

COMPARAÇÃO DA COMPLEXIDADE LINGUÍSTICA EM TRADUÇÕES HUMANAS E AUTOMÁTICAS: UM ESTUDO DE CASO COM CONTOS DE OSCAR WILDE

Maria Luiza Martins (PIC/UEM), André Luis Antonelli (Orientador). E-mail: alantonelli@uem.br.

Universidade Estadual de Maringá, Centro de Ciências Humanas, Letras e Artes, Maringá, PR.

Área e subárea do conhecimento: Linguística, Letras e Artes/Teoria e Análise Linguística).

Palavras-chave: Complexidade Linguística; Tradução Automática; Métricas Avaliativas.

RESUMO

O projeto de pesquisa busca comparar a complexidade linguística entre traduções realizadas por humanos e pelo modelo de inteligência artificial ChatGPT. Mais especificamente, esse trabalho tem como objetivo principal identificar e analisar as distinções linguísticas entre as traduções humanas e automáticas, focando em aspectos como estrutura sintática, coesão textual e uso de vocabulário. O material de análise é composto por vinte traduções de contos do autor britânico Oscar Wilde: dez traduções realizadas por humanos e dez traduções feitas pelo ChatGPT. Todas as traduções foram submetidas à ferramenta Nilc-Metrix, que analisa a coesão e complexidade textual para a língua portuguesa. Essa ferramenta, desenvolvida por pesquisadores da Universidade de São Paulo, permite a análise detalhada de fatores como densidade lexical e conectividade referencial. A justificativa para este estudo reside na crescente utilização de ferramentas como o ChatGPT, que, embora mostrem fluência e rapidez, ainda enfrentam desafios na manutenção da coesão e precisão em textos de alta complexidade. A pesquisa contribuirá para um entendimento mais profundo das capacidades e limitações da tradução automática em relação às traduções humanas, fornecendo eventuais insights sobre como essas ferramentas podem ser aprimoradas.

INTRODUÇÃO

Nos últimos anos, a utilização de ferramentas de inteligência artificial, como o ChatGPT, tem se expandido rapidamente no domínio da tradução automática. Esses sistemas, fundamentados em modelos de linguagem de larga escala, possuem a capacidade de processar e traduzir textos de maneira eficiente, permitindo a produção em massa e de forma ágil, oferecendo uma solução prática para indivíduos e organizações que demandam traduções imediatas. A acessibilidade dessas tecnologias, aliada à melhoria contínua na qualidade das traduções, tem

impulsionado sua adoção crescente, posicionando-as como uma alternativa viável às traduções realizadas por tradutores humanos (BAKER, 2018). Embora essas tecnologias tenham evoluído substancialmente, existem distinções importantes quando comparadas às traduções realizadas por humanos. Tradutores humanos, com sua capacidade de captar nuances culturais, expressões idiomáticas e contextos específicos, frequentemente produzem traduções mais precisas e adequadas ao público-alvo (NEWMARK, 1988). Em contraste, modelos de inteligência artificial, como o ChatGPT, demonstram grande eficiência em termos de fluência e rapidez, mas ainda enfrentam desafios significativos ao lidar com estruturas sintáticas complexas e na manutenção da coesão textual (WU et al., 2016). Um ponto fundamental nessa comparação é a complexidade linguística, na qual as abordagens humanas e automáticas podem divergir significativamente (HAWKINS, 2004).

Compreender as diferenças práticas entre essas duas abordagens de tradução é crucial, especialmente em contextos nos quais a complexidade linguística do texto traduzido pode influenciar diretamente a clareza e a compreensão pelo leitor.

MATERIAIS E MÉTODOS

Corpus

Para esse projeto, selecionamos dez contos do autor britânico Oscar Wilde, listados a seguir:

The devoted friend; The remarkable rocket; The sphinx without a secret; The model millionaire; The young king; The birthday of the infant; The fisherman and his soul; The star-child; Lord Arthur Savile's crime – chapter I; The Canterville ghost – chapter I.

