,84%	1,14%
Gramática	7	7,45%	0,81%	Gramática	5	4,76%	0,62%	Gramática	17	11,04%	2,15%
Semântica	37	39,36%	4,30%	Semântica	55	52,38%	6,79%	Semântica	83	53,90%	10,52%
Discurso	16	17,02%	1,86%	Discurso	18	17,14%	2,22%	Discurso	19	12,34%	2,41%

Fonte: Elaboração própria, 2025.

Olhando para as categorias e etiquetas, na escrita em português brasileiro, os principais problemas concentraram-se em três categorias. No caso da primeira delas, Ortografia, a principal etiqueta foi Pontuação. Isso mostra que, ainda que estejam escrevendo na língua materna e que parte desses problemas possam ter advindo da ausência de uma revisão do texto produzido, os sinais de pontuação se impõem como um problema aos

doutorandos. Assim, provam-se um aspecto linguístico a ser trabalhado não só durante a autopós-edição, mas durante a pré-edição, de modo a contribuir não só para a melhora da tradução automática gerada, mas para a qualidade dos textos que sejam redigidos em português para publicações nacionais.

Na passagem pela DeepL durante a tradução automática, os erros de pontuação caíram. Assim, comparando os números absolutos e relativos do texto fonte àqueles do texto final, o DeepL auxiliou no processo de redução do número de problemas pontuacionais. No entanto, durante a autopós-edição, os doutorandos acabaram por introduzir novos erros de grafia e letra maiúscula, de forma que a queda nos erros ortográficos se tornou menos expressiva.

A outras duas categorias a reunirem mais erros foram Semântica e Discurso. Nelas, os principais problemas relacionam-se às previamente mencionadas construção e conexão dos significados do texto, essenciais à estrutura textual; ou seja: coerência/coesão e ambiguidade (no caso da Semântica) e repetição (no caso do Discurso). Durante a tradução automática, o DeepL não foi capaz de diminuir suas ocorrências, acabando por acrescentar outras, ligadas ao uso de colocações.

É interessante observar, no entanto, que, apesar de o DeepL ter introduzido erros colocacionais, durante a autopós-edição, os doutorandos foram bem sucedidos na sua redução, o que sugere um nível de conhecimento do uso do inglês acadêmico. Isso não ocorreu, porém, com as demais etiquetas semânticas, que viram seu número crescer com a autopós-edição. “Estilística-repetição” (Discurso), porém, permaneceu constante, somando 12 ocorrências absolutas em todos os casos.

Em seguida, passando para CH, a área com a segunda maior taxa de erros, dessa vez, ocorre um aumento significativo de *tokens* e de erros durante a autopós-edição (Tabela 7). Durante a escrita do resumo em português, o principal problema, como em todos os demais casos, resumiu-se às categorias Semântica e Discurso, responsáveis por 87,83% dos erros. Como ocorreu com CET, a construção e conexão dos significados do texto foram o maior desafio, o que levou a uma taxa elevada, principalmente, de problemas de coerência e coesão e de repetição. Na verdade, em comparação à primeira área analisada, a frequência da etiqueta coerência/coesão foi bastante superior, registrando 14,1% E/T, contra 2,21% de CET.

Tabela 7 - Erros por modo de produção (CH)

CH											
EL1				TA				PE L2			
Total	230			Total	226			Total	322		
Tokens	1057			Tokens	1068			Tokens	1130		
E/T	21,76%			E/T	21,16%			E/T	28,50%		
Categoria	Cat./Total	E/T		Categoria	Cat./Total	E/T		Categoria	Cat./Total	E/T	
Ortografia	20	8,70%	1,89%	Ortografia	4	1,77%	0,37%	Ortografia	13	4,04%	1,15%
Léxico	6	2,61%	0,57%	Léxico	4	1,77%	0,37%	Léxico	12	3,73%	1,06%
Gramática	2	0,87%	0,19%	Gramática	3	1,33%	0,28%	Gramática	24	7,45%	2,12%
Semântica	165	71,74%	15,61%	Semântica	134	59,29%	12,55%	Semântica	203	63,04%	17,96%
Discurso	37	16,09%	3,50%	Discurso	81	35,84%	7,58%	Discurso	70	21,74%	6,19%

Fonte: Elaboração própria, 2025.

Depois de os textos serem traduzidos automaticamente, houve uma redução de 16 ocorrências de erros ortográficos e de 31 ocorrências de erros semânticos, o que acabou por culminar na redução da taxa de E/T em 0,5%. Essa redução, no entanto, foi freada pela introdução de erros discursivos, uma vez que o tradutor automático elevou a taxa de 16,06% E/T para 34,84% ao elevar as ocorrências da etiqueta ligada ao registro acadêmico. Dessa vez, no caso da Semântica, houve uma redução absoluta de 65 problemas de coerência/coesão; mas, novamente, a interferência do sistema de tradução automática resultou na inclusão de problemas colocacionais (33 ao todo).

Durante a autopós-edição, a taxa de E/T sofre um salto de 7,34% que se reflete em todas as categorias, menos Discurso, a única na qual ocorre uma redução, ainda que apenas de 1,39% E/T. Gramática, que passou pelo segundo maior aumento, registrou mais ocorrências de erros ligados à concordância verbal. No caso da Semântica, há um aumento não só dos problemas de coerência/coesão quanto dos colocacionais, o que sugere um menor domínio do inglês acadêmico, reforçado pela persistência dos problemas de registro.

Finalmente, no que tange à CSA, a área com a maior taxa de erros, é durante a tradução automática que é introduzida a maior parte dos erros, conforme se observa na Tabela 8. Isso, combinado à redução dos *tokens*, faz com que o salto na taxa de erros seja de 13,46% em relação à EL1. Mais uma vez, Semântica é o maior desafio para o tradutor automático, que acaba por acrescentar 73 erros desse tipo durante a tradução automática. Assim, embora Gramática, Léxico e Ortografia registrem uma queda nas ocorrências e Discurso se mantenha quase constante, o aumento exponencial dos problemas semânticos resulta na elevação da taxa de E/T.

Tabela 8 - Erros por modo de produção (CSA)

CSA											
EL1				TA				PE L2			
Total		100		Total		158		Total		150	
Tokens		527		Tokens		487		Tokens		514	
E/T		18,98%		E/T		32,44%		E/T		29,18%	
Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T
Ortografia	22	22,00%	4,17%	Ortografia	13	8,23%	2,67%	Ortografia	16	10,67%	3,11%
Léxico	9	9,00%	1,71%	Léxico	5	3,16%	1,03%	Léxico	7	4,67%	1,36%
Gramática	6	6,00%	1,14%	Gramática	1	0,63%	0,21%	Gramática	10	6,67%	1,95%
Semântica	40	40,00%	7,59%	Semântica	113	71,52%	23,20%	Semântica	89	59,33%	17,32%
Discurso	23	23,00%	4,36%	Discurso	26	16,46%	5,34%	Discurso	28	18,67%	5,45%

Fonte: Elaboração própria, 2025.

