

## Chat GPT vs. Google AI Studio: Uma Análise Lexicográfica da Homonímia em Dicionários Gerados por Inteligência Artificial

Raquel de Oliveira<sup>1</sup>

Instituto Federal de Mato Grosso do Sul

### IV Seminário de Pós-graduação do IFMS – SEMPOG IFMS 2024

**Resumo.** A Inteligência Artificial (IA) tem impactado todas as áreas e suscitado diferentes reflexões acerca de suas potencialidades. Os holofotes estão voltados, atualmente, sobretudo, para as IA gerativas de modelo de linguagem, entre as quais o Chat GPT tem se destacado. Com efeito, as máquinas agora são capazes de imitar a linguagem humana e geram textos de diferentes tipologias com grande agilidade e precisão, sendo capazes de produzir, por meio de prompts (comandos) simples, inclusive dicionários de língua. Assim, este estudo compara o tratamento lexicográfico da homonímia em dois dicionários de inglês criados por IA: um criado pelo Chat GPT 3.5 e outro pelo Google AI Studio. Utilizando princípios da Lexicografia Pedagógica, o estudo analisa a estrutura dos dicionários, o tratamento de homófonos não homógrafos em seus níveis macro e microestruturais e avalia se atendem aos critérios de dicionários pedagógicos de acordo com autores como Prado Aragonés (2005) e Azorín Fernández (2009). Ademais, a comparação entre as informações geradas pelas IA busca identificar semelhanças e originalidades no tratamento desses homófonos. Nossa análise, com os dados colhidos até o presente momento, demonstra a capacidade da IA de gerar dicionários personalizados, abrindo caminho para uma nova era de lexicografia digital e impulsionando pesquisas nesse campo.

**Palavras-chave.** Lexicografia, dicionários, IA.

**Abstract.** Artificial Intelligence (AI) has impacted all areas and raised different reflections about its potential. The spotlight is currently focused mainly on language model generative AI, among which GPT Chat has stood out. In effect, machines are now capable of imitating human language and generate texts of different types with great agility and precision, being able to produce, through simple prompts (commands), including language dictionaries. Therefore, this study compares the lexicographic treatment of homonymy in two English dictionaries created by AI: one created by Chat GPT 3.5 and the other one by Google AI Studio. Using principles of Pedagogical Lexicography, the study analyzes the structure of dictionaries, the treatment of non-homograph homophones at their macro and microstructural levels and evaluates whether they meet the criteria of pedagogical dictionaries according to authors such as Prado Aragonés (2005) and Azorín Fernández (2009). Furthermore, the comparison between the information generated by AI seeks to identify similarities and originalities in the treatment of these homophones. Our analysis, with the data collected so far, demonstrates AI's ability to generate personalized dictionaries, paving the way for a new era of digital lexicography and boosting research in this field.

**Key words.** Lexicography, dictionaries, AI.

## Introdução

A Inteligência Artificial (IA) tem impactado todas as áreas e suscitado diferentes reflexões acerca de suas potencialidades. Os holofotes estão voltados, atualmente, sobretudo, para as IA gerativas de modelo de linguagem, entre as quais o Chat GPT tem se destacado. Com efeito, as máquinas agora são capazes de imitar a linguagem humana e geram textos de diferentes tipologias com grande agilidade e precisão, sendo capazes de

---

<sup>1</sup> Mestre em Letras pela UFMS/CPTL: raquel.de.oliveira@outlook.com.

produzir, por meio de *prompts* (comandos) simples, inclusive dicionários de língua. Assim, conforme aponta Lew (2023), editor do *International Journal of Lexicography*, apesar de o surgimento do Chat GPT ter representado um marco significativo na área de IA, o ônus disso é que, ao mesmo tempo, ele desafiou a tradicional posição ocupada pelos dicionários como recursos confiáveis de aprendizagem de línguas. Veremos os desdobramentos disso neste trabalho.

Essas reflexões são todas muito recentes, afinal, o Chat GPT foi lançado em dezembro de 2022 pela Open AI. E, desde então, já passou por várias atualizações e modificações. No momento de escrita deste artigo, por exemplo, já foi lançado o Chat GPT 4.0, que, segundo afirmam seus criadores, é infinitamente mais rápido, mais preciso e capaz de muito mais que o Chat GPT 3.5, analisado nesta pesquisa. Entretanto, nossa escolha inicial pelas ferramentas Google AI Studio (lançado em 2024 pelo Google) e Chat GPT 3.5 se deveu principalmente ao fato de serem ambas gratuitas, enquanto atualizações, como o Chat GPT 4.0, são pagas.

Dissemos que o Google AI Studio e o Chat GPT 3.5 são exemplos de ferramentas de IA gerativas porque, com efeito, consiste em uma tradução inadequada chamá-las de “gerativas”, como vários cursos disponíveis no mercado têm feito. A confusão parece ter surgido por meio de uma tradução literal feita do seu equivalente em inglês: AI (GenAI), mas em português, não existe o verbo “generar” e, sim, “gerar”.

