

Evolução de Padrões Linguísticos na Escrita Científica em Português: Uma Análise com NILC-Metrix

Thiago Ruiz Lobo

Institute of Computing
Federal University of Mato Grosso
Brazil, 78060-900
thiago.lobo@sou.ufmt.br

Claudia Aparecida Martins

Institute of Computing
Federal University of Mato Grosso
Brazil, 78060-900
claudia@ic.ufmt.br

Resumo

Este trabalho analisa a evolução de padrões linguísticos em resumos de artigos em português da Sociedade Brasileira de Computação entre 2020 e 2025, com base em métricas linguísticas do NILC-Metrix. Foram aplicadas 72 métricas a um conjunto de mais de 10 mil resumos, e comparações estatísticas (t-test) foram realizadas entre o período de referência (2020–2022) e os anos subsequentes. Os resultados indicam transformações a partir de 2023, incluindo simplificação estrutural, aumento da densidade lexical, reconfiguração de estratégias discursivas e mudanças no uso de conectivos. Em 2024 e 2025, mais de 95% dos artigos apresentam múltiplas métricas significativamente distintas em relação ao período de referência.

1 Introdução

A pesquisa científica vem sofrendo grande impacto com a popularização de Large Language Models (LLMs), frequentemente utilizadas para polimento e geração de textos em produções científicas. Seu uso, em especial na escrita científica possibilita celeridade na produção acadêmica, diminuindo também a barreira existente na escrita científica entre Native English Speakers (NES) e Non Native English Speakers (NNES) (Lin et al., 2025).

Contudo, apesar de aparentemente inofensivas, essas ferramentas podem se apropriar do estilo de escrita de um autor, criando brechas para violações de direitos autorais (Cranor et al., 2024). Seu uso desenfreado também levanta questionamentos sobre o impacto cognitivo, o qual pode afetar diretamente o desenvolvimento do pensamento crítico, pilar fundamental da pesquisa científica e da academia (Kosmyna et al., 2025).

O cenário exposto acima trouxe luz a estudos que buscam compreender como o uso de tais ferramentas vem reestruturando e remodelando a pesquisa e em especial, a escrita científica ao redor do mundo. Por exemplo, Kobak et al. (2025) analisa mudanças

no vocabulário de resumos de artigos de Biomedicina entre 2010 e 2024, identificando palavras e estilos linguísticos inéditos após a introdução das Inteligência Artificial Generativa (IAGs).

No Brasil, a Sociedade Brasileira de Computação (SBC), entidade responsável pela promoção e organização de diversos eventos científicos na área de Computação no país, registrou recordes no número de submissões de artigos científicos nos anos de 2024 e 2025. Nesse sentido, o aumento das submissões pode estar associado ao processo de aceleração da produção de textos científicos por meio das ferramentas de IAGs.

Sendo assim, este trabalho analisa, em caráter exploratório, padrões linguísticos presentes em resumos de artigos científicos em português da SBC produzidos em períodos anteriores e posteriores à popularização das LLMs. Logo, a seção abaixo apresenta trabalhos relacionados ao escopo deste artigo. Após, apresenta-se a metodologia, a qual usa a ferramenta NILC-Metrix para extração de variadas métricas linguísticas dos resumos. Por fim, por meio de análises estatísticas aplicadas aos dados buscou-se identificar possíveis indícios de transformações linguísticas associadas a esse marco tecnológico.

2 Trabalhos Correlatos

Uma primeira vertente de estudos foca na quantificação do uso de Large Language Models (LLMs) em grandes repositórios. Kobak et al. (2025) analisaram 15,1 milhões de resumos do PubMed (2010–2024) por meio de excess word analysis, indicando que ao menos 13,5% dos resumos de 2024 apresentam sinais de processamento por IA.

Seguindo essa linha, Astarita et al. (2024) rastreou palavras características em 1 milhão de artigos da NASA ADS, observando um aumento estatisticamente significativo a partir de 2024, especialmente em artigos não revisados por pares.

De forma complementar, [Geng and Trotta \(2024\)](#) examinaram o arXiv (2018–2024) com um classificador estilístico, notando que 35% dos resumos em Ciência da Computação exibem traços de LLMs, enquanto a Matemática apresenta impacto mínimo.

