

# Análise de ambiguidade linguística em modelos de linguagem de grande escala (LLMs)

## Linguistic ambiguity analysis in large language models (LLMs)

Lavínia de Carvalho Moraes \*, Irene Cristina Silvério †, Rafael Alexandre Sousa Marques ‡, Bianca de Castro Anaia §, Dandara Freitas de Paula ¶, Maria Carolina Schincariol de Faria ||, Iury Cleveston \*\*\*, Alana de Santana Correia †† e Raquel Meister Ko Freitag ‡‡

<sup>1</sup> Attenty Sistemas de Software, Análise de dados, Campinas, SP, Brasil.

<sup>2</sup> Universidade Federal de Sergipe, Departamento de Letras Vernáculas, Sergipe, SE, Brasil.

## Resumo

A ambiguidade linguística ainda é um grande desafio para sistemas de processamento de linguagem natural (NLP) apesar dos avanços em arquiteturas como Transformers e BERT. Inspirado pelo êxito recente dos modelos instrucionais ChatGPT (versão 3.5) e Gemini (denominado Bard até 2023), este trabalho visa analisar e discutir a ambiguidade linguística nesses modelos a partir de três tipos de ambiguidade no Português Brasileiro: semântica, sintática e lexical. Para isso, foi desenvolvido um corpus com 120 frases ambíguas e não ambíguas, submetidas aos modelos para tipificação, explicação e desambiguação. Também foi explorada a capacidade de geração de frases ambíguas, solicitando a geração de conjuntos de frases para cada tipo de ambiguidade. Os resultados foram analisados qualitativamente, com base em referenciais linguísticos reconhecidos, e quantitativamente pela acurácia das respostas obtidas. Evidenciamos que equívocos e deficiências nas respostas permanecem mesmo em modelos mais sofisticados, como ChatGPT e Gemini, com explicações frequentemente inconsistentes. A acurácia foi de no máximo 49,58%, apontando a necessidade de estudos descritivos para o aprendizado supervisionado.

*Palavras-chave:* Ambiguidade. Modelos de linguagem. ChatGPT. Gemini.

## Abstract

Linguistic ambiguity continues to represent a significant challenge for natural language processing (NLP) systems, notwithstanding the advancements in architectures such as Transformers and BERT. Inspired by the recent success of instructional models like ChatGPT and Gemini (named Bard in 2023), this study aims to analyze and discuss linguistic ambiguity within these models, focusing on three types prevalent in Brazilian Portuguese: semantic, syntactic, and lexical ambiguity. We created a corpus comprising 120 sentences, both ambiguous and unambiguous, for classification, explanation, and disambiguation. The models' capability to generate ambiguous sentences was also explored by soliciting sets of sentences for each type of ambiguity. The results underwent qualitative analysis, drawing on recognized linguistic references, and quantitative assessment based on the accuracy of the responses obtained. It was evidenced that even the most sophisticated models, such as ChatGPT and Gemini, exhibit errors and deficiencies in their responses, often providing inconsistent explanations. Furthermore, the accuracy peaked at 49.58%, indicating the need for descriptive studies for supervised learning.

*Keywords:* Ambiguity. Language models. ChatGPT. Gemini.

\*Email: l237294@dac.unicamp.br

†Email: i169329@dac.unicamp.br

‡Email: rafaelalexandre001@gmail.com

§Email: biancaanaia@gmail.com

¶Email: d206236@dac.unicamp.br

||Email: mariaschincariol@outlook.com

\*\*\*Email: iury@attenty.com.br

††Email: alana@attenty.com.br

‡‡Email: rkofreitag@uol.com.br

## 1 Introdução

A ambiguidade linguística, caracterizada pela possibilidade de uma palavra ou frase ter dois ou mais significados distintos em uma sentença, é um fenômeno complexo para os modelos de linguagem natural (Ortega-Martín *et al.*, 2023). Essa complexidade deriva da riqueza e sutilezas inerentes à estrutura e ao uso das línguas humanas, a partir da multiplicidade de significados que palavras e estruturas linguísticas podem assumir, dependendo do contexto em que são utilizadas. Mesmo os modelos de linguagem mais avançados, tais como Transformer (Vaswani *et al.*, 2017) e BERT (Devlin *et al.*, 2018), enfrentam desafios ao lidar com diversos tipos de ambiguidade, devido à necessidade de considerar uma ampla variedade de contextos, conhecimentos prévios e nuances culturais que influenciam a interpretação das palavras e frases. Discernir o significado correto em um contexto específico demanda não apenas uma compreensão profunda da língua, mas também uma capacidade de inferência e abstração historicamente desafiadoras de replicar em sistemas computacionais (Ortega-Martín *et al.*, 2023).

A partir de 2010, a área de NLP testemunhou um grande avanço tecnológico com a evolução de técnicas de deep learning, principalmente em funções como sumarização de texto, classificação de tópicos, análise de sentimentos e sintetização de voz. Assim, foi possível experienciar tarefas com uma qualidade que ainda não era esperada até o momento. Desde o lançamento do ChatGPT e do Gemini<sup>1</sup>, modelos de linguagem instrucionais que têm revolucionado o mercado desde 2022, em consonância com o avanço tecnológico e o aumento da produtividade em diversas áreas do cotidiano, a inteligência artificial generativa tem impactado significativamente diversos setores, desde a educação até o mercado de trabalho. Essa influência se manifesta na distribuição de informações, na comunicação de ideias e na compreensão de dados discursivos, redefinindo a forma como interagimos e lidamos com o vasto espectro de informações disponíveis.

Embora modelos avançados como o ChatGPT e o Gemini tenham apresentado progressos significativos na geração de linguagem natural, ainda subsistem incertezas quanto à sua capacidade de compreender fenômenos linguísticos complexos. Esses modelos, baseados em padrões estatísticos e processamento de grandes volumes de dados, não configuram uma verdadeira interlocução com o usuário, uma vez que anulam componentes fundamentais da comunicação humana, como gestos, entonação e expressões faciais. Além disso, eles não possuem uma compreensão do senso comum, o que frequentemente resulta na geração de dados que não refletem adequadamente a realidade compartilhada pelos seres humanos (Itankar; Raza, 2020; Singh; Siddiqui, 2015; Chaplot; Bhattacharyya; Paranjape, 2015; Agirre; Rigau, 1997).

Por outro lado, a capacidade de processamento de um grande volume de dados empíricos provida pelos LLMs também tem oferecido oportunidades de reflexão sobre a relação entre teoria e dados, que é diferente a depender da abordagem assumida, seja indutivo, seja dedutivo, seja de orientação funcionalista, seja de orientação formalista. Mas há também abordagens que descrevem padrões fenômenos sem assumir uma teoria para língua, como é o caso de LLMs. Piantadosi (2023) afirma que LLMs têm conseguido resultados de desenvolvimento da gramática sem usar nenhum dos métodos assumidos como necessários para a descrição linguística por humanos. Por outro lado, resultados de estudos que observam como LLMs lidam com tarefas linguísticas cotidianas, como a leitura de palavras que tiveram letras sistematicamente trocadas por números, como "l3tr4 tr0c4d4" (letra trocada) (Marcus; Leivada; Murphy, 2023), sugerem que, por não terem cognição fundamentada, LLMs não podem tirar proveito desses recursos e, em vez disso, dependem apenas de associações fixas entre palavras representadas e vetores de palavras. Os autores evidenciam um paradoxo: se os LLMs supostamente emulam a linguagem humana, e seu desempenho supostamente mostra habilidades que superaram as dos seres humanos, ao mesmo tempo que estes mesmos LLMs não conseguem fornecer respostas precisas em tarefas muito básicas de compreensão da linguagem, como é o caso da tarefa letra trocada, sendo necessários mais estudos em mais tarefas, como a de resolução de ambiguidades, que podem contribuir para a compreensão da base cognitiva da faculdade humana da linguagem.

Do ponto de vista linguístico, é improvável que esses modelos de linguagem consigam reproduzir,

<sup>1</sup> O termo ChatPGT e Gemini serão tratados ao longo deste trabalho com pronomes masculinos por ser uma convenção adotada pelas demais literaturas brasileiras.

em sua totalidade, as capacidades linguísticas humanas. A significação, por exemplo, é intrinsecamente multimodal, envolvendo não apenas a linguagem verbal, mas também elementos contextuais como expressões faciais, postura corporal e até detalhes sutis como o vestuário do interlocutor, no caso de interações orais (Freitag; Cardoso; Tejada, 2022). No contexto da comunicação escrita, aspectos como a cor da fonte, o tipo de papel e o suporte do texto podem influenciar a interpretação do significado. Além disso, fatores como variação linguística, tom de voz e o uso de linguagem figurativa são determinantes para a construção de sentido, o que torna a completa emulação da experiência linguística humana por modelos artificiais uma tarefa extremamente desafiadora.

Essas dificuldades afetam a capacidade dos modelos de linguagem de processar adequadamente ambiguidades linguísticas e superar desafios já observados em modelos tradicionais, que também enfrentam limitações. Além disso, a maior parte dos estudos desenvolvidos até o momento têm como foco a língua inglesa (Ortega-Martín *et al.*, 2023). Análises mais aprofundadas no contexto do português brasileiro, uma língua de baixos recursos linguísticos (Finger, 2021; Freitag, 2022), ainda não foram conduzidas, revelando a importância de investigações neste campo.

Nesse contexto, este trabalho tem como objetivo responder as seguintes perguntas:

1. Qual é a precisão dos modelos na detecção de ambiguidade linguística em frases do Português Brasileiro?
2. Os modelos conseguem desambiguar adequadamente as sentenças?
3. Qual dos modelos percebe melhor os fenômenos de homonímia e polissemia?
4. Quais padrões de ambiguidade os modelos ChatGPT e Gemini demonstram conhecer na geração de frases ambíguas?

O estudo de ambiguidade é particularmente complexo, por envolver uma gama de variáveis que interferem, desde a natureza do item lexical, passando pela sintaxe da sentença, e envolvendo o conhecimento de mundo e experiência pessoal de cada falante. Além disso, há um paradoxo a ser superado: um humano pode perceber ambiguidade e não saber explicá-la, enquanto a IA pode saber explicar o que é uma ambiguidade, mas não saber reconhecê-la. Assumindo que as IAs são capazes de imitar em grande medida o processamento da linguagem humana e que têm o potencial de fornecer informações sobre a forma como as pessoas aprendem e utilizam a linguagem (Cai *et al.*, 2023) conduzimos um estudo com a realização de quatro tarefas. Utilizando um conjunto de sentenças com e sem ambiguidade criado por nós, verificamos a consistência das respostas dos modelos ao fazer as mesmas perguntas duas vezes, em momentos distintos, e contrastamos algumas inconsistências nas respostas obtidas.

## 2 Processamento da ambiguidade linguística e o processamento da linguagem natural

O processamento semântico é um domínio de investigação fundamental para os modelos linguísticos pré-treinados e dos grandes modelos de linguagem. No campo das tarefas de processamento semântico, a desambiguação do sentido das palavras demanda parâmetros definidos. No entanto, o estudo da ambiguidade é multidimensional na linguística, como veremos a seguir.

Esta pesquisa fundamenta-se em disciplinas que estão interligadas, demonstrando a natureza interdisciplinar das áreas envolvidas. Portanto, o estudo aproveita-se dos princípios teóricos relacionados à ambiguidade linguística (Seção 2.1) e da aplicação dos modelos computacionais de grande escala, ChatGPT e Gemini (Seção 2.2).

