

# A coerência na predição de verbos *dicendi* em resumos: comparação entre capacidades humanas e de atuais modelos de linguagem

*Coherence in the Prediction of Dicendi Verbs in Abstracts:  
A Comparison Between Human Capabilities and Current  
Language Models*

**Osmar de Oliveira Braz Junior**  
Universidade do Estado de Santa  
Catarina (UDESC) | Florianópolis | SC | BR  
osmar.braz@udesc.br  
<https://orcid.org/0000-0003-0321-4787>

**Roberlei Alves Bertucci**  
Universidade Tecnológica Federal do  
Paraná (UTFPR) | Curitiba | PR | BR  
Universidade Federal de Santa Catarina  
(UFSC) | Curitiba | PR | BR  
Fundação Araucária | Curitiba | PR | BR  
bertucci@utfpr.edu.br  
<https://orcid.org/0000-0003-4014-5610>

**Renato Fileto**  
Universidade Federal de Santa Catarina  
(UFSC) | Curitiba | PR | Brasil  
r.fileto@ufsc.br  
<https://orcid.org/0000-0002-7941-6281>

**Resumo:** A relação entre a tecnologia e a linguagem humana é tema de debate crescente, especialmente com o surgimento de modelos de linguagem de grande escala (LLMs) e sua capacidade de produção textual. Essa pesquisa justifica-se pela importância de verificar a proximidade entre as produções linguísticas dos LLMs e as humanas, partindo do pressuposto de que a linguagem humana é moldada pelo conhecimento epilingüístico. Para isso, conduziu-se um estudo comparativo em que se pediu a estudantes e a LLMs que sugerissem verbos (*dicendi*) para completar coerentemente lacunas em um resumo de um texto-base. A escolha de um resumo escolar foi motivada pela necessidade de profunda compreensão do texto original e coerência textual, o que representa um desafio tanto para humanos quanto para LLMs. A metodologia incluiu a comparação das sugestões dos alunos e 5 LLMs via chat, buscando avaliar a coerência com o texto-fonte. Os resultados mostram que tanto LLMs quanto humanos apresentaram boas sugestões nas lacunas. No entanto, os resultados contrariam a hipótese inicial de que os humanos apresentariam uma maior diversidade lexical. Por outro lado, os resultados indicam também que LLMs têm um desempenho melhor quanto às coerências local e global. Estes fatos sugerem que os LLMs podem ser uma ferramenta de apoio para os estudos relacionados à intersecção entre linguagem e tecnologia, inclusive para o ensino.



**Palavras-chave:** coerência; resumo escolar; predição de palavras; verbos *dicendi*; modelo de linguagem de grande escala.

**Abstract:** The relationship between technology and human language is a topic of growing debate, especially with the emergence of large-scale language models (LLMs) and their textual production capacity. This research is justified by the importance of verifying the proximity between the linguistic productions of LLMs and humans, based on the assumption that human language is shaped by epilinguistic knowledge. To this end, we conducted a comparative study in which we asked students and LLMs to suggest verbs (*dicendi*) to coherently fill gaps in a summary of a source text. The choice of a school summary was motivated by the need for a deep understanding of the original text and textual coherence, which represents a challenge for both humans and LLMs. The methodology included the comparison of the students' and 5 LLMs suggestions via chat, seeking to assess coherence with the source text. The results show that LLMs and humans presented good suggestions in the gaps. However, the results contradict the initial hypothesis that humans would present greater lexical diversity. On the other hand, the results also indicate that LLMs perform better in terms of local and global coherence. These facts suggest that LLMs can be a support tool for studies related to the intersection between language and technology, including for teaching.

**Keywords:** coherence; school abstract; word prediction; *dicendi* verbs; large language model.

## 1 Introdução

Enquanto a discussão sobre o impacto da tecnologia nas vidas humanas é algo relativamente antigo na história, o modo como a tecnologia se relaciona com a linguagem humana tem ganhado mais força na atualidade. A presença de aplicativos de conversação, equipamentos que obedecem a comando de voz e (claro) os modelos de linguagem de grande escala (do inglês, *Large Language Models*—LLMs) colocam em discussão o modo como as práticas linguageiras serão afetadas por essas tecnologias. Nesse contexto, a capacidade de sumarizar é uma rea-

lidade para os LLMs, que apresentam bons textos produzidos a partir de *prompts* (Paes; Freitas, 2023; Paiola, 2022). Por outro lado, como esses modelos são probabilísticos, questiona-se o quanto eles seriam capazes de se aproximar da criatividade humana nos processos de escrita (Nunes, 2024). Nesta pesquisa, assumimos que o ser humano produz linguagem a partir de seu conhecimento epilingüístico (Culioli; Normand, 2005). Entendemos que é essa capacidade que permite que ele faça as escolhas adequadas para a produção de um enunciado (e.g., a construção de um texto falado ou escrito). Com esse pressuposto, propomos um experimento cuja intenção é verificar a proximidade dos LLMs com a capacidade humana. Conduzimos um estudo comparativo investigando a capacidade dos LLMs em sugerir verbos *dicendi* adequados em resumos. Selecioneamos um texto utilizado em uma prova de vestibular e um resumo deste texto que obteve uma nota alta (i.e., 19,25/20 ou 96% da nota total). Posteriormente solicitamos a um especialista em Língua Portuguesa que destacasse os verbos *dicendi* do resumo e os substituímos por lacunas. Em seguida, apresentamos o texto-base e o texto com lacunas a alunos de graduação do curso de Letras e solicitamos que sugerissem um verbo para cada lacuna do resumo, a partir da interpretação do texto-fonte. Em paralelo, o texto-base e o resumo com lacunas foi submetido a LLMs para realizar a mesma tarefa de predizer verbos para as lacunas<sup>1</sup>. Por fim, comparamos as sugestões dos alunos com as previsões dos modelos.

Nossa decisão pelo resumo escolar se deve ao fato de ser um texto que exige uma profunda compreensão do texto-base e, ao mesmo tempo, requerer uma escrita que esteja coerente com esse mesmo texto. Nesse sentido, ele apresenta um desafio duplo ao estudante, já que, além de ficar atento à leitura, ele deverá selecionar informações relevantes e apresentá-las de forma coerente ao texto original. Para garantir essa coerência, entre outras exigências, é fundamental que o estudante utilize adequadamente verbos *dicendi*. Esses verbos indicam quem é o responsável por determinada ideia. Ao utilizar os verbos *dicendi* de forma precisa, o estudante demonstra que comprehendeu a posição do autor do texto-base e evita a atribuição errônea de ideias (Bragagnollo, 2017; Clark; Gerrig, 1990; Machado; Lousada; Abreu-Tardelli, 2005).

Com isso em mente, apontamos uma questão-problema: dada uma tarefa de preenchimento de lacunas em um texto, quais as diferenças observadas entre as previsões humanas e as de LLMs? Assim, nosso objetivo é tecer uma comparação entre as sugestões de preenchimento de lacunas apontadas por humanos e por LLMs, verificando a coerência dessas sugestões com o que o texto-base permite. Nossa hipótese inicial é de que os seres humanos proporcionalmente oferecerão mais alternativas de preenchimento (i.e., maior diversidade lexical) do que os LLMs em virtude de: *i*) utilizarem de seu conhecimento epilingüístico para a tarefa, algo que os LLMs não possuem; e *ii*) realizarem uma interpretação global do texto, para além das palavras ali presentes, ao contrário dos LLMs, que são probabilísticos.

A avaliação automática de textos tem se mostrado uma ferramenta promissora para auxiliar estudantes e professores no processo de escrita e correção. Nesse contexto, este trabalho contribui ao investigar a capacidade de LLMs em identificar verbos *dicendi* adequados para o preenchimento de lacunas em resumos. A hipótese é que, ao analisar a estrutura e o conteúdo de um texto, um LLM seja capaz de sugerir verbos *dicendi* que refletem o objetivo do texto original. Essa pesquisa se alinha a estudos anteriores que exploram o potencial das

<sup>1</sup> Neste trabalho, utilizaremos “texto original”, “texto lido”, “texto-fonte” e “texto-base” indefinidamente para nos referirmos ao texto que dá origem ao resumo.

inteligências artificiais (IAs) na geração e análise de textos (Paes; Freitas, 2023; Nunes, 2024), buscando aprofundar a compreensão do papel desses modelos na avaliação de textos acadêmicos, como ressaltado por Meira *et al.* (2023) e Rassi e Lopes (2023). A capacidade de um LLM em sugerir verbos *dicendi* adequados pode contribuir significativamente para a melhoria da qualidade dos resumos, uma vez que esses verbos desempenham um papel crucial na expressão do objetivo do texto original.

Com esse recorte de pesquisa, temos o intuito de contribuir para a discussão das possibilidades de uso de IAs em contextos de produção e avaliação de textos, sobretudo no ensino. Nesse viés, o artigo contribui essencialmente com:

- ◆ o enfoque aos verbos *dicendi* na coerência de resumos, comparando sugestões humanas com versões de LLMs;
- ◆ a reflexão sobre as capacidades e limitações do emprego de IAs na produção e avaliação de textos, o que pode ter impacto no campo da educação; e
- ◆ a disponibilização da projeção online dos *embeddings* das palavras do texto para análise e comparações de sua proximidade semântica.

O restante do artigo está organizado assim: a Seção 2 apresenta os fundamentos da pesquisa; a Seção 3 descreve a metodologia utilizada para os experimentos; a Seção 4 apresenta os resultados e as discussões; finalmente, na Seção 5, concluímos o trabalho e apresentamos pontos a serem ainda investigados.

## 2 Fundamentos

A produção de resumos exige uma compreensão profunda dos mecanismos de coerência textual. Nesta seção, aprofundaremos a discussão sobre a coerência semântica, explorando como ela se manifesta na construção de resumos escolares e como a seleção de verbos *dicendi* influencia a percepção de coerência entre o resumo e o texto original.

### 2.1 Coerência semântica

Seguindo as premissas de De Beaugrande e Dressler (1981) sobre a natureza textual, Wang e Guo (2014) destacam a coerência como o atributo que distingue discursos coesos e significativos daqueles que apresentam incoerências ou lacunas na construção de sentidos. De modo mais específico, Koch e Travaglia (2021) caracterizam a coerência semântica como a compatibilidade de significado tanto entre elementos textuais adjacentes (i.e., nível local) quanto entre todos os elementos de um texto (i.e., nível global).

