



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS QUIXADÁ**  
**CURSO DE GRADUAÇÃO EM ENGENHARIA DE SOFTWARE**

**JOÃO PEDRO PEREIRA HOLANDA**

**LINGUAGEM SIMPLES COM LLMS: AVALIAÇÃO DE SIMPLIFICAÇÃO TEXTUAL  
NO CONTEXTO ADMINISTRATIVO DA UFC INOVA**

**QUIXADÁ**  
**2026**

JOÃO PEDRO PEREIRA HOLANDA

LINGUAGEM SIMPLES COM LLMS: AVALIAÇÃO DE SIMPLIFICAÇÃO TEXTUAL NO  
CONTEXTO ADMINISTRATIVO DA UFC INOVA

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Engenharia de Software  
do Campus Quixadá da Universidade Federal  
do Ceará, como requisito parcial à obtenção do  
grau de bacharel em Engenharia de Software.

Orientador: Prof Dr. Regis Pires Magalhães.

Coorientador: Prof Dr. Camilo Camilo  
Almendra.

QUIXADÁ

2026

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

H669l Holanda, João Pedro Pereira.  
Linguagem simples com llms : avaliação de simplificação textual no contexto administrativo da UFC Inova / João Pedro Pereira Holanda. – 2026.  
86 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Engenharia de Software, Quixadá, 2026.  
Orientação: Prof. Dr. Regis Pires Magalhães.  
Coorientação: Prof. Dr. Camilo Camilo Almendra.

1. Linguagem simples. 2. Grandes modelos de linguagem. 3. Simplificação textual. 4. Sumarização textual. 5. Complexidade textual. I. Título.

CDD 005.1

---

JOÃO PEDRO PEREIRA HOLANDA

LINGUAGEM SIMPLES COM LLMS: AVALIAÇÃO DE SIMPLIFICAÇÃO TEXTUAL NO  
CONTEXTO ADMINISTRATIVO DA UFC INOVA

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Engenharia de Software  
do Campus Quixadá da Universidade Federal  
do Ceará, como requisito parcial à obtenção do  
grau de bacharel em Engenharia de Software.

Aprovada em: 23/01/2026.

BANCA EXAMINADORA

---

Prof Dr. Regis Pires Magalhães (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof Dr. Camilo Camilo Almendra (Coorientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Luís Gustavo Coutinho do Rêgo  
Instituto Federal de Educação, Ciência e Tecnologia  
do Ceará (IFCE)

À minha família, por toda a confiança depositada em mim durante toda a graduação.

## **AGRADECIMENTOS**

Agradeço a meus pais, irmã e amigos por todo o apoio prestado durante toda a minha graduação.

Aos professores Dr. Regis Pires Magalhães e Dr. Camilo Camilo Almendra por compartilharem sua empolgação e conhecimento durante todo o desenvolvimento do projeto.

Ao professor Dr. Luis Gustavo Coutinho do Rêgo pela participação na banca e por todas as valiosas sugestões.

À equipe da Igeos por todo o incentivo dado ao meu desenvolvimento acadêmico e como pessoa.

À PROINTER pelo acesso a uma máquina que permitiu a execução dos experimentos com uma quantidade satisfatória de modelos.

"A pen can be sharper than a scalpel to a patient.  
To an open heart, it's a daily operation"  
(ANNEN, SHINGO 2013)

## RESUMO

A Linguagem Simples é um conjunto de práticas de escrita que busca facilitar a leitura de documentos públicos para transmitir informações importantes mais rapidamente, além de promover a acessibilidade e transparência. No ano de 2025 o Governo Brasileiro sancionou a Lei nº 15.263, que institui a Política Nacional de Linguagem Simples. A UFC Inova já publica versões em linguagem simplificada de parte dos seus editais desde 2021, mas a escrita deles é realizada manualmente, exigindo bastante tempo e esforço extra de seus colaboradores. A partir desses pressupostos, este trabalho discorre sobre a utilização de Inteligência Artificial Generativa (IAG) para realizar a simplificação textual, comparando o desempenho de Grandes Modelos de Linguagem (LLMs) proprietários e abertos. A avaliação utilizou índices de inteligibilidade consolidados, métricas morfossintáticas relacionadas à complexidade textual e métricas de similaridade semântica dos textos produzidos. Foram realizadas duas iterações do processo de simplificação textual, na primeira avaliação os modelos proprietários e abertos com melhores resultados nos índices e métricas linguísticas foram o Gemini 2.5 Flash e o Qwen2.5:14b, respectivamente, mas o Gemma3:4b e Gemini 2.5 Pro foram os que obtiveram maior similaridade semântica em relação à versão completa original, com BERTScore de 0,826 e 0,760, respectivamente. Após a avaliação pelo público-alvo em um questionário *online*, o processo foi adaptado para realizar a sumarização e simplificação dos documentos conjuntamente e novos textos foram gerados, os resultados dessa segunda iteração mostram a eficiência da sumarização combinada com simplificação, com 7 dos 9 modelos obtendo melhorias significativas no tamanho médio das sentenças e o Qwen3:14b tendo performance comparável com o Gemini 2.5 Pro. Os textos gerados foram comparados com os documentos simplificados já existentes e os melhores resultados da similaridade semântica foram, novamente, o Gemini 2.5 Pro e o Gemma3:4b, ambos com BERTScore de 0,665.

**Palavras-chave:** linguagem simples; complexidade textual; simplificação textual; sumarização textual; análise morfossintática; context embeddings; grandes modelos de linguagem.

## ABSTRACT

Plain Language consists of a set of practices to write easier to read public documents, aiming to expose relevant information quicker, and also promote accessibility and transparency. In 2025, the Brazilian government sanctioned the law n° 15.263, which establishes the National Policy of Plain Language. The UFC Inova institution already publishes plain language versions of some of its documents since 2021, but they are written manually, requiring extra time and work of the collaborators. Using that background, this work discusses the use of Generative Artificial Intelligence (GenAI) to perform text simplification, comparing the performance of proprietary and open Large Language Models (LLMs). The evaluation used consolidated readability indexes, morphosyntactic metrics related to text complexity and semantic similarity metrics on the generated texts. Two iterations of text simplification were performed, in the first one, the best proprietary and open models were Gemini 2.5 Flash and Qwen2.5:14b, respectively, but Gemma3:4b and Gemini 2.5 Pro had the best semantic similarity, with BERTScores of 0.826 and 0.760. After the audience evaluation in an online survey, the process was adapted to perform summarization and simplification jointly. The results of this second iteration show the efficiency of summarization combined with simplification, having 7 out of 9 models with significant improvements on the average sentence size and Qwen3:14b having a performance on par with Gemini 2.5 Pro's. Generated texts were compared with the existing simplified version and the models with the best results were, again, Gemini 2.5 Pro and Gemma3:4b, both with a 0.665 BERTScore, the similarity scores were lower in general, sitting around 0.622 and 0.684.

**Keywords:** plain language; text complexity; text simplification; text summarization; morphosyntactical analysis; context embeddings; large language models.

## LISTA DE FIGURAS

Figura 1 – Exemplo de tokenização do modelo Deepseek-R1 com cada <i>token</i> sendo representado por diferentes cores. . . . .	21
Figura 2 – Grafo de dependências sintáticas da frase ‘Uma andorinha africana ou europeia?’. . . . .	23
Figura 3 – Tarefas de processamento de linguagem natural relacionadas à complexidade textual. . . . .	31
Figura 4 – Exemplo de rede neural direta . . . . .	36
Figura 5 – Diagrama da arquitetura <i>Transformer</i> . Em violeta está o codificador e em verde o decodificador . . . . .	38
Figura 6 – Representação do processo metodológico. . . . .	44
Figura 7 – Fluxo de extração das métricas selecionadas. . . . .	47
Figura 8 – Fluxograma do processo de simplificação textual. . . . .	52
Figura 9 – Fluxograma do processo de simplificação textual aprimorado. . . . .	64

## LISTA DE TABELAS

Tabela 1	– Quantidade de documentos gerados com alucinações por modelo. . . . .	53
Tabela 2	– Valor médio e intervalo de confiança de 95% para os índices de inteligibilidade de todos os documentos em cada modelo . . . . .	54
Tabela 3	– Comparação da média e intervalo de confiança de 95% das métricas morfosintáticas para todos os documentos em cada modelo . . . . .	56
Tabela 4	– Acurácia, revocação, precisão e <i>score</i> F1 na métrica <i>foreign word ratio</i> por documento. . . . .	57
Tabela 5	– Valores médios das métricas de similaridade entre todos os documentos para cada modelo, ordenados pelo <i>BERTScore</i> . . . . .	58
Tabela 6	– Perfis dos participantes do questionário <i>online</i> . . . . .	58
Tabela 7	– Valores médio dos índices de inteligibilidade em todos os documentos da segunda iteração com intervalo de confiança de 95% . . . . .	65
Tabela 8	– Valores médios para todos os documentos da segunda iteração nas métricas morfossintáticas com intervalo de confiança de 95% . . . . .	67
Tabela 9	– Precisão, Revocação, Acurácia e <i>Score</i> F1 da métrica <i>foreign word ratio</i> para o segundo conjunto de dados considerado, . . . . .	68
Tabela 10	– Média e intervalo de confiança de 95% das métricas de similaridade para todos os documentos em cada modelo . . . . .	68

## LISTA DE QUADROS

Quadro 1 – Métricas morfossintáticas selecionadas e suas respectivas categorias. . . . .	35
Quadro 2 – Características de cada trabalho relacionado . . . . .	43
Quadro 3 – Documentos onde a anotação sintática com o PortParser V2 falhou . . . . .	64

## LISTA DE CÓDIGOS-FONTE

Código-fonte 1	– Simplificação com a maior similaridade semântica – Edital N° 12/2021 - Hackathon Inovando UFC – Texto simplificado original . . . . .	80
Código-fonte 2	– Simplificação com a maior similaridade semântica – Edital N° 12/2021 - Hackathon Inovando UFC – Gerada pelo Qwen3:14b . . . . .	81
Código-fonte 3	– Simplificação com o melhor resultado na métrica <i>long sentence ratio</i> – Edital Prointer PIBI N° 07/2023 – Texto Original . . . . .	82
Código-fonte 4	– Simplificação com o melhor resultado na métrica <i>long sentence ratio</i> – Edital Prointer PIBI N° 07/2023 – Gerada pelo Qwen3:14b . . . . .	82
Código-fonte 5	– Simplificação com alucinações – Edital Prointer PIBI N° 07/2023 – Texto original . . . . .	83
Código-fonte 6	– Simplificação com alucinações – Edital – Gerada pelo Lllama-3.2- instruct:3b . . . . .	84



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>16</b>
<b>1.1</b>	<b>Objetivos</b>	<b>17</b>
<b>1.1.1</b>	<i>Objetivos Específicos</i>	<b>18</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>19</b>
<b>2.1</b>	<b>Linguagem Simples</b>	<b>19</b>
<b>2.2</b>	<b>Processamento de Linguagem Natural</b>	<b>20</b>
<b>2.2.1</b>	<i>Terminologia</i>	<b>21</b>
<b>2.2.2</b>	<i>Interpretação de Linguagem Natural</i>	<b>22</b>
<b>2.2.2.1</b>	<i>Parte do Discurso (POS) e Relações de Dependência</i>	<b>23</b>
<b>2.2.2.2</b>	<i>Semântica Distribucional e Embeddings</i>	<b>24</b>
<b>2.2.3</b>	<i>Geração de Linguagem Natural</i>	<b>25</b>
<b>2.2.3.1</b>	<i>Sumarização Textual</i>	<b>26</b>
<b>2.2.3.2</b>	<i>Simplificação Textual</i>	<b>27</b>
<b>2.3</b>	<b>Complexidade Textual</b>	<b>29</b>
<b>2.3.1</b>	<i>Avaliação de Complexidade</i>	<b>30</b>
<b>2.3.1.1</b>	<i>Índices de Inteligibilidade</i>	<b>31</b>
<b>2.3.1.2</b>	<i>Métricas Morfossintáticas</i>	<b>33</b>
<b>2.4</b>	<b>Grandes Modelos de Linguagem</b>	<b>35</b>
<b>2.4.1</b>	<i>Redes Neurais</i>	<b>36</b>
<b>2.4.2</b>	<i>Transformers</i>	<b>37</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>39</b>
<b>3.1</b>	<i>SimplifyMyText: An LLM-Based System for Inclusive Plain Language Text Simplification (Färber et al., 2025)</i>	<b>39</b>
<b>3.2</b>	<i>Evaluating GenAI for Simplifying Text for Education: Improving Accuracy and Consistency for Enhanced Readability (Day et al., 2025)</i>	<b>40</b>
<b>3.3</b>	<i>Is It Possible to Modify Text to a Target Readability Level? An Initial Investigation Using Zero-Shot Large Language Models (Farajidizaji et al., 2024)</i>	<b>40</b>
<b>3.4</b>	<i>Assessing AI Simplification of Medical Texts: Readability and Content Fidelity (Picton et al., 2025)</i>	<b>41</b>

3.5	Comparação entre os trabalhos . . . . .	42
4	<b>METODOLOGIA</b> . . . . .	44
4.1	<b>Identificação de Métricas de Complexidade Textual Relacionadas à Linguagem Simples.</b> . . . . .	44
4.2	<b>Extração de Documentos Abertos</b> . . . . .	45
4.3	<b>Geração de Documentos Simplificados com Diferentes Grandes Modelos de Linguagem (LLMs)</b> . . . . .	45
4.4	<b>Avaliação com Métricas Estabelecidas</b> . . . . .	46
4.4.1	<i>Avaliação de Similaridade Semântica das simplificações</i> . . . . .	46
4.5	<b>Questionário com Público-Alvo</b> . . . . .	47
4.6	<b>Aprimoramento da Geração de Documentos Simplificados com Base nas Avaliações Iniciais</b> . . . . .	48
4.7	<b>Análise Final dos Resultados</b> . . . . .	49
5	<b>RESULTADOS</b> . . . . .	50
5.1	<b>Identificação de Métricas de Complexidade Textual Relacionadas à Linguagem Simples</b> . . . . .	50
5.2	<b>Extração de Documentos</b> . . . . .	51
5.3	<b>Geração de Documentos Simplificados com Diferentes LLMs</b> . . . . .	52
5.4	<b>Avaliação com Métricas Estabelecidas</b> . . . . .	53
5.4.1	<i>Índices de Inteligibilidade</i> . . . . .	53
5.4.2	<i>Métricas Morfossintáticas</i> . . . . .	54
5.4.2.1	<i>Eficiência da métrica proposta - foreign word ratio</i> . . . . .	55
5.4.3	<i>Similaridade Semântica</i> . . . . .	56
5.5	<b>Questionário com Público-Alvo</b> . . . . .	57
5.5.1	<i>SRQ1 - Quão Fácil é a Compreensão do Documento Simplificado em Comparação com o Documento Original?</i> . . . . .	58
5.5.2	<i>SRQ2 - Qual a utilidade da versão gerada automaticamente para apoiar uma rápida compreensão das principais informações?</i> . . . . .	59
5.5.3	<i>SRQ3 - Você considera que o documento simplificado está pronto para ser publicado como complemento da versão original?</i> . . . . .	59
5.5.4	<i>SRQ4 - Que aspectos de ambas as versões simplificadas podem melhorar a compreensão por parte do seu público-alvo?</i> . . . . .	59

5.5.5	<i>SRQ5 - Você identificou algum risco ou limitação no conteúdo destas versões simplificadas?</i> . . . . .	60
5.6	<b>Aprimoramento da Geração de Textos</b> . . . . .	60
5.7	<b>Análise Final dos Resultados</b> . . . . .	61
5.7.1	<i>Extração dos Documentos</i> . . . . .	61
5.7.2	<i>Geração de Documentos Simplificados com Diferentes LLMs</i> . . . . .	62
5.7.3	<i>Avaliação com as métricas</i> . . . . .	63
5.7.3.1	<i>Índices de Inteligibilidade</i> . . . . .	65
5.7.3.2	<i>Métricas Morfossintáticas</i> . . . . .	66
5.7.3.2.1	Eficiência da métrica proposta - <i>foreign word ratio</i> . . . . .	66
5.7.3.3	<i>Similaridade Semântica</i> . . . . .	67
6	<b>CONCLUSÕES E TRABALHOS FUTUROS</b> . . . . .	69
	<b>REFERÊNCIAS</b> . . . . .	71
	<b>APÊNDICE A –PROMPTS UTILIZADOS PARA SIMPLIFICAÇÃO</b>	
	<b>TEXTUAL</b> . . . . .	77
A.1	<b>Primeiro Prompt - apenas simplificação</b> . . . . .	77
A.2	<b>Segundo Prompt - sumarização e simplificação</b> . . . . .	78
	<b>APÊNDICE B –EXEMPLOS DE SIMPLIFICAÇÕES REALIZADAS</b> .	79
	<b>APÊNDICE C –QUESTIONÁRIO SOBRE SIMPLIFICAÇÃO TEXTUAL</b>	85

## 1 INTRODUÇÃO

Entre as principais responsabilidades de órgãos governamentais está a necessidade de uma comunicação clara com a população, e um dos primeiros vetores de contato na maioria desses órgãos se dá por documentos disponibilizados para consulta pública (editais, pareceres, entre outros). Um aspecto fundamental da comunicação é ser transparente e inteligível por **todo** o público-alvo para realmente cumprir sua função (Brasil, 2025a, Art. 5, inc. XXXIII).

Entretanto, no Brasil ainda existem muitos desafios a serem superados para garantir uma comunicação realmente acessível na esfera administrativa, principalmente pelo elevado índice de analfabetismo funcional no país. Em 2024, 7% da população era analfabeta e 22% tinha uma alfabetização de nível rudimentar (com 12% possuindo o ensino superior completo) (INDICADOR DE ALFABETISMO FUNCIONAL, 2024).

