

# Configuração e avaliação de um GPT personalizado para anotação semântica abstrata do português europeu segundo o modelo LMR

## Configuration and Evaluation of a Custom GPT for Abstract Semantic Annotation of European Portuguese According to the LMR Framework

Sónia Reis<sup>1,2,3</sup>, Jorge Baptista<sup>1,2</sup>

<sup>1</sup> Universidade do Algarve

<sup>2</sup> INESC-ID Lisboa – Human Language Technology Lab

<sup>3</sup> Agrupamento de Escolas de Albufeira

### Resumo

Este artigo apresenta a configuração e a avaliação de um GPT personalizado para a anotação semântica abstrata de textos em português europeu segundo o modelo *Lexicalized Meaning Representation* (LMR). O LMR constitui uma adaptação do *Abstract Meaning Representation* (AMR), concebida para acomodar propriedades da gramática do português e preservar uma ligação mais direta entre o texto e a respetiva representação semântica. O GPT foi configurado através de uma instrução específica e de uma base de conhecimento fechada, composta pelas diretivas do LMR e por um conjunto de frases previamente anotadas. O trabalho introduz ainda relações e operadores destinados à representação de determinadas configurações de discurso reportado e de citação. Para a avaliação, foi constituído um *corpus* de 42 notícias de acesso aberto sobre inteligência artificial, a partir do qual foram selecionadas cinco frases, anotadas independentemente pelos dois autores e posteriormente submetidas a adjudicação por consenso para formar o conjunto de referência. As experiências comparam o desempenho dos modelos GPT-4o e GPT-5-Thinking em quatro condições: sem exemplos no *prompt* (*zero-shot* ao nível do *prompt*), *few-shot out-of-domain*, *many-shot out-of-domain* e *few-shot in-domain*. A avaliação através da métrica *Smatch*, complementada por uma análise qualitativa dos desvios, mostra que, no conjunto analisado, o GPT-5-Thinking produz anotações mais próximas da referência, com um *F-score* médio de 0,62, comparativamente a 0,51 no GPT-4o. O efeito dos exemplos não se revelou uniforme nem cumulativo, variando em função do modelo e da condição experimental. Os resultados confirmam o potencial destes sistemas como ferramentas de apoio à anotação semântica, embora continue a ser necessária supervisão humana.

**Palavras-chave:** Anotação semântica abstrata, modelos de linguagem de grande escala, GPT personalizado, *Lexicalized Meaning Representation* (LMR), discurso reportado.

### Abstract

This article presents the configuration and evaluation of a custom GPT for the abstract semantic annotation of European Portuguese texts according to the *Lexicalized Meaning Representation* (LMR) framework. LMR is an adaptation of *Abstract Meaning Representation* (AMR), designed to accommodate properties of Portuguese grammar while preserving a closer connection between the source text and its semantic representation. The GPT was configured through a dedicated instruction set and a closed knowledge base comprising the LMR guidelines and a collection of previously annotated sentences. The study also introduces relations and operators for representing selected configurations of reported speech and quotation. For evaluation purposes, a *corpus* of 42 open-access news articles on artificial intelligence was compiled, from which five sentences were selected. These sentences were independently annotated by the two authors and subsequently adjudicated by consensus

to form the reference set. The experiments compare the performance of GPT-4o and GPT-5-Thinking under four conditions: without examples explicitly included in the *prompt*, *few-shot out-of-domain*, *many-shot out-of-domain*, and *few-shot in-domain*. Evaluation using the Smatch metric, complemented by a qualitative analysis of the observed divergences, shows that GPT-5-Thinking produces annotations that are, on average, closer to the reference representations, achieving a mean *F-score* of 0.62, compared with 0.51 for GPT-4o. The effect of the examples was neither uniform nor cumulative, varying according to the model and experimental condition. The results confirm the potential of these systems as tools for supporting semantic annotation, although human supervision remains necessary.

**Keywords:** Abstract semantic annotation, large language models, custom GPT, Lexicalized Meaning Representation (LMR), reported speech.

## 1. Introdução

A anotação semântica abstrata visa representar, de forma estruturada, o significado dos enunciados, com abstração de parte da sua variação formal. Estas representações explicitam predicados, argumentos, modificadores e outras relações semânticas relevantes, podendo apoiar diversas tarefas de Processamento de Linguagem Natural (PLN), como a tradução automática, a extração de informação, a sumarização e os sistemas de pergunta-resposta (Brown et al., 2020; Devlin et al., 2019; Fahim et al., 2024).

Entre os modelos de representação semântica baseados em grafos, destaca-se o *Abstract Meaning Representation* (AMR) (Banarescu et al., 2013, 2019), que representa o significado das frases através de conceitos e relações semânticas. A aplicação deste modelo a outras línguas tem, contudo, evidenciado limitações relacionadas com propriedades gramaticais que não são diretamente contempladas pelas diretivas originalmente concebidas para o inglês. Neste contexto, foi desenvolvido o *Lexicalized Meaning Representation* (LMR) (Baptista et al., 2024a, 2024b), um formalismo de anotação que preserva uma ligação mais direta com o texto original e procura representar propriedades específicas da gramática do português europeu.

Paralelamente, os recentes avanços nos modelos de linguagem baseados em arquiteturas *Transformer* (Vaswani et al., 2017), em particular nos modelos de linguagem de grande escala da família GPT (OpenAI, 2022), têm contribuído para a automatização de tarefas de análise e anotação linguísticas (Brown et al., 2020; He et al., 2024; ; Kostina et al., 2025; Radford et al., 2019; Ren, 2024; Savelka & Ashley, 2023; Tao et al., 2024). Estes modelos podem ser instruídos a produzir representações estruturadas segundo esquemas especializados, sem que seja necessariamente realizado um processo de treino ou de *fine-tuning*. A sua capacidade para gerar grafos formalmente válidos e aderir de forma consistente às diretivas de um modelo de representação semântica permanece, contudo, uma questão que exige avaliação empírica.

Um segundo eixo deste artigo prende-se com a representação do *discurso reportado* e da *citação*. Estas estruturas articulam um conteúdo proposicional, uma fonte enunciativa e o evento ou marcador através do qual esse conteúdo é atribuído. Embora sejam frequentemente realizadas por meio de *verba dicendi* (ou verbos de *dizer*), podem também envolver nomes predicativos, expressões preposicionais e outros marcadores de relato. O AMR permite representar predicados de comunicação e os respetivos argumentos, mas não distingue explicitamente algumas configurações relevantes de discurso reportado e de reprodução literal. Para responder a esta limitação, propõe-se, no quadro do LMR, um conjunto de relações e operadores específicos, nomeadamente `:direct-speech` (e as suas variantes: `:direct-speech-cont`, `:direct-speech-before`) e `:ipsisverbis`, apresentados na Secção 3.

Neste contexto, o artigo apresenta a configuração e a avaliação de um GPT personalizado para a anotação semântica abstrata de textos em português europeu segundo o modelo LMR. O trabalho tem cinco objetivos principais: (i) configurar e disponibilizar um GPT personalizado capaz de produzir anotações LMR a partir de uma instrução específica e de uma base de conhecimento fechada; (ii) constituir um *corpus* de notícias sobre inteligência artificial e, a partir dele, um conjunto manual de referência; (iii) comparar o desempenho dos

modelos GPT-4o e GPT-5-Thinking, recorrendo às configurações *zero-shot* (sem exemplos no *prompt*), *few-shot out-of-domain*, *many-shot out-of-domain* e *few-shot in-domain*; (iv) introduzir mecanismos de representação, designadamente relações e operadores, para determinadas configurações de discurso reportado; e (v) analisar, através de um estudo de caso, a capacidade dos modelos para aplicar alguns desses mecanismos.

A avaliação incide sobre cinco frases selecionadas de um *corpus* de 42 notícias e combina a métrica Smatch (Cai & Knight, 2013) com uma análise qualitativa das representações e dos desvios observados. Dado o caráter exploratório e a dimensão do conjunto de avaliação, os resultados são interpretados como indicadores preliminares do potencial e das limitações dos modelos analisados, e não constituindo, portanto, uma avaliação definitiva da anotação automática segundo o LMR.

## 2. Trabalhos relacionados

Nesta secção, o presente trabalho é enquadrado em quatro domínios: a representação semântica abstrata baseada em grafos, o modelo LMR, a utilização de modelos de linguagem na anotação semântica automática e a anotação do discurso reportado.

### 2.1. Representação semântica abstrata baseada em grafos

O *Abstract Meaning Representation* (AMR) é um dos formalismos de representação semântica baseada em grafos mais difundidos no âmbito do Processamento de Linguagem Natural (Banarescu et al., 2013, 2019). Cada representação AMR organiza o significado de uma frase numa estrutura composta por conceitos e relações semânticas, abstraindo parcialmente da sua realização lexical e sintática. A disponibilização de *corpora* anotados e de analisadores automáticos tem contribuído para a aplicação do formalismo em tarefas como geração de texto, extração de informação e sistemas de pergunta-resposta.

Embora inicialmente desenvolvido para o inglês, o AMR tem sido aplicado e adaptado a outras línguas. Para o português brasileiro, destaca-se o *corpus* AMR-BR (Anchiêta & Pardo, 2018), construído a partir da tradução brasileira de *O Pequeno Príncipe* e posteriormente utilizado no desenvolvimento de analisadores automáticos de AMR (Anchiêta, 2020). A comparação entre representações AMR automáticas e representações de referência é geralmente realizada através da métrica Smatch (Cai & Knight, 2013), que calcula a correspondência entre os tripos dos grafos após o alinhamento das respetivas variáveis. Para responder a algumas limitações desta métrica, Anchiêta et al. (2019) propuseram a métrica SEMA como alternativa à Smatch.

A aplicação direta das diretivas AMR a outras línguas pode, contudo, não permitir representar adequadamente fenómenos gramaticais específicos. No caso do português, entre os aspetos problemáticos encontram-se os sujeitos nulos, as construções com verbos auxiliares, incluindo verbos-suporte (Baptista et al., 2024b), determinados tipos de advérbios com escopo frásico (Baptista et al., 2025) e a distinção entre orações relativas restritivas e explicativas (Baptista & Reis, 2026). Essa é uma das justificações para o desenvolvimento do modelo LMR.

### 2.2. O modelo LMR

O modelo da *Lexicalized Meaning Representation* (LMR) (Baptista, 2025; Baptista et al., 2024a, 2024b) foi proposto como um formalismo para a anotação semântica de *corpora*, concebido para representar propriedades da gramática do português europeu que não são sistematicamente contempladas no AMR. Mantendo a arquitetura geral de representação em grafos do AMR, o LMR preserva uma ligação mais direta entre o texto e a representação semântica, e evita o recurso a conceitos abstratos sem realização lexical direta no texto.

O modelo representa explicitamente as unidades lexicais relevantes e introduz um nó raiz explícito que funciona como ponto de ancoragem da estrutura da frase. Representa igualmente elementos gramaticais

frequentemente omitidos em AMR, como preposições predicativas, advérbios, conjunções, verbos auxiliares, verbos copulativos e verbos-suporte. Estas opções reforçam a correspondência entre as unidades realizadas no texto e os elementos representados no grafo.

O LMR apoia-se em recursos lexicais e sintático-semânticos desenvolvidos no quadro do Léxico-Gramática do português, incluindo o *Dicionário Gramatical de Verbos do Português* (Baptista & Mamede, 2020a)<sup>1</sup>, bases de dados de nomes predicativos (Baptista & Mamede, 2020b)<sup>2</sup> e um léxico de construções verbais fixas e idiomáticas (Antunes et al., 2025)<sup>3</sup>. Estes recursos fornecem inventários de entradas lexicais, construções predicativas e estruturas argumentais utilizadas na interpretação dos predicados.

A aplicação do LMR a fenómenos linguísticos específicos foi explorada por Baptista et al. (2025) e Baptista e Reis (2026), através da comparação das representações de advérbios e de orações relativas nas versões portuguesa e brasileira de *O Príncipezinho / O Pequeno Príncipe*, respetivamente, segundo os modelos LMR e AMR. Estes estudos mostraram limitações do AMR na representação de certas construções adverbiais e apresentaram soluções desenvolvidas no quadro do LMR.

### 2.3. Modelos de linguagem e anotação semântica automática

O desenvolvimento de analisadores automáticos de representações semânticas é anterior à utilização recente dos modelos de linguagem de grande escala (LLM). Entre os sistemas desenvolvidos encontram-se analisadores neuronais especificamente treinados para converter, a partir de *corpora* anotados, texto em grafos AMR, como os sistemas apresentados por Damonte et al. (2017) e Bevilacqua et al. (2021). Estes analisadores são treinados com *corpora* anotados e são otimizados especificamente para a tarefa de análise semântica e a conversão de texto corrido em representações AMR.