Inclusive na área da Computação, a coerência semântica, conforme abordado por Braz Jr e Fileto (2021), consiste na organização de componentes textuais (palavras, frases, etc.) de modo a estabelecer relações de significado compatíveis, permitindo ao leitor construir uma interpretação consistente do texto. A incoerência semântica, por sua vez, ocorre quando elementos textuais próximos apresentam significados incompatíveis, ou seja, quando há um

distanciamento semântico entre esses elementos ou entre eles e o contexto geral (i.e., frase, parágrafo, documento).

Focando na educação básica, Gomes (2020) argumenta que a análise da coerência textual exige uma avaliação abrangente, incluindo o contexto comunicativo em que o texto está inserido. A interação entre autor e leitor, com seus objetivos específicos, influencia diretamente a avaliação da adequação do texto. Segundo o autor, o resumo – texto foco da presente pesquisa – demanda uma coerência tanto global quanto situacional. A coerência global exige que o resumo seja fiel ao texto-base, enquanto a coerência situacional considera o contexto de produção do resumo.

Desenvolvendo esses pontos e focando em resumos escolares produzidos em contexto de vestibular, Costa e Silva (2013) defendem que a análise da coerência textual exige uma avaliação cuidadosa da situação comunicativa na qual os textos são produzidos. A interação entre autor e leitor, caracterizada por objetivos específicos de leitura e escrita, deve ser considerada para se avaliar a adequação do texto. Os autores defendem que o resumo, em particular, demanda uma coerência tanto global quanto situacional. A coerência global se refere à compatibilidade do conteúdo do resumo com as informações presentes no texto -base, enquanto a coerência situacional diz respeito à consideração da situação de produção do resumo. No caso específico do vestibular, o avaliador julga tanto a coerência interna do resumo, quando a coerência deste com o texto-fonte. Nesse sentido, cabe ao candidato ficar atento a essa situação de produção específica.

Numa perspectiva diferente, na área de Computação, Souza (2022) propõe um modelo inovador para a geração automática de resumos de textos científicos extensos, com foco na sumarização (i.e., resumo) extrativa. O estudo investiga a viabilidade de combinar resumos extraídos de diferentes seções de um documento para produzir um resumo mais completo e coerente. A autora reconhece a importância da coerência e da coesão textual na construção de bons resumos e pretende explorar novas estratégias para combiná-las com outras métricas de avaliação. Ela destaca ainda a importância da coerência na sumarização, mas não mede o impacto dos verbos *dicendi* no texto.

De tudo o que se apresentou aqui, destacamos que, ao se elaborar um resumo, especialmente em contextos avaliativos como o descrito por Costa e Silva (2013), é crucial considerar a perspectiva do leitor, entendido como um avaliador. Ao contrário da produção de textos autônomos, onde a coerência interna é primordial, na avaliação de um resumo, exige-se também uma análise comparativa com o texto original. Nesse sentido, o leitor, que já possui conhecimento prévio do texto-base, atua como um “juiz”, verificando se as informações do resumo estão em consonância com o conteúdo original. Diante desse cenário, a coerência externa, ou seja, a relação do resumo com o texto-fonte, ganha maior relevância. No presente estudo, essa coerência será investigada a partir da análise dos verbos *dicendi*, que indicam os atos de fala realizados no texto-base e expressos no resumo, a partir da interpretação do produtor. A hipótese é que a seleção e o emprego desses verbos influenciam diretamente a percepção de coerência entre o resumo e o texto original.

## 2.2 Resumos escolares

O resumo escolar pode assumir diferentes funções, como ser um auxiliar para a compreensão de um texto ou ser um instrumento de avaliação das habilidades de leitura e escrita. Entre suas características, estão: a seleção de informações essenciais; a sumarização delas em uma estrutura textual; e a correta indicação das vozes que aparecem no texto original. Para que seja materializado, é igualmente importante que o produtor sintetize, por meio de um texto escrito, a interpretação global e os tópicos que sumarizam o texto ou evento a ser resumido.

No ambiente de ensino, o resumo é considerado um importante meio de avaliação de duas habilidades: a leitura, já que pode medir o grau de compreensão de um texto; e a escrita, por requerer do escritor que exponha essa interpretação em um texto independente, com coesão e coerência. Além disso, pode ser útil como um aliado ao estudo, já que auxilia a retenção das informações presentes nos textos. Bragagnollo (2017) destaca a relevância do resumo para o ensino por estar diretamente relacionado a práticas cotidianas diversas. Ela ressalta ainda que a elaboração do resumo precisa expor informações fiéis ao texto original, focando naquilo que é essencial no texto-base.

Devemos sublinhar também que o emprego da paráfrase é uma estratégia essencial para o resumo. Isso ocorre porque é esperado um texto novo (i.e., resumo), mas que deve ser fruto do texto-base. Essa estratégia deve se afastar de uma mera cópia ou alteração pontual do texto-fonte (e.g., com o simples uso de sinônimos). Idealmente, é melhor que ela se constitua, sobretudo, como uma apresentação dos sentidos que emergiram a partir da leitura realizada. Por isso, não se espera que o resumo seja uma sumarização pura e simples do texto, mas que indique marcas (i.e., coerentes) de leitura do produtor.

Nessa abordagem, assumimos que o resumo é originado pelo próprio texto original, uma vez que é organizado a partir deste (Machado; Lousada; Abreu-Tardelli, 2005). Portanto, ele é evidentemente acarretado pelo texto-fonte e suas informações devem estar presentes no texto que lhe deu origem. Sob essa perspectiva, é possível que se considere um resumo mais ou menos coerente em relação ao texto-base, em especial no caso do resumo escolar utilizado como item de avaliação, como é o foco neste trabalho. Assim, é essencial que o resumidor compreenda que, ao

resumir o que o autor do texto de base diz, estamos, de fato, interpretando o seu agir, atribuindo-lhe a efetivação de determinados atos. Ou seja, quando dizemos que o autor *nega* algo, *afirma* algo, *questiona* algo, estamos inferindo, através de nossa compreensão do conteúdo do texto, que o autor está realizando esses atos (Machado; Lousada; Abreu-Tardelli, 2005, p. 98-99, grifos no original).

As autoras confirmam que a interpretação é crucial para o texto, sendo um movimento anterior ao da escrita. Nesse sentido, a boa produção de um resumo exige uma escolha adequada do verbo *dicendi* para expressar o que o texto-base apresenta. Afinal, o ato de resumir contorna aquilo que o autor do texto-base promove, por meio da compreensão do autor do resumo.

Nesse contexto, podemos analisar com um pouco mais de detalhes a relevância dos verbos *dicendi* para o resumo. Do ponto de vista de sua estrutura, é comum aparecerem em períodos compostos, sendo eles os verbos da oração principal, seguidos por uma subordi-

nada, que representa o conteúdo interpretado no texto-base. Sua característica fundamental é apresentar um discurso reportado (Clark; Gerrig, 1990), ou que alguns autores consideram serem atos ou vozes do autor presentes no texto-fonte (Machado; Lousada; Abreu-Tardelli, 2005; Bragagnollo, 2017; Campos; Ribeiro, 2013). Em geral, nessa função de reportar uma ação/voz de um texto anterior eles evidenciam o modo como o resumidor compreendeu determinados trechos do texto. Por exemplo, se o autor de um texto está apresentando seus pontos de vista sobre um assunto, é mais aceitável que o resumidor use verbos como “argumenta”, “critica” ou “defende” do que outros como “expõe”, “esclarece” ou “narra”.

Focando no resumo escolar, entende-se que o resumo materializa um processo de retextualização, entendido como o processo de desenvolvimento de um novo texto a partir de um texto-base. Nesse sentido, é obrigatório que o resumidor deixe claro a quem pertencem as informações ali presentes, evitando possíveis conflitos de vozes. Nesse sentido, os verbos *dicendi* exercem um papel fundamental, uma vez que ajudam nessa distinção. Assim, o emprego desses verbos não se resume a um efeito de estilo, mas de construção de sentido no texto, em sua mais profunda função.

É exatamente por isso que a produção de um resumo e, consequentemente, a escolha dos verbos *dicendi* exige que o falante faça uso de seu conhecimento epilingüístico. Ao escolher expressões específicas para desenvolver seus raciocínios, o ser humano orienta seu discurso para determinado fim. Neste sentido, pode-se afirmar que todo falante nativo tem a capacidade de escolher expressões adequadas aos objetivos que pretende em uma dada situação de interação do cotidiano. A essa capacidade de uso linguístico adequado no cotidiano, Culioli e Normand (2005) denominam “epilinguismo”.

No recorte de uma situação de produção de resumo no contexto de avaliação (i.e., o vestibular de 2024 da UFPR, no caso desta pesquisa), há alguns pontos indicados por Costa e Silva (2013) que devemos trazer aqui. Primeiro, é necessária uma coerência entre as informações do texto-fonte e o resumo, de modo a não haver acréscimos de informações ou opiniões neste em relação àquele. Depois, o avaliador verifica a pertinência entre a interpretação do resumidor e as informações do texto-fonte, o que uma simples sumarização não faz. Em ambos os casos, as autoras consideram que é preciso se considerar uma coerência global na produção, para que a avaliação seja realizada.

A partir desta pesquisa, entendemos que os verbos *dicendi* possam ser cruciais para a métrica de avaliação de resumos, seja por humanos, seja por LLMs. Assim, além de critérios básicos exigidos para a boa avaliação de um resumo, como objetividade, clareza e, sobretudo, respeito ao conteúdo do texto original, consideramos que a escolha dos verbos *dicendi* é uma estratégia linguística fundamental, pois o seu uso garante que as vozes do texto original sejam devidamente representadas e que o leitor compreenda as informações principais do texto. Consequentemente, as chances de o resumo produzido apresentar coesão e coerência são maiores conforme o nível de adequação dos verbos *dicendi* escolhidos para referenciar as ações do autor do texto-base. Por isso, assumimos que a análise das sugestões realizadas por humanos e por LLMs a um texto com lacunas de verbos *dicendi* pode tanto dar pistas das capacidades linguísticas de humanos e máquinas, quando ser parte de uma tarefa para a avaliação desse gênero.

## 2.3 Modelos de Linguagem de Grande Escala

Atualmente, pesquisadores destacam que os LLMs ampliam significativamente as capacidades dos sistemas na geração de texto. Estudos apontam que os LLMs, treinados com vastas quantidades de dados textuais, apresentam uma notável habilidade para produzir textos que se assemelham àqueles criados por humanos (Kasneci *et al.*, 2023). Desde então, diversos LLMs foram desenvolvidos, como o GPT-3 (Radford *et al.*, 2018) da OpenAI, o LLaMA (Touvron *et al.*, 2023a,b) da Meta, o Gemini (Gemini Team *et al.*, 2023) do Google, Saibá-3 (Abonizio *et al.*, 2024) da Maritaca-AI (i.e., treinado em um grande *corpus* centrado no Brasil) e o Claude da Anthropic (Anthropic, 2024). Usar esses modelos, por meio de prompts apropriados, pode ser uma estratégia eficaz para orientar a avaliação e melhoria da coerência do texto.