Uma das formas de tratar esse problema é simplificar a forma como transmitimos informações ao público geral. Uma abordagem de simplificação da escrita que vem ganhando força ao longo dos anos é o conceito de **linguagem simples**, ou clara (*plain language*).

Segundo a INTERNATIONAL PLAIN LANGUAGE FEDERATION (2019), “uma comunicação está em linguagem clara quando o texto, a estrutura e o design são tão claros que o público-alvo consegue encontrar facilmente o que procura, compreender o que encontrou e usar essa informação.”. Além de ser mais acessível (para pessoas que não tem o Português como primeiro idioma, por exemplo), uma escrita mais simples transmite as informações mais rapidamente e de formas menos ambígua, podendo reduzir a necessidade por suporte especializado, diminuir o tempo gasto em leitura pelo público, evitar erros em submissões de formulários, entre outras consequências positivas, como mostrado no relatório do programa de linguagem simples da Colômbia em Cuesta *et al.* (2019).

A UFC Inova é um órgão da Universidade Federal do Ceará (UFC) com o propósito de apoiar a inovação tecnológica, propriedade intelectual e o empreendedorismo na universidade <sup>1</sup>, que já realiza a publicação de parte de seus editais com versões em linguagem simplificada. Mas de acordo com relatos de servidores envolvidos nas ações da agência, ainda existem desafios para redigir versões simplificadas dos seus documentos publicados de forma prática, principalmente pelo tempo demandado em escrever uma versão resumida e uma versão completa de um mesmo documento. A proposta desse trabalho é explorar abordagens de automatização da sumarização e simplificação textual.

---

<sup>1</sup> <https://ufcinova.ufc.br/pt/nossa-historia/>

Um processo de sumarização e simplificação textual de forma automatizada envolve a geração de textos com técnicas de Processamento de Linguagem Natural (PLN) apoiadas no uso de Inteligência Artificial Generativa (IAG). A sumarização textual consiste em selecionar as partes mais importantes de um texto sem grandes mudanças na estrutura (extrativa) ou criar um novo texto que resuma as principais informações (abstrativa), já a simplificação textual é uma tarefa de reescrita de um texto para facilitar a interpretação, alterando a ordem das frases, substituindo palavras por sinônimos mais simples, removendo informações redundantes, etc. (Sun *et al.*, 2021; Giarelis *et al.*, 2023; Yadav *et al.*, 2021).

Um desafio da automatização desse processo é avaliar como a inteligibilidade, isto é, a facilidade que o leitor tem de extrair as informações apresentadas, melhora ao aplicar a linguagem simples em um documento, sendo necessário um conjunto de métricas que possam avaliar a inteligibilidade de um texto.

Além disso, ainda é necessário considerar as restrições no uso de IAG para a realização dessa tarefa: modelos proprietários acessíveis por meio de subscrições, como o ChatGPT, Google Gemini e Claude, possuem performance de estado da arte, mas geralmente incorrem em altos custos e impõem uma dependência ao fornecedor desse serviço. Os modelos abertos, que são disponibilizados ao público geral e podem ser utilizados em uma infraestrutura própria, são uma alternativa aos modelos proprietários, mas não possuem tanta capacidade de resolução de tarefas quanto eles.

## 1.1 Objetivos

O objetivo geral deste trabalho é avaliar simplificações textuais feitas com o uso de IAG em editais públicos da UFC Inova segundo três questões de pesquisa:

- QP1** A sumarização e simplificação textual realizadas com o uso de IAG realmente melhoram a inteligibilidade de um texto?
- QP2** Modelos abertos do Ollama conseguem realizar a simplificação textual com eficiência comparável a grandes modelos proprietários?
- QP3** Como a comunidade acadêmica da UFC avalia a qualidade da simplificação automática de documentos administrativos e quais riscos são identificados nessa atividade?

### 1.1.1 *Objetivos Específicos*

Tendo em vista o objetivo geral deste trabalho, elencam-se a seguir os seguintes objetivos específicos que permitem a realização do objetivo geral do trabalho.

- a) Estruturar um *dataset* com documentos existentes em versão completa e simplificada da UFC Inova.
- b) Realizar a geração textos simplificados de editais com o uso de diferentes LLMs abertos e proprietários.
- c) Avaliar a inteligibilidade dos dois formatos de texto utilizando índices de inteligibilidade consolidados e aspectos linguísticos.
- d) Medir a similaridade de versões simplificadas por IAG e os documentos originais.
- e) Avaliação com público-alvo da inteligibilidade dos textos originais e simplificados.

A organização deste trabalho se dá da seguinte forma: o Capítulo 2 detalha os conceitos que fundamentam esse trabalho, principalmente o que são PLN, sumarização e simplificação textual, representação semântica com *Embeddings*, e como a complexidade de um texto é calculada sob diferentes métricas; o Capítulo 3 compara as soluções existentes para a simplificação textual com o uso de LLMs; o Capítulo 4 trata da extração de dados, do processo de simplificação realizado, como cada métrica de complexidade foi medida e como foram realizadas as comparações entre documentos completos e simplificados; O Capítulo 5 mostra os resultados das avaliações quantitativas e qualitativas realizadas; por fim, o Capítulo 6 traz as últimas considerações sobre as contribuições deste trabalho e define os trabalhos futuros a serem realizados.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Linguagem Simples

O principal objetivo da linguagem simples é a garantia de acessibilidade na leitura por parte do público-alvo. A principal entidade reguladora da linguagem simples é a *International Plain Language Federation* que é composta pela *Plain Language Association International* (PLAIN), *Clarity* e *Center for Plain Language*<sup>1</sup>. Entre as principais diretrizes para acessibilidade dos textos estão as descritas por INTERNATIONAL PLAIN LANGUAGE FEDERATION (2019):

- a) redigir frases em ordem direta, Sujeito-Verbo-Objeto (SVO);
- b) redigir frases curtas;
- c) desenvolver uma ideia por parágrafo;
- d) usar palavras comuns, de fácil compreensão;
- e) usar sinônimos de termos técnicos e de jargões ou explicá-los no próprio texto;
- f) evitar palavras estrangeiras que não sejam de uso corrente;
- g) redigir o nome completo antes das siglas;
- h) organizar o texto de forma esquemática, onde for possível, com o uso de listas, de tabelas e de gráficos;

Atualmente, mais de 30 países na Europa e Américas do Sul e Norte possuem alguma forma de regulação na redação de documentos oficiais com inspiração nessas convenções, mas a formalização ainda é um processo em formação e cada país adapta as convenções existentes para seu cenário específico (Martins *et al.*, 2023). Como exemplo de padronização internacional, temos o padrão da *International Organization for Standardization* (ISO), ISO 24495-1:2023, que está atualmente em rascunho e visa centralizar a definição e implementação da linguagem simples (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2023).

Na realidade brasileira, os esforços na regulação da linguagem simples ainda são recentes, e dentre as principais ações promulgadas está a Lei nº 15.263 de 2025 (Brasil, 2025b), sancionada em 14 de novembro de 2025, que regulamenta o uso da linguagem simples em órgãos e entidades públicas; e o Pacto Nacional pela Linguagem simples (CONSELHO NACIONAL DE JUSTIÇA, 2023). Em cada uma dessas ações existem especificidades da área de atuação da instituição que impõem a necessidade de novas regras, como no caso do Conselho Nacional de

---

<sup>1</sup> <https://www.iplfederation.org/members/>

Justiça (CNJ), que possui regras específicas como:

- a) organizar o texto a fim de que as informações mais importantes apareçam primeiramente;
- b) não usar novas formas de flexão de gênero e de número das palavras da língua portuguesa, em contrariedade às regras gramaticais consolidadas, ao Vocabulário Ortográfico da Língua Portuguesa (Volp) e ao Acordo Ortográfico da Língua Portuguesa, promulgado pelo Decreto nº 6.583, de 29 de setembro de 2008 (Brasil, 2008).

Além de instituições jurídicas, muitas universidades hoje já dedicam esforços para publicarem seus documentos direcionados ao público geral também com versões simplificadas, como a UNIVERSIDADE ESTADUAL DE GOIÁS (2023), UNIVERSIDADE ESTADUAL DE CAMPINAS (2024), INSTITUTO FEDERAL DO PERNAMBUCO (2023), INSTITUTO FEDERAL DE MATO GROSSO (2021) e a UFC, que possui uma grande demanda por uma forma mais prática para simplificar os documentos administrativos em suas coordenadorias. A UFC Inova publica parte de seus editais em formato simplificado e completo desde 2021, mas a proporção de documentos com versão simplificada ainda é baixa: apenas 11 dos 98 editais publicados utilizam as práticas da linguagem simples.

## 2.2 Processamento de Linguagem Natural

O PLN é uma área da computação que se dedica a solucionar problemas que envolvem processar a linguagem humana escrita e falada com o uso de computadores. Divide-se em duas áreas principais: Interpretação de Linguagem Natural ou *Natural Language Understanding* (NLU), que diz respeito a análises e interpretações da linguagem (como a análise da sintaxe, análise do léxico e relações de dependência entre os termos de frases) realizadas pelos modelos computacionais, e a Geração de Linguagem Natural ou *Natural Language Generation* (NLG), utilizada comumente em aplicações de *chatbots*.

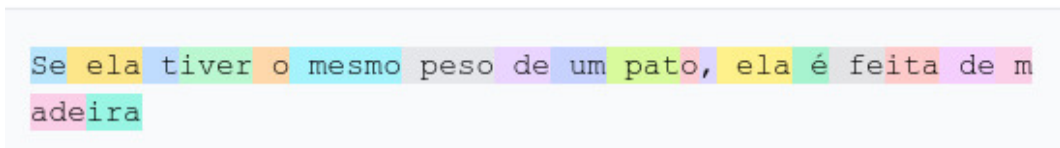
Pelo fato da NLU ser utilizada na análise do conteúdo da simplificação, e a NLG ser utilizada para criar esses documentos simplificados, as duas áreas estão presentes na fundamentação deste trabalho. A NLU é detalhada na Subseção 2.2.2 e a NLG é detalhada na Subseção 2.2.3. A Subseção 2.2.1 traz as definições dos principais termos relativos ao PLN que são utilizados nesse trabalho.

### 2.2.1 Terminologia

Nesta Seção detalham-se os principais termos linguísticos e específicos de PLN que serão utilizados no decorrer deste trabalho.

*tokens* são a unidade básica do PLN e correspondem a uma ocorrência específica de uma sequência de caracteres, incluindo pontuações e números em um texto. No Português essas sequências são normalmente uma única palavra ou pontuação e são separadas por espaços gráficos. Por exemplo, na frase “Ela tinha uma maçã e uma goiaba.” há oito *tokens* diferentes: { ‘Ela’, ‘tinha’, ‘uma’, ‘maçã’, ‘e’, ‘uma’, ‘goiaba’, ‘.’ }. Vale salientar que *tokens* podem ser maiores ou menores do que uma palavra a depender da implementação; modelos de linguagem mais recentes usam algoritmos de tokenização que não separam as palavras com base em espaços, mas sim em fragmentos que se repetem em todo o texto, para reduzir o tamanho dos vocabulários criados, como ilustrado na Figura 1 (Finatto *et al.*, 2024).

Figura 1 – Exemplo de tokenização do modelo Deepseek-R1 com cada *token* sendo representado por diferentes cores.



Fonte: Frase presente em MONTY Python (1975). Elaborado pelo autor no site [tiktokenizer.vercel.app](https://tiktokenizer.vercel.app)

O termo *type* é referente a um *token* único no texto, no exemplo da frase anterior, a única palavra que se repete é ‘uma’ e portanto ‘uma’ é o único *type* que possui dois *tokens*, enquanto os outros possuem apenas um *token* (Finatto *et al.*, 2024).

O conceito de **lema** representa a raiz morfológica de uma palavra, sua forma mais básica sem nenhuma flexão gramatical (Finatto *et al.*, 2024). Seguindo com a frase de exemplo, nós teríamos os seguintes lexemas: { ‘ele’, ‘ter’, ‘um’, ‘maçã’, ‘e’, ‘um’, ‘goiaba’, ‘.’ } . Pode-se observar que não são consideradas nenhuma informação de quantidade e tempo nos verbos e nem de gênero nos artigos, mas os pronomes, conjunções e substantivos já estão em sua forma mais básica nesse caso, por isso não se alteram.

A **paráfrase** é o processo de reescrita de uma sentença, onde o significado geral da sentença é mantido, mas com variações na estrutura gramatical e léxica, mas o nível de mudanças que geram uma paráfrase de um texto varia de acordo com cada autor. Com uma definição mais flexível entre linguistas (BARRÓN-CEDEÑO *et al.*, 2013; Vila *et al.*, 2014), e mais fixa entre cientistas da computação (Bhagat; Hovy, 2013). O Exemplo 1 traz duas versões de um

mesmo texto, em que a segundo é mais formal do que a primeira, mas ambas trazem o mesmo significado.

Exemplo de paráfrase: primeira versão

É apenas uma maquete.

(1)

Fonte: (MONTY Python, 1975)

Exemplo de paráfrase: segunda versão

Trata-se de uma representação em miniatura.

Fonte: Elaborado pelo autor.

Na Subseção 2.2.3.2 são mostrados mais exemplos de paráfrases específicos para o processo de simplificação textual.

**Anáfora** é uma referência a um termo de uma sentença anterior (**antecedente** ou **referente**) por meio de um ou mais termos **anafóricos**, cuja interpretação depende do significado do antecedente, A **Correferência** por sua vez, ocorre quando duas ou mais referências se referem a uma mesma entidade, por isso anáforas podem ou não realizar a correferência (Fonseca *et al.*, 2024). No Exemplo 2 tem-se uma anáfora que não realiza correferência, onde o fragmento sublinhado tem relação com a primeira entidade em negrito, mas se refere a uma outra loja que está implícita no contexto da frase.

Exemplo de anáfora sem correferência

**A loja de doces** na rua A vende os melhores chocolates, mas a da outra esquina faz os melhores bolos.

(2)

Fonte: Elaborado pelo autor.

## 2.2.2 Interpretação de Linguagem Natural

No âmbito da interpretação da linguagem natural, este trabalho faz uso tanto de conceitos da análise sintática do texto, incluindo as funções gramaticais dos termos e as dependências existentes entre eles, quanto da análise léxica, onde o significado das palavras e sentenças é interpretado por uma abordagem computacional. Na Subseção 2.2.2.1 são detalhadas os conceitos de análise sintática que são utilizados no cálculo das métricas morfossintáticas

e a Subseção 2.2.2.2 traz a fundamentação da semântica distribucional e representações por *embeddings*.

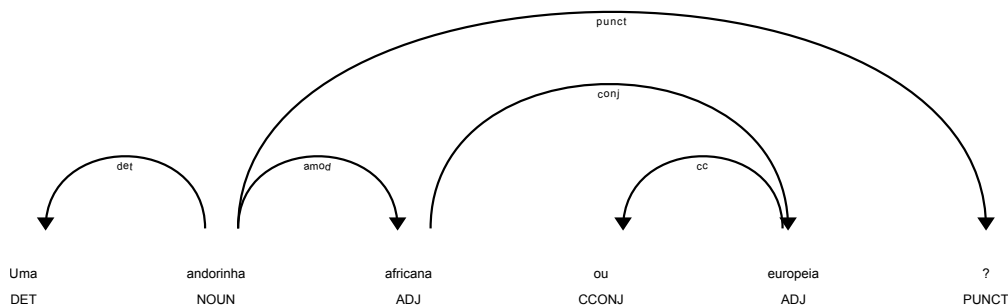
### 2.2.2.1 Parte do Discurso (POS) e Relações de Dependência

**Parte do Discurso - *Part of Speech* (POS)**, são as classes gramaticais atribuídas a cada *token* de uma frase, tais como substantivos, verbos, ou advérbios. As classes gramaticais existentes vão variar de idioma para idioma, mas existem formatos que universalizam essas classes entre múltiplos idiomas, como o formato UPOS do projeto *Universal Dependencies* (UD)<sup>2</sup> (DE MARNEFFE *et al.*, 2021; Finatto *et al.*, 2024).

**Dependências sintáticas** descrevem a relação de cada *token* em uma sentença entre eles, como um verbo que é auxiliar de outro, artigos que determinam o gênero de um substantivo, objetos de um verbo e etc. Essas relações de dependência são descritas na sua forma básica como conexões binárias direcionadas entre uma palavra base (*head*) e uma dependente, comumente apresentadas por uma seta que parte da base e chega à dependente. todas as relações em uma sentença formam um grafo direcionado (Jurafsky; Martin, 2026, cap. 19) :

Como exemplo de análise da POS e dependências sintáticas, incluí-se um exemplo de uma frase tirada de MONTY Python (1975) na Figura 2, abaixo de cada palavra está a sua classe gramatical UPOS e as setas definem os relacionamentos entre os termos no formato de *universal dependency relations*<sup>3</sup>.

Figura 2 – Grafo de dependências sintáticas da frase ‘Uma andorinha africana ou europeia?’.



Fonte: Elaborado pelo autor com a biblioteca Spacy (Honnibal *et al.*, 2020).

<sup>2</sup> <https://universaldependencies.org/u/pos/index.html>

<sup>3</sup> <https://universaldependencies.org/u/dep/index.html>

### 2.2.2.2 Semântica Distribucional e Embeddings

A análise semântica é uma área da linguística e NLU que, definindo de forma simples, busca entender os significados das palavras e frases. Ainda que seja uma definição simples, a própria noção do que é um significado diverge entre as perspectivas teóricas adotadas. Os estudos linguísticos nessa área dividem-se em duas correntes principais: a representacionista e a pragmática (Freitas, 2024).