Examinando-se as etiquetas de Semântica, é possível perceber que os erros de colocação saltaram de 4 (0,76% E/T) para 17 ocorrências (3,49% E/T), enquanto os de coerência/coesão passaram de 35 (6,6% E/T) para 85 (17,45% E/T). Dentre todas as três áreas, essa foi a única que testemunhou um aumento desse tipo de erro, e não uma redução. Durante a autopós-edição, há um acréscimo de *tokens* e uma diminuição de erros, o que faz com que a taxa de E/T caia para 29,18%. Essa redução concerne unicamente os erros semânticos, uma vez que todas as demais categorias sofrem aumentos: Ortografia testemunha o aumento de problemas de grafia; Léxico, de omissão de palavras gramaticais; Gramática, de concordância verbal; Discurso, de registro. Semântica, no entanto, embora registre um aumento de 50% nas ocorrências de colocação incorreta, também contabiliza uma queda de 39 das ocorrências de coerência/coesão, o que mostra que, ainda que os doutorandos tenham apresentado dificuldades ao lidar com o uso do inglês acadêmico, conseguiram aprimorar a construção do texto como um todo.

Desse modo, o que se observa é que, para todas as três áreas, a passagem do resumo em português para a autopós-edição implicou um aumento na taxa dos erros; mas, enquanto no caso do CSA, esse salto ocorreu via interferência do sistema de tradução automática, no caso de CET e de CH, esse salto veio diretamente durante a autopós-edição. Na verdade, ao contrário das outras duas grandes áreas, na PE, CSA reduziu em parte a taxa de E/T, justamente porque os participantes dessa área conseguiram diminuir as ocorrências ligadas à coerência/coesão, que o DeepL havia mais do que dobrado. Em pesquisas futuras, seria interessante investigar, então, que aspectos textuais dos resumos em português dessa área, na comparação com os das demais, podem ser os mais relevantes para determinar este resultado.

4.7 Contagem dos erros por ordem e modo de produção

A última etapa da análise dos dados quantitativos voltou-se para os erros contabilizados, simultaneamente, por ordem e modo de produção. No caso da Ordem 1 (EL1 → EL2 → PEL2), responsável pela maior quantidade de erros dentre as ordenações (25,26%), a inclusão de erros aconteceu principalmente na passagem no texto pelo DeepL, e a taxa não foi reduzida pelos participantes durante autopós-edição, conforme mostra a Tabela 9.

Assim, no resumo em português, há um total de 19,75 erros a cada 100 *tokens*, enquanto nos textos gerados pelo DeepL o valor é de 23,76% E/T. Essa elevação de 4,1% acontece não só pela proliferação das etiquetas, mas pela redução na quantidade de texto em que se distribuem. Na autopós-edição, o aumento dos erros é ainda mais significativo: 79 na relação com os textos do TA. Ao aumento dos erros, corresponde um crescimento no tamanho dos textos, o que faz com que a taxa de E/T permaneça em 23,88%. Considerando-se que essa é a única ordem na qual a EL2 ocorre imediatamente antes da PE L2 (e a única em que o número de *tokens* cresce), cabe perguntar-se em que medida a redação de um segundo resumo, diretamente em língua inglesa, impactou na decisão de acrescentar palavras ao resumo durante a autopós-edição.

Tabela 9 - Erros por modo de produção (Ordem 1)

ORDEM 1											
EL1			TA			PE L2					
Total	142		Total	168		Total	245				
Tokens	719		Tokens	707		Tokens	1026				
E/T	19,75%		E/T	23,76%		E/T	23,88%				
Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T			
Ortografia	20	14,08%	2,78%	Ortografia	2	1,19%	0,28%	Ortografia	11	4,49%	1,07%
Léxico	5	3,52%	0,70%	Léxico	5	2,98%	0,71%	Léxico	12	4,90%	1,17%
Gramática	5	3,52%	0,70%	Gramática	3	1,79%	0,42%	Gramática	23	9,39%	2,24%
Semântica	94	66,20%	13,07%	Semântica	107	63,69%	15,13%	Semântica	152	62,04%	14,81%
Discurso	18	12,68%	2,50%	Discurso	51	30,36%	7,21%	Discurso	47	19,18%	4,58%

Fonte: Elaboração própria, 2025.

No que tange às categorias, o sistema de tradução automática foi responsável por, simultaneamente, reduzir os erros de Gramática e de Ortografia e elevar os semânticos e discursivos, padrão que vinha se mostrando nos dados anteriores. Diante disso, na autopós-edição, os doutorandos acabaram por introduzir mais ocorrências em todas as

categorias, com exceção de Semântica e Discurso, que viram uma queda de suas taxas de E/T.

Especificamente no caso dessas duas últimas categorias, as etiquetas mais afetadas pelo tradutor automático foram “colocação incorreta” e “estilística - registro”. Assim, embora tenha havido uma queda nos problemas de coerência/coesão, a introdução de 31 problemas colocacionais fez com que o saldo final fosse negativo, levando a um aumento de 2,06% da E/T. No caso de Discurso, a taxa foi elevada em 4,71%, em razão da dificuldade do DeepL em lidar com inadequações de registro, que passaram de apenas 4 marcações na EL1 para 31 nos textos traduzidos automaticamente.

A autopós-edição resulta, mais uma vez, em um aumento principalmente dos problemas de concordância verbal, elevando a taxa de Gramática. Léxico e Ortografia também testemunham elevações, devido ao acréscimo de ocorrências ligadas à omissão de palavras gramaticais e a erros de grafia. No caso de Semântica, embora o número relativo de colocações incorretas cresça, o de erros de coerência/coesão cai levemente, levando à redução geral da taxa de E/T da categoria. No caso de Discurso, é a redução dos problemas de repetição que resulta na diminuição na quantidade de erros a cada 100 *tokens*. Assim, ainda que os participantes dessa ordem tenham apresentado problemas com a língua inglesa e com o registro acadêmico, foram bem sucedidos em reduzir parte dos problemas ligados à estruturação dos sentidos do texto. Dessa forma, cabe se perguntar de que forma a EL2 imediatamente anterior à PE L2 influenciou a qualidade do texto resultante, pergunta que só poderia ser respondida em uma futura análise que investigasse os erros dos resumos escritos diretamente em inglês .

Ao contrário do que acontece na Ordem 1, na Ordem 2 (EL2 → EL1 → PEL2), não há uma inclusão de erros significativa por parte do sistema de tradução automática, de forma que o maior aumento ocorre na realização da autopós-edição (11%). Além disso, diferentemente da ordem anterior, há uma redução progressiva no número de *tokens*, o que contribui para um aumento ainda maior da taxa, mesmo que o número absoluto de erros não se eleve tanto. Essas informações podem ser conferidas na Tabela 10.

Como dito anteriormente, dentre as três ordens, a 2 foi a que apresentou a menor taxa de E/T: 16,62%. Também foi a única na qual o DeepL reduziu o número absoluto de erros durante a tradução automática.

