



10° EBBC

Encontro Brasileiro de
Bibliometria e Cientometria



A ESCRITA ACADÊMICA MUDOU? uma análise pré e pós inteligência artificial

Natasha Duarte Amarante

<https://orcid.org/0000-0003-3837-6583>
natasha.damarante@gmail.com
Universidade Federal do Rio Grande do Sul (UFRGS)

Juliana Cristina Sanches

<https://orcid.org/0000-0002-1888-1593>
juliana.sanches@ufrgs.br
Universidade Federal do Rio Grande do Sul (UFRGS)

Diego Tonello

<https://orcid.org/0000-0002-8351-9566>
diego-tonello@hotmail.com
Universidade Federal do Rio Grande do Sul (UFRGS)

Rene Faustino Gabriel Junior

<https://orcid.org/0000-0003-1021-3360>
rene.gabriel@ufrgs.br
Universidade Federal do Rio Grande do Sul (UFRGS)

Sônia Elisa Caregnato

<https://orcid.org/0000-0002-5676-2763>
sonia.caregnato@ufrgs.br
Universidade Federal do Rio Grande do Sul (UFRGS)

Resumo:

O estudo tem como objetivo investigar a possível influência das IAs na escrita acadêmica a partir de dissertações de mestrado da área de Ciências Sociais Aplicadas da Universidade Federal do Rio Grande do Sul (UFRGS). Quanto a metodologia, apresenta abordagem quantitativa e caráter exploratório, com análise lexical e cálculo da Razão de excesso (r) de palavras entre os períodos pré-LLM (2020-2022) e pós-LLM (2023-2025), com corpus de 917 dissertações da UFRGS. Foram selecionadas 20 palavras de estilo (verbos, advérbios e adjetivos) com maiores valores de (r) presentes em mais de 39,76% das dissertações. Os resultados sugerem aumento na frequência de termos de estilo na redação acadêmica, possivelmente influenciada pelo uso de IA,, destacando-se a palavra *crucial* ($r=4,05$), representando 305% de aumento no uso em relação ao período Pré-LLM. Conclui-se que há indícios de mudança estilística no corpus analisado, em razão da influência dos LLMs na escrita acadêmica, embora não seja possível afirmar causalidade direta. Investigações futuras, com corpus ampliado e análise contextual, poderão refinar essa compreensão.

Palavras-Chave: Escrita acadêmica. Inteligência artificial. Estilo de escrita.

EIXO TEMÁTICO:

Ciência, Tecnologia e Inovação

MODALIDADE:

Resumo expandido

DOI: 10.22477/x.ebbc.1019

COMO CITAR ESTE ARTIGO:

AMARANTE, Natasha Duarte; SANCHES, Juliana Cristina; TONELLO, Diego; GABRIEL JUNIOR, Rene Faustino; CAREGNATO, Sônia Elisa. A escrita acadêmica mudou? Uma análise pré e pós inteligência artificial. In: EBBC - Encontro Brasileiro de Bibliometria e Cientometria. 10. **Anais [...]**. Curitiba, 2026. DOI: 10.22477/x.ebbc.1019. Disponível em: <https://doi.org/10.22477/x.ebbc.1019>.



1 INTRODUÇÃO

Um texto gerado por um tipo de *Large Language Model* (LLM) é indistinguível de um texto produzido por uma pessoa? Se um leitor não for capaz de fazer essa distinção, podemos dizer que o modelo de linguagem terá então atingido o critério de inteligência funcional de Turing (1950).¹ A concretização de alguns dos postulados de Turing (1950) pode ser vista com o *Transformer*, uma arquitetura de rede neural baseada em autoatenção (*Self-Attention*) proposta Vaswani *et al.* (2017), o mecanismo relaciona diferentes posições de uma mesma sequência para computar uma representação precisa do sentido de cada termo conforme as palavras que o cercam, conferindo aos LLMs a capacidade de capturar dependências de longo alcance e nuances contextuais.