Assim, de acordo com um guia elaborado pela Universidade de Cambridge (2023, p. 3, tradução nossa) e destinado a professores de idiomas, “IA gerativa é uma forma de Inteligência Artificial que utiliza grandes quantidades de informações e poderosos modelos de aprendizado de máquina para criar conteúdo original, incluindo texto, áudio, código, imagens, simulações e vídeo”<sup>2</sup>. De fato, uma das grandes novidades dessas ferramentas de IA é sua capacidade de imitar a linguagem humana, o que as torna capazes de criar textos de diversos gêneros, incluindo, aqui, dicionários, como será demonstrado nesta pesquisa.

Todavia, segundo alguns críticos, como o sociólogo Nick Couldry, professor da *London School of Economics and Political Science*, no Reino Unido, essas ferramentas receberam uma denominação equivocada. “Não é inteligência, nem é artificial, porque

---

<sup>2</sup> No original: “GenAI is a form of Artificial Intelligence which uses vast quantities of information and powerful machine learning models to create original content, including text, audio, code, images, simulations and vídeo”.

depende do trabalho humano para treiná-la. É apenas probabilística, não é criativa”<sup>3</sup>. Com efeito, um dos atributos de seres inteligentes, como seres humanos, por exemplo, é a vida orgânica. Sob esse aspecto, portanto, uma máquina não poderia ser classificada como “inteligente”, visto que não possui vida.

Em todo caso, considerando que esse é o termo mais amplamente difundido, utilizaremos “Inteligência Artificial” para designar esse campo do Processamento de Linguagem Natural (PLN), que possibilita a computadores não apenas entender como manipular a linguagem humana. Assim, tanto o Chat GPT quanto o Google AI Studio operam como Modelos de Linguagem Grande (MLG) e são compostos por algo que se assemelha a uma rede neural. Ambos foram treinados com enormes quantidades de textos e permitem que possamos conversar com eles. De fato, “*Chat*” significa “bate-papo” em inglês, o que indica que essa ferramenta pode manter diálogos com seres humanos. Já GPT é uma sigla para *Generative Pre-Trained Transformer*, o que já sinaliza para o fato de a ferramenta ter sido treinada para gerar conteúdo.

Além disso, gostaríamos de destacar, inclusive, que, como o treinamento dessas IA foi originalmente feito em inglês, elas costumam apresentar um desempenho melhor quando são chamadas a produzir textos nesse idioma. Isso não impede que possamos também interagir com elas em outras línguas, mas apenas indica que o domínio da língua inglesa oferece uma vantagem para quem for elaborar o *prompt*.

Nesse contexto, decidimos testar a habilidade dessas máquinas em produzir pequenos dicionários. Outros estudos, como os conduzidos por Jakubíček e Rundell (2023) e Lew (2023) também já se lançaram a essa tarefa. Ambos demonstraram o potencial de dicionários criados por IA, que, apesar de não terem a qualidade de um restaurante cinco estrelas, como aponta Lew (2024), podem saciar a fome da maneira como um restaurante de comida rápida o faria. Desse modo, ao contrário do que se pode pensar, a IA não parece apontar para o fim da lexicografia, que se dedica à criação de dicionários, mas pode ser usada por lexicógrafos experientes para reduzir o tempo e o custo de criação dessas obras.

Vale recordar, por exemplo, que o *Oxford English Dictionary* (OED) consumiu 40 anos da vida do filólogo James Murray para ficar pronto, conforme aponta Winchester (2009). Se uma obra dessa grandiosidade tivesse sido feita na atualidade, com os recursos de que dispomos hoje, talvez em poucos anos ela já estivesse concluída. Esse fato ilustra a

---

<sup>3</sup> Artigo original disponível em: <https://www.abc.org.br/2024/05/11/nick-couldry-ia-nao-e-inteligencia-nem-e-artificial/>

rapidez com que as ferramentas de IA conseguem trabalhar e estimula várias pesquisas nessa área.

Este estudo, portanto, tem como objetivo geral comparar o tratamento lexicográfico da homonímia em dois dicionários de inglês criados por IA: um criado pelo Chat GPT 3.5 e outro pelo Google AI Studio, verificando se atendem aos critérios de dicionários pedagógicos definidos por autores como Prado Aragonés (2005) e Azorín Fernández (2009). Ademais, Ademais, a comparação entre as informações geradas pelas IA busca identificar semelhanças e originalidades no tratamento desses homófonos.

### Referencial teórico

Os homófonos não homógrafos são palavras que possuem pronúncia idêntica, mas grafia e significados distintos. São, pois, um exemplo de homonímia, fenômeno linguístico em que um nome se equivale a outro, seja em sua forma oral ou escrita. Berruto (1979) considera haver três tipos de homônimos: os homógrafos e homófonos, os homógrafos não homófonos (também chamados de homófonos heterófonos) e os homófonos não homógrafos ou heterógrafos, tema desta pesquisa.