É importante destacar que o arcabouço teórico sobre ambiguidade linguística, empregado nesta seção, servirá como alicerce para a construção do nosso conjunto de sentenças. Entre várias abordagens teóricas disponíveis, decidimos adotar a perspectiva de Cançado (Cançado, 2005) dada a sua compatibilidade com a metodologia da pesquisa e o objeto de estudo, os modelos de linguagem, que não se valem dos aspectos de interlocução para compreender ambiguidade, mas focam nos componentes linguísticos do enunciado. Dessa forma, a fundamentação dos princípios de ambiguidade dessa taxonomia permitiu que os dados fossem analisados de forma mais objetiva, evitando depender de fatores contextuais externos. Além disso, há trabalhos bem recentes de outros países que fazem estudos do ponto de vista computacional de análise de ambiguidade mantendo uma coesão similar

com aspectos da teoria da Cançado em termos de ambiguidade lexical, semântica e sintática (Haber; Poesio, 2021; Ortega-Martín *et al.*, 2023; Liu *et al.*, 2023; Goel, 2023).

## 2.1 Ambiguidade linguística

A ambiguidade é um fenômeno semântico no qual uma palavra, expressão ou sentença pode ter mais de uma interpretação válida. A resolução desse tipo de fenômeno depende fortemente do contexto, que orienta a escolha do sentido adequado entre os que possíveis. No entanto, os modelos de linguagem, por apresentarem restrições na compreensão dos componentes multimodais presentes no contexto, não captam sinais extralingüísticos, fundamentais para processar ambiguidade ou desambiguar enunciados.

Historicamente, a ambiguidade é uma questão antiga e amplamente estudada em tarefas de PLN através de análises simbólicas de sentenças, como o parsing sintático e semântico (Church; Patil, 1982; Koller *et al.*, 2008) ou resolução de correferência (Poesio; Artstein, 2005). Entretanto, com o avanço recente no uso de redes neurais profundas e modelos de linguagem pré-treinados, o foco do campo tem se deslocado para problemas de compreensão em níveis mais altos, incluindo o raciocínio e a geração de texto. Nesse novo contexto, a ambiguidade continua sendo um desafio já que os modelos aprendem a lidar com ambiguidade de forma implícita, por meio de grandes quantidade de dados de treinamento.

Seguimos a proposta de Cançado (2005) para o português brasileiro, que considera o nível de uso da língua e inclui a ambiguidade lexical, com casos de homonímia, polissemia, ambiguidade semântica e ambiguidade sintática.

A **ambiguidade lexical** compreende uma sentença com dupla interpretação incidente em um item lexical, podendo ser gerada por homonímia ou polissemia. A homonímia se dá quando os sentidos do item lexical não são relacionados, como na oração “Eu estou indo para o banco” em que a palavra “banco” possui significados diferentes, pode corresponder à instituição financeira e ao assento. Já a polissemia ocorre quando os sentidos do termo identificado como ambíguo contém relação entre si, por exemplo na sentença “O Frederico esqueceu a sua concha”, neste caso “concha” pode significar uma concha do mar ou uma concha de cozinha, em que ambos os objetos possuem o mesmo formato, e por isso, pode ocorrer uma associação polissêmica, o mesmo fenômeno ocorre com palavras como: rede (de internet, de deitar, de pescar) e pilha (de comida, de bateria).

Além disso, na taxonomia de Cançado (2005), a ambiguidade lexical também pode ser causada por meio de preposições, classificando-se assim como ambiguidade preposicional. Por exemplo, na sentença “O burro do Paulo anda doente” permite duas interpretações: “burro” pode se referir ao animal que Paulo possui, caracterizando uma ambiguidade literal, ou pode ser uma expressão figurativa, referindo-se a Paulo como sendo uma pessoa “burra”. Essa dupla interpretação ocorre por causa da homonímia, onde a palavra “burro” tem dois significados não relacionados que são possíveis de serem inferidos por conta da presença da preposição. Esse tipo de ambiguidade preposicional se soma às outras duas subclassificações de ambiguidade lexical identificadas por Cançado que são tratadas na pesquisa, homonímia e polissemia.

Embora a ambiguidade preposicional seja uma categoria relevante, o conjunto de dados criado para testar os modelos neste trabalho se restringiu a casos de polissemia e homonímia, excluindo assim a ambiguidade gerada por preposições, mas esses casos apareceram na Tarefa 4 de geração de frases por parte dos modelos.

Nessa mesma esteira teórica, Lyons (Lyons, 1977) considera a homonímia e a polissemia como casos de ambiguidade lexical, assim como (Cançado, 2005). A homonímia ocorre quando dois ou mais significados não possuem relação semântica entre si, sendo historicamente distintas no desenvolvimento da língua, o que reforça a independência dos sentidos em termos lexicais. Já a polissemia envolve uma relação intrínseca entre os diferentes sentidos de uma palavra, onde esses sentidos compartilham um núcleo comum de significado. Assim, no exemplo “banco”, os diferentes sentidos (assento e instituição financeira) são homônimos, uma vez que suas origens etimológicas e semânticas são divergentes. Por outro lado, no caso de “concha”, os diferentes usos podem ser entendidos como polissemia, já que ambos os sentidos remetem a uma semelhança de forma e função, refletindo a natureza interconectada das diversas acepções do termo.

É importante destacar que essa distinção de subcategorias é fundamental, visto que, para o português, a ambiguidade lexical é a que apresenta mais recursos descritivos para suporte computacional (Laporte, 2001), necessário aos LLMs.

Na taxonomia de Cançado, a **ambiguidade semântica** é abordada como uma questão de correferencialidade, em que os pronomes podem ter vários antecedentes. Consideremos, por exemplo, a seguinte frase: “José falou com seu irmão?” Esta sentença ilustra claramente esse tipo de ambiguidade, na qual não é possível determinar se o irmão mencionado é o irmão de José ou o irmão da pessoa para quem a pergunta é dirigida, ou seja, um terceiro elemento. Nesse contexto, as interpretações possíveis são atribuídas à natureza da ligação entre os pronomes presentes na sentença.

A gramática gerativa compartilha da mesma percepção sobre pronomes de referência como fator de ambiguidade. (Chomsky, 1981) trata da correferencialidade por meio do conceito de *binding*, que estabelece princípios para a associação entre pronomes e seus possíveis antecedentes dentro de uma estrutura sintática. A interpretação de um pronome é, portanto, condicionada por sua posição hierárquica na sentença e pelas relações gramaticais que mantém com os demais constituintes. No exemplo “José falou com seu irmão”, a ambiguidade semântica decorre da indefinição sobre a correferência do pronome possessivo “seu”, permitindo que ele se refira tanto a José quanto a um terceiro participante no discurso. Dessa forma, a teoria de Chomsky contribui para a análise da ambiguidade pronominal ao demonstrar como a estrutura sintática subjacente pode resultar em múltiplas possibilidades interpretativas.

A perspectiva assumida por (Cançado, 2005) e (Chomsky, 1981) sobre ambiguidade semântica, ou anafórica, também é confirmada por outros estudos da linguística computacional, uma vez que esses também apresentam padrões regulares na retomada de antecedentes a partir de pronomes (Bruscato; Baptista, 2021; Nogueira Teixeira; Micelli Fonseca; Soares, 2014; Godoy; Carvalho, 2020; Lima Silva; Melo; França, 2023).

A **ambiguidade sintática** é um fenômeno de imprecisão de sentidos que não é ocasionado pela interpretação de uma palavra individual, mas se atribui às distintas estruturas sintáticas que originam diferentes interpretações: a frase concebe diferentes análises a partir dos seus possíveis sintagmas, que são divisões existentes dentro da frase em grupos de palavras. A sentença “O magistrado julga as pessoas culpadas” é um exemplo em que a organização da sentença pode ser: (a) O magistrado [julga] [as pessoas culpadas] ou (b) O magistrado [julga] [culpadas] [as pessoas]. A ambiguidade sintática envolve as diversas possibilidades de interpretação da sentença apenas reorganizando a posição das expressões envolvidas na oração, o que não ocorre nos tipos de ambiguidade tratados anteriormente. A ambiguidade sintática é um fenômeno amplamente estudado, com padrões obedecendo a princípios como o da aposição mínima e da aposição local (Maia, s.d.), que atuam em diferentes línguas, incluindo o português brasileiro (Maia, M. A. R. et al., 2003; Maia, M. et al., 2004; Brito Bezerra; Martins Leitão, 2013; Machado, 1996).

Os enunciados processados com esse tipo de ambiguidade corroboram com o que é descrito por (Jurafsky; Martin, 2000), ao se basearem na decomposição dos constituintes por meio do *parsing sintático*. A estrutura dos sintagmas permite identificar diferentes possibilidades de interpretação para o adjunto. Por exemplo, na frase “Ele saiu da loja de carro”, há a possibilidade de o adjunto modificar o verbo (1) ou de modificar o objeto indireto (2):

1. [Ele] [saiu [da loja] [de carro.]]
2. [Ele] [saiu [da loja [de carro.]]]

Essa variação estrutural é o que dá origem à ambiguidade na sentença, dependendo da relação entre o adjunto e os outros elementos da frase.

## 2.2 Modelos de linguagem

Os modelos de linguagem grandes (LLMs), ChatGPT 3.5 (OpenAI, 2023) e o Gemini (Ahmed et al., 2023), contam com aproximadamente 175 bilhões e 1,5 trilhão de parâmetros, respectivamente. Eles funcionam a partir de redes atencionais do tipo Transformer (Vaswani et al., 2017) que são pré-treinadas de forma auto-supervisionada em grandes conjuntos de dados. Posteriormente, eles são refinados através do aprendizado instrucional baseado em contexto e através do aprendizado por

reforço baseado em feedback humano (RLHF) (Ouyang *et al.*, 2022). Os mecanismos atencionais do tipo *self-attention* (Vaswani *et al.*, 2017) presentes nas arquiteturas permitem a captura de dependências de longas distâncias de forma computacionalmente eficaz, minimizando o esquecimento dos modelos em sequências longas. Por fim, a estratégia de indução de pensamento em cadeia (CoT) (Wei *et al.*, 2023) aplicada após os treinamentos permite que os modelos usem a sua última saída como entrada para gerar uma saída ainda mais refinada, melhorando a qualidade das respostas dadas.

No contexto da ambiguidade linguística, sabe-se que os mecanismos de *self-attention* aprimoram a capacidade dos modelos em lidar com a ambiguidade semântica por meio do aprendizado da correferência, naturalmente presente na estrutura do mecanismo atencional (Ortega-Martín *et al.*, 2023). Entretanto, ainda não há evidências de que o RLHF, CoT e treinamentos instrucionais exclusivamente presentes na ChatGPT e Gemini impactam no processamento da ambiguidade.

Assumimos a hipótese de que o aprendizado instrucional permite, de forma implícita, que esses modelos compreendam instruções pelo contexto e sigam direções específicas sobre a intenção do usuário, mitigando ou resolvendo certos casos de ambiguidade. Além disso, o RLHF pode ser particularmente valioso para lidar com situações ambíguas, já que ele provê o alinhamento correto do modelo com as intenções do usuário, por meio de recompensas dadas durante um treinamento de ajuste fino. Por fim, é possível que o CoT também tenha um papel crucial na resolução de ambiguidade, pois pode auxiliar os modelos a decompor o problema em etapas intermediárias mais gerenciáveis, permitindo abordar a resolução de ambiguidade de forma gradual, em vez de tentar resolvê-la de uma só vez. Diante dessas considerações, surge a questão: como esses elementos se adaptam durante o treinamento dos modelos para lidar com as sutilezas e complexidades do processamento de ambiguidades no contexto do português brasileiro? Explorar essa questão pode fornecer informações valiosas sobre como otimizar esses modelos para atender às nossas necessidades linguísticas e culturais.