Diante desse potencial, pesquisadores têm explorado formas de direcionar a geração de texto nos LLMs (Chen *et al.*, 2023; Marvin *et al.*, 2024). Um elemento fundamental nesse processo é o “prompt”, que pode ser definido como uma instrução ou entrada fornecida a um modelo de linguagem, solicitando uma resposta ou a continuação de um texto. Eles esclarecem que um *prompt* direciona o modelo para gerar textos relevantes ou executar tarefas específicas (Radford *et al.*, 2018). Segundo os pesquisadores, os *prompts* podem incluir perguntas, descrições de problemas, fragmentos de texto ou outras entradas que ajudem a moldar as respostas do modelo (Radford *et al.*, 2018). Para (Gu *et al.*, 2023), os *prompts* permitem submeter informações adicionais ou dicas a um LLM para guiar seu comportamento ou auxiliar na execução de uma tarefa específica. Essa interação entre o modelo e o usuário, mediada pelo *prompt*, deu origem a um novo paradigma no processamento de linguagem natural (PLN), conhecido como aprendizado baseado em prompt ou *prompt-learning* e segue o processo “pré-treinar, *prompt* e predizer” (do inglês, *pre-train, prompt, and predict*) (Liu *et al.*, 2023). Eles destacam que o uso de *prompts* possibilita guiar o comportamento dos LLMs por meio de informações contextuais ou exemplos relevantes relacionados à tarefa.

Apesar dessas promissoras aplicações, os LLMs ainda apresentam desafios. Uma das principais limitações é a falta de criatividade, decorrente de sua base probabilística, limita sua capacidade de gerar respostas verdadeiramente inovadoras (Paes; Freitas, 2023). A tarefa de preencher lacunas, por exemplo, é tratada como um problema de modelagem de linguagem, onde o modelo prevê probabilisticamente a palavra mais adequada com base no contexto fornecido no *prompt*. No entanto, essa abordagem pode ser insuficiente quando se exige maior nuance ou inovação textual. Essa característica dos LLMs, baseada em probabilidade, tem implicações importantes para diversas tarefas, como a de preencher lacunas. Nesse sentido, propomos a hipótese de que as respostas e avaliações feitas pelos LLMs tendem a ser inferiores às esperadas de humanos do ponto de vista da diversidade de opções. A ausência de criatividade impacta diretamente a interpretação de textos ou situações, o que pode interferir nos resultados experimentais. Assim, a capacidade de predição de elementos textuais, como verbos *dicendi*, pode ser limitada em tarefas que exigem maior flexibilidade ou originalidade na geração de respostas.

## 2.4 Trabalhos relacionados

Em nossa pesquisa, não encontramos trabalhos que analisassem o preenchimento de lacunas em resumos, nem que comparassem o desempenho de humanos e máquinas em contextos similares. Na literatura, o preenchimento de lacunas é comum, por exemplo, em experimentos sobre aquisição de escrita (Busnello *et al.*, 2011) e sobre julgamentos de aceitabilidade ou produção (Soto; Augusto; Abend, 2023).

Diversos trabalhos tratam sobre o gênero resumo escolar em português (Bicudo; Hila, 2015; Biral, 2003; Bragagnollo, 2017; Campos; Ribeiro, 2013; Costa; Silva, 2013; Machado; Lousada; Abreu-Tardelli, 2005) entre outros. Apesar disso, poucos dão enfoque aos verbos *dicendi*. Souza (2017) sublinha a relevância deles ao propor um modelo sociorretórico para a produção do resumo. Para a autora, a reflexão sobre o emprego adequado do verbo é uma estratégia fundamental para o gênero. Além dela, Bicudo e Hila (2015) recuperam o emprego de verbos *dicendi* em algumas redações de vestibulares. Os autores sustentam que, nos dados analisados, que eram redações bem avaliadas no certame, os candidatos selecionaram verbos coerentes para a exposição das ações que o autor do texto-base realizava.

No restante dos textos pesquisados, os verbos *dicendi* são reconhecidos como uma característica do resumo, não recebem um destaque especial. Além disso, vale ressaltar que, em nenhum deles: *i*) houve algum experimento em relação à predição de uso de verbos *dicendi* em resumos, como fizemos; nem *ii*) houve uma comparação entre as escolhas humanas e de LLMs. Dadas essas considerações, entendemos que nossa pesquisa é uma contribuição inédita para o estudo do gênero textual em questão, com foco em um elemento central que são os verbos.

## 3 Metodologia

Nos experimentos, utilizamos um texto-base que tem origem no vestibular da UFPR de 2024<sup>2</sup>. Um resumo, escolhido aleatoriamente entre os três disponíveis em um repositório de candidatos aprovados no curso de Medicina no vestibular da UFPR de 2024, foi utilizado no experimento<sup>3</sup>. O texto recebeu a nota 19,25/20, ou seja, obteve 96% da nota total esperada e está transscrito na Seção 4.

Com o objetivo de analisar a importância dos verbos *dicendi* na construção de resumos e avaliar a capacidade de Humanos e LLMs de preenchê-los, solicitamos a um especialista da área que removesse verbos *dicendi* do resumo, o que gerou assim um total de dez (10) lacunas. Esse especialista é doutor em linguística, professor universitário e pesquisador, com enfoque em produção de texto (i.e., inclusive em correções de processos seletivos). Depois, solicitamos a avaliadores humanos, estudantes da disciplina de Semântica do referido especialista, que realizassem a predição das palavras para as lacunas, com base no texto-fonte. A mesma tarefa

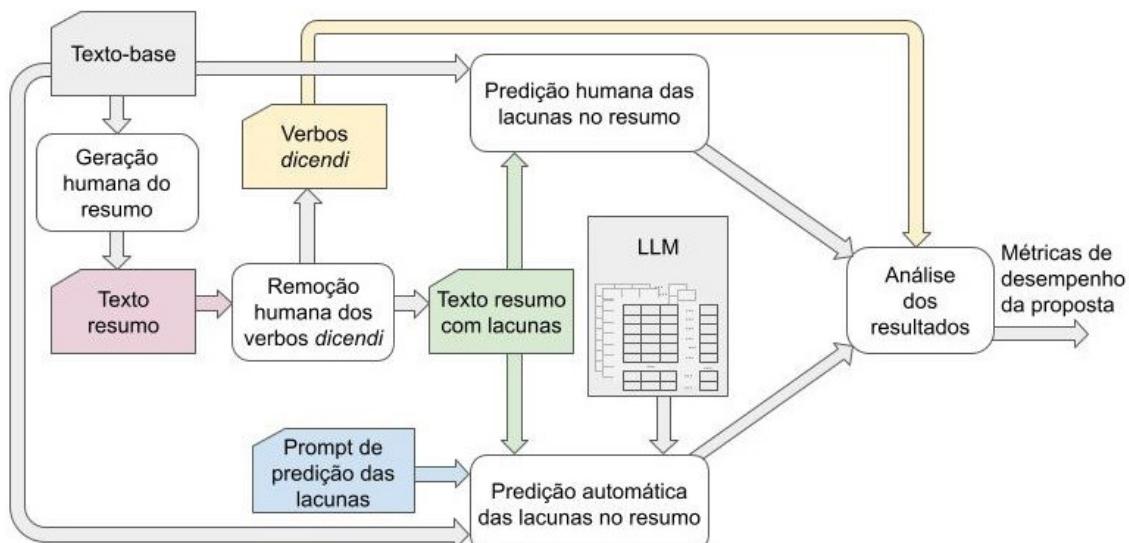
<sup>2</sup> Texto disponível em: [https://servicos.nc.ufpr.br/documentos/ps2024/provas/2fase/001-Compreensao\\_e\\_Producao\\_de\\_Texto.pdf](https://servicos.nc.ufpr.br/documentos/ps2024/provas/2fase/001-Compreensao_e_Producao_de_Texto.pdf)

<sup>3</sup> Originalmente disponível no drive do Desempenhos Med:<https://drive.google.com/drive/folders/1fCj-ZklwowmZSCOSfVUsUVaC3GvZaFyRoAcesso em 13 nov. 2024>.

de predição foi solicitada a cinco LLMs. Finalmente, a lista dos verbos *dicendi* do resumo foi analisada com as predições humana e automatizada. Vale ressaltar que os avaliadores humanos eram todos estudantes de Letras, cursando o 6º semestre (ou superior). Após a realização da tarefa, eles poderiam autorizar (ou não) a coleta dos textos produzidos com as lacunas preenchidas, para fins de pesquisa. Essa coleta seguiu o que se prevê em projeto aprovado no Comitê de Ética em Pesquisa (i.e., parecer CAAE 88328218.6.0000.5547).

A Imagem 1 ilustra o fluxo de trabalho da predição das lacunas, desde a remoção dos verbos *dicendi* até a comparação dos resultados das predições humana e automática. Como podemos observar, o fluxo se divide em duas etapas principais, uma de predição humana e outro com a predição automática com o uso de LLMs. A análise final dos resultados, incluindo métricas de desempenho, possibilita a avaliação da eficácia do fluxo de trabalho e a identificação de áreas para melhoria.

Imagem 1 – Fluxo de trabalho proposto



Fonte: elaboração própria.

Para entender melhor como os humanos realizam a tarefa de preenchimento das lacunas, descreveremos a seguir as etapas da tarefa “Remoção humana dos verbos *dicendi*” e da tarefa “Predição humana das lacunas no resumo”. Na tarefa “Remoção humana dos verbos *dicendi*”, um especialista humano em linguística identifica “Verbos *dicendi*” no resumo para gerar o “texto resumo com lacunas”. Já a tarefa “Predição humana das lacunas no resumo” apresenta o resumo com lacunas e o texto-base como uma atividade aos estudantes de Letras do 6º período ou superior, com as seguintes instruções: 1. O candidato da UFPR (2024) recebeu a seguinte instrução para o resumo (abaixo dessa afirmação, inserimos o enunciado completo da atividade). 2. Considere que a coerência de um resumo está diretamente relacionada com aquilo que o texto-base permite acarretar dele. Com isso em mente, complete os espaços vazios do resumo a seguir, considerando a coerência dos possíveis verbos *dicendi* com o texto-base.