Tabela 10 - Erros por modo de produção (Ordem 2)

EL1			TA			PE L2					
Total	136		Total	135		Total	161				
Tokens	893		Tokens	848		Tokens	606				
E/T	15,23%		E/T	15,92%		E/T	26,57%				
Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T			
Ortografia	22	16,18%	2,46%	Ortografia	10	7,41%	1,18%	Ortografia	15	9,32%	2,48%
Léxico	7	5,15%	0,78%	Léxico	7	5,19%	0,83%	Léxico	7	4,35%	1,16%
Gramática	6	4,41%	0,67%	-	-	-	-	Gramática	8	4,97%	1,32%
Semântica	77	56,62%	8,62%	Semântica	93	68,89%	10,97%	Semântica	104	64,60%	17,16%
Discurso	24	17,65%	2,69%	Discurso	25	18,52%	2,95%	Discurso	27	16,77%	4,46%

Fonte: Elaboração própria, 2025.

Na verdade, fora Semântica, que viu um aumento nos casos de colocação incorreta, nenhuma outra categoria testemunhou um aumento expressivo de suas ocorrências, que ou foram reduzidas, ou foram mantidas. Ademais, de todos os casos analisados até aqui, somente neste o tradutor automático não acrescentou um problema de registro que fosse. Assim, interessa, no futuro, investigar que aspectos da redação e/ou da pré-edição do texto em EL1 podem ter contribuído para que o texto resultante não destoasse do registro acadêmico, por mais que o DeepL seja um tradutor automático para fins gerais, sem especialização neste domínio particular, tendo mesmo introduzido problemas colocacionais.

Durante a autopós-edição, embora os erros de Gramática tenham sido reintroduzidos (novamente, por causa da concordância verbal), o maior problema para os participantes foram os erros semânticos, mais especificamente aqueles relacionados ao uso das colocações; portanto, ao uso do inglês acadêmico. Apesar disso, a adoção e manutenção do registro nessa língua não foi um empecilho, de forma que foram registradas apenas três ocorrências desse tipo. É interessante observar que, além disso, os doutorandos da Ordem 2 foram os únicos dentre as três ordens a reduzir as ocorrências de coerência/coesão em números absolutos durante a autopós-edição, ainda que, na relação com a totalidade dos *tokens*, a taxa de E/T tenha aumentado.

A seu turno, a Ordem 3 (EL1 → PEL2 → EL2) foi a que apresentou a segunda maior taxa de E/T, com uma porcentagem de 22,57%. Além disso, diferente do que ocorre nas ordens anteriores, o aumento entre os modos de produção é progressivo: na passagem do texto em L1 pelo DeepL, há um crescimento de 5,43% na taxa de E/T, e, na passagem da TA para a PE L2, de mais 4,44% (ver Tabela 11). Essa foi também a única ordem em que a PE L2 ocorreu antes da EL2. Assim, interessa investigar, por um lado, os motivos pelos quais os

textos em EL1 redigidos antes da EL2 são mais problemáticos para o sistema de tradução automática (caso das Ordens 1 e 3) e, por outro, as razões pelas quais a EL1 imediatamente antes da PE L2 torna a autopós-edição mais problemática (caso das Ordens 2 e 3), menos pelo aumento dos erros que pela redução dos tokens

Olhando para as categorias, que podem ser conferidas na Tabela 11, os maiores desafios na EL1 relacionam-se à Ortografia (grafia e pontuação), à Semântica (coerência/coesão) e ao Discurso (repetição). Diferentemente do que ocorre nas ordens anteriores, dessa vez, o tradutor automático não consegue lidar de forma efetiva com as questões ortográficas, aumentando as ocorrências de grafia, ainda que reduza as de pontuação. Além disso, introduz 20 novos problemas colocacionais, eleva os casos de coerência/coesão, duplica os de ambiguidade e introduz 21 novos problemas de registro. Dentre todas as ordens, inclusive, foi a única na qual o tradutor automático elevou os casos de problemas gramaticais. Isso sugere que, embora tanto a Ordem 1 quanto a Ordem 3 tenham apresentado resumos em português problemáticos para o tradutor automático, eles foram problemáticos de formas diferentes.

Tabela 11 - Erros por modo de produção (Ordem 3)

ORDEM 3											
EL1			TA			PE L2					
Total	34		Total	49		Total	43				
Tokens	833		Tokens	810		Tokens	803				
E/T	17,53%		E/T	22,96%		E/T	27,40%				
Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T	Categoria	Cat./Total	E/T			
Ortografia	31	21,23%	3,72%	Ortografia	27	14,52%	3,33%	Ortografia	29	13,18%	3,61%
Léxico	6	4,11%	0,72%	Léxico	2	1,08%	0,25%	Léxico	9	4,09%	1,12%
Gramática	4	2,74%	0,48%	Gramática	6	3,23%	0,74%	Gramática	20	9,09%	2,49%
Semântica	71	48,63%	8,52%	Semântica	102	54,84%	12,59%	Semântica	119	54,09%	14,82%
Discurso	34	23,29%	4,08%	Discurso	49	26,34%	6,05%	Discurso	43	19,55%	5,35%

Fonte: Elaboração própria, 2025.

Assim, resumindo o que foi visto, a ordem das tarefas parece influenciar na qualidade 1) do texto gerado pelo DeepL e 2) da autopós-edição. Quando a EL1 ocorreu antes da EL2, a o tradutor automático gerou traduções de menor qualidade. Já quando a EL2 ocorreu imediatamente antes da PE L2, a autopós-edição resultou em textos de maior qualidade, devido ao aumento dos tokens sem que isso levasse a um aumento dos erros. Inclusive, a ordem a registrar menos erros foi aquela na qual a EL2 foi a primeira tarefa de todas,

sugerindo que os textos de EL1 também foram melhor produzidos e/ou pré-editados quando a escrita em língua inglesa a precedeu.

4.8 Síntese

Com isso, todas as perguntas de pesquisa encontram-se respondidas, conforme síntese mostrada no Quadro 14.