### 3 Metodologia

Para avaliar o desempenho dos modelos no processamento de ambiguidades quanto aos parâmetros da ambiguidade lexical, sintática e semântica, conduzimos tarefas utilizando um conjunto de dados que foi elaborado e analisado pelos autores do presente artigo, um grupo de seis estudantes de Letras e Linguística, cada um com um conhecimento maior em áreas distintas da linguística, como fonética, tradução, sintaxe e análise do discurso. Esses estudantes serão referidos no trabalho como juízes-humanos por se demonstrarem aptos a produzir e julgar de forma adequada os dados, além disso, vale salientar que os dados e resultados das tarefas foram avaliados, posteriormente, por uma coautora que é especialista na área de variação linguística.

Para garantir a consistência e a confiabilidade das frases geradas pelo grupo, foi adotado o procedimento de *inter-annotator agreement*. As sentenças que não receberam consenso absoluto entre os juízes-humanos foram revisadas ou excluídas do corpus. Como resultado, não foi necessário calcular o coeficiente de concordância<sup>1</sup>, uma vez que todas as sentenças, tanto as ambíguas quanto as não ambíguas, só foram incluídas após alcançarem 100% de aprovação. Esse critério rigoroso assegurou que o conjunto de dados utilizado nas tarefas fosse confiável ao máximo para minimizar divergências.

O experimento foi composto por um grupo de 120 sentenças, distribuídas de forma balanceada entre os três tipos de ambiguidade. Desses, 60 apresentam algum tipo de ambiguidade, seja semântica, lexical ou sintática. As frases ambíguas foram elaboradas com o objetivo de provocar especificamente um dos três tipos de ambiguidade (semântica, lexical ou sintática). No entanto, é possível que algumas

<sup>1</sup> Ao criar o corpus, as frases foram revisadas por todos os autores do artigo, e apenas aquelas com 100% de aprovação no procedimento de *inter-annotator agreement* foram disponibilizadas para os testes, com o objetivo de minimizar a presença de frases problemáticas. As frases geradas pelos modelos também passaram por uma revisão conjunta e foram discutidas entre os autores para garantir consenso absoluto a partir do mesmo procedimento. Por exemplo, na Tabela 4, que corresponde à geração de frases por parte do modelo, a sentença “O homem viu o acidente com os próprios olhos” contém um componente favorável à criação de ambiguidade sintática, o adjunto “com os próprios olhos”, no entanto, com base na teoria de Cañado, essa ambiguidade é válida quando acompanhada por uma ambiguidade semântica, que permite o duplo sentido na compreensão pragmática do contexto dentro do enunciado. Já na Tabela 3, a frase “Pedi o prato principal ao garçom, era filé!” é parte do corpus das sentenças distratoras, pois a estrutura do enunciado fornece informações suficientes para evitar outras interpretações, ou seja, entendemos que “prato” não se refere ao utensílio doméstico e “filé” não é um elogio metafórico. Sendo assim, essa frase foi considerada não ambígua pelos seis juízes-humanos e pela especialista em linguística que revisou o trabalho.

sentenças apresentem mais de um tipo de ambiguidade, que não foi avaliado durante a criação das mesmas. No decorrer de nossa análise, buscamos isolar ao máximo cada frase, de modo que apenas um elemento causador de ambiguidade estivesse presente. Essa abordagem nos permite um controle mais rigoroso das variáveis observadas nos experimentos, uma vez que a ambiguidade é um fenômeno complexo e multifacetado.

É importante ressaltar que, dentro de uma única sentença, podem existir múltiplos fatores que contribuem para a ambiguidade, o que torna a sua identificação e análise ainda mais desafiadoras, então, para um estudo inicial, de caráter pioneiro com LLM's formamos dados linguísticos mais artificiais, mas planejamos trabalhar futuramente com dados provenientes de corpora ou textos reais, incorporando materiais autênticos que nos permitam uma compreensão mais abrangente das ambiguidades linguísticas em contextos variados.

Foram criadas 20 sentenças com ambiguidade lexical — que abrangem casos de homônimia e polissemia, sem distinção de categoria na análise das frases —, 20 sentenças com ambiguidade semântica, nas quais o referente dos pronomes não está claro, e, por fim, 20 sentenças com ambiguidade sintática, envolvendo adjuntos adnominais ou adverbiais ambíguos que provocam duplo sentido devido às diferentes organizações estruturais que a frase pode ter (Tabela 1).

As outras 60 sentenças distratoras tiveram sua ambiguidade barrada ao máximo, não sendo verificadas pelos juízes-humanos durante a elaboração dos dados linguísticos. As sentenças distratoras foram criadas com a intenção de evitar qualquer forma de ambiguidade. Da mesma forma que foi feito o julgamento das frases ambíguas, as frases não ambíguas foram avaliadas em um nível de significado mais isolado, sem considerar diversos contextos enunciativos figurativos, na maioria das vezes, nas quais ela poderia fazer sentido, logo assumimos um posicionamento de interpretação mais pragmático linguístico, que corrobora com Cançado. Essas mesmas considerações são válidas para as avaliações feitas em relação às frases que foram geradas pelos modelos de linguagem na Tarefa 4.

Um exemplo para ilustrar essa forma de análise é a frase “Pedi o prato principal ao garçom, era filé!” (Tabela 3) em que em um contexto muito específico, poderia significar que o cliente pediu o utensílio principal de servir comida e este utensílio era “filé”, um termo popular para se referir a algo bom, porém, a princípio, a frase foi escrita tendo em vista o significado mais óbvio, que se determinou a partir das pistas interpretativas deixadas dentro da frase na qual um sentido lexical confirmava o outro sem a necessidade de buscar condições exofóricas à sentença que justificassem uma polissemia.

**Tabela 1**

Id	Sentença	Classe
1	A rede caiu.	Ambiguidade lexical
2	Ela não gosta da amiga dela.	Ambiguidade semântica
3	O menino viu o incêndio do prédio.	Ambiguidade sintática

*Fonte:* <https://XXXXXXXXXX>.

Para responder as nossas perguntas de pesquisa, foram conduzidas quatro tarefas<sup>2</sup> distintas com as sentenças criadas. Em todas, foram realizadas coletas duplicadas das interações para cada frase, reiniciando o *console* entre cada coleta para evitar qualquer influência do contexto que pudesse gerar respostas tendenciosas. Essa abordagem permitiu avaliar a consistência dos modelos nas respostas fornecidas.

A tarefa 1 visava identificar se os modelos conseguem detectar a presença de ambiguidade em cada sentença por meio da seguinte instrução: **A sentença “[sentença]” é ambígua? Responda, sim, não ou não sei.** Foram apresentadas individualmente todas as sentenças e registradas as respostas

<sup>2</sup> As respostas dadas pelos modelos de linguagem nas tarefas foram coletadas entre julho de 2023 até janeiro de 2024. Coletas posteriores podem levar a resultados diferentes devido às atualizações dos modelos de linguagem. Para evitar ao máximo o enviesamento dos modelos, as perguntas foram formuladas da forma mais objetiva possível, evitando dar pistas sobre a resposta correta ou sinalizando qual era o resultado esperado. A preocupação em minimizar testes com viés tinha o objetivo de garantir uma avaliação mais precisa do desempenho dos modelos de linguagem, afastar casos de generalização nas respostas, favorecer a imparcialidade dos resultados e alcançar um conjunto de respostas o mais transparente possível.

dos modelos, comparando-as com a nossa classificação prévia. As respostas foram cuidadosamente avaliadas quanto à correção e abrangência das explicações fornecidas por seis juízes-humanos que as julgaram independentemente. A partir dos resultados, foi gerada uma matriz de confusão para computar a quantidade de verdadeiros positivos (sentenças que são ambíguas e que os modelos classificaram como ambíguas), falsos positivos (sentenças que não o são e que os modelos classificaram como ambíguas), verdadeiros negativos (sentenças que não são ambíguas e que os modelos assim classificaram como não ambíguas), e falsos negativos (sentenças que são ambíguas e que os modelos classificaram como não ambíguas).

Na tarefa 2, foi realizado um teste para avaliar a capacidade dos modelos em distinguir corretamente entre as três classes de ambiguidade estudadas neste trabalho, formulando a seguinte pergunta para cada modelo: **“Qual o tipo de ambiguidade?”**. A tarefa consistiu em perguntar qual o tipo de ambiguidade da sentença que foi classificada anteriormente como ambígua ou não ambígua. Na tarefa 3, foi verificada a capacidade dos modelos em desambiguar as sentenças que foram fornecidas a eles. Com esse propósito, foram apresentadas frases que incluem tanto sentenças ambíguas, quanto sentenças não ambíguas, e solicitado aos modelos a seguinte instrução: **Faça a desambiguação da frase: “[sentença]”**. A tarefa busca testar a habilidade dos modelos em compreender e interpretar o contexto, escolhendo a interpretação mais apropriada quando a ambiguidade está presente.

Na tarefa 4, foi avaliada a capacidade dos modelos em gerar frases ambíguas na categoria solicitada. Para isso, pedimos para cada modelo gerar frases da seguinte forma: **Gere 20 frases com ambiguidade “[categoria]”**. Em seguida, as respostas obtidas foram avaliadas por juízes-humanos, buscando compreender quão preciso é o ChatGPT o Gemini ao criarem frases que apresentam múltiplas interpretações contextuais.

Para mensurar quantitativamente o desempenho dos modelos, foi utilizada a métrica de acurácia, a qual já é amplamente empregada na área de aprendizado de máquina (Naser; Alavi, 2021; Freitag; Tejada *et al.*, 2021). A acurácia, no contexto da classificação, representa a proporção de frases corretamente classificadas pelos modelos em relação ao total de frases apresentadas na tarefa, como apresentado na equação 1.

$$acc = \frac{\text{Número de previsões corretas}}{\text{Total de previsões}} \quad (1)$$

Todas as sentenças criadas por nós e geradas pelos modelos durante as tarefas estão disponíveis no Apêndice A. As respostas dos modelos durante as tarefas estão disponíveis para download<sup>3</sup> em nosso repositório.