De forma similar à tarefa realizada pelos humanos, a “Predição automática das lacunas no resumo” também envolveu o fornecimento do texto-base e do resumo com lacunas aos modelos de linguagem, porém, utilizando *prompts* específicos para direcionar a geração do texto. A tarefa seguiu um parâmetro similar ao anterior, com humanos: foram seleciona-

dos 5 LLMs, aos quais foram oferecidos o texto-base e o resumo com lacunas. Em seguida, foi solicitado o preenchimento dessas lacunas por meio de *prompts* específicos (i.e., Seção 4.2), a fim de se analisar as ocorrências e se comparar com o resultado dos humanos.

Como se vê, a tarefa é baseada em um contexto natural de produção de texto, com o objetivo de verificar a variabilidade das respostas e a possível aproximação com a redação original. Nesse sentido, a metodologia sugere que, além das respostas coincidentes ao texto original (i.e., as quais poderíamos chamar de “exatas”), humanos e LLMs poderiam atribuir palavras bem diferentes (i.e., nem sequer sinônimas), mas que tivessem coerentes ao texto original (i.e., a ser resumido) e ao texto com lacunas (i.e., a ser preenchido). Nesse sentido, a tarefa permite uma ampla amplitude semântica, algo que justamente estamos investigando. Porém, entendemos que essa mesma tarefa impede a utilização de métricas comuns em trabalhos da área (como precisão e acurácia) como já fizemos em um trabalho similar (Braz Jr. et al. 2025). Por essa razão, propomos uma medida de aproximação por meio de *embeddings*, apresentados na discussão.

Após a coleta dos dados das duas tarefas, procedemos à análise dos resultados, comparando as predições realizadas pelos humanos com as geradas pelos modelos de linguagem. Na Seção 4.3, apresentamos uma análise detalhada dos resultados, incluindo as avaliações quantitativas e qualitativas, tanto humanas quanto automáticas, bem como a lista dos verbos *dicendi*. Essa análise nos permite avaliar a qualidade das predições e a eficácia do LLM.

Os experimentos foram realizados com 5 LLMs via chat. Cada prompt foi submetido uma única vez a cada LLM, com o chat sendo reiniciado a cada nova submissão para garantir independência nas respostas. Para estes LLMs usamos as configurações de hiperparâmetros padrão, incluindo *temperature* e *top\_p*. Esses LLMs incluem LLaMA-3.1 70B Instruct<sup>4</sup>, GPT-4o mini<sup>5</sup>, Gemini<sup>6</sup>, Sabiá-3<sup>7</sup> e Claude-3.5 Sonnet<sup>8</sup>.

Finalmente, vale dizer que este trabalho tem semelhanças com o teste de Cloze (Taylor, 1953), cujo objetivo é medir habilidade de leitura. Explicamos que, do ponto de vista da escolha feita, há uma aproximação e um distanciamento desse método. Na metodologia de Cloze, o mesmo texto lido é também aquele cujas lacunas são abertas. Isso possibilita uma verificação da compreensão textual, sobretudo da capacidade dos leitores de produzirem inferências. O julgamento pode ser feito com base na exatidão do termo que poderia estar ali ou mesmo de seus sinônimos. No presente trabalho, porém, há uma transposição do conteúdo do texto original (i.e., a ser resumido) para o texto com lacunas (i.e., a ser preenchido). Nesse caso, é uma tarefa de verificação da capacidade de os graduandos encontrarem verbos *dicendi* adequados à situação apresentada, de um para outro texto. Esses verbos podem ser muito diferentes entre si (i.e., não sinônimos), mas, ainda assim, serem adequados ao contexto, o que torna a tarefa mais complexa do que a metodologia padrão de Cloze (e mais interessante para a análise de estudantes de graduação que serão professores de Língua Portuguesa).<sup>9</sup>

<sup>4</sup> <https://huggingface.co/meta-llama/Llama-3.1-70B-Instruct/>

<sup>5</sup> <https://www.chatgpt.com>

<sup>6</sup> <https://gemini.google.com/>

<sup>7</sup> <https://www.maritaca.ai/>

<sup>8</sup> <https://claude.ai/>

<sup>9</sup> Agradecemos a um parecerista deste periódico que nos indicou a relação com o teste de Cloze. Entre diversos trabalhos sobre o tema no Brasil, mencionamos os de Abreu et al. (2017) e Cunha et al. (2020).

## 4 Experimentos e Resultados

Nesta seção, adicionamos detalhes aos dois experimentos e apresentamos os seus resultados. Ao final, analisamos e discutimos esses resultados. Antes disso, queremos apresentar o Exemplo 1 com o resumo original com suas 249 palavras, no qual foram numerados (i.e., sobrescritos) e destacados 10 verbos. A remoção desses verbos resultou na criação de lacunas que foram preenchidas com o marcador “\_\_\_\_\_” (i.e., lacuna).

### Exemplo 1 – Texto resumo com os verbos selecionados e numerados

No texto “Agrotóxicos são detectados em cera e mel de abelha” (Jornal da Unicamp 07 a 20/08/2023), Liana Coll discorre<sup>1</sup> sobre a morte expressiva de abelhas no Brasil e como os agrotóxicos podem ser os responsáveis por esse fenômeno. Para isso, Coll destaca<sup>2</sup> que as abelhas são fundamentais para a polinização de 70% das plantas no planeta e, por isso, a morte desses insetos é preocupante. Nessa lógica de atenção aos possíveis impactos, Liana Coll apresenta<sup>3</sup> a pesquisa da química Ana Paula de Souza da Unicamp, a qual aborda<sup>4</sup> a presença de agrotóxicos no mel e na cera de abelhas comuns. Nessa tese de doutorado para a Faculdade de Engenharia de Alimentos, Souza avaliou<sup>5</sup> os produtos apícolas – bioindicadores de contaminação – em vez dos insetos polinizadores, já que estes são de difícil manejo por causa do pequeno tamanho e a baixa massa. Como resultado da pesquisa, Coll expõe<sup>6</sup> que a pesquisadora conclui<sup>7</sup> que das 40 amostras de mel 6 estavam contaminadas com herbicidas e 90% das ceras analisadas continham um ou mais agrotóxicos. Além disso, Coll expõe<sup>8</sup> que Ana Paula de Souza se interessou pelo tema por causa da preocupação frente à tendência de mortes de abelhas no Brasil, fenômeno que se repete na Europa e nos Estados Unidos, já que esses insetos são indispensáveis para as plantações. Com isso, Souza afirma<sup>9</sup> que é indispensável a atenção aos níveis de contaminação do mel e da cera, já que esta é matéria prima para a indústria de cosméticos e aquele é um alimento de amplo uso. Por fim, Liana Coll apresenta<sup>10</sup> a recomendação da química Ana Paula de Souza para que as práticas agrícolas possuam mais controle no uso de agrotóxicos, a fim de evitar que esses produtos continuem prejudicando as abelhas e a biodiversidade.

Na sequência, apresentamos os detalhes dos dois experimentos realizados com o resumo anterior.

### 4.1 Preenchimento humano das lacunas

Nesta primeira tarefa, solicitamos a estudantes de Letras que prenchessem as lacunas com verbos *dicendi* coerentes com o contexto de produção. Assim, esperávamos que os verbos ali presentes estivessem de acordo com o texto-base, uma vez que precisa ser coerente com ele, mas também com o texto em lacunas, uma vez que os verbos fazem parte desse resumo. Os resultados estão expostos a seguir. O Quadro 1 mostra as predições de 33 alunos para completar as 5 primeiras lacunas do texto resumido. No cabeçalho do quadro, estão os 5 primeiros verbos removidos e, nas linhas, os verbos preditos pelos alunos. No Quadro 2 continuamos com as predições apresentando os 5 verbos restantes. Para complementar a exposição, a Imagem 2 apresenta os 15 verbos mais recorrentes dos 86 informados por humanos.

Quadro 1 – Predições humanas para as 5 primeiras lacunas no texto resumo

<b>Verbo Aluno</b>	<b>discorre<sup>1</sup></b>	<b>destaca<sup>2</sup></b>	<b>apresenta<sup>3</sup></b>	<b>aborda<sup>4</sup></b>	<b>avaliou<sup>5</sup></b>
Aluno 1	Inicia comentando	expõe	apresenta	analisou	escolheu
Aluno 2	disserta	defende	menciona	encontrou	analisa
Aluno 3	<b>discorre</b>	alertou	utilizou	relata	analisou
Aluno 4	<b>discorre</b>	aponta	<b>apresenta</b>	destaca	analisou
Aluno 5	<b>discorre</b>	indica	apresentou	estudou	investigou
Aluno 6	informa	esclarece	traz	busca analisar	pesquisou
Aluno 7	informa	alega	<b>apresenta</b>	expõe	analisa
Aluno 8	fala	explica	utiliza	apresenta	analisa
Aluno 9	fala	diz	menciona	indica	usa
Aluno 10	informa	declara	recomenda	enfatiza	indica
Aluno 11	fala	afirma	cita	indica	analisa
Aluno 12	comenta	argumenta	menciona	indica	apresenta
Aluno 13	questiona	indica	menciona	enfatiza	sugere
Aluno 14	comenta	<b>destaca</b>	mostra	relata	comenta
Aluno 15	analisou	explica	destaca	apresenta	demonstra
Aluno 16	alerta	expressa	<b>apresenta</b>	indica	analisa
Aluno 17	explica	afirma	descreve	confirma	escolhe
Aluno 18	<b>discorre</b>	explicita	cita	estuda	analisou
Aluno 19	<b>discorre</b>	demonstra	destaca	<b>aborda</b>	analisa
Aluno 20	<b>discorre</b>	expõe	relaciona	analisa	destaca
Aluno 21	discute	expõe	cita	considera	destaca
Aluno 22	disserta	aponta	cita	confirma	analisou
Aluno 23	disserta	argumenta	<b>apresenta</b>	detecta	analisa
Aluno 24	<b>discorre</b>	afirma	cita	analisa	estuda
Aluno 25	fala	afirma	menciona	indica	analisa
Aluno 26	alerta	afirma	cita	analisa	analisa
Aluno 27	<b>discorre</b>	afirma	aponta	indica	estuda
Aluno 28	<b>discorre</b>	afirma	<b>apresenta</b>	indica	estuda
Aluno 29	disserta	afirma	aponta	aponta	estuda
Aluno 30	<b>discorre</b>	ressalta	<b>apresenta</b>	indica	cita
Aluno 31	discursa	indica	<b>apresenta</b>	relata	amostra
Aluno 32	noticia	explicita	destaca	estuda	analisa
Aluno 33	disserta	explica	menciona	declara	indica

Fonte: elaboração própria.