Quadro 14 - Síntese da pesquisa com as perguntas respondidas

Perguntas	Respostas
Qual é a influência da autopós-edição no número e nos tipos de erros de tradução em textos de diferentes áreas do conhecimento?	No geral, o sistema de tradução automática reduz os erros ortográficos, gramaticais e lexicais, mas eleva os erros semânticos e discursivos, especialmente os de colocação e registro, devido à sua adaptação limitada ao domínio acadêmico. Durante a autopós-edição, todos os erros, exceto os discursivos, aumentam, mesmo aqueles que o sistema de tradução havia eliminado. Destacam-se os problemas com a estruturação dos verbos, da coerência/coesão (presentes já na EL1) e das colocações. Além disso, a posição que a EL2 ocupa na sequência das tarefas parece impactar na quantidade de tokens e de erros presentes nos resumos.
Em que medida o número e os tipos de erros encontrados nos textos em português brasileiro influenciaram os erros encontrados nos textos traduzidos automaticamente para o inglês?	<ul style="list-style-type: none"> ● Na passagem pelo tradutor automático, há um aumento dos erros totais, embora a quantidade de tokens diminua. ● Os erros ortográficos, gramaticais e lexicais são reduzidos, mas os semânticos e discursivos são elevados. Isso porque o tradutor automático não é eficaz em reduzir os problemas anteriores e, ademais, introduz problemas de colocação e registro, mostrando-se pouco adaptada ao registro acadêmico, dado ser um tradutor de uso geral.
Em que medida o número e os tipos de erros encontrados nos textos traduzidos automaticamente para o inglês influenciaram os erros encontrados nos textos autopós-editados em inglês?	<ul style="list-style-type: none"> ● Durante a autopós-edição há um aumento dos erros totais e da quantidade de tokens. ● Todas as categorias, com exceção de Discurso, tem seu número de ocorrências elevado. <ul style="list-style-type: none"> ○ Gramática: há um crescimento, especialmente, dos problemas ligados ao verbo, apontando-o como um ponto nevrálgico. ○ Semântica: evidencia-se uma dificuldade com os erros colocacionais e com as questões ligadas à coerência/coesão, o que indica uma dificuldade em estruturar o texto em ambas as línguas.
Em que medida o número e os tipos de erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente para o inglês e nos resumos autopós-editados em inglês variam entre as diferentes áreas do conhecimento?	<ul style="list-style-type: none"> ● Nas três áreas, crescem os erros na passagem da EL1 para a PE L2, via interferência do tradutor automático (CSA) ou autopós-edição (CET e CH). ● No geral, o sistema de tradução automática foi efetivo em lidar com erros gramaticais, lexicais e ortográficos, porém apresentou problemas consideráveis em lidar com aspectos semânticos e de registro (Discurso). Os erros semânticos, em geral, foram aprofundados durante a autopós-edição, especialmente aqueles ligados ao uso do inglês (colocacionais) e, com exceção de CSA, à coerência/coesão. Os discursivos, com exceção do registro, foram melhor geridos, tendo sido diminuídos (CH e CSA) ou bem pouco aumentados (CET).

Fonte: Elaboração própria, 2025.

Quadro 14 - Síntese da pesquisa com as perguntas respondidas

Perguntas	Respostas
Em que medida o número e os tipos de erros encontrados nos resumos acadêmicos em português, nos resumos traduzidos automaticamente para o inglês e nos resumos autopós-editados em inglês variam conforme a ordem de execução das tarefas?	<ul style="list-style-type: none"> • Nas Ordens 1 e 3, a EL1 antes da PE L2 parece resultar em textos em português brasileiro mais problemáticos para o tradutor automático. No entanto, em cada uma das ordens, eles são problemáticos de formas distintas. • Na Ordem 1, o tradutor automático conseguiu reduzir os erros de gramática e ortografia, mas elevou os discursivos (registro e repetição) e semânticos, pois, apesar de diminuir os de coerência/coesão, elevou os de colocação. • Na Ordem 3, o tradutor automático elevou os erros ortográficos e todos os erros ligados à estruturação dos significados (menos repetições) e ao uso do inglês acadêmico. • Nas Ordens 2 e 3, a EL1 antes da PE L2 leva a uma redução do texto sem que haja redução dos erros, conduzindo a autopós-edições mais problemáticas. • Na Ordem 2, a EL2 imediatamente antes da EL1 culmina em resumos em português brasileiro com uma escrita (ou pré-edição) de maior qualidade.

Fonte: Elaboração própria, 2025.

Além disso, os resultados forneceram outras informações interessantes a respeito da produção textual dos sujeitos envolvidos nas coletas. Na análise da TTR do corpus e dos subcorpora, por exemplo, os dados evidenciaram que, em relação à EL1, o texto de PE L2 não registra uma perda grande de densidade lexical, apesar de haver um aumento dos *tokens* durante a autopós-edição sem que haja um aumento proporcional dos *types*. Quanto à análise por modo de produção, além de responder a uma das perguntas específicas, ela mostrou que os doutorandos, ao contrário do sistema de tradução automática, tenderam a acrescentar texto durante a autopós-edição, abrindo margem para que se questione em que medida isso se relaciona à ordem.

Na análise de erros por área, por sua vez, foi possível perceber diferentes padrões em cada uma das grandes áreas. CET, por exemplo, apresentou uma taxa de erros bem inferior à das demais, abrindo espaço à investigação das razões por trás disso; CH e CSA demonstraram um menor domínio do registro acadêmico, apontando esse como um aspecto essencial de ser trabalhado nessas áreas. Quanto à análise dos erros por ordem, a sequência em que se realiza as tarefas indicou guardar relação com a quantidade de erros, posto que a EL2 parece facilitar a pós-edição, enquanto a EL1 inicial, resultar em mais erros. Assim, resta avaliar, então, como escrever na L2 afeta a EL1 e a autopós-edição.

Por fim, embora Vilar *et al.* (2006) afirmem que, no contexto de seu trabalho, as inadequações estilísticas são de menor importância, na presente dissertação ela foi uma das etiquetas que mais apareceu e influenciou nos resultados, o que já havia ocorrido em estudos anteriores (Popović, 2018). Na verdade, junto aos erros semânticos e aos de concordância

verbal, mostram-se aspectos cruciais a serem trabalhados em um curso de autopós-edição. Isso posto e dando continuidade ao trabalho, a próxima e última seção se volta às considerações finais, na intenção de retomar tudo que foi apresentado na presente dissertação e prospectar os passos futuros.

5 CONSIDERAÇÕES FINAIS

O presente trabalho propôs-se a investigar preliminarmente, por meio de análise de erros, a influência que a autopós-edição tem na qualidade linguística de resumos acadêmicos produzidos por doutorandos de diferentes áreas do conhecimento. Para atingir esse objetivo geral, foram estabelecidos quatro objetivos específicos, voltados à análise dos erros por modo de produção, por área e por ordem.

De forma a viabilizar seu cumprimento, o trabalho adotou uma metodologia que implicava, inicialmente, a compilação e preparação do corpus; na sequência, a anotação manual dos erros com base em uma versão adaptada da taxonomia de Costa *et al.* (2015); e, por fim, a contabilização e análise dos erros. Na discussão dos resultados, foi possível, então, responder às perguntas de pesquisa por trás da investigação proposta na dissertação.

O primeiro objetivo específico propunha a análise da relação entre os erros encontrados nos resumos acadêmicos escritos em português brasileiro e aqueles presentes nos resumos traduzidos automaticamente para o inglês. Nesse ponto, a análise dos dados mostrou que, apesar de diminuir os erros ortográficos, gramaticais e lexicais, o tradutor automático tem tendência a elevar a frequência dos erros semânticos e discursivos relacionados ao uso do inglês acadêmico, como os colocacionais e de registro.