Os LLMs são sistemas avançados de inteligência artificial (IA) projetados para compreender, gerar e manipular a linguagem humana (Microsoft do Brasil [...], 2026). Com o lançamento do ChatGPT em novembro de 2022, e posteriormente de outros modelos de IA como o Gemini do Google e o Copilot da Microsoft, tarefas como escrever, resumir e traduzir passaram a ser passíveis de automatização. Diante disso, defrontamo-nos com uma ruptura em determinadas práticas convencionais do fazer acadêmico.

Segundo Kobak *et al.* (2024), o estilo de escrita dos LLMs é caracterizado como “floreado”, com predomínio de verbos e adjetivos. Na análise dos autores, em 2024 cerca de 13,5% dos resumos da base de dados PubMed, foram escritos por IAs generativas. Bao *et al.* (2025) também detectaram a preferência no uso de certos advérbios e adjetivos específicos na língua inglesa após o lançamento do ChatGPT. Um indicativo de mudança no tom e no estilo da escrita acadêmica, ca-

¹ Turing (1950) propôs a ideia de que, se uma máquina pudesse produzir respostas indistinguíveis das de um ser humano, ela teria atingido um patamar de inteligência funcional.

racterizado pela simplificação das estruturas das frases, declínio na sobreposição lexical e de conectivos básicos.

Essas transformações apontam para uma influência profunda e crescente das IAs. Embora essas tecnologias apresentem benefícios, a exemplo da superação de barreiras apontada por Bao *et al.* (2025), a sua utilização pode comprometer a coesão e legibilidade dos textos acadêmicos, além de alterar o estilo de escrita característico de cada campo do conhecimento.

Nesse contexto, o objetivo deste estudo é investigar a possível influência das IAs na escrita acadêmica a partir de dissertações de mestrado da área de Ciências Sociais Aplicadas da Universidade Federal do Rio Grande do Sul (UFRGS). Para atingir o objetivo proposto, executou-se análise lexical e cálculos estatísticos para detecção do aumento de palavras de estilo da língua portuguesa, em dissertações disponíveis em texto completo no repositório institucional Lume. A escolha justifica-se pela curadoria digital mediada por bibliotecários(as), o que garante que os documentos, depositados no Lume recebam metadados qualificados, assegurando a facilidade de recuperação.

Apoiada em dados empíricos, a relevância desta pesquisa está na discussão sobre como essas ferramentas podem estar contribuindo para a descaracterização dos estilos próprios da redação científica e para a ruptura de práticas convencionais de produção textual acadêmica. Além disso, estudos sobre a formação de possíveis traços característicos de “assinatura textual” e padronização de recursos retóricos associados ao uso de IA ainda são pouco explorados na literatura em língua portuguesa, por se tratar de um fenômeno recente.

2 METODOLOGIA

Este estudo adota uma abordagem predominantemente quantitativa, de caráter exploratório, se justifica pela necessidade de combinarmos procedimentos estatísticos de análise lexical com interpretação qualitativa dos padrões identificados. Ela estrutura-se em quatro etapas: coleta e limpeza dos dados; processamento computacional dos textos; análise empírica de vocabulário a partir do cálculo de Razão de excesso; e classificação estilística dos termos identificados.

A coleta de dados foi realizada em 16/12/25 no SABI (sabi.ufrgs.br)², escolhido por disponibilizar registros bibliográficos em formato MARC, possibilitando a recuperação de campos específicos, incluindo link direcionado ao texto completo em formato .pdf do Lume. O levantamento foi conduzido através da pesquisa CCL (*Common Command Language*), integrando operadores booleanos e linguagem de comandos, para seleção de dissertações dos doze programas de mestrado da instituição que pertencem às Ciências Sociais Aplicadas, publicadas entre 2020 a 2025, resultando na seguinte expressão de busca: *(WPG=ADMP or ARQP or CINP or COMP or CONP or DESP or DRDP or DIRP or ECOP or EEIP or MUPP or URBP) and (WYR=2020 or 2021 or 2022 or 2023 or 2024 or 2025) and (WPP=pe)*³.