## 4 Resultados

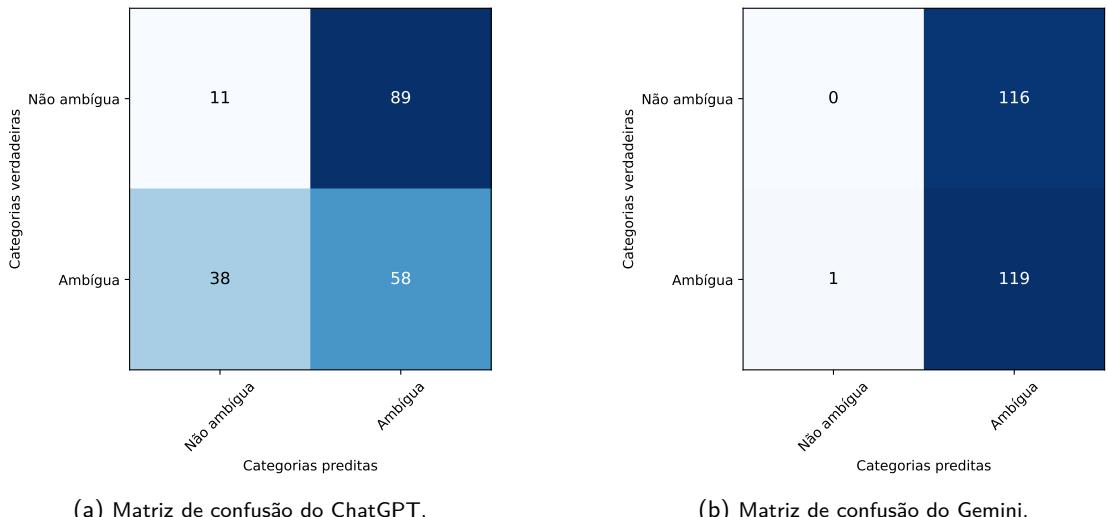
Nesta seção, são apresentados os resultados obtidos. Cada seção responde uma das nossas perguntas de pesquisa através de quatro tarefas. Todas as análises dos resultados descritos foram realizadas por seis juízes-humanos do curso de Letras ou Linguística com conhecimento na área.

### 4.1 Qual é a precisão dos modelos na detecção de ambiguidade linguística em frases do Português Brasileiro?

Para avaliar a precisão dos modelos na detecção de ambiguidade seguimos a proposta de (Freitag; Tejada *et al.*, 2021), foram comparadas as acurárias e as matrizes de confusão obtidas, usando os dados da tarefa 1. Esta análise concentrou-se exclusivamente na detecção da presença ou ausência de ambiguidade, sem levar em consideração o tipo específico identificado pelos modelos posteriormente. Assim, os dados foram divididos em dois grupos distintos, totalizando 60 frases com ambiguidade e 60 frases distratoras. Foram realizadas duas coletas por frase, sendo obtidas 240 previsões para cada modelo.

Os resultados da matriz de confusão na Figura 1 revelam que o ChatGPT registrou uma acurácia de apenas 28,75%, enquanto o Gemini alcançou 49,58%, indicando que estas versões dos modelos

<sup>3</sup> <https://docs.google.com/spreadsheets/d/1AOff1GJmh3oWIuKdfBGGHWeox-7HtuHQW0LkaT5yFVI/edit?usp=sharing>



**Figura 1.** Matrizes de Confusão dos modelos ChatGPT e Gemini.

não conseguem detectar ambiguidade com precisão confiável. Os resultados revelam que ambos os modelos exibem uma quantidade significativa de falsos positivos, identificando ambiguidade em frases que carecem dela. Enquanto o ChatGPT demonstra erros distribuídos em todos os quadrantes da matriz, o Gemini tende a rotular quase todas as frases como ambíguas, resultando em uma taxa maior de falsos positivos. Para computar as matrizes de confusão e a acurácia foram consideradas apenas as respostas em que os modelos responderam “Sim” e “Não”, de modo que todas as respostas “Não Sei” foram descartadas. Assim, o ChatGPT destacou-se por apresentar mais dúvidas, conseguindo responder 196 perguntas enquanto o Gemini respondeu 236.

A diferença de acurácia entre os dois modelos pode ser atribuída ao fato do ChatGPT expressar dúvidas ao detectar frases ambíguas, declarando não saber ou negando a presença de ambiguidade. Analisando os três tipos de ambiguidade, observa-se que o ChatGPT lida melhor com ambiguidades semânticas e sintáticas, cometendo mais erros quando a ambiguidade é apenas lexical. Uma explicação é devida à estrutura da ambiguidade sintática ser descrita em estudos de processamento linguístico (Maia, s.d.) (1) e de processamento de linguagem natural (Padovani, 2022)(2). Em contrapartida, o Gemini apresentou apenas um caso de falso negativo, acertando todos os outros testes em frases ambíguas. Entretanto, o Gemini tem a tendência de não distinguir entre frases ambíguas e não ambíguas, pois, em todos os testes, indica a presença de ambiguidade.

#### 4.2 Qual dos modelos percebe melhor os fenômenos de homonímia e polissemia?

Neste estudo, foram examinadas as explicações fornecidas pelos modelos na tarefa 2, analisando apenas as respostas do conjunto de frases com ambiguidade lexical. Durante a análise, foi considerado que os modelos perceberam a homonímia e a polissemia através da explicação dada pelos modelos para justificar o tipo de ambiguidade identificada. Se a explicação dada pelos modelos foi referente a ambiguidade gerada devido aos diferentes significados que o item lexical pode assumir na frase, e se realmente os juízes-humanos enxergariam os diferentes significados do item lexical da mesma maneira, foi considerado acerto, caso contrário foi tido como um erro por parte dos modelos.

Para demonstrar temos as frases **“Isso não é legal!”** e **“A carteira foi danificada.”** que foram classificadas como ambiguidade lexical de homonímia pelos juízes-humanos. Tais frases foram testadas no ChatGPT de modo que a primeira frase foi considerada correta, pois recebeu, em um dos testes, a seguinte resposta **“A frase “Isso não é legal!” pode apresentar ambiguidade de sentido, pois a palavra “legal” possui múltiplas interpretações, dependendo do contexto em que é usada.”** e a segunda frase foi classificada como incorreta, em um dos testes, por apresentar uma explicação incoerente **“A frase “a carteira foi danificada” pode ser considerada ambígua devido à ambiguidade estrutural. Isso ocorre porque não está claro se a carteira sofreu dano físico ou se está se referindo a uma carteira de identidade ou pertencente a alguém. Portanto, a**

## **ambiguidade está relacionada à interpretação da frase em termos de sua estrutura sintática.**

Ambos os modelos demonstraram uma boa compreensão dos fenômenos de homônima e polissêmia. O ChatGPT obteve sucesso em 75% das vinte frases testadas, enquanto o Gemini alcançou uma taxa de acerto de 80%. Vale ressaltar a notável similaridade na forma como os modelos interpretaram esses fenômenos, uma vez que ambos obtiveram sucesso quase no mesmo subconjunto de frases.

### **4.3 Os modelos conseguem desambiguar adequadamente as sentenças?**

Nesta seção, a análise foi baseada nas respostas obtidas na tarefa 3, na qual foi solicitada a desambiguação de todas as sentenças. As respostas obtidas foram divididas em três categorias: 1) Correta ocorreu quando o modelo explica corretamente duas ou mais interpretações possíveis para a sentença, de forma semelhante à interpretação que um ser humano teria. No caso das sentenças sem ambiguidade, o modelo esclarece a ausência de ambiguidade e observa que não é necessário executar a tarefa solicitada; 2) Incompleta ocorreu quando o modelo apresentou pelo menos uma interpretação correta para a sentença, errando nas demais interpretações possíveis; e 3) Incorreta ocorreu quando o modelo apresentou opções de desambiguação sem nenhum sentido.

Os resultados mostrados na Figura 2 revelam que ambos os modelos enfrentam dificuldades ao tentar desambiguar corretamente as sentenças, uma vez que propõem interpretações equivocadas, identificando ambiguidade em elementos pouco ou nada relacionados à ambiguidade real. Por exemplo, no conjunto de frases sem ambiguidade, ambos os modelos tentaram desambiguá-las oferecendo opções de interpretação.

A frase “*Como o pássaro estava cansado, Ricardo o alimentou.*” obteve uma desambiguação incorreta tanto na tarefa do ChatGPT como do Gemini. No primeiro modelo, a resposta foi a seguinte: “*Para desambiguar a frase e torná-la mais clara, podemos reescrevê-la de forma a eliminar a ambiguidade de referência. Uma maneira de fazer isso é substituindo o pronome ‘o’ por seu referente explícito. Vejamos: 1. Como Ricardo estava cansado, o pássaro foi alimentado por ele. Nesta versão reescrita, fica claro que o pronome ‘o’ se refere ao pássaro, e não a Ricardo. Agora a frase não possui mais ambiguidade de referência.*”

O segundo modelo apresentou a seguinte explicação: “*Claro, aqui estão duas possíveis desambiguações da frase ‘Como o pássaro estava cansado, Ricardo o alimentou’: Ricardo alimentou o pássaro porque estava cansado. Ricardo alimentou o pássaro porque estava cansado de ver o pássaro cansado. A primeira interpretação é a mais provável, pois é mais comum alimentar alguém quando está cansado. No entanto, a segunda interpretação também é possível, pois é possível que Ricardo estivesse preocupado com o bem-estar do pássaro e quisesse ajudá-lo a se recuperar. Sem mais informações, não é possível dizer com certeza qual é a interpretação correta.*”

Os modelos desambiguaram sem necessidade mais de 90% dos casos de teste do conjunto de frases sem ambiguidade. É relevante destacar que, mesmo o ChatGPT - que afirmou em perguntas anteriores que algumas frases não tinham ambiguidade - ofereceu opções de desambiguação, sendo incongruente com suas respostas anteriores. Uma exceção notável ocorreu em uma resposta relacionada à frase “*Gostamos da cor rosa*”, na qual o ChatGPT afirmou: “*Peço desculpas, mas a frase ‘Gostamos da cor rosa’ não é intrinsecamente ambígua e, portanto, não requer desambiguação. Ela expressa claramente uma preferência pela cor rosa. Se houver um contexto específico causando ambiguidade, por favor, forneça mais informações para que eu possa ajudar a esclarecer a situação*”. Observa-se que os modelos têm a tendência de realizar superinterpretações das sentenças sem ambiguidade, oferecendo soluções sem lógica apenas para cumprir a tarefa solicitada, demonstrando ainda não haver o conhecimento explícito de regras do funcionamento da língua.

Na desambiguação das frases com ambiguidade sintática, os modelos também apresentaram erros na maioria dos testes. Em 62,5% dos testes, as análises apresentadas estavam incorretas ou incompletas e os critérios usados nas frases classificadas com ambiguidade sintática, também, pendiam para uma abordagem semântica, lexical ou voltada para a necessidade de um contexto favorável para desambiguar as sentenças. No caso da sentença “*Pedro me mandou um cartão postal de Veneza*”, o ChatGPT apresentou a seguinte desambiguação para a frase: “*Pedro me mandou um cartão postal de Veneza*”; “*Pedro mandou um cartão postal de Veneza para a minha amiga Maria, e ela me mostrou*

Sem Amb.	12	0	108
Sintática	15	19	6
Semântica	31	3	6
Lexical	29	7	4

Corretas      Incompletas      Incorretas

(a) Resultados quantitativos da tarefa de desambiguação realizada pelo ChatGPT.

Sem Amb.	3	6	108
Sintática	20	13	7
Semântica	23	3	14
Lexical	29	5	6

Corretas      Incompletas      Incorretas

(b) Resultados quantitativos da tarefa de desambiguação realizada pelo Gemini.

