



ANÁLISE DA AMBIGUIDADE LINGUÍSTICA NA “CHATGPT”

Palavras-Chave: AMBIGUIDADE, INTELIGÊNCIA ARTIFICIAL, CHATGPT

Autores(as):

Lavinia de Carvalho Moraes, IEL – UNICAMP

Alana de Santana Correia, Iury Cleveston - Attenty

Coautores/as:

Bianca de Castro Anaia [IEL], Dandara Freitas de Paula [IEL], Irene Cristina Silvério [IEL], Maria Carolina Schincariol de Faria [IEL], Rafael Alexandre Sousa Marques [IEL]

INTRODUÇÃO:

Em março de 2023, a Startup Open.AI anunciou o lançamento da versão GPT-4 do ChatGPT¹, uma inteligência artificial revolucionária que entrou no mercado em 2022. Essa tecnologia atraiu e conquistou usuários ao redor do mundo por ser um recurso digital que contempla uma base de dados altamente aperfeiçoada. Com a ChatGPT é possível solucionar dúvidas de variadas áreas do conhecimento, desenvolver códigos de programação, elaborar trabalhos acadêmicos, dentre muitas outras operações, logo suas funcionalidades atingem diretamente a distribuição de informação, a comunicação de ideias e a compreensão de dados discursivos.

Diversos debates têm sido gerados sobre os impactos dessa ferramenta no que diz respeito aos seus propósitos e aos seus resultados, pois mais do que decodificar e redigir textos, é uma inteligência ligada diretamente com os fenômenos da linguagem e do conhecimento humano. Dentre as respostas dadas pela ChatGPT, emergem incertezas em relação ao seu comportamento nos campos da linguística, assim antigos desafios para estudiosos da área, como a gramática, a morfologia, a sintaxe e a semântica, servem como alvo de análise do seu desempenho.

Diante desse cenário, a ambiguidade linguística aparece como um fenômeno que reflete bem aspectos da qualidade da performance dessa tecnologia, pois testa sua capacidade de construção de diferentes sentidos e compreensão sobre os variados significados presentes em uma mesma frase ou texto, atributos essenciais para entregar bons resultados nas suas funções de leitura e escrita. Logo, o presente estudo visa investigar a capacidade da ChatGPT em detectar e lidar com ambiguidades da Língua Portuguesa.

Para alcançar esse objetivo serão apresentados quatro experimentos distintos, cada um focado em uma etapa específica da análise. Além disso, a pesquisa propõe o desenvolvimento de um

¹ O termo ChatPGT será tratado ao longo do resumo por meio de artigo e pronome no feminino e no masculino, pois pode-se referir ao ChatGPT como uma inteligência artificial, palavra feminina, ou pelo uso recorrente, no Brasil, do termo no masculino .

conjunto de dados composto por 120 frases, sendo 60 delas ambíguas e as outras 60 não ambíguas, abrangendo diferentes contextos e tópicos para verificarmos de que forma a inteligência opera para distinguir as sentenças e qual é o seu nível de competência para cumprir essa atividade.

METODOLOGIA:

Para avaliar como a ChatGPT lida com a ambiguidade, conduzimos uma série de experimentos utilizando um conjunto de dados compilados, uma vez que não identificamos conjuntos disponíveis para esta finalidade. Esse fato destaca a natureza pioneira do nosso estudo, no contexto da resolução de ambiguidades linguísticas da língua portuguesa.

O nosso conjunto é composto por 120 sentenças, sendo 60 com ambiguidade lexical², sintática³, e semântica (20 de cada), e 60 sem ambiguidade (Tabela 1). Essa classificação foi feita por meio do reconhecimento do elemento gramatical que causa as múltiplas interpretações dentro da frase. Logo, no caso da ambiguidade lexical foram considerados os casos em que a palavra assume dois ou mais significados diferentes e pode se manifestar à nível fonético e sintático. O Dado 1 (Tabela 1), por exemplo, traz a palavra mangueira, que pode indicar uma mangueira de água ou a árvore de uma mangueira, pois ambas são semelhantes foneticamente.

A próxima ambiguidade tratada é a semântica, as ocorrências classificadas como tal se caracterizam por haver mais de uma interpretação possível para o termo, considerando o sentido geral da frase. À vista do Dado 2 (Tabela 1), há duas explicações possíveis para a frase, a primeira seria que a pessoa saiu de uma loja que vende automóveis e a segunda seria que ele saiu da loja dirigindo um veículo. Por fim, para a ambiguidade sintática, foram considerados os enunciados que fazem referência ao duplo sentido causado pela organização estrutural da sentença, como a posição de uma preposição, uso de vírgula, dentre outros aspectos. No Dado 3 (Tabela 1), é possível notar a ambiguidade causada pelo uso da preposição “com”, que não determina quem estava com a arma, o policial ou o criminoso.

Amostras de sentenças do conjunto de dados.		
	Sentença	Classe
Dado 1	João foi à mangueira, assim como Maria.	Ambiguidade lexical
Dado 2	Ele saiu da loja de carro.	Ambiguidade semântica
Dado 3	Policial prende criminoso com arma de brincadeira.	Ambiguidade sintática

Tabela 1 – Amostra das sentenças criadas para realizar a análise.