Os resultados foram consolidados em uma planilha .csv contendo os seguintes metadados de cada dissertação: data, nome do Programa de Pós-Graduação (PPG) e link do texto completo. No tratamento e limpeza dos dados foram excluídas dissertações sem link ao texto completo e datas fora do escopo da pesquisa, resultando em um corpus de 917 arquivos. O processamento dos textos e os cálculos estatísticos foram realizados no ambiente *Jupyter Notebook*, que fornece

² Catálogo online das bibliotecas da UFRGS, que fornece os registros bibliográficos das publicações depositadas no Lume.

³ Comandos da expressão de busca disponíveis em: <https://www.ufrgs.br/documenta/d/anexo-4-campos-recuperacao-informacao/>, <https://www.ufrgs.br/documenta/d/tabela-k-programas-pos-graduacao-ufrgs/>

ferramentas e padrões para computação interativa, sendo os comandos gerados com auxílio da plataforma de inteligência artificial *Adapta One*⁴. Os PDFs foram convertidos em .txt e submetidos a processos de normalização textual, tais como: conversão para letras minúsculas, remoção de pontuação e eliminação de caracteres especiais, além da remoção de *stopwords* (artigos, preposições e pronomes).

Após esse procedimento, a contagem de palavras dos arquivos .txt, resultou na extração de 502.949 termos, acompanhados pela respectiva frequência e ano, organizados em um arquivo .csv. Para a análise empírica dos resultados, trabalhamos com o corpus em dois períodos:

- a. Pré-LLM (2020–2022): utilizado para estabelecer a frequência esperada das palavras, funcionando como base histórica contrafactual, composto por 497 dissertações;
- b. Pós-LLM (2023–2025): utilizado para observar a frequência real dos termos após a popularização do ChatGPT, composto por 420 dissertações.

A partir dessa divisão, calculamos o excesso de vocabulário por meio do indicador estatístico de Razão de Excesso (r), cuja fórmula é definida como a razão entre a frequência observada de um termo no período pós-LLM (p) e sua frequência esperada para esse mesmo período (q), projetada a partir da tendência pré-LLM, representada pela fórmula $r=p/q$, onde (r) significa razão de excesso. Para garantir a estabilidade estatística e evitar divisões por zero, as frequências foram calculadas utilizando a fórmula de suavização $p=(a+1)/(b+1)$, em que (a) corresponde ao número de dissertações em que o termo aparece e (b) corresponde ao total de dissertações consideradas no período analisado, conforme proposto por Kobak *et al.* (2025).

⁴ Comandos e arquivos gerados no Jupyter disponíveis em: <https://cedapdados.ufpr.br/dataset.xhtml?persistentId=doi:10.48472/CEDAP/BOYDFS>

A partir desse cálculo, identificou-se as palavras que passaram a apresentar crescimento abrupto no período pós-LLM. Valores de r maiores que 1 indicam aumento relativo, desta forma, o ponto de corte deste estudo foram as palavras com ($r > 2$). Adicionalmente, foi aplicado o critério de frequência total no período Pós-LLM, considerando as palavras que tiveram frequência superior a 100 ocorrências, totalizando 1.478 palavras. Destas, foram selecionadas manualmente pelos pesquisadores as palavras de estilo: verbos, adjetivos e advérbios de caráter retórico, associados à construção argumentativa do texto, conforme Kobak *et al.* (2025)⁵. A triagem manual seguiu critérios de: (a) função potencial no discurso acadêmico; (b) recorrência em diferentes dissertações; e (c) frequência elevada no corpus. Destaca-se que, devido à escala do corpus e às limitações de tempo e espaço do estudo, não foi realizada a verificação contextual individual de cada ocorrência nos textos originais. A validação da classificação gramatical foi feita através da consulta de cada palavra aos dicionários online: Priberam, Michaelis e Dicio - Dicionário online de português.

Essa triagem restringiu o conjunto a 134 termos, para os quais calculou-se o total de ocorrências e a porcentagem em dissertações diferentes em que cada um ocorre, a fim de evitar que valores elevados de r fossem influenciados por usos concentrados em poucos documentos. Foram selecionados para esta análise os 20 termos de maior abrangência, uma vez que estão presentes em mais de 39,76% das dissertações.