No mesmo sentido, [Liang et al. \(2024\)](#) realizou uma análise sistemática em quase 1 milhão de artigos do arXiv, bioRxiv e Nature, sugerindo que o crescimento do uso de LLMs já representa uma incorporação estrutural na escrita científica, especialmente em preprints e artigos curtos.

Além do mapeamento volumétrico, outra linha de investigação concentra-se nas diferenças qualitativas entre a produção humana e a sintética. [Bagdasarov and Alves \(2025\)](#) analisaram características léxicas e sintáticas de textos científicos, observando que humanos mantêm maior variabilidade sintática, enquanto modelos como GPT-4o e Llama 3.1 apresentam maior variabilidade léxica.

Já [Zanotto and Aroyehun \(2024\)](#) caracterizou textos de 11 modelos diferentes em múltiplos domínios, revelando que a escrita humana tende a estruturas sintáticas mais simples e conteúdo semântico diverso, contrastando com a tendência à homogeneização estilística dos modelos recentes.

Corroborando essa visão, [Culda et al. \(2025\)](#) propôs um framework para analisar introduções de textos, destacando que modelos como o Qwen2.5 frequentemente dependem de padrões repetitivos e clichês, o que reduz a originalidade em comparação ao engajamento humano.

Por fim, estudos mais recentes exploram como essa distinção se manifesta em componentes gramaticais e contextos específicos. [Georgiou \(2025\)](#) investigou a representação de componentes fonológicos, morfológicos e sintáticos em ensaios, encontrando diferenças significativas no uso de substantivos, adjetivos e palavras complexas entre humanos e o ChatGPT.

Em uma análise de notícias, [Muñoz-Ortiz et al. \(2024\)](#) observou que textos humanos possuem maior variedade vocabular e carga emocional negativa, enquanto os LLMs utilizam mais símbolos e pronomes.

Por fim, [Rodrigues et al. \(2026\)](#) comparou conteúdos em português, concluindo que a desinformação gerada por IA é mais formal e positiva que a humana, o que desafia a eficácia de modelos tradicionais de detecção de fake news.

3 Metodologia

A metodologia compreendeu três etapas: (1) coleta de resumos científicos do repositório selecionado; (2) aplicação das métricas linguísticas do NILC-Matrix para extração dos indicadores; (3) comparação entre períodos através do *t-test* para identificar variações estatisticamente significativas.

3.1 Coleta de Dados

O processo de coleta de dados foi realizado por meio de técnicas de web scraping aplicadas aos artigos disponíveis no repositório da Sociedade Brasileira de Computação (SOL-LIB).¹

Inicialmente, para cada artigo, foram extraídas as informações: categoria, título, URL do artigo, autores, evento e data de publicação, totalizando 34.318 registros. Em seguida, foi realizada uma requisição à página individual da publicação, a partir da qual foram extraídos os campos de resumo e palavras-chave.

Após a coleta, os dados passaram pela etapa de pré-processamento. Inicialmente, foram removidos os registros cujos títulos continham os termos “Apresentação”, “Apresentação e Organização” e “Cover e Frontmatter”, uma vez que tais itens não correspondiam propriamente a artigos científicos, mas sim a materiais introdutórios ou de apresentação de eventos. Adicionalmente, artigos duplicados, decorrentes da publicação de versões estendidas, também foram excluídos do conjunto de dados.

Após, os campos resumo, palavras-chave e data passaram por tratamento textual. Nos resumos, foi removida a ocorrência inicial dos termos “Resumo” ou “Abstract”. De forma análoga, no campo de palavras-chave, foram eliminados os termos iniciais “Palavras-chave” ou “Keywords”. No campo data, foram removidos os prefixos “Publicado” ou “Published”. Por fim, obteve-se um conjunto de 29.678 registros tratados.

Ainda, considerando que o objetivo deste trabalho é a análise de resumos em língua portuguesa, foi utilizada a biblioteca langdetect para a identificação automática do idioma dos textos. Foram identificados 17.490 resumos em português, distribuídos entre os anos de 1987 e 2025. Para as análises realizadas neste estudo, foram selecionados apenas os artigos compreendidos entre 2020 e 2025, totalizando 10.066 resumos em português.