Na corrente representacionista as palavras são consideradas manifestações de entidades do mundo real, o significado das palavras é determinado puramente pela entidade que as descreve e duas palavras são sinônimas quando estão associadas a uma mesma entidade. O processamento computacional baseado nessa corrente é realizado por técnicas simbólicas que envolvem relações complexas de hipo e hiperonímia, sinonímia, antonímia, etc. (no caso de *Wordnets*. Ou conexões lógicas dos elementos textuais a papéis semânticos e afirmações estruturadas em bases de conhecimento (Freitas, 2024; Seno *et al.*, 2024b).

Já na corrente pragmática, o significado da palavra está relacionado ao seu uso em situações concretas, e como essas situações variam no tempo, espaço, participantes e muitos outros aspectos, o significado de uma palavra sempre vai depender do contexto. A semântica distribucional é, atualmente, a forma mais utilizada de representação lexical em PLN e se enquadra na corrente pragmática Freitas (2024), Seno *et al.* (2024a).

A semântica distribucional tem como base a hipótese distribucional, que, segundo Freitas (2024), Seno *et al.* (2024a) e Jurafsky e Martin (2026, cap. 5), aponta para a tendência de palavras com significado similar ocorrerem em contextos semelhantes. A representação computacional das palavras é feita utilizando vetores multidimensionais, também chamados de *embeddings*, em que cada palavra passa a ser um ponto no espaço vetorial e palavras semelhantes são posicionadas próximas. Em *embeddings* estáticos esses vetores são construídos com matrizes de coocorrência. Essas matrizes de coocorrência contabilizam as ocorrências de cada palavra em um contexto escolhido, normalmente o contexto é um documento do *corpus* ou uma janela de caracteres próximos. Após a criação dessa matriz, os *embeddings* só serão atualizados caso essa matriz seja recalculada (Jurafsky; Martin, 2026, cap. 5).

Atualmente, os principais LLMs utilizam *embeddings* dinâmicos ou contextualizados, onde, além da representação vetorial do texto, os *tokens* também possuem um vetor que codifica a posição dele, o que é descrito juntamente com a arquitetura *Transformer* na Subseção 2.4.2. Essa representação é utilizada juntamente com os mecanismos de atenção para representar palavras de

forma diferente a depender do contexto delas no texto, fazendo com que duas palavras escritas da mesma forma sejam representadas de forma diferente a depender do contexto que cerca elas Paes *et al.* (2024).

Essa sensibilidade ao contexto torna a corrente pragmática mais apropriada para análises semânticas que não requerem resoluções de hierarquias, anotação de papéis semânticos ou resolução de ambiguidades, além de poder medir a similaridade de dois textos de forma concisa e facilmente computável: a **distância do cosseno** é uma fórmula vetorial que permite avaliar a similaridade de vetores multidimensionais com base no cosseno do ângulo formado entre eles, podendo ir de -1 em vetores totalmente opostos, até 1 em vetores que são congruentes (Seno *et al.*, 2024a).

Uma das métricas mais relevantes para avaliar a similaridade semântica em PLN é o *BERTScore*, descrito em Zhang *et al.* (2020). Voltada para comparar uma sentença de referência, com outra, chamada de candidata. A comparação é feita de forma gulosa, utilizando *embeddings* contextuais das duas sentenças e a medida de similaridade do cosseno ( $\frac{x_i^\top \hat{x}_j}{\|x_i\| \|\hat{x}_j\|}$ ), simplificada para  $x_i^\top \hat{x}_j$  pelo fato dos vetores serem normalizados.

O *BERTScore* consiste de três sub-métricas:  $R_{BERT}$ ,  $P_{BERT}$  e  $F_{BERT}$ . Representando, respectivamente, revocação, precisão e a média harmônica entre as duas anteriores.

A precisão é a média da similaridade do cosseno de cada *token* da sentença de referência com o *token* mais similar na sentença candidata, e a revocação avalia a similaridade no sentido contrário, casando cada *token* da sentença candidata com o mais próximo na sentença de referência. A média harmônica é utilizada para uma representação mais balanceada das duas medidas em uma única métrica.

As fórmulas para as métricas  $P_{BERT}$ ,  $R_{BERT}$  e  $F_{BERT}$  estão presentes na Equação 2.3, onde  $x$  são os vetores de *embedding* da sentença de referência e  $\hat{x}$  os da sentença candidata.

$$P_{BERT} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max_{x_i \in x} x_i^\top \hat{x}_j \quad R_{BERT} = \frac{1}{|x|} \sum_{x_i \in x} \max_{\hat{x}_j \in \hat{x}} x_i^\top \hat{x}_j \quad F_{BERT} = 2 \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}} \quad (2.3)$$

### 2.2.3 Geração de Linguagem Natural

Uma das atividades mais importantes dentro do PLN é a geração textual, ou NLG, que consiste em criar conteúdos textuais ou de áudio em linguagem natural a partir de dados

linguísticos, imagens e outras formas de comunicação (Ferreira; Paraboni, 2024). A principal atividade de NLG realizada neste trabalho é a simplificação textual, descrita na Subseção 2.2.3.2.

### 2.2.3.1 Sumarização Textual

A sumarização textual consiste na atividade de representar as informações mais relevantes de um texto original em um novo texto mais curto que condensa as informações do original. A sumarização pode ser feita com três abordagens distintas: a extrativa, onde o novo texto é constituído apenas de partes selecionadas do texto original, a abstrativa, onde o novo texto pode conter conteúdos que não estavam presentes no texto original e a híbrida, que combina as duas abordagens (Souza *et al.*, 2024; Zhang *et al.*, 2025).

Inicialmente, as abordagens para sumarização textual envolviam o reconhecimento de padrões linguísticos, como a *Term Frequency-inverse document frequency* (TF-IDF), para encontrar informações relevantes para a versão sumarizada. A arquitetura de *Transformers*, explicada na Subseção 2.4.2 possibilitou a uma grande melhoria do processo de sumarização, favorecendo a sumarização abstrativa (Zhang *et al.*, 2025).

As aplicações da sumarização textual são muitas em diferentes áreas, como na criação de chamadas de notícias, sinopses de filmes, etc. As principais vantagens de aplicar a sumarização de forma automática incluem: o acesso mais rápido às informações de um documento, maior interesse por parte do público, redução de vieses quando comparada a redatores humanos e a maior velocidade para gerar conteúdo quando comparada à sumarização manual (Sharma *et al.*, 2025).

As métricas de avaliação da sumarização automática incluem as avaliações por juízes humanos, métricas baseadas em similaridade (como o BERTScore, detalhado em 2.2.2.2), métricas de consistência factual e métricas de coerência e redundância (Zhang *et al.*, 2025).

As métricas baseadas em similaridade buscam avaliar como o sumário gerado se compara com um sumário de referência, um exemplo de métrica consolidada é a ROUGE (Lin; Hovy, 2003), que compara a quantidade de n-gramas do texto (normalmente n é 1 ou 2) de referência que está no texto avaliado. A desvantagem dessa métrica é que sumarizações que utilizam palavras diferentes são penalizadas, mesmo que o significado geral seja o mesmo, algo que não ocorre em métricas como o BERTScore.

Neste trabalho realiza-se a sumarização textual de forma abstrativa em documentos individuais por meio de uma abordagem dividir para conquistar, onde a sumarização é realizada

em múltiplas seções do documentos separadamente e os resultados são combinados depois, ela é descrita em maiores detalhes na Subseção 5.7.2.

### 2.2.3.2 *Simplificação Textual*

A simplificação textual é um tipo específico de tarefa de geração de textos (*text2text-generation*), com o foco específico em reescrever o texto de uma forma que as sentenças sejam mais fáceis de serem entendidas pelos leitores (Leal; Aluísio, 2024). Sun *et al.* (2021) descreve as principais diferenças dessa tarefa para a sumarização de textos:

- a) **Remoção de sentenças:** para realização da simplificação textual podem ser realizadas remoções de sentenças não relevantes. Enquanto que, segundo Giarelis *et al.* (2023), na sumarização extrativa apenas sentenças mais relevantes são selecionadas sem alterações, além disso na sumarização abstrativa o texto original será utilizado para gerar um novo texto que apresenta apenas as informações mais relevantes e não segue o mesmo formato do original.
- b) **Reordenação de sentenças:** outra operação realizada apenas na simplificação, onde as sentenças podem mudar sua posição no texto para ficarem próximas a outras sentenças que possuam alguma relação. Na sumarização, a ordem das informações também pode mudar, com o conteúdo mais relevante podendo ser apresentado primeiro, mas, como dito no primeiro ponto, esse texto sumarizado não se relaciona com o original.
- c) **Reescrita das sentenças:** outras operações que podem ocorrer apenas na simplificação são a concatenação de sentenças curtas e a reescrita de sentenças complexas, em um processo de paráfrase. Enquanto que na sumarização não ocorrem mudanças no texto original.

Existem ainda diferentes tipos de simplificação que podem ser realizadas em um texto, incluindo a **simplificação lexical**, **simplificação sintática** e a **elaboração textual** (Leal; Aluísio, 2024).

A simplificação lexical acontece quando palavras complexas são substituídas por outras, como no caso de Swanson *et al.* (2024), em que termos médicos complexos eram substituídos por hiperônimos mais simples, como no Exemplo 4.

#### Exemplo de simplificação lexical

O paciente apresentou um quadro de hipertermia  
→ O paciente estava com febre.

(4)

Fonte: Elaborado pelo autor.

A simplificação sintática é um processo que envolve reescrever orações complexas e quebrar orações longas em outras menores, como no Exemplo 5 em Leal e Aluísio (2024), onde o segundo texto é uma versão simplificada do primeiro, mas que possui uma estrutura diferente e um texto maior, mostrando que a simplificação não está sempre relacionada a tornar um texto menor.

#### Exemplo de simplificação sintática: texto original

O uso de forragem conservada, cujas formas mais comuns são: ensilagem e fenação, é uma solução para alimentar o rebanho.

(5)

#### Exemplo de simplificação sintática: texto simplificado

O uso de forragem conservada é uma solução para alimentar o rebanho. As formas mais comuns para conservar forragens são: ensilagem e fenação.

Fonte: (Leal; Aluísio, 2024)

Finalmente, a elaboração textual consiste em adicionar mais conteúdos para descrever determinadas palavras, com o intuito de esclarecer termos complexos que não possuem substitutos adequados em algum contexto ou para ampliar o vocabulário do leitor (Leal; Aluísio, 2024). No Exemplo 6, a palavra “enantiomorfa” é seguida por uma definição mais aprofundada dentro do próprio texto na sua versão elaborada.

#### Exemplo de elaboração textual: texto original

O reflexo que vemos no espelho é uma imagem enantiomorfa.

(6)

#### Exemplo de elaboração textual: texto com elaboração

O reflexo que vemos no espelho é uma imagem enantiomorfa, isto é, que não pode ser sobreposta com o que está sendo refletido.

Fonte: (Asth, *[s.d.]*), com reestruturação do autor

A simplificação textual pode variar ainda no objeto base de sua simplificação, podendo ocorrer apenas a nível de sentença (Xu *et al.*, 2016), em parágrafos de texto (Farajidizaji *et al.*, 2024) ou em documentos inteiros (Sun *et al.*, 2021). Este trabalho foca na simplificação textual a nível de documento.

O Exemplo 7 mostra uma simplificação textual seguindo princípios da Linguagem Simples, onde uma frase foi reescrita para facilitar a interpretação.

Exemplo de aplicação da Linguagem Simples: texto original

É preciso entregar uma manifestação escrita a próprio punho declarando seu endereço de residência domiciliar.

(7)

Exemplo de aplicação da Linguagem Simples: texto simplificado

Você pode entregar uma declaração de residência escrita à mão.

Fonte: (UNIVERSIDADE ESTADUAL DE CAMPINAS, 2024)

## 2.3 Complexidade Textual

O estudo da complexidade textual é uma área do PLN que investiga a forma como nós, humanos, percebemos a linguagem escrita e o quão difícil é interpretar um texto com base em aspectos linguísticos como o léxico e a sintaxe, ou seja, quão **inteligível** um texto é. A inteligibilidade pode ser definida como a facilidade de entender um texto, o que é diferente da legibilidade, que é a facilidade de ler o conteúdo e está relacionada com o tamanho da fonte, estruturação dos elementos, etc. (Leal; Aluísio, 2024).

Um aspecto fundamental da complexidade textual é o fato de que diferentes pessoas interpretam textos de forma singular, com base não só em seu conhecimento de mundo, educação formal e área de atuação profissional, mas também em fatores psicológicos, como o seu interesse no texto. Por isso a complexidade de um texto depende não só de fatores linguísticos, mas também da sua adequação ao público-alvo (Leal *et al.*, 2024; Leal; Aluísio, 2024).

Entre os principais fatores linguísticos que, segundo Leal *et al.* (2024), afetam a complexidade textual, estão:

- a) Tamanho geral das sentenças, quanto mais longa mais difícil tende a ser a leitura.

- b) Quantidade de palavras anteriores ao verbo, quanto maior, mais carga cognitiva é necessária para interpretar uma sentença.
- c) Uso da voz passiva aumenta a complexidade quando comparado com a voz ativa.
- d) Frases fora da ordem básica das palavras (no português, SVO) são mais difíceis de serem interpretadas.
- e) O uso de palavras incomuns ao leitor, como jargões e estrangeirismos, prejudica o entendimento geral do texto.
- f) O excesso de referentes possíveis em anáforas dificulta a leitura por que entidades mencionadas anteriormente não são mantidas por tanto tempo na memória sem referências anafóricas e o excesso de possibilidades dificulta o entendimento de qual é a entidade que está sendo referenciada.

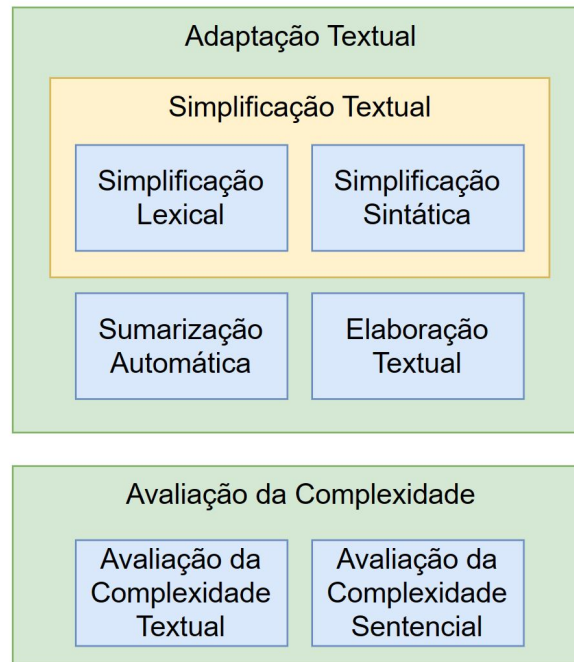
Entre as diferentes tarefas relacionadas à área da complexidade textual pode-se delimitar dois grupos principais: as tarefas de adaptação textual, que buscam transformar um texto em uma versão mais legível, e as tarefas de avaliação da complexidade, que tem como foco o desenvolvimento e análise de métricas que capturem a complexidade desses textos eficientemente para que seja possível compará-los de forma automática.

Na figura Figura 3 apresenta-se um diagrama que mostra como as diferentes tarefas de complexidade textual são divididas conceitualmente. Dentro do escopo deste trabalho, serão aplicadas tarefas de simplificação textual, tanto em léxico quanto em sintaxe, juntamente com a avaliação da complexidade textual.

### **2.3.1 Avaliação de Complexidade**

Os primeiros esforços para avaliar a complexidade de textos escritos consistiam em cálculos de implementação computacional simples baseados apenas na quantidade de caracteres, sílabas, palavras e sentenças e em alguns casos na quantidade de palavras complexas segundo algum critério. Com base nesses fatores, uma pontuação numérica é atribuída a um texto (Leal *et al.*, 2024; Moreno *et al.*, 2023). Esses índices, descritos na Subseção 2.3.1.1, são úteis para representar uma parcela dos fatores que influenciam na complexidade de um texto, mas não tem como objetivo avaliar o conteúdo nem a estrutura interna do texto (Leal *et al.*, 2024). Por isso, também foram incluídas neste trabalho métricas que se relacionam com aspectos morfológicos e sintáticos do discurso, descritas na Subseção 2.3.1.2.

Figura 3 – Tarefas de processamento de linguagem natural relacionadas à complexidade textual.



Fonte: Leal e Aluísio (2024)

### 2.3.1.1 Índices de Inteligibilidade

Esta seção descreve cada um dos índices de inteligibilidade utilizados neste trabalho, com os seus valores constantes adaptados para o Português segundo o trabalho de Moreno *et al.* (2023). Tais índices utilizam diferentes escalas para medir a inteligibilidade de um texto: dois deles medem em uma escala centígrada, onde 100 é a inteligibilidade máxima e 0 a mínima, e 4 deles medem em uma escala de nível de instrução, em que os números correspondem a quantidade de anos de educação formal necessárias para compreender um texto.

As variáveis utilizadas no cálculo desses índices são:  $N_{sentencas}$ , quantidade de sentenças em um texto (separadas por ponto final e ponto e vírgula);  $N_{palavras}$ , quantidade de palavras separadas por espaço no texto, excluindo-se pontuações;  $N_{silabas}$ , a quantidade total de sílabas no texto;  $N_{caracteres}$ , quantidade de caracteres no texto, incluindo pontuações; e  $N_{palavrascomplexas}$ , a quantidade de *types* que são considerados complexos sob algum critério.