Devido a flexibilidade inerente à língua portuguesa (Bechara, 2019), determinadas palavras podem transitar entre classes gramaticais, como ocorre com substantivos que assumem função adjetiva em contextos específicos. Considerando que

⁵ Segundo Kobak *et al.* (2025), marcadores retóricos correspondem a palavras de estilo (verbos, adjetivos e advérbios) que contribuem para a fluidez, ênfase e legibilidade do texto acadêmico, sem alterar diretamente seu conteúdo temático.

a natureza quantitativa desta etapa inviabiliza uma análise contextual exaustiva de cada ocorrência, a seleção dos termos baseou-se em sua elevada probabilidade de atuação como marcadores retóricos e de coesão textual, considerando os termos com alta dispersão no corpus, reduzindo a influência de usos isolados. O aprofundamento qualitativo dessas ocorrências será desenvolvido em investigações posteriores.

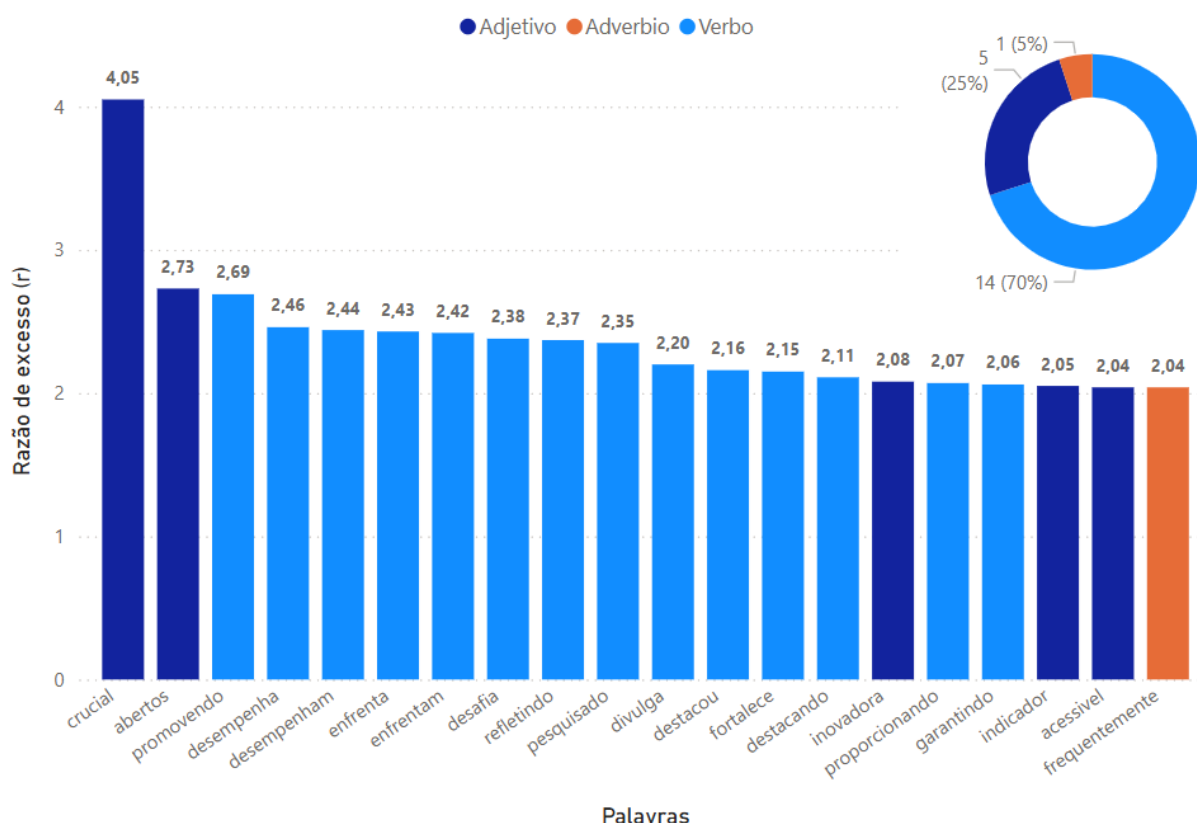
3 RESULTADOS E ANÁLISES

Em relação ao corpus analisado, a análise dos dados sinaliza mudanças no padrão de escrita acadêmica no corpus estudado, entre 2023 e 2025, período marcado pela difusão dos LLMs. Observa-se o aumento da incidência de palavras de estilo, isto é, termos que não introduzem novos conteúdos temáticos, mas atuam na modulação retórica e na formalização do discurso científico. Salienta-se que esses resultados devem ser entendidos como indícios de mudança no vocabulário do corpus analisado, e não como evidência conclusiva de causalidade associada ao uso dessas tecnologias nas dissertações em geral.

O conjunto analisado permitiu identificar tendências linguísticas com maior representatividade dentro do recorte metodológico proposto. Entre os 20 termos selecionados do corpus, verifica-se a predominância dos verbos (70%): *enfrenta, pesquisado, desempenha, divulga, destacando, fortalece, desafia, enfrentam, promovendo, garantindo, refletindo, desempenham, destacou e proporcionando*; seguidos pelos adjetivos (25%): *indicador, crucial, acessível, abertos, inovadora* e advérbios (5%): *frequentemente*. Esse conjunto indica um reforço de estruturas linguísticas associadas à coesão, qualificação e ênfase argumentativa e sugere a intensificação de recursos retóricos que contribuem para a construção de uma escrita mais uniforme e formalizada no período e corpus analisado, conforme pode ser visua-

lizado no Gráfico 1. onde a Razão de excesso (r) representa o quanto o uso da palavra se destaca em excesso no período Pós-LLM em relação ao período Pré-LLM.