Visando observar como a ChatGPT lida com a ambiguidade nas sentenças criadas, realizamos quatro experimentos distintos. Em todos os experimentos, nós conduzimos coletas duplicadas das interações para cada frase. Entre cada coleta, reiniciamos o console para evitar qualquer influência do

² A ambiguidade lexical e léxica são sinônimos, logo ambos os termos são usados na literatura.

³ A ambiguidade sintática também costuma ser chamada de ambiguidade estrutural dentro da literatura.

contexto que pudesse gerar respostas tendenciosas. Essa abordagem permitiu avaliar a consistência do modelo nas respostas fornecidas.

Iniciamos as coletas com o **experimento 1**, cujo objetivo é verificar se a ChatGPT consegue identificar a presença ou ausência de ambiguidade em cada sentença. Para isso, apresentamos individualmente as sentenças com e sem ambiguidade à ChatGPT e registramos suas respostas, comparando-as com a classificação prévia para medir a precisão da detecção (Tabela 2). As respostas foram cuidadosamente avaliadas quanto à correção e abrangência das explicações fornecidas, classificando-as entre corretas e incorretas.

Exemplo de prompt de entrada para verificar a presença de ambiguidade na ChatGPT.						
Exemplo de prompt na chatGPT	Classe verdadeira	Resposta da ChatGPT	Classificação da resposta dada pela ChatGPT após análise			
			Verdadeiro positivo	Falso positivo	Verdadeiro negativo	Falso negativo
A sentença “[sentença]” é ambígua? Responda, sim, não ou não sei.	Ambiguidade	Sim, a sentença é ambígua ...	X			

Tabela 2 – Exemplo de prompt para coletar informações da ChatGPT para o experimento 1.

Após a coleta e a análise das respostas, nós computamos a taxa de verdadeiros positivos (sentenças que são ambíguas e que a ChatGPT classificou como ambíguas), falsos positivos (sentenças que não são ambíguas e que a ChatGPT classificou como ambíguas), verdadeiros negativos (sentenças que não são ambíguas e que a ChatGPT classificou como não ambíguas), e falsos negativos (sentenças que são ambíguas e que a ChatGPT classificou como não ambíguas) e discutimos os resultados.

No **experimento 2**, realizamos um teste para avaliar a capacidade da ChatGPT em distinguir corretamente entre as três classes de ambiguidade estudadas neste trabalho. O experimento consistiu em perguntar qual o tipo de ambiguidade da sentença que foi classificada anteriormente por ela como ambígua ou não ambígua. Para mensurar a qualidade da classificação, nós usamos a métrica acurácia para medir o quão preciso é um modelo em suas previsões ou classificações. Em nosso contexto, a acurácia representa a proporção de frases corretamente classificadas pela ChatGPT para um determinado tipo de ambiguidade em relação ao total de frases apresentadas no experimento (ambíguas de uma determinada categoria e não ambíguas). Por exemplo, se o modelo classificou corretamente 80 frases de um total de 100 para um determinado tipo de ambiguidade, a acurácia é determinada por:

$$A = \frac{\text{número de frases corretamente classificadas}}{\text{total de frases no experimento}} \times 100 = \frac{80}{100} \times 100 = 80 \%$$

No **experimento 3**, nosso objetivo é verificar a capacidade da ChatGPT em desambiguar as sentenças que foram fornecidas a ela. Com esse propósito, apresentamos uma série de frases que incluem tanto sentenças ambíguas quanto sentenças não ambíguas. O experimento busca testar a

habilidade do modelo em compreender e interpretar o contexto, escolhendo a interpretação mais apropriada quando a ambiguidade está presente. Por fim, no **experimento 4** nós avaliamos a capacidade da ChatGPT em gerar frases ambíguas na categoria correta. Para isso, pedimos para ela gerar 15 frases de cada categoria de ambiguidade e avaliamos as taxas de erros e acertos de cada resposta. Ao realizar essa avaliação, buscamos compreender quão precisa é a ChatGPT ao criar frases que apresentam múltiplas interpretações contextuais.

RESULTADOS E DISCUSSÃO:

Os primeiros testes foram feitos a partir de frases com ambiguidade sintática, na qual a ChatGPT foi perguntada sobre a ambiguidade de 40 sentenças em que metade das proposições eram ambíguas e a outra metade não apresentavam mais de um sentido. Destaca-se que cada frase foi testada duas vezes para verificar a consistência dos resultados, resultando em 80 respostas obtidas. Todos os resultados obtidos no **experimento 1** são mostrados na matriz de confusão da Figura 1.

