

IDENTIFICANDO ANÁFORAS ENCAPSULADORAS: COMPARAÇÃO ENTRE ANÁLISE AUTOMATIZADA E HUMANA

IDENTIFYING ENCAPSULATING ANAPHORAS: COMPARISON BETWEEN AUTOMATED AND HUMAN ANALYSIS

Osmar de Oliveira Braz Junior (UDESC)¹

Roberlei Alves Bertucci (UTFPR)²

Renato Fileto (UFSC)³

Resumo: As anáforas encapsuladoras constituem o foco de investigação deste estudo, destacando sua relevância para a análise e ensino da escrita. Considerando os avanços no uso de Inteligência Artificial (IA) na linguística computacional, a pesquisa justifica-se pela necessidade de integrar metodologias inovadoras na análise textual. A motivação surge da lacuna na aplicação de grandes modelos de linguagem (LLMs) para identificar e categorizar encapsulamentos descritivos e opinativos em redações. O estudo propõe uma abordagem híbrida que combina a análise humana, rica em nuances contextuais, com as capacidades e escalabilidade dos LLMs utilizando *prompts zero-shot* e *few-shot*. Os experimentos realizados com redações do Enem mostram que o uso de *prompts few-shot* melhora significativamente a identificação de anáforas encapsuladoras pelos LLMs, quando comparado a *prompts zero-shot*, mas ainda aquém do observado em análises humanas. Em suma, este trabalho busca contribuir para o avanço da pesquisa em linguística e IA, oferecendo uma nova perspectiva para análise de textos e demonstrando o potencial da combinar métodos humanos e computacionais para identificar padrões linguísticos complexos.

Palavras-chave: Anáforas Encapsuladoras; Análise Textual; Redação do Enem; Grandes Modelos de Linguagem.

Abstract: Encapsulating anaphoras are the focus of this study, highlighting their relevance for the analysis and teaching of writing. Considering the advances in the use of Artificial Intelligence (AI) in computational linguistics, the research is justified by the need to integrate innovative methodologies in textual analysis. The motivation arises from the gap in the application of large language models (LLMs) to identify and categorize descriptive and argumentative encapsulations in essays. The study proposes a hybrid approach that combines human analysis, rich in contextual nuances, with the capabilities and scalability of LLMs using zero-shot and few-shot prompts. The experiments carried out with Enem essays show that the use of few-shot prompts significantly improves the identification of encapsulating anaphoras by LLMs, when compared to zero-shot prompts, but still below that observed in human analyses. In short, this work seeks to contribute to the advancement of research in linguistics and AI, offering a new perspective for the analysis of texts and demonstrating the potential of combining human and computational methods to identify complex linguistic patterns.

Keywords: Encapsulated Anaphoras; Textual Analysis; Enem Essay; Large Language Model.

¹ Doutor em Ciência da Computação pela Universidade Federal de Santa Catarina (UFSC). Professor adjunto do Departamento de Educação Científica e Tecnológica (DECT) da Universidade do Estado de Santa Catarina (UDESC). E-mail: osmar.braz@udesc.br

² Doutor em Linguística pela Universidade de São Paulo (USP). Professor associado do Departamento Acadêmico de Linguagem e Comunicação (DALIC) da Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: bertucci@utfpr.edu.br

³ Doutor em Ciência da Computação pela Universidade Estadual de Campinas (UNICAMP). Professor associado do Departamento de Informática e Estatística (INE) da Universidade Federal de Santa Catarina (UFSC). E-mail: r.fileto@ufsc.br.

Introdução

Em seus intensos anos acadêmicos, a professora Mônica Magalhães Cavalcante realizou um frutífero trabalho na área de Linguística Textual, com pesquisas, orientações e parcerias relacionadas a temas como argumentação, coerência e ensino (Cavalcante, 2016a; Custódio Filho; Cavalcante, 2023; Cavalcante, Nascimento e Dias, 2023). Aliás, em texto recentemente publicado, Custódio Filho e Elias (2024) fazem uma excelente retomada sobre a contribuição da professora Cavalcante para a área. Contudo, é no campo da referencião que a professora deixou um imenso legado, analisando tanto os elementos básicos nos processos de retomada, quanto a importância desse fenômeno para o texto e o discurso (Cavalcante, 2016b, 2003a; Cavalcante e Santos, 2012; Cavalcante, 2015).

O presente trabalho enfoca o tema das anáforas encapsuladoras, discutido pela saudosa professora em diferentes momentos (Cavalcante, 2003a; Cavalcante, 2003b; Cavalcante e Brito, 2016). Nesse contexto, pretendemos acrescentar um elemento que certamente, mais cedo ou mais tarde, Mônica Cavalcante tomaria como um desafio para a pesquisa na área de texto: o papel das inteligências artificiais na análise de textos.

As anáforas são essenciais para a construção da tessitura textual, por sua capacidade de retomar direta (1) ou indiretamente (2) um referente no discurso, como exemplificado pelas sentenças a seguir.

- (1) *Um cachorro latiu. O cão estava bravo.*
- (2) *Tá tendo um casamento ali. Os convidados estão eufóricos.*

Enquanto em (1), “*o cão*” retoma diretamente “*um cachorro*”, em (2), “*os convidados*” retomam indiretamente “*um casamento*”.

Já as anáforas encapsuladoras (i.e., também denominadas encapsulamentos ou rótulos) são mais complexas. Além da retomada de uma parte do texto apresentado, elas ainda podem ter um caráter mais descriptivo (3) ou mais avaliativo/argumentativo (4).

- (3) *A Maria criticou o aumento do ônibus. Esse acréscimo* começa amanhã.
- (4) *A Maria criticou o aumento do ônibus. Esse absurdo* começa amanhã.

Nota-se que, em (3), “*esse acréscimo*” retoma “*o aumento de ônibus*” quase de forma direta, sem avaliar o fato. Por sua vez, “*esse absurdo*”, em (4), é uma avaliação subjetiva sobre o mesmo aumento.

Conscientes da complexidade e, ao mesmo tempo, da importância desse fenômeno para a coesão e coerência textuais, o presente trabalho foca nesse tipo de anáforas. Nesta pesquisa, nos propomos a analisar a ocorrência de encapsulamentos em redações do Enem com o auxílio de grandes modelos de linguagem (do inglês, *Large Language Models - LLMs*). Assim, nossa principal questão é: os LLMs são capazes de identificar e classificar as ocorrências de anáforas encapsuladoras em redações do Enem, tal como especialistas humanos? Além disso, ao analisar a ocorrência e as características dessas estruturas em redações do Enem, utilizando LLMs, este estudo enfatiza a contribuição delas para a qualidade da escrita.

A identificação precisa de anáforas encapsuladoras, sobretudo em grandes volumes de dados, demanda uma abordagem híbrida que combine a intuição linguística da análise humana com a capacidade de processamento dos LLMs. Esses modelos, quando treinados com dados de alta qualidade e bem elaborados, podem fornecer resultados consistentes e escaláveis. A comparação entre as duas abordagens permite identificar os pontos fortes e fracos de cada uma, o que refina os modelos de linguagem e aumenta a confiança nos resultados obtidos. A combinação dessas

perspectivas, portanto, oferece uma solução mais robusta e precisa para a identificação de anáforas encapsuladoras, contribuindo para avanços tanto na Linguística Computacional quanto na análise de textos, neste caso, para desdobramentos na Linguística Textual.

Nesta pesquisa, propomos a tarefa de identificação de encapsulamentos em um contexto avaliativo: a redação do Enem. A complexidade semântica e a variedade de formas que essas estruturas podem assumir exigem uma abordagem multifacetada. A motivação para este trabalho reside na necessidade de desenvolver ferramentas e metodologias que permitam uma análise mais precisa e eficiente desses elementos linguísticos, contribuindo para uma melhor compreensão da coesão e coerência textual e da construção do sentido, sobretudo em análise de *corpora* maiores, como o caso das redações. Além disso, os resultados desta pesquisa podem ter aplicações práticas em diversas áreas, como a educação, os estudos em linguística, a análise e a geração de texto, impulsionando o desenvolvimento de tecnologias mais inteligentes e personalizadas, a partir da intersecção entre áreas.

O objetivo principal deste estudo é investigar a eficácia da combinação de avaliação humana e LLMs na identificação de anáforas encapsuladoras em redações do Enem. Busca-se, especificamente, comparar a performance desses dois métodos, descrever os tipos de anáforas mais facilmente identificados por eles e analisar a influência de diferentes fatores, como a complexidade sintática e semântica das anáforas, na precisão da identificação. Consideramos também que os LLMs podem não ser capazes de identificar anáforas, desta forma necessitam de contexto com definições e exemplos no *prompt* — texto pelo qual se atribui uma tarefa ao LLM. Nossas hipóteses iniciais são, considerando a complexidade do fenômeno: *i*) a avaliação humana será mais precisa na tarefa de identificação dos encapsulamentos; *ii*) os LLMs, por sua vez, necessitam de informações mais robustas para realizar a mesma tarefa; *iii*) tenderão a apresentar casos não esperados; e *iv*) terão dificuldade em restringir (i.e., sintaticamente) o sintagma anafórico. Por outro lado, os LLMs, baseados em probabilidades e padrões estatísticos, tendem a escolher as opções que mais se ajustam aos dados de treinamento, podendo limitar a diversidade de respostas. Além disso, as aplicações poderão selecionar casos não correspondentes, em virtude de similaridades morfossintáticas.