**Figura 2.** Estatísticas obtidas na tarefa de desambiguação. Os resultados mostram que ambos os modelos apresentam muita dificuldade para identificar que as sentenças distratoras não tem ambiguidade. Em 108 casos de teste, ambos os modelos forneceram explicações que não refletiam a interpretação humana, justificando a presença de ambiguidade em sentenças que, na verdade, não a possuíam.

o cartão”. Contudo, a ambiguidade no adjunto “de Veneza” permanece. Outra solução proposta é mudar a alocação do pronome: “Pedro mandou um cartão postal de Veneza para mim”, mantendo a frase ambígua. Nesse caso, o Gemini ofereceu melhores opções, como “Pedro está atualmente em Veneza e me enviou um cartão postal que foi produzido lá.” ou “Pedro não está atualmente em Veneza, mas me enviou um cartão postal que mostra uma imagem da cidade.”, demonstrando aplicar os princípios da aposição mínima e da aposição local (Maia, s.d.) para resolver a ambiguidade. No entanto, o modelo, na maioria das vezes, não conseguiu identificar com precisão qual era o elemento causador da ambiguidade.

Por outro lado, no grupo de frases com ambiguidade semântica, na maior parte dos testes, ambos os modelos de linguagem sugeriram frases adequadas para a desambiguação das sentenças, sempre destacando a importância do contexto para a correta interpretação. Um exemplo pode ser visto com a frase “Ana me contou um segredo sobre ela”, em que o ChatGPT propôs a desambiguação esperada, compreendendo que “ela” pode se referir tanto à Ana, quanto a uma outra pessoa. Em relação às respostas do Gemini, a maioria também foi apropriada porém com precisão menor que a ChatGPT. Por exemplo, em “Paulo não entrou na universidade de novo”, o Gemini indicou corretamente que as possíveis reescritas da sentença seriam “Paulo não foi aceito na universidade de novo” e “Paulo não entrou na universidade de novo depois de ter se formado”. Já em “A carteira foi danificada”, o Gemini sugeriu opções que especificassem informações não necessariamente ambíguas, visto que ele interpretou que a ambiguidade estaria em “danificado”, assim, deveria ser especificado se “A carteira foi danificada fisicamente” ou “A carteira foi danificada financeiramente”. Estas interpretações são bastante distantes do que se encontra em estudos psicolinguísticos sobre processamento (Machado, 1996; Brito Bezerra; Martins Leitão, 2013).

Os modelos tiveram, simultaneamente, melhor desempenho nos casos de ambiguidade lexical (Figura 2), entretanto, o ChatGPT não conseguiu desambiguar corretamente algumas frases que envolvem polissemia. Na frase “Pedi um prato ao garçom.”, o modelo apresentou duas opções de desambiguação com a seguinte resposta à tarefa 3: “Claro, aqui estão as duas possíveis interpretações desambiguadas da frase: Eu pedi um prato ao garçom. Eu pedi ao garçom um prato. Essas reestruturações eliminam a ambiguidade, deixando claro quem realizou a ação de pedir e o que foi pedido.”.

Percebe-se, então, que o modelo tende a encontrar problemas na estrutura, ao trocar a posição dos elementos sintáticos e não nas palavras isoladamente. O Gemini apresentou mais facilidade de identificação e explicação nos casos de homônimias, haja vista que todas as suas explicações nestes casos estão corretas. Por outro lado, também predominaram justificativas equivocadas em alguns casos de polissemia.

#### 4.4 Quais padrões de ambiguidade os modelos ChatGPT e Bard demonstram conhecer na geração de frases ambíguas?

A investigação sobre os padrões de ambiguidade usados pelos modelos ChatGPT e Gemini durante a geração de frases ambíguas é abordada por meio da análise das respostas obtidas na tarefa 4. Para conduzir essa análise, seis juízes-humanos qualificados julgaram se as frases geradas pelos modelos continham elementos que induzem ambiguidade perceptível por seres humanos e, em caso afirmativo, se esses elementos se alinhavam corretamente com a categoria de ambiguidade.

**Ambiguidade Lexical.** Na geração de frases com ambiguidade lexical, tanto o ChatGPT quanto o Gemini revelaram não conhecer os padrões geradores desse tipo de ambiguidade, resultando, em sua maioria, em frases sem qualquer forma de ambiguidade identificada pelos avaliadores humanos. O ChatGPT, em particular, não gerou nenhuma frase com ambiguidade lexical, tendo 18 frases avaliadas sem qualquer notificação de ambiguidade. Por exemplo, foram geradas frases como “O pássaro voou em direção à árvore mais alta.” e “Elá encontrou uma bela maçã na floresta.”, em que para um observador humano não há ambiguidade, revelando clareza na expressão. Em apenas duas das frases geradas, foram identificados padrões de adjunto ambíguo, referindo-se a ambiguidade estrutural ou sintática, não se enquadrando corretamente na categoria de ambiguidade lexical.

O Gemini também enfrentou dificuldades ao gerar frases com ambiguidade lexical. Das 20 frases, em 14 delas nenhum avaliador humano conseguiu identificar qualquer tipo de ambiguidade, resultando em uma taxa de sucesso de apenas 20% para o Gemini e de 0% para o ChatGPT. Um exemplo é a frase “O advogado defendeu o criminoso.” que não apresenta ambiguidade perceptível por um humano, pois com o contexto da frase é facilmente inferido que o advogado defendeu legalmente o criminoso. Foram observados apenas dois casos em que a ambiguidade residia na homônimia e polissemia, como na frase “O professor ensinou a classe.”. Além disso, o Gemini, em alguns casos, utilizou padrões que se alinham com a ambiguidade situacional, que é considerada parte da ambiguidade lexical por Cançado, por exemplo ao sugerir a frase “A casa está vazia” acompanhada da explicação que a casa pode estar vazia de pessoas ou de móveis. O desempenho nesta tarefa reforça o resultado de os modelos já possuem uma certa compreensão de ambiguidade, mas que ambos os modelos ainda enfrentam desafios na geração de frases em que o elemento gerador de ambiguidade é um item lexical.

**Ambiguidade Sintática.** Na análise da geração de frases com ambiguidade sintática, o ChatGPT apresentou um desempenho relativamente superior, com uma taxa de sucesso de 65%. Ou seja, das 20 frases geradas, 13 efetivamente incorporaram ambiguidade sintática por meio de complementos sintáticos, adjuntos adverbiais ou adnominais ambíguos, demonstrando que o modelo conseguiu aprender tais padrões de forma mais satisfatória. Por exemplo, na frase “Ela viu o homem com o telescópio,” o ChatGPT explorou a ambiguidade gerada pelo complemento sintático “com o telescópio,” permitindo interpretações tanto de ela ter utilizado um telescópio para ver o homem quanto de o homem estar com um telescópio quando foi visto. O modelo também utilizou ambiguidade oriunda de adjuntos adverbiais e adnominais, como em “A mãe deu um presente para a filha com uma fita bonita,” na qual o adjunto “com uma fita bonita” possibilita interpretações sobre o presente ter uma fita bonita ou a filha estar com uma fita bonita. Estes são os casos clássicos de ambiguidade sintática, amplamente descritos na literatura e explicados a partir dos princípios da aposição mínima e da aposição local (Maia, s.d.), com descrições no português brasileiro (Maia, M. A. R. et al., 2003; Maia, M. et al., 2004; Brito Bezerra; Martins Leitão, 2013).

Apesar dos resultados satisfatórios obtidos na geração das frases, foram encontradas inconsistências na interpretação do elemento gerador da ambiguidade em 7 das 20 frases, mesmo quando a ambiguidade sintática estava de fato presente. Além disso, o modelo gerou 6 frases sem identificação de qualquer tipo de ambiguidade pelos juízes-humanos. Em 20% das frases, o ChatGPT personificou

elementos não-humanos, prejudicando a interpretação correta. Por exemplo, na frase “O jogador marcou um gol com a camisa amarela.”, o ChatGPT colocou o adjunto, gerando uma sentença com potencial de ser sintaticamente ambígua, no entanto, a compreensão de senso comum barra a possibilidade da frase ter ambiguidade para humanos, pois a interpretação mais intuitiva linguisticamente será que o jogador estava vestindo uma camisa amarela e descartará o sentido de que o gol vestia uma camisa amarela.

Embora a nossa comunicação seja potencialmente ambígua, a ambiguidade não parece ser um problema, pois quando há interferência, a ambiguidade é resolvida com o esclarecimento, reparo ou correção. A IA atua como o analista, como explicam (Freitag; Cruz; Cunha Nascimento, 2021), que dispõem de um grande conjunto de dados desprovidos de contexto, e que portanto são propensos a gerar ambiguidade na compreensão, mas que só existem do ponto de vista da IA.

O Gemini demonstrou um conhecimento ainda mais limitado dos padrões geradores de ambiguidade sintática, gerando apenas 4 das 20 frases com adjuntos adnominais e adverbiais como fontes de ambiguidade, resultando em uma taxa de acertos de apenas 20%. Das 16 frases sem ambiguidade sintática, 11 não apresentam nenhum tipo de ambiguidade, como em “O homem comprou o livro que estava na prateleira,” em que erroneamente ela atribuiu ambiguidade à expressão “na prateleira.”, algo que um ser humano não faria, embora seja possível por conta dos princípios da aposição mínima e da aposição local (Brito Bezerra; Martins Leitão, 2013; Maia, s.d.). Além disso, em quatro sentenças, o Gemini confundiu o uso de palavras polissêmicas e homônimas, classificando erroneamente a ambiguidade lexical como sintática. Também é notável que, ao contrário do ChatGPT, o Gemini não incorporou a personificação de elementos inanimados em suas gerações.

Em 80% das frases geradas, as explicações fornecidas pelo Gemini foram incoerentes demonstrando que o modelo ainda não consegue explicar corretamente a causa da ambiguidade, como na frase “O homem viu a mulher na janela,” onde atribuiu a ambiguidade ao verbo “ver,” ignorando que a verdadeira fonte era o adjunto “na janela,” permitindo que o homem ou a mulher estivessem na janela. Outro problema observado foi a dificuldade dos modelos em lidar com ambiguidades sintáticas envolvendo a concordância de gênero, como no caso de adjuntos que podem concordar tanto com sujeito como o objeto. Na frase “O rapaz encontrou a carteira perdida no chão do parque”, o adjetivo “perdida” só pode ocupar a função de adjunto adnominal por se referir à carteira, objeto da sentença, devido à flexão de gênero. O termo não pode ser adjunto adverbial do verbo “encontrar”, porque não concorda com o sujeito do verbo, que está no masculino. Essa ocorrência dificilmente ocorreria em inglês, por ser uma língua sem uso generalizado de gênero, logo essa ambiguidade seria mantida com o uso do termo “lost”. Isso mostra que os resultados poderiam ter sido melhores se as tarefas fossem realizadas em inglês, já que o modelo é majoritariamente treinado nessa língua e evidencia a necessidade de mais estudos e investimento sobre os dados linguísticos dos modelos de linguagem na língua portuguesa.

**Ambiguidade Semântica.** É importante ressaltar que a ambiguidade semântica, em determinados referenciais teóricos, pode se assemelhar à ambiguidade lexical ou, em alguns casos, não é reconhecida como uma categoria distinta, resultando em uma linha tênue de separação entre ela e outros tipos de ambiguidade (Zavaglia, 2003). Contudo, os resultados gerados pelo ChatGPT e pelo Gemini não se manifestam apenas na mistura de diferentes padrões de ambiguidade. Em vez disso, destaca-se na geração de frases que carecem de qualquer ambiguidade, as quais, foram consideradas como possuidoras de ambiguidade semântica. O ChatGPT, por exemplo, produziu 7 frases que não apresentaram ambiguidade para os juízes-humanos, enquanto o Gemini gerou 15 frases nessas condições. As demais sentenças, em sua maioria, foram geradas com base em padrões de adjunto ambíguo ou, residindo apenas em elementos lexicais.