Dessa forma, devido a suas características de grafia e/ ou pronúncia semelhantes, a homonímia causa ambiguidade lexical (Zavaglia, 2003), logo, merecendo atenção de quem estiver aprendendo, sobretudo, uma língua estrangeira. Por esse motivo, é interessante que informações sobre homonímia sejam incluídas em obras lexicográficas, com o intuito de auxiliar seus consulentes a sanar suas dúvidas acerca dessas palavras. Essa questão também justifica, por exemplo, a criação de um dicionário unicamente composto de homófonos não homógrafos, que reúna em seu interior vários exemplos de palavras com essas características, a fim de explicitar essa economia linguística (uso de uma mesma palavra para expressar significados diferentes) ao consulente.

Sob esse viés, as tecnologias atuais prestam um grande benefício porque colocam nas mãos dos usuários o poder de criação de conteúdos que, anteriormente, eram de exclusividade de certas profissões. Por exemplo, compartilhando *prompts* da área de *design* e *marketing*, pessoas comuns podem ser capazes de criar imagens no Copilot (ferramenta do Google para criar imagens com IA) para depois usar em postagens mais atraentes no Instagram. Ou, de posse de *prompts* como os sugeridos neste artigo, podem criar dicionários para fins pessoais. Portanto, uma tarefa que durante séculos ficou restrita

aos compiladores ou lexicógrafos se torna agora acessível a qualquer um com acesso à *internet*.

Ademais, a IA confere protagonismo ao consulente que, para atender suas próprias necessidades poderia criar suas próprias obras com as informações que julgar relevantes. Essa perspectiva é uma grande novidade, pois o consulente é, via de regra, um “ilustre desconhecido”, uma vez que o compilador de uma obra lexicográfica comumente escreve para um público com o qual não tem contato direto, precisando supor quais seriam suas necessidades linguísticas. Inclusive, de acordo com Hartmann (2001), existem quatro atores principais nesse cenário de contato com o texto dicionarístico, sendo eles o consulente, o professor, o pesquisador e o compilador. O detalhe, contudo, é que, como já dito, eles não costumam dialogar entre si.

Gostaríamos de destacar, todavia, que o trabalho de um lexicógrafo é meticuloso e envolve, entre outras questões, conhecimentos sobre *verb patterns*, a frequência de uso das palavras, a ordenação de diferentes acepções dentro do verbete, bem como a origem dessas palavras, entre outras informações. São conhecimentos mais profundos e que requerem estudos específicos na área. Mas, para fins de aquisição/ fixação de vocabulário, o consulente poderia criar obras mais simples, contendo, por exemplo, apenas uma dentre as definições possíveis, um ou dois exemplos e assim por diante. Com base nesses conhecimentos, um consulente poderia solicitar ao Chat GPT (bem como ao Google AI Studio ou a outra ferramenta de IA), que criasse: i) um minidicionário com 100 verbos em inglês, ii) um minidicionário com 50 sentimentos, iii) um minidicionário com 50 palavras que possuam letras mudas e assim por diante.

No tocante ao Google AI Studio, em específico, que possui em sua interface recursos de multimodalidade, permitindo também anexar à solicitação de *prompt* arquivos de áudio, vídeo e imagem, o usuário poderia ir, inclusive, além. Por exemplo: i) poderia anexar uma imagem de uma bicicleta e solicitar que seja criado um dicionário contendo todas as partes dela, ii) poderia anexar uma cópia em PDF de um livro do Harry Potter e solicitar que fosse criado um dicionário de criaturas mágicas, iii) poderia anexar um trecho de um vídeo do filme “As branqueles” e solicitar que fosse criado um dicionário de gírias. Logo, existem infinitas possibilidades de criação de dicionários ou até de glossários usando IA.

Desse modo, nesta pesquisa, nossa atenção foi concentrada na criação de verbetes. Afinal, um dicionário consiste, segundo Krieger (2007) em um catálogo que expõe aos falantes as palavras de que ele dispõe para se comunicar em uma certa língua. Em um dicionário, essas palavras surgem em blocos, ou seja, listas de verbetes (junção da palavra entrada ou lema com as informações subsequentes), dispostos um após o outro. Nossa escolha pelo verbete não é fortuita, pois, conforme aponta Garriga Escribano (2003, p. 105, tradução nossa), “considera-se que o verbete é a unidade mínima autônoma em torno da qual o dicionário está organizado”<sup>4</sup>. Então, se a IA conseguisse criar, de maneira satisfatória, um verbete, seria de se esperar que também conseguisse gerar outros 10, ou 100, ou 1.000. Todavia, destacamos que um dicionário, principalmente se impresso, possui também outras partes além de apenas verbetes, como a *Front Matter* (páginas iniciais) e a *Back Matter* (páginas finais) que, para os fins deste trabalho, não foram consideradas.

Assim, nesta pesquisa, levamos em consideração apenas a macroestrutura e a microestrutura. A organização macroestrutural compreende essa disposição dos verbetes de A a Z. Já a organização microestrutural diz respeito a quais informações são citadas dentro desse verbete, logo após a palavra-entrada. E, ao analisar essa microestrutura, nos propusemos a avaliar, segundo os critérios da lexicografia pedagógica, se esses dicionários “amadores”, criados por consultantes poderiam ser considerados como dicionários pedagógicos ou não.