**Índice de Inteligibilidade de Flesch:** é um dos índices mais antigos a ser utilizado para quantificar a facilidade de compreensão de um texto. Foi elaborado em 1943 e revisado em 1948. Definido em uma escala que varia de 0 a 100 (ainda que os limites teóricos descrevam um intervalo maior, valores fora dessa escala são incomuns), com 100 representando o máximo de inteligibilidade (Flesch, 1948). Este índice é calculado pela Equação 2.8.

$$\text{Flesch} = 227 - 1,04 \cdot \frac{N_{\text{palavras}}}{N_{\text{sentencas}}} - 72 \cdot \frac{N_{\text{silabas}}}{N_{\text{palavras}}} \quad (2.8)$$

**Nível de Instrução de Flesch-Kincaid:** Em 1975, a fórmula de Flesch foi reescrita para utilizar a escala de nível de instrução por J. Peter Kincaid, e por isso a métrica foi batizada dessa forma (Kincaid *et al.*, 1975). A fórmula para esse índice está presente na Equação 2.9.

$$\text{Flesch Kincaid} = 0,36 \cdot \frac{N_{\text{palavras}}}{N_{\text{sentencas}}} + \frac{N_{\text{silabas}}}{N_{\text{palavras}}} \quad (2.9)$$

**Índice de Nebulosidade de Gunning:** elaborado em 1952, esse índice mede a complexidade na escala de nível de instrução com base no número total de palavras, número de palavras complexas e número de sentenças (Gunning, 1968). O autor considerou como palavras complexas todas as que possuíam mais de duas sílabas, mas essa visão não se reflete da mesma forma no Português, que possui, em média, mais sílabas nas suas palavras do que o inglês (Moreno *et al.*, 2023; Leal *et al.*, 2024). Calculado segundo a Equação 2.10.

$$\text{Nebulosidade de Gunning} = 0,49 \cdot \frac{N_{\text{palavras}}}{N_{\text{sentencas}}} + 19 \cdot \frac{N_{\text{palavras complexas}}}{N_{\text{palavras}}} \quad (2.10)$$

**Índice de Inteligibilidade Automatizado - Automated Readability Index (ARI):** desenvolvido em 1967, com o objetivo de fornecer um índice que pudesse ser usado como referência para diminuir o tempo de extração de informação de documentos da Força Aérea dos Estados Unidos, esse índice também utiliza a escala de nível de instrução (Smith; Senter, 1967). O cálculo é dado na Equação 2.11.

$$\text{ARI} = 0,44 \cdot \frac{N_{\text{palavras}}}{N_{\text{sentencas}}} - 4,6 \cdot \frac{N_{\text{caracteres}}}{N_{\text{palavras}}} - 20 \quad (2.11)$$

**Índice de Coleman-Liau:** semelhante ao ARI, com foco na simplicidade de implementação, mas utiliza uma razão do número de sentenças pelo número de palavras, o inverso dos índices anteriores (Coleman; Liau, 1975). Sua fórmula é a descrita na Equação 2.12.

$$\text{Coleman Liau} = 5,4 \cdot \frac{N_{\text{caracteres}}}{N_{\text{palavras}}} - 21 \cdot \frac{N_{\text{sentencas}}}{N_{\text{palavras}}} - 14 \quad (2.12)$$

**Índice Gulpease:** é um outro índice que usa uma escala centígrada, mas foi criado especificamente para avaliar a inteligibilidade de textos em Italiano (Lucisano; Piemontese,

1988). No trabalho de Moreno *et al.* (2023), os valores originais já estavam dentro da margem esperada na conversão, o que significa que esse índice não necessitou de alterações para ser portado para o português. O cálculo original é o da Equação 2.13.

$$\text{Gulpease} = 89 + \frac{300 \cdot N_{\text{sentencas}} - 10 \cdot N_{\text{caracteres}}}{N_{\text{palavras}}} \quad (2.13)$$

A principal desvantagem em todas estas métricas é o fato de que apenas estatísticas básicas de sentenças e palavras não representam muitos dos fatores de complexidade textual mostrados na Seção 2.3. Nenhuma delas considera informações detalhadas da sintaxe, semântica e concatenação lógica de termos em uma ou várias sentenças. A próxima seção apresenta o outro conjunto de métricas selecionado que considera os aspectos morfológicos e sintáticos.

### 2.3.1.2 Métricas Morfossintáticas

Muitos dos fatores descritos no início deste capítulo não dependem somente da frequência e de palavras no texto, mas também de aspectos complexos que são percebidos de forma subjetiva pelos leitores.

Neste trabalho, foram selecionadas 7 das métricas descritas em Leal *et al.* (2024) que avaliam diferentes aspectos sintáticos dos textos distribuídas nas categorias de 1. complexidade sintática 2. informação morfossintática das palavras 3. coesão referencial e 4. simplicidade textual com base nas diretrizes presentes em UNIVERSIDADE ESTADUAL DE CAMPINAS (2024), UNIVERSIDADE ESTADUAL DE GOIÁS (2023), Sousa *et al.* (2024), Almeida *et al.* (2024). A seguir estão descritas cada uma das métricas selecionadas de Leal *et al.* (2024).

**Non SVO ratio:** é a proporção de frases que estão fora da ordem básica do português, SVO, ou seja, todas as frases onde o sujeito aparece depois do verbo principal ou o objeto aparece antes, essa proporção é definida na Equação 2.14.

$$\text{non\_svo\_ratio} = \frac{N_{\text{sentencas\_nao\_svo}}}{N_{\text{sentencas}}} \quad (2.14)$$

**Passive ratio:** proporção de orações em voz passiva no texto, em Leal *et al.* (2024) apenas a voz passiva analítica é reconhecida, e o padrão buscado nas sentenças é de uma conjugação do verbo 'ser' seguida diretamente de um verbo no particípio passado. Sua fórmula está descrita na Equação 2.15.

$$\text{passive\_ratio} = \frac{N_{\text{sentencas\_passivas}}}{N_{\text{sentencas}}} \quad (2.15)$$

**Words before main verb:** quantidade de palavras antes do verbo principal da oração. Neste trabalho utilizou-se o valor médio dessa métrica entre todas as sentenças de um texto, segundo a Equação 2.16, em que  $AVP_i$  é o número de palavras antes do verbo na sentença  $i$  e  $|S_i|$  o tamanho da sentença  $i$ .

$$words\_before\_main\_verb = \frac{\sum_{i=1}^{N_{sentencas}} \frac{AVP_i}{|S_i|}}{N_{sentencas}} \quad (2.16)$$

**Personal pronouns:** razão entre o número de pronomes pessoais e o número de palavras no texto. O cálculo é mostrado na Equação 2.17.

$$personal\_pronouns = \frac{N_{pronomes\ pessoais}}{N_{palavras}} \quad (2.17)$$

**Coreference pronoun ratio:** número de possíveis correferências na sentença anterior pelos pronomes anafóricos do caso reto: ‘ele’, ‘ela’, ‘eles’, ‘elas’. Neste trabalho a média das médias entre todos os pares de sentenças é utilizada para descrever as correferências no documento inteiro. Compartilha das mesmas limitações de Leal *et al.* (2024), reconhecendo apenas substantivos e nomes próprios que concordem em gênero e número com os pronomes encontrados. O cálculo dessa métrica é o da Equação 2.18, em que  $N_{spcr}$  é o número de sentenças com pronomes anafóricos do caso reto,  $PR_i$  o número de pronomes anafóricos na sentença  $i$  e  $Ref_j$  é o número de possíveis referentes da sentença  $j$ .

$$coreference\_pronoun\_ratio = \frac{\sum_{i=2}^{N_{spcr}} \frac{Ref_{i-1}}{PR_i}}{N_{spcr}} \quad (2.18)$$

**Demonstrative pronoun ratio:** mesma lógica da métrica anterior (e portanto as mesmas limitações), mas buscando anáforas em pronomes demonstrativos em vez do caso reto. Na Equação 2.19  $N_{spd}$  é a quantidade de sentenças com pronomes demonstrativos e  $PD_i$  o número de pronomes demonstrativos na sentença  $i$ , e o restante do cálculo é similar à Equação 2.18, mas utilizando  $N_{spd}$  em vez de  $N_{spcr}$  e  $PD_i$  no lugar de  $PR_i$ .

$$demonstrative\_pronoun\_ratio = \frac{\sum_{i=2}^{N_{spd}} \frac{Ref_{i-1}}{PD_i}}{N_{spd}} \quad (2.19)$$

**Long sentence ratio:** Diferentemente de Leal *et al.* (2024), neste trabalho consideram-se sentenças longas aquelas que possuem mais de 20 palavras, mas a razão utilizada é a mesma de Leal *et al.* (2024) como mostrado na Equação 2.20.

$$long\_sentence\_ratio = \frac{N_{sentencas\ longas}}{N_{sentencas}} \quad (2.20)$$

Adicionalmente, uma nova métrica foi elaborada para este trabalho, a razão de lemas estrangeiros comparados com todos os lemas do texto, ou *foreign word ratio*. A motivação para inclusão dessa métrica veio dos guias de linguagem simples supracitados, onde existe o consenso de que um número grande de termos estrangeiros que não são de uso comum prejudica o entendimento pelo público. A fórmula elaborada para essa métrica é a da Equação 2.21.

$$foreign\_word\_ratio = \frac{N_{lemas\ estrangeiros}}{N_{lemas}} \quad (2.21)$$

Quadro 1 – Métricas morfossintáticas selecionadas e suas respectivas categorias.

Métrica	Categoria	Princípios de Linguagem Simples Relacionados
<i>non SVO ratio</i>	Complexidade Sintática	Ordem direta das palavras
<i>passive ratio</i>		Voz ativa em vez de passiva
<i>words before main verb</i>		Sentenças curtas Ordem direta das palavras
<i>personal pronouns</i>	Informação Morfossintática das Palavras	Comunicação próxima do leitor
<i>coreference pronoun ratio</i> <i>demonstrative pronoun ratio</i>	Coesão Referencial	Uma ideia por frase
<i>long sentence ratio</i>	Simplicidade Textual	Sentenças curtas
<i>foreign word ratio</i>	Diversidade Lexical <sup>5</sup>	Evitar termos estrangeiros que não sejam de uso corrente

Fonte: Leal *et al.* (2024)

## 2.4 Grandes Modelos de Linguagem

Os modelos de linguagem consistem, basicamente, de sistemas computacionais capazes de prever uma palavra mais provável a partir de palavras anteriores. Sendo utilizados em diversas atividades de PLN. Esses modelos se distinguem em modelos probabilísticos e modelos de linguagem neurais (Paes *et al.*, 2024).

Nos modelos probabilísticos a geração da linguagem é modelada com base na probabilidade de ocorrência de uma dada sequência de palavras a partir de uma entrada (*prompt*). Essa probabilidade utiliza a suposição de Markov de que na probabilidade apenas o passado mais recente importa para o futuro, dessa forma a próxima palavra mais provável é sempre escolhida

<sup>5</sup> Como a *foreign word ratio* não está incluída em Leal *et al.* (2024), a diversidade lexical foi considerada a categoria mais adequada.

com base na imediatamente anterior. A probabilidade de uma palavra pode ser determinada, tomando como exemplo modelos de n-gramas, pela coocorrência dela e da palavra anterior no *corpus* considerado para o modelo (Paes *et al.*, 2024; Zhao *et al.*, 2025).

Já os LLMs são construídos principalmente sobre a arquitetura de rede neural, e a maioria dos modelos recentes, como os modelos BERT e GPT, utilizam o mecanismo de *Transformers*, descrito primeiramente em Vaswani *et al.* (2017).

### 2.4.1 Redes Neurais

Uma rede neural consiste de uma sequência de camadas de neurônios, que estão conectados entre si, o termo neurônio é utilizado por causa do neurônio de McCulloch-Pits, um modelo computacional que buscava imitar um neurônio biológico por meio da lógica proposicional (Jurafsky; Martin, 2026, cap. 6).

Esses neurônios consistem de unidades computacionais que recebem um vetor de valores de entrada e retornam uma única saída. Cada um deles possui pesos determinados para os valores de entrada e um termo de viés (Jurafsky; Martin, 2026, cap. 6).

O tipo de rede neural mais simples é a rede neural direta (*feedforward*), onde os neurônios são agrupados em múltiplas camadas e cada uma delas envia informações apenas para a próxima camada, nenhuma informação é passada de volta para camadas anteriores (Jurafsky; Martin, 2026, cap. 6).

A Figura 4 ilustra uma rede neural profunda, que possui múltiplas camadas interligadas, consistindo de uma camada de entrada, que recebe os dados externo, uma ou mais camadas ocultas densamente conectadas, onde os padrões complexos são aprendidos, e uma camada de saída com os resultados das computações realizadas entre as camadas. (Jurafsky; Martin, 2026)

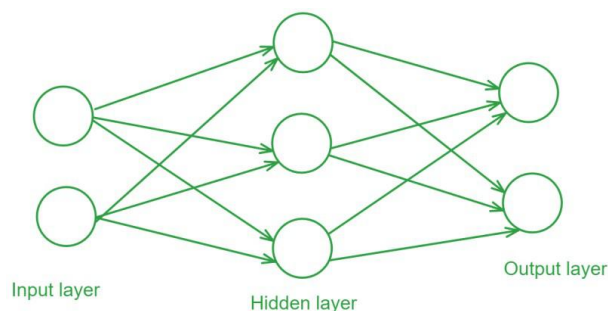


Figura 4 – Exemplo de rede neural direta

Uma das etapas realizadas na construção de uma rede neural é o pré-treinamento,

onde o modelo irá “aprender” os pesos iniciais de seus neurônios ao realizar uma determinada tarefa em uma grande quantidade de dados. Essa tarefa é realizada por um certo número de iterações no conjunto de dados (épocas) e o algoritmo que atualiza os pesos (taxa de aprendizado) pode ser personalizado para mudar a forma como os pesos de cada neurônio são atualizados. Todos esses parâmetros que controlam o aprendizado do modelo são chamados de **hiperparâmetros** (Paes *et al.*, 2024)

O tipo de tarefa realizado durante o treinamento pode variar bastante a depender de quais dados são fornecidos ao modelo, sua arquitetura, objetivo, etc. Por exemplo, o modelo BERT realiza uma tarefa de *masked language modeling* para prever um *token* escondido utilizando apenas os outros tokens da sequência.

Também há problemas que podem ocorrer nesse pré-treinamento, sendo o mais notável deles o enviesamento: onde o modelo pode priorizar informações que não são verdadeiras e criar padrões que não, por exemplo: . Por isso, é indispensável que o conjunto de dados seja diverso para evitar a associação de padrões negativos à indivíduos (Paes *et al.*, 2024).

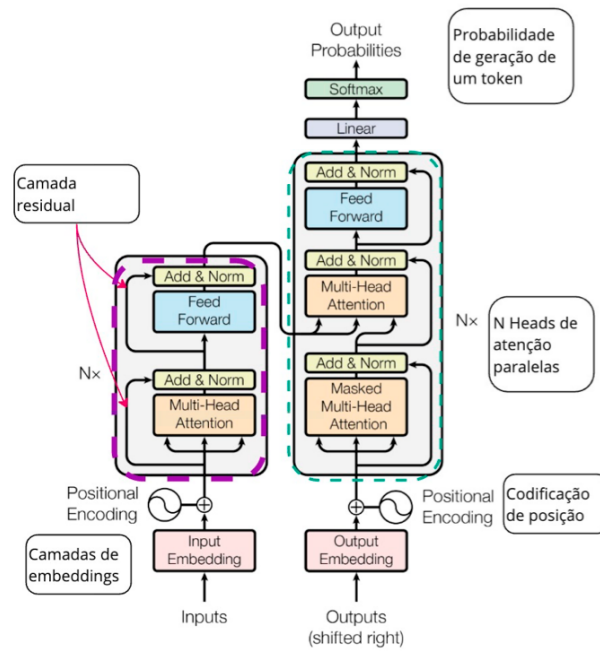
#### 2.4.2 *Transformers*

O *Transformer* é o principal componente dos LLMs, e tem como responsabilidade o processamento dos *tokens* de forma paralelizável. Composto por uma pilha de codificadores e uma pilha de decodificadores. Cada sub-codificador da pilha possui um mecanismo de auto-atenção e uma rede neural completamente conectada de uma camada (Paes *et al.*, 2024).

Dentro do *transformer* está um mecanismo de auto-atenção, que é responsável por atribuir pesos para cada *token* de uma camada anterior com base nos *tokens* próximos (Paes *et al.*, 2024; Jurafsky; Martin, 2026). Os *transformers* utilizam múltiplos pontos de atenção com seus próprio pesos e o valor propagado como resultado é a combinação da saída de cada ponto de atenção. Na camada de decodificação, o mecanismo de atenção só tem acesso aos *tokens* que vêm antes do *token* que está sendo processado.

A Figura 5 mostra de forma gráfica a arquitetura geral do *Transformer*, nela pode-se observar as camadas de atenção e redes neurais, juntamente das camadas de normalização.

Figura 5 – Diagrama da arquitetura *Transformer*. Em violeta está o codificador e em verde o decodificador



Fonte: Paes *et al.* (2024)

### 3 TRABALHOS RELACIONADOS

Este Capítulo apresenta trabalhos relacionados ao proposto que realizem a simplificação textual com LLMs.

O primeiro trabalho, apresentado na Seção 3.1, trata de um sistema para simplificação textual realizado por LLM segundo regras da linguagem simples, com diferentes níveis de simplificação para públicos específicos. O segundo trabalho, descrito na Seção 3.2, apresenta diferentes estratégias de *prompting* que melhoram a performance da simplificação textual voltada para textos da área educacional. A Seção 3.3 discorre sobre o terceiro trabalho, que tem como principal objetivo realizar a simplificação textual de forma que um mesmo texto possa ser adaptado para diferentes níveis de escolaridade. Na Seção 3.4 o último trabalho relacionado é descrito, esse trabalho realizou a simplificação textual em resumos de artigos acadêmicos da área da saúde e materiais educativos para pacientes.

Na sequência esses trabalhos são aprofundados individualmente. Ao final desse Capítulo, apresenta-se uma tabela comparando as características de cada um dos trabalhos mostrados com o presente trabalho.