¹Disponível em: [Repositório da SOL-LIB SBC](#).

3.2 Obtenção das Métricas Linguísticas

O processo de obtenção das métricas linguísticas dos resumos foi realizado com base no repositório do NILC-Metrix (Leal et al., 2024), disponibilizado no GitHub. O NILC-Metrix é uma ferramenta de análise linguística que fornece um amplo conjunto de métricas destinadas à avaliação da complexidade textual em diferentes níveis linguísticos.

De modo geral, na Tabela 1 é apresentado a quantidade de métricas analisadas por grupo, bem como uma breve descrição do grupo linguístico em questão. Foram analisadas 72 métricas que abrangem 11 dos 14 grupos linguísticos do NILC-Metrix, com ausência de métricas dos grupos de medidas psicolinguísticas, simplicidade textual e densidade de padrões sintáticos.

Embora o NILC-Metrix disponibilize cerca de 200 métricas, este estudo utiliza 72 métricas extraídas pelo script `run_minimal`, seguindo o procedimento do repositório oficial da ferramenta. A execução do conjunto completo (`run_all`) apresentou limitações técnicas que inviabilizaram sua implementação no momento da construção deste trabalho.

3.3 T-test Welch

Para identificar variações estatisticamente significativas entre o grupo de resumos anterior à adoção de LLMs (2020-2022) e as médias anuais dos anos subsequentes, foi aplicado o teste *t* de Welch às métricas linguísticas. Essa variação do teste *t* clássico não assume homogeneidade de variâncias, sendo adequada a cenários com tamanhos amostrais distintos e variâncias populacionais desconhecidas (Welch, 1947). A estatística *t* foi calculada a partir da diferença entre as médias, com graus de liberdade aproximados pela fórmula de Welch-Satterthwaite.

Para cada métrica, calculou-se inicialmente a média ponderada do grupo de referência (2020-2022), utilizando como pesos os respectivos tamanhos amostrais, bem como a variância combinada desse período, formando a distribuição de referência pré-LLMs. Em seguida, cada ano posterior (2023, 2024 e 2025) foi comparado individualmente a esse grupo por meio do teste *t*, obtendo-se a estatística *t*, os graus de liberdade e o valor de *p*. Valores de *p* inferiores a 0,05 foram considerados estatisticamente significativos, e inferiores a 0,001, altamente significativos. O sinal da estatística *t* indicou a direção da variação em relação ao período de referência.

Por fim, os resultados de cada grupo e métrica, bem como o conjunto de dados principal e o código utilizado para análise podem ser observados por meio do [Github](#) do projeto.

4 Resultados

A distribuição dos níveis de significância ao longo dos anos revela um padrão de fortalecimento estatístico a partir de 2024. Na Tabela 2 é observado que em 2023 há predomínio de resultados não significativos, convergindo para o momento de início de uso, ainda que tímido de IAGs na Escrita Científica. As diferenças em relação ao grupo de referência ainda eram limitadas ou incipientes, com grande parte das métricas não apresentando evidência estatística suficiente para indicar mudanças sistemáticas.

Nos anos posteriores, há predomínio das métricas altamente significativas. Esse deslocamento pode indicar a adesão em maior escala de ferramentas generativas na produção de resumos científicos, tornando as transformações linguísticas observadas mais consistentes e recorrentes. Tal padrão aponta para uma possível estabilização das mudanças linguísticas ao longo do tempo, na qual um conjunto substancial de métricas passa a apresentar diferenças robustas e estatisticamente consolidadas em relação ao período de referência.

A análise da direção dos efeitos mostra os resultados de significância e permite compreender o sentido predominante das transformações linguísticas ao longo do período analisado. Na Figura 1 é ilustrado uma predominância nos anos de 2024 e 2025 de métricas com direção de diminuição, em comparação às métricas que indicam aumento.