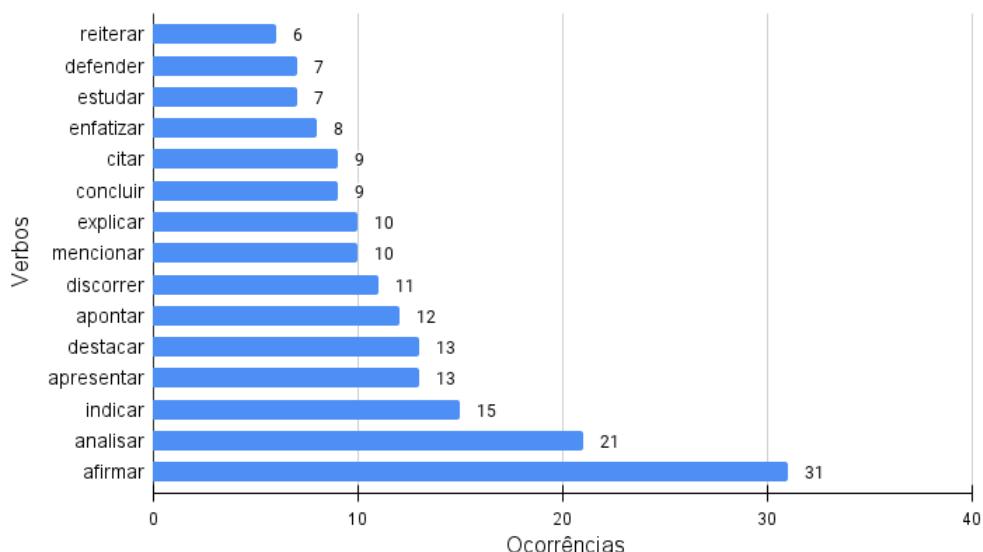
Quadro 2 – Predições humanas para as 5 últimas lacunas no texto resumo

Aluno \ Verbo	expõe <sup>6</sup>	conclui <sup>7</sup>	expõe <sup>8</sup>	afirma <sup>9</sup>	apresenta <sup>10</sup>
Aluno 1	mostra	descobriu	acrescenta	conclui	traz
Aluno 2	mostra	constatou	demonstra	defende	aconselha
Aluno 3	afirma	concluiu	aponta	constata	defende
Aluno 4	informa	traz	afirma	defende	compartilha
Aluno 5	aponta	constatou	justifica	defende	expôs
Aluno 6	constatou	concluiu	apontou	defende	dá destaque
Aluno 7	discorre	acusá	aponta	defende	indica
Aluno 8	declara	descobriu	cita	afirma	institui
Aluno 9	fala	afirma	comenta	diz	enfatiza
Aluno 10	reconhece	enfatiza	sugere	afirma	enfatiza
Aluno 11	informa	alega	explica	destaca	reforça
Aluno 12	menciona	afirma	argumenta	afirma	menciona
Aluno 13	expressa	diz	afirma	declara	reconhece
Aluno 14	explica	constata	afirma	confirma	explica
Aluno 15	<b>expõe</b>	mostra	destaca	apresenta	afirma
Aluno 16	expressa	afirma	lembra	reforça	reitera
Aluno 17	analisa	aponta	especifica	explica	indicou
Aluno 18	menciona	viu	explica	conclui	reforça
Aluno 19	ressalta	descobriu	destaca	afirma	reitera
Aluno 20	enfatiza	afirma	relata	reitera	retifica
Aluno 21	conclui	analisa	explica	afirma	menciona
Aluno 22	afirma	constata	destacou	reiterou	retificou
Aluno 23	mostra	<b>conclui</b>	destaca	adverte	perpetua
Aluno 24	diz	afirma	conta	reitera	reforça
Aluno 25	observou	demonstra	afirma	conclui	aborda
Aluno 26	destaca	confirmou	afirma	evidencia	compartilha
Aluno 27	afirma	reveia	argumenta	conclui	frisa
Aluno 28	aponta	descobriu	postula	argumenta	enfatiza
Aluno 29	compartilha	afirma	adiciona	conclui	replica
Aluno 30	aponta	diz	afirma	enfatiza	retoma
Aluno 31	disserta	expressa	afirma	resulta	faz
Aluno 32	relata	diz	fala	afirma	exalta
Aluno 33	confirma	reveia	cita	salienta	reitera

Fonte: elaboração própria.

Nos Quadros 1 e 2 ressaltamos em negrito os casos em que o estudante conseguiu preencher exatamente com o mesmo verbo do resumo original. Das 10 predições, apenas 3 delas não coincidiram em alguma linha (i.e., os verbos, “avaliou<sup>5</sup>”, “expõe<sup>8</sup>” e “apresenta<sup>10</sup>”). Isso sugere que as lacunas restantes estavam mais alinhadas com as expectativas dos alunos em termos de coerência entre contexto e verbo.

Imagen 2 – Os 15 verbos no infinitivo com maiores ocorrências preditos por humanos



Fonte: elaboração própria.

## 4.2 Preenchimento automático das lacunas

No segundo experimento, utilizamos LLMs para predição das lacunas no texto resumido. A tarefa “Predição automática das lacunas no resumo” preenche o prompt de predição das lacunas com o texto-base e resumo.

O Exemplo 2 apresenta o prompt usado para solicitar ao LLM para predição das lacunas no resumo. O LLM ao analisar este prompt deve considerar somente o texto-base compreendido entre os marcadores “####”, substituindo <TEXTO\_BASE>. A seguir a tarefa solicita o preenchimento das lacunas do texto resumido. O texto resumido com lacunas está compreendido entre os marcadores “####”, substituindo <TEXTO\_RESUMO\_LACUNA>. Em “Resposta:” será apresentado o resultado do LLM.

### Exemplo 2 – Prompt de predição das lacunas

```
Considerar como texto-base o texto que está entre os marcadores ####.
#####
"<TEXTO_BASE>"
#####
Tarefa: Você é um assistente útil responsável que completa com uma
palavra as lacunas("_____") no texto a seguir delimitado
pelos marcadores ####. As palavras que completam as lacunas devem ser
coerentes e manter o objetivo principal do texto-base.
#####
"<TEXTO_RESUMO_LACUNA>"
#####
Resposta:
```

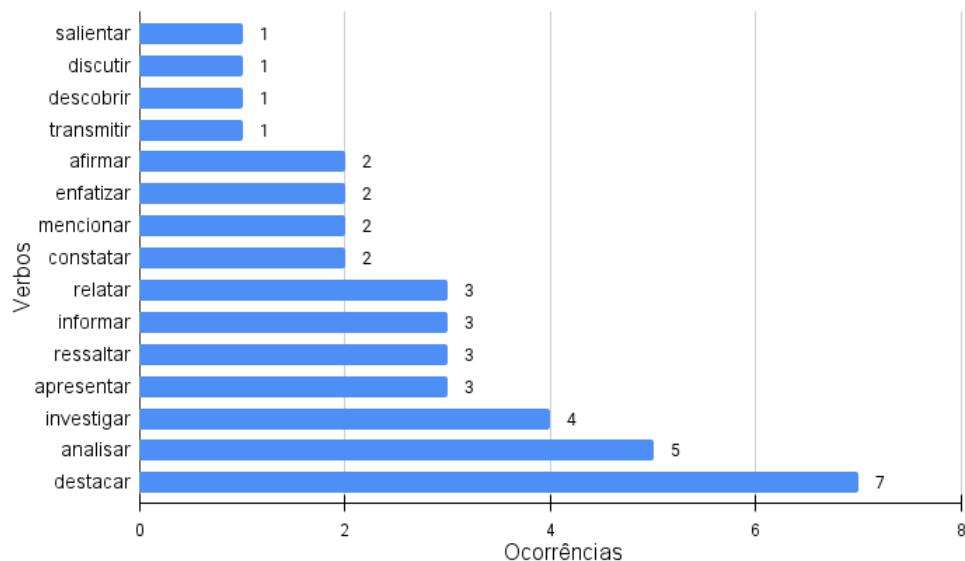
O Quadro 3 mostra as previsões de diferentes LLMs (i.e., LLaMA 3.170B Instruct, GPT-4o mini, Gemini, Sabiá-3 e Claude-3.5 Sonnet) para completar lacunas no texto resumido. A primeira coluna indica os verbos originais substituídos. As demais colunas mostram as sugestões de cada modelo de linguagem para cada lacuna. Já a Imagem 3 mostra os 15 verbos mais recorrentes dos 25 no preenchimento dos LLMs.

Quadro 3 – Previsões dos LLMs para as lacunas no texto resumo

Verbo	LLaMA 70B	GPT-4o mini	Gemini	Sabiá-3	Claude-3.5
discorre <sup>1</sup>	escreve	relata	alerta	relata	discute
destaca <sup>2</sup>	afirma*	menciona	ressalta	explica	destaca
apresenta <sup>3**</sup>	destaca**	descreve	apresenta	destaca	apresenta
aborda <sup>4</sup>	Investiga	investigou	investigou	analisou	analisou
avaliou <sup>5</sup>	analisa	analisou	analisou	utilizou	investigou
expõe <sup>6*</sup>	afirma*	destaca	informa	informa	relata
conclui <sup>7</sup>	encontrou	constatou	conclui	constatou	descobriu
expõe <sup>8*</sup>	destaca**	informa	destaca	comenta	menciona
afirma <sup>9</sup>	enfatiza	enfatiza	defende	salienta	ressalta
apresenta <sup>10**</sup>	destaca**	ressalta	reforça	apresenta	transmite

Fonte: elaboração própria.

Imagen 3 – Os 15 verbos no infinitivo com maiores ocorrências preditos por LLMs



Fonte: elaboração própria.

Observa-se que as escolhas de verbos variam entre os modelos, embora alguns padrões se repitam, como “destaca”, “informa”, “enfatiza” e “conclui”. Essas diferenças ilustram como cada modelo pode interpretar e estruturar a apresentação de ideias e ações no texto, com variações que indicam nuances nas abordagens de cada LLM para transmitir a informação ou intenção dos autores do texto-base. Além disso, ressalta-se (i.e., em negrito) que quase

todos os modelos acertaram com precisão pelo menos um verbo do texto original (i.e., com exceção do GPT-40 mini). Outro ponto importante identificado é quanto à diversidade lexical das palavras a qual indicamos com o símbolo “\*\*” e “\*\*\*” as repetições. O texto resumo possui uma repetição da palavra “expõe”, LLaMA possui duas repetições de “afirma” e três de “destaca” enquanto o Gemini, Sabiá-3 e o Claude-3.5 não possuem repetições.