#### 3.1 *SimplifyMyText: An LLM-Based System for Inclusive Plain Language Text Simplification* (Färber *et al.*, 2025)

Em Färber *et al.* (2025), apresenta-se uma ferramenta para realizar a simplificação textual com LLM segundo princípios da linguagem simples, possuindo diferentes níveis de simplificação para 5 grupos de usuários diferentes: 1. Cientistas e Pesquisadores 2. Estudantes e Acadêmicos 3. Profissionais da Indústria 4. Jornalistas e membros midiáticos 5. Público Geral . Os modelos utilizados pelos autores foram o Llama 3.1:70b e GPT-4o.

O ponto de interseção desse trabalho com o aqui presente é a simplificação textual realizada por LLMs apoiada na linguagem simples, com ambos os trabalhos implementando um modelo de *zero-shot learning*, mas diferentemente de Färber *et al.* (2025), neste trabalho avaliam-se diversos modelos Ollama menores que o Llama 3.1 com 70 bilhões de parâmetros para encontrar respostas sobre a possibilidade de usar modelos menores nessa tarefa com eficiência similar à de grandes LLMs proprietários.

Também foram consideradas mais convenções da linguagem simples durante a geração de documentos, Färber *et al.* (2025) considerou apenas a preferência de sentenças curtas

em vez de longas nos *prompts* enviados aos modelos, enquanto que o presente trabalho adicionou também a recomendação de frases na ordem direta, em voz ativa e sem termos complexos ou jargões.

Além disso, as métricas utilizadas neste trabalho ampliam a análise do discurso nos artefatos gerados, visto que também foram consideradas características morfossintáticas e utilizou-se outros índices de inteligibilidade além do nível de escolaridade Flesch-Kincaid. Enquanto que em Färber *et al.* (2025) utiliza-se somente o SARI (Xu *et al.*, 2016), *Bilingual Evaluation Understudy* (BLEU) (Papineni *et al.*, 2002) e o nível de escolaridade Flesch-Kincaid (Kincaid *et al.*, 1975).

### **3.2 *Evaluating GenAI for Simplifying Text for Education: Improving Accuracy and Consistency for Enhanced Readability* (Day *et al.*, 2025)**

O trabalho de Day *et al.* (2025) também avalia a simplificação textual realizada por LLMs, ainda que não considere os princípios da linguagem simples. Nele são investigadas 4 estratégias de *prompt* diferentes: encadeamento de *prompts*, cadeia de pensamento (*Chain-of-Thought*), *prompting* de estímulo direcional e *zero-shot prompting*, além uma arquitetura multi-agentes na simplificação de textos do 12º ano escolar para o 8º, 6º e 4º ano realizada pelos modelos GPT-4-Turbo, Claude 3 e Mixtral 8x22B.

Um ponto de destaque desse trabalho é a avaliação de como as palavras relevantes são mantidas após a simplificação, isso foi verificado por meio da comparação com avaliações humanas e os resultados da métrica BERTScore (Zhang *et al.*, 2020), também utilizada neste trabalho.

A avaliação da simplificação em si foi realizada com base no nível de escolaridade Flesch-Kincaid e na mudança percentual do número de palavras.

### **3.3 *Is It Possible to Modify Text to a Target Readability Level? An Initial Investigation Using Zero-Shot Large Language Models* (Farajidizaji *et al.*, 2024)**

No trabalho de Farajidizaji *et al.* (2024) avalia-se a simplificação textual a nível de parágrafo em um *corpora* de textos de diferentes domínios, com o objetivo geral de criar múltiplas versões de um mesmo parágrafo simplificadas para um conjunto pré-definido de valores do índice de inteligibilidade de Flesch, que demonstram um nível de adequação para grupos de

usuários distintos. Foram usados 8 valores para cada paráfrase, representando os pontos médios dos intervalos de 0-10 (5), 10-30 (20), 30-50 (40), 50-60 (55), 60-70 (65), 80-90 (85) e 90-100 (95).

Após a geração das 8 versões de cada parágrafo, a ordem da complexidade cada paráfrase foi comparada com a ordem do seu intervalo esperado utilizando a correlação de Spearman e a divergência no valor do índice de inteligibilidade de Flesch esperado e encontrado foi calculada com a raiz do erro quadrático médio (RMSE). As divergências lexicais foram calculadas utilizando o Taxa de erros de palavras - *Word Error Rate* (WER) e o BERTScore (Zhang *et al.*, 2020).

O trabalho de Farajidizaji *et al.* (2024) relaciona-se com o trabalho proposto na medida que ambos utilizam LLMs para realizar a simplificação textual e realizam uma comparação de modelos menores com grandes modelos proprietários, no caso de Farajidizaji *et al.* (2024), o Llama2-7b-chat-hf é comparado ao gpt-3.5-turbo. Mas existe um contraste na forma que essa simplificação é realizada, visto que o presente trabalho opera a nível de documento, enquanto que Farajidizaji *et al.* (2024) atua a nível de parágrafos.

### **3.4 Assessing AI Simplification of Medical Texts: Readability and Content Fidelity (Picton *et al.*, 2025)**

Picton *et al.* (2025) apresenta uma discussão sobre a eficácia do modelo ChatGPT-4 na simplificação de resumos de artigos em neurologia e neurocirurgia junto de materiais educativos para pacientes (PEM).

A avaliação foi conduzida em 100 resumos e 340 PEMs, considerando o índice de escolaridade Flesch-Kincaid e o índice de inteligibilidade de Flesch como medidas de inteligibilidade.

A similaridade do conteúdo simplificado e a versão original foi medida com a similaridade do cosseno em representações feitas com a Análise semântica latente - *Latent Semantical Analysis* (LSA) como métrica de similaridade entre as diferentes versões de um mesmo texto. Essas medidas foram comparadas com a avaliação de similaridade realizada por revisores humanos, em uma escala de 0 a 1, através do coeficiente de correlação de Pearson.

Em relação ao processo de simplificação, o *prompt* utilizado solicitava que o texto fosse simplificado para o nível de compreensão do 5º ano do ensino básico (4º ano do ensino fundamental brasileiro) sem adicionar ou remover informações no texto. Ou seja, as instruções

especificavam ao modelo para realizar a simplificação textual sem a sumarização, o que era desejado pelos autores, dado que a maioria dos textos eram pouco extensos (média de  $377,53 \pm 99,84$  caracteres nos resumos e  $581,57 \pm 726,42$  nos PEMs). O ChatGPT-4 conseguiu reduzir a média do índice de escolaridade de Flesch-Kincaid de aproximadamente 12 nos resumos e 13 nos PEMs para aproximadamente 5 nos dois casos.

A interseção do trabalho de Picton *et al.* (2025) com este se dá por ambos considerarem índices de inteligibilidade e métricas de similaridade semântica para avaliar as simplificações, mas o primeiro não incluiu a avaliação de modelos abertos ou o uso de práticas simples no seu escopo.

### **3.5 Comparação entre os trabalhos**

No Quadro 2 resumem-se os principais pontos de comparação entre os trabalhos. Nele o trabalho de Färber *et al.* (2025) e o trabalho proposto utilizam princípios da linguagem simples, com ambos considerando a prática de sentenças curtas em vez de longas, mas apenas o trabalho proposto considera práticas relacionadas à ordem direta das palavras, uso da voz passiva e presença de termos complexos. Färber *et al.* (2025) e Farajidizaji *et al.* (2024) realizam a simplificação textual em um domínio geral, Day *et al.* (2025) em um contexto educacional, Picton *et al.* (2025) foca especificamente em textos médicos e o trabalho proposto discorre sobre a simplificação apenas no contexto administrativo da UFC Inova.

Quadro 2 – Características de cada trabalho relacionado

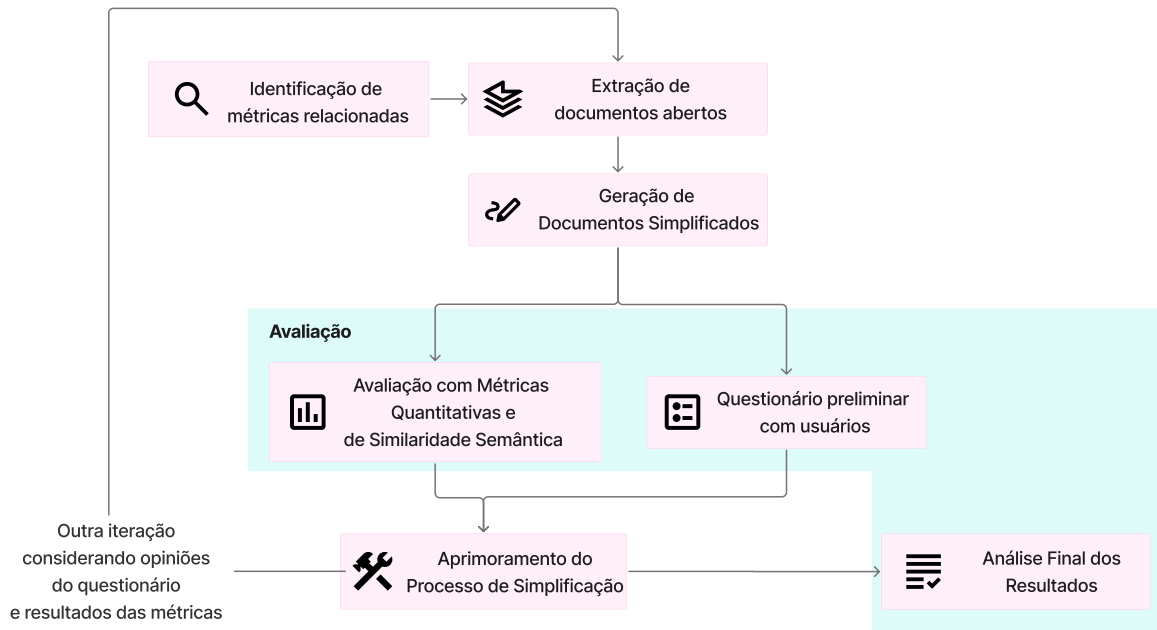
	Färber <i>et al.</i> (2025)	Day <i>et al.</i> (2025)	Farajidizaji <i>et al.</i> (2024)	(Picton <i>et al.</i> , 2025)	Trabalho Proposto
<b>Domínio</b>	Geral	Educacional	Geral	Saúde	Administrativo
<b>Público-Alvo</b>	Público geral	Estudantes de ensino fundamental e médio	Público geral	Acadêmicos da área da saúde e pacientes	Estudantes e Professores Universitários
<b>Unidade de Simplificação</b>	Documento	Documento	Parágrafo	Documento	Documento
<b>Práticas da Linguagem Simples utilizadas</b>	Sentenças curtas em vez de longas	Nenhuma	Nenhuma	Nenhuma	Sentenças curtas em vez de longas Frases na ordem Direta Utilizar a voz ativa em vez da voz passiva Evitar termos complexos e jargões
<b>Modelos Utilizados</b>	GPT-4o Llama3.1:70b	GPT-4-Turbo Claude 3 Mixtral 8x22B	gpt-3.5-turbo Llama2-7b-chat-hf	GPT-4	Gemini 2.5 Pro Gemini 2.5 Flash Phi4:14.7b Phi3:3.8b Cow/Gemma2-tools:2b Llama3.2:3.2b Gemma3:4b Qwen2.5:14b Qwen2.5-coder:32b Deepseek-r1:14b Granite-code:8b Granite3-dense:2b Granite3-dense:8b
<b>Estratégias de Prompting</b>	<i>Zero-shot</i>	<i>Zero-shot</i> Estímulo direcional Cadeia de pensamento Encadeamento de <i>prompts</i> (isoladamente)	<i>Zero-shot</i>	<i>Zero-shot</i>	<i>Zero-shot</i> (isolado)  Estímulo direcional Cadeia de pensamento Encadeamento de <i>prompts</i> (conjuntamente)
<b>Métricas de avaliação</b>	BLEU Nível de escolaridade Flesh-Kincaid	BERTScore Nível de escolaridade Flesh-Kincaid Acurácia de palavras-chave Mudança percentual do número de palavras	BERTScore WER Índice de inteligibilidade de Flesch	LSA Índice de inteligibilidade de Flesch Nível de escolaridade Flesch-Kincaid	BERTScore Índice de inteligibilidade de Flesch Nível de escolaridade Flesch-Kincaid Índice de nebulosidade de Gunning ARI Índice Coleman-Liau Índice <i>Gulpease non svo ratio passive ratio words before main verb personal pronoun ratio coreference pronoun ratio demonstrative pronoun ratio long sentence ratio foreign word ratio</i>

Fonte: Elaborado pelo autor.

## 4 METODOLOGIA

A metodologia deste trabalho divide-se em 7 etapas que estão ilustradas na Figura 6. As próximas seções deste Capítulo descrevem a forma como cada etapa foi realizada para este trabalho.

Figura 6 – Representação do processo metodológico.



Fonte: Elaborado pelo autor.

### 4.1 Identificação de Métricas de Complexidade Textual Relacionadas à Linguagem Simples.

Nessa etapa foi realizada uma revisão da literatura acerca da linguagem simples e do processo de simplificação textual. Durante essa análise, foram identificados os principais pontos priorizados na criação de textos em linguagem simplificada por diferentes instituições, e os pontos mais relevantes para o contexto administrativo da UFC Inova serão priorizados.

Ao fim dessa revisão da literatura, foram selecionadas as métricas a serem utilizadas neste trabalho, que incluem índices de inteligibilidade adaptados para o Português em Moreno *et al.* (2023) (ver Subseção 2.3.1.1); métricas morfossintáticas com a maior relação com os princípios de linguagem simples, oriundas do trabalho de Leal *et al.* (2024) e apresentadas na Subseção 2.3.1.2; e também a métrica de similaridade semântica BERTScore, detalhado na 2.2.2.2

Algumas métricas morfossintáticas tiveram sua implementação alterada para uma maior precisão com o uso de atributos presentes na análise sintática do UD, na Seção 5.1 são apresentadas as mudanças realizadas nessas métricas.

## 4.2 Extração de Documentos Abertos

Nessa etapa são extraídos os documentos a serem simplificados, dando prioridade aos editais de competições e programas de bolsa, por serem os documentos que são acessados pelo público com a maior frequência. Além disso, os documentos são convertidos para o formato *Markdown* antes de serem enviados aos LLMs.

## 4.3 Geração de Documentos Simplificados com Diferentes LLMs

Após a extração dos documentos, novas versões simplificadas por IAG são geradas com diferentes modelos, mas com as mesmas configurações de *hardware* (no caso dos modelos do Ollama), *prompt* e temperatura para uma maior reprodutividade nos testes realizados.

Os LLMs utilizados na primeira rodada de simplificação são:

- a) gemini-2.5-flash
- b) gemini-2.5-pro
- c) phi4:14.7b
- d) phi3:3.8b
- e) gemma2\_tools:2b
- f) llama3.2:3.2b
- g) gemma3:4b
- h) qwen2.5:14b
- i) deepseek-r1:14b
- j) granite3-dense:2b
- k) granite3-dense:8b

Onde as duas versões do Google Gemini são os LLMs de maior escala que serão comparados com o restante dos modelos vindos do Ollama.

Os textos *Markdown* dos documentos originais e simplificados pela IAG foram convertidos para PDF com a ferramenta Pandoc (MacFarlane *et al.*, 2006) com a mesma configuração de formatação e fonte, e esse documento reconvertido foi mostrado para os usuários

durante a avaliação qualitativa descrita em 4.5. A motivação para converter os documentos originais em vez de utilizar o PDF original era a de incentivar os usuários a considerarem apenas o conteúdo durante o questionário descrito em Seção 4.5.

O acesso aos modelos Gemini é feito pela API da Google, e os modelos Ollama estão hospedados em uma máquina da UFC que os expõe através da interface Open WebUI. As configurações de *hardware* são:

- a) **Memória RAM:** 128 GB DDR4
- b) **Processamento** Intel Core I9-2900F 12th 24 *cores*
- c) **Placa de Vídeo:** Nvidia Geforce RTX 3080 Ti 12228 MB

#### 4.4 Avaliação com Métricas Estabelecidas

Para garantir que apenas o conteúdo de texto plano seja considerado durante a análise sintática utilizado nessa etapa, é necessário que ele receba uma entrada pronta para análise. Por isso, são realizados tratamentos específicos em todos os documentos analisados, em suas versões completas, sumarizadas (caso existam) e geradas por IAG, esse tratamento envolve a aplicação das seguintes alterações no texto:

- a) Remoção de tabelas *markdown*.
- b) Remoção de caracteres especiais (\* e #).
- c) Remoção de espaços em branco extras.
- d) Remoção de prefixos em enumerações de seções (III., a), etc.).