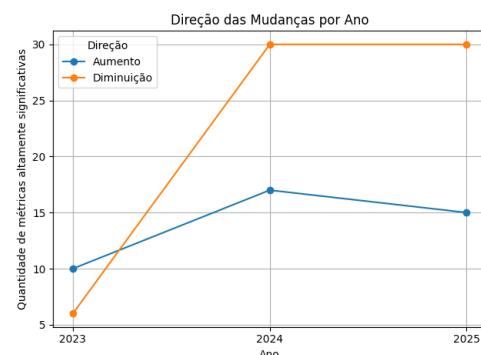


Figura 1: Mudanças por Período

A análise dos resultados pelo categoria de significância e sentido de direção sugerem que, no conjunto das métricas analisadas, há uma tendência de

Grupo Linguístico	Descrição	Métricas Analisadas
Coesão Referencial	Métricas que capturam a presença de elementos necessários para a formação de cadeias de correferência no texto.	7
Coesão Semântica	Métricas que medem a similaridade semântica entre sentenças, indicando a continuidade temática do texto.	11
Complexidade Sintática	Métricas baseadas em análises sintáticas de dependência e de constituintes que caracterizam a estrutura das sentenças.	3
Conectivos	Métricas de palavras que estabelecem relações lógicas e coesivas entre diferentes partes do texto.	12
Diversidade Lexical	Métricas de proporção entre palavras distintas (types) e o total de palavras (tokens) no texto.	2
Frequência de Palavras	Métricas que medem a frequência média das palavras com base em corpora de referência, indicando o grau de especialização do vocabulário.	10
Índices de Leiturabilidade	Métricas que estimam o nível de facilidade ou dificuldade de leitura do texto.	3
Informações Morfosintáticas de Palavras	Métricas sobre densidade e distribuição de palavras de conteúdo e funcionais no texto e nas sentenças.	9
Informações Semânticas de Palavras	Características semânticas do vocabulário utilizado, como classes ou relações conceituais.	7
Léxico Temporal	Métricas de frequência relativa de tempos e modos verbais em relação ao total de verbos do texto.	2
Medidas Descritivas	Estatísticas básicas do texto, como contagem de palavras e sentenças.	6

Tabela 1: Grupos linguísticos analisados, suas descrições e a quantidade de métricas consideradas em cada grupo.

Ano	Tipo	Total
2023	Altamente Significativo	16
2023	Significativo	13
2023	Não Significativo	43
2024	Altamente Significativo	47
2024	Significativo	10
2024	Não Significativo	15
2025	Altamente Significativo	45
2025	Significativo	7
2025	Não Significativo	20

Tabela 2: Métricas pelo Tipo de Significância

redução em propriedades linguísticas associadas, em muitos casos, à extensão textual, à complexidade estrutural ou à densidade de determinados fenômenos linguísticos.

4.1 Quantificação de Métricas nos Artigos

Na Figura 2 é apresentado o comportamento das métricas mais significativas dos anos de 2024 e 2025. Entre 2020 e 2022, nenhum dos artigos alcança os limiares considerados (10, 15, 20, 25 ou 27 métricas), indicando ausência do uso de IAGs no período.

A partir de 2023 observa-se uma incidência expressiva e consistente, quase a totalidade dos artigos atende ao critério mínimo de 10 métricas (acima de 95% em todos os anos), e proporções igualmente elevadas são observadas para o limiar de 15 métricas, variando de aproximadamente 92% a 95%. À medida que o critério se torna mais restritivo, há uma redução gradual, porém estável, das proporções, cerca de 70% a 74% dos artigos alcançam pelo menos 20 métricas, enquanto apenas uma

fração menor atende aos limiares mais elevados, como 25 métricas (entre 19% e 29%) e 27 métricas (entre 3% e 4%).