Mas, quais seriam, então, os critérios mínimos para se considerar um dicionário como pedagógico? Em primeiro lugar, os dicionários chamados de pedagógicos, diferentemente dos dicionários gerais (voltados para falantes nativos de um idioma), têm como público-alvo aqueles que estão aprendendo uma língua estrangeira. Adotam, portanto, uma postura bastante didática, ou seja, procuram tratar as informações com clareza e objetividade, sem rodeios.

Acerca desse assunto, citamos os trabalhos de Prado Aragonés (2005) e Azorín Fernández (2009) que procuram elencar as características mínimas para um dicionário ser considerado como pedagógico. Destacamos que são critérios definidos para dicionários impressos, mas, diante da não existência de obras atuais voltadas unicamente para dicionários virtuais, também os aplicaremos a esses dicionários gerados por IA. Não listaremos todos aqui, porém, gostaríamos de mencionar que estão presentes as definições,

---

<sup>4</sup> No original: “[...] se considera que el artículo lexicográfico es la unidad mínima autónoma en que se organiza el diccionario”.

em caso de dicionários monolíngues (ou os equivalentes, em caso de dicionários bilíngues), a pronúncia e a ortografia dessa palavra, os exemplos de uso, expressões idiomáticas, sinônimos, *collocations* (preposições e adjetivos, por exemplo, que são usados junto com essa palavra), entre outras informações.

Dessa maneira, cabe ao criador do dicionário decidir quais informações ele julga relevantes para incluir no verbete. Nesta pesquisa, inicialmente, o modelo solicitado foi o de construção de um dicionário que se assemelhasse bastante ao modelo atualmente existente para obras bilingualizadas (também chamadas de semibilíngues), ou seja, que apresentassem tanto a definição quanto o equivalente da palavra-entrada (Duran & Xatara, 2005). São, portanto, uma junção “do melhor dos dois mundos”, uma vez que reúnem em um mesmo verbete características de dicionários monolíngues (como a definição) e bilíngues (como a tradução do lema). O motivo de termos eleito esse formato de verbete em nossa primeira tentativa foi sua simplicidade, pois, normalmente, sua estrutura é a seguinte:

**lema** *classe gramatical* **acepção 1** (apresentando a definição do primeiro significado da palavra-entrada): exemplos de uso; **equivalente (s) da palavra-entrada**

Na sequência, dedicamo-nos à elaboração de um *prompt* mais sofisticado, que tinha por objetivo criar um verbete semelhante ao que havíamos construído em nossa pesquisa de Mestrado, conforme Oliveira (2023). A criação de verbetes com base nos *prompts* que usamos se mostrou muito promissora e isso poderá ser conferido mais adiante, na sessão que trata dos resultados. Ficamos, por ora, com a metodologia.

## Metodologia

Nesta pesquisa, procuramos verificar como seria o desempenho de duas ferramentas gratuitas de IA a saber: o Chat GPT 3.5 e o Google AI Studio quando solicitadas a produzir dicionários de inglês. Ademais, também observamos a estrutura dos verbetes e sua organização macroestrutural, de forma a avaliar se esses dicionários produzidos por máquinas poderiam ser considerados dicionários pedagógicos. Para tanto, analisamos os verbetes obtidos e, após feita a coleta dos dados, comparamos nossos resultados. Como nossa pesquisa de Mestrado, concluída em 2023, versava sobre os homófonos não homógrafos, esse também foi o tema que elegemos para construir nossos dicionários. Assim, inicialmente, elaboramos um *prompt* (comando) que seria enviado de

maneira idêntica às duas ferramentas de IA, com o intuito de observar como elas reagiriam ao comando solicitado. Este foi o *prompt* usado com ambas as ferramentas:

**Figura 1 - *Prompt* usado tanto com o ChatGPT 3.5 quanto com o Google AI Studio para criação de um minidicionário contendo 100 exemplos de homófonos não homógrafos**

Hi. Please help me create a dictionary. Make a list of 100 homophones (words that sound the same but have different meanings and different spellings). Then provide to me their definitions, phonemic symbol (how to pronounce them) and give me one example of those words in a sentence. Also translate the headword to me (in Portuguese) and translate the example sentences to Portuguese.

**Fonte: a autora**

Analisando o comando acima, é possível notar que havíamos desenhado um verbete com as seguintes informações: lema, definição, transcrição fonêmica do lema, exemplo de uso, tradução do lema e tradução do exemplo. Destacamos que essa estrutura solicitada, embora não se caracterize como um exemplo de verbete-padrão encontrado em dicionários semibilíngues, serve para ilustrar a margem de liberdade de que o criador do *prompt* dispõe. Assim, em vez de criarmos verbetes segundo o formato tradicional existente no mercado, como “lexicógrafos virtuais” temos à nossa disposição uma ampla gama de possibilidades, podendo ajustar a criação de um verbete que atenda melhor nossas necessidades. Nossas reflexões acerca dos resultados obtidos podem ser conferidas a seguir.