#### 4.3 Discussão

Uma análise atenta dos dados apresentados anteriormente revela pontos interessantes e que dialogam com as questões centrais deste artigo. Um primeiro ponto de destaque é com relação à coerência semântica global. Quando se observam as escolhas, ainda que diferentes, a maioria delas é pertinente ao contexto do resumo (e do texto-base). Isso sugere que a tarefa de lacunas e a interpretação foi bem realizada na maioria dos casos, tanto por humanos quanto pelos LLMs. Estes, aliás, elencaram sugestões muito assertivas para o contexto (i.e., coerência global). Isso reitera a interpretação de que essas aplicações são ótimas ferramentas para tarefas precisas, como a sumarização, por sua capacidade probabilística (Paes; Freitas, 2023).

Soma-se a isso também o fato de os LLMs sugerirem verbos que de fato indicam uma atitude linguística que se encontra no texto original. As possíveis exceções são “escrever”, “analisar”, “investigar” e “encontrar”. Explicamos: o primeiro é muito genérico e, por isso, pouco representativo para o resumo; os demais parecem prioritariamente descrever uma ação realizada pela pesquisadora referida no texto-base, mais do que uma ação linguística ali presente. Ainda assim, no contexto, não são sugestões descartáveis.

As previsões humanas, por sua vez, apresentam alguns equívocos quando se considera a coerência global. É o que se vê em “comenta” (e.g., Aluno 12), que não descreve o objetivo geral do texto-base, “acusa” (e.g., Aluno 7), “retifica” (e.g., Alunos 20 e 22), “perpetua” (e.g., Aluno 23) e “amostra” (e.g., Aluno 31), todas escolhas incoerentes. Além disso, alguns verbos também são genéricos e inadequados para o contexto de resumo escolar, sobretudo em contexto de avaliação. É o caso de “dizer” e “falar”, que aparecem 6 vezes cada no corpus.

Já do ponto de vista da coerência local, ou seja, da pertinência da escolha de verbos para a cada uma das orações, os LLMs fizeram previsões mais corretas. Todas as 50 ocorrências são pertinentes ao contexto sintático (i.e., local) em que poderiam aparecer. As previsões humanas, por sua vez, apresentam resultados incoerentes, como se vê com: “comenta” (e.g., Aluno 14, verbo 5); “analisou” (e.g., Aluno 15); “faz” e “amostra” (e.g., Aluno 31).

Assim, fechando o aspecto da coerência, podemos dizer que o resultado refuta a nossa hipótese inicial de que os LLMs trariam respostas inferiores às dos humanos, sobretudo por se considerar a capacidade epilingüística destes. O que se constatam são escolhas adequadas ao contexto, em ambos os grupos analisados.

Um segundo ponto a se observar é com relação às previsões exatas, ou seja, aquelas sugestões que coincidiram com o resumo original. A Tabela 1 apresenta as quantidades e percentuais das previsões exatas dos humanos e dos LLMs para as lacunas do texto resumo. Percebe-se pelos resultados que os verbos “avaliou<sup>5</sup>”, “expõe<sup>8</sup>” e “apresenta<sup>10</sup>” foram os únicos que não convergiram. Apesar disso, é importante dizer que os dois últimos estão entre os 10 mais frequentes no corpus analisado, o que significa que são do vocabulário dos estudantes. Por outro lado, “avaliou<sup>5</sup>” não aparece na lista das 330 ocorrências, o que pode sugerir que os estudantes não o consideraram relevante no contexto do resumo.

Tabela 1 – Percentuais de acertos humanos e dos LLMs para as lacunas no texto resumo

Verbo	Humano		LLM	
	Predição exata	%	Predição exata	%
discorre1	10	30,30%	0	0,00%
destaca2	1	3,03%	1	20,00%
apresenta3	8	24,24%	2	40,00%
aborda4	1	3,03%	0	0,00%
avaliou5	0	0,00%	0	0,00%
expõe6	1	3,03%	0	0,00%
conclui7	1	3,03%	1	20,00%
expõe8	0	0,00%	0	0,00%
afirma9	6	18,18%	0	0,00%
apresenta10	0	0,00%	1	20,00%

Fonte: elaboração própria.

Uma terceira questão diz respeito à variação lexical. A maioria dos verbos sugeridos, seja pelos humanos, seja pelos LLMs, de fato é diferente do resumo original, como se vê na Tabela 1. No entanto, a olho nu, os dados de informantes parecem indicar uma maior variedade (comparam-se os dados das Tabelas 1, 2 e 3). Isso corroboraria a hipótese de que os falantes teriam um repertório mais criativo que os LLMs, por fazerem uso de seu conhecimento epilingüístico.

No entanto, quando se observam os dados proporcionais relativos à diversidade lexical, na Tabela 2, percebe-se que os LLMs superaram a variedade de sugestões verbais esperadas para o contexto. Considera-se a diversidade lexical como a razão entre os diferentes lexemas apresentados e o número total de ocorrências da palavra (Martins, 2016). Aplicando essa equação no resumo original, observa-se que ele obteve 80% de variação nos verbos *dicendi*, devido as repetições de “apresenta” e “expõe”. Já os resultados entre os informantes e os LLMs são diferentes.

Tabela 2 – Diversidade lexical das predições humanas e LLMs

DiL	Humano	LLM
100%	57,58% (19)	80,00% (4)
90%	21,21 % (7)	0,00% (0)
80%	18,18% (6)	0,00% (0)
70%	0,00% (0)	20,00% (1)
60%	3,03% (1)	0,00% (0)
Total	100,00% (33)	100,00% (5)

Fonte: elaboração própria.

Para a predição humana a análise da diversidade lexical revela que cada informante utilizou um vocabulário variado para suas predições, com poucas repetições. Isso indica que os

participantes possuem um repertório verbal relativamente amplo e que são capazes de mobilizar diferentes sinônimos para expressar a mesma ideia. A exceção é para um dos informantes (i.e., Aluno 12) que obteve 60% de diversidade lexical em suas sugestões. Ao analisarmos a Imagem 2 (ou os Quadros 1 e 2), identificamos padrões interessantes nas predições dos alunos. Verbos com o “afirmar”, “analisar”, “indicar”, “destacar” e “apresentar” receberam entre 30 e 13 ocorrências cada, sendo muito frequentes no corpus. Considerando o contexto de produção, são verbos apropriados para a proposta. Apesar disso, é preciso dizer que verbos genéricos, como o próprio “afirmar” (i.e., 30 ocorrências) quanto “discorrer”, “mencionar” e “explicar”, com 10 ocorrências cada, foram frequentemente utilizados. Essa preferência por verbos mais genéricos sugere que os participantes os percebem como mais versáteis para expressar uma variedade de ações em diferentes contextos. Tal estratégia pode ser uma forma de garantir a coerência com o texto-base, evitando escolhas lexicais mais específicas que poderiam comprometer a compreensão do texto. No entanto, a precisão das predições pode ter sido influenciada pelo contexto em que os verbos *dicendi* foram removidos. Verbos *dicendi* em contextos mais específicos ou que exigem um conhecimento mais profundo do assunto podem ter sido mais desafiadores de prever, exigindo um repertório lexical mais especializado.

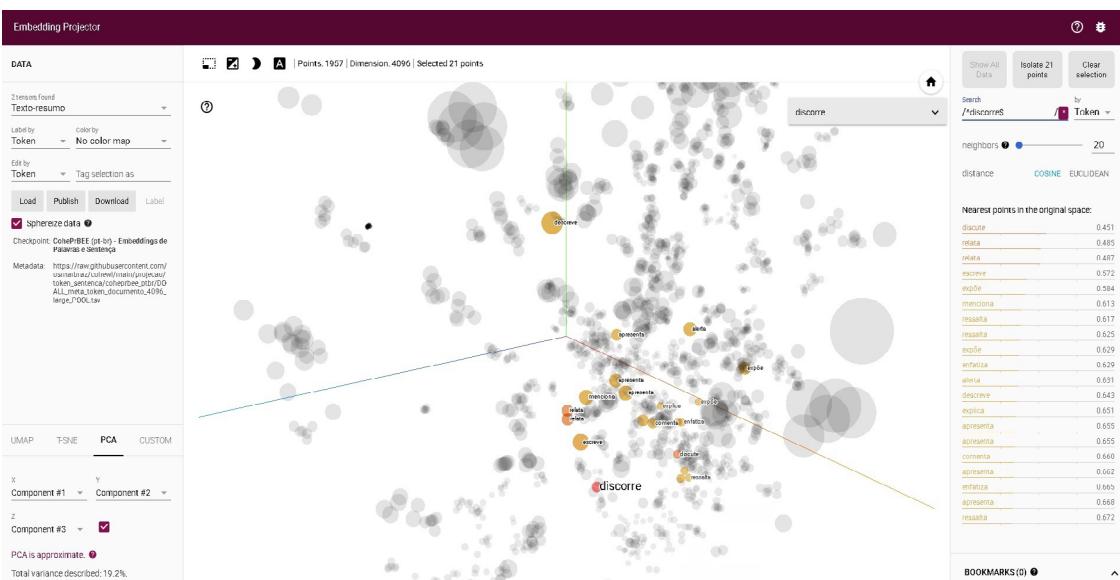
No caso dos LLMs, quatro deles tiveram 100% de diversidade, com exceção do LLaMA, que atingiu 70%. Desta forma GPT, Gemini, Sabiá-3 e Claude apresentam uma diversidade lexical considerável, sugerindo que esses modelos são capazes de gerar textos menos repetitivos, incluindo com relação ao próprio texto humano avaliado. Além disso, ressalta-se que diversos verbos que estão na lista dos mais frequentes para os LLMs (i.e., Imagem 3) são os mesmos daqueles que se vê nas predições humanas (i.e., Imagem 2), tais como “analisar”, “apresentar” e “destacar”.

Vale indicar, ainda, que quando se agrupam os verbos sugeridos pelos LLMs e pelos humanos a diferença também é bastante grande: enquanto os LLMs sugeriram 54,0% de verbos diferentes entre si (i.e., 27 verbos das 50 sugestões eram diferentes), os humanos atingiram a marca de 30,3% de variação (i.e., 100 dos 330 verbos eram distintos). Isso reforça a capacidade da máquina de relacionar conteúdos e buscar padrões também não repetitivos para as tarefas encomendadas.