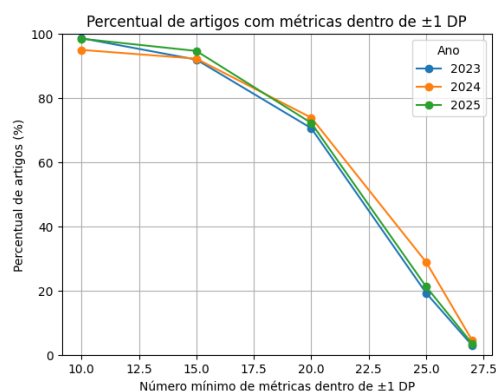


Figura 2: Percentual de Artigos com Métricas Altamente Significativas entre 2023 a 2025

Esse padrão sugere que, após 2023, a grande maioria dos resumos passa a apresentar múltiplas métricas linguisticamente alteradas em relação ao período de referência, ainda que apenas uma parcela reduzida concentre um número muito elevado de alterações simultâneas, indicando um efeito disseminado sobre o conjunto de métricas analisadas.

4.2 Resultados por Grupo Linguístico

A análise das métricas revelou padrões distintos de transformação nos textos ao longo do período estudado. Em termos de **Coesão Referencial**, observou-se uma redução em todas as métricas, indicando menor repetição de elementos referenciais

e ligações explícitas mais fracas entre ideias. Essa tendência é reforçada pela **Coesão Semântica**, que apresentou queda na similaridade entre sentenças, com impacto em 10 das 11 métricas em 2024 e 7 em 2025, sugerindo menor continuidade temática e progressão discursiva. No âmbito da **Complexidade Sintática**, houve uma redução consistente em todos os indicadores, apontando para estruturas frasais mais simples e com menor profundidade de encadeamento. Quanto aos **Conectivos**, embora as mudanças sejam estatisticamente significantes em diversos tipos, não foi observada uma tendência única de aumento ou diminuição.

No que diz respeito ao vocabulário, a **Diversidade Lexical** apresentou aumento significativo em 2024 e 2025, indicando léxico mais variado e maior proporção de palavras de conteúdo, apesar da redução na coesão. A **Frequência de Palavras** revelou o uso de termos mais comuns em 2023 e 2024, com uma leve reversão para palavras menos frequentes em 2025, alterando o grau de familiaridade do texto. Paralelamente, os **Índices de Leiturabilidade** mostraram queda nos índices Brunet e Flesch, acompanhada por um aumento no índice Honore em 2025, sugerindo textos menos uniformes e com legibilidade oscilante.

As **Informações Morfossintáticas** indicaram um aumento na proporção de adjetivos e substantivos em contraposição à redução de pronomes e palavras funcionais, resultando em textos mais informativos e menos dependentes de ligações gramaticais. Nas **Informações Semânticas**, verificou-se menor ambiguidade em verbos e adjetivos, maior profundidade verbal e aumento de termos positivos, embora em 2024 tenha ocorrido um pico de termos negativos e ambiguidade em substantivos. Já o **Léxico Temporal** sofreu diminuição nas conexões temporais positivas em 2024 e 2025, sugerindo menor uso de marcadores de sequência. Por fim, as **Medidas Descritivas** confirmaram uma redução no total de palavras, de sentenças e na extensão das frases, resultando em textos mais curtos, porém compostos por palavras com maior número médio de sílabas.

Os resultados de cada grupo e métrica, bem como o conjunto de dados principal e o código utilizado para análise podem ser observados por meio do [Github](#) do projeto.

5 Conclusão

Este estudo mostrou que após 2023 transformações sistemáticas e mensuráveis nos resumos científicos da SBC em português são notadas, caracterizadas por simplificação estrutural, densidade lexical elevada e reconfiguração das estratégias discursivas. A análise estatística revela que em 2024-2025 a maioria dos artigos apresenta múltiplas métricas alteradas em relação ao período anterior a 2023. Esse equilíbrio entre economia estrutural e complexidade informacional pode ser um novo delimitador no padrão de escrita acadêmica brasileira.

É importante destacar a necessidade de analisar os resumos do mesmo período sob a ótica das 200 métricas do NILC-Metrix, visando identificar de forma aprofundada as transformações na escrita científica e as correlações existentes entre os diferentes indicadores linguísticos.

Ainda, diante da carência de dados validadas para o treinamento de classificadores, destaca-se a necessidade de construir um conjunto de dados textuais científicos com curadoria e validação linguística rigorosa das amostras humanas e sintéticas.