Ambos os modelos geraram frases semelhantes ou idênticas às produzidas para ambiguidade sintática. O ChatGPT gerou a sentença “Ela viu o homem com o telescópio.”, e o Gemini, “A menina viu o homem com o binóculo”. Ambas continham adjunto ambíguo, e padrões similares foram reproduzidos quando solicitadas frases com ambiguidade sintática, evidenciando a falta de distinção clara entre os dois tipos de ambiguidade por parte dessas versões dos modelos.

**Caso particular.** Apesar do Manual de Semântica (Cançado, 2005) não tratar diretamente de

substantivos compostos cristalizados pelo uso social e do nosso cotidiano, vale destacar duas sentenças geradas por ambos os modelos na Tarefa 4. O ChatGPT gerou a frase “Ele pegou a maçã verde do chão” e o Gemini produziu “O carro está parado no sinal vermelho.” Ambas as frases podem ser consideradas ambíguas devido aos termos compostos “maçã verde” e “sinal vermelho”. No primeiro caso, “maçã verde” pode se referir tanto à cor da fruta quanto a uma variedade específica de maçã, criando uma ambiguidade lexical. Já no segundo exemplo, “sinal vermelho” pode ser interpretado como a luz vermelha do semáforo, um alerta de perigo ou o poste do semáforo pintado de vermelho, resultando em diferentes interpretações.

Esses dois exemplos evidenciam um padrão reconhecido pelos modelos na geração de sentenças classificadas como ambiguidade semântica, que classificamos como lexical no nosso julgamento, por estar em conformidade com o que a autora propõe em sua taxonomia em relação aos itens com múltiplas inferências de significado. Os componentes lexicais “maçã verde” e “sinal vermelho” provocam ambiguidade ao apresentarem mais de um sentido possível para o termo, mas essa distinção só é válida para fins de sistematização da análise dos dados. Reforçamos que não presumímos que os modelos deveriam seguir a classificação adotada no estudo. Por conseguinte, não julgamos a nomenclatura como correta ou errada, mas somente a explicação que acompanhava as frases que foram geradas. Em vista dessas considerações, tal evidência pode servir como uma provocação para futuras pesquisas sobre a relação entre ambiguidade e modelos de linguagem, especialmente em como esses modelos lidam com a interpretação de termos cristalizados que carregam múltiplos sentidos lexicais.

## 5 Conclusão

Nossos resultados indicaram melhorias significativas entre os diferentes modelos, assim como diversas vantagens e limitações. Nesse sentido, nosso trabalho apresenta as seguintes contribuições: 1) disponibiliza um conjunto de dados formatado para testar a ambiguidade linguística em modelos de linguagem natural no português brasileiro, que até o nosso conhecimento é o primeiro proposto na literatura; 2) é o primeiro trabalho a informar à comunidade científica sobre as limitações do ChatGPT e do Gemini em compreender fenômenos linguísticos complexos, como a ambiguidade na língua portuguesa; 3) apresenta uma metodologia para avaliar esses modelos quanto ao fenômeno da ambiguidade; e 4) demonstra, por meio de resultados qualitativos e quantitativos, qual dos dois modelos lida melhor com esses fenômenos linguísticos.

A análise do fenômeno de ambiguidade linguística nos modelos instrucionais ChatGPT e Gemini, cujas versões 3.5 e Bard, respectivamente, foram submetidos a quatro tarefas referentes à detecção, tipificação, desambiguação e geração de frases ambíguas. Os resultados obtidos mostraram que ambiguidade linguística ainda é um grande desafio para estas versões de modelos de processamento de linguagem natural, demandando ainda estudos e implementações para o aprimoramento.

Os modelos apresentaram baixa acurácia e baixo desempenho em praticamente todas as tarefas testadas. Na detecção de ambiguidade, o ChatGPT conseguiu uma acurácia de 28,75% e o Gemini 49,58%. Os modelos também apresentaram uma superinterpretação de sentenças não ambíguas, detectando e desambiguando frases que não tinham qualquer tipo de ambiguidade e que seres humanos facilmente conseguem interpretar apenas um sentido nas sentenças. Os melhores resultados obtidos ocorreram na tarefa de desambiguação e classificação onde a ambiguidade residia apenas no item lexical, demonstrando que é o tipo de ambiguidade em que os modelos têm mais facilidade para lidar.

Merece destaque o fato de que, apesar da maior facilidade em lidar com ambiguidade lexical, os modelos tiveram o pior desempenho na geração de frases com ambiguidade desse tipo. Algo similar ocorreu com as frases de categoria semântica, em que os modelos confundiram com os padrões de ambiguidade estrutural ou geraram a maioria das frases sem ambiguidade. Por outro lado, ocorreu um melhor desempenho na geração de frases com ambiguidade sintática, porém ainda com uma interpretação errada sobre a origem da ambiguidade em várias frases, demonstrando que os modelos conseguiram gerar algumas frases corretamente mas ainda não conseguem explicar com clareza as causas da ambiguidade sintática. Outro ponto que chamou a atenção foi a tendência da ChatGPT em personificar alguns elementos inanimados para atribuir ambiguidade às frases, algo que um ser humano jamais faria. Os resultados relativos à identificação de ambiguidades se alinham ao que

(Marcus; Leivada; Murphy, 2023) obtiveram com a tarefa de letra trocada: enquanto seres humanos são excelentes nessa tarefa, LLMs apresentam dificuldades. A resolução de ambiguidades é então uma das habilidades que ainda estão faltando no estado atual de desenvolvimento desses modelos.

Em suma, os resultados mostram que estas versões dos modelos instrucionais ainda estão distantes de emular plenamente a capacidade cognitiva dos seres humanos, não só envolvendo a relação entre linguagem e identidade social (Freitag, 2021), mas também o uso na interação social, o que requer a compreensão de ambiguidades. No entanto, os resultados também sinalizam um progresso inicial na compreensão e aquisição do senso comum a respeito de como a linguagem humana funciona e reiteram a importância dos estudos descritivos em línguas ainda com poucos recursos, como é o caso do português (Finger, 2021; Freitag, 2022), para aprimoramento.

## Referências

- AGIRRE, Eneko; RIGAU, German. A proposal for word sense disambiguation using conceptual distance. *Amsterdam Studies In The Theory And History Of Linguistic Science Series 4*, John Benjamins Bv, p. 161–172, 1997.
- AHMED, Imtiaz; KAJOL, Mashrafi; HASAN, Uzma; DATTA, Partha Protim; ROY, Ayon; REZA, Md Rokonuzzaman. ChatGPT vs. Bard: A Comparative Study. *UMBC Student Collection*, 2023.
- BRITO BEZERRA, Gitanna; MARTINS LEITÃO, Márcio. O processamento de argumentos e adjuntos em construções sem ambiguidade estrutural. *Revista Veredas*, v. 17, n. 2, 2013.
- BRUSCATO, Amanda Maraschin; BAPTISTA, Jorge. Resolução de ambiguidade anafórica em português, inglês e espanhol (estudo-piloto). *Alfa: Revista de Linguística (São José do Rio Preto)*, SciELO Brasil, v. 65, e13626, 2021.
- CAI, Zhenguang G; HASLETT, David A; DUAN, Xufeng; WANG, Shuqi; PICKERING, Martin J. Does ChatGPT resemble humans in language use? *arXiv preprint arXiv:2303.08014*, 2023.
- CANÇADO, Márcia. Manual de semântica. Belo Horizonte: Editora UFMG, 2005.
- CHAPLOT, Devendra; BHATTACHARYYA, Pushpak; PARANJAPE, Ashwin. Unsupervised word sense disambiguation using markov random field and dependency parser. In: 1. PROCEEDINGS of the AAAI Conference on Artificial Intelligence. [S. I.: s. n.], 2015. v. 29.
- CHOMSKY, Noam. *Lectures on Government and Binding: The Pisa Lectures*. Berlin, New York: Mouton de Gruyter, 1981. (Studies in Generative Grammar).
- CHURCH, Kenneth W.; PATIL, Ramesh. Coping with Syntactic Ambiguity or How to Put the Block in the Box on the Table. In: PROCEEDINGS of the 20th Annual Meeting on Association for Computational Linguistics. [S. I.]: Association for Computational Linguistics, 1982. p. 139–146. DOI: 10.3115/981311.981336.
- DEVLIN, Jacob; CHANG, Ming-Wei; LEE, Kenton; TOUTANOVA, Kristina. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- FINGER, Marcelo. Inteligência Artificial e os rumos do processamento do português brasileiro. *Estudos Avançados*, SciELO Brasil, v. 35, p. 51–72, 2021.
- FREITAG, Raquel; TEJADA, Julian; PINHEIRO, Bruno; CARDOSO, Paloma. Função na língua, generalização e reproduzibilidade. *Revista da ABRALIN*, p. 1–27, 2021.
- FREITAG, Raquel Meister Ko. Preconceito linguístico para humanizar as máquinas. *Cadernos de Linguística*, v. 2, n. 4, e495–e495, 2021.
- FREITAG, Raquel Meister Ko. Sociolinguistic repositories as asset: challenges and difficulties in Brazil. *The Electronic Library*, Emerald Publishing Limited, v. 40, n. 5, p. 607–622, 2022.

FREITAG, Raquel Meister Ko; CARDOSO, Paloma Batista; TEJADA, Julian. Linguistic and paralinguistic constraints on the function of (eu) acho que as DM in Brazilian Portuguese: A multilevel approach. *Pragmatics & Cognition*, John Benjamins Publishing Company Amsterdam/Philadelphia, v. 29, n. 2, p. 324–346, 2022.

FREITAG, Raquel Meister Ko; CRUZ, Regina Célia Fernandes; CUNHA NASCIMENTO, Thiago da. A gramática no corpo: dos recursos corporificados na construção e negociação dos sentidos. *Cadernos de Linguística*, v. 2, n. 1, e354–e354, 2021.

GODOY, Mahayana; CARVALHO, Renata Sabrinne Souza de. Efeitos sintáticos e temáticos na resolução de pronomes ambíguos em Português Brasileiro. *Revista do GELNE*, v. 22, n. 2, p. 131–142, 2020.

GOEL, Anmol. *Beyond the Surface: A Computational Exploration of Linguistic Ambiguity*. 2023. Tese (Doutorado) – International Institute of Information Technology Hyderabad.

HABER, Janosch; POESIO, Massimo. Patterns of lexical ambiguity in contextualised language models. *arXiv preprint arXiv:2109.13032*, 2021.

ILARI, Rodrigues; GERALDI, João. *Semântica: Implicaturas e Ambigüidade Situacional*. São Paulo: Editora da Unicamp, 1987.

ITANKAR, Prashant Y; RAZA, Nikhat. Ambiguity resolution: An analytical study. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, v. 6, n. 2, p. 471–479, 2020.

JURAFSKY, Daniel; MARTIN, James H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 1st. Upper Saddle River, NJ: Prentice Hall, 2000.

KOLLER, Alexander; ALEVEN, Vincent E.; PIETQUIN, Olivier; RIEZLER, Stefan J. Learning to Parse for Belief Update. In: PROCEEDINGS of the 22nd International Conference on Computational Linguistics (Coling 2008). Manchester, UK: Coling 2008 Organizing Committee, 2008. p. 569–576. DOI: 10.5555/1599081.1599145.