Gráfico 1 - Razão de excesso (r) e palavras de estilo do conjunto analisado



Fonte: elaborado pelos próprios autores na plataforma Power BI.

Nota: os dados foram coletados durante o 4º trimestre de 2025.

Dentre as dissertações analisadas, a palavra *crucial* obteve destaque ao apresentar aumento de 305% em relação ao esperado historicamente ($r=4,05$), configurando o principal marcador de excesso entre aqueles com maior dispersão no corpus. Entre os verbos, destacamos os termos *promovendo* ($r=2,69$) com aumento de 169% e *enfrenta* ($r=2,43$) com aumento de 143%, enquanto, no grupo dos advérbios e qualificadores, *frequentemente* registra ($r=2,04$) com aumento de 104% e *inovadora* apresenta crescimento de 108%, ($r=2,08$). Embora superem a Razão de excesso de *crucial*, alguns termos como *estacionárias* ($r=22,51$) e *perfilado* ($r=8,45$)

foram excluídos pela concentração em poucas dissertações, o que sugere especificidade temática, não constituindo uma tendência estilística transversal no corpus.

De acordo com Bechara (2019), a flexibilidade da língua permite a transposição funcional entre classes gramaticais conforme o contexto sintático. No corpus analisado, termos como *indicador* e *abertos*, originalmente substantivos em determinados usos, foram identificados com função adjetiva, qualificando conceitos e objetos de pesquisa (por exemplo, fator indicador e dados abertos). Essa mobilidade reforça seu papel como elementos de coesão e qualificação textual. Formas verbais e participípios, como *pesquisado* ($r=2,35$), foram consideradas marcadores de estilo em razão de seu uso recorrente como qualificadores que conferem autoridade discursiva. Advérbios de frequência e ênfase atuam de modo semelhante, ampliando a densidade retórica sem alterar o conteúdo informacional.

A Razão de excesso (r) observada nos termos analisados corrobora as tendências observadas por Kobak *et al.* (2025) e Bao *et al.* (2025), sugerindo a influência do uso dos LLMs no período de 2023 a 2025 no *corpus* analisado, indicando que eles possuem uma assinatura textual específica. O uso crescente e significativo de certas palavras de estilo aponta para uma intensificação de recursos de ênfase e coesão que, conforme discutido por Bao *et al.* (2025), podem sinalizar uma tendência à padronização e simplificação do discurso acadêmico.