Todos esses contos estão disponíveis na edição bilíngue português/inglês *Contos Completos de Oscar Wilde*, publicada pela editora Landmark em 2012, e com tradução de Luciana Salgado. As respectivas traduções em português que utilizaremos são as que aparecem nessa mesma edição citada. As traduções automáticas foram feitas pelo ChatGPT a partir dos dez contos originais.

Métricas avaliativas

Em termos de métricas avaliativas, utilizamos a ferramenta Nilc-Metrix, desenvolvida pelo Núcleo Interinstitucional de Linguística Computacional (NILC) da Universidade de São Paulo (<http://fw.nilc.icmc.usp.br:23380/nilcmetrix>). Trata-se de uma ferramenta computacional projetada para analisar a coesão e a complexidade dos textos com aplicação especificamente para a língua portuguesa. A ferramenta oferece uma ampla gama de métricas que avaliam diferentes aspectos linguísticos,

como conectividade referencial, estrutura sintática, densidade lexical e características discursivas.

Uma vez submetidos à análise os dois grupos de traduções, fizemos uma análise comparativa dos resultados, buscando avaliar, a partir do conjunto de métricas selecionadas, qual tipo de tradução apresentava um maior nível de complexidade.

RESULTADOS E DISCUSSÃO

Dentre as dez métricas selecionadas, seis apresentaram um nível maior de complexidade nas traduções humanas, enquanto quatro demonstraram maior complexidade nas traduções realizadas pelo ChatGPT. A título de ilustração, observamos que a tradução humana, em comparação às traduções da ferramenta de IA, usou palavras mais complexas que destacam as características textuais e estilísticas. Em termos de processamento linguístico, isso gera uma maior dificuldade, o que evidencia um maior nível de complexidade das traduções humanas nesse quesito. O menor nível de complexidade do ChatGPT aqui talvez se explique pelo fato de que essa ferramenta, ainda que alimentada com uma grande quantidade de dados, limita sua capacidade de geração linguística a partir de modelos que lhe serviram de base. Ou seja, tem-se aí uma restrição ao processo criativo, já que a ferramenta se comporta dentro de um padrão estatístico alinhado aos dados de treinamento. Os textos humanos, por sua vez, experimentam uma maior liberdade já que a criatividade linguística é um dos traços definidores da linguagem humana (uso infinito a partir de recursos finitos). O resultado disso é um nível de complexidade maior.

CONCLUSÕES

Os resultados deste estudo evidenciam que, embora as traduções automáticas apresentem fluência e rapidez, ainda se distinguem significativamente das traduções humanas quanto à complexidade linguística. A análise mostrou que as produções humanas tendem a explorar, por exemplo, maior diversidade lexical, refletindo escolhas criativas que tornam os textos mais densos e ricos em nuances. Já o ChatGPT, apesar de eficiente, mantém-se mais próximo de padrões estatísticos, o que limita sua variação e reduz a complexidade em certos aspectos. Assim, torna-se claro que a tradução automática pode ser uma ferramenta útil em contextos que exigem praticidade e agilidade, mas não substitui a profundidade interpretativa e a flexibilidade criativa próprias do trabalho humano. Em última instância, compreender essas diferenças é fundamental para orientar o uso adequado de cada abordagem, especialmente em situações nas quais a qualidade e a sofisticação textual são determinantes para a compreensão do leitor.

REFERÊNCIAS

BAKER, M. In other words: **A coursebook on translation**. Abingdon: Routledge, 2018.

HAWKINS, J. A. **Efficiency and complexity in grammars**. New York: Oxford University Press, 2004.

NEWMARK, P. **A textbook of translation**. Hertfordshire: Prentice Hall, 1988.

WILDE, O. **Contos completos de Oscar Wilde**. Tradução de Luciana Salgado. São Paulo: Editora Landmark, 2012.

WU, Y. et al. **Google's neural machine translation system: Bridging the gap between human and machine translation**. arXiv preprint arXiv:1609.08144, 2016.