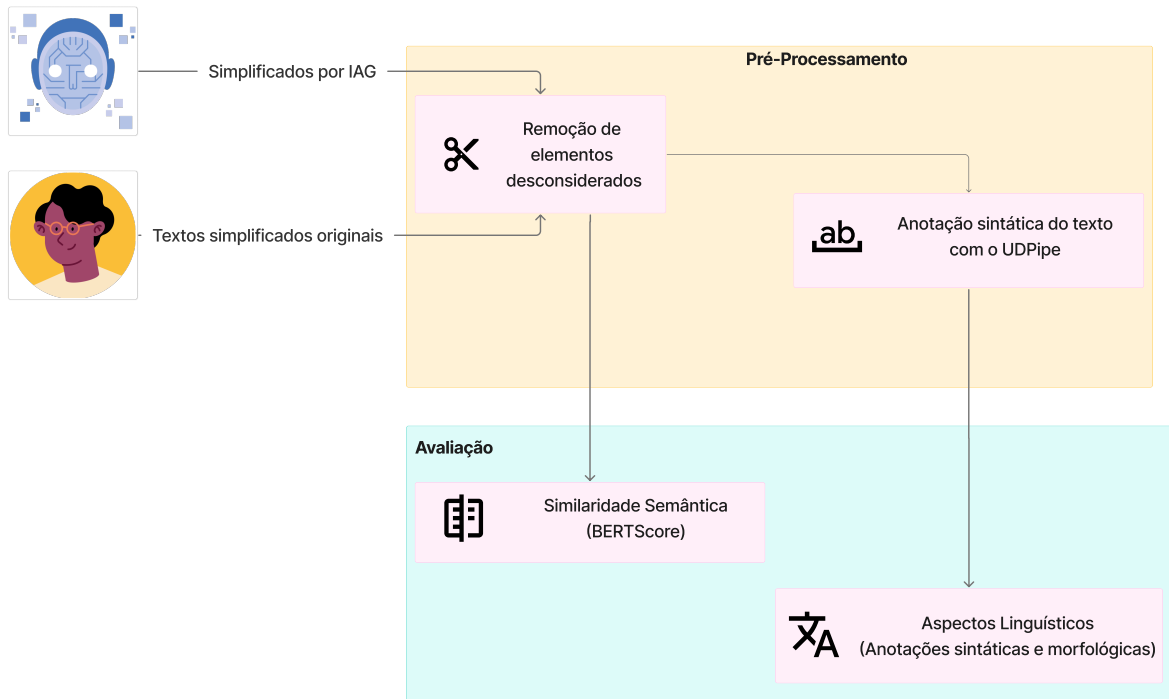
As anotações sintáticas foram geradas com o modelo UDPipe 2.0 Straka (2018), treinado no *treebank* Porttinari-base contendo textos jornalísticos do *treebank* completo de Pardo *et al.* (2021).

O processo de avaliação das métricas está descrito na Figura 7, com um fluxo que inicia ao ler os arquivos markdown gerados pelos LLMs, avança para a remoção de elementos não interpretáveis, e prossegue na extração separada das métricas linguísticas e similaridade semântica.

##### 4.4.1 Avaliação de Similaridade Semântica das simplificações

A avaliação do quão próximas as simplificações realizadas pela IAG estavam das simplificações originais foi feita com uma abordagem de comparação dos textos representados

Figura 7 – Fluxo de extração das métricas selecionadas.



Fonte: Elaborado pelo autor.

através de *embeddings*. A métrica BERTScore de Zhang *et al.* (2020) foi utilizada pelo fato de ser mais robusta do que estratégias de casamento de unigramas como o BLEU (Papineni *et al.*, 2002), e por possuir uma biblioteca Python que implementa o cálculo da métrica. O modelo de *embeddings* utilizado foi o *bert-base-multilingual-cased* de (Devlin *et al.*, 2018).

Para calcular essa similaridade, ambos os textos foram divididos em segmentos de no máximo 510 *tokens* (o tamanho total do modelo era de 512 *tokens*, mas 2 *tokens* extras eram criados para marcar o início e o fim da sequência). Após essa divisão, os segmentos da versão de referência e candidata eram agrupados par a par e o BERTScore era calculado para cada par.

O valor das 3 métricas BERTScore para o documento inteiro era a média dos valores para cada par avaliado. O cálculo sempre considerava o maior número possível de pares de segmentos, caso uma das versões possuísse mais segmentos que a outra, um par não poderia ser formado e esse segmentos eram desconsiderados.

#### 4.5 Questionário com Público-Alvo

O objetivo nessa etapa é avaliar se o público geral considera as versões simplificadas por IAG mais legíveis do que as versões originais dos documentos e se a simplificação realizada foi suficiente, ou é necessário complementar com uma sumarização do conteúdo. Para isso

esse trabalho se utiliza de um questionário auto-administrado, como descrito em Kitchenham e Pfleeger (2008), onde o público-alvo específico são membros da comunidade acadêmica da UFC, incluindo docentes, discentes e servidores técnicos-administrativos. Encaixando-se assim como uma pesquisa de levantamento (Easterbrook *et al.*, 2008).

As questões desse questionário envolvem pares de documentos de exemplo, em que um documento é a versão original e outro é a versão simplificada do modelo que possui melhor resultado nos índices de inteligibilidade escolhidos.

As seguintes questões de pesquisa foram definidas para esta etapa:

- a) **SRQ1** – Quão fácil é a compreensão do documento simplificado em comparação com o documento original?
- b) **SRQ2** – Qual a utilidade da versão gerada automaticamente para apoiar uma rápida compreensão das principais informações?
- c) **SRQ3** – Você considera que o documento simplificado está pronto para ser publicado como complemento da versão original?
- d) **SRQ4** – Que aspectos de ambas as versões simplificadas podem melhorar a compreensão por parte do seu público-alvo?
- e) **SRQ5** – Você identificou algum risco ou limitação no conteúdo destas versões simplificadas?

A descrição completa de todas as questões do questionário está presente no Apêndice C.

#### **4.6 Aprimoramento da Geração de Documentos Simplificados com Base nas Avaliações Iniciais**

Após a realização da primeira avaliação com opiniões de participantes e as métricas escolhidas, as diferenças entre os artefatos originais e as versões geradas por IAG serão medidas com as métricas descritas anteriormente e as avaliações dos participantes serão consideradas para definir se a simplificação textual foi considerada como suficiente. Caso os participantes enxerguem que somente a simplificação não facilite significativamente a leitura, então o processo é incrementado para realizar conjuntamente simplificação e sumarização dos textos.

Como a forma de comparar a sumarização é fundamentalmente diferente da simplificação, essas mudanças fazem necessário uma mudança no conjunto de dados utilizado, visto que os documentos precisam de uma versão sumarizada de referência para que o resultado do

modelo possa ser comparado com um valor esperado.

#### **4.7 Análise Final dos Resultados**

Nesta etapa final são analisados os resultados da execução dos passos metodológicos após o processo de aprimoramento descrito Seção 4.6, para determinar se houve uma simplificação satisfatória nos textos gerados por IAG, e se os modelos menores do Ollama são capazes de atingir resultados próximos dos modelos proprietários nessa tarefa.

## 5 RESULTADOS

Neste capítulo estão presentes os resultados obtidos pela aplicação dos procedimentos metodológicos.

O processo de aprimoramento resultou em uma nova iteração das etapas de extração, simplificação e avaliação com métricas quantitativas. Por isso, as Seções 5.2, 5.3, 5.4 descrevem os resultados dos procedimentos antes da etapa de aprimoramento e a Seção 5.7 traz os resultados após o aprimoramento com base nas considerações da avaliação qualitativa pelos participantes do questionário e outros resultados da primeira iteração.

O código-fonte desenvolvido durante a realização dos procedimentos metodológicos está disponível em um repositório do Github neste link: <https://github.com/Joao-Pedro-P-Holanda/text-simplification>. Na descrição do repositório está a referência para o *script* utilizado para o processamento das métricas com o PortParser V2.

### 5.1 Identificação de Métricas de Complexidade Textual Relacionadas à Linguagem Simples

Os índices de inteligibilidade e as métricas morfossintáticas descritas na Subseção 2.3.1.1 e Subseção 2.3.1.2 foram implementados seguindo sua fundamentação teórica, e no caso das métricas morfossintáticas foram realizadas mudanças para aproveitar as capacidades do *framework* UD.

Assim como em Moreno *et al.* (2023), neste trabalho usa-se a concepção de que as palavras complexas são as que não estão entre as 5.000 mais frequentes no *corpus* de textos “todos juntos” da Linguateca (2025) para o cálculo do índice de nebulosidade de Gunning descrito na Equação 2.10. Porém neste trabalho, sinais de pontuação e números foram ignorados na contagem.

Para a separação silábica necessária no índice de inteligibilidade de Flesch e no nível de escolaridade Flesch-Kincaid utilizou-se a biblioteca Pyphen<sup>1</sup> da linguagem Python.

As seguintes métricas morfossintáticas tiveram alterações na sua implementação: *non SVO ratio*, *personal pronouns*, *passive ratio*, *words before main verb*, *coreference pronoun ratio* e *demonstrative pronoun ratio*. Nos próximos parágrafos são descritas as suas implementações utilizando funcionalidades do UD.

<sup>1</sup> <https://doc.courtbouillon.org/pyphen/stable/>

A *non SVO ratio* foi implementada de forma iterativa sobre todas as palavras em uma sentença, uma sentença fora da ordem comum ocorria quando a posição de uma relação de dependência de sujeito ocorria depois do verbo principal, ou um objeto ocorria antes do verbo principal. O verbo principal era o *token* com POS ‘VERB’ e relação de dependência ‘root’. As relações de dependência consideradas para sujeito e objeto foram, respectivamente: { ‘nsubj’, ‘nsubj:pass’, ‘nsubj:outer’, ‘csubj’, ‘csubj:pass’, ‘csubj:outer’ } e ‘obj’.

Na *personal pronouns* em vez de realizar o casamento de padrões com as palavras ‘eu’, ‘tu’, ‘ele’, ‘ela’, ‘nós’, ‘vós’, ‘eles’, ‘elas’, ‘você’ e ‘vocês’, utilizou-se a *feature* PronType do UD com o valor ‘Prs’.

As sentenças em voz passiva consideradas para a *passive ratio* eram aquelas que possuíam no mínimo uma das seguintes relações de dependência do UD: { ‘csubj:pass’, ‘nsubj:pass’, ‘aux:pass’, ‘expl:pass’, ‘obl:agent’ }. Sendo uma forma mais robusta do que reconhecer apenas ocorrências do padrão sintático da voz passiva analítica (verbo ‘ser’ seguido de um verbo no particípio passado) como na implementação original presente em Leal *et al.* (2024).

O cálculo da *words before main verb* foi feito pela contagem de palavras em uma sentença que ocorriam antes do elemento textual que possuísse a relação de dependência *root* e POS ‘VERB’.

A detecção de pronomes pessoais anafóricos do caso reto na *coreference pronoun ratio* foi feita buscando palavras com os lemas ‘ele’ e ‘ela’ em cada sentença, já na *demonstrative pronoun ratio*, utilizou-se a *feature* PronType igual à ‘Dem’. Os *tokens* candidatos à anáfora na sentença anterior foram encontrados pela concordância em gênero e número com base nas *features* Gender e Number.

## 5.2 Extração de Documentos

Para a primeira etapa, 10 editais publicados em 2025, sendo 7 deles versões completas, 2 versões simplificadas e 1 era um edital de alteração do Edital PROINTER Nº 02/2025 - UFC Mundo, foram extraídos de forma manual. A conversão deles para o formato *Markdown* foi feita com o Gemini 2.5 Pro seguindo o *prompt* “Converta este documento para *Markdown*”, após isso o texto foi ajustado manualmente para remover citações geradas durante a interação com o Gemini e recriar a estrutura de cabeçalhos caso estivesse presente nos metadados do documento PDF.

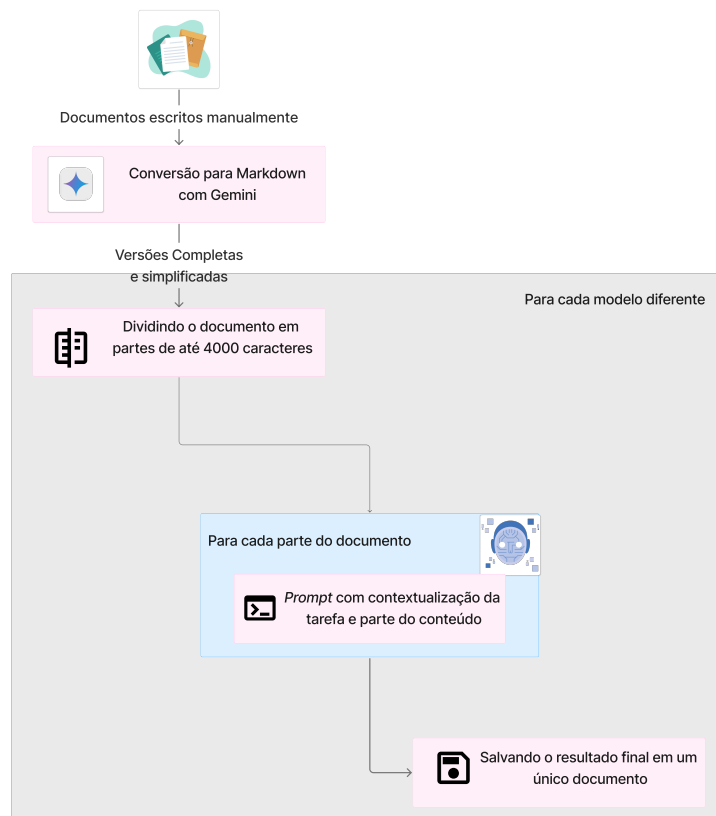
### 5.3 Geração de Documentos Simplificados com Diferentes LLMs

Nesta primeira etapa, a temperatura selecionada foi 0 para garantia de reprodutibilidade com o *prompt* utilizado

Com os documentos para análise selecionados, o passo seguinte foi executar, utilizando uma interface comum da ferramenta Langchain, a simplificação textual propriamente dita por meio de sequências de *prompts* com pedaços do conteúdo do texto com até 4000 caracteres criados com o *MarkdownTextSplitter*<sup>2</sup> disponibilizado pelo Langchain, ao final de cada sequência de *prompts* de um documento específico os resultados eram concatenados em um único arquivo *Markdown*. Nessa etapa foram gerados 130 documentos simplificados por IAG, 10 por cada modelo, totalizando 14 versões diferentes para cada documento.

A Figura 8 descreve a forma como a simplificação foi realizada.

Figura 8 – Fluxograma do processo de simplificação textual.



Fonte: Elaborado pelo autor.

Algo notável é o fato de que 4 modelos diferentes tiveram alucinações na geração de um ou mais documentos, como mostrado na Tabela 1, onde o modelo Phi3 alucinou em 70%

<sup>2</sup> [https://python.langchain.com/api\\_reference/text\\_splitters/markdown/langchain\\_text\\_splitters.markdown.MarkdownTextSplitter.html#markdowntextsplitter](https://python.langchain.com/api_reference/text_splitters/markdown/langchain_text_splitters.markdown.MarkdownTextSplitter.html#markdowntextsplitter)

dos documentos. Foram consideradas alucinações respostas que tivessem algum desses pontos: 1) repetição parcial ou total do conteúdo do *prompt*; 2) conteúdo ininteligível; 3) mais de três parágrafos em Inglês.

Tabela 1 – Quantidade de documentos gerados com alucinações por modelo.

Modelo	Número de documentos com alucinações
phi3:3.8b	7
granite3-dense:8b	3
granite3-dense:2b	1
qwen2.5:14b	1

Fonte: Elaborado pelo autor.

## 5.4 Avaliação com Métricas Estabelecidas

Na sequência da geração dos documentos, os índices de inteligibilidade e métricas morfossintáticas foram calculados em cada versão dos documentos. Os resultados nas próximas seções mostram como os resultados obtidos por cada modelo se comparam com os valores originais.

Um ponto importante de se destacar é que a versão Flash do Gemini teve resultados melhores do que a versão Pro nos dois conjuntos de métricas avaliados.

### 5.4.1 Índices de Inteligibilidade

Com relação aos índices de inteligibilidade, a Tabela 2 apresenta a melhoria percentual média de todos os modelos obtida para cada um dos modelos testados, a melhoria nos índices de Flesch e no índice *Gulpease* ocorria quando o valor medido para o texto simplificado era maior do que na versão original, enquanto que nos outros índices existia uma relação inversamente proporcional. Os modelos com melhores resultados foram o Gemini 2.5 Flash e Pro, respectivamente, seguidos pelos modelos do Ollama Phi4:14.7b e Qwen2.5:14b.

Outro resultado relevante é o fato de que o Phi3:3.8b obteve a maior melhoria no índice de inteligibilidade de Flesch, mas obteve o pior resultado no índice de nebulosidade de Gunning. Essa discrepância pode ser atribuída a grande quantidade de alucinações para esse modelo, e como 3 dessas consistiam da inclusão de pelo menos três parágrafos em inglês, vocábulos estrangeiros foram adicionados e afetaram negativamente as métricas. Da mesma forma, todos os modelos que tiveram uma piora no índice de nebulosidade de Gunning alucinaaram em mais de um documento, indicando que essas alucinações introduziam palavras mais incomuns ao

léxico considerado, como no caso do Phi3:8b que gerou 7 documentos (todos os que contiveram alucinações) com textos escritos em boa parte em Inglês.