## **Resultados e discussão**

Após solicitar à IA que produzisse os verbetes, procedemos à análise dos resultados. As Figuras 2 (a seguir) e 3 (disposta mais adiante) revelam a organização inicial dos verbetes obtida em ambas as ferramentas:

**Figura 2 - Exemplo de resposta à solicitação feita ao ChatGPT 3.5 para construir um minidicionário, juntamente com os quatro primeiros verbetes obtidos**

Sure, here's a list of 100 homophones with their definitions, phonemic symbols, example sentences in English, and translations into Portuguese:

1. **write** /raɪt/ - to put words on paper
  - I will write a letter to my friend.
  - *escrever* - Eu vou escrever uma carta para meu amigo.
2. **right** /raɪt/ - opposite of left, correct
  - Turn right at the next intersection.
  - *direita* - Vire à direita na próxima interseção.
3. **cell** /sɛl/ - basic unit of life, small room
  - The human body is made up of billions of cells.
  - *célula* - O corpo humano é composto por bilhões de células.
4. **sell** /sɛl/ - to exchange for money



- She wants to sell her old car.
- *vender* - Ela quer vender o carro velho dela.

**Fonte: a autora**

Analisando a Figura 2 e a Figura 3 (mais à frente), pudemos constatar que, em seu primeiro comando, as máquinas já tiveram relativa facilidade em elaborar uma microestrutura bastante próxima às encontradas no mercado em dicionários convencionais. Além disso, o primeiro aspecto que nos saltou aos olhos foi, naturalmente, a velocidade com que elaboraram esses verbetes, ambos em menos de dois minutos. O segundo aspecto relevante foi que não fornecemos a lista de homófonos para a ferramenta. Apenas solicitamos que fosse criado um minidicionário com homófonos não homógrafo (e explicamos que se tratavam de palavras com grafia diferente, mas pronúncia idêntica). E ambos foram bem-sucedidos na tarefa, listando 100 palavras que realmente são classificadas como tal.

Esses verbetes iniciais muito nos surpreenderam porque o ChatGPT 3.5<sup>5</sup>, por exemplo, demonstrou ser capaz de organizar as informações de maneira muito didática. Pode-se notar que a ferramenta grafou o lema em negrito (embora não houvésemos solicitado especificamente por esse detalhe) e usou marcadores (em formato de círculos) para sinalizar o exemplo de uso, bem como o equivalente do lema e a tradução do exemplo de uso fornecido. Também foi interessante notar que, diferentemente do que havia sido solicitado no comando, a ferramenta produziu um verbeito dispondo a transcrição fonêmica imediatamente após o lema (e não após a definição, como nosso *prompt* havia sugerido), que é, por sinal, o modelo mais comum usado em dicionários tanto impressos quanto da *internet*.

Essas constatações parecem indicar que a ferramenta possa ter tido algum treino com dicionários durante sua fase de testes. Também é possível que a ferramenta tenha “aprendido” algo sobre construção de dicionários conosco mesmo, porque esse não foi o primeiro dicionário que solicitamos ao ChatGPT 3.5 fazer. Dando sequência à análise, em termos macroestruturais, notamos que o ChatGPT 3.5 enumerou os verbetes de 1 a 100 e os dispôs verticalmente, segundo o formato tradicional, porém, não colocou os verbetes em ordem alfabética (embora ele possivelmente poderia ter feito assim, se tivéssemos solicitado por isso expressamente). A seguir, é possível notar que os resultados obtidos

---

<sup>5</sup> O dicionário em questão foi construído em uma segunda-feira, no dia 6 de maio de 2024, às 13:31:31.

com o Google AI Studio<sup>6</sup> foram semelhantes aos do Chat GPT 3.5, embora também seja possível notar algumas diferenças:

**Figura 3 - Exemplo de resposta à solicitação feita ao Google AI Studio para construir um minidicionário, juntamente com os quatro primeiros verbetes obtidos**

### 100 Homophones with Definitions, Pronunciation, Example Sentences, and Portuguese Translations

Here is a list of 100 homophones with their definitions, phonetic transcriptions, example sentences, and Portuguese translations.

Note: The phonetic transcriptions use the International Phonetic Alphabet (IPA).

Word 1	Definition 1	IPA 1	Example Sentence 1	Portuguese Translation 1	Word 2	Definition 2	IPA 2	Example Sentence 2	Portuguese Translation 2
Allowed	Permitted	/ə'laud/	The dog is allowed to sleep on the couch.	O cão tem permissão para dormir no sofá.	Aloud	In a voice that can be heard	/ə'laud/	He read the poem aloud to the class.	Ele leu o poema em voz alta para a turma.
Bare	Uncovered, naked	/beər/	The tree branches were bare in winter.	Os galhos da árvore estavam despidos no inverno.	Bear	A large furry mammal	/beər/	A bear emerged from the forest.	Um urso saiu da floresta.