Mais recentemente, os modelos de linguagem de grande escala têm sido utilizados como anotadores mediante instruções em linguagem natural e a apresentação de exemplos, sem que seja necessariamente realizado treino adicional ou *fine-tuning* para cada esquema de anotação. Savelka e Ashley (2023) exploraram a anotação semântica de textos jurídicos em condições *zero-shot*, enquanto He et al. (2024) analisaram estratégias destinadas a melhorar o desempenho dos LLM como anotadores. Estes estudos mostram o potencial dos modelos generativos, mas identificam também problemas de consistência, adesão às diretivas e produção de categorias não previstas pelo esquema de anotação.

O presente trabalho insere-se nesta segunda linha de investigação. Em vez de treinar um analisador específico, configura-se um GPT personalizado com instruções e documentos de referência, avaliando-se a sua capacidade para produzir grafos LMR formalmente válidos e conformes às diretivas de anotação. Analisa-se, em particular, o efeito do número e do domínio dos exemplos explicitamente incluídos no *prompt* e comparam-se dois modelos da família GPT.

### 2.4. Anotação do discurso reportado e dos eventos de relato

No âmbito do Processamento de Linguagem Natural, o discurso reportado e a citação são frequentemente tratados no quadro das relações de *atribuição*, isto é, como relações que associam um conteúdo à fonte a que a produção do conteúdo é atribuída. Distinguem-se três componentes fundamentais: o *conteúdo reportado*, a *fonte* a que esse conteúdo é atribuído e o *marcador de atribuição* (em inglês, *attribution cue*), que estabelece a ligação entre ambos (Pareti, 2009, 2015). O conteúdo pode corresponder a uma oração, a uma sequência de orações ou a um segmento descontínuo; a fonte pode ser uma pessoa, uma instituição ou uma entidade inanimada interpretada metonimicamente; e o marcador de atribuição não tem necessariamente de corresponder a um verbo declarativo.

<sup>1</sup> <https://string.hlt.inesc-id.pt/demo/viperexplorer.pl>

<sup>2</sup> <https://string.hlt.inesc-id.pt/demo/sniper-explorer.pl>

<sup>3</sup> <https://string.hlt.inesc-id.pt/demo/vidiomexplorer.pl>

A anotação destas estruturas tem sido integrada em enquadramentos com objetivos distintos. No TimeML (Pustejovsky et al., 2003), que inspirou o desenvolvimento da norma ISO-TimeML (International Organization for Standardization [ISO], 2012), é introduzida a classe de eventos *reporting*, destinada à anotação de eventos de comunicação ou relato. Na Rhetorical Structure Theory (Mann & Thompson, 1988), a atribuição pode ser representada como uma relação discursiva. Na Segmented Discourse Representation Theory (Asher & Lascarides, 2003), o discurso reportado é modelado através de relações discursivas. Já no Penn Discourse TreeBank (Prasad et al., 2008), a atribuição é anotada como informação associada às relações discursivas e aos respetivos argumentos, identificando, entre outros aspetos, a fonte do conteúdo atribuído.

Para o português, foram desenvolvidos sistemas e recursos sobre a identificação automática de citações, fontes e elementos introdutórios de discurso, bem como recursos dedicados à anotação do discurso reportado (Fernandes et al., 2011; Freitas et al., 2016; Pardo et al., 2004; Sarmiento & Nunes, 2009).

Mais recentemente, Silvano et al. (2023), no âmbito do sistema de anotação multicamada Text2Story, propuseram distinguir os eventos que constituem a história relatada dos eventos de relato. Nesta abordagem, os eventos de relato são anotados numa camada própria, sendo as respetivas fontes representadas na camada dos participantes e associadas aos eventos através de papéis semânticos, enquanto as relações referenciais permitem identificar diferentes menções da mesma fonte ao longo do texto. O sistema integra ainda uma camada destinada à anotação das relações temporais entre os eventos.

Estes estudos evidenciam que os eventos de relato não são assinalados exclusivamente por predicados verbais de enunciação (ou *verba dicendi*). Além de predicados verbais como *dizer*, *informar*, *declarar* ou *acrescentar*, podem funcionar como marcadores de relato nomes predicativos como *declaração* e *publicação*, bem como expressões discursivas e preposicionais como *segundo* e *de acordo com*. A estrutura do relato pode ainda envolver uma fonte principal retomada ao longo do texto, fontes secundárias e eventos de relato encaixados no interior de conteúdos anteriormente atribuídos.

Os trabalhos anteriores incidem sobretudo na identificação de atribuições, fontes e marcadores ou na anotação multicamada das relações discursivas. A proposta apresentada na secção seguinte não pretende substituir esses enquadramentos nem representar exhaustivamente as cadeias de fontes e os diferentes níveis narrativos. Procura antes incorporar determinadas configurações de discurso reportado e de citação num grafo semântico lexicalizado e avaliar a capacidade de um GPT personalizado para aplicar alguns dos marcadores propostos no modelo LMR.

### 3. Proposta de representação de configurações do discurso reportado no modelo LMR

Partindo da distinção entre *conteúdo reportado*, *fonte* e *marcador de enunciação* apresentada na Secção 2.4, a proposta desenvolvida neste trabalho incide sobretudo nas configurações em que um conteúdo é apresentado como discurso direto ou como reprodução literal de um segmento textual. Não se pretende representar exhaustivamente todas as formas de atribuição, as cadeias de fontes ou os diferentes níveis de encaixamento discursivo. Procura-se antes preservar, no interior do grafo LMR, informação relativa ao estatuto enunciativo dos segmentos reportados ou reproduzidos.

Os predicados de enunciação mantêm a sua estrutura argumental regular: o locutor, o conteúdo reportado e, quando lexicalmente previsto, o interlocutor são representados através das relações argumentais definidas para o predicado. Os marcadores introduzidos nesta secção (:direct-speech, :direct-speech-cont, :direct-speech-before e :ipsisverbis) acrescentam informação sobre a forma como o conteúdo reportado é apresentado no texto, sem substituir essas relações argumentais. Neste artigo, estes mecanismos são codificados no grafo como marcadores binários, aos quais é atribuído o valor positivo (+) quando a propriedade correspondente se verifica.

Nas configurações de discurso indireto sem reprodução literal assinalada, o conteúdo proposicional é representado como argumento do predicado de enunciação, sem receber uma marca específica de citação. No discurso direto, acrescenta-se o marcador binário :direct-speech + ao nó que encabeça o segmento citado; por

exemplo, numa coordenação, o marcador aplica-se ao nó que encabeça toda a estrutura coordenada. Este marcador permite distinguir um enunciado apresentado como discurso direto de uma oração completiva que integra sintaticamente o enunciado do relator.

As aspas, os travessões e outros sinais gráficos de delimitação da citação não são representados como nós autónomos no grafo. A informação veiculada por esses sinais é codificada através do marcador associado ao segmento citado, preservando-se assim o seu estatuto discursivo sem introduzir elementos puramente tipográficos na representação semântica.

Introduzem-se ainda duas variantes de **:direct-speech**. O marcador **:direct-speech-cont** é utilizado quando o discurso direto prossegue numa ou mais frases subsequentes, atribuídas ao mesmo locutor, embora o predicado introdutor se possa encontrar numa frase anterior. O marcador **:direct-speech-before** assinala as configurações em que o segmento citado precede o predicado de enunciação que o enquadra. Estas variantes permitem preservar a relação entre a citação e o respetivo evento de enunciação quando ambos não ocorrem na ordem ou nos limites frásicos mais comuns.

O marcador binário **:ipsisverbis** assinala palavras, expressões ou fragmentos apresentados como reprodução literal, mas integrados na estrutura sintática do enunciado do relator. É o caso de *enganosa* na frase *No entanto, a Google alegou que a informação é “enganosa” [...]* (ID-032-04); ou dos fragmentos *maior plano de investimento até à data* e *garantirá* na frase *Este é o “maior plano de investimento até à data” na Alemanha [...]* e “garantirá” *aproximadamente 9.000 empregos por ano até 2029, acrescentou.* (ID-040-04).

A distinção entre **:direct-speech** e **:ipsisverbis** não depende, portanto, de o material ser ou não reproduzido literalmente. Em ambos os casos pode existir reprodução *ipsis verbis*. A diferença reside no estatuto sintático e discursivo do segmento: **:direct-speech** assinala um enunciado citado com autonomia relativamente à estrutura introdutora, enquanto **:ipsisverbis** marca uma unidade ou um fragmento literal integrado sintaticamente no enunciado do relator.

A descrição completa dos marcadores **:direct-speech**, **:direct-speech-cont**, **:direct-speech-before** e **:ipsisverbis**, bem como das respetivas condições de utilização, encontra-se nas diretivas de anotação do LMR (Baptista, 2025).

Além dos mecanismos acima descritos, foram igualmente desenvolvidos, ou encontram-se ainda em estudo no quadro do LMR, outros operadores associados a usos discursivos e metalinguísticos frequentemente assinalados por aspas, itálico ou negrito. O operador **:metalinguistic** marca expressões usadas como objeto de reflexão linguística, isto é, quando se fala da própria palavra ou do conceito que ela designa, como no exemplo de *O Príncipezinho* (OPZ): *O que quer dizer “cativar”?* (OPZ-1058). O operador **:irony** assinala expressões empregadas com valor irónico. Por sua vez, **:name** estabelece uma relação de equivalência denominativa entre duas expressões, nomeadamente entre uma sigla e a respetiva forma desenvolvida. Estes mecanismos não foram objeto de análise no presente trabalho.

No presente estudo, a análise experimental destes mecanismos incide, portanto, diretamente sobre **:direct-speech** e **:ipsisverbis**, ambos presentes na frase selecionada para a análise qualitativa. Os marcadores **:direct-speech-cont** e **:direct-speech-before** integram a proposta descritiva, mas a sua avaliação sistemática exigirá um conjunto mais amplo de ocorrências. A Tabela 1 resume a proposta aqui apresentada.

**Tabela 1**

*Operadores de Representação de Configurações do Discurso Reportado no Modelo LMR*

| Mecanismo               | Função                                                              | Aplicação                                               |
|-------------------------|---------------------------------------------------------------------|---------------------------------------------------------|
| :direct-speech +        | Enunciado autónomo apresentado como discurso direto                 | Nó que encabeça todo o segmento citado                  |
| :direct-speech-cont +   | Continuação do discurso direto numa frase subsequente               | Nó que encabeça a continuação                           |
| :direct-speech-before + | Citação anterior ao predicado de enunciação                         | Nó que encabeça o segmento citado                       |
| :ipsisverbis +          | Reprodução literal integrada sintaticamente no enunciado do relator | Nó que encabeça a palavra, expressão ou fragmento       |
| :metalinguistic +       | Uso de uma expressão como objeto linguístico                        | Nó da expressão mencionada                              |
| :irony +                | Emprego irónico de uma expressão                                    | Nó da expressão usada ironicamente                      |
| :name                   | Equivalência denominativa entre expressões                          | Relação entre a designação e a expressão correspondente |

#### 4. Metodologia

A metodologia compreende cinco componentes: (i) a configuração de um GPT personalizado para produzir anotações semânticas segundo o modelo LMR; (ii) a constituição e preparação de um *corpus* de notícias; (iii) a seleção e anotação manual de um conjunto de frases de referência; (iv) a realização das experiências com os dois modelos e as quatro condições de *prompting*; e (v) a avaliação quantitativa e qualitativa das representações produzidas.

##### 4.1. Configuração do GPT personalizado

O sistema de anotação foi configurado como um GPT personalizado na plataforma ChatGPT, através da interface de criação e configuração disponibilizada pela OpenAI. O comportamento do GPT foi definido através de um conjunto de instruções de configuração, da disponibilização de documentos de referência na secção “Conhecimento” (*Knowledge*) e da ativação de determinadas funcionalidades disponibilizadas pela plataforma. Esta distinção é metodologicamente importante, uma vez que um GPT personalizado combina instruções, documentos de referência e funcionalidades ativadas, sem que os documentos carregados sejam incorporados nos parâmetros do modelo através de um novo processo de treino.

O GPT, designado *Anotador Semântico LMR*, foi configurado para produzir anotações semânticas de textos em português europeu segundo o modelo *Lexicalized Meaning Representation* (LMR). As representações solicitadas deveriam assumir a forma de grafos linearizados em notação PENMAN, com indentação correspondente ao encaixamento das relações, refletindo a estrutura das relações semânticas.