O segundo objetivo específico voltava-se à relação entre os textos gerados pelo DeepL e os textos autopós-editados. A análise dos erros encontrados indicou que a autopós-edição resultou em um aumento de todas as categorias de erros, exceto Discurso.

O terceiro objetivo específico, por sua vez, destinava-se à observar a variação dos erros em cada um dos modos de produção conforme a área do conhecimento a que seus autores pertenciam (CET, CH ou CSA). A esse respeito, os resultados evidenciaram o mesmo que as análises anteriores quanto à influência do sistema de tradução automática e dos doutorandos no crescimento dos erros.

Finalmente, o quarto objetivo específico tinha por alvo a variação dos erros presentes nos resumos escritos em L1, traduzidos automaticamente e autopós-editados conforme a ordem em que foram produzidos. Nesse aspecto, o que se observou foi que a posição da EL2

em relação às tarefas de EL1 e PE L2 parece diretamente relacionada à qualidade dos resumos em português brasileiro e das autopós-edições.

Os resultados, entretanto, devem ser analisados à luz das limitações da pesquisa. Primeiramente, além de o corpus ter dimensões reduzidas, a variação de tamanho entre os resumos elevou o desvio padrão. Assim, não há uniformidade na distribuição de *tokens* e de erros por resumo. Essa limitação, como foi colocado na discussão dos resultados, adveio do próprio desenho experimental de Nunes (2021) e, para que as suas consequências fossem reduzidas, optamos por trabalhar com todos os dados normalizados, de forma a tornar as comparações mais justas. No entanto, qualquer generalização a partir dos nossos resultados seria equivocada.

Além disso, as conclusões aqui obtidas devem ser consideradas dentro do contexto de produção dos resumos, que era empírico-experimental e, portanto, apesar de simular uma situação real de escrita e trabalhar para reduzir os fatores estressantes, não corresponde a uma em tempo, espaço ou propósito. Ademais, os fatos de o nível de inglês ser autodeclarado e de haver uma margem de 100 tokens nas instruções de produção, adaptações que precisaram ser feitas para viabilizar as coletas, também devem ser levados em consideração.

Outra limitação diz respeito a qualquer análise de erros com base em anotação manual: humanos estão suscetíveis a falhas. Por mais que se tenha recorrido ao auxílio do Chat GPT e de pares, o erro durante a anotação das ocorrências é pressuposto e esperado. Ademais, anotações manuais estão sujeitas ao julgamento dos anotadores e, por isso, o ideal é que os textos sejam anotados por mais de uma pessoa, sem que uma tenha acesso à anotação da outra. Contudo, nesta dissertação, devido a restrições de tempo e recursos humanos, todo o processo de anotação teve de ser realizado por uma única pessoa. Assim, os dois revisores a conferirem as marcações já estavam, obrigatoriamente, viesados pelo olhar da anotadora.

Para diminuir o prejuízo gerado por esse aspecto, adotou-se a taxonomia baseada em categorias gerais. Ela foi, então, adaptada de modo a acrescentar o mínimo de categorias possível sem que, com isso, houvesse um prejuízo grande das informações obtidas. Todas as etiquetas foram detalhadas, de forma a se explicitar ao que atendiam. Essas informações foram armazenadas em planilha e entregue aos dois revisores com antecedência. Conforme discutido na Revisão da Literatura e na Metodologia, todas essas foram estratégias para validar o processo de anotação manual de erros.

As limitações da pesquisa, contudo, não invalidam um de seus maiores acertos: a partir de seus achados, abrir caminhos para que outras pesquisas se desenvolvam. Com base nas lacunas que ainda deixa, abre espaço para que outros pesquisadores anotem esses erros, de

forma a validar a taxonomia aqui adaptada e/ou a sua aplicação nos presentes textos. A validação da adaptação da taxonomia pode ser feita, ainda, anotando-se os resumos acadêmicos em EL2 ou mesmo outros textos do par português brasileiro - inglês.

Já com base nos resultados aqui obtidos, uma possibilidade, por exemplo, seria, adotando-se uma metodologia quali-quantitativa, estudar a quantidade e os tipos de erros encontrados na sua correlação com as respostas fornecidas pelos doutorandos em seus questionários e relatos retrospectivos, sejam os pré ou pós-tarefa. Outro caminho viável seria analisar categorias específicas, verificando as suas ocorrências nos diferentes textos; o foco poderia, mesmo, ser naquelas que representaram mais dificuldade para o tradutor automático e /ou os doutorandos. Uma quarta possibilidade é analisar os erros encontrados na comparação com aqueles presentes nos resumos em EL2 para todas as áreas e ordens, respondendo às perguntas que ficaram em aberto sobre o impacto dessa tarefa nas demais.

Ademais, considerando que, conforme ressaltam Kirchoff *et al.* (2012) e Federico *et al.* (2014), não existe uma relação direta entre a frequência de um erro e a má-qualidade do texto que o contém, um quinto caminho possível seria averiguar a relação entre os tipos de erros identificados no corpus de resumos acadêmicos e a percepção de qualidade por parte de pesquisadores em anos mais avançados de carreira. Farrús *et al.* (2010), por exemplo, em seu trabalho com o par inglês-catalão, apontaram os erros lexicais e semânticos como aqueles que mais têm influência nessa percepção. No entanto, os próprios autores reconheceram a necessidade de expandir a investigação para outros pares linguísticos.

O segundo mérito desta dissertação foi, ao investigar a produção textual com auxílio de sistemas de tradução automática, fornecer informações que podem contribuir para a aplicação efetiva da autopós-edição no contexto da produção textual acadêmica de pós-graduandos e da tradução especializada. Conforme ressaltado na introdução, os erros, humanos e de tradutores automáticos, permitem distinguir aspectos linguísticos que se impõem como dificuldades. Assim, tornam-se tópicos a serem abordados em oficinas e cursos de autopós-edição, os quais podem ser destinados tanto aos pesquisadores acadêmicos em geral quanto aos tradutores em formação.

Isso pode conduzir, por exemplo, a doutorandos que, realizando pré e pós-edições mais eficientes, obtenham textos de qualidade cada vez superior; e a tradutores capacitados para lidar com textos especializados de distintas áreas do conhecimento. Em última instância, uma formação adequada em autopós-edição pode levar a tradutores com um leque mais amplo de conhecimentos e a uma maior inclusividade da comunidade acadêmica. Ademais, como contribuição adicional, a pesquisa expandiu a taxonomia de Costa *et al.* (2015) que, além de

ter sido adaptada à tarefa em mãos, foi aplicada na análise de erros humanos e de tradutores automáticos simultaneamente, testando um dos propósitos com os quais foi inicialmente concebida.

Em suma, o presente estudo trouxe resultados aplicáveis ao Ensino e Aprendizagem de Línguas e aos Estudos da Tradução, ao propor uma análise de erros de tradução e pós-edição no par inglês/português brasileiro, ainda carente de estudos semelhantes, e, com isso, oferecer mais informações sobre a influência dos tradutores automáticos na produção de textos acadêmicos por estudantes universitários. Conclui-se, assim, que o estudo cumpriu o seu objetivo não só enquanto pesquisa acadêmica, mas enquanto parte orgânica do todo que é a universidade.