**Fonte: a autora**

Observando a Figura 3, pode-se perceber que esta também atendeu satisfatoriamente à solicitação feita, embora sua organização não tenha sido tão didática como aquela do ChatGPT 3.5. Em primeiro lugar, é possível notar que a disposição dos verbetes na macroestrutura se deu horizontalmente, algo incomum de se encontrar em dicionários, tanto impressos quanto da *internet*. Ele também não enumerou os verbetes de 1 a 100 (embora possivelmente tivesse feito isso, se tivéssemos solicitado por essa informação de maneira expressa). Mas, em seu favor, o Google AI Studio dispôs os verbetes em ordem alfabética mesmo que isso tenha sido omitido do comando. Esse resultado parece indicar que a ferramenta também possa ter tido algum contato com dicionários anteriormente, durante sua fase de treinos.

Logo no início da resposta ao comando, a ferramenta Google AI Studio também inseriu uma informação interessante: a de que os símbolos usados para indicar a pronúncia haviam sido retirados do *International Phonetic Alphabet (IPA)*, embora tenha cometido um pequeno equívoco ao chamar essas transcrições de “fonéticas” (normalmente representadas entre colchetes) e não de “fonêmicas” (representadas entre barras transversais), conforme a terminologia correta.

<sup>6</sup> O dicionário em questão foi construído em uma segunda-feira, no dia 6 de maio de 2024, às 13:35:32.

Quanto à microestrutura, pode-se observar que o Google AI Studio grafou de negrito a palavra-entrada, estando essa mesma palavra também em negrito no exemplo fornecido, bem como na tradução desse exemplo. Ao proceder dessa forma, a ferramenta economizou, automaticamente, espaço, pois não necessitou traduzir a palavra-entrada separadamente. Ademais, o Google AI Studio se mostrou mais “obediente” que o ChatGPT 3.5, uma vez que dispôs as informações no verbete exatamente como a solicitação havia sido feita, ou seja, com a definição logo depois do lema e antes da transcrição fonêmica.

Após feitas essas análises iniciais, dedicamo-nos a tentar aprimorar os verbetes em questão. Para tanto, criamos uma estrutura mais sofisticada, adaptada do nosso verbete usado na dissertação (Oliveira, 2023) e solicitamos às ferramentas que reescrevessem um dos verbetes criados. Assim, o novo *prompt* pode ser conferido na Figura 4:

**Figura 4 - Prompt usado tanto com o ChatGPT 3.5 quanto com o Google AI Studio para criação de um verbete sofisticado**

Hi. Please help me create a very sophisticated entry for the word “aloud”. This entry will be part of a list of 100 homophones (words that sound the same but have different meanings and different spellings). So, you will actually have to provide to me two entries: entry 1 for the word “aloud” and entry 2 for another word that sounds the same but has a different meaning and different spelling. The entries have to appear vertically on the screen, that is, entry 1 must be followed by entry 2, which is going to be below entry 1 and not beside it. Each entry must have the following information (exactly in this order and organized in a paragraph structure): the headword (in bold and using a dot to separate syllables), its phonemic symbol (how to pronounce it), its part of speech, its definition and one example sentence using that headword (use a small circle before the example). Also indicate two example sentences highlighting the collocations (in bold), the idioms and the phrasal verbs (if there are any) you can usually find with that headword. Finally, translate the headword to me (in Portuguese) and translate the example sentences to Portuguese. Don’t use labels.

**Fonte: a autora**

A escolha pela palavra-entrada “*aloud*” foi intencional porque trata-se de uma palavra com duas sílabas em inglês e gostaríamos de ver a capacidade de a ferramenta indicar ou não isso no verbete. A seguir, é possível observar como foi elaborado o verbete pelo Chat GPT 3.5:

**Figura 5 - Exemplos de verbetes sofisticados fornecidos pelo Chat GPT 3.5**

Certainly! Here are the entries for "aloud" and "allowed" presented with the requested structure:

**Entry 1:** a·loud

/ə'laʊd/, adverb, in a voice that can be clearly heard. • She read the letter aloud to the class. **Read aloud:** She often **reads aloud** to her children at bedtime. **Speak aloud:** He couldn't help but **speak aloud** his thoughts during the meeting.

Translation (Portuguese): em voz alta. Ela leu a carta em voz alta para a turma. Ela frequentemente **lê em voz alta** para

seus filhos na hora de dormir. Ele não pôde deixar de **falar em voz alta** seus pensamentos durante a reunião.

**Entry 2:** al·lowed

/ə'laʊd/, verb (past tense of "allow"), gave permission for something to happen. • She was allowed to go to the party. **Not allowed:** Pets are **not allowed** in the hotel. **Be allowed to:** Children should **be allowed to** play outside.

Translation (Portuguese): permitido. Ela foi permitida a ir à festa. Animais de estimação **não são permitidos** no hotel. As crianças devem **ter permissão para** brincar fora.