4 CONSIDERAÇÕES FINAIS

A identificação de palavras de estilo é fundamental para a detecção de possíveis assinaturas textuais associadas ao uso de IA, visto que esses termos estruturam a formalização e padronização do discurso (Kobak *et al.*, 2025). Nesse contexto, embora a associação entre o aumento desses termos e o uso de modelos de linguagem deva ser interpretada com cautela, a natureza dos termos avaliados, centra-

dos em palavras de estilo, com ampla dispersão no corpus e aumento consistente entre os períodos pré e pós LLMs, sugerem que o aumento dessas ocorrências pode não estar relacionado ao surgimento de novos tópicos de pesquisa, mas a transformações na forma de comunicar resultados, indicando uma mudança estilística em relação às tendências históricas do vocabulário acadêmico.

Embora o estudo identifique que o uso de determinadas palavras de estilo passou a ocorrer com uma frequência superior no período Pós-LLM, esses achados devem ser compreendidos com cautela, reconhecendo que as transformações lexicais observadas possam ter outras razões, como mudanças temáticas nos programas de Ciências Sociais Aplicadas da UFRGS, variações editoriais, estilos de orientadores ou o impacto do período-pandêmico, entre outras.

Mesmo que não seja possível afirmar causalidade direta, os resultados apontam para uma mudança estilística associada ao período pós-LLM. Diante da complexidade da língua portuguesa, novas investigações apoiadas em um corpus mais extenso e diversificado, juntamente com análises lexicais, amostragens de contexto e de concordância, poderão captar com maior precisão as transformações em curso na escrita acadêmica.

Dessa forma, o trabalho contribui para o debate sobre a emergência de uma possível “assinatura textual” da IA, ao mesmo tempo em que ressalta a necessidade de investigações futuras e amostragens mais amplas, com análises qualitativas para validar a extensão dessas mudanças estilísticas na escrita científica brasileira.

REFERÊNCIAS

BAO, T.; ZHAO, Y.; MAO, J.; ZHANG, C. Examining linguistic shifts in academic writing before and after the launch of ChatGPT: a study on preprint papers. **Scientometrics**, [Cham], v. 130, p. 3597-3627, 2025. DOI: 10.1007/s11192-025-05341-y. Disponível em: <https://doi.org/10.1007/s11192-025-05341-y>. Acesso em: 16 fev. 2026.

BECHARA, E. **Moderna gramática portuguesa**. 39 ed. Rio de Janeiro: Lucerna, 2019.

KOBAC, D.; GONZÁLEZ-MÁRQUEZ, R.; HORVÁT, E.-Á.; LAUSE, J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. **Science Advances**, Washington, v. 11, n. 27, p. 1-8, July 2025. Artigo eadt3813. DOI: 10.1126/sciadv.adt3813. Disponível em: <https://www.science.org/doi/10.1126/sciadv.adt3813>. Acesso em: 10 dez. 2025.

MICROSOFT do Brasil [...]. **O que são os grandes modelos de linguagem (LLMs)?** São Paulo, c2026. Disponível em: <https://azure.microsoft.com/pt-br/resources/cloud-computing-dictionary/what-are-large-language-models-llms>. Acesso em: 20 fev. 2026.

PROJECT JUPYTER. **Jupyter notebook documentation**. c2015. Disponível em: <https://jupyter-notebook.readthedocs.io/en/stable/>. Acesso em 13 jan. 2026.

TURING, A. M. Computing machinery and intelligence. **Mind, New Series**, Oxford, v. 59, n. 236, p. 433-460, Oct. 1950. DOI: <https://doi.org/10.1093/mind/LIX.236.433>. Disponível em: <http://www.jstor.org/stable/2251299>. Acesso em: 10 dez. 2025.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. Attention is all you need. *In*: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 31., 2017, Long Beach, CA. **Proceedings** [...]. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Acesso em: 23 dez. 2025.