A análise quantitativa apresentada nas Tabelas 1 e 2 nos fornece uma visão geral das frequências e da diversidade lexical dos verbos utilizados. No entanto, para aprofundar a compreensão das relações semânticas entre os verbos *dicendi* e identificar padrões mais sutis, realizamos uma análise qualitativa utilizando projeção de *embeddings* e medidas de distância. Tais medidas contribuem para se estipular o grau de proximidade semântica entre as palavras de um determinado contexto, no nosso caso, a proximidade dos verbos *dicendi* que foram sugeridos no experimento, numa simulação matemática da intuição sobre a relação de sinonímia entre os termos. A Imagem 4, que apresenta a projeção de componentes principais (do inglês *Principal Component Analisys* – PCA) dos *embeddings* de palavras do texto resumo e dos textos preenchidos pelos 5 LLMs. Esta projeção permite visualizar de forma mais intuitiva as relações semânticas entre os termos utilizados. Utilizamos o *Embedding Projector* (Smilkov et al., 2016) para reduzir a dimensionalidade de *embeddings* e visualizar sua projeção em um plano. Os *embeddings* das palavras foram gerados através do modelo de linguagem BERTimbau (Souza; Nogueira; Lotufo, 2020) no tamanho Large (i.e., dimensões 1.024) e no formato cased (i.e., com caracteres maiúsculos e minúsculos). Cada texto foi submetido ao LM e os *embeddings* das palavras recuperados para gerar a projeção. A palavra “discorre” foi esco-

Ihida como ponto focal, e a distância das 20 palavras vizinhas foram identificadas utilizando a similaridade do cosseno. Quanto maior o número, maior a distância do ponto focal, ou seja, menos relacionada a ela. Disponibilizamos na internet a projeção<sup>10</sup> dos *embeddings* das palavras dos textos resumos preenchidos pelos LLMs para outras análises. Nesse link, é possível consultar outras palavras, ampliar a imagem da projeção, visualizar as palavras próximas utilizando diferentes medidas de distância (i.e., similaridade do cosseno e distância euclidiana) entre outras funcionalidades.

Imagen 4 – Projeção PCA dos embeddings das palavras dos textos preenchidos pelo LLM próximas a “discorre” utilizando o Embedding Projector



Fonte: elaboração própria.

A visualização sugere que as palavras mais próximas de “discorre” estão semanticamente relacionadas ao ato de falar, apresentar ou discutir um assunto. Isso é esperado, pois “discorre” é um verbo que se refere a expor ideias de forma extensa e detalhada. Percebe-se também que as predições para a palavra “discorre” dos LLMs (i.e., “relata”, “escreve” e “alerta”) aparecem entre as 20 palavras mais próximas e nessa ordem de distância. Desta forma demonstram um agrupamento semântico coerente, uma vez que isso coincide com nossa intuição de falantes do português brasileiro (i.e., “relatar” parece de fato mais próximo de “discorrer” do que “alertar”). Tal fato sugere que os *embeddings* do BERT e a medida de similaridade do cosseno foram eficazes em capturar as relações semânticas entre as palavras.

## 5 Conclusão e Trabalhos Futuros

Este estudo investigou a tarefa de preenchimento de lacunas de verbos *dicendi* em resumos, com o objetivo de avaliar a capacidade de LLMs em simular a produção humana de texto. Ao

<sup>10</sup> [https://projector.tensorflow.org/?config=https://raw.githubusercontent.com/osmarbraz/cohewl/main/proj-cao/token\\_sentenca/config\\_cohesrbee\\_ptbr\\_token\\_sentenca.json](https://projector.tensorflow.org/?config=https://raw.githubusercontent.com/osmarbraz/cohewl/main/proj-cao/token_sentenca/config_cohesrbee_ptbr_token_sentenca.json)

comparar as predições de LLMs com as de estudantes de Letras, buscamos identificar similaridades e diferenças nas escolhas lexicais e na coerência das respostas.

A metodologia proposta envolveu a criação de um corpus de resumos com lacunas, a partir da remoção de verbos *dicendi*. Posteriormente, esses resumos foram apresentados a alunos e LLMs para que fossem preenchidos. A análise dos resultados permitiu constatar que as escolhas de ambos, humanos e aplicações foram predominantemente adequadas ao contexto. No entanto, os LLMs superaram os informantes humanos na coerência e na diversidade lexical.

Os resultados discutidos sugerem que pode haver uma maior precisão na tarefa de preenchimento de lacunas com verbos *dicendi* por parte do LLMs do que por humanos, sobretudo quando se considera a coerência semântica global ou local. De algum modo, isso refuta a hipótese de que o conhecimento epilingüístico seria um diferencial em favor das escolhas humanas, tal como se indicou no início da pesquisa. Uma possível alegação seria o fato de que os humanos sugerem verbos menos frequentes ou esperados e isso estaria diretamente relacionado à capacidade epilingüística (e.g., “reiterar” e “frisar”). Além disso, o resultado também afasta uma possível desconfiança em relação aos LLMs quanto a sua capacidade criativa, em contextos de preenchimento de lacunas, como se indicou na discussão dos modelos. O que se viu é que cumpriram melhor as mesmas tarefas indicadas a humanos.

No mesmo contexto, também indicamos que os verbos *dicendi* poderiam ser um ponto importante para a discussão da coerência nos resumos. Acreditamos que os resultados reforçam tal fato e ressaltam a importância de investigações sobre esse tema, inclusive para tarefas humanas e automáticas referentes à correção do gênero resumo escolar. Nesse ponto é importante ressaltar que o trabalho do professor de Língua Portuguesa é justamente o de apontar caminhos para a proficiência de leitura e escrita, considerando todas as potencialidades e os contextos de leitura e produção. Os verbos *dicendi* são, certamente, uma ferramenta para esse processo, uma vez que estão diretamente relacionados com a coerência global e local no resumo.

Os resultados reforçam que os LLMs são capazes de gerar textos coerentes e contextualmente relevantes (Paes; Freitas, 2023). No contexto dos resumos, o presente trabalho sugere que eles podem ser uma ferramenta de apoio no estudo, análise e comparação de textos, tais como o resumo escolar. Nesse sentido, o trabalho reforça a ideia de que é preciso se investigar com maior afinco as possíveis intersecções entre os estudos da linguagem e os novos modelos de linguagem generativa, a fim de se entender os desafios e perspectivas que eles apresentam, inclusive no que diz respeito ao ensino (Paes; Freitas, 2023; Nunes, 2024; Fernandes, 2024). A análise dos *embeddings* de palavras, por sua vez, revela um agrupamento semântico coerente entre as palavras mais próximas (e.g., com o verbo “discorrer”), indicando que os modelos de linguagem são capazes de capturar relações semânticas entre as palavras. Tal fato reforça ainda mais as potencialidades de intersecção já referidas.

Embora este estudo tenha contribuído para o avanço da compreensão sobre a capacidade linguística humana em tarefas de predição, bem como dos LLMs em simular a produção humana de texto, algumas limitações devem ser consideradas. A primeira diz respeito à análise, que se concentrou em um tipo específico de lacuna (i.e., verbos *dicendi*), o que limita a generalização dos resultados para outros tipos de construções linguísticas. Outra limitação reside na natureza estática dos dados utilizados, o que não permite capturar a dinâmica da linguagem em situações de interação real. Por fim, a avaliação do desempenho dos modelos privilegiou métricas quantitativas, o que pode não capturar nuances qualitativas da produção textual, ainda que tenhamos indicado alguns pontos. Finalmente, não foi feita uma compara-

ção entre possíveis notas a serem atribuídas aos resumos a partir do preenchimento de lacunas, com o fito de garantir a relação entre a pertinência da escolha verbal e a nota obtida no texto.

Trabalhos futuros relacionados com o preenchimento coerente de lacunas com verbos *dicendi* e os resultados apresentados neste artigo incluem: *i*) comparar os resultados obtidos com diferentes tipos de textos (e.g., científicos, literários, jornalísticos) para verificar se a dificuldade em prever os verbos varia de acordo com o gênero textual; *ii*) utilizar técnicas de PLN para identificar padrões nas predições dos alunos e para avaliar a qualidade das substituições verbais; *iii*) ampliar a lacuna a fim de verificar outras nuances de coerência global e local, ainda com o pressuposto do conhecimento epilingüístico humano; e *iv*) comparar a atribuição de notas a resumos com as lacunas, realizada por humanos e LLMs.

Finalmente, trabalhos que adaptassem a metodologia aqui empregada para o teste de Cloze também poderiam trazer ótimos resultados. Abreu *et al.* (2017) mostram que alunos de fundamental e médio fazem escolhas coerentes com relação ao sintagma (i.e., sintática e semanticamente), mas sem observar o significado na frase. Além disso, é comum se comparar a performance em relação a diferentes contextos sintáticos. Isso poderia ser checado com graduandos, em um trabalho similar ao presente, com alguns juízes que analisassem essas escolhas, em comparação aos textos originais (i.e., no caso de resumos) ou mesmo de gêneros diferentes. Tais pontos nos parecem fundamentais para uma análise sobre a capacidade interpretativa de futuros docentes de Língua Portuguesa.

## Agradecimentos

Este trabalho foi apoiado pela bolsa Universal do CNPq 2022, pela FAPESC sob o processo 2021TR1510, pela Universidade Estadual de Santa Catarina (UDESC), pelo Projeto Print CAPES-UFSC Automação 4.0 e indiretamente pelo projeto Céos, financiado pelo Ministério Público do Estado de Santa Catarina (MPSC), que tem contribuído significativamente para o aprimoramento do nosso Laboratório (LISA) e da infraestrutura de processamento de alto desempenho da UFSC. Um dos coautores obteve apoio da Fundação Araucária (PR) para realizar parte do trabalho. Agradecemos ainda aos estudantes que cederam os dados para a pesquisa e aos pareceristas, cujas observações foram fundamentais para sua versão final.

## Contribuições dos autores

Osmar de Oliveira Braz Junior: Conceitualização, Metodologia, Validação, Análise formal, Investigação, Recursos, Curadoria de dados, Redação – rascunho original, Redação – revisão e edição, Visualização. Roberlei Alves Bertucci: Conceitualização, Metodologia, Validação, Análise formal, Investigação, Recursos, Curadoria dos Dados, Redação – rascunho original, Redação – revisão e edição. Renato Fileto: Conceitualização, Metodologia, Validação, Análise formal, Investigação, Recursos, Redação – rascunho original, Redação – revisão e edição, Visualização, Supervisão, Administração do projeto.