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 e classificar Anáforas Encapsuladoras Descritivas (AED) e Anáforas Encapsuladoras Opinativas (AEO) em redações do Enem. Poderemos observar se, ao analisar a estrutura e o conteúdo de um texto, um LLM é capaz de reconhecer e categorizar essas anáforas, refletindo o papel delas no encadeamento e na coesão textual. Essa pesquisa se alinha a estudos anteriores que exploram o potencial das inteligências artificiais (IAs) na análise de textos e na identificação de padrões linguísticos (Paes e Freitas, 2023; Nunes, 2024), buscando aprofundar a compreensão do papel desses modelos na avaliação de aspectos discursivos, como ressaltado por Meira *et al.* (2023); Rassi e Lopes (2023). A capacidade de um LLM em classificar AED e AEO pode contribuir significativamente para o desenvolvimento de estratégias de escrita mais eficazes, uma vez que essas estruturas desempenham um papel central na organização e na construção argumentativa dos textos. Essa mesma classificação também pode ajudar pesquisadores que estão em busca de modelos de anotação automática (ou semiautomática) para análise de grande volume de dados.

Os experimentos realizados aqui mostram que o uso de *prompts few-shot* — instruções mais detalhadas e com exemplos — melhora significativamente a identificação de AEO pelos LLMs em relação ao *zero-shot* — instrução direta, sem exemplificação —, mas a identificação das anáforas encapsuladoras ainda apresenta desafios, com desempenho aquém do observado em análises humanas. 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 redações do Enem, sobretudo no ensino. Nesse viés, as contribuições do presente artigo se verificam:

- na retomada e contextualização da contribuição da professora Mônica Magalhães Cavalcante na área de Linguística Textual, com destaque para seu trabalho sobre anáforas encapsuladoras e sua relevância para coesão e coerência textuais;
- na introdução da utilização de LLMs para analisar anáforas encapsuladoras em redações do Enem, explorando uma interseção entre Linguística Textual e Inteligência Artificial;
- na proposta de uma abordagem que combina análise humana e processamento automatizado por LLMs para identificar e classificar encapsulamentos, destacando as vantagens e limitações de cada método;
- no apontamento de aplicações práticas em áreas como educação e geração automática de textos, além da contribuição para avanços teóricos em Linguística Textual e Linguística Computacional;
- na comparação de estratégias de *prompting* (i.e., *zero-shot* vs. *few-shot*) na identificação de anáforas; e
- no impacto que as IAs podem apresentar no campo da educação.

1 Fundamentos teóricos

1.1 Anáforas encapsuladoras

As anáforas são essenciais para materializar a capacidade da linguagem de estabelecer referências. Elas são entendidas como uma atividade voluntária do falante, uma vez que ele escolhe o modo de apresentar seu discurso e de estabelecer referências entre os elementos sobre os quais enuncia algo (Koch e Elias, 2017). Nessa escolha, além de serem essenciais para a manutenção da coerência do texto, por garantirem a referenciação, permitem também a adequada coesão durante a progressão do texto.

Como já se observou anteriormente na introdução deste trabalho, as anáforas se dividem em dois grandes grupos: diretas e indiretas. Repetimos a seguir os exemplos anteriormente apresentados de anáforas desses grupos e de anáforas encapsuladoras para conveniência do leitor.

- (1) *Um cachorro latiu. O cão estava bravo.*
- (2) *Tá tendo um casamento ali. Os convidados estão eufóricos.*
- (3) *A Maria criticou o aumento do ônibus. Esse acréscimo começa amanhã.*
- (4) *A Maria criticou o aumento do ônibus. Esse absurdo começa amanhã.*

As anáforas diretas (e.g., sentenças do item 1 acima) retomam elementos já apresentados no texto, enquanto as indiretas recuperam elementos no texto a partir de menções a coisas relacionadas ao elemento retomado (e.g., item 2 acima). Enquanto em (1) a relação entre os elementos em destaque é direta (i.e., por sinônímia), em (2) a coesão e a coerência do trecho são realizadas pela relação entre “*casamento*” e “*convidados*”, a partir daquilo que se conhece no mundo.

Como também já se antecipou, as anáforas encapsuladoras são estruturas de referenciação complexas, uma vez que mesclam as características indicadas acima. Isso significa que podem recuperar total ou parcialmente um referente no texto ou ter parte de sua composição indicando a recuperação e outra parte indicando uma novidade. Além disso, essas anáforas podem apresentar um caráter predominantemente descritivo (3) ou opinativo (4).

Vemos que, em (3), a anáfora em destaque (i.e., “*esse acréscimo*”) recupera de forma direta, por sinônímia, o trecho “*o aumento do ônibus*”. Por sua vez, o encapsulamento em (4) indica a recuperação indireta e híbrida de um referente: enquanto o “*esse*” aponta para um elemento já introduzido, o adjetivo “*absurdo*” é um modo de se classificar o elemento recuperado. Pelo contexto, essa recuperação (i.e., indireta) parece ser mesmo a de “*o aumento do ônibus*”.

Do ponto de vista linguístico, os encapsulamentos são entendidos como formas nominais que recategorizam, recuperam ou sumarizam partes do texto, com a finalidade de atribuir-lhes rótulos (Conte, 2003; Apothéloz, 2003; Koch, 2005; Cavalcante, 2003a). Cavalcante (2003a, p. 155), no entanto, é mais precisa ao afirmar que “encapsular consiste em resumir proposições do discurso empacotando-as numa expressão referencial, que pode ser um sintagma nominal (...), ou pode ser um pronome, geralmente demonstrativo”. Ela acrescenta que as anáforas encapsuladoras “recuperam o que há no co(n)texto”, mas, ao rotularem, produzem um novo sentido ao referente relacionado (Cavalcante, 2003a, p. 116). É por isso que Apothéloz (2003, p. 72) sugere que essa retomada pode ser a oportunidade de o falante “manifestar a sua subjetividade ou repetir a de outrem.” Isso nos leva a interpretar que a definição do autor é uma consequência do ato de encapsular: se a sua finalidade é dar um rótulo a um segmento textual, tal rótulo não é vazio ou neutro, uma vez que as escolhas linguísticas nunca o são. Tais fatos colocariam os encapsulamentos no rol dos “dêiticos discursivos”, conforme sustenta Cavalcante (2003a).

De volta aos exemplos em (3) e (4), o que se pode acrescentar é que os encapsulamentos ali presentes revelam um ponto de vista apresentado sobre o “*o aumento do ônibus*”. No primeiro caso, parece-nos que não há necessariamente um juízo de valor em relação ao aumento (i.e., apesar da crítica). O que se lê é simplesmente a descrição de quanto o referido aumento entra em vigor. Já em (4) percebe-se um juízo de valor bastante claro e negativo em relação ao aumento, caracterizado pelo uso de “absurdo”.

Sendo assim, vamos assumir, seguindo trabalho anterior (Bertucci, Malheiros e Lopes 2020), que os encapsulamentos podem ser essencialmente descritivos, como em (3), ou opinativos, como em (4). A principal diferença entre eles é que, no primeiro caso, não se observa juízo de valor em relação ao objeto retomado, enquanto no segundo isso está presente.

Vamos retomar o trabalho de Cavalcante (2003a) e acrescentar um fato importante: é muito comum que as anáforas encapsuladoras sejam compostas por pronomes demonstrativos (como nos exemplos acima). Outro ponto importante é que o próprio demonstrativo *isso* pode igualmente ser considerado um encapsulamento, ainda que não tenha a função de rotular as partes retomadas. Por tudo isso, neste trabalho, vamos enfocar os casos de anáforas encapsuladoras encabeçadas por demonstrativos, sem enfatizar, no entanto, as ocorrências com *isso*.

Em Bertucci, Malheiros e Lopes (2020), discutiu-se sobre o tipo de encapsulamento mais comum em redações no Enem. Na ocasião, os autores apontaram que, no *corpus* de 31 redações analisadas, as anáforas encapsuladoras foram algo constante nos textos de nota máxima no Enem (29 dos 31 apresentaram encapsulamentos). Neles, o tipo descritivo (79%) superou o opinativo (21%). Mas, ainda que tal classificação possa ser sustentada por uma certa “semântica” dos sintagmas, cabe ressaltar o papel sempre argumentativo dos encapsulamentos. Nesse sentido, Cavalcante e Brito (2016) defendem que

[...] as anáforas encapsuladoras, assim como qualquer processo referencial, cumprem uma função **argumentativa** importante. Com isso, estamos assumindo que todas as anáforas encapsuladoras, inclusive as pronominais, contribuem, de algum modo, para o desenvolvimento argumentativo de um texto, sobretudo porque ajudam a organizar os argumentos na arquitetura textual (Cavalcante e Brito, 2016, p. 30, grifo no original).

Tendo tal observação em mente, a seguir, trataremos da redação do Enem e do modo como as anáforas podem contribuir para a composição de tal gênero.

1.2 Redação do Enem

A escola, como mediadora dos processos de conhecimento, pode contribuir para que o estudante reflita sobre as melhores escolhas para promover uma argumentação consistente, afinal, como afirmam Antonio e Nuss (2018, p. 520), “dada a importância social dos textos argumentativos, o trabalho com esses textos merece lugar de destaque na sala de aula e em vestibulares e avaliações nacionais como o Enem”.

É preciso considerar que a redação do Enem (texto dissertativo-argumentativo) é um gênero misto, composto basicamente por um movimento em torno da dissertação, compreendida como uma análise sobre um tema, e um movimento em torno do argumento, caracterizado pelo posicionamento do autor (Coroa, 2017; Cantarin, Bertucci e Almeida, 2017). Em documentos e textos teóricos sobre a redação do Enem, o que se percebe é uma relevância em torno da progressão textual coerente, realizada por meio de uma estrutura geral apresentada (introdução, desenvolvimento e conclusão), repertório adequado ao tema, consistência na argumentação, uso adequado e variado de recursos linguísticos e uma proposta de intervenção também relacionada ao tema. Esses pontos são detalhados na orientação sobre as cinco competências que são avaliadas na redação do Enem (Brasil, 2024; Garcez e Corrêa, 2017).

Com relação à Competência IV, exclusiva para a avaliação de elementos linguísticos, a Cartilha do Participante (Brasil, 2024) deixa claro que os elementos avaliados na competência são aqueles relativos à coesão, como a referenciação, os quais são essenciais para a construção argumentativa. E, na análise de redações presentes no mesmo material, há comentários explícitos quanto aos pronomes, mas não aos encapsulamentos.

Garcez (2017) sustenta que as escolhas linguísticas são essenciais para se garantir a composição do gênero em questão, o que se observa em diferentes trabalhos empíricos sobre o tema (Garcez e Corrêa, 2017; Bertucci, 2021). Entende-se que os diferentes elementos escolhidos pelo candidato podem ser cruciais para que o avaliador reconstrua o caminho dissertativo-argumentativo do texto, avaliando-o conforme a sua compreensão em relação a essa reconstrução. Nesse sentido, é possível se questionar sobre o papel das anáforas encapsuladoras como promotoras dos movimentos de composição que o texto em tela exige. Alguns autores como Aquino (2017); Elias (2017); Val e Mendonça (2017) consideram que as anáforas têm um papel crucial em textos do Enem. Mas, apenas Elias (2017, p. 31) apresenta explicitamente o papel dos encapsulamentos, considerando o rótulo como “importante sinalizador da orientação argumentativa que o produtor imprime ao texto”.

Voltando à proposta de Bertucci, Malheiros e Lopes (2020), os autores sugerem que essas anáforas estejam relacionadas diretamente a dois subitens de avaliação da redação do Enem, mais especificamente as Competências 3 e 4: no primeiro caso, entendem que elas ajudam na organização das informações e argumentos na defesa de um ponto de vista; no segundo, relacionam as anáforas diretamente ao conhecimento de recursos linguísticos por parte do candidato. Para eles, os encapsulamentos garantem a progressão textual pela coesão, como defende a literatura, e, especificamente na redação do Enem, tem uma função de organizar o discurso em torno do tema solicitado, ajudando na composição da dissertação (análise) e da argumentação.

Por ora, vale destacar que o *corpus* desta pesquisa é composto por textos que, segundo a banca avaliadora, atingiram a excelência em cada uma das competências. Assim, poderíamos dizer que são textos “prototípicos” desse gênero avaliado, que é a redação do Enem. Nesse sentido, esta pesquisa descreve aquilo que é comum nesse gênero, no que diz respeito às anáforas encapsuladoras.

1.3 Grandes Modelos de Linguagem

Atualmente, pesquisadores afirmam que os LLMs ampliam significativamente as capacidades dos sistemas na geração de texto. Esses modelos, treinados com vastas quantidades de dados, demonstram uma notável habilidade para produzir textos que se assemelham àqueles criados por humanos (Kasneci *et al.*, 2023). Nesse cenário, diversas organizações desenvolveram LLMs, como o GPT-3 (Radford *et al.*, 2018) da OpenAI, o LLaMA (Touvron *et al.*, 2023a,b) da Meta, o Gemini (Team *et al.*, 2023) do Google, o Saibá-3 (Abonizio *et al.*, 2024) da Maritaca-AI, treinado em um grande *corpus* centrado no Brasil, e o Claude da Anthropic (Anthropic, 2024). Utilizar esses modelos com *prompts* apropriados possibilita orientar de forma eficaz a avaliação e a melhoria da coerência do texto.

Pesquisadores têm explorado formas de direcionar os LLMs para identificar e classificar textos, aproveitando o potencial desses modelos (Chen *et al.*, 2023; Marvin *et al.*, 2024). Um elemento central nesse processo é o “*prompt*”, definido como uma instrução ou entrada que orienta o modelo a gerar textos relevantes ou realizar tarefas específicas. Ao fornecer perguntas, descrições de problemas, fragmentos de texto ou outras entradas contextuais, os *prompts* moldam as respostas dos LLMs, tornando-os capazes de executar tarefas de forma direcionada (Radford *et al.*, 2018). Pesquisadores como Gu *et al.* (2023) destacam que os *prompts* permitem incluir informações adicionais ou dicas que guiam o modelo, facilitando a execução de tarefas complexas. Essa interação com os LLMs introduziu o paradigma do aprendizado baseado em *prompt* (i.e., *prompt-learning*), que segue o processo “pré-treinar, instruir e predizer” (do inglês, *pre-train, prompt, and predict*) (Liu *et al.*, 2023). Por meio de informações contextuais e exemplos relevantes, os *prompts* aprimoram a capacidade dos modelos em identificar e classificar elementos textuais, promovendo avanços significativos no processamento de linguagem natural (PLN). O presente trabalho testa o desempenho desse modelo no fenômeno das anáforas encapsuladoras em textos escritos do português brasileiro.

Apesar de suas promissoras aplicações, os LLMs ainda enfrentam desafios significativos, especialmente na identificação e classificação de textos. Esses modelos, baseados em probabilidades, muitas vezes demonstram uma criatividade limitada, o que prejudica sua capacidade de gerar respostas verdadeiramente inovadoras (Paes e Freitas, 2023). Por exemplo, ao lidar com tarefas como a identificação de anáforas, o modelo prevê probabilisticamente as opções mais adequadas com base no contexto fornecido no *prompt*. Essa abordagem, porém, pode se mostrar insuficiente quando a tarefa exige maior nuance ou originalidade, como no caso dos encapsulamentos, os quais também são sintaticamente mais complexos. A falta de criatividade dos LLMs influencia diretamente sua habilidade de interpretar textos ou situações com maior complexidade, afetando os resultados experimentais. Nesse contexto, formulamos a hipótese de que, ao aplicarmos os modelos para identificar e classificar Anáforas Encapsuladoras Descritivas (AED) e Anáforas Encapsuladoras Opinativas (AEO), a diversidade das opções geradas deverá ser inferior àquelas esperadas de humanos. Essa limitação, centrada na predição probabilística, compromete a flexibilidade necessária para tarefas que exigem maior adaptabilidade e inovação na geração de respostas.

Nosso trabalho considera a possibilidade de uma comparação entre uma análise crua, ou seja, sem apresentar ao LLM definições e exemplos do que se quer (i.e., análise intitulada “*zero-shot*”) e uma aprendizagem contextual, um processo que envolve a introdução de informações adicionais para facilitar o aprendizado em tarefas de PLN (i.e., intitulada “*few-shot*”). Este último método permite que os LLMs resolvam novos problemas sem a necessidade de treinamento adicional, utilizando instruções ou demonstrações que exemplifiquem a execução de tarefas específicas (Liu *et al.*, 2023; Gu *et al.*, 2023). Dessa forma, este estudo busca investigar as limitações dos LLMs com relação a criatividade e diversidade, explorando a aprendizagem contextual e

ampliando a capacidade dos LLMs de lidar com problemas complexos ou com o mínimo de contexto oferecido.

1.4 Trabalhos relacionados

Além das muitas contribuições que a professora Mônica Cavalcante deu aos estudos sobre o tema, também já trouxemos a contribuição teórica de diferentes autores para a análise dos encapsulamentos na linguística textual. Nesta seção, apresentamos alguns trabalhos sobre o tema que podem estar relacionados à proposta aqui apresentada.

A análise do fenômeno em redações é realizada por Cordeiro (2006), que analisa 60 redações de estudantes do terceiro ano do ensino médio. Ela observa que, nos textos dissertativos, o bom uso das anáforas contribui muito para a força argumentativa, sobretudo quando encabeçam o período. A partir de sua observação sobre a importância dos encapsulamentos na produção textual, Cordeiro (2006) entende que esse conteúdo deveria receber mais atenção no ensino, a fim de oferecer aos estudantes uma estratégia eficiente para a coesão textual, contribuindo significativamente para a argumentação. Paes (2013) destaca que as anáforas encapsuladoras contribuíram para a organização textual e discursiva (i.e., argumentativa) das redações de pré-vestibulandos. Para a autora, esse tipo de análise também contribui para estratégias de correção e orientação dos textos, em busca de oferecer melhores caminhos aos estudantes. Sobre redações nota 1000 (i.e., redações que obtiveram nota máxima), Santos (2023) pesquisa a ocorrência de encapsulamentos em seu trabalho de conclusão de curso. A autora analisa apenas 3 redações nota 1000 de 2021 e considera que é um fenômeno muito importante para o ensino de argumentação, sobretudo para o texto em questão.

Nesta pesquisa, vamos usar o trabalho de Bertucci, Malheiros e Lopes (2020) como a base de comparação para a tarefa com os LLMs. Na ocasião, os autores analisaram 31 redações nota 1000, referentes a 2018, transcritas em Felpi (2019). Os autores encontraram 116 anáforas encapsuladoras, divididas em descriptivas (80, 17%) e opinativas (19,83%). Vale destacar que eles focaram sua pesquisa em encapsulamentos encabeçados por pronomes demonstrativos (e.g., *esse*, *desse*, *tal*, etc.), sem focar no encapsulamento puramente pronominal (e.g., *isso*), ainda que esse caso tivesse relativa recorrência nos textos. Do ponto de vista da construção textual, os autores entendem que os encapsulamentos, no conjunto de textos analisado, têm uma maior função dissertativa, relacionada com a coesão necessária para a construção do texto. Essa estratégia é prioritária em relação à opinião, uma vez que as anáforas de valor mais subjetivo são menores.

Tendo isso em mente, a presente pesquisa toma como parâmetro os dados do trabalho apresentado, a fim de realizar uma tarefa similar com os LLMs. O objetivo é verificar o quanto os LLMs se aproximam da identificação realizada em Bertucci, Malheiros e Lopes (2020).

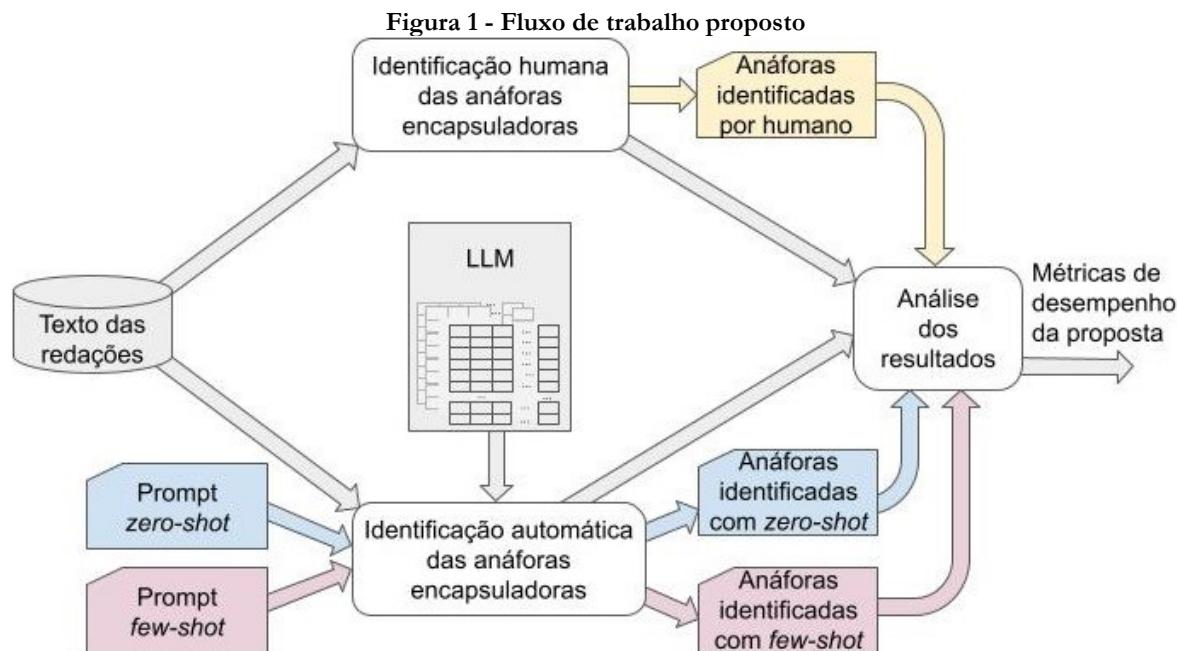
2 Metodologia

Do ponto de vista metodológico, utilizamos os mesmos 31 textos da redação do Enem de 2018, apresentados em Felpi (2019) e analisados por Bertucci, Malheiros e Lopes (2020). Além disso, também tomamos as duas categorias básicas da classificação presente no trabalho destes autores. Tabelados os dados, fizemos as verificações dos resultados.

O objetivo é avaliar a capacidade de um LLM em identificar anáforas encapsuladoras de forma automática, comparando os resultados com a análise realizada por especialistas humanos. Essa avaliação é fundamental para o desenvolvimento e aprimoramento de modelos de linguagem que possam auxiliar em tarefas de PLN, como a análise de textos e a extração de informações.

A Figura 1 ilustra o fluxo de trabalho da identificação e classificação de anáforas, desde a análise dos textos até a comparação dos resultados das identificações e classificações humanas e automáticas. Como podemos observar, o fluxo se divide em duas etapas principais: uma de análise

humana e outra de análise 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.



Fonte: Autoria própria (2024)

O processo inicia com a análise humana de um *corpus* de textos, onde especialistas identificam manualmente as anáforas encapsuladoras descritivas e opinativas, estabelecendo uma regra ouro. Em seguida, um LLM processa os mesmos textos, utilizando algum *prompt* específico para identificar automaticamente as anáforas. Utilizamos duas estratégias de *prompt*: *zero-shot* e *few-shot*. Os resultados da identificação humana são comparados com os resultados gerados pelo LLM usando as duas estratégias, permitindo avaliar a precisão e eficácia do modelo. Métricas como acurácia e precisão são calculadas para avaliar a qualidade da análise automática. A acurácia é a proporção de anáforas classificadas corretamente pelos LLMs em relação ao total de anáforas. Enquanto a precisão avalia a proporção de anáforas corretamente classificadas dentre as anáforas classificadas.

Os experimentos fazem uso de 5 LLMs. Para estes LLMs usamos as respectivas configurações de hiperparâmetros padrão, incluindo *temperature* e *top_p*. Esses LLMs incluem LLaMA-3.1 70B Instruct⁴(*temperature* = 0,75 e *top_p* = 0,9), GPT-4o Mini⁵(*temperature* = 0,7 e *top_p* = 1,0), Gemini⁶(*temperature* = 1,0 e *top_p* = 0,95), Sabiá-3⁷(parâmetros *temperature* e *top_p* não divulgados) e Claude-3.5 Haiku⁸(*temperature* = 1,0 e *top_p* não divulgado).

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

⁵ <https://www.chatgpt.com/>

⁶ <https://gemini.google.com/>

⁷ <https://www.maritaca.ai/>

⁸ <https://claude.ai/>

3 Experimentos e Resultados

Nesta seção, esmiuçamos alguns detalhes dos dois experimentos e apresentamos os seus resultados. No final, analisamos e discutimos esses resultados. Primeiramente descrevemos os dois tipos de *prompts* utilizados nas tarefas de identificar as anáforas nas redações. O *prompt* do tipo *zero-shot* os modelos não recebem exemplos ou definições prévias de treinamento e *few-shot* são expostos a exemplos e definições na tarefa. O *prompt zero-shot* usado para solicitar ao LLM identificar e classificar as anáforas encapsuladoras na redação é transscrito a seguir. O LLM, ao analisar este *prompt*, deve considerar somente o texto compreendido entre os marcadores "####", substituindo <TEXTO_BASE> pela redação a ser analisada. A seguir o *prompt* define a forma de apresentar o resultado.

Você é um assistente linguístico. Sua tarefa é identificar todas as anáforas encapsuladoras no texto dissertativo delimitado por "####". Classifique cada anáfora encontrada como "descriptiva" ou "opinativa" e forneça uma justificativa breve da sua classificação.

```
#####
"<TEXTO_BASE>"
```

Apresente suas respostas em uma lista com as seguintes informações: Anáfora, Tipo e Justificativa. Se nenhuma anáfora encapsuladora for identificada, responda: "Nenhuma anáfora encapsuladora foi encontrada."

A Tabela 1 mostra as quantidades de anáforas encapsuladoras descriptivas (D) e opinativas (O) de humanos e de diferentes LLMs (i.e., LLaMA 3.1 70B Instruct, GPT-4o mini, Gemini, Sabiá-3 e Claude-3.5 Haiku) utilizando a estratégia de *prompt zero-shot*. Na última linha da tabela, estão os totais de cada tipo de anáfora. Como as ocorrências de anáforas descriptivas foram superiores à de opinativas na maioria dos textos analisados por humanos (Bertucci, Malheiros e Lopes 2020), usamos o "*" para sinalizar as anáforas encapsuladoras opinativas identificadas em maior quantidade que as descriptivas por redação. Os percentuais deste quantitativo podem ser observados na Tabela 3 para o especialista humano e na Tabela 4 para os LLMs.

Tabela 1 - Encapsulamentos identificados por humano e LLMs usando zero-shot

#Texto	Humano		LLaMA		GPT 4o		Gemini		Sabiá-3		Claude	
	D	O	D	O	D	O	D	O	D	O	D	O
1	8	0	3	2	2	4*	4	4	9	0	2	2
2	2	2	7	0	1	3*	0	0	3	3	1	4*
3	3	4*	3	2	4	2	0	1	3	2	2	2
4	4	3	5	0	3	3	3	2	5	0	2	3*
5	1	0	3	0	6	3	0	0	2	1	2	2
6	0	0	6	0	3	2	4	1	8	1	2	2
7	2	0	0	8*	8	4	0	0	12	0	1	6*
8	2	1	5	0	8	3	0	8*	4	1	2	3*
9	2	0	5	0	4	7*	2	2	4	5*	1	4*
10	6	0	7	0	8	4	5	0	4	4	2	3*
11	2	0	2	0	2	2	9	1	2	2	2	2
12	0	0	4	1	2	3	4	0	2	3*	1	2*
13	3	1	7	0	2	2	0	6*	0	0	1	3*
14	2	0	0	0	3	3	7	9*	0	0	1	3*
15	2	0	7	0	4	4	1	1	0	0	3	3
16	3	0	1	2	4	2	5	2	6	3	1	3
17	4	2	7	1	4	1	12	0	2	1	1	1
18	1	0	6	0	3	4*	1	5*	4	2	2	3*
19	5	0	0	0	4	4	7	1	0	0	1	2*
20	5	3	0	5*	4	2	10	0	0	0	3	3
21	2	0	2	0	3	2	6	1	0	0	2	3*
22	2	1	0	0	3	1	3	4*	2	9*	2	2

23	1	0	7	5	3	2	5	11*	0	0	2	2
24	6	1	1	4*	4	2	13	5	4	1	4	1
25	5	0	0	0	4	2	3	0	0	0	2	1
26	4	0	0	7*	1	6*	4	0	0	0	1	2*
27	5	3	8	0	3	3	2	2	0	0	1	3*
28	2	1	0	0	5	0	4	2	15	0	1	3*
29	3	0	0	0	2	3*	13	0	0	0	2	2
30	5	1	2	7	3	2	7	0	1	2*	3	2
31	1	0	5	0	10	0	12	0	0	0	2	2
Total	95	23	103	44	120	85	146	68	92	40	55	79

Fonte: Autoria própria (2024)

O *prompt* da estratégia *few-shot* descrito a seguir é semelhante ao anterior (i.e., *zero-shot*), mas possui a inclusão de definições e exemplos de anáforas encapsuladoras descriptivas e opinativas. Estas definições e exemplos têm o objetivo de ensinar o contexto da tarefa ao modelo.

Você é um assistente linguístico. Sua tarefa é identificar todas as anáforas encapsuladoras no texto dissertativo delimitado por "####". Classifique cada anáfora encontrada como "descriptiva" ou "opinativa" e forneça uma justificativa breve da sua classificação.

####

"<TEXTO_BASE>"

####

Definições: Anáforas encapsuladoras (ou encapsulamentos), são estruturas de referênciação complexas. As anáforas são estruturas que referenciam um trecho anterior do texto, recuperando-o total ou parcialmente. As anáforas encapsuladoras podem ser classificadas como descriptivas e opinativas. As anáforas descriptivas recuperam o trecho anterior de maneira objetiva, sem expressar opinião. E as anáforas opinativas recuperam o trecho anterior de forma subjetiva, expressando uma opinião.

Exemplos:

Exemplo 1 anáfora encapsuladora descriptiva:

Texto: "A Maria criticou o aumento do ônibus. Esse acréscimo começa amanhã."

Anáfora: "Esse acréscimo"

Classificação: Descriptiva

Justificativa: Recupera "o aumento do ônibus" de forma objetiva.

Exemplo 2 anáfora encapsuladora opinativa:

Texto: "A Maria criticou o aumento do ônibus. Esse absurdo começa amanhã."

Anáfora: "Esse absurdo".

Classificação: Opinativa

Justificativa: Recupera "o aumento do ônibus" expressando uma opinião. Apresente suas respostas em uma lista com as seguintes informações: Anáfora, Tipo e Justificativa. Se nenhuma anáfora encapsuladora for identificada, responda: "Nenhuma anáfora encapsuladora foi encontrada."

A Tabela 2 é semelhante à Tabela 1, mas com resultados dos LLMS utilizando a estratégia *few-shot*. Repetimos os dados de humanos para facilitar comparações com os LLMs.

Tabela 2 - Encapsulamentos identificados por humano e LLMS usando *few-shot*

#Texto	Humano		LLaMA		GPT 4o		Gemini		Sabiá-3		Claude	
	D	O	D	O	D	O	D	O	D	O	D	O
1	8	0	9	4	3	5*	10	0	2	5*	6	1
2	2	2	5	5	3	2	6	0	12	1	1	4*
3	3	4*	6	2	3	5*	5	1	1	2*	4	2
4	4	3	3	8*	3	5	8	3	6	1	4	4
5	1	0	1	4*	2	3*	3	3	6	3	1	4*
6	0	0	3	0	2	11*	2	2	7	0	4	1
7	2	0	8	0	14	0	10	0	5	1	0	5*
8	2	1	2	5*	12	1	7	2	8	7	2	3*
9	2	0	17	0	10	1	3	0	0	2*	4	1
10	6	0	12	1	1	6*	1	6*	0	5*	0	5*

11	2	0	3	0	3	6*	0	2	2	2	2	3*
12	0	0	1	5*	3	3	9	0	2	0	2	3*
13	3	1	8	1	8	3	8	1	4	1	2	3*
14	2	0	7	0	3	4*	5	0	12	6	1	3*
15	2	0	8	0	6	2	3	3	5	0	2	3*
16	3	0	4	2	6	2	3	3	5	0	2	3*
17	4	2	4	0	6	1	3	1	4	2	1	2*
18	1	0	1	3*	12	0	1	3	4	4	3	2
19	5	0	8	0	11	0	8	0	3	7*	2	3*
20	5	3	9	0	5	1	6	1	3	3	0	6
21	2	0	6	3	4	3	12	1	2	3*	0	5
22	2	1	8	1	10	2	4	1	1	6*	1	3
23	1	0	5	2	8	0	4	2	2	5*	2	2
24	6	1	11	0	9	1	11	1	9	3	2	6*
25	5	0	7	2	5	0	6	3	6	2	2	6*
26	4	0	28	0	10	0	6	0	1	5*	4	1
27	5	3	2	7*	7	1	9	0	2	7*	3	0
28	2	1	1	5*	2	4*	1	3*	4	4	0	5*
29	3	0	8	0	6	0	5	0	5	7*	0	4*
30	5	1	4	5*	7	1	9	1	1	10*	0	6*
31	1	0	8	2	8	2	12	0	11	0	2	2
Total	95	23	207	67	192	75	183	41	133	109	60	99

Fonte: Autoria própria (2024)

Ao compararmos as Tabelas 1 e 2, evidencia-se o impacto do uso de *prompts few-shot* na identificação de AED e AEO pelos LLMs. Na Tabela 2, observa-se um aumento consistente na quantidade de anáforas opinativas marcadas com “*”, indicando que os modelos identificam mais casos nos quais as anáforas opinativas superam as descriptivas em textos individuais. Por exemplo, o GPT-4o e o Claude apresentaram maior frequência de opinativas destacadas na Tabela 2 em comparação com a Tabela 1, como nos textos 7 e 10. Além disso, modelos como o LLaMA e o Sabiá-3 demonstraram aumento significativo no número total de encapsulamentos detectados em vários textos, sugerindo que os exemplos fornecidos no *few-shot* ajudam os modelos a reconhecerem padrões mais complexos. Esses resultados reforçam a vantagem do contexto adicional para melhorar o desempenho dos LLMs em tarefas linguísticas desafiadoras.

A Tabela 3 sintetiza resultados obtidos por especialistas humanos na identificação de encapsulamentos AED e AEO. A segunda coluna apresenta as quantidades e os percentuais (entre parênteses) de textos em que não foram observadas as anáforas em questão (i.e., Quantidades de textos com 0 (zero) anáforas). A terceira coluna lista as quantidades totais e respectivos percentuais (entre parênteses) de anáforas identificadas pelos humanos. Estas quantidades e percentuais são base para comparação com o desempenho dos LLMs.

Tabela 3 - Quantidade de anáforas AED e AEO identificadas por humano

Anáfora	Quantidades de textos com 0 (zero) anáforas	Quantidades totais de anáforas nos textos
AED	2 (6,45%)	93 (80,17%)
AEO	19 (61,29%)	23 (19,83%)

Fonte: Autoria própria (2024)

A Tabela 4 sintetiza resultados obtidos por LLMs na identificação de encapsulamentos AED e AEO utilizando as estratégias *zero-shot* e *few-shot*. Cada entrada inclui as quantidades e os percentuais (i.e., entre parênteses) de textos em que não foram observadas as anáforas em questão (i.e., Quantidades de textos com 0 (zero) anáforas) e as quantidades totais e respectivos percentuais (entre parênteses) de anáforas identificadas pelos LLMs considerando cada estratégia.

Tabela 4 - Quantidade de anáforas identificadas por LLMs usando *zero-shot* e *few-shot*

LLM	Anáfora	<i>Zero-shot</i>		<i>Few-shot</i>	
		Quantidades de textos com 0 (zero) anáforas	Quantidades totais de anáforas nos textos	Quantidades de textos com 0 (zero) anáforas	Quantidades totais de anáforas nos textos
LLaMA	AED	9 (29,03%)	103 (70,07%)	0 (0,00%)	207 (75,55%)
	AEO	20 (64,52%)	44 (29,93%)	12 (38,71%)	67 (24,54%)
GPT 4o	AED	0 (0,00%)	120 (58,54%)	0 (0,00%)	192 (71,91%)
	AEO	2 (6,45%)	85 (41,46%)	7 (22,58%)	75 (28,09%)
Gemini	AED	6 (19,35%)	146 (68,22%)	1 (3,23%)	183 (81,70%)
	AEO	12 (38,71%)	68 (31,78%)	11 (35,48%)	41 (18,30%)
Sabiá-3	AED	12 (38,71%)	93 (69,70%)	2 (6,45%)	133 (54,96%)
	AEO	16 (51,61%)	40 (30,30%)	4 (12,90%)	109 (45,04%)
Claude	AED	0 (0,00%)	55 (41,04%)	7 (22,58%)	60 (37,74%)
	AEO	0 (0,00%)	79 (58,96%)	1 (3,23%)	99 (62,26%)

Fonte: Autoria própria (2024)

Ao analisarmos a Tabela 4, percebemos que os LLMs apresentam maior tendência em identificar anáforas AEO do que AED. Tratemos primeiro dos casos *zero-shot*. No que diz respeito ao índice de redações sem anáfora (i.e., Quantidades de textos com 0 (zero) anáforas), o modelo que mais se aproxima dos índices humanos é o LLaMA. Já o modelo Sabiá-3 se destaca com 38,71% de AED e 51,61% de AEO, índices altos para os dois tipos de anáfora, o que não se verifica com os humanos. Na direção oposta, o desempenho do Claude é notavelmente baixo neste cenário, com 0% em ambos os tipos, resultado similar ao do GPT-4o. O Gemini teve um resultado intermediário em relação aos humanos. A identificação de anáforas (i.e., Quantidades totais de anáforas nos textos), na estratégia *zero-shot* apresenta resultados melhores. O LLaMA obteve índices mais próximos da identificação humana, seguido por Sabiá-3 e Gemini. GPT-4o e Claude apresentaram percentuais muito mais distoantes do padrão humano.

Com relação à estratégia *few-shot*, percebe-se que a adição de exemplos melhora a identificação de anáforas, inclusive as AEO em modelos como o GPT-4o e o Gemini. Quando são observados os dados de redações sem anáforas (i.e., Quantidades de textos com 0 (zero) anáforas), percebe-se uma melhora significativa dos modelos. O Sabiá-3 acertou o índice de AED, enquanto Gemini chegou perto. LLaMA e GPT-4o também melhoraram seus índices em relação ao *zero-shot*. O Claude, por sua vez, piorou o percentual em relação à mesma estratégia. As AEO, por outro lado, ainda são um desafio para os LLMs, pois nenhum modelo chegou perto dos 61,29% dos casos anotados por humanos. O índice geral de identificação das anáforas (i.e., Quantidades totais de anáforas nos textos), por outro lado, apresentou resultados muito mais interessantes para a maioria dos modelos. Gemini obteve resultados praticamente iguais aos humanos; o LLaMA também foi bem, seguido pelo GPT-4o, com um índice próximo. Por outro lado, Sabiá-3 e Claude tiveram desempenhos muito abaixo dos demais.

No entanto, a identificação de AED continua sendo um desafio, com taxas próximas de zero para a maioria dos modelos, exceto o Claude (22,58%). Gemini e Sabiá-3 apresentam os melhores desempenhos gerais, mostrando potencial em ambos os cenários. GPT-4o e Claude têm desempenhos inconsistentes, indicando necessidade de ajustes ou treinamento adicional para lidar com AEDs. Os resultados indicam que os LLMs possuem limitações na identificação de anáforas AED, especialmente no cenário *zero-shot*. O cenário *few-shot* melhora o desempenho em anáforas AEO, mas não resolve completamente as dificuldades.

A Tabela 5 apresenta o desempenho de diferentes LLMs (i.e., LLaMA 3.1 70B Instruct, GPT-4o mini, Gemini, Sabiá-3 e Claude-3.5 Haiku) na identificação de anáforas descritivas (AED) e opinativas (AEO), utilizando estratégias de *prompt zero-shot* e *few-shot*, comparados a desempenho humano. Nessas estratégias, a acurácia (Ac.) mede a proporção de anáforas corretas em relação ao

total de tentativas, enquanto a precisão (Prec.) avalia a proporção de anáforas corretas dentre as respostas fornecidas. Tais métricas são básicas quando se trata de aprendizado de máquina, pois sugerem o mínimo de confiabilidade e precisão dos modelos.

Tabela 5 - Desempenho dos LLMs considerando classificações de anáforas AED e AEO feitas por humanos como regra ouro

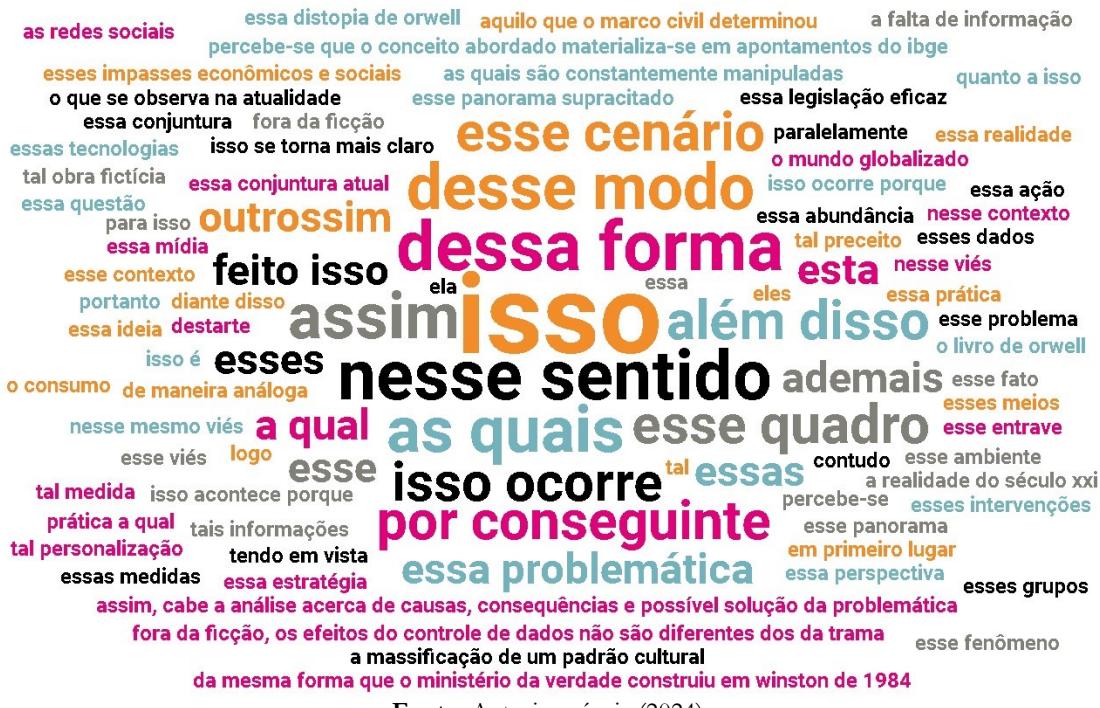
LLM	Zero-shot				Few-shot			
	AED		AEO		AED		AEO	
	Ac.	Prec.	Ac.	Prec.	Ac.	Prec.	Ac.	Prec.
LLaMA	1,81%	5,68%	0,00%	0,00%	9,01%	11,86%	4,52%	5,48%
GPT 4o	15,10%	23,52%	0,00%	0,00%	10,51%	17,00%	6,32%	6,56%
Gemini	7,35%	9,59%	0,00%	0,00%	9,24%	14,51%	10,75%	16,13%
Sabiá-3	2,51%	4,07%	3,71%	4,84%	15,15%	21,76%	6,56%	11,29%
Claude	12,29%	30,91%	4,62%	6,18%	10,28%	15,59%	6,21%	7,63%

Fonte: Autoria própria (2024)

Os resultados da Tabela 5 indicam variações significativas no desempenho dos modelos, tanto em AED quanto em AEO, com percentuais bastante baixos. No cenário *zero-shot*, GPT-4o obteve a maior acurácia e precisão para AED (i.e., 15,10% e 23,52%), mas não identificou corretamente AEO. O modelo Claude demonstrou melhor desempenho geral em *zero-shot* para AEO (i.e., 4,62% Ac. e 6,18% Prec.). Em *few-shot*, o modelo Sabiá-3 apresentou os melhores resultados de acurácia e precisão para AED (i.e., 15,15% e 21,76%), enquanto o modelo Gemini se destacou em AEO (i.e., 10,75% Ac. e 16,13% Prec.). Isso demonstra que os LLMs respondem de forma distinta às configurações de *prompts* e às categorias de anáforas.

Complementando essa análise, a Figura 2 ilustra a nuvem das 162 AED identificadas incorretamente pelo GPT-4o no conjunto das 31 redações, levando em conta a análise humana (Bertucci, Malheiros e Lopes 2020). Dentre as ocorrências destacamos que a anáfora “isso” apareceu 20 vezes, seguida de “nesse sentido” e “dessa forma”, ambas com 5 ocorrências. Outras expressões como “as quais”, “assim” e “desse modo” também se repetiram, com 4 ocorrências cada. Algumas anáforas foram incorretamente identificadas como expressões longas, como por exemplo: “da mesma forma que o ministério da verdade construiu em winston de 1984” (na base da Figura), que ocorre na redação 16. É importante observar que, olhando apenas as expressões na nuvem, algumas parecem absolutamente corretas, como “esse cenário”, por exemplo. No entanto, a classificação como “incorrecta” está baseada na análise humana, que anotou a anáfora como “nesse cenário” (i.e., ver exemplo da redação 13 do *corpus*). Esses padrões sugerem desafios específicos na identificação correta de anáforas descritivas por parte dos LLMs. Decidimos ilustrar apenas o caso com o ChatGPT, por ser o LLM mais conhecido. Também decidimos focar nas AED por serem mais numerosas no *corpus* e gerarem uma nuvem mais complexa.

Figura 2 - Nuvem de AED identificadas incorretamente pelo GPT-4o



Fonte: Autoria própria (2024)

Com os resultados apresentados, observou-se que os LLMs possuem limitações significativas, embora o uso de exemplos tenha melhorado alguns aspectos de seu desempenho. A seguir, discutiremos de forma crítica as implicações desses resultados, avaliando as diferenças no desempenho humano e dos LLMs, os desafios enfrentados e os possíveis caminhos para avanços na tarefa.

4 Discussão

O primeiro ponto que já indicamos e podemos reforçar nesta seção é com relação à diferença de desempenho dos modelos quando se passa da estratégia *zero-shot* para *few-shot*. Como se viu na seção anterior, a inclusão de uma explicação e um contexto maior favoreceu a identificação das anáforas encapsuladoras pelos LLMs. Este ponto leva-nos a considerar que, sem dúvida, é preciso que os profissionais que desejam utilizar as ferramentas de inteligência artificial estejam bem preparados para tirar o máximo de proveito do modelo.

O segundo ponto a ser destacado é com relação à diferença entre os índices gerais e às métricas de avaliação dos modelos. Quando se observam os dados na Tabela 4, verifica-se que o percentual de identificação geral dos três primeiros modelos é bastante alto e próximo ao humano. Por outro lado, a Tabela 5 indica a baixa acurácia e precisão, métricas muito relevantes para a avaliação de modelos de classificação, como o que se verifica aqui. Esse ponto nos indica que é sempre preciso um cuidado grande quando se observam índices gerais que comparam atividades humanas e aquelas desempenhadas de forma similar por LLMs.

Nesse mesmo contexto, vale a pena indicar alguns casos bastante interessantes de erros e acertos na análise dos modelos. Quanto aos erros, destaca-se que os três melhores modelos no índice geral (i.e., LLaMA, Gemini e GPT-4o) não acertaram nenhuma anáfora opinativa. Mesmo casos como “esse assédio comercial” ou “essa problemática” não foram identificados. Por outro lado, tanto o Sabiá-3 quanto o Claude identificaram pelo menos um desses casos. No caso dos acertos, é importante observar que eles foram melhores nas descriptivas. Encapsulamentos como

“esse cenário” e “essa realidade” foram os mais corretamente identificados pelos modelos. Outro ponto a se destacar é a grande quantidade de identificação de encapsulamentos pronominais puros, como “isso”, realizada pelos LLMs. Do ponto de vista empírico, daquilo que esta pesquisa pretendia, esses casos não foram considerados. No entanto, do ponto de vista teórico, não se deve considerar um erro do modelo, mas um potencial para identificação e anotação de anáforas nos textos.

É importante destacar ainda que houve muita identificação de conectivos como encapsulamentos (e.g., “assim”, “portanto”). Além disso, houve algumas identificações consideradas imprecisas, tais como: “Nesse cenário, a falsa liberdade de escolha e a padronização dos pensamentos emergem como empecilhos para a manutenção de uma sociedade dialógica e igualitária” (i.e., Gemini, redação 13). Observa-se que, nesse caso, a simples indicação de “nesse cenário” levaria o modelo a acertar a identificação, seguindo o padrão humano. Logo, podemos concluir que a complexidade sintática, que caracteriza o tipo de anáfora em questão, interfere no processo de identificação. Apesar disso, dado que alguns erros podem ser facilmente corrigidos com supervisão (ver comentário sobre a Figura 2), o modelo apresenta potencial para contribuir com a anotação de *corpus*, tal como se questionava em relação ao potencial.

Em todos os casos vistos, no entanto, o que parece ficar evidente é que os LLMs precisam de muito mais informações sobre o que são as anáforas encapsuladoras para que possam proceder a uma correta identificação. Considerando o que se apresentou na introdução deste trabalho, podemos afirmar que a maior facilidade de identificação de anáforas ficou com o demonstrativo “isso” e a maior dificuldade com as anáforas opinativas. Assim, avaliar a sutileza da subjetividade das anáforas pode ser um ponto fraco dos modelos. De qualquer modo, apesar dos muitos problemas apontados, percebe-se que os LLMs têm um potencial como ferramenta de apoio para a linguística.

Conclusão e Trabalhos Futuros

Este estudo investigou a capacidade de LLMs em identificar automaticamente anáforas encapsuladoras em textos dissertativos, comparando os resultados com análises realizadas por especialistas humanos, e explorando uma interseção entre Linguística Textual e Inteligência Artificial. A pesquisa explorou duas estratégias de *prompt*, *zero-shot* e *few-shot*, visando avaliar o impacto do fornecimento de contexto adicional no desempenho dos modelos.

Quando se analisam as hipóteses apresentadas, a primeira delas, relativa à maior qualidade das avaliações humanas, foi corroborada. O trabalho mostrou uma clara dificuldade dos LLMs na identificação precisa de encapsulamentos, quando se compara o desempenho deles com o de humanos na tarefa. Tal fato pode ser resultado de uma dificuldade em se precisar as restrições e condições de referência desse complexo fenômeno das línguas humanas, ou ao menos de se criar condições objetivas sobre elas para apresentar *prompts* adequados aos LLMs. Além disso, os resultados obtidos demonstram que os LLMs apresentam um potencial significativo para auxiliar na identificação de anáforas, especialmente quando são fornecidos exemplos e definições prévias (i.e., *few-shot*). No entanto, os modelos ainda enfrentam desafios na distinção entre anáforas descritivas e opinativas, bem como na identificação precisa de anáforas mais complexas. Esses pontos corroboram as hipóteses de que os modelos apresentariam casos inesperados e teriam dificuldade com a complexidade sintática das anáforas. A análise também revelou que a escolha do modelo e da estratégia de *prompt* influencia significativamente o desempenho na tarefa. Este fato corrobora a hipótese de que os LLMs necessitam de informações robustas para a tarefa. Acrescenta-se que o trabalho revela as muitas lacunas apresentadas pelos LLMs na análise de texto, o que sugere uma contínua pesquisa e aplicação das IAs relacionadas à Linguística Computacional e ao Ensino de Língua Portuguesa.

As limitações deste estudo incluem a pequena quantidade de textos analisados, a subjetividade na classificação de anáforas por parte dos especialistas e a complexidade inerente à identificação automática de elementos linguísticos. Além disso, a pesquisa se concentrou em um tipo específico de texto dissertativo, o que limita a generalização dos resultados para outros gêneros textuais. Outro ponto limitante é que os resultados indicam que solicitar aos LLMs a anotação de dados sem supervisão humana (e.g., para um banco de dados) pode comprometer a qualidade da anotação. Por outro lado, uma anotação supervisionada poderia ser testada, com potencial de facilitar o trabalho humano.

Um possível e interessante trabalho a ser feito, com vistas a se explorar o que se diz anteriormente, é o teste com o preenchimento de lacunas. A partir de um texto apresentado, pode-se deixar lacunas de encapsulamentos para que os LLMs apresentem sugestões de preenchimento. Tal tarefa pode jogar mais luz sobre a capacidade desses sistemas lidarem com as peculiaridades da linguagem humana. Outros trabalhos futuros relacionados com a identificação e classificação de anáforas encapsuladoras descriptivas e opinativas e os achados 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 identificar anáforas descriptivas e opinativas varia de acordo com o gênero textual; e *ii*) a criação de conjuntos de dados anotados de forma mais consistente e detalhada pode contribuir para o desenvolvimento de modelos mais precisos.

Por fim, é essencial pontuar que o presente trabalho é uma forma de homenagear e reiterar o incalculável legado deixado pela professora Mônica Magalhães Cavalcante, cuja memória se encapsula no rol das mais importantes para a área no Brasil.

Agradecimentos

Este trabalho foi apoiado pela bolsa Universal do CNPq 2022, pela FAPESC sob o processo 2021TR1510, pela Universidade Estadual de Santa Catarina (UDESC) 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.

Referências

- ABONIZIO, Hugo *et al.* Sabiá-3 Technical Report. *arXiv preprint arXiv:2410.12049*, 2024.
- ANTHROPIC. *Introducing Claude 3.5 Sonnet*. 2024. <https://www.anthropic.com/news/clause-3-5-sonnet>. Acesso em: 25 nov. 2024.
- ANTONIO, Juliano Desiderato.; NUSS, Virginia Maria. Relações retóricas e tipos de argumentos: um estudo das correlações possíveis. *Estudos Linguísticos*, 47(2), p. 519–531, 2018.
- APOTHÉLOZ, Denis. Papel e funcionamento da anáfora na dinâmica textual. In: *Referenciação*. São Paulo: Contexto, 2003. P. 53–84.
- AQUINO, Maria de Fátima de Souza. coesão na tessitura textual: avaliação do emprego de recursos coesivos. In: TEXTOS dissertativo-argumentativos: subsídios para a qualificação de avaliadores. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2017. P. 213–220.

BERTUCCI, Roberlei Alves. Propriedades linguísticas da redação Enem: uma análise computacional. *Revista de Estudos da Linguagem*, v. 29, n. 2, p. 999–1032, out. 2021. DOI: 10.17851/2237-2083.29.2.999-1032.

BERTUCCI, Roberlei Alves; MALHEIROS, Andréa Jacqueline; LOPES, Wanderlei de Souza. Ocorrências de anáforas encapsuladoras em redações do Enem. *Filologia e Linguística Portuguesa*, Universidade de São Paulo, v. 22, n. 1, p. 81–102, 2020. DOI: 10.11606/issn.2176-9419.v22i1p81-102.

BRASIL. *A redação do Enem - Cartilha do Participante*. Relatório técnico, Ministério da Educação: Instituto Nacional de Estudos e Pesquisas Anísio Teixeira. 2024.

CANTARIN, Márcio Matiassi; BERTUCCI, Roberlei Alves; ALMEIDA, Rogério Caetano. A análise do texto dissertativo-argumentativo. In: TEXTOS dissertativo-argumentativos: subsídios para a qualificação de avaliadores. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2017. P. 81–91

CAVALCANTE, Monica Magalhães. Expressões referenciais – uma proposta classificatória. *Caderno de Estudos Linguísticos*, v. 44, p. 105–118, 2003a.

CAVALCANTE, Mônica Magalhães. Encapsulamento anafórico. In: REFERENCIAÇÃO. Contexto, 2003b. v. 1. P. 177–190

CAVALCANTE, Mônica Magalhães. Referenciação – uma entrevista com Mônica Magalhães Cavalcante. *ReVEL*, v. 13, n. 25, 2015.

CAVALCANTE, Mônica Magalhães. Abordagens da argumentação nos estudos de Linguística Textual. *ReVEL*, v. 14, n. 12, 2016a.

CAVALCANTE, Mônica Magalhães. Expressões indiciais anafóricas. *Revista do GELNE*, v. 2, n. 1/2, p. 1–4, fev. 2016b.

CAVALCANTE, Mônica Magalhães; BRITO, Mariza Angélica Paiva. Anáforas Encapsuladoras - traços peculiares aos rótulos. *Revista de Letras*, v. 1, n. 32, jan. 2016.

CAVALCANTE, Mônica Magalhães; NASCIMENTO, Celina Aparecida Garcia de Souza; DIAS, Silvelena Cosmo. Apresentação Dossiê. *Revista Ensin@ UFMS*, v. 4, n. 8, p. 14–25, dez. 2023. DOI: 10.55028/revens.v4i8.19945.

CAVALCANTE, Mônica Magalhães; SANTOS, Leonor Werneck dos. Referenciação e marcas de conhecimento partilhado. *Linguagem em (Dis)curso*, Universidade do Sul de Santa Catarina, v. 12, n. 3, p. 657–681, set. 2012. ISSN 1518-7632. DOI: 10.1590/S1518-76322012000300002.

CHEN, Banghao *et al.* Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review. *arXiv preprint arXiv:2310.14735*, 2023.

CONTE, M. E. Encapsulamento anafórico. In: *Referenciação*. São Paulo: Contexto, 2003. P. 177–190.

CORDEIRO, Maria Audirene de Souza. *A contribuição da anáfora encapsuladora para a organização da informação em dissertações produzidas por alunos do 3º ano do ensino médio de Santarém*. 2006. Diss. (Mestrado) – Universidade Federal do Pará, Instituto de Letras e Comunicação, Programa de Pós-graduação em Letras, Belém.

COROA, Maria Luiza. O texto dissertativo-argumentativo. In: TEXTOS dissertativo-argumentativos: subsídios para a qualificação de avaliadores. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2017. P. 59–71.

CUSTÓDIO FILHO, Valdinar; CAVALCANTE, Mônica Magalhães. Ponto de Vista em Linguística Textual: Efeitos Argumentativos e Aplicações no Ensino de Língua Portuguesa. *Revista Ensin@ UFMS*, v. 4, n. 8, p. 379–403, dez. 2023. DOI: 10.55028/revens.v4i8.19104.

CUSTÓDIO FILHO, Valdinar; ELIAS, Vanda Maria. O Legado de Mônica Cavalcante para a linguística textual brasileira. *Todas as Letras - Revista de Língua e Literatura*, v. 26, n. 2, p. 1–17. set. 2024.

ELIAS, Vanda Maria da Silva. Texto e argumentação. In: TEXTOS dissertativo-argumentativos: subsídios para a qualificação de avaliadores. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2017. P. 117–136.

FELPI, Lucas. *Cartilha Redação Nota Mil 1.0. 2019*. Disponível em: <<https://www.lucasfelpi.com.br/redamil>>. Acesso em: 10 dez. 2024.

GARCEZ, Lucília Helena do Carmo. Gênero e tipo de texto. In: TEXTOS dissertativo-argumentativos: subsídios para a qualificação de avaliadores. Brasília: Ministério da Educação: Instituto Nacional de Estudos e Pesquisas Anísio Teixeira, 2017. P. 51–58.

GARCEZ, Lucília Helena do Carmo; CORRÊA, Vilma Reche. *Textos dissertativo-argumentativos: subsídios para a qualificação de avaliadores*. Brasília: Ministério da Educação: Instituto Nacional de Estudos e Pesquisas Anísio Teixeira, 2017.

GU, Jindong *et al.* A systematic survey of prompt engineering on vision-language foundation models. *arXiv preprint arXiv:2307.12980*, 2023.

KASNECI, Enkelejda *et al.* ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, v. 103, p. 102274, 2023. ISSN 1041-6080. DOI: <https://doi.org/10.1016/j.lindif.2023.102274>.

KOCH, Ingredore Grunfeld Villaça. Referenciação e orientação argumentativa. In: *Referenciação e discurso*. São Paulo: Contexto, 2005. P. 33–52.

KOCH, Ingredore Grunfeld Villaça; ELIAS, Vanda Maria. *Ler e Compreender: os sentidos do texto*. São Paulo: Contexto, 2017.

LIU, Pengfei *et al.* Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, ACM New York, NY, v. 55, n. 9, p. 1–35, 2023.

MARVIN, Ggaliwango *et al.* Prompt Engineering in Large Language Models. In: *International conference on data intelligence and cognitive informatics*. Singapore: Springer Nature Singapore, jan. 2024. P. 387–402. DOI: 10.1007/978-981-99-7962-2_30.

MEIRA, Ricardo Radaelli *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.

NUNES, Paula Ávila. Escrever não é útil. *Revista da ABRALIN*, p. 192–213, 2024.

PAES, Aline; FREITAS, Cláudia. ChatGPT, MariTalk e outros agentes de conversação. *Processamento de linguagem natural: conceitos, técnicas e aplicações em português*, 2023.

PAES, Dayhane Alves Escobar Ribeiro. Estruturas discursivas: o encapsulamento anafórico em redações de pré-vestibulandos. In: ANAIS do SILEL. Uberlândia: EDUFU, 2013. v. 3.

RADFORD, Alec *et al.* *Improving language understanding by generative pre-training*. 2018.

RASSI, Amanda Pontes; LOPES LOPES, Priscilla de Abreu. Capítulo 19 Correção automática de redação. In: PROCESSAMENTO de Linguagem Natural: conceitos, técnicas e aplicações em Português. BPLN, 2023.

SANTOS, Simone Gregório dos. *Anáfora encapsuladora em redações nota 1000 do Enem*. 2023. Diss. (Mestrado) – Universidade Federal do Rio Grande do Norte, Centro de Ciências Humanas, Letras e Artes, Curso de Letras (Trabalho de Conclusão de Curso), Natal.

TEAM, Gemini *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>>.

TOUVRON, Hugo; LAVRIL, Thibaut *et al.* LLaMA: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023a.

TOUVRON, Hugo; MARTIN, Louis *et al.* LLaMA 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023b.

VAL, Maria da Graça da Costa.; MENDONÇA, Márcia. Coesão nominal: relacionar, categorizar e... argumentar. In: TEXTOS dissertativo-argumentativos: subsídios para a qualificação de avaliadores. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2017. P. 205–212.

Submetido em 10/01/2025
Aceito em 28/05/2025