Tabela 2 – Valor médio e intervalo de confiança de 95% para os índices de inteligibilidade de todos os documentos em cada modelo

Modelo	FE↑	IG↑	FK↓	ARI↓	GF↓	CL↓
Original	51,54 (47,43 - 55,66)	49,77 (47,00 - 52,54)	10,54 (9,54 - 11,55)	14,08 (12,98 - 15,18)	10,78 (9,58 - 11,98)	16,39 (15,75 - 17,02)
gemma2	52,54 (47,95 - 57,13)	51,05 (47,98 - 54,13)	10,17 (9,17 - 11,17)	13,59 (12,66 - 14,52)	10,30 (9,45 - 11,15)	16,29 (15,69 - 16,89)
deepseek-rl:14b	56,11 (51,01 - 61,22)	53,91 (50,35 - 57,47)	9,02 (8,01 - 10,03)	13,43 (12,52 - 14,34)	9,31 (8,66 - 9,95)	17,29 (16,44 - 18,13)
gemi-2.5-flash	<b>74,33</b> <b>(72,22 - 76,45)</b>	<b>62,02</b> (60,40 - 63,63)	<b>6,25</b> (5,89 - 6,61)	<b>9,71</b> (9,23 - 10,19)	<b>7,46</b> (7,11 - 7,81)	<b>13,22</b> (12,74 - 13,71)
gemi-2.5-pro	<b>68,35</b> <b>(66,16 - 70,54)</b>	<b>60,71</b> (58,30 - 63,12)	<b>7,07</b> (6,68 - 7,46)	<b>10,48</b> (9,87 - 11,09)	<b>7,49</b> (7,01 - 7,97)	<b>14,19</b> (13,66 - 14,71)
gemma3:4b	53,02 (48,05 - 57,99)	51,16 (47,73 - 54,60)	10,32 (9,08 - 11,55)	13,52 (12,12 - 14,92)	10,45 (9,32 - 11,58)	15,75 (14,92 - 16,58)
granite3-dense:2b	50,69 (45,96 - 55,43)	48,93 (46,38 - 51,47)	10,75 (9,68 - 11,82)	14,30 (13,31 - 15,29)	11,17 (9,98 - 12,37)	16,49 (16,00 - 16,99)
granite3-dense:8b	50,86 (47,06 - 54,67)	49,35 (46,81 - 51,89)	10,76 (9,82 - 11,71)	14,13 (13,12 - 15,14)	11,21 (10,04 - 12,38)	16,21 (15,48 - 16,93)
llama3.2	51,77 (48,21 - 55,33)	50,68 (48,62 - 52,75)	10,28 (9,49 - 11,06)	13,69 (12,96 - 14,42)	10,25 (9,48 - 11,02)	16,42 (15,85 - 16,99)
phi3:latest	<b>76,85</b> <b>(57,56 - 96,15)</b>	57,35 (47,30 - 67,39)	6,32 (2,59 - 10,06)	14,18 (12,36 - 16,00)	12,40 (11,08 - 13,71)	17,17 (16,26 - 18,08)
phi4	57,72 (54,25 - 61,18)	<b>56,60</b> <b>(53,84 - 59,35)</b>	<b>8,65</b> <b>(7,94 - 9,35)</b>	12,47 (11,86 - 13,08)	<b>8,58</b> <b>(8,10 - 9,06)</b>	16,42 (15,82 - 17,02)
qwen2.5	56,53 (51,75 - 61,31)	<b>57,45</b> <b>(53,62 - 61,27)</b>	8,72 (7,89 - 9,55)	12,55 (11,87 - 13,23)	<b>8,75</b> <b>(8,17 - 9,33)</b>	16,65 (15,93 - 17,36)

Fonte: Elaborada pelo autor.

Nota: A seta para cima indica que uma métrica é diretamente proporcional à complexidade e por isso valores menores são melhores, a seta para baixo indica métricas inversamente proporcionais, onde valores maiores são melhores

Nota: FE = índice de inteligibilidade de Flesch; FK = índice de escolaridade Flesch-Kincaid; IG = Índice Gulpease; ARI - Índice de Inteligibilidade Automatizado; GF - Índice de Nebulosidade de Gunning; CL - Índice de Coleman-Liau

## 5.4.2 Métricas Morfossintáticas

Os resultados da métricas morfossintáticas estão descritos na Tabela 3, pelo fato de muitos valores estarem em uma escala pequena, a melhoria é medida pela diferença na diferença na escala *Minmax* em vez do percentual e os números estão representados na sua parte decimal. Novamente, percebe-se o feito das alucinações na performance das métricas: 3 dos modelos tiveram uma piora em relação a métrica *foreign word ratio*, o que pode ser associado às respostas dadas em Inglês pela IAG dentro dos documentos simplificados, principalmente no caso do Phi3, que teve uma piora de 0,58.

Em relação ao tamanho das sentenças: apenas os três modelos granite (que continham

alucinações) tiveram uma piora, o restante dos modelos teve melhorias e o resultado mais positivo foi o do Gemini 2.5 Pro, com 0,3327 pontos de melhoria nessa escala.

Os modelos Gemini tiveram a maior introdução de pronomes pessoais nos textos gerados, com Gemini Flash atingindo o maior resultado entre todos os modelos, 0,5561 e o Gemini Pro 0,3231. Portanto, esses modelos criaram os textos com o maior nível de pessoalidade, algo que é desejado em UNIVERSIDADE ESTADUAL DE CAMPINAS (2024). Os modelos Ollama com resultados mais próximos foram o Gemma3 e Phi4, mas ainda eram mais de 10 vezes menores do que os valores dos modelos Gemini.

A *passive voice ratio* teve resultados negativos apenas nos 4 modelos com piores resultados negativos, já a *non svo ratio* teve resultados nos três piores modelos e no Gemini Flash (segundo melhor modelo).

5 modelos diferentes conseguiram obter intervalos superiores ao dos textos originais para a métrica *long sentence ratio* e 4 deles também tiveram intervalos superiores para a *words before main verb*, mostrando que a simplificação textual conseguiu simplificar a estrutura do texto em parte dos casos.

#### 5.4.2.1 Eficiência da métrica proposta - *foreign word ratio*

Em relação à métrica que foi elaborada para este trabalho, foram encontradas limitações na detecção de termos estrangeiros: no documento com a maior quantidade de termos estrangeiros, foram detectados apenas 8 termos de 32. Isso pode ser atribuído à performance do modelo utilizado pelo UDPipe, que não reconheceu muitas palavras de origem estrangeira e uso cotidiano, p. ex. as palavras ‘*online*’ e ‘*on-line*’ foram reconhecidas como estrangeiras em nenhum e em um documento pelo UDPipe, mas estavam presentes em três e dois documentos, respectivamente. A palavra ‘*link*’ também foi reconhecida em apenas um dos documentos originais, mas estava presente em 5 deles.

A Tabela 4 traz os resultados da acurácia, revocação, precisão e *score* F1 da métrica em cada um dos documentos originais da UFC Inova selecionados, os valores de referência foram extraídos através de verificações manuais no conteúdo de cada documento buscando por palavras estrangeiras que estivessem definidas no dicionário de Cambridge<sup>3</sup> ou termos que sejam empréstimos do latim (apenas *campus*, *campi* e *alumni* ocorreram em algum dos 10 documentos).

A tabela mostra que houveram 4 documentos em que não foi reconhecido nenhum

<sup>3</sup> <https://dictionary.cambridge.org/us/dictionary/>

Tabela 3 – Comparação da média e intervalo de confiança de 95% das métricas morfossintáticas para todos os documentos em cada modelo

Modelo	PP ↑	CP ↓	DP ↓	NS ↓	PV ↓	WB ↓	LS ↓	FW ↓
Original	,00 (,00 – ,00)	,70 (-,47 – 1,87)	1,05 (,68 – 1,42)	,05 (,03 – ,07)	,21 (,16 – ,27)	5,12 (4,10 – 6,13)	,28 (,19 – ,38)	,00 (,00 – ,01)
gemma2	,00 (,00 – ,01)	1,40 (,09 – 2,72)	1,18 (,71 – 1,65)	,04 (,02 – ,06)	,20 (,15 – ,25)	4,81 (3,93 – 5,69)	,26 (,20 – ,33)	,00 (,00 – ,01)
deepseek- r1:14b	,00 (,00 – ,00)	,63 (,02 – 1,25)	,77 (,23 – 1,31)	,03 (,01 – ,05)	,15 (,10 – ,20)	<b>3,29</b> <b>(2,50 – 4,08)</b>	<b>,13</b> <b>(,08 – ,18)</b>	,00 (,00 – ,00)
gemi- 2.5- flash	<b>,02</b> <b>(,02 – ,02)</b>	1,12 (,86 – 1,37)	,93 (,66 – 1,19)	,07 (,05 – ,09)	<b>,09</b> <b>(,07 – ,11)</b>	<b>3,23</b> <b>(2,84 – 3,62)</b>	<b>,09</b> <b>(,07 – ,11)</b>	,00 (,00 – ,00)
gemi- 2.5-pro	<b>,01</b> <b>(,01 – ,02)</b>	1,35 (1,15 – 1,55)	1,03 (,86 – 1,19)	,04 (,03 – ,05)	<b>,08</b> <b>(,07 – ,09)</b>	<b>3,25</b> <b>(2,95 – 3,56)</b>	<b>,08</b> <b>(,06 – ,10)</b>	,00 (,00 – ,00)
gemma3	,00 (,00 – ,01)	1,13 (,27 – 1,98)	1,13 (,75 – 1,50)	,08 (,05 – ,10)	,22 (,17 – ,26)	5,33 (4,26 – 6,40)	,28 (,17 – ,38)	,00 (,00 – ,00)
granite3- dense:2b	,00 (,00 – ,00)	,24 (-,04 – ,53)	1,42 (,86 – 1,97)	,05 (,03 – ,06)	,23 (,18 – ,28)	4,55 (3,35 – 5,75)	,30 (,23 – ,37)	,00 (,00 – ,01)
granite3- dense:8b	,00 (,00 – ,00)	1,15 (-,14 – 2,44)	1,26 (,84 – 1,67)	,06 (,04 – ,08)	,22 (,17 – ,28)	4,44 (3,42 – 5,46)	,29 (,22 – ,36)	,01 (,00 – ,02)
llama3.2	,00 (,00 – ,01)	,35 (-,19 – ,89)	1,14 (,62 – 1,66)	,05 (,04 – ,06)	,20 (,17 – ,24)	4,65 (3,65 – 5,65)	,26 (,20 – ,33)	,00 (,00 – ,00)
phi3:3.8b	,01 (,00 – ,01)	,43 (-,28 – 1,13)	,47 (,03 – ,91)	,04 (,01 – ,06)	,11 (,01 – ,21)	3,32 (1,77 – 4,86)	,25 (,08 – ,43)	,29 (,15 – ,44)
phi4	,00 (,00 – ,01)	,90 (,16 – 1,64)	1,05 (,68 – 1,41)	,03 (,02 – ,05)	,12 (,09 – ,15)	3,30 (2,38 – 4,22)	<b>,13</b> <b>(,09 – ,17)</b>	,00 (,00 – ,00)
qwen2.5	,00 (,00 – ,01)	,33 (-,21 – ,86)	,47 (,09 – ,85)	,03 (,01 – ,04)	,14 (,10 – ,18)	<b>2,71</b> <b>(2,16 – 3,25)</b>	<b>,10</b> <b>(,07 – ,13)</b>	,00 (,00 – ,00)

Fonte: Elaborada pelo autor.

Nota: Valores em negrito estavam em um intervalo melhor do que o intervalo original

Nota: A seta para cima indica que uma métrica é inversamente proporcional à complexidade e por isso valores maiores são melhores, a seta para baixo indica métricas diretamente proporcionais, onde valores menores são melhores

Nota: CP = *Coreference Pronoun Ratio*; DP = *Demonstrative Pronoun Ratio*; NS = *Non SVO Ratio*; PV = *Passive Voice Ratio*; WB = *Words before main verb*; LS = *Long Sentence Ratio*; FW = *Foreign Word Ratio*; PP = *Personal Pronoun Ratio*.

termo estrangeiro dado a revocação, precisão e *score* F1 iguais a 0, mas um deles (Edital Prointer/UFC N° 03/2025 - PIBI) não possuía nenhum termo estrangeiro e por isso a acurácia medida foi de 100%. Nos casos onde houveram verdadeiros positivos (revocação > 0), a revocação atingiu valores baixos, com um máximo de 33.33%, mas a precisão atingida foi de 100% em 5 dos 6 casos, mostrando que o reconhecimento de termos estrangeiros foi incompleto, mas na maior parte das vezes não contabilizava termos não estrangeiros.

### 5.4.3 Similaridade Semântica

O cálculo da similaridade semântica utilizou a biblioteca `bert_score`<sup>4</sup> e o modelo `bert-base-multilingual-cased` de (Devlin *et al.*, 2018) para os *embeddings*. Os documentos passaram pelas mesmas etapas de filtragem usada para as métricas morfossintáticas.

A implementação foi feita de forma que ambos os documentos eram divididos em

<sup>4</sup> [https://github.com/Tiiiger/bert\\_score](https://github.com/Tiiiger/bert_score)

Tabela 4 – Acurácia, revocação, precisão e *score* F1 na métrica *foreign word ratio* por documento.

Documento	Acurácia (%)	Revocação (%)	Precisão (%)	Score F1 (%)
Edital Prointer/UFC N° 09/2025 - PROMISAES	99,56	33,33	100,00	66,67
Hackathon Inovan' do UFC	99,07	25,00	100,00	54,55
Edital Prointer PIBI N° 04/2025 - Empreende UFC - Simplificado	96,81	25,00	100,00	44,44
Edital Prointer PIBI N° 04/2025 - Empreende UFC	98,74	28,57	66,67	42,86
Edital Conjunto UFC - TRE/CE N° 01/2025 - Simplificado	96,57	21,05	100,00	36,36
Edital Conjunto UFC - TRE/CE N° 01/2025	98,77	5,00	100,00	9,52
Edital Prointer/UFC N° 02/2025 - UFC Mundo	98,90	0,00	0,00	0,00
Altera o Edital Prointer/UFC N° 02/2025 - UFC Mundo	99,38	0,00	0,00	0,00
Edital Prointer/UFC N° 03/2025 - PIBI	100,00	0,00	0,00	0,00
Edital Prointer/UFC N° 08/2025 - Bolsas de Internacionalização	99,17	0,00	0,00	0,00
<b>Média</b>	98,967	13,79	56,67	25,44

Fonte: Elaborado pelo autor.

Nota: Valores indefinidos foram considerados 0.

Nota: Não foram encontrados termos estrangeiros no texto do Edital Prointer/UFC N° 03/2025 - PIBI.

blocos de até 510 *tokens*. O tamanho do modelo era de 512 *tokens*, mas toda sentença precisava de dois tokens especiais [CLS] e [SEP], por isso o tamanho máximo era de 510 *tokens*.

Como todos os documentos possuíam mais de 512 *tokens*, a comparação foi feita em pares de blocos, e o resultado final de um documento correspondia à média entre os valores medidos em cada par. Nos casos onde um documento possuía mais blocos que outro, a comparação considerava o máximo de pares possíveis.

Na Tabela 5 estão os valores de precisão, revocação e *BERTScore* médios entre todos os documentos da primeira iteração para cada modelo. Dois modelos abertos tiveram resultados superiores ao Gemini: o Gemma3:4b obteve resultados melhores do que as duas versões e o modelo cow/gemma2\_tools:2b superou a versão Flash.

## 5.5 Questionário com Público-Alvo

Na primeira rodada de questionários realizada, foram obtidas respostas de 5 pessoas vinculadas à UFC, 3 docentes e dois discentes de graduação ou Pós-graduação. Os editais simplificados foram selecionados a partir do modelo com melhor resultado nos índices de inteligibilidade, o Phi4, e os dois pares de editais comparados foram: Edital Prointer PIBI N° 04/2025 - Empreende UFC e Edital Prointer/UFC N° 09/2025 - PROMISAES.

Tabela 5 – Valores médios das métricas de similaridade entre todos os documentos para cada modelo, ordenados pelo *BERTScore*

Modelo	BERTScore	Revocação	Precisão
gemma3:4b	0,826	0,827	0,826
gemini-2.5-pro-preview-05-06	0,760	0,809	0,731
cow-gemma2_tools:2b	0,745	0,763	0,727
gemini-2.5-flash-preview-04-17	0,737	0,750	0,725
phi4:14.7b	0,728	0,733	0,723
llama3.2:3.2b	0,726	0,724	0,729
granite3-dense:8b	0,714	0,728	0,702
qwen2.5-14b	0,708	0,710	0,707
deepseek-r1:14b	0,677	0,681	0,674
phi3:3.8b	0,653	0,665	0,642
granite3-dense:2b	0,617	0,638	0,599

Fonte: Elaborada pelo autor.

O perfil de cada participante está detalhado na Tabela 6. O recurso com a descrição de todas as questões realizadas nessa etapa está presente no Figshare<sup>5</sup>. Nas próximas seções são descritos os resultados de cada uma das questões de pesquisa do questionário presentes na Seção 4.5.

Tabela 6 – Perfis dos participantes do questionário *online*.

Pergunta	P. 1	P. 2	P. 3	P. 4	P. 5
Vínculo com a universidade	Docente	Docente	Discente de Graduação ou Pós-graduação	Docente	Discente de Graduação ou Pós-graduação
Com que frequência você lê editais públicos	Sempre	Sempre	Raramente	Ocasionalmente	Raramente
Qual sua experiência na escrita, revisão e divulgação de editais públicos?	Participo Frequentemente	Participo Muito Frequentemente	Raramente participo	Nunca participei	Participo Ocasionalmente

Fonte: Elaborada pelo autor.

### 5.5.1 SRQ1 - Quão Fácil é a Compreensão do Documento Simplificado em Comparação com o Documento Original?

As respostas para a **SRQ1** foram, em sua maioria, positivas. Para o primeiro edital, 4 dos 5 participantes afirmaram que a versão simplificada por IAG era ligeiramente mais legível do que a versão original, e um dos participantes afirmou ser muito mais legível. No segundo edital, dois participantes responderam "Muito mais clara", dois responderam "Ligeiramente mais

<sup>5</sup> <https://figshare.com/s/56c1234fe972ccd0ecc7>

clara" e um respondeu "Sem diferença significativa". Essas avaliações positivas mostram que o objetivo principal da simplificação foi atingido, mas a presença de uma resposta neutra e o fato de que a maior parte das respostas descreviam apenas uma pequena melhoria indicam que o processo ainda não alcançou um efeito expressivo na percepção dos usuários.

#### ***5.5.2 SRQ2 - Qual a utilidade da versão gerada automaticamente para apoiar uma rápida compreensão das principais informações?***

Para a **SRQ2**, as opiniões foram mais divididas: P3 concordou completamente que a versão simplificada era mais útil para leitores que precisam encontrar as informações mais relevantes rapidamente; P3 e P4 concordaram parcialmente; P5 se manteve neutro nos dois editais; e P2 discordou parcialmente para o primeiro edital e manteve opinião neutra para o segundo.