**Fonte: a autora**

Ficamos bastante satisfeitos com a compreensão que a máquina demonstrou ter do que queríamos. A organização dos verbetes ficou melhor em vários aspectos, sobretudo porque, agora, seguiu-se uma estrutura de parágrafo e novas informações foram incluídas, como a indicação da separação de sílaba, por exemplo. Informações gramaticais, como o fato de “*allowed*” ser o passado de “*allow*” também foram mencionadas, bem como *collocations*, a exemplo de “*speak*” ser um verbo comumente utilizado junto com “*aloud*”. Em seu desfavor, apontamos, apenas que a tradução apresentada para “*She was allowed to go to the party*” (Ela foi permitida a ir à festa) não foi tão precisa. Uma tradução melhor poderia ter sido: Foi permitido que ela fosse à festa. A seguir pode ser conferida a resposta fornecida pelo Google AI Studio para o mesmo *prompt*:

**Figura 6 - Exemplos de verbetes sofisticados fornecidos pelo Google AI Studio**

Entry 1: Aloud

**a·loud** /ə'laʊd/ *adverb* Spoken out loud, rather than silently or in one's mind.

The teacher asked the students to read the passage aloud.

**Collocations:** read **aloud**, speak **aloud**, think **aloud**

**Idiom:** **Speak your mind**

**Portuguese Translation:** Em voz alta

**Portuguese Translation of Example Sentence:** O professor pediu aos alunos para lerem a passagem em voz alta.

Entry 2: Allowed

**al·lowed** /ə'laʊd/ *adjective* Permitted, authorized, or given permission.

Smoking is not allowed in this building.

**Collocations:** **strictly allowed**, **not allowed**, **fully allowed**

**Idiom: Allowed to do something**

**Portuguese Translation:** Permitido

**Portuguese Translation of Example Sentence:** Fumar não é permitido neste prédio.

**Fonte: a autora**

Avaliamos também como satisfatória a habilidade de o Google AI Studio gerar verbetes, porém, pudemos perceber que certos aspectos solicitados no *prompt* não foram postos em prática, como a indicação do exemplo ser precedida por um círculo (ou bolinha em formato de marcador), ou a solicitação de que não fossem usadas “etiquetas” (dizendo que uma *collocation* era uma *collocation*, por exemplo). Talvez sua resistência em cumprir com as ordens se deva ao fato de que ele tenha sido treinado para inserir tais informações, porque muitos dicionários, de fato, as trazem assim. Em seu favor, apontamos que as traduções com o Google AI Studio foram melhores.

Se compararmos esses verbetes com verbetes do *Oxford Advanced Learner’s Dictionary* (OALD), por exemplo, poderemos perceber que embora as informações presentes no último possuam mais rigor e complexidade, sendo, portanto, mais confiáveis, as informações do Chat GPT 3.5 e do Google AI Studio podem ser úteis para tarefas menos complexas. Voltando à analogia de Lew (2023), são como exemplos de comida rápida que preenchem o estômago de quem tem fome, mas não está preocupado se sua comida foi preparada por um chefe de cozinha conceituado ou não.

Então, se as ferramentas (Chat GPT 3.5 e Google AI Studio, respectivamente) foram capazes de elaborar verbetes sofisticados, elas também seriam capazes de adaptar todos aqueles 100 verbetes criados anteriormente para o mesmo formato. Nossa avaliação final é que esses minidicionários criados por IA podem ser classificados como pedagógicos, embora estejamos falando de obras muito pequenas, de até 100 verbetes. Um dicionário monolíngue elaborado por lexicógrafos, como o OALD, por exemplo, tem centenas de milhares de entradas.

O que fica evidente, contudo, é que as máquinas já estão conseguindo reproduzir exemplos bastante próximos do que poderia ser considerado um “verbe ideal” e, com um pouco mais de treino, poderão aprender e aprimorar rapidamente sua “*expertise*”.

Sobretudo, se lembrarmos que elas são operadas por seres humanos, perceberemos que, com *prompts* melhores, elas poderão produzir melhores resultados e serem capazes de grandes feitos.

### Considerações finais

Neste estudo, dedicamo-nos a comparar a qualidade da produção de verbetes por IA gratuitas, a exemplo do Chat GPT 3.5 e do Google AI Studio. Para tanto, solicitamos que ambas gerassem verbetes contendo exemplos de homófonos não homógrafos. Nossa experiência revelou que: i) as ferramentas de IA testadas conseguem inserir informações satisfatórias dentro dos verbetes; ii) tanto o Chat GPT quanto o Google AI Studio conseguem aprimorar os resultados caso sejam fornecidos *prompts* mais específicos e iii) os minidicionários gerados, apesar de serem apenas amostras do que dicionários maiores, como o OALD conseguem produzir, podem ser classificados como exemplos de dicionários pedagógicos.

Por fim, acrescentamos que, a nosso ver, o mais importante não é a avaliação atual quanto aos resultados gerados, mas o potencial de uso dessas ferramentas no futuro. Pesquisas nessa área ainda são muito recentes e a IA está passando por um processo de desenvolvimento muito acelerado, mas já podemos vislumbrar que, em pouco tempo, possivelmente a qualidade dos verbetes poderá se equiparar de igual para igual aos tradicionais dicionários pedagógicos.