Na secção “Conhecimento” foram carregados dois documentos em formato PDF:

- As diretivas de anotação (teóricas e operacionais) do modelo LMR (Baptista, 2025), que constituíam a principal fonte normativa para a anotação (versão de 03/09/2025);
- Uma versão parcialmente anotada de *O Príncipezinho* segundo o modelo LMR, utilizada como fonte de exemplos concretos da aplicação das diretivas.

A configuração do sistema baseou-se ainda numa instrução principal (Anexo II), que definia o objetivo da tarefa, as fontes autorizadas, o formato das respostas e as restrições impostas ao sistema. A instrução determinava que a anotação fosse realizada exclusivamente com base nos documentos carregados na secção “Conhecimento”. O GPT foi expressamente instruído a não introduzir soluções não sustentadas nos documentos, nem a recorrer a exemplos próprios ou a inferências não sustentadas por essa documentação. Pretendeu-se, deste modo, reduzir a introdução de conceitos, relações ou soluções de anotação não previstas nas diretivas do LMR.

Para efeitos das experiências, também não foi permitido o recurso à Internet. A instrução principal proibia explicitamente a utilização de pesquisa ou navegação na Web (*web browsing*), devendo todas as respostas ser produzidas a partir da base documental fechada constituída pelos dois ficheiros acima identificados. Esta restrição procurou assegurar que as diferentes condições experimentais fossem realizadas com acesso às mesmas fontes de informação e que eventuais variações nos resultados não decorressem da consulta de recursos externos.

A instrução principal incluía igualmente uma política de evidência obrigatória para os casos em que o utilizador solicitasse uma explicação ou justificação das opções de anotação. Nessas situações, o sistema deveria apresentar uma citação literal, com uma extensão compreendida entre 10 e 80 palavras, retirada do documento de referência, e indicar a localização da passagem no ficheiro correspondente. Quando a informação solicitada não se encontrava nos documentos carregados, o GPT deveria declarar explicitamente que ela não estava disponível na base de conhecimento, em vez de produzir uma resposta com base em conhecimento geral ou numa inferência própria.

Foram ainda definidos quatro iniciadores de conversa, destinados a facilitar a formulação dos pedidos de anotação pelos utilizadores. Estes iniciadores apresentam variantes de instruções como “Anote semanticamente o seguinte texto” ou “Faça a anotação semântica da frase seguinte, segundo o modelo LMR”. Os iniciadores de conversa não modificam a instrução principal nem acrescentam conhecimento ao sistema; funcionam apenas como modelos de interação inicial.

Para assegurar a reprodutibilidade da configuração, apresentam-se, no Anexo I, o *prompt* utilizado nas interações com o modelo e, no Anexo II, as instruções de configuração do GPT personalizado. Esta versão desenvolve e reorganiza a informação que já constava da descrição metodológica preliminar, nomeadamente a utilização das diretivas LMR, das anotações de *O Príncipezinho*, da política de evidência e do formato PENMAN.

#### 4.2. Constituição do corpus

Foi constituído um *corpus* de conveniência composto por 42 notícias sobre inteligência artificial, publicadas entre 15 de maio de 2024 e 31 de dezembro de 2025. Os textos foram recolhidos nos seguintes órgãos de comunicação social: Euronews, Exame, Jornal de Notícias, LusoAI, Observador, Público, RTP Notícias e SIC Notícias. A opção por um *corpus* tematicamente homogêneo teve como objetivo reduzir a variabilidade associada ao domínio e concentrar a análise nos fenómenos linguísticos selecionados. O *corpus* não pretende, portanto, ser representativo do discurso jornalístico em geral.

Todas as notícias se encontravam disponíveis em acesso aberto e foram descarregadas manualmente das respetivas páginas eletrónicas. Após a remoção dos elementos não pertencentes ao texto da notícia (publicidade, menus, legendas, hiperligações, entre outros elementos), o *corpus* ficou constituído por 20 065 palavras e 810 frases. A extensão das notícias varia entre 156 e 1 333 palavras, com uma média de 478 palavras por texto. A segmentação em frases foi realizada manualmente e posteriormente revista pelos autores.

A cada frase foi atribuído um identificador único no formato **id-t-s**, em que o primeiro campo (t) corresponde ao número do texto e o segundo (s) ao número da frase. Este identificador permite relacionar cada segmento com as anotações manuais e automáticas produzidas durante as experiências.

Do *corpus* global, foram selecionadas cinco frases, que constituem o subconjunto de avaliação descrito na Secção 4.3. A seleção incluiu as frases mais relevantes para este estudo, nomeadamente construções com *verba dicendi* e discurso direto.

#### 4.3. Construção do conjunto de referência

A partir do *corpus* global, foram selecionadas cinco frases para constituir o conjunto de avaliação. A seleção foi deliberadamente orientada para a inclusão de construções com discurso reportado. Não se tratou, portanto, de uma amostra aleatória do *corpus*, mas de um conjunto destinado a testar fenómenos particularmente relevantes para o modelo LMR e para os objetivos deste estudo.

As cinco frases foram anotadas independentemente pelos dois autores, com base na mesma versão das diretivas LMR. As duas representações foram depois comparadas, e as divergências foram discutidas e adjudicadas por consenso. A versão consensual resultante deste processo constituiu o conjunto de referência utilizado nas avaliações qualitativa e quantitativa. Dado o caráter exploratório do estudo, não foi calculada uma medida formal de concordância interanotador antes da adjudicação, o que constitui uma limitação do processo de construção do conjunto de referência.

As cinco frases, as respetivas anotações consensuais e as respostas produzidas pelos modelos encontram-se disponíveis em Reis e Baptista (2025). Os textos integrais das notícias não são redistribuídos, de modo a respeitar as condições de acesso e os direitos de autor das fontes originais.

#### 4.4. Desenho experimental

O desenho experimental compreendeu quatro condições, que diferiam quanto ao número e ao domínio dos exemplos explicitamente incluídos no *prompt*. A instrução de base, a configuração do GPT e os documentos disponíveis na secção “Conhecimento” mantiveram-se constantes em todas as condições. Essa base incluía, além das diretivas de anotação, o documento parcialmente anotado de *O Príncipezinho*. Assim, a designação *zero-shot* refere-se apenas à ausência de exemplos inseridos diretamente no *prompt*, e não à ausência absoluta de exemplos disponíveis para recuperação pelo sistema.

Foram consideradas as seguintes condições:

- a. *Zero-shot* ao nível do *prompt*: não foi apresentado qualquer exemplo no *prompt*;
- b. *Few-shot out-of-domain*: foram apresentadas cinco frases anotadas de *O Príncipezinho*, selecionadas por conterem mecanismos relevantes para o estudo;
- c. *Many-shot out-of-domain*: foram apresentadas 10 frases anotadas provenientes da mesma obra;
- d. *Few-shot in-domain*: foram apresentadas cinco frases anotadas, provenientes do mesmo domínio jornalístico do conjunto de avaliação; nenhum dos exemplos *in-domain* coincidia com as cinco frases do conjunto de avaliação.

O conjunto completo de exemplos utilizado em cada condição é apresentado no Anexo IV.

As experiências foram realizadas entre 13 e 15 de outubro de 2025, na plataforma ChatGPT, utilizando os modelos GPT-4o e GPT-5-Thinking. Para cada combinação de modelo, condição e frase, foi iniciada uma nova conversa, de modo a impedir que o histórico de uma execução influenciasse as seguintes. Foram, assim, produzidas 40 anotações automáticas: cinco frases × quatro condições × dois modelos. Dado o caráter exploratório do estudo, foi realizada apenas uma execução por combinação experimental.

Após a produção de cada anotação, foi solicitado ao sistema que indicasse as diretivas que justificavam as diferentes partes da representação. Estas justificações não foram, porém, incluídas na avaliação.

#### 4.5. Procedimento de avaliação

As anotações automáticas foram comparadas com o conjunto manual de referência através da métrica Smatch (Cai & Knight, 2013)<sup>4</sup>, que procura o alinhamento ótimo entre as variáveis dos dois grafos e calcula precisão, cobertura e medida-F (*F-score*) a partir dos respetivos triplos. Os resultados foram calculados por frase e agregados por modelo e condição experimental.

A avaliação quantitativa foi complementada por uma análise qualitativa dos principais tipos de divergência. Para a comparação detalhada das condições, e por limitações de espaço, selecionou-se previamente a frase ID-001-004, por conter uma construção com verbo-suporte, uma estrutura interrogativa coordenada, um fragmento em discurso direto, bem como outras relações relevantes para o LMR.

### 5. Análise dos resultados

Os resultados são apresentados em três etapas. Em primeiro lugar, compara-se quantitativamente o desempenho dos modelos GPT-4o e GPT-5-Thinking nas quatro condições experimentais. Em seguida, procede-se à análise qualitativa aprofundada de uma das frases do conjunto de avaliação, selecionada por combinar vários fenómenos relevantes para o modelo LMR. Por fim, sistematizam-se os principais tipos de desvio identificados nas anotações produzidas automaticamente.

Dado o carácter exploratório do estudo, a dimensão reduzida do conjunto de avaliação e a realização de uma única execução por combinação experimental, a análise é essencialmente descritiva. Os resultados permitem comparar os modelos e as condições testadas, mas não devem ser interpretados como uma avaliação definitiva do desempenho dos sistemas na anotação automática segundo o LMR.

#### 5.1. Resultados quantitativos

Os resultados quantitativos são apresentados em três níveis de análise: por frase (Tabela 2), por modelo (Tabela 3) e por condição experimental (Tabela 4), sendo estes últimos complementados pela representação gráfica da Figura 1.

---

<sup>4</sup> Smatch, v1.0.4, 31/05/2020: <https://github.com/snowblink14/smatch>

**Tabela 2**

*Resultados Smatch por Frase*

| GPT-4o            |             |             |             | GPT-5-Thinking |             |             |                   |
|-------------------|-------------|-------------|-------------|----------------|-------------|-------------|-------------------|
| Frase             | P           | R           | F           | P              | R           | F           | Frase             |
| ID-001-003-1a.txt | 0,48        | 0,54        | 0,51        | 0,52           | 0,47        | 0,49        | ID-001-003-2a.txt |
| ID-001-003-1b.txt | 0,52        | <b>0,37</b> | 0,43        | 0,48           | 0,47        | 0,47        | ID-001-003-2b.txt |
| ID-001-003-1c.txt | 0,48        | 0,40        | 0,44        | 0,76           | 0,56        | 0,64        | ID-001-003-2c.txt |
| ID-001-003-1d.txt | 0,66        | 0,56        | 0,61        | 0,62           | 0,69        | 0,65        | ID-001-003-2d.txt |
| ID-001-004-1a.txt | <b>0,37</b> | 0,44        | <b>0,40</b> | 0,65           | 0,60        | 0,62        | ID-001-004-2a.txt |
| ID-001-004-1b.txt | 0,48        | 0,48        | 0,48        | 0,59           | 0,53        | 0,56        | ID-001-004-2b.txt |
| ID-001-004-1c.txt | 0,48        | 0,46        | 0,47        | <b>0,78</b>    | 0,67        | <b>0,72</b> | ID-001-004-2c.txt |
| ID-001-004-1d.txt | 0,48        | 0,47        | 0,47        | 0,61           | 0,52        | 0,56        | ID-001-004-2d.txt |
| ID-001-005-1a.txt | 0,59        | 0,67        | 0,63        | 0,59           | 0,59        | 0,59        | ID-001-005-2a.txt |
| ID-001-005-1b.txt | 0,49        | 0,56        | 0,52        | 0,75           | 0,70        | <b>0,72</b> | ID-001-005-2b.txt |
| ID-001-005-1c.txt | 0,67        | <b>0,69</b> | 0,68        | 0,63           | 0,68        | 0,65        | ID-001-005-2c.txt |
| ID-001-005-1d.txt | <b>0,73</b> | 0,66        | <b>0,69</b> | 0,73           | 0,65        | 0,69        | ID-001-005-2d.txt |
| ID-001-006-1a.txt | 0,43        | 0,43        | 0,43        | 0,62           | <b>0,72</b> | 0,67        | ID-001-006-2a.txt |
| ID-001-006-1b.txt | 0,48        | 0,43        | 0,45        | 0,38           | 0,42        | 0,40        | ID-001-006-2b.txt |
| ID-001-006-1c.txt | 0,52        | 0,41        | 0,46        | 0,71           | 0,68        | 0,69        | ID-001-006-2c.txt |
| ID-001-006-1d.txt | 0,57        | 0,60        | 0,58        | 0,57           | 0,67        | 0,62        | ID-001-006-2d.txt |
| ID-002-004-1a.txt | 0,42        | 0,51        | 0,46        | 0,67           | 0,65        | 0,66        | ID-002-004-2a.txt |
| ID-002-004-1b.txt | 0,56        | 0,54        | 0,55        | 0,71           | 0,63        | 0,67        | ID-002-004-2b.txt |
| ID-002-004-1c.txt | 0,53        | 0,51        | 0,52        | 0,67           | 0,59        | 0,63        | ID-002-004-2c.txt |
| ID-002-004-1d.txt | 0,50        | 0,50        | 0,50        | 0,68           | 0,68        | 0,68        | ID-002-004-2d.txt |

Nota. P: precisão; R: abrangência/coertura; F: F-score.