REFERÊNCIAS

- ALLEN, J. Post-editing. In: SOMERS, Harold. (Ed.). *Computers and Translation: a translator's guide*. Amsterdã-Filadélfia: John Benjamins, 2003. p. 297-318.
- ALTBACH, P. G.; KNIGHT, J. The Internationalization of Higher Education: Motivations and Realities. *Journal of Studies in International Education*, v. 11, n. 3-4, p. 290–305, set. 2007.
- ALTMAN, S. (2022). ChatGPT launched on Wednesday. Today it crossed 1 million users![tweet]. Twitter. Disponível em: <https://twitter.com/sama/status/1599668808285028353>. Acesso em 20 de agosto de 2024.
- ALVES, F. A triangulação como opção metodológica em pesquisas empírico-experimentais em tradução. In: PAGANO, Adriana Silvino (org.). *Metodologias de pesquisa em tradução*. Série Estudos Linguísticos, v. 3. Belo Horizonte: FALE-UFMG, 2001. p. 69-92.
- ALVES, F. Cognitive load and potential cost of ELF in academic contexts. Projeto de pesquisa (Pós-graduação em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2018-2022.
- ALVES, F. *et al.* Looking into translation process data: academic writing and post-editing. 2019a. (Apresentação de Trabalho/Simpósio).
- ALVES, F. *et al.* Cognitive load of academic writing in L2 English vs. in L1 plus translation. 2019b. (Apresentação de Trabalho/Comunicação).
- ARENAS-CASTRO, H. *et al.* Academic publishing requires linguistically inclusive policies. *Proceedings - Royal Society. Biological sciences/Proceedings - Royal Society. Biological Sciences*, v. 291, n. 2018, 13 mar. 2024.
- ARRUDA, C. F. B. *et al.* Internacionalização de instituições de ensino superior no Brasil e na América do Norte: implicações e desafios para o ensino de idiomas. In: GÓES-NETO, A.; SALIBA, A. T.; LOPES, D. B. (Org.). *América do Norte*. Belo Horizonte: Editora UFMG, 2021. p. 173–200.
- AVELAR, L. H. D. *Esforço cognitivo e esforço emocional na escrita e pós-edição de resumos acadêmicos: uma análise exploratória*. 2023. Monografia (Graduação em Letras) - Faculdade de Letras, Universidade Federal de Minas Gerais – Belo Horizonte, 2023.
- BECK, K. Globalization/s: Reproduction and Resistance in the Internationalization of Higher Education. *Canadian Journal of Education/Revue canadienne de l'éducation*, v. 35, n. 3, p. 133-148, 2012. Disponível em: <http://journals.sfu.ca/cje/index.php/cje-rce/article/view/1077>. Acesso em: 2 fev. 2024.
- BERBER-SARDINHA, T. *Linguística de Corpus*. Barueri, SP: Manole, 2004.
- BIBER, D. *et al.* *Longman grammar of spoken and written English*. 2000.
- BLAIN, F. *et al.* Qualitative analysis of post-editing for high quality machine translation. In: *Proceedings of Machine Translation Summit XIII: Papers*. 2011.

BOJAR, O. Analyzing error types in English-Czech machine translation. *The Prague Bulletin of Mathematical Linguistics*, v. 95, p. 63, 2011. APA

BOWKER, L.; CIRO, J. B. *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Bingley, Reino Unido: Emerald Publishing Limited, 2019.

BRASIL. Presidência da República. Decreto n.º 7.642, de 13 de dezembro de 2011. Institui o Programa Ciência sem Fronteiras. *Diário Oficial da União*, 13 dez. 2011. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/decreto/d7642.htm#:~:text=DECRETO%20N%207.642%2C%20DE%2013,o%20Programa%20Ciência%20sem%20Fronteiras>. Acesso em: 29 nov. 2024.

BRASIL. Ministério da Educação. *Plano Nacional de Educação 2014-2024: Lei n.º 13.005*, de 25 de junho de 2014. Disponível em: <https://pne.mec.gov.br/18-planos-subnacionais-de-educacao/543-plano-nacional-de-educacao-lei-n-13-005-2014>. Acesso em: 29 nov. 2024.

BRASIL. Ministério da Educação. *Portaria n.º 973, de 14 de novembro de 2014. Regulamenta o Programa Idiomas sem Fronteiras*. Disponível em: https://isf.mec.gov.br/images/pdf/novembro/Portaria_973_Idiomas_sem_Fronteiras.pdf. Acesso em: 29 nov. 2024.

CONEGLIAN, A. V. L. O modelo das dependências universais: assentando bases teóricas e revisando diretrizes metodológicas. *Revista da Abralin*, v. 22, n. 2, p. 1-26, 2023.

COSTA, A. *et al.* A linguistically motivated taxonomy for Machine Translation error analysis. *Machine Translation*, v. 29, p. 127-161, 2015.

COSTA, F. de A. *et al.* A Human-machine Cooperation Protocol for Machine Translation Output Edit Annotation. *Tradumàtica: tecnologies de la traducció*, v. 19, n. 19, p. 148-170, 31 dez. 2021. Acesso em: 10 maio 2022.

DOMBEK, M. *QT LaunchPad Project on Quality Translation: A Survey*. Disponível em: <<https://www.iatis.org/index.php/news/other-news/item/610-qt-launchpad-project-on-quality-translation-a-survey>>. Acesso em: 5 jan. 2025.

DULAY, H.; BURT, M.; KRASHEN, S. *Language two*. Oxford University Press, 200 Madison Ave., New York, NY 10016, 1982.

DUTRA, D. P. *et al.* Institutional internationalisation through academic literacies in English: teaching and learning written genres in the Brazilian higher education context. *The Specialist*, v. 40, n. 2, 10 dez. 2019. Acesso em: 20 jan. 2022.

ELLIOTT, D.; HARTLEY, A.; ATWELL, E.. A fluency error categorization scheme to guide automated machine translation evaluation. In: *Machine Translation: From Real Users to Research: 6th Conference of the Association for Machine Translation in the Americas*, AMTA 2004, Washington, DC, USA, September 28-October 2, 2004. Proceedings 6. Springer Berlin Heidelberg, 2004. p. 64-73.