## Referências

- ABONIZIO, H. et al. *Sabiá-3 Technical Report*. arXiv preprint arXiv:2410.12049, 2024. Disponível em: <<https://arxiv.org/pdf/2410.12049.pdf>>. Acesso em: 02 dez. 2024.
- ABREU, K. O teste de Cloze como instrumento de medida da proficiência em leitura: fatores linguísticos e não linguísticos. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 25, n. 3, p. 1767-1799, 2017. DOI: 10.17851/2237-2083.25.3.1767-1799.
- ANTHROPIC. *Introducing Claude 3.5 Sonnet*. [s.l.], 2024. Disponível em: <https://www.anthropic.com/news/clause-3-5-sonnet>. Acesso em: 25 nov. 2024.
- BICUDO, C.; HILA, C. O bom resumo em situação de vestibular. *Claraboia*, Jacarezinho, v. 2, n. 2, p. 102–114, 2015. Disponível em: <<https://core.ac.uk/download/pdf/229002452.pdf>>. Acesso em: 02 dez. 2024.
- BIRAL, J. *Operações recorrentes na produção de resumos*. 2003. Dissertação (Mestrado em Letras) – Universidade Federal do Paraná, Curitiba, 2003. Disponível em: <<https://acervodigital.ufpr.br/handle/1884/24499>>. Acesso em: 02 dez. 2024.
- BRAGAGNOLLO, R. M. A produção textual do gênero resumo escolar. In: ANTONIO, J. D.; NAVARRO, P. (orgs.). *Gêneros textuais em contexto de vestibular*. Maringá: Eduem, 2017. p. 235-251.
- BRAZ JR, O. O.; FILETO, R. Investigando coerência em postagens de um fórum de dúvidas em ambiente virtual de aprendizagem com o BERT. In: *Simpósio Brasileiro Informática na Educação*. Online. Anais do XXXII Simpósio Brasileiro Informática na Educação: SBC, 2021. P. 749–759. DOI: 10.5753/sbie.2021.217397.
- BRAZ JR, O. O.; SANCHUKI, A.; BERTUCCI, R.; FILETO, R. Verbos Dicendi como Indicadores de Coerência em Resumos: uma análise humana e automatizada. *Linguamática*, Lisboa, v. 17, n. 1, p. 3-16, 2025. DOI: 10.21814/lm.17.1.461.
- BUSNELLO, R. H. D. et al. Passos e listas na investigação do priming ortográfico. *Psicologia: Reflexão e Crítica*, Porto Alegre, v. 24, n. 2, p. 352–354, 2011. DOI: 10.1590/S0102-79722011000200017.
- CAMPOS, C. M.; RIBEIRO, J. Gêneros. In: COSTA, I. B; FOLTRAN, M. J. (orgs.). *A tessitura da escrita*. São Paulo: Contexto, 2013. p.23-44. ISBN: 978-85-7244-782-9.
- CHEN, B. et al. *Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review*. arXiv preprint arXiv:2310.14735, 2023. Disponível em: <<https://arxiv.org/pdf/2310.14735.pdf>>. Acesso em: 02 dez. 2024.
- CLARK, H. H; GERRIG, R. J. Quotations as demonstrations. *Language*, Washington, v. 66, n. 4, p. 764–805, 1990. DOI: 10.2307/414729.
- COSTA, I.; SILVA, L. P. Coerência. In: COSTA, I. B; FOLTRAN, M. J. (orgs.). *A tessitura da escrita*. São Paulo: Contexto, 2013. p. 64-81. ISBN: 978-85-7244-782-9.
- CULIOLI, A.; NORMAND, C. *Onze rencontres sur le langage et les langues*. Paris: Ophrys, 2005. p. 211-213. DOI: 10.4000/linx.479.

- CUNHA, N.; LIMA, T.; SANTOS, A.; OLIVEIRA, K. Teste de Cloze: evidência de validade por processo de resposta. *Psicologia Escolar e Educacional*, São Paulo, v. 24, p. 1-10, 2020. DOI: 10.1590/2175-35392020191537.
- DE BEAUGRANDE, R.; DRESSLER, W. U. *Introduction to Text Linguistics*. London: Longman, 1981. v. 1. DOI: 10.4324/9781315835839.
- FERNANDES, C. A autoria em textos produzidos por inteligência artificial e por alunos em uma perspectiva discursiva. *Revista da ABRALIN*, v. 23, n. 2, p. 214–235, 2024. DOI: 10.25189/rabralin.v23i2.2183.
- GEMINI TEAM *et al.* *Gemini: a family of highly capable multimodal models*. arXiv preprint arXiv:2312.11805, 2023. Disponível em: <<https://arxiv.org/pdf/2312.11805.pdf>>. Acesso em: 02 dez. 2024.
- GOMES, D. P. Coerência Textual: Analisando uma Produção de texto na Educação Básica. *Humanidades & Inovação*, Palmas, v. 7, n. 1, p. 131–142, 2020. Disponível em: <<https://revista.unitins.br/index.php/humanidadeseinovacao/article/download/1932/1337>>. Acesso em: 02 dez. 2024.
- GU, J. *et al.* *A systematic survey of prompt engineering on vision-language foundation models*. arXiv preprint arXiv:2307.12980, 2023. Disponível em: <<https://arxiv.org/pdf/2307.12980.pdf>>. Acesso em: 02 dez. 2024.
- KASNECI, E. *et al.* ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, v. 103, p. 102274, 2023. DOI: 10.1016/j.lindif.2023.102274.
- KOCH, I. G. V.; TRAVAGLIA, L. C. *A coerência textual*. 18. ed. São Paulo: Editora Contexto, 2021.
- LIU, P. *et al.* Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, New York, v. 55, n. 9, p. 1–35, 2023. DOI: 10.1145/3560815.
- MACHADO, A. R.; LOUSADA, E.; ABREU-TARDELLI, L. S. O Resumo Escolar: Uma Proposta de Ensino de Gênero. *Signum: Estudos da Linguagem*, v. 8, n. 1, p. 89–101, jul. 2005. DOI: 10.5433/2237-4876.2005v8n1p89.
- MARTINS, M. A diversidade lexical na escrita de textos escolares. *Fórum Linguístico*, Florianópolis, Universidade Federal de Santa Catarina, v. 13, n. 1, p. 1068–1082, 2016. Disponível em: <<https://dialnet.unirioja.es/descarga/articulo/6125260.pdf>>.
- MARVIN, G. *et al.* Prompt Engineering in Large Language Models. In: *International conference on data intelligence and cognitive informatics*. Singapore: Springer, 2024, p. 387–402, DOI: 10.1007/978-981-99-7962-2\_30.
- MEIRA, R. *et al.* A Analítica da Escrita para Identificação de Indicadores de Qualidade Textual. *Revista Novas Tecnologias na Educação*, v. 21, n. 2, p. 342–351, 2023. Disponível em: <<https://seer.ufrgs.br/renote/article/download/137756/90942>>. Acesso em: 02 dez. 2024.
- NUNES, P. Escrever não é útil. *Revista da ABRALIN*, v. 23, n. 2, p. 192–213, 2024. DOI: 10.25189/rabralin.v23i2.2190.
- PAES, A.; FREITAS, C. ChatGPT, MariTalk e outros agentes de conversação. *Processamento de linguagem natural: conceitos, técnicas e aplicações em português*. São Carlos: BPLN, 2023. Disponível em: <<https://repositorio.usp.br/directbitstream/2d278e81-8e44-41ee-9788-26db07aa38cd/3165894.pdf>>. Acesso em: 02 dez. 2024.
- PAIOLA, P. H. *Sumarização abstrativa de textos em português utilizando aprendizado de máquina*. 2022. Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual Paulista. São Paulo,

2022. Disponível em: <<https://repositorio.unesp.br/server/api/core/bitstreams/43581702-4eco-4cba-8048-4bb418950b5a/content>>. Acesso em: 02 dez. 2024.

RADFORD, A. et al. *Improving language understanding by generative pre-training*. [S.l.]: OpenAI, 2018. Disponível em: <<https://hayate-lab.com/wp-content/uploads/2023/05/43372bfa750340059ad87a-c8e538c53b.pdf>>. Acesso em: 02 dez. 2024.

RASSI, A. P.; LOPES, P. A. Capítulo 19 Correção automática de redação. In: CASELI, H.; NUNES, M. (orgs.). *Processamento de Linguagem Natural: conceitos, técnicas e aplicações em Português*. BPLN, 2023. Disponível em: <<https://brasileiraspln.com/livro-pln/1a-edicao/parte8/cap19/cap19.pdf>>. Acesso em: 02 dez. 2024.

SMILKOV, D. et al. Embedding projector: Interactive visualization and interpretation of embeddings. *CoRR*, abs/1611.05469, 2016. Disponível em: <<http://arxiv.org/abs/1611.05469>>. Acesso em: 02 dez. 2024.

SOTO, M.; AUGUSTO, M. R. A.; ABEND, M. C. Processamento de orações relativas de objeto direto com pronomes resumptivos: o caso do Português Brasileiro e o efeito de lacuna preenchida. *Veredas-Revista de Estudos Linguísticos*, Juiz de Fora, v. 27, n. 1, 2023. Disponível em: <https://periodicos.ufjf.br/index.php/veredas/article/view/40741/26857>. Acesso em: 02 dez. 2024.

SOUZA, C. M. *Proposta de uma abordagem para sumarização extrativa de textos científicos longos*. 2022. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2022. Disponível em: <[https://repositorio.ufmg.br/bitstream/1843/51324/1/Cinthia%20Mikaela%20de%20Souza\\_final%20%281%29.pdf](https://repositorio.ufmg.br/bitstream/1843/51324/1/Cinthia%20Mikaela%20de%20Souza_final%20%281%29.pdf)>. Acesso em: 02 dez. 2024.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In: *Brazilian Conference on Intelligent Systems*. Rio Grande, Brazil, 2020. P. 403–417. DOI: 10.1007/978-3-030-61377-8\_28.

SOUZA, R. Modelo de estrutura retórica para leitura e escrita de resumo escolar no ensino médio técnico. pt, en. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, v. 33, p. 911–943, set. 2017. DOI: 10.1590/0102-445046525302137346.

TAYLOR, W. L. Recent developments in the use of “Cloze Procedure”. *Journalism Quarterly*, v. 33, n. 1, p. 42-48, 1956. DOI: 10.1177/107769905603300106.

TOUVRON, H.; LAVRIL, T. et al. *LLaMA*: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971, 2023. Disponível em: <<https://arxiv.org/pdf/2302.13971>>. Acesso em: 02 dez. 2024.

TOUVRON, H.; MARTIN, L. et al. *LLaMA 2*: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288, 2023. Disponível em: <<https://arxiv.org/pdf/2307.09288>>. Acesso em: 02 dez. 2024.

WANG, Y.; GUO, M. A short analysis of discourse coherence. *Journal of Language Teaching and Research*, Citeseer, v. 5, n. 2, p. 460-465, 2014. DOI: 10.4304/jltr.5.2.460-465.