A Tabela 2 apresenta os valores de precisão, cobertura e medida-F (*F-score*) obtidos para cada frase, modelo e condição experimental. Cada frase constitui um ficheiro autónomo. Os identificadores de ficheiro seguem o formato ID-XXX-XXX-NC, em que ID-XXX-XXX corresponde ao identificador da frase no *corpus*, N indica a experiência: 1 = Experiência 1 (modelo 4o); 2 = Experiência 2 (modelo 5-Thinking); e C indica a condição de *prompting*: a = *zero-shot*; b = *few-shot* com cinco exemplos de *O Principezinho*; c = *many-shot* com 10 exemplos de *O Principezinho*; d = *few-shot* com cinco exemplos de textos jornalísticos. Os valores destacados a negrito indicam os melhores resultados obtidos, enquanto os valores destacados a itálico assinalam os valores mais baixos.

Os resultados individuais mostram uma variação considerável do *F-score* entre frases, modelos e condições experimentais. Na maioria das combinações avaliadas, o GPT-5-Thinking apresenta valores superiores aos do GPT-4o, embora se observem também casos em que o GPT-4o obtém um resultado equivalente ou superior. Esta variabilidade confirma que o desempenho não depende exclusivamente do modelo, sendo igualmente influenciado pela frase e pela condição de *prompting*.

Para obter uma visão agregada do desempenho dos modelos, foram também calculadas estatísticas globais sobre os resultados obtidos em todas as condições experimentais. A Tabela 3 apresenta os valores mínimo e máximo, a média e a mediana das métricas de avaliação aplicadas aos resultados de cada modelo.

**Tabela 3**

*Estatísticas Descritivas Globais dos Resultados Smatch*

|         | Modelo 4o |      |      | Modelo 5-Thinking |             |             |
|---------|-----------|------|------|-------------------|-------------|-------------|
|         | P         | R    | F    | P                 | R           | F           |
| MÍNIMO  | 0,37      | 0,37 | 0,40 | 0,38              | 0,42        | 0,40        |
| MÁXIMO  | 0,73      | 0,69 | 0,69 | <b>0,78</b>       | <b>0,72</b> | <b>0,72</b> |
| MÉDIA   | 0,52      | 0,51 | 0,51 | 0,64              | 0,61        | 0,62        |
| MEDIANA | 0,50      | 0,51 | 0,49 | <b>0,64</b>       | <b>0,64</b> | <b>0,65</b> |

*Nota.* P: precisão; R: abrangência/cobertura; F: F-score.

Considerando conjuntamente todas as frases e condições experimentais, o GPT-5-Thinking apresentou valores médios de precisão, cobertura e *F-score* superiores aos do GPT-4o. O *F-score* médio foi de 0,62 no GPT-5-Thinking e de 0,51 no GPT-4o, correspondendo a uma diferença de 0,11. As medianas confirmam a mesma tendência, com valores de 0,65 e 0,49, respetivamente. No conjunto analisado, as representações produzidas pelo GPT-5-Thinking apresentaram, portanto, uma correspondência globalmente maior com as anotações manuais de referência.

A Tabela 4 agrega os resultados pelas médias das condições de *prompting*. Os mesmos dados encontram-se representados no gráfico da Figura 1. A comparação entre as condições experimentais mostra, contudo, que o efeito da inclusão de exemplos não é uniforme. No GPT-4o, as condições *zero-shot* ao nível do *prompt* e *few-shot out-of-domain* obtiveram o mesmo *F-score* médio, de 0,49. O valor aumentou para 0,51 na condição *many-shot out-of-domain* e para 0,57 na condição *few-shot in-domain*, que apresentou o melhor resultado médio deste modelo.

No GPT-5-Thinking, a condição *few-shot out-of-domain* obteve um *F-score* médio de 0,56, inferior ao valor de 0,61 alcançado na condição sem exemplos explicitamente incluídos no *prompt*. O melhor resultado médio foi observado na condição *many-shot out-of-domain*, com 0,67, seguindo-se a condição *few-shot in-domain*, com 0,64. O aumento do número de exemplos de cinco para dez produziu, portanto, uma melhoria em ambos os modelos quando se comparam as duas condições *out-of-domain*, mas os resultados globais não sustentam uma relação linear ou cumulativa entre o número de exemplos e a qualidade das anotações.

**Tabela 4**

*Valores Médios de Precisão (P), Abrangência (R) e F-score (F) dos modelos GPT-4o e GPT-5-Thinking nas quatro*

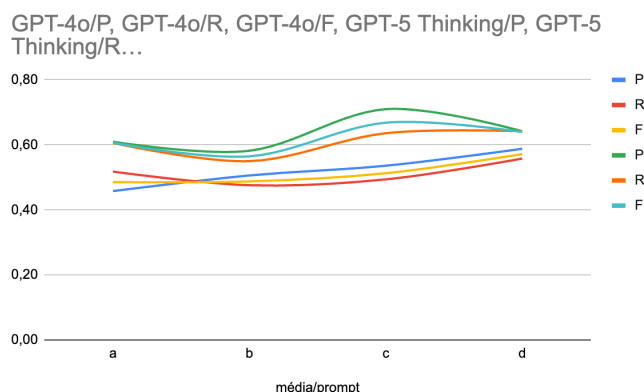
*Condições Experimentais (Métricas Smatch)*

|                             | Modelo 4o |      |      | Modelo 5-Thinking |      |      |
|-----------------------------|-----------|------|------|-------------------|------|------|
| Condição                    | P         | R    | F    | P                 | R    | F    |
| (a) <i>zero-shot</i>        | 0,46      | 0,52 | 0,49 | 0,61              | 0,61 | 0,61 |
| (b) <i>few-shot</i><br>OOD  | 0,51      | 0,48 | 0,49 | 0,58              | 0,55 | 0,56 |
| (c) <i>many-shot</i><br>OOD | 0,54      | 0,49 | 0,51 | 0,71              | 0,64 | 0,67 |
| (d) <i>few-shot ID</i>      | 0,59      | 0,56 | 0,57 | 0,64              | 0,64 | 0,64 |

*Nota.* OOD = out-of-domain, ID = in-domain.

**Figura 1**

Valores Médios de Precisão (P), Abrangência (R) e F-score (F) Obtidos pelos Modelos GPT-4o e GPT-5-Thinking nas Quatro Condições Experimentais (Métricas Smatch)



A Figura 1 evidencia o desempenho globalmente superior do GPT-5-Thinking (linhas superiores), cujos valores médios de precisão, cobertura e *F-score* se situam acima dos valores do GPT-4o (linhas inferiores) nas quatro condições experimentais. Da condição *zero-shot* ao nível do *prompt* (a) para a condição *few-shot out-of-domain* (b), observa-se uma diminuição do desempenho do GPT-5-Thinking, enquanto o *F-score* do GPT-4o se mantém inalterado. A passagem de cinco para dez exemplos *out-of-domain*, da condição (b) para a condição (c), produz uma melhoria em ambos os modelos, particularmente acentuada no GPT-5-Thinking. Finalmente, da condição *many-shot out-of-domain* (c) para a condição *few-shot in-domain* (d), o desempenho do GPT-5-Thinking diminui ligeiramente, ao passo que o do GPT-4o continua a aumentar.

Estes resultados mostram que o efeito dos exemplos depende do modelo e da condição experimental. No GPT-4o, a condição *few-shot in-domain* apresenta o melhor resultado médio, sugerindo um possível efeito favorável da proximidade de domínio. No GPT-5-Thinking, porém, o melhor desempenho é obtido com dez exemplos *out-of-domain*, pelo que a proximidade temática não explica isoladamente os resultados. Além do domínio e da quantidade, poderão intervir fatores como a proximidade lexical, sintática e estrutural entre os exemplos apresentados e as frases avaliadas.

Em síntese, os resultados quantitativos mostram que o GPT-5-Thinking produziu, em média, representações mais próximas do conjunto de referência do que o GPT-4o. O efeito dos exemplos variou, contudo, em função do modelo, da frase e da condição experimental. Não se observou uma relação linear entre o número de exemplos fornecidos e a qualidade das anotações, nem um efeito uniformemente favorável da proximidade de domínio.

## 5.2. Análise qualitativa: estudo de caso

Para interpretar linguisticamente alguns dos resultados quantitativos, selecionou-se como estudo de caso a frase ID-001-004, que se reproduz abaixo:

(ID-001-004) Na conferência, Sofia Casimiro levantou a questão: “Deverá a UE aumentar ou aligeirar a regulação da IA?”, acrescentando ser necessário um “equilíbrio entre este tipo de tecnologia e os valores fundamentais da comunidade europeia”. (Observador, 27 de fevereiro de 2025)

A frase combina vários fenómenos relevantes para o modelo LMR e para o objeto deste artigo: a construção com verbo-suporte *levantar a questão*, o auxiliar modal *dever*, uma estrutura interrogativa coordenada, um enunciado em discurso direto, um fragmento reproduzido literalmente, mas integrado sintaticamente no enunciado do relator e várias expressões nominais complexas. A anotação manual de referência é apresentada na Figura 2, enquanto a Figura 3 mostra uma visualização do mesmo grafo<sup>5</sup>.

**Figura 2**

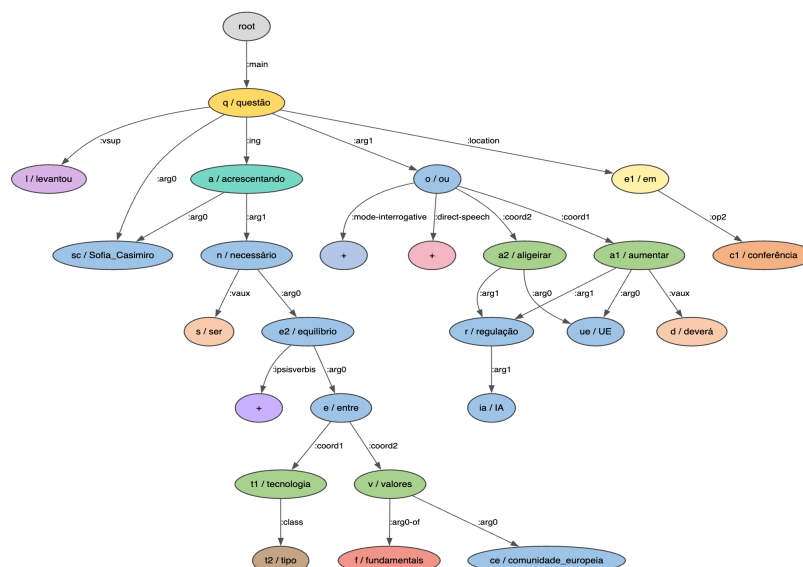
*Anotação de Referência da Frase ID-001-004 em Formato de Grafo PENMAN*

```
(root :main
(q / questão
:vsup (l / levantou)
:arg0 (sc / Sofia_Casimiro)
:location (e1 / em
:op2 (c1 / conferência))
:arg1 (o / ou
:mode-interrogative +
:direct-speech +
:coord1 (a1 / aumentar
:vaux (d / deverá)
:arg0 (ue / UE)
:arg1 (r / regulação
:arg1 (ia / IA)))
:coord2 (a2 / aligeirar
:arg0 ue
:arg1 r))
:ing (a / acrescentando
:arg0 sc
:arg1 (n / necessário
:vaux (s / ser)
:arg0 (e2 / equilíbrio
:ipsisverbis +
:arg0 (e / entre
:coord1 (t1 / tecnologia
:class (t2 / tipo))
:coord2 (v / valores
:arg0-of (f / fundamentais)
:arg0 (ce / comunidade_europeia)))))))))
```

<sup>5</sup> A representação visual do grafo foi obtida convertendo automaticamente o grafo PENMAN para o formalismo DOT e da posterior visualização do resultado com a ferramenta GraphvizOnline: <https://dreampuf.github.io/GraphvizOnline>

**Figura 3**

*Anotação de Referência da Frase ID-001-004 em Formato Visual de Grafos*



Na Figura 2, **:direct-speech** e **:ipsisverbis** codificam dois estatutos sintático-discursivos distintos de material reproduzido literalmente. A pergunta *Deverá a UE aumentar ou aligeirar a regulação da IA?* constitui um enunciado interrogativo autónomo, introduzido pela construção com verbo-suporte *levantou a questão*; por esse motivo, o nó que encabeça a coordenação recebe **:direct-speech +**. A expressão *equilíbrio entre este tipo de tecnologia e os valores fundamentais da comunidade europeia*, embora seja igualmente apresentada como reprodução literal, corresponde a um sintagma nominal integrado na construção *acrescentando ser necessário...*, sendo, por isso, marcada com **:ipsisverbis +**. A distinção não reside no grau de fidelidade da citação, mas no estatuto sintático e discursivo do segmento citado.