Por meio de nossa análise, pudemos perceber que, diferentemente do passado, quando compiladores sem contato direto com consulentes criavam obras que buscavam atender as necessidades desses últimos segundo critérios previamente estabelecidos, agora os próprios consulentes, de posse de *prompts* como os demonstrados nessa pesquisa podem criar dicionários com as informações que julgarem mais pertinentes. Assim, essa pesquisa demonstra a capacidade de a IA gerar dicionários personalizados, abrindo portas para uma nova era na lexicografia digital e impulsionando pesquisas nessa área.

### Referências

AZORÍN FERNÁNDEZ, D.; MARTÍNEZ EGIDO, José Joaquín. Hacia una evaluación de las prestaciones didácticas del diccionario desde la perspectiva de los usuarios. In: GARCÍA PLATERO, Juan Manuel; CASTILLO CARBALLO, Maria Auxiliadora (coords.). **Investigaciones lexicográficas para la enseñanza de lenguas**. Málaga: Servicio de Publicaciones de la Universidad de Málaga, 2009.

BERRUTO, G. **La Sémantica**. México: Editorial Nueva Imagen S.A. 1979.

CAMBRIDGE. **English language education in the era of generative AI: our perspective**. 2023.

COWIE, A. **A. S. Hornby**: a Centenary Tribute. EURALEX '98 PROCEEDINGS. PLENARY LECTURES. University of Leeds, 1998.

CRUZ, E. O. **Nick Couldry**: “IA não é inteligência, nem é artificial”. REUNIÃO MAGNA ABC | 11 de maio de 2024 – Disponível em: <https://www.abc.org.br/2024/05/11/nick-couldry-ia-nao-e-inteligencia-nem-e-artificial/> Acesso em 27 de jun. 2024.

DURAN, M.S.; XATARA, C.M. Dicionários semibilíngües: uma inovação? **Rev. Est. Ling.**, Belo Horizonte, v. 13, n. 1, p. 45-57, jan./jun. 2005.

GARRIGA ESCRIBANO, C. La microestructura del diccionario: las informaciones lexicográficas. In: MEDINA GUERRA, A. M. (Coord.). **Lexicografía española**. Barcelona: Ariel Lingüística, 2003, p. 103-126.

HARTMANN, R. R. K. **Teaching and Researching Lexicography**. – London: Pearson Education Limited, 2001.

JAKUBÍČEK, M.; RUNDELL, M. The End of Lexicography? Can Chat GPT Outperform Current Tools for Post-Editing Lexicography? In: MEDVED, M., MECHURA, M., KOSEM, I.; KALLAS, J.; TIBERIUS, C.; JAKUBÍČEK, M. (eds). **Electronic lexicography in the 21st century (eLex 2023)**: Invisible Lexicography. Proceedings of the eLex 2023 conference. Brno: Lexical Computing CZ s.r.o., 518-533.

KRIEGER, M. da G. O dicionário de língua como potencial instrumento didático. In: ISQUERDO, A. N.; ALVES, I. M. (Orgs.). **As ciências do léxico**: lexicologia, lexicografia e terminologia, v. III. Campo Grande: Editora da UFMS, 2007, p. 295-309.

LEW, R. ChatGPT as a COBUILD Lexicographer. **Humanities and Social Sciences Communications**. 10.1. Palgrave: 1-10. Doi: 10.1057/s41599-023-02119-6

OLIVEIRA, R. **OS HOMÓFONOS NÃO HOMÓGRAFOS EM DICIONÁRIOS PEDAGÓGICOS DE LÍNGUA INGLESA**: possíveis sugestões lexicográficas. Dissertação (Mestrado em Letras). Universidade Federal do Mato Grosso do Sul, Faculdade de Letras: Três Lagoas, 2023, 352 p.

PRADO ARAGONÉS, J. El uso del diccionario para la enseñanza de la lengua: consideraciones metodológicas. En: Káñina, **Rev. Artes y Letras**, Univ. Costa Rica. Vol. XXIX (Especial), 2005.

REES, G. P.; LEW, R. The Effectiveness of OpenAI GPT- Generated Definitions Versus Definitions from an English Learners' Dictionary in a Lexically Orientated Reading Task. **International Journal of Lexicography**, 37, 50-74. 2024. doi: 10.1093/ijl/ecad030

WINCHESTER, S. **O professor e o louco**: uma história de assassinato e loucura durante a elaboração do dicionário Oxford. Tradução Flávia Villas-Boas. – São Paulo: Companhia das Letras, 2009.

ZAVAGLIA, C. Ambiguidade gerada pela homonímia: revisitação teórica, linhas limítrofes com a polissemia e proposta de critérios distintivos. **DELTA**: Documentação de Estudos em Linguística Teórica e Aplicada, São Paulo, v. 1, n. 19, 2003, p. 337-266.