Conseqüentemente, faz-se necessário investigar a eficácia de variadas técnicas de Aprendizado de Máquina, Aprendizado Profundo e de arquiteturas modernas baseadas em *Transformers*, para a identificação de textos gerados ou alterados por IAGs.

Por fim, pretende-se expandir a investigação para artigos na íntegra, a fim de compreender como a influência da ferramentas de IAG se distribui por parágrafos e seções específicas. Isso permitirá identificar não apenas as métricas mais sensíveis, mas também quais partes de um artigo científico (Introdução, Metodologia, Resultados) são mais suscetíveis a alterações estilísticas e estruturais.

6 Uso de Inteligência Artificial Generativa

As ferramentas de Inteligência Artificial Generativa ChatGPT e Gemini foram utilizadas para sintetizar os resultados da seção de trabalhos correlatos, considerando o limite no número de páginas.

References

Simone Astarita, Sandor Kruk, Jan Reerink, and Pablo Gómez. 2024. [Delving into the utilisation of ChatGPT in scientific publications in astronomy](#). *Proceedings of SPAICE2024: The First Joint European Space Agency / IAA Conference on AI in and for Space*, pages 241–246.

- Sergei Bagdasarov and Diego Alves. 2025. [Like a human? a linguistic analysis of human-written and machine-generated scientific texts](#). In *Proceedings of the First on Natural Language Processing and Language Models for Digital Humanities*, pages 38–47, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.
- Lorrie Cranor, Kim Hazelwood, Daniel Lopresti, and Amanda Stent. 2024. [Conference submission and review policies to foster responsible computing research](#). *arXiv preprint arXiv:2408.09678*.
- Lia Cornelia Culda, Raluca Andreea Nerişanu, Marian Pompiliu Cristescu, Dumitru Alexandru Mara, Adela Bâra, and Simona-Vasilica Oprea. 2025. [Comparative linguistic analysis framework of human-written vs. machine-generated text](#). *Connection Science*, 37(1):2507183.
- Mingmeng Geng and Roberto Trotta. 2024. [Is ChatGPT transforming academics’ writing style?](#) *ICML 2024 Next Generation of AI Safety Workshop*.
- Georgios P. Georgiou. 2025. [Differentiating between human-written and AI-generated texts using automatically extracted linguistic features](#). *Information*, 16(11):979.
- Dmitry Kobak, Rita González-Márquez, Emőke Ágnes Horvát, and Jan Lause. 2025. [Delving into LLM-assisted writing in biomedical publications through excess vocabulary](#). *Science Advances*, 11(27):eadt3813.
- Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Berensitzky, Iris Braunstein, and Pattie Maes. 2025. [Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task](#). *arXiv preprint arXiv:2506.08872*.
- S. E. Leal, M. S. Duran, C. E. Scarton, and 1 others. 2024. [NILC-Metrix: assessing the complexity of written and spoken language in brazilian portuguese](#). *Language Resources and Evaluation*, 58:73–110.
- Weixin Liang, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D. Manning, and James Y. Zou. 2024. [Mapping the increasing use of LLMs in scientific papers](#). *arXiv preprint arXiv:2404.01268*.
- Dingkang Lin, Naixuan Zhao, Dan Tian, and Jiang Li. 2025. [ChatGPT as linguistic equalizer? quantifying LLM-driven lexical shifts in academic writing](#). *arXiv preprint arXiv:2504.12317*.
- A. Muñoz-Ortiz, C. Gómez-Rodríguez, and D. Vilares. 2024. [Contrasting linguistic patterns in human and LLM-generated news text](#). *Artificial Intelligence Review*, 57:265.
- Flávia A. Rodrigues, Niclas F. Sturm, and Flávio L. Pinheiro. 2026. [A linguistic comparison between human- and AI-generated content](#). *iScience*, 29(3).
- Bernard L. Welch. 1947. The generalization of “student’s” problem when several different population variances are involved. *Biometrika*, 34(1–2):28–35.
- Sergio E. Zanotto and Segun Aroyehun. 2024. [Human variability vs. machine consistency: A linguistic analysis of texts generated by humans and large language models](#). *arXiv preprint arXiv:2412.03025*.