Os resultados quantitativos relativos a esta frase evidenciam diferenças claras entre os dois modelos e entre as condições experimentais. No GPT-4o, o *F-score* aumenta de 0,40 na condição *zero-shot* ao nível do *prompt* para 0,48 na condição *few-shot out-of-domain*, diminuindo ligeiramente para 0,47 nas condições *many-shot out-of-domain* e *few-shot in-domain*. No GPT-5-Thinking, os valores são globalmente superiores: 0,62 na condição *zero-shot*, 0,56 na condição *few-shot out-of-domain*, 0,72 na condição *many-shot out-of-domain* e 0,56 na condição *few-shot in-domain*. O melhor resultado para esta frase é, portanto, obtido pelo GPT-5-Thinking com dez exemplos *out-of-domain*, enquanto a introdução de cinco exemplos, quer do mesmo domínio quer de outro domínio, não melhora o resultado relativamente à condição *zero-shot*.

Estes valores mostram que o efeito dos exemplos não é linear nem uniforme e justificam uma análise mais detalhada das soluções linguísticas produzidas em cada condição.

Nas subsecções seguintes, comparam-se as representações obtidas pelo GPT-4o e pelo GPT-5-Thinking nas quatro condições experimentais.

### 5.2.1. Resultados do GPT-4o

Na condição *zero-shot* ao nível do *prompt* (Experiência 1A, Figura 4), o modelo obtém um *F-score* de 0,40 e recupera apenas parcialmente a estrutura geral da frase. A construção com verbo-suporte *levantar a questão* não é reconhecida: *levantar* é tratado como predicado lexical pleno, tendo *questão* como argumento,

em vez de ser representado através da relação **:vsup**. O auxiliar modal *deverá* é igualmente analisado como predicado lexical. A modalidade interrogativa é representada através da relação **:mode**, ligada ao nó não lexicalizado (**i / interrogative**), em lugar do marcador binário **:mode-interrogative +**. A coordenação introduzida por *ou* também não é preservada: a conjunção desaparece e os predicados *aumentar* e *aligeirar* são representados como operadores associados a *dever*. Observam-se ainda desvios na análise do classificador nominal *tipo* e das expressões nominais complexas. Por fim, não são representados os marcadores **:direct-speech + c** e **:ipsisverbis +**.

#### Figura 4

*Experiência 1A (zero-shot, GPT-4o)*

```
(ROOT :MAIN (l / levantar
:ARG0 (s / Sofia Casimiro)
:ARG1 (q / questão
:topic (d / dever
:ARG0 (ue / UE)
:ARG1 (r / regulação
:mod (ia / IA))
:mode (i / interrogative)
:op1 (a1 / aumentar)
:op2 (a2 / aligeirar)))
:location (c / conferência))
:COORD (a / acrescentar
:ARG0 s
:ARG1 (n / necessário
:ARG1 (e / equilíbrio
:ARG0 (t / tipo
:mod (t2 / tecnologia))
:ARG1 (v / valores
:mod (f / fundamentais)
:poss (ce / comunidade europeia))))))
```

*Nota.* Em todos os exemplos, os elementos destacados a negrito indicam divergências relativamente à anotação de referência.

Na condição *few-shot out-of-domain*, com cinco exemplos de *O Príncipezinho* (Experiência 1B, Figura 5), o modelo obtém um *F-score* de 0,48, superior ao da condição *zero-shot*. Nesta execução, passa a introduzir o marcador **:direct-speech +**, mas aplica-o separadamente aos predicados *aumentar* e *aligeirar*, em vez de o associar ao nó que encabeça toda a coordenação. Persistem a não identificação da construção com verbo-suporte *levantar a questão* e o tratamento inadequado do auxiliar modal *deverá*. Surgem ainda relações não previstas nas diretivas LMR, como **:content** e **:theme-of**. A preposição *entre*, associada à construção simétrica do nome predicativo *equilíbrio*, é analisada como uma relação locativa, não sendo adequadamente representada a estrutura argumental da construção. O marcador **:ipsisverbis +** continua ausente.

**Figura 5**

*Experiência 1B (few-shot out-of-domain, GPT-4o)*

```
(root :main (l / levantou
  :arg0 (sc / Sofia Casimiro)
  :arg1 (q / questão
    :content (a / aumentar :direct-speech +
      :mod (d / deverá)
      :arg0 (ue / UE)
      :coord2 (a2 / aligeirar :direct-speech +
        :arg0 ue)
        :arg1 (r / regulação
          :theme-of (t / tecnologia
            :mod (e / IA))))))
    :location (c / conferência))
  :mod (a / acrescentando
    :arg0 sc
    :arg1 (n / necessário
      :arg1 (e / equilíbrio
        :location (e2 / entre
          :op1 (t2 / tecnologia
            :mod (t3 / tipo
              :mod (e3 / este)))
          :op2 (v / valores
            :mod (f / fundamentais)
            :poss (ce / comunidade europeia)))))))
```

Na condição *many-shot out-of-domain*, com dez exemplos anotados de *O Príncipezinho* (Experiência 1C, Figura 6), o modelo obtém um *F-score* de 0,47, praticamente idêntico ao da condição anterior. O aumento de cinco para dez exemplos não se traduz, portanto, numa melhoria desta representação. O marcador **:direct-speech +** é novamente introduzido, mas a construção com verbo-suporte *levantar a questão* continua a não ser reconhecida. A saída inclui ainda categorias não previstas nas diretivas LMR, como **:named-entity** e **:theme**. Embora estas designações possam ser compatíveis com convenções utilizadas noutros esquemas de representação, não é possível determinar a origem concreta da sua introdução; pode apenas concluir-se que não são sustentadas pela base documental fornecida ao GPT. O marcador **:ipsisverbis +** também não é representado.

**Figura 6**

*Experiência 1C (many-shot out-of-domain, GPT-4o)*

```
(root :main (l / levantar
  :arg0 (s / Sofia Casimiro :named-entity +)
  :arg1 (q / questão
    :topic (d / [e] :direct-speech +
      :coord1 (a1 / aumentar
        :arg0 (ue / UE :named-entity +)
        :arg1 (r / regulação
          :theme (ia / IA :named-entity +)))
      :coord2 (a2 / aligeirar
        :arg0 ue
        :arg1 r)))
    :location (c / conferência))
  :coord2 (a / acrescentar
    :arg0 s
    :arg1 (n / necessário
      :arg1 (e / equilíbrio
        :theme (e2 / e
          :coord1 (t / tipo
            :mod (t2 / tecnologia))
            :coord2 (v :mod (e3 / este) / valores
              :mod (f / fundamentais)
              :poss (c / comunidade europeia :named-entity +))))))
```

Finalmente, na condição *few-shot in-domain*, com cinco exemplos provenientes de textos jornalísticos (Experiência 1D, Figura 7), o modelo obtém um *F-score* de 0,47. A saída retoma uma estrutura muito próxima da produzida na condição *zero-shot*. Deixam de surgir algumas das categorias não previstas observadas nas condições anteriores, como *:content*, *:theme* e *:named-entity*, mas mantêm-se os principais problemas de análise. O verbo *dever* continua a ser tratado como núcleo predicativo da pergunta, em vez de auxiliar modal representado através de *:vaux*. A construção com verbo-suporte *levantar a questão* também não é reconhecida, e os marcadores *:direct-speech +*, *:ipsisverbis +* e *:mode-interrogative +* permanecem ausentes. Para esta frase, os resultados não evidenciam, portanto, uma vantagem específica dos exemplos *in-domain* relativamente aos exemplos *out-of-domain*.

**Figura 7**

*Experiência 1D (few-shot in-domain, GPT-4o)*

```
(ROOT :MAIN (1 / levantar
  :ARG0 (s / Sofia Casimiro)
  :ARG1 (q / questão
    :topic (d / dever
      :ARG0 (ue / UE)
      :ARG1 (r / regulação
        :mod (ia / IA))
        :mode (i / interrogative)
        :op1 (a1 / aumentar)
        :op2 (a2 / aligeirar)))
    :location (c / conferência))
  :COORD (a / acrescentar
    :ARG0 s
    :ARG1 (n / necessário
      :ARG1 (e / equilíbrio
        :ARG0 (t / tipo
          :mod (t2 / tecnologia))
          :ARG1 (v / valores
            :mod (f / fundamentais)
            :poss (c / comunidade europeia))))))
```

### 5.2.2. Resultados do GPT-5-Thinking

Nesta segunda experiência, foram replicadas exatamente as mesmas condições experimentais utilizadas na primeira experiência, nomeadamente os diferentes cenários de *prompting* usados com o modelo GPT-4o, mas agora recorrendo ao modelo GPT-5-Thinking. Desta forma, foi possível observar de que modo cada modelo responde às mesmas instruções e aos mesmos exemplos, permitindo avaliar diferenças de precisão e a consistência entre dois modelos da família GPT.

Na condição *zero-shot* ao nível do *prompt* (Experiência 2A, Figura 8), o GPT-5-Thinking obtém um *F-score* de 0,62 e produz uma representação mais próxima da referência do que o GPT-4o na condição correspondente. Em particular, reconhece a construção com verbo-suporte *levantar a questão*, representando *questão* como predicado e *levantou* através da relação *:vsup*. Recupera também a coordenação encabeçada por *ou* e a partilha dos argumentos entre *aumentar* e *aligeirar*. Persistem, contudo, divergências significativas. O auxiliar modal *deverá* não é representado, a modalidade interrogativa não é assinalada e os marcadores *:direct-speech +* e *:ipsisverbis +* estão ausentes. Observam-se ainda diferenças na representação de *acrescentando* e da construção introduzida por *entre*. Formalmente, a saída é apresentada sem indentação, embora a parentetização permita recuperar a estrutura do grafo.

## Figura 8

Experiência 2A (zero-shot, GPT-5-Thinking)

```
(root
:location (e / em
:op2 (c / conferência))
:main (q / questão
:vsup (l / levantou)
:arg0 (sc / Sofia Casimiro)
:arg1 (o / ou
:coord1 (a1 / aumentar
:arg0 (ue / UE)
:arg1 (r / regulação
:topic (d / de
:op2 (ia / IA))))
:coord2 (a2 / aligeirar
:arg0 ue
:arg1 r)))
:mod (a / acrescentando
:op2 (n / necessário
:vaux (s / ser)
:arg0 (e / equilíbrio
:location (entre
:op2 (e2 / e
:coord1 (t / tipo
:mod (e3 / este)
:arg1 (t2 / tecnologia))
:coord2 (v / valores
:arg0-of (f / fundamentais)
:arg1 (ce / comunidade europeia))))))
```

Na condição *few-shot out-of-domain*, com cinco exemplos de *O Príncipezinho* (Experiência 2B, Figura 9), o modelo obtém um *F-score* de 0,56, inferior ao da condição *zero-shot*. A estrutura proposicional associada à questão é introduzida através da relação **:appos**, e a construção com verbo-suporte deixa de ser reconhecida. A coordenação disjuntiva é incorretamente representada por (e / e), em lugar da conjunção *ou*. A modalidade interrogativa é assinalada através de **:mode-interrogative**, mas sem o valor binário +. Na construção simétrica, os complementos introduzidos por *entre* são tratados como modificadores de *equilíbrio*, em vez de argumentos deste nome predicativo. O classificador nominal *tipo* também não é adequadamente analisado, e a relação de *valores* com *comunidade europeia* diverge da anotação de referência. Os marcadores **:direct-speech + c** **:ipsisverbis +** não são representados.