LAPORTE, Éric. Resolução de ambiguidades. *Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações*, Caminho, p. 49–89, 2001.

LIMA SILVA, Henrique Miguel de; MELO, Rita de Cássia Freire de; FRANÇA, Rayane Bezerra de. Interpretação Preferencial de Pronomes Nulos em Frases Ambíguas no Português Brasileiro por Falantes Monolíngues Universitários do Curso de Letras: Um Estudo Off-Line. *ID on line. Revista de psicologia*, v. 17, n. 67, p. 145–152, 2023.

LIU, Alisa; WU, Zhaofeng; MICHAEL, Julian; SUHR, Alane; WEST, Peter; KOLLER, Alexander; SWAYAMDIPTA, Swabha; SMITH, Noah A; CHOI, Yejin. We're afraid language models aren't modeling ambiguity. *arXiv preprint arXiv:2304.14399*, 2023.

LYONS, John. *Semantics*. 1. [S. I.]: Cambridge University Press, 1977.

MACHADO, Maria Paula de Oliveira. Sintaxe dos advérbios de modo em Português. Porto:[Edição do Autor], 1996.

MAIA, Marcus. Dimensões do Processamento Sintático. *MOTA, MB; NAME, C.(org.)*

MAIA, Marcus; COSTA, Armanda; FERNÁNDEZ, Eva; LOURENÇO-GOMES, Maria do Carmo. A compreensão de orações relativas ambíguas em Português Brasileiro e Europeu: Um estudo comparativo. *Revista ABRALIN*, v. 3, n. 1, p. 2, 2004.

MAIA, Marcus Antonio Rezende; ALCANTARA, Shelen Nascimento de; BUARQUE, Simone Peres; SOUZA FARIA, Fernanda de. O processamento de concatenações sintáticas em três tipos de estruturas frasais ambíguas em português. *Fórum Linguístico*, v. 3, n. 1, p. 13–53, 2003.

MARCUS, Gary; LEIVADA, Evelina; MURPHY, Elliot. A sentence is worth a thousand pictures: Can large language models understand human language? *arXiv preprint arXiv:2308.00109*, 2023.

NASER, MZ; ALAVI, Amir H. Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences. *Architecture, Structures and Construction*, Springer, p. 1–19, 2021.

NOGUEIRA TEIXEIRA, Elisângela; MICELLI FONSECA, Maria Cristina; SOARES, Maria Elias. Resolução do pronomé nulo em Português Brasileiro: Evidência de movimentação ocular. *Revista Veredas*, v. 18, n. 1, 2014.

OPENAI. *GPT-4 Technical Report*. [S. I.: s. n.], 2023. arXiv: 2303.08774 [cs.CL].

ORTEGA-MARTÍN, Miguel; GARCÍA-SIERRA, Óscar; ARDOIZ, Alfonso; ÁLVAREZ, Jorge; ARMENTEROS, Juan Carlos; ALONSO, Adrián. Linguistic ambiguity analysis in ChatGPT. *arXiv preprint arXiv:2302.06426*, 2023.

OUYANG, Long; WU, Jeffrey; JIANG, Xu; ALMEIDA, Diogo; WAINWRIGHT, Carroll; MISHKIN, Pamela; ZHANG, Chong; AGARWAL, Sandhini; SLAMA, Katarina; RAY, Alex et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, v. 35, p. 27730–27744, 2022.

PADOVANI, Djalma. *Um método adaptativo para análise sintática do Português Brasileiro*. 2022. Tese (Doutorado) – Universidade de São Paulo.

PIANTADOSI, Steven. Modern language models refute Chomsky's approach to language. *LingBuzz Preprint, lingbuzz*, v. 7180, 2023.

POESIO, Massimo; ARTSTEIN, Ron. The Reliability of Anaphoric Annotation, Reconsidered: Taking Ambiguity into Account. In: *PROCEEDINGS of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*. [S. I.]: Association for Computational Linguistics, 2005. p. 76–83. DOI: 10.3115/1608890.1608900.

SINGH, Satyendr; SIDDIQUI, Tanveer J. Role of karaka relations in Hindi word sense disambiguation. *Journal of Information Technology Research (JITR)*, IGI Global, v. 8, n. 3, p. 21–42, 2015.

VASWANI, Ashish; SHAZER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N; KAISER, Łukasz; POLOSUKHIN, Illia. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017.

WEI, Jason; WANG, Xuezhi; SCHUURMANS, Dale; BOSMA, Maarten; ICHTER, Brian; XIA, Fei; CHI, Ed; LE, Quoc; ZHOU, Denny. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. [S. I.: s. n.], 2023. arXiv: 2201.11903 [cs.CL].

ZAVAGLIA, Claudia. Ambigüidade gerada pela homonímia: revisitação teórica, linhas limítrofes com a polissemia e proposta de critérios distintivos. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, SciELO Brasil, v. 19, p. 237–266, 2003.

## Contribuições dos autores

**Lavínia de Carvalho Moraes:** Conceituação, Curadoria de dados, Investigação, Escrita – rascunho original, Escrita – revisão e edição; **Irene Cristina Silvério:** Curadoria de dados, Investigação, Escrita – rascunho original; **Rafael Alexandre Sousa Marques:** Curadoria de dados, Investigação, Escrita – rascunho original; **Bianca de Castro Anaia:** Curadoria de dados, Investigação, Escrita – rascunho original; **Dandara Freitas de Paula:** Curadoria de dados, Investigação, Escrita – rascunho original; **Maria Carolina Schincariol de Faria:** Curadoria de dados, Investigação, Escrita – rascunho original; **Iury Cleveston:** Validação, Recursos, Aquisição de financiamento; **Alana de Santana Correia:** Conceituação, Validação, Recursos, Metodologia, Supervisão, Escrita – rascunho original, Escrita – revisão e edição; **Raquel Meister Ko Freitag:** Conceituação, Validação, Metodologia, Escrita – revisão e edição.

## A Apêndice

**Tabela 2.** Conjunto de sentenças com ambiguidade por nós desenvolvido. As categorias foram atribuídas de acordo com o referencial teórico da seção 2.

sentenças		categoria
João foi à mangueira, assim como Maria.	João quer mangas, assim como Maria.	L
João fez uma rezinha.	Isso não é legal!	L
Essa dama é linda!	Que gato!	L
Ana vai ao banco.	Acende logo!	L
O papel foi bem feito.	Ele está com aquela matraca.	L
A carteira foi danificada.	Ela estava perto da mangueira.	L
O homem esperava no banco.	Gostamos de rosa.	L
Eu gosto de damas.	Era o ponto certo.	L
Pedi um prato ao garçom.	Ficamos sem rede.	L
Pegue a pilha, por favor.	A rede caiu.	L
Minha mãe e minha irmã ficaram chateadas depois que ela gritou com ela.	Ela não gosta da amiga dela.	SE
João falou comigo e com sua mãe.	Maria e sua mãe falaram comigo.	SE
Ele perguntou para ela se estava bem.	Ana me contou um segredo sobre ela.	SE
Ele não sabia que era o dia do seu aniversário.	O professor dele escreveu várias coisas em seu caderno.	SE
O Estado deve ajudar o povo para que ele prospere.	O padre bateu o carro dele.	SE
Paulo não entrou na sua casa.	Ela viu o ônibus passar na rua dela.	SE
O policial prendeu o bandido em sua casa.	Para ele ficar satisfeito, o chefe preparou um prato para o convidado.	SE
O homem matou seu tigre.	A professora proibiu que o aluno utilizasse seu dicionário.	SE
Ele a viu com a sua amiga.	A moça colocou sua mão na tinta.	SE
José comprou pão para Maria perto de sua casa.	Os empregados se revoltaram contra os superiores por causa dos seus salários.	SE
Ele saiu da loja de carro.	Joana falou com a Maria brava.	SI
Carlos observou Tiago malhando.	Ele viu a moça com um binóculo.	SI
Abandonei meu irmão contrariado.	Soube do emprego novo de Janaína no restaurante.	SI
O menino viu o incêndio do prédio.	Pedro me mandou um cartão postal de Veneza.	SI
Idosa é presa por matar homem em crise na avenida.	A filha ligou para a mãe que tinha batido o carro.	SI
Policial prende criminoso com arma de brincadeira.	Eu li sobre a greve dos estudantes na universidade.	SI
Eu avisei à Júllia que estava atrasada.	Ricardo alimentou o pássaro cansado.	SI
O jogador o viu de tênis.	A mãe pegou o bebê chorando.	SI
A moça viu o rapaz andando com a amiga na rua.	Homens e mulheres inteligentes alcançam sucesso.	SI
Vendo carros e caminhões usados.	Ele foi atrás do táxi apressado.	SI

Fonte: Fonte: Própria.

**Tabela 3.** Conjunto de sentenças sem ambiguidade por nós desenvolvido.

sentenças	
Michel Teló tem uma sorveteria.	Eu cavo com a pá.
O lápis ficou desesperado quando viu que o papel acabou.	O padre bateu o carro dando ré.
João foi à árvore da mangueira, assim como Maria.	João e Maria gostam da fruta que se chama manga.
João engatou a ré.	Viajem à praia vocês!
Viajem vocês!	Números não é uma boa temática para filmes.
O homem, que ama sertanejo, estava ouvindo arrocha.	A conduta dele é inadequada, porque ele vende drogas.
João ganhou a partida de Canastra com uma dama na mão.	O gato fez cocô fora da caixinha de areia.
Eu gosto de damas, são tão educadas.	Ana vai ao banco sacar dinheiro.
Acende logo! Eu quero cozinar!	Ascenda e tenha prestígio.
Pedi o prato principal ao garçom, era filé!	Ontem ficamos sem rede, não dava mais pra navegar na internet.
A rede caiu, ninguém mais balançou!	A minha mãe, que estava brava, perguntou que horas eu cheguei.
Carlos, que estava malhando, observou Thiago.	Ele, por meio de um binóculo, viu a moça.
Proibido ultrapassar.	Janaína começará a trabalhar num restaurante.
O bandido foi preso em flagrante pelo policial.	Pegaram emprestado o livro muito culto dele.
Enquanto o menino estava no prédio, viu um incêndio.	Em Veneza, Pedro comprou um cartão postal pra mim.
A idosa estava em uma crise e matou, na avenida, o homem.	A filha, após sofrer um acidente de carro, ligou para a mãe.
Um policial, portando uma arma de brinquedo, prende criminoso.	Na universidade, os estudantes faziam uma greve, que eu vi.
Eu avisei atrasadamente à Júlia.	Como o pássaro estava cansado, Ricardo o alimentou.
O Estado deve ajudar o povo, que prosperará.	O tigre matou o homem.
Falei que estava passando mal à chefe.	O chefe preparou um prato para que o convidado ficasse satisfeito.
A filha gritou com a mãe e ambas ficaram chateadas.	Ela não gosta daquela mulher.
João falou comigo depois de falar com a mãe pelo telefone.	Conversei com José e a mãe dele, a qual o acompanhava.
Ele foi embora da loja dirigindo.	Ela estava doente e avisou o chefe.
Ele perguntou se ela estava bem.	Ana me contou um segredo.
Ele não sabia que já era o dia do próprio aniversário.	O professor fez várias correções no caderno do aluno.
Ele está conversando com aquela mulher.	Paulo não foi até a universidade.
Ela não viu o ônibus que passou.	A porta foi danificada.
Ela estava perto da árvore.	Pegue as pilhas de roupas, por favor.
Gostamos da cor rosa.	O homem esperava sentado.
Ela é profissional em tênis de mesa.	Era o ponto de ônibus certo.