- ESCARTÍN, C. P. *et al.* Machine translation as an academic writing aid for medical practitioners. In: *Proceedings of Machine Translation Summit XVI: Research Track*. 2017. p. 254-267.
- EUROPEAN ASSOCIATION FOR MACHINE TRANSLATION. *What is Machine Translation?* Disponível em: <<https://eamt.org/what-is-machine-translation/>>. Acesso em: 3 jan. 2024.
- FARRÚS CABECERAN, M. *et al.* Linguistic-based evaluation criteria to identify statistical machine translation errors. In: *14th Annual Conference of the European Association for Machine Translation*. 2010. p. 167-173.
- FEDERICO, M. *et al.* Assessing the impact of translation errors on machine translation quality with mixed-effects models. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014. p. 1643-1653.
- FIRTH, A.. The discursive accomplishment of normality: On “lingua franca” English and conversation analysis. *Journal of Pragmatics*, v. 26, n. 2, p. 237-259, 1996.
- FLANAGAN, M. Error Classification for MT Evaluation. In: *AMTA*. 1994.
- FLOWERDEW, J. Some thoughts on English for Research Publication Purposes (ERPP) and related issues. *Language Teaching*, v. 48, n. 2, p. 250-262, 2015. DOI: 10.1017/S0261444812000523.
- FONSECA, N. B. de L.; NUNES, L. P.; VIEIRA, M. R. Esforço temporal em processos de escrita e de pós-edição acadêmicas na área de educação. *Cadernos de Tradução*, v. 44, n. 1, p. e96629, 2024.
- FRANKENBERG-GARCIA, A. Investigating the collocations available to EAP writers. *Journal of English for Academic Purposes*, v. 35, p. 93-104, 2018.
- GARCIA, I.; PENA, M. I. Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, v. 24, n. 5, p. 471-487, 2011. DOI: 10.1080/09588221.2011.582687.
- GIRARDI, C. *et al.* Mt-equal: a toolkit for human assessment of machine translation output. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*. 2014. p. 120-123.
- HALLIDAY, M. A. K.; MATTHIESSEN, C. *Halliday's introduction to functional grammar*. Milton Park, Abingdon, Oxon: Routledge, 2014.
- HEINE, B.; NARROG, H.; BIBER, D. Corpus-Based and Corpus-Driven Analyses of Language Variation and Use. In: HEINE, B.; NARROG, H. (ed.). *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 2015.
- HOLMES, J. S. The name and nature of translation studies. Translated. *Papers on literary translation and translation studies*, v. 2, p. 67-80, 1988.
- HOUAISS, A. *Minidicionário Houaiss da língua portuguesa*. 5. ed. São Paulo: Editora Moderna, 2019. 956 p.

- HU, K. (2023). *ChatGPT sets record for fastest-growing user base - analyst note*. Reuters. Disponível em: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>. Acesso em 25 de setembro de 2024.
- HUTCHINS, J. Two precursors of machine translation: Artsrouni and Trojanskij. *International Journal of Translation*, v. 16, n. 1, p. 11-31, 2004.
- HUTCHINS, J.. Machine translation: A concise history. *Computer aided translation: Theory and practice*, v. 13, n. 29-70, 2007.
- JAKOBSEN, A. L. Research methods in translation—Translog. In: *Computer Key-Stroke Logging and Writing*. Brill, 2006. p. 95-105.
- JAMES, C. *Errors in language learning and use: Exploring error analysis*. Routledge, 2013.
- KIRCHHOFF, K.; CAPURRO, D.; TURNER, A.. Evaluating user preferences in machine translation using conjoint analysis. In: *Proceedings of the 16th Annual conference of the European Association for Machine Translation*. 2012. p. 119-126.
- KLIE, J-C *et al.* The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, 2018, p. 5-9.
- KNIGHT, J. Internationalisation: Management Strategies and Issues. *International Education Magazine*, CBIE, Ottawa, v. 9, n. 1, p. 191-203, 1993.
- KNIGHT, J. *Internationalization: Elements and Checkpoints*. Ottawa, Canada: Canadian Bureau for International Education, 1994. (Research Monograph, 7).
- KNIGHT, J. *Higher education in turmoil: The changing world of internationalization*. Brill, 2008.
- KOEHN, P. The State of Neural Machine Translation (NMT). *Omniscien Technologies*, c2022. Disponível em: <<https://omniscien.com/blog/state-neural-machine-translation-nmt>>. Acesso em 07 nov. 2022.
- KOGLIN, A. Uma investigação empírica do esforço cognitivo despendido para pós-editar metáforas traduzidas automaticamente em comparação com a tradução manual de metáforas. Tradução de Willian Henrique Cândido Moura e João Gabriel Pereira da Silveira. *Caletrosópio*, Mariana (MG), v. 8, número especial sobre tradução, p. 53-74, 2020.
- KOPONEN, M. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, v. 25, n. 2, p. 131-148, 2016.
- KRINGS, H. P. *Repairing texts: Empirical investigations of machine translation post-editing processes*. Kent State University Press, 2001.
- LENHARO, M. Scientists who don't speak fluent English get little help from journals, study finds. *Nature*, v. 620, n. 7976, p. 931-931, 2023.

- LLITJÓS, A. F.; CARBONELL, J. G.; LAVIE, A. A framework for interactive and automatic refinement of transfer-based machine translation. In: *Proceedings of the 10th EAMT Conference: Practical applications of machine translation*. 2005.
- LOMMEL, A. *et al.* Using a new analytic measure for the annotation and analysis of MT errors on real data. In: *Proceedings of the 17th Annual conference of the European Association for Machine Translation*. 2014. p. 165-172.
- LUFT, C. P. *Dicionário prático de regência nominal*. São Paulo: Editora Ática Didáticos, 2013. 550 p. ISBN 978-8508127641.
- LUFT, C. P. *Dicionário prático de regência verbal*. São Paulo: Editora Ática, 2010. 544 p. ISBN 978-8508127641.
- MATTHIESSEN, C.; TERUYA, K.; WU, C. Multilingual studies as a multi-dimensional space of interconnected language studies. In: WEBSTER, J. (Ed.). *Meaning in Context: implementing intelligent applications of language studies*. London and New York: Continuum, 2008.
- MESA-LAO, B. Introduction to post-editing–The CasMaCat GUI. *Retrieved March*, v. 1, n. 2015, p. 305-314, 2013.
- MONTGOMERY, S. L. *Does science need a global language?: English and the future of research*. University of Chicago Press, 2013.
- MUNDAY, J. *Introducing Translation Studies: Theories and Applications*. 4. ed. London: Routledge, 2016.
- NATIONAL GEOGRAPHIC SOCIETY. *Globalization*. Disponível em: <<https://education.nationalgeographic.org/resource/globalization/>>. Acesso em: 29 nov. 2024.
- NEVES, M. H. de M. *Gramática de usos do português*. Unesp, 2000.
- NIÑO, A. Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, v. 21, n. 1, p. 29-49, 2008.
- NUNES, L. P. *A produção textual acadêmica em distintos domínios: investigando a escrita e a pós-edição em língua inglesa sob a perspectiva linguística e cognitiva*. Projeto de pesquisa (Pós-graduação em Estudos Linguísticos). 29 p. Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2021.
- O'BRIEN, S.; SIMARD, M.; GOULET, M-J. Machine translation and self-post-editing for academic writing support: Quality explorations. *Translation quality assessment: From principles to practice*, p. 237-262, 2018.
- O'BRIEN, S. Teaching post-editing: a proposal for course content. In: *Proceedings of the 6th EAMT workshop: Teaching machine translation*. 2002.
- ORFANÒ, B. Malveira; OLIVEIRA, A. L. A. M.; DA SILVA, S. B.. The role of pragmatic markers in academic spoken interlanguage: A corpus-based study of a group of Brazilian EFL university students. *Diacrítica*, v. 32, n. 3, p. 207-226, 2018.