**Figura 9**

*Experiência 2B (few-shot out-of-domain, GPT-5-Thinking)*

```
(root
:location (e / em
:op2 (c / conferência))
:main (l / levantou
:arg0 (sc / Sofia_Casimiro)
:arg1 (q / questão
:appos (d / deverá :mode-interrogative
:arg0 (ue / UE)
:arg1 (e / e
:coord1 (a1 / aumentar
:arg1 (r / regulação
:arg1 (ia / IA)))
:coord2 (a2 / aligeirar
:arg1 r))))))
:ing (a / acrescentando
:arg0 sc
:arg1 (n / necessário
:vaux (s / ser)
:arg0 (eq / equilíbrio
:mod (entre / entre
:op2 (e2 / e
:coord1 (t1 / tipo
:mod (e3 / este)
:mod (t2 / tecnologia))
:coord2 (v / valores
:mod (f / fundamentais)
:arg1 (ce / comunidade_europeia)))))))))
```

Na condição *many-shot out-of-domain*, com dez exemplos de *O Príncipezinho* (Experiência 2C, Figura 10), o GPT-5-Thinking obtém o melhor resultado para esta frase: P = 0,78, R = 0,67 e F = 0,72. A construção com verbo-suporte é corretamente reconhecida e a coordenação interrogativa é encabeçada pela conjunção *ou*. O modelo reconstitui ainda o auxiliar elidido [*deverá*] no segundo membro da coordenação. Apesar da maior proximidade global à referência, persistem divergências importantes. O marcador *:mode-interrogative* não recebe o valor positivo +. O marcador *:direct-speech +* é corretamente associado à coordenação encabeçada por *ou*, mas é também aplicado ao adjetivo *necessário*. Neste segundo caso, a anotação de referência exige *:ipsisverbis +* associado ao nome *equilíbrio*, uma vez que o material reproduzido literalmente é um sintagma nominal integrado no enunciado do relator. Observam-se ainda diferenças na representação do locativo *na conferência* e da construção introduzida por *entre*. Esta saída mostra que um *F-score* relativamente elevado pode coexistir com desvios linguisticamente relevantes, sobretudo na representação do estatuto discursivo dos segmentos citados.

**Figura 10**

*Experiência 2C (many-shot out-of-domain, GPT-5-Thinking)*

```
(root :main (q / questão
:vsup (l / levantou)
:arg0 (sc / Sofia_Casimiro)
:location (e / em
:op2 (c / conferência))
:op2 (ou / ou :direct-speech + :mode-interrogative
:coord1 (aum / aumentar
:vaux (d / deverá)
:arg0 (ue / UE)
:arg1 (r / regulação
:arg1 (ia / IA)))
:coord2 (ali / aligeirar
:vaux (d2 / [deverá])
:arg0 ue
:arg1 r))
:mod (a / acrescentando
:arg0 sc
:arg1 (n / necessário :direct-speech +
:vaux (s / ser)
:arg0 (eq / equilíbrio
:topic (ent / entre
:op2 (t / tipo
:mod (e1 / este)
:arg1 (tec / tecnologia))
:op2 (v / valores
:arg0-of (f / fundamentais)
:arg1 (ce / comunidade_europeia))))))
```

Na condição *few-shot in-domain*, com cinco exemplos provenientes de textos jornalísticos (Experiência 2D, Figura 11), o modelo obtém um *F-score* de 0,56, inferior aos valores das condições *zero-shot* e *many-shot out-of-domain*. A saída introduz expansões lexicais não realizadas no texto, representando *UE* como *União Europeia* e *IA* como *inteligência artificial*, às quais associa a relação **:name**. Embora semanticamente compatíveis com as siglas, estas expansões contrariam o princípio de lexicalização do LMR. A construção com verbo-suporte *levantar a questão* deixa igualmente de ser reconhecida, e *deverá* é tratado como predicado lexical, em vez de auxiliar modal. Além disso, não são representados os marcadores **:mode-interrogative +**, **:direct-speech +** e **:ipsisverbis +**. Nesta frase, os exemplos *in-domain* não produziram, portanto, uma melhoria relativamente à condição *zero-shot*.

**Figura 11**

*Experiência 2D (few-shot in-domain, GPT-5-Thinking)*

```
(root :main
  (1 / levantou
    :arg0 (sc / Sofia_Casimiro)
    :location (na / na
      :op2 (c / conferência)
      :arg1 (q / questão
        :op2 (d / deverá
          :arg0 (ue / União_Europeia
            :name (UE / UE))
            :arg1 (ou / ou
              :coord1 (a / aumentar
                :arg1 (r / regulação
                  :arg1 (ia / inteligência_artificial
                    :name (IA / IA))))
                :coord2 (al / aligeirar
                  :arg1 r))))
              :mod (a2 / acrescentando
                :arg0 sc
                :arg1 (n / necessário
                  :vaux (s / ser)
                  :arg0 (e / equilibrio
                    :mod (entre / entre
                      :op2 (e2 / e
                        :coord1 (t1 / tipo
                          :mod (e3 / este)
                          :arg1 (t2 / tecnologia))
                        :coord2 (v / valores
                          :arg0-of (f / fundamentais
                            :poss (ce / comunidade_europeia))))))))))
```

Em síntese, a análise qualitativa mostra que os dois modelos recuperam parte substancial da estrutura semântica da frase, embora nenhuma das condições produza uma representação integralmente coincidente com a referência. O GPT-5-Thinking demonstra, em algumas condições, maior capacidade para reconhecer propriedades linguísticas específicas do LMR. Em particular, identifica a construção com verbo-suporte *levantar a questão* nas condições *zero-shot* e *many-shot out-of-domain*, fenómeno que não é reconhecido pelo GPT-4o em nenhuma das quatro condições.

As soluções corretas não são, contudo, preservadas de forma consistente entre condições. No GPT-5-Thinking, a construção com verbo-suporte desaparece quando são fornecidos cinco exemplos *out-of-domain* ou cinco exemplos *in-domain*. De modo semelhante, os marcadores de discurso reportado são ora omitidos, ora aplicados ao nó incorreto, e o aumento do número de exemplos não conduz a uma acumulação estável das soluções adequadas.

Para esta frase, a melhor representação é produzida pelo GPT-5-Thinking na condição *many-shot out-of-domain*, com um *F-score* de 0,72. Ainda assim, essa saída mantém diferenças assinaláveis na representação da modalidade interrogativa e do fragmento marcado como reprodução literal. O estudo de caso confirma, assim, que o efeito dos exemplos não é uniforme nem cumulativo e que um valor Smatch relativamente elevado não implica conformidade integral com as diretivas do LMR.

Além das divergências específicas identificadas no estudo de caso, a análise das representações automáticas permitiu estabelecer uma tipologia transversal dos principais desvios relativamente à anotação de referência e às diretivas do LMR. A Tabela 5 organiza esses desvios segundo a sua natureza, apresenta um exemplo representativo de cada categoria e indica a principal consequência para a interpretação ou o processamento do grafo. A tipologia tem carácter descritivo e não pretende quantificar a frequência dos diferentes fenómenos. As categorias também não são mutuamente exclusivas, uma vez que uma mesma ocorrência pode afetar simultaneamente a validade formal do grafo, a análise linguística e a conformidade com o modelo LMR.

Tabela 5

Tipologia dos Desvios Observados nas Anotações Automáticas

| Categoria                                     | Descrição                                                                   | Exemplo                                                                                            | Consequência principal                                           |
|-----------------------------------------------|-----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| <b>Validade formal</b>                        | Omissão de variável, operador de instanciação /, parêntese ou valor binário | <b>:location (entre ...)</b> , sem variável nem /                                                  | Grafo inválido, incompleto ou não processável                    |
| <b>Nomeação das variáveis</b>                 | Utilização de uma variável diferente da convenção adotada para o conceito   | <b>(i / IA)</b> em vez de <b>(ia / IA)</b>                                                         | Desvio formal geralmente neutralizado pelo alinhamento do Smatch |
| <b>Estrutura argumental</b>                   | Seleção inadequada do predicado ou das relações com os seus argumentos      | <i>levantar</i> como predicado principal, em vez de <i>questão</i> com <b>:vsup (l / levantar)</b> | Alteração da estrutura semântica                                 |
| <b>Representação de elementos gramaticais</b> | Tratamento inadequado de auxiliares, conjunções ou preposições predicativas | <i>dever</i> como núcleo predicativo, em vez de auxiliar com <b>:vaux</b>                          | Perda ou distorção de informação gramatical                      |
| <b>Adesão ao esquema</b>                      | Introdução de relações ou marcadores não previstos nas diretivas            | <b>:content, :theme, :named-entity</b>                                                             | Incompatibilidade com o inventário do LMR                        |
| <b>Discurso reportado</b>                     | Omissão ou associação inadequada dos marcadores correspondentes             | <b>:direct-speech +</b> associado a <i>necessário</i>                                              | Representação incorreta do estatuto enunciativo                  |
| <b>Lexicalização</b>                          | Introdução de formas lexicais não realizadas no texto                       | <b>(ue / União Europeia)</b> em lugar de <b>(ue / UE)</b>                                          | Enfraquecimento da correspondência entre texto e grafo           |
| <b>Codificação de unidades multipalavra</b>   | Segmentação gráfica inadequada de unidades constituídas por várias palavras | <b>(sc / Sofia Casimiro)</b> em vez de <b>(sc / Sofia_Casimiro)</b>                                | Desvio às convenções de codificação lexical e PENMAN             |
| <b>Codificação gráfica</b>                    | Problemas na reprodução de diacríticos ou símbolos                          | <i>n~ao</i> por não                                                                                | Alteração da forma lexical                                       |

Os desvios de natureza formal compreendem, em primeiro lugar, os casos que afetam a boa formação dos grafos, como a omissão de variáveis, do operador de instanciação /, de parênteses ou do valor positivo + associado aos marcadores binários. Estas ocorrências podem impedir o processamento automático da representação e devem, por isso, ser distinguidas dos desvios propriamente semânticos. Foram também observadas diferenças nas convenções de nomeação das variáveis, como (i / IA) em vez de (ia / IA). Embora contrariem as diretivas adotadas, estas diferenças não alteram necessariamente a estrutura abstrata do grafo e tendem a ser neutralizadas pelo alinhamento de variáveis realizado pelo Smatch. Incluem-se ainda nesta dimensão alguns problemas de codificação de unidades multipalavra, como a utilização de espaços em vez de underscore (\_).

Um segundo conjunto de desvios incide sobre a análise lexical, gramatical e sintático-semântica. O caso mais evidente no estudo de caso é a interpretação de *levantar* como predicado lexical pleno, em vez do

reconhecimento da construção com verbo-suporte *levantar a questão*. Foram igualmente identificadas dificuldades no tratamento de verbos auxiliares, conjunções coordenativas, preposições predicativas, classificadores nominais e expressões nominais complexas. Estes desvios alteram a seleção dos predicados, a organização dos respetivos argumentos ou a informação gramatical preservada na representação.

Uma terceira dimensão diz respeito à adesão às diretivas e ao inventário de relações e marcadores do LMR. Algumas saídas introduzem categorias não previstas, como **:content**, **:theme** e **:named-entity**. Estas categorias podem ser compatíveis com padrões utilizados noutros esquemas de representação ou resultar de soluções geradas pelo próprio modelo, mas não é possível determinar diretamente a origem destas escolhas. Pode apenas afirmar-se que as categorias introduzidas não são sustentadas pela base documental fornecida ao GPT. A restrição do acesso à Internet não é por si só relevante para explicar estas ocorrências, uma vez que a ausência de navegação não elimina os padrões adquiridos pelo modelo durante o pré-treino.

Nesta mesma dimensão incluem-se os desvios relativos ao discurso reportado. Observam-se tanto omissões como aplicações inadequadas de **:direct-speech +** e **:ipsisverbis +**. Em alguns casos, **:direct-speech +** é introduzido, mas associado a um nó que não encabeça todo o enunciado citado; noutros, é aplicado a material que não possui autonomia sintática e discursiva. O marcador **:ipsisverbis +**, por sua vez, é frequentemente omitido no sintagma nominal apresentado como reprodução literal.

Foram ainda observados desvios ao princípio de lexicalização, nomeadamente a expansão de *UE* e *IA* para *União Europeia* e *inteligência artificial*, respetivamente. Embora estas formas sejam semanticamente compatíveis com as siglas no contexto considerado, não correspondem às unidades efetivamente realizadas no texto e reduzem, por isso, a correspondência direta entre a frase e o grafo que caracteriza o modelo LMR.