#### ***5.5.3 SRQ3 - Você considera que o documento simplificado está pronto para ser publicado como complemento da versão original?***

Em relação à prontidão das versões simplificadas que era o tema da **SRQ3**, P3 e P5 consideraram os dois editais como prontos; P4 afirmou que o primeiro edital era útil como complemento, mas não como substituto do original, e que o segundo edital simplificado simplificava a maior parte do conteúdo, mas uma versão completa ainda seria necessária; P1 discordou para os dois editais, no caso do primeiro afirmou-se que ele poderia ser ainda mais simples e visualmente atrativo, já no segundo foram destacados os mesmos pontos, com a adição de uma sentença encontrada com informações incorretas; por fim, P2 discordou da afirmação nos dois editais, para o primeiro o motivo estava no fato de não possuir elementos gráficos para as informações mais relevantes ou uma estrutura de tópicos adequada, já o segundo edital foi identificado como idêntico à sua versão simplificada.

#### ***5.5.4 SRQ4 - Que aspectos de ambas as versões simplificadas podem melhorar a compreensão por parte do seu público-alvo?***

Na **SRQ4**, os participantes eram solicitados a destacar quais características dos textos que facilitassem a compreensão foram encontrados nos dois editais simplificados, mas a maior parte das respostas tratavam de características processo de simplificação textual em

geral, e não dos resultados encontrados nos dois editais apresentados. P3 destacou o rápido entendimento; P4 pontuou que o texto simplificado tende a ser mais objetivo e portanto mais fluido e menor; P5 afirmou que textos menores tem menos chances de serem mal interpretados. Dois dos participantes responderam exclusivamente sobre os dois documentos apresentados como era esperado: P1 afirmou que uma simplificação podia ser observada, mas ainda poderia melhorar; e P2 sugeriu estruturas para uma melhor simplificação a nível de discurso: o uso de elementos gráficos, links clicáveis, divisões de tópicos no texto e responsividade que permita a visualização também em redes sociais.

#### **5.5.5 SRQ5 - *Você identificou algum risco ou limitação no conteúdo destas versões simplificadas?***

Por fim, a questão 5 está relacionada aos riscos e/ou limitações percebidas pelos participantes em relação ao conteúdo da versão simplificada. Os participantes P3 e P4 não identificaram nenhum risco, mas P4 afirmou que algumas seções de textos foram mais bem simplificadas nos dois documentos simplificados; P5 respondeu que alguns detalhes poderiam ser excluídos pelos documentos por IAG; P1 identificou um erro em uma sentença do item 6.3 no primeiro edital; por fim, P2 destacou que o primeiro edital poderia não cumprir o papel de facilitar a compreensão dos usuários, e o segundo edital (referindo-se à versão sumarizada) pode passar a sensação de que o edital completo não é necessário.

### **5.6 Aprimoramento da Geração de Textos**

Partindo das avaliações qualitativas, foi identificada a necessidade de aprimorar o processo de simplificação textual para também realizar a sumarização dos documentos, considerados como excessivamente extensos.

Dessa forma, os documentos simplificados originais não passaram pelo processo de reescrita pela IAG, em vez disso eles foram utilizados como padrão de referência para avaliar as simplificações + sumarizações realizadas.

As mudanças realizadas no processo de simplificação incluíram:

- a) Uso de funcionalidade de *Chat* em vez da geração de texto pura, incluindo uma lógica para manter apenas um certo número de mensagens para não preencher a janela de contexto

- b) Alterações no processo de quebrar o texto alvo em múltiplas partes
- c) Mudança na temperatura utilizada
- d) Reformulação do *prompt* passado para os modelos

Em relação ao processo de análise morfossintática, o modelo PortParser, definido em Lopes e Pardo (2024), foi utilizado na sua segunda versão. Esse modelo atingiu resultados melhores do que o UDPipe2 treinado no *dataset* Porttinari (utilizado na primeira avaliação) nos *benchmarks* realizados pelos autores.

Também foram adicionadas mais etapas para a limpeza dos dados com base nos resultados anteriores, onde a interpretação de uma sentença era prejudicada por elementos *Markdown* que deixaram de ser removidos. As novas etapas de pré-processamento adicionadas foram:

- a) Remoção de legendas de imagens duplicadas como descrito na Subseção 5.7.1.
- b) Remoção de referências de imagens, no formato `![texto alternativo](url)`.
- c) Substituição dos links com texto alternativo para utilizar apenas o texto alternativo presente na expressão `[texto alternativo](url)`.
- d) Remoção de *tags* HTML.

Essas etapas foram realizadas nesta ordem, após a remoção de caracteres especiais da sintaxe *Markdown* e antes da remoção de espaços em branco extras.

A métrica *foreign word ratio* também foi aprimorada para considerar lemas em vez de *types*, principalmente por que o PortParser V2 conseguiu diferenciar o lema de algumas palavras estrangeiras, algo que o UDPipe2 não conseguiu.

## 5.7 Análise Final dos Resultados

### 5.7.1 Extração dos Documentos

Na segunda iteração, os editais públicos disponibilizados publicamente no site da UFC Inova<sup>6</sup> foram extraídos manualmente, considerando apenas documentos que possuísem uma versão simplificada. Por isso, publicações de resultados e alterações e anexos de editais não foram considerados. No caso do edital N° 01 2025 do I Hackathon TRE/CE – UFC s retificações foram publicadas em uma nova versão do documento completo, por isso apenas a versão retificada foi considerada.

<sup>6</sup> <https://ufcinova.ufc.br/pt/editais-e-resultados/>

Os editais do programa Inovando UFC dos anos de 2022 até 2025 foram publicados apenas na versão aplicando a linguagem simples, por isso não foram considerados já que não poderiam servir como um ponto de partida para a simplificação do modelo e um resultado a ser comparado ao mesmo tempo.

O conjunto de documentos selecionados foi de 7 pares de documentos, com 3 pares de documentos em 2021, 1 em 2022, 1 em 2023 e 2 em 2025, somados com o restante dos documentos de 2025, à exceção do edital do Hackaton Inovando UFC 2025. Totalizando 7 documentos com um simplificação de referência e 3 apenas com o texto original.

A conversão do documentos em PDF para o formato *Markdown* foi realizada com o modelo Chandra OCR por meio da interface *web*, as configurações utilizadas foram: modo de uso balanceado, conversão em lotes de até 10 páginas, exclusão de cabeçalhos e rodapés e preservação dos *hyperlinks* e imagens presentes no PDF.

Um ponto negativo da conversão com o Chandra OCR estava na conversão das imagens para o formato *Markdown*, onde o texto alternativo era duplicado no documento, em vez de aparecer apenas na declaração de *hyperlink* no formato `![texto alternativo](url)`. Um exemplo desse resultado está mostrado no Código-fonte 1. Durante a análise das métricas esse texto foi eliminado com o uso de expressões regulares.

### 5.7.2 Geração de Documentos Simplificados com Diferentes LLMs

Após o aprimoramento, o processo de simplificação foi realizado através da plataforma OpenRouter<sup>7</sup>, devido a incompatibilidade da máquina utilizada na primeira etapa com as funcionalidades de *Chat* da biblioteca Langchain.

Parte dos modelos não estava disponível nessa plataforma, por isso novos modelos foram considerados pensando em uma distribuição mais homogênea na quantidade de parâmetros. Os novos modelos considerados foram:

- a) Gemini 2.5 Pro
- b) Gemini 2.5 Flash
- c) Gemma-3n:4b
- d) Phi4:14b
- e) Llama-3.2-instruct:3b
- f) Qwen3:14b

---

<sup>7</sup> <https://openrouter.ai/>

- g) Granite-4.0-micro:4b
- h) Deepseek-v3.2:670b
- i) Ministral:8b

Como o Deepseek é o maior dentre os modelos abertos, o resultado esperado é de que sua performance seja a mais próxima dos modelos Gemini.

A nova configuração utilizada passava inicialmente um *prompt* de sistema presente na Seção A.2 e mantinha no máximo 4 mensagens do usuário no contexto, de forma que a janela de contexto não fosse transbordada. A temperatura dos modelos nessa etapa foi definida para 1, visto que a sumarização abstrativa permite mais liberdade no texto que será gerado.

No novo processo de segmentação, o texto alvo era quebrado primeiro com bases nas marcações de seções com o *MarkdownHeaderTextSplitter* do Langchain, após essa separação por contextos, os textos menores eram combinados até terem *MarkdownTextSplitter*, mas utilizando um máximo de 8000 caracteres, finalmente, seções com mais de 8000 caracteres eram quebradas utilizando o *RecursiveCharacterTextSplitter*, também com 8000 caracteres como limite. A escolha desse limite se deu com base na menor janela de contexto entre os modelos, 16384 tokens para o Phi4, e na afirmação de que um *token* corresponde a, aproximadamente, 4 caracteres<sup>8</sup>. Com isso, o máximo de 4 mensagens mantidas no contexto ocuparia 32000 caracteres ( $\approx$  8000 tokens), fazendo com que a janela de contexto não seria extravasada, mesmo que as respostas do modelo fossem do mesmo tamanho do texto passado.

O processo de simplificação textual aprimorado está descrito na Figura 9, nela tem-se o início do processo com a extração dos documentos e sua divisão em completo e resumido, ambas as categorias são convertidas para *Markdown* e simplificadas pelos LLMs, a formatação das versões originais e simplificadas por IAG é homogeneizada e então são gerados os documentos em PDF.

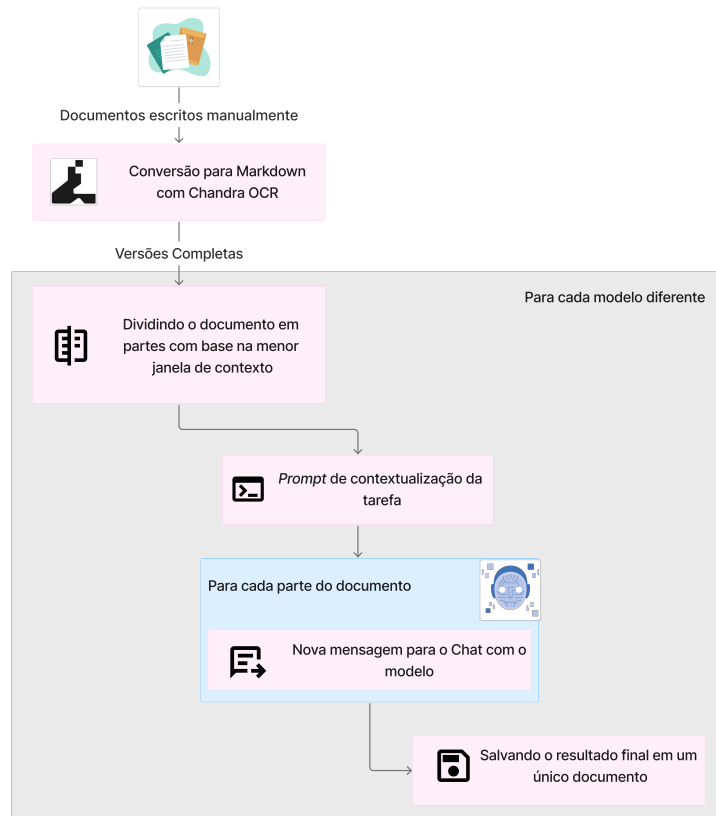
### 5.7.3 Avaliação com as métricas

O processo de avaliação das métricas teve de ser alterado após a mudança para o PortParser V2, como ele não realiza a segmentação do texto automaticamente, um arquivo .conllu deve ser fornecido para o modelo, em vez de um arquivo de texto plano.

Por isso, na segunda avaliação das métricas, após a remoção dos elementos não considerados, o texto é convertido para o formato Conllu em duas etapas: primeiro, as sentenças

<sup>8</sup> <https://platform.openai.com/tokenizer>

Figura 9 – Fluxograma do processo de simplificação textual aprimorado.



Fonte: Elaborada pelo autor.

são ordenadas em linhas separadas através da ferramenta PortSentencer<sup>9</sup>; o segundo passo é tokenizar as palavras de cada sentença, utilizando o projeto PortTokenizer<sup>10</sup>.

O modelo PortParser V2 não conseguiu realizar as anotações sintáticas de 5 dos documentos gerado pela IAG, mesmo com uma etapa de pós-processamento extra que removia os caracteres *Unicode* que não existem no Português (e estavam presentes em resultados com alucinações). Os documentos que não foram mostrados estão presentes no Quadro 3.

Quadro 3 – Documentos onde a anotação sintática com o PortParser V2 falhou

Documento simplificado	Modelo	Possuía alucinações	Possuía caracteres inexistentes em Português
Edital PIBI/UFC N° 02/2021	Llama 3.2	Sim	Sim
Edital PROINTER PIBI N° 02/2022	Llama 3.2	Não	Não
Edital Prointer PIBI N° 07/2023	Llama 3.2	Sim	Sim
Edital PIBI/UFC N° 02/2021	Gemini 2.5 Flash	Sim	Não
Edital PROINTER PIBI N° 02/2022	Granite 4	Não	Não

Fonte: Elaborado pelo autor.

<sup>9</sup> <https://github.com/LuceleneL/portTokenizer>

<sup>10</sup> <https://github.com/LuceleneL/portSentencer>

### 5.7.3.1 Índices de Inteligibilidade

A implementação dos índices de inteligibilidade não sofreu alterações após o aprimoramento. Os resultados dos índices de inteligibilidade para a segunda iteração das avaliações das métricas estão descritos na Tabela 7.

Nela é possível ver que os modelos Gemini 2.5 Pro e Deepseek v3.2 tiveram valores superiores aos editais completos e simplificados originais, mostrando que, no aspecto de tamanho das frases e palavras, eles superaram os textos de referência. Além disso, o modelo Gemma3, de apenas 4 bilhões de parâmetros teve resultados melhores do que o Gemini 2.5 Flash, o Qwen3, com 14 bilhões teve resultados levemente inferiores ao Gemini Flash, mas também superou os valores medidos no editais completos em 4 das 6 métricas.

Apenas três dos modelos não tiveram diferença estatística significativa em relação aos documentos completos originais, o Phi4 apresentou um intervalo pior para a métrica Coleman Liau, mas em todos os outros casos o intervalo de confiança tinha interseção com os intervalos originais, indicando que a diferença da simplificação não foi significativa nesses casos.

Tabela 7 – Valores médio dos índices de inteligibilidade em todos os documentos da segunda iteração com intervalo de confiança de 95%

Fonte	FE↑	IG↑	FK↓	ARI↓	GF↓	CL↓
Completos	43,72 (40,53–46,90)	43,75 (42,27–45,23)	12,74 (11,81–13,66)	16,77 (15,86–17,68)	12,76 (11,42–14,11)	17,29 (16,88–17,70)
Simplificados	46,81 (41,81–51,80)	44,82 (40,63–49,01)	12,81 (10,74–14,88)	17,15 (14,36–19,94)	15,37 (11,78–18,96)	16,49 (15,49–17,49)
gemini-2.5-pro	<b>59,24</b> <b>(57,35–61,13)</b>	<b>62,69</b> <b>(60,32–65,06)</b>	<b>8,27</b> <b>(7,98–8,56)</b>	<b>10,07</b> <b>(9,76–10,38)</b>	<b>7,28</b> <b>(6,65–7,91)</b>	<b>13,89</b> <b>(13,54–14,25)</b>
deepseek-v3.2	<b>54,15</b> <b>(48,85–59,44)</b>	<b>57,71</b> <b>(53,81–61,61)</b>	<b>9,21</b> <b>(8,25–10,16)</b>	<b>11,65</b> <b>(10,47–12,84)</b>	<b>8,18</b> <b>(7,27–9,09)</b>	<b>15,42</b> <b>(14,37–16,48)</b>
gemma3	49,83 (44,79–54,86)	<b>63,50</b> <b>(59,36–67,64)</b>	<b>9,29</b> <b>(8,54–10,05)</b>	<b>11,84</b> <b>(11,18–12,50)</b>	<b>7,23</b> <b>(6,20–8,25)</b>	16,36 (15,23–17,49)
gemini-2.5-flash	50,62 (46,62–54,62)	<b>57,16</b> <b>(49,35–64,97)</b>	<b>9,83</b> <b>(8,66–11,00)</b>	<b>12,49</b> <b>(10,72–14,27)</b>	<b>8,58</b> <b>(6,79–10,38)</b>	16,08 (15,04–17,11)
qwen3	49,75 (45,18–54,32)	<b>57,07</b> <b>(53,98–60,16)</b>	<b>9,64</b> <b>(8,83–10,44)</b>	<b>12,63</b> <b>(11,81–13,45)</b>	<b>8,02</b> <b>(7,36–8,68)</b>	16,91 (16,15–17,67)
ministral	45,05 (41,59–48,51)	<b>48,97</b> <b>(45,95–51,99)</b>	11,09 (10,32–11,86)	<b>14,64</b> <b>(13,70–15,57)</b>	<b>9,13</b> <b>(7,95–10,31)</b>	17,86 (17,11–18,62)
granite-4.0	43,71 (34,89–52,53)	47,88 (41,64–54,11)	12,48 (9,93–15,04)	15,32 (12,42–18,22)	12,41 (8,96–15,87)	16,07 (14,85–17,29)
llama-3.2	41,62 (39,99–43,25)	45,75 (43,44–48,06)	12,66 (11,93–13,40)	15,62 (14,36–16,88)	12,14 (10,71–13,56)	16,78 (16,37–17,20)
phi-4	41,45 (34,09–48,81)	47,90 (41,74–54,05)	11,46 (10,01–12,92)	15,54 (13,81–17,27)	12,36 (10,10–14,62)	19,17 (17,99–20,36)

Fonte: Elaborada pelo autor.

Nota: A seta para cima indica que uma métrica é diretamente proporcional à complexidade e por isso valores menores são melhores, a seta para baixo indica métricas inversamente proporcionais, onde valores maiores são melhores

Nota: Valores em negrito obtiveram um intervalo de frequência superior ao dos documentos completos

Nota: FE = índice de inteligibilidade de Flesch; FK = índice de escolaridade Flesch-