- PAGANO, A. STV em Linguística Aplicada Abordagens multidisciplinares da tradução. Apresentação em PowerPoint. 2023. 29 slides. *Primeira aula de Abordagens Multidisciplinares da Tradução*. Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 3 de abril de 2023.
- POPOVIĆ, M.; NEY, H. Error analysis of verb inflections in spanish translation output. In: *TCSTAR Workshop on Speech-to-Speech Translation*. 2006. p. 99-103.
- POPOVIĆ, M.; NEY, H. Towards automatic error analysis of machine translation output. *Computational Linguistics*, v. 37, n. 4, p. 657-688, 2011.
- POPOVIĆ, M. Error Classification and Analysis for Machine Translation Quality Assessment. *Machine Translation: Technologies and Applications*, p. 129–158, 2018.
- PYM, A. Exploring Translation Theories. *Cadernos de Tradução*. Tradução de Eduardo César Godarth, Yéo N'gana e Bernardo Sant'Anna. , v. 36, n. 3, p. 214–268, 2016.
- RICHARDS, J. C. Second language acquisition: Error analysis. *Annual Review of Applied Linguistics*, v. 1, p. 91-107, 1980.
- SNOVER, M. *et al.* A study of translation edit rate with targeted human annotation. In: *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*. 2006. p. 223-231.
- STYMNE, S.; AHRENBERG, L. On the practice of error analysis for machine translation evaluation. In: *LREC*. 2012. p. 1785-1790.
- TERRA, L. de F. *O uso de tecnologias digitais para o desenvolvimento de habilidades orais em inglês: um estudo de caso no contexto acadêmico*. 2019. 113 f. Dissertação (Mestrado em Linguística Aplicada) – Universidade Federal de Minas Gerais, Belo Horizonte, 2019.
- VAN DIJK, T. A. Academic nationalism. *Discourse & Society*, v. 5, n. 3, p. 275-276, 1994.
- VASCONCELLOS, M; LEÓN, M. SPANAM e ENGSPAN: machine translation at the Pan American Health Organization. *Computational Linguistics*, v. 11, n. 2-3, p. 122-136, 1985.
- VIEIRA, E. R. P.. Some remarks on comparative stylistics applied to translation from English into Portuguese. *Estudos Germânicos*, v. 3, n. 1, p. 147-161, 1982.
- VIEIRA, M. R. *Investigando o esforço temporal em tarefas de escrita e de pós-edição acadêmica em domínio específico: um estudo exploratório*. 2022. Monografia (Graduação em Letras) - Faculdade de Letras, Universidade Federal de Minas Gerais – Belo Horizonte, 2022.
- VILAR, D. *et al.* Error Analysis of Statistical Machine Translation Output. In: *LREC*. 2006. p. 697-702.
- WANG, H. *et al.* Progress in machine translation. *Engineering*, v. 18, p. 143-153, 2022.
- ZHANG, J.; ZONG, C.. Neural machine translation: Challenges, progress and future. *Science China Technological Sciences*, v. 63, n. 10, p. 2028-2050, 2020.

APÊNDICE A - Taxonomia de Costa *et al.* (2015) com exemplos dos autores

Quadro 15 - Taxonomia de Costa *et al.* (2015) com exemplos dos autores

Categoria Subcategoria		Exemplo				
		ING	PT-EUR (incorreto)	PT-EUR(correto)		
Ort.	Pontuação		green tea	chá*, verde	chá verde	
	Letra maiúscula		... on time, I can console myself...	... a tempo, *Posso consolar-me...	... a tempo, posso consolar-me...	
	Digitação		Basilica of the Martyrs	Basilica dos *Mátires	Basilica dos Mártires	
Léx.	Omissão	Palavra lexical	In his inaugural address, Barack Obama	No seu * inaugural, Barack Obama	No seu discurso inaugural, Barack Obama	
		Palavra gramatical	In India	Em Índia	Na Índia	
	Adição	Palavra lexical	This time I'm not going to miss	Desta vez *correr não vou perder	Desta vez não vou perder	
		Palavra gramatical	highlights the work	*Já destaca-se o trabalho	destaca-se o trabalho	
	Não traduzido		in the world of botany	no mundo da *botany	no mundo da botânica	
Gram.	Escolha incorreta	Classe de palavras		world	mundial (worldwide)	mundo
		Verbos	Tempo/ Modo	Even though this is a long list	Mesmo que esta *é uma longa lista	Mesmo que esta seja uma longa lista
			Pessoa/ Número	Theater-goers can discover	As pessoas que vão ao teatro *pode descobrir	As pessoas que vão ao teatro podem descobrir
			Ambos (<i>blend</i>)	If I go to see the Dario Fo play	Se *vai ver a peça de Dário Fo	Se eu for ver a peça de Dário Fo
		Concordância (nome)	Gênero	The German artist Thomas Schutte	*(A artista alemã) Thomas Schutte	O artista alemão Thomas Schutte
			Número	moral skills	capacidades *moral	capacidades morais
			Pessoa	learn from our failures	aprender com os *vossos fracassos	aprender com os nossos fracassos
			Mais de um (<i>blend</i>)	funky clothes shops	lojas *simpático de roupa	lojas simpáticas de roupa
		Contração		in an environment	em um ambiente	num ambiente
		Ordem incorreta		A person is wise.	Uma pessoa sábia *é.	Uma pessoa é sábia.

Fonte: adaptado e traduzido de Costa *et al.* (2015)

Quadro 15 - Taxonomia de Costa *et al.* (2015) com exemplos dos autores

Categoria Subcategoria		Exemplo		
		ING	PT-EUR (incorreto)	PT-EUR(correto)
Sem.	Sentido incorreto	the authentic tea set that includes a tray, teapot and glasses	um autêntico jogo de chá que inclui bandeja, bule e *óculos	um autêntico jogo de chá que inclui bandeja, bule e copos
	Palavra incorreta	in the same quarter	no mesmo *histórica	no mesmo bairro
	Colocação incorreta	high wind	vento alto	vento forte
	Expressões idiomáticas	kick the bucket	dar um pontapé ao balde	esticou o pernil
Disc.	Estilística	permission to be allowed to improvise	autorização para ser autorizado a improvisar	permissão para ser autorizado a improvisar
	Dialeto	in his speech	em seu discurso	no seu discurs
	Não deveria traduzir	Havia um Menino que era Pessoa	Havia hum Menino era Opaco Pessoa	Havia um Menino que era Pessoa

Fonte: adaptado e traduzido de Costa *et al.* (2015)