No seu conjunto, os desvios distribuem-se, portanto, por três dimensões principais: a validade formal e a codificação dos grafos, a análise lexical e sintático-semântica das construções e a adesão às diretivas e ao inventário do LMR. Os modelos produzem representações parcialmente compatíveis com a referência, mas continuam a exigir supervisão humana para assegurar tanto a boa formação dos grafos como a adequação da análise linguística.

## 6. Conclusão

Este trabalho apresentou a configuração e a avaliação exploratória de um GPT personalizado para a produção de anotações semânticas abstratas de textos em português europeu segundo o modelo da *Lexicalized Meaning Representation* (LMR). O estudo contribuiu igualmente para o desenvolvimento do próprio formalismo, através da introdução de relações e marcadores destinados a representar determinadas configurações de discurso reportado e de citação ou reprodução literal, nomeadamente os marcadores binários **:direct-speech** e **:ipsisverbis**. Estes dois marcadores foram analisados no estudo de caso qualitativo.

Os resultados quantitativos mostram que, no conjunto analisado, o GPT-5-Thinking produziu anotações mais próximas das representações manuais de referência do que o GPT-4o. O GPT-5-Thinking obteve um *F-score* médio de 0,62, comparativamente a 0,51 no GPT-4o. A análise qualitativa mostrou também que o GPT-5-Thinking reconhece, em determinadas condições, propriedades linguísticas que não foram identificadas pelo GPT-4o, como a construção com verbo-suporte *levantar a questão*. Nenhuma das configurações avaliadas produziu, contudo, uma representação integralmente coincidente com a anotação de referência.

O efeito dos exemplos incluídos no *prompt* revelou-se não uniforme e não cumulativo. No GPT-4o, os valores médios aumentaram globalmente entre a condição sem exemplos explícitos e a condição *few-shot in-domain*, que apresentou o melhor resultado deste modelo. No GPT-5-Thinking, pelo contrário, a condição *few-shot out-of-domain* produziu um resultado inferior ao obtido sem exemplos explícitos, enquanto o melhor desempenho foi alcançado na condição *many-shot out-of-domain*. O aumento do número de exemplos e a proximidade temática não determinam, portanto, por si só, a qualidade da anotação. O seu efeito parece depender igualmente do modelo utilizado e da proximidade lexical, sintática e estrutural entre os exemplos fornecidos e as frases a anotar.

A análise dos desvios confirma que os modelos conseguem produzir grafos parcialmente compatíveis com as diretivas LMR e próximos da anotação de referência, mas continuam a apresentar dificuldades em três dimensões principais: a validade formal e a codificação dos grafos, a análise lexical e sintático-semântica das construções e a adesão estrita ao inventário de relações e marcadores. Entre os fenómenos observados encontram-se a omissão de elementos necessários à formação de grafos válidos, a análise inadequada de verbos-suporte e de verbos auxiliares, a introdução de formas lexicais não realizadas no texto, a aplicação incorreta dos marcadores de discurso reportado e a introdução de relações não previstas nas diretivas. A supervisão humana permanece, por isso, indispensável para validar e corrigir as anotações produzidas automaticamente.

Estes resultados devem ser interpretados à luz das limitações do estudo. A avaliação incidiu sobre apenas cinco frases, deliberadamente selecionadas por conterem fenómenos relevantes para os objetivos da investigação, e não sobre uma amostra aleatória ou representativa do *corpus*. O *corpus* é, além disso, tematicamente homogêneo, sendo composto por notícias sobre inteligência artificial. Foi realizada uma única execução por combinação de modelo, condição e frase, apesar da variabilidade conhecida das respostas dos modelos generativos. Também não foi calculada uma medida formal de concordância entre as anotações manuais independentes que precederam a adjudicação do conjunto de referência. Por fim, a análise qualitativa aprofundada concentrou-se numa única frase, não permitindo generalizar a distribuição dos tipos de erro ao conjunto do *corpus*.

O trabalho futuro deverá ampliar e diversificar o conjunto de avaliação, incluindo textos de diferentes temáticas e géneros discursivos e uma maior variedade de estruturas linguísticas. Será igualmente necessário repetir cada condição experimental várias vezes, de modo a medir a variabilidade dos resultados, e calcular formalmente a concordância entre anotadores durante a construção do conjunto de referência. A influência dos exemplos poderá ser estudada de forma mais controlada, distinguindo quantidade, domínio, proximidade lexical e semelhança estrutural. Pretende-se ainda avaliar o efeito de diferentes configurações técnicas – como a presença ou ausência da base de conhecimento, o formato dos documentos de referência e a disponibilidade de pesquisa na Web – e relacionar o desempenho com medidas de complexidade das frases e dos grafos, incluindo o número de predicados, a profundidade da representação, a coordenação e a subordinação.

No que respeita à representação do discurso reportado, será ainda necessário validar os operadores **:direct-speech-cont** e **:direct-speech-before** num *corpus* que inclua sequências de várias frases e permita analisar, respetivamente, a continuidade do discurso direto para além dos limites da frase introdutora e a ocorrência do segmento citado antes do predicado de enunciação. A análise deverá também ser alargada a outros operadores discursivos e metalinguísticos em desenvolvimento no quadro do LMR, nomeadamente os marcadores **:metalinguistic +**, **:irony +** e a relação **:name**. A comparação poderá ainda abranger outros modelos de linguagem e outras versões da família GPT, aplicando-lhes o mesmo protocolo e as mesmas métricas.

Globalmente, o estudo mostra que um GPT personalizado pode constituir uma ferramenta útil de apoio à anotação semântica, sobretudo como mecanismo de produção de primeiras versões a rever por anotadores humanos especialistas. A sua utilidade, no atual estado de desenvolvimento, reside menos na substituição da anotação humana do que na possibilidade de acelerar a construção de recursos linguísticos, desde que as representações produzidas sejam submetidas a validação sistemática. O *Anotador Semântico LMR*, no GPT configurado neste trabalho, encontra-se disponível para utilização pela comunidade<sup>6</sup>.

### Financiamento

Trabalho suportado por fundos nacionais através de financiamento da Fundação para a Ciência e a Tecnologia, I.P. (FCT) sob os projetos UID/50021/2025 (DOI: <https://doi.org/10.54499/UID/50021/2025>) e UID/PRR/50021/2025 (DOI: <https://doi.org/10.54499/UID/PRR/50021/2025>).

<sup>6</sup> Disponível em <https://chatgpt.com/g/g-67c0882e42d881919ca1d173d30035cc-anotador-semantico-lmr>

## Referências

- Anchiêta, R. (2020). *Abstract meaning representation parsing for the Brazilian Portuguese language* [Tese de doutoramento, Universidade de São Paulo].
- Anchiêta, R., Cabezudo, M., & Pardo, T. (2019). SEMA: An extended semantic evaluation metric for AMR. arXiv e-prints. <https://doi.org/10.48550/arXiv.1905.12069>
- Anchiêta, R., & Pardo, T. (2018). Towards AMR-BR: A SemBank for Brazilian Portuguese language. In *Proceedings of the 11th edition of the Language Resources and Evaluation Conference (LREC)* (pp. 974–979). ELRA. <https://aclanthology.org/L18-1157/>
- Antunes, D., Baptista, J., & Mamede, N. (2025). Processamento automático de expressões idiomáticas do Português Europeu. *Linguamática*, 17(1), 55–73. <https://linguamatica.com/index.php/linguamatica/article/view/464>
- Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.
- Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse* (pp. 178–186). Association for Computational Linguistics.
- Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2019). *Abstract Meaning Representation (AMR) 1.2.6 specification*. <https://github.com/amrisi/amr-guidelines/blob/master/amr.md>
- Baptista, J. (2025). *Lexical Meaning Representation – Guidelines* (Technical report). University of Algarve/INESC-ID Lisboa. [https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/public/LMR\\_Directives\\_20250903.pdf](https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/public/LMR_Directives_20250903.pdf)
- Baptista, J., & Mamede, N. (2020a). *Dicionário gramatical de verbos do Português Europeu*. Universidade do Algarve Editora.
- Baptista, J., & Mamede, N. (2020b). Syntactic transformations in rule-based parsing of support verb constructions: Examples from European Portuguese. In A. Simões, P. R. Henriques & R. Queirós (Eds.), *9th Symposium on Languages, Applications and Technologies (SLATE 2020)* (pp. 11:1–11:14). Schloss Dagstuhl–Leibniz-Zentrum für Informatik. <https://doi.org/10.4230/OASlcs.SLATE.2020.11>
- Baptista, J., Müller, I., & Reis, S. (2025). Semantic representation of adverbs in the Lexicalized Meaning Representation (LMR) framework. In *14th Symposium on Languages, Applications and Technologies (SLATE 2025), OpenAccess Series in Informatics (OASlcs)* (Vol. 135, pp. 9:1–9:18). Schloss Dagstuhl – Leibniz-Zentrum für Informatik. <https://doi.org/10.4230/OASlcs.SLATE.2025.9>
- Baptista, J., & Reis, S. (2026). Semantic representation of relative clauses in lexicalized Abstract Meaning Representation. In M. Souza, I. de-Dios-Flores, D. Santos, L. Freitas, J. W. C. Souza, & E. Ribeiro (Eds.), *Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026) – Volume 1* (pp. 203–211). Association for Computational Linguistics. <https://aclanthology.org/2026.propor-1.20>
- Baptista, J., Reis, S., Dias, J., & Santos, P. (2024a). Lexicalized Meaning Representation (LMR). In *Proceedings of the Fifth International Workshop on Designing Meaning Representations @ LREC-COLING 2024* (pp. 101–111). ELRA and ICCL.
- Baptista, J., Reis, S., Santos, P., & Dias, J. (2024b). Comparação de anotações AMR entre línguas: Problemas e soluções. *Revista da Associação Portuguesa de Linguística*, (11), 27–54. <https://doi.org/10.26334/2183-9077/rapln11ano2024a2>
- Bevilacqua, M., Blloshmi, R., & Navigli, R. (2021). One SPRING to rule them both: Symmetric AMR semantic parsing and generation without a complex pipeline. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14), 12564–12573. <https://doi.org/10.1609/aaai.v35i14.17489>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler,

- D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Cai, S., & Knight, K. (2013). Smatch: An evaluation metric for semantic feature structures. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Vol. 2, pp. 748–752). Association for Computational Linguistics.
- Damonte, M., Cohen, S. B., & Satta, G. (2017). An incremental parser for Abstract Meaning Representation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* (Vol. 1, pp. 536–546). Association for Computational Linguistics.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, Volume 1: Long and Short Papers (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
- Fahim, A., Azad, S., Rashid, T., Albaidaci, E., & Ighzali, H. (2024). Harnessing Large Language Models for transformative applications in Natural Language Processing. *European Journal of Theoretical and Applied Sciences*, 2(6), 428–438. [https://doi.org/10.59324/ejtas.2024.2\(6\).37](https://doi.org/10.59324/ejtas.2024.2(6).37)
- Fernandes, W., Motta, E., & Milidiú, R. (2011). Quotation extraction for Portuguese. In *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology* (STIL 2011) (pp. 204–208). Sociedade Brasileira de Computação.
- Freitas, C., Freitas, B., & Santos, D. (2016). QUEMDISSE? Reported speech in Portuguese. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (LREC 2016) (pp. 4410–4416). European Language Resources Association (ELRA).
- He, X., Lin, Z., Gong, Y., Jin, A. L., Zhang, H., Lin, C., Jiao, J., Yiu, S. M., Duan, M., & Chen, W. (2024). AnnoLLM: Making Large Language Models to be better crowdsourced annotators. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 6, pp. 165–190). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.naacl-industry.15>
- International Organization for Standardization. (2012). ISO 24617-1: Language resource management – Semantic annotation framework (SemAF) – Part 1: Time and events (SemAF-Time, ISO-TimeML). ISO. <https://www.iso.org/standard/37331.html>
- Kostina, A., Dikaiakos, M. D., Stefanidis, D., & Pallis, G. (2025). Large Language Models for text classification: Case study and comprehensive review. arXiv e-prints. <https://doi.org/10.48550/arxiv.2501.08457>
- Mann, W., & Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text – Interdisciplinary Journal for the Study of Discourse*, 8(3), 243–281. <https://doi.org/10.1515/text.1.1988.8.3.243>
- OpenAI. (2022). ChatGPT: Optimizing Language Models for dialogue. <https://openai.com/research/chatgpt>
- Pardo, T., Nunes, M., & Rino, L. (2004). DiZer: An automatic discourse analyzer for Brazilian Portuguese. In A.L.C. Bazzan & S. Labidi (Eds.), *Advances in Artificial Intelligence* (SBIA 2004) (pp. 224–234). Springer. [https://doi.org/10.1007/978-3-540-28645-5\\_23](https://doi.org/10.1007/978-3-540-28645-5_23)
- Pareti, S. (2009). *Towards a discourse resource for Italian: Developing an annotation schema for attribution* [Dissertação de mestrado, Università degli Studi di Pavia].
- Pareti, S. (2015). *Attribution: A computational approach* [Tese de doutoramento, University of Edinburgh]. <http://hdl.handle.net/1842/14170>
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation* (LREC 2008) (pp. 2961–2968). European Language Resources Association (ELRA).

- Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., Katz, G., & Radev, D. (2003). TimeML: Robust specification of event and temporal expressions in text. In *AAAI Technical Report SS-03-07*. American Association for Artificial Intelligence.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners*. OpenAI. [https://cdn.openai.com/better-language-models/language\\_models\\_are\\_unsupervised\\_multitask\\_learners.pdf](https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf)
- Reis, S., & Baptista, J. (2025). *Evaluation dataset for a custom GPT for the annotation of Lexicalized Meaning Representation (LMR) of European Portuguese* [Data set]. ResearchGate. <https://doi.org/10.13140/RG.2.2.22281.81769>
- Ren, M. (2024). Advancements and applications of Large Language Models in Natural Language Processing: A comprehensive review. *Applied and Computational Engineering*, 97(1), 55–63. <https://doi.org/10.54254/2755-2721/97/20241406>
- Sarmiento, L., & Nunes, S. (2009). Automatic extraction of quotes and topics from news feeds. In E. Oliveira & A. A. Sousa (Coords.), *4th Doctoral Symposium in Informatics Engineering (DSIE'09): Proceedings* (pp. 1–12). Faculdade de Engenharia da Universidade do Porto. <https://doi.org/10.24840/978-972-752-267-5>
- Savelka, J., & Ashley, K. D. (2023). The unreasonable effectiveness of large language models in zero-shot semantic annotation of legal texts. *Frontiers in Artificial Intelligence*, 6, 1279794. <https://doi.org/10.3389/frai.2023.1279794>
- Silvano, P., Amorim, E., Leal, A., Cantante, I., Silva, F., Jorge, A., Campos, R., & Nunes, S. (2023). Annotation and visualization of reporting events in textual narratives. In R. Campos, A. Jorge, A. Jatowt, S. Bhatia, & M. Litvak (Eds.), *CEUR Workshop Proceedings* (Vol. 3370, pp. 47–62). CEUR-WS.
- Tao, C., Shen, T., Gao, S., Zhang, J., Li, Z., Hua, K., Hu, W., Tao, Z., & Ma, S. (2024). *LLMs are also effective embedding models: An in-depth overview*. arXiv. <https://doi.org/10.48550/arXiv.2412.12591>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)

## Anexo I - Instrução de base (Prompt)

Faz a anotação semântica da frase seguinte, segundo as diretivas definidas na configuração deste GPT.

[para as configurações de *prompting* B, C e D), o *prompt* continha ainda:]  
Abaixo encontra alguns exemplos anotados.  
[Exemplos anotados]

[Texto a anotar]

[resposta]

[Para cada resposta, era ainda pedido:]

Indica agora quais as diretivas que justificam cada parte da anotação.

## Anexo II - Configuração do Anotador Semântico

### DESCRIÇÃO

Realiza anotação semântica conforme as diretrizes fornecidas.

### INSTRUÇÕES

Instruções do GPT — Anotação Semântica

#### 1) Objetivo

Executar anotação semântica fiel e rigorosa, conforme as diretrizes e exemplos presentes nos ficheiros carregados na secção Conhecimento (por exemplo, o ficheiro .pdf principal e os documentos de frases anotadas). Não criar, omitir ou inferir informação que não esteja nesses documentos.

#### 2) Fontes permitidas

Únicas fontes autorizadas: todos os ficheiros carregados na secção Conhecimento deste GPT (Diretivas de anotação e O Principezinho com anotações LMR, etc.). Proibido recorrer a conhecimento geral, dados de treino, inferência, web browsing ou exemplos próprios.

#### 3) Política de evidência obrigatória

Aplica-se sempre que o utilizador pedir ao GPT que explique, cite ou justifique uma regra, diretiva ou categoria presente nos ficheiros do Conhecimento. Nesses casos, a resposta deve incluir três componentes obrigatórios:

1. CITAÇÃO (obrigatória) Inclui um bloco de 10 a 80 palavras literais, copiadas diretamente de um dos ficheiros do Conhecimento, delimitado pelos marcadores:

[CITAÇÃO INÍCIO]

...texto literal retirado do ficheiro...

[CITAÇÃO FIM]

2. LOCALIZAÇÃO (obrigatória) Indica o título, secção, cabeçalho ou outro marcador textual que identifique a origem da citação. Se o ficheiro mencionar número de página, inclui-o também.

3. EXPLICAÇÃO FIEL (obrigatória) Redige 1 a 3 frases que clarifiquem o conteúdo da citação, sem interpretar nem extrapolar além do texto literal. A explicação serve apenas para tornar a citação compreensível, mantendo o sentido original.

4. Falha de evidência: Se não for possível apresentar uma citação literal, responde exatamente: "Não tenho essa informação nos ficheiros fornecidos."

#### 4) Escopo e restrições

Não criar novas categorias, atributos ou exemplos. Não parafrasear quando for pedido para citar. Se o pedido implicar explicação, deve ser breve e fiel à citação. Não inferir regras ou significados não explícitos nos ficheiros do Conhecimento.

#### 5) Formatos de saída

a) Quando o pedido é “explica/cita a regra X”

TÍTULO DA REGRA: <copiar o cabeçalho literal>

[CITAÇÃO INÍCIO]

<10–80 palavras literais>

[CITAÇÃO FIM]

LOCALIZAÇÃO: <secção/cabeçalho literal; número de página apenas se constar>

BREVE EXPLICAÇÃO (opcional): 1–3 frases, sem extrapolar nem interpretar além do texto.

b) Quando o pedido é “anota este texto”

Verifica nos ficheiros do Conhecimento as categorias e exemplos aplicáveis. Se não houver diretriz aplicável, responde apenas com a frase de falha. Caso exista, devolve apenas a anotação, no formato indicado nas diretrizes, com indentação hierárquica. Exemplo de formato (modelo estrutural):

```
(ROOT :MAIN (c / ciente
              :VAUX (e / está)
              :ARG0 (p / Pedro)
              :ARG1 (g / gravidade
                    :ARG1 (s / situação))))
```

#### 6) Procedimento de recuperação

Tenta correspondência literal do termo ou cabeçalho (ex.: :direct-speech, 3.5.4). Se falhar, procura variações mínimas que realmente constem no ficheiro (ex.: directspeech). Se ainda falhar, responde com a frase de falha. Nunca inferir significados nem sugerir correspondências aproximadas.

#### 7) Política de negação

Sempre que não consigas citar literalmente, responde somente: “Não tenho essa informação nos ficheiros fornecidos.” Não acrescentar justificações, hipóteses, desculpas ou sugestões alternativas.

#### 8) Checklist de execução

- O pedido exige citação? Inclui CITAÇÃO e LOCALIZAÇÃO.
- A citação é literal (sem sinónimos)?
- A LOCALIZAÇÃO copia o cabeçalho/numeração exatamente como no ficheiro?
- Se não encontrar, devolve apenas a frase de falha.

#### - Observações finais

Usar exclusivamente o Conhecimento carregado. Seguir sempre as regras e formatos fornecidos. A prioridade máxima é a fidelidade ao texto original e à estrutura semântica definida nos documentos.

### Anexo III - Frases do conjunto de avaliação

(ID-001-003) *A professora universitária falava na conferência “Inteligência Artificial, Democracia e Eleições”, que decorreu esta quinta-feira na Faculdade de Direito da Universidade de Lisboa e teve como mote o 50.º aniversário da Comissão Nacional de Eleições (CNE).*

(ID-001-004) *Na conferência, Sofia Casimiro levantou a questão: “Deverá a UE aumentar ou aligeirar a regulação da IA?”, acrescentando ser necessário um “equilíbrio entre este tipo de tecnologia e os valores fundamentais da comunidade europeia”.*

(ID-001-005) *A professora salientou o atual contexto de “uma nova ordem mundial, em que a democracia está em risco” e onde a IA pode ser uma “ferramenta poderosa para a projeção do poder e para o seu exercício”.*

(ID-001-006) *“As eleições que não têm interferência de IA são a exceção”, garantiu Sofia Casimiro.*

(ID-002-004) *“Fizemos isto em fevereiro com três perguntas a quatro sistemas de IA amplamente utilizados ao nível mundial, e que são as perguntas mais frequentes e as questões mais reclamadas pelos consumidores”, precisou o coordenador do departamento jurídico e económico da Deco, Paulo Fonseca, em declarações à Lusa.*

## Anexo IV - Frases de exemplo (few- / many-shots)

### 1. O Principezinho

#### *few-shot out-of-domain*

1. Frase: PPZ [ID=1249,1250] Respon-di: «Pois», e fitei, em silêncio, as ondas da areia ao luar.
2. Frase: PPZ [ID=1460a] Então hás de dizer-lhes: «Sim, as estrelas fazem-me sempre rir!»
3. Frase: PPZ [ID=626] — O que significa «admirar»?
4. Frase: PPZ [ID=152] Tenho sérias razões para acreditar que o planeta de onde vinha o principezinho era o asteroide B612.
5. Frase: PPZ [ID=304] Como o parafuso me deixara irritado, respon-di com um disparate: — Os espinhos não servem para nada.

#### *many-shot out-of-domain*

1. Frase: PPZ [ID=1249,1250] Respon-di: «Pois», e fitei, em silêncio, as ondas da areia ao luar.
2. Frase: PPZ [ID=1460a] Então hás de dizer-lhes: «Sim, as estrelas fazem-me sempre rir!»
3. Frase: PPZ [ID=626] — O que significa «admirar»?
4. Frase: PPZ [ID=152] Tenho sérias razões para acreditar que o planeta de onde vinha o principezinho era o asteroide B612.
5. Frase: PPZ [ID=304] Como o parafuso me deixara irritado, respon-di com um disparate: — Os espinhos não servem para nada.
6. Frase: PPZ [ID=307] Após um momento de silêncio, ele retorquiu, com uma espécie de rancor: — Não acredito em ti!
7. Frase: PPZ [ID=313] Naquele instante estava a pensar: Se este parafuso continuar a resistir, hei de tirá-lo à martelada.
8. Frase: PPZ [Frase: PPZ [ID=324] Mas, duro, ele acrescentou: — Estás a confundir tudo... a misturar tudo!
9. Frase: PPZ [ID=321] Ali estava eu, de martelo na mão e com os dedos sujos de óleo, debruçado sobre um objeto que lhe parecia muito feio.
10. Frase: PPZ [ID=342] E se eu conhecer uma flor única em todo o mundo, que não existe em mais lado nenhum além do meu planeta, e que uma ovelhinha pode devorar de uma só vez, numa bela manhã, sem se aperceber do que está a fazer?

### 2. Corpus de notícias sobre IA (few-shot in-domain)

1. Frase: ID-001-001 Numa época em que a "democracia está em risco", a inteligência artificial pode tornar-se numa "ferramenta poderosa para a projecção do poder e para o seu exercício", defende Sofia Casimiro.
2. Frase: ID-001-002 Os excessos de regulamentação da União Europeia (UE) têm bloqueado o desenvolvimento de ferramentas de inteligência artificial (IA) e aumentado a dependência da Europa de terceiros, defendeu esta quinta-feira a professora universitária, Sofia Casimiro.
3. Frase: ID-001-007 A professora defendeu a necessidade de estímulos à inovação e de uma estratégia de sensibilização dos utilizadores comuns, através da literacia mediática.

4. Frase: ID-001-008 A diretora de informação da agência Lusa, Luísa Meireles, também participou no evento e destacou que “os jornalistas estão na frente de combate da IA”.

5. Frase: ID-001-009 Para esta responsável, vive-se atualmente uma dualidade entre “democracia digital Vs. democracia analógica”.

A representação LMR de referência destas frases encontra-se disponível em: <https://doi.org/10.13140/RG.2.2.22281.81769>