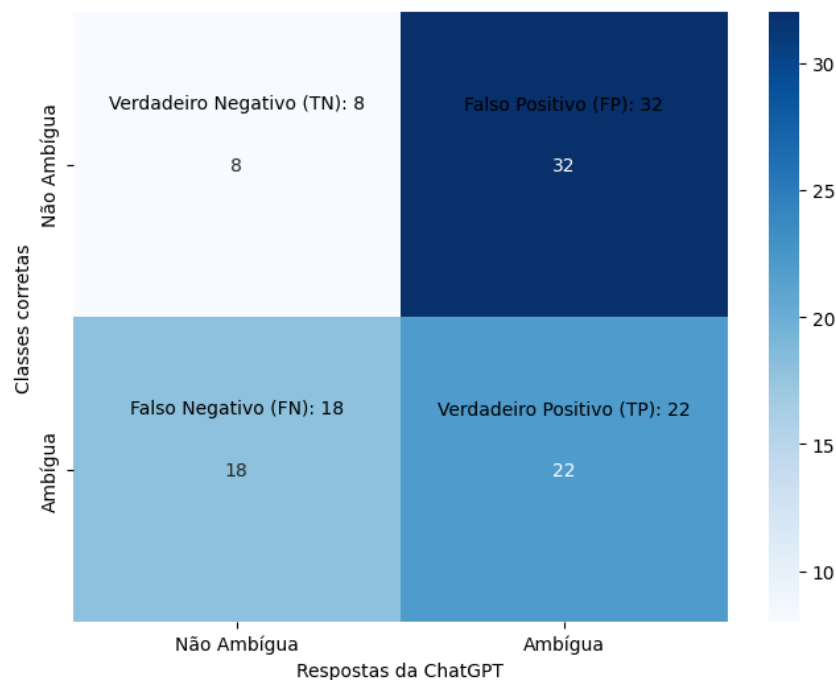


Figura 1 – Matriz de confusão mostrando os resultados do experimento 1.

Os dados revelaram 8 resultados de verdadeiro negativo, 32 de falso positivo, 18 de falso negativo e 22 de verdadeiro positivo. Esses resultados demonstram a dificuldade da ChatGPT em classificar sentenças não ambíguas corretamente. Adicionalmente, durante o experimento 2, a ChatGPT classificou erroneamente essas sentenças em categorias equivocadas de ambiguidade. Além disso, ela demonstrou considerável dificuldade em identificar a presença de ambiguidade em sentenças verdadeiramente ambíguas (falsos negativos). Tais observações ressaltam a necessidade de aprimoramentos em sua capacidade de lidar com a ambiguidade linguística.

Com relação à consistência das respostas quando questionada pela segunda vez sobre a detecção da ambiguidade na frase, das 40 perguntas repetidas, a ChatGPT apresentou respostas

inconsistentes em 12 delas. Ou seja, ao ser questionada pela primeira vez, ela forneceu uma resposta, mas ao ser questionada novamente sobre a presença ou não de ambiguidade, sua resposta foi diferente, apesar de termos redefinido o console para eliminar o contexto e realizado a segunda pergunta após um intervalo de tempo da primeira.

Com relação às frases que não apresentam ambiguidade, dentre os 20 casos, a ChatGPT só registrou 3 verdadeiros negativos, conseguindo reconhecer que não havia duplo sentido, porém desses resultados, obteve-se 2 casos em que ao solicitar a classificação da ambiguidade, ela o fez. Em outras 15 frases, ela gerou um falso positivo por avaliar que não havia ambiguidade ou dizer que não sabia, porém ao ser perguntada posteriormente sobre o tipo de ambiguidade, ela informou o que estava causando a ambiguidade em todos, contrariando seu resultado inicial. Dentro deste grupo de falsos positivos, o ChatGPT mudou sua resposta de um teste para outro, em relação à classificação da ambiguidade, 8 vezes. Essas alterações de resposta resultaram em 16 nomenclaturas diferentes de classificação ao longo de todo o teste com as 40 frases, demonstrando assim, que apesar de ter uma base de dados muito completa, ela gera alguns resultados inconsistentes.

No **experimento 2**, a ChatGPT conseguiu classificar corretamente a ambiguidade como “sintática” ou “estrutural” em apenas 40 tentativas de 80, apresentando uma acurácia de apenas 50%. Ela acertou apenas em 4 casos que a sentença não tem ambiguidade e 4 sentenças ambíguas ela não conseguiu gerar uma classificação.

CONCLUSÕES:

É válido destacar que apesar do variado uso de nomenclatura, a ChatGPT seguiu uma linha coerente de análise em alguns dos casos de ambiguidade, apresentando aspectos que de fato constituem o quadro de causas para a ambiguidade sintática. Porém, como ela não assume uma metodologia específica do campo da linguística, as informações são explicadas de formas diferentes a cada nova consulta feita, sem haver um processamento lógico embasado em um referencial teórico. Além disso, ainda existe bastante dificuldade do modelo em classificar corretamente frases não ambíguas, demonstrando que as respostas do modelo são enviesadas em muitos casos de teste e não são totalmente confiáveis.

BIBLIOGRAFIA

TESCARI NETO, Aquiles. **Constituição sintática, ambiguidade estrutural e aula de português: o lugar da teoria gramatical no ensino e na formação do professor**. Working papers em Linguística, Campinas, v. 18, p. 129-152, 2017.

ORTEGA-MARTÍN, Miguel. et al. **Linguistic ambiguity analysis in ChatGPT**. arXiv preprint arXiv:2302.06426, 2023.