Fonte: Fonte: Própria.

**Tabela 4.** Conjunto de sentenças com ambiguidade lexical geradas pelos modelos de linguagem no experimento 4. A coluna de categoria real indica a classificação da sentença conforme a percepção humana da ambiguidade. Sentenças categorizadas como sem ambiguidade são aquelas em que seis seres humanos não detectaram qualquer ambiguidade. As demais categorias foram atribuídas com base no referencial teórico adotado neste trabalho.

modelo	categoria solicitada	sentença	categoria real	sentença	categoria real
ChatGPT	lexical	O pássaro voou em direção à árvore mais alta.	sem ambiguidade	Ele viu o filme no cinema com os óculos escuros.	sem ambiguidade
ChatGPT	lexical	A fuga do preso foi rápida e espetacular.	sem ambiguidade	Ela encontrou uma bela maçã na floresta.	sem ambiguidade
ChatGPT	lexical	O livro estava na mesa do professor.	sem ambiguidade	O jogador ferido foi substituído por um novato.	sem ambiguidade
ChatGPT	lexical	Ele comprou uma camisa azul e uma vermelha.	sem ambiguidade	O médico recomendou repouso e exercício.	sem ambiguidade
ChatGPT	lexical	A aula de biologia foi sobre células e tecidos.	sem ambiguidade	O rapaz correu para a casa da avó.	sem ambiguidade
ChatGPT	lexical	Ela pintou a sala com as mãos.	sem ambiguidade	O barco ancorou na ilha deserta.	sem ambiguidade
ChatGPT	lexical	O cachorro late quando o carteiro chega.	sem ambiguidade	O guarda fechou a porta do parque.	sem ambiguidade
ChatGPT	lexical	Ele comeu o bolo com gosto.	sintática	O homem viu o acidente com os próprios olhos.	sem ambiguidade
ChatGPT	lexical	O carro saiu da garagem com dificuldade.	sem ambiguidade	O escritor escreveu um romance emocionante.	sem ambiguidade
ChatGPT	lexical	O cantor entrou no palco com uma guitarra.	sintática	A mãe preparou a comida com amor.	sem ambiguidade
Gemini	lexical	O cachorro do vizinho mordeu a criança.	sintática e lexical	O homem comprou uma casa no campo.	sintática
Gemini	lexical	O advogado defendeu o criminoso.	sem ambiguidade	O soldado matou o inimigo.	sem ambiguidade
Gemini	lexical	O médico operou o paciente.	sem ambiguidade	O professor ensinou a classe.	lexical
Gemini	lexical	O juiz sentenciou o réu.	sem ambiguidade	O político falou para a multidão.	sem ambiguidade
Gemini	lexical	O artista pintou um quadro.	sem ambiguidade	O escritor escreveu um livro.	sem ambiguidade
Gemini	lexical	O carro preto passou por nós.	sem ambiguidade	A menina gorda está sentada ali.	sem ambiguidade
Gemini	lexical	A casa está vazia.	situacional <sup>4</sup>	A rua está molhada.	sem ambiguidade

<sup>4</sup> A ambiguidade situacional ocorre quando o contexto é crucial para determinar a interpretação de uma sentença, já que o enunciado não fornece pistas suficientes para uma única interpretação, e diferentes implicaturas podem ser geradas a partir das circunstâncias. No caso da frase "A casa está vazia", que pode significar que a casa está sem moradores ou está sem móveis e outros objetos dentro dela, a ambiguidade surge a partir da falta de informações contextuais claras. Segundo (Ilari; Geraldi, 1987), essa é uma forma de ambiguidade que depende fortemente do contexto situacional e das inferências do ouvinte, já que ele precisará de outros indícios, como a situação em que a frase foi dita ou o tom de voz do falante, para escolher entre as possíveis implicaturas da sentença.

Gemini	lexical	O livro está fechado.	sem ambiguidade	A porta está aberta.	situacional <sup>5</sup>
Gemini	lexical	O copo está cheio.	sem ambiguidade	A mesa está limpa.	situacional <sup>6</sup>
Gemini	lexical	O homem velho está dormindo na rua.	sem ambiguidade	A cama está arrumada.	sem ambiguidade

Fonte: Fonte: Própria.

**Tabela 5.** Conjunto de sentenças com ambiguidade sintática que foram geradas pelos modelos de linguagem no experimento 4. A coluna de categoria real indica a classificação da sentença conforme a percepção humana da ambiguidade. Sentenças categorizadas como sem ambiguidade são aquelas em que seis seres humanos não detectaram qualquer ambiguidade. As demais categorias foram atribuídas com base no referencial teórico adotado neste trabalho.

modelo	categoria solicitada	sentença	categoria real	sentença	categoria real
ChatGPT	sintática	Ela viu o homem com o telescópio.	sintática	Ele trouxe o livro para o amigo do professor.	sem ambiguidade
ChatGPT	sintática	O rapaz encontrou o carro dirigindo na estrada.	sintática	A professora elogiou o aluno com boas notas.	sintática
ChatGPT	sintática	A mãe deu um presente para a filha com uma fita bonita.	sintática	O cachorro mordeu o homem com a camisa rasgada.	sem ambiguidade
ChatGPT	sintática	O pintor viu o quadro com os olhos cansados.	sintática	O jogador marcou um gol com a camisa amarela.	sem ambiguidade
ChatGPT	sintática	O médico disse ao paciente que o tratamento seria difícil devido à sua saúde frágil.	sintática e semântica	O rapaz encontrou a carteira perdida no chão do parque.	lexical
ChatGPT	sintática	Vi o homem com o telescópio.	sintática	Ele viu o pássaro do vizinho com binóculos.	sem ambiguidade
ChatGPT	sintática	Eu gosto de estudar mais do que meu irmão.	sintática	Ela viu o homem com a luneta sentado na varanda.	sintática
ChatGPT	sintática	O cachorro do meu amigo late durante a noite.	sintática e lexical	Maria ama sua irmã mais do que eu.	sintática
ChatGPT	sintática	Comi a pizza com queijo.	sintática	Vi a fotografia da Maria com o cachorro.	sintática
ChatGPT	sintática	Ele quebrou a janela com a bola de beisebol jogando no quintal.	sem ambiguidade	O professor disse ao aluno que a matéria era fácil.	sem ambiguidade
Gemini	sintática	A menina viu o cachorro do vizinho correndo pela rua.	sintática e lexical	O homem levou a esposa ao médico.	sem ambiguidade
Gemini	sintática	O menino caiu da árvore e quebrou o braço.	semântica e lexical	A mulher comprou o vestido verde na loja.	sem ambiguidade
Gemini	sintática	O homem viu a mulher na janela.	sintática e lexical	O gato comeu o rato que estava no armário.	sem ambiguidade

<sup>5</sup> É o mesmo caso de "A casa está vazia", pois a porta pode estar destrancada ou estar entreaberta, ou seja, visivelmente aberta.

<sup>6</sup> É o mesmo caso de "A casa está vazia", pois a mesa pode estar sem nenhum objeto em cima ou não estar com sujeira.

Gemini	sintática	O menino jogou a bola para o cachorro.	sem ambiguidade	A menina estava brincando com a boneca quando a mãe chegou.	sem ambiguidade
Gemini	sintática	O homem comprou o livro que estava na prateleira.	sem ambiguidade	O menino tirou a roupa e foi tomar banho.	sem ambiguidade
Gemini	sintática	O cachorro do meu tio só dá trabalho.	sintática e lexical	Carla falou pro marido que tinha chegado.	sintática
Gemini	sintática	Eu fiquei te esperando no banco.	lexical	Ela puxou a cadeira e acabou quebrando o braço.	semântica e lexical
Gemini	sintática	A galinha brincava com uma xícara até que quebrou a asa.	semântica e lexical	O guarda deteve o suspeito em sua casa.	semântica
Gemini	sintática	O cachorro comeu o osso e o menino.	sem ambiguidade	O homem comprou um carro novo e uma casa.	sem ambiguidade
Gemini	sintática	A mulher estava com medo do cachorro.	sem ambiguidade	O menino estava andando de bicicleta na rua.	sem ambiguidade

Fonte: Fonte: Própria.

**Tabela 6.** Conjunto de sentenças com ambiguidade semântica que foram geradas pelos modelos de linguagem no experimento 4. A coluna de categoria real indica a classificação da sentença conforme a percepção humana da ambiguidade. Sentenças categorizadas como sem ambiguidade são aquelas em que seis seres humanos não detectaram qualquer ambiguidade. As demais categorias foram atribuídas com base no referencial teórico adotado neste trabalho.

modelo	categoria solicitada	sentença	categoria real	sentença	categoria real
ChatGPT	semântica	Ela viu o homem com o telescópio.	sintática	O banco estava vazio, exceto por um único ocupante.	lexical
ChatGPT	semântica	Ele pegou a maçã verde do chão.	lexical	Vi o pássaro da janela enquanto dirigia.	sintática e lexical
ChatGPT	semântica	O jogador atirou a bola para o goleiro.	sem ambiguidade	Ele encontrou a carta de Maria na gaveta.	lexical
ChatGPT	semântica	A aula foi interessante para os alunos cansados.	sem ambiguidade	A luz do sol iluminava a sala escura.	sem ambiguidade
ChatGPT	semântica	A pele do caçador era macia.	lexical	O cachorro viu o gato no telhado com binóculos.	sintática
ChatGPT	semântica	Ela viu o homem com o telescópio.	sintática	Ele encontrou a mulher de chapéu na festa.	sintática
ChatGPT	semântica	O pato é bom para comer.	sintática	As crianças viram o filme no parque.	sem ambiguidade
ChatGPT	semântica	O ladrão roubou o banco com uma máscara.	sintática e lexical	Ele leu o livro com os óculos escuros.	sem ambiguidade
ChatGPT	semântica	O cachorro viu o gato na varanda.	sintática	Ela pintou o quadro com a mão.	sintática
ChatGPT	semântica	O jogador ferido saiu do campo mancando.	sem ambiguidade	A situação melhorou após a conversa com o terapeuta.	sem ambiguidade
Gemini	semântica	A menina viu o homem com o binóculo.	sintática	O homem viu a mulher com o cachorro	sintática.

Gemini	semântica	O menino quebrou o braço da cadeira.	sem ambiguidade	A casa do João é azul.	sem ambiguidade
Gemini	semântica	O livro do menino está na mesa.	sem ambiguidade	O homem está morto.	lexical
Gemini	semântica	O menino está com fome.	sem ambiguidade	A mulher está cansada.	sem ambiguidade
Gemini	semântica	O cachorro está feliz.	sem ambiguidade	O gato está dormindo.	sem ambiguidade
Gemini	semântica	O homem está procurando um cachorro.	sem ambiguidade	O menino foi ao banco.	lexical
Gemini	semântica	O guarda prendeu o suspeito em sua casa.	semântica e sintática	O professor deu uma bronca na aluna.	sem ambiguidade
Gemini	semântica	O carro está parado no sinal vermelho.	lexical	O homem está com fome.	sem ambiguidade
Gemini	semântica	A mulher está com frio.	sem ambiguidade	O menino está com medo.	sem ambiguidade
Gemini	semântica	A menina está com alegria.	sem ambiguidade	O homem está com raiva.	sem ambiguidade

Fonte: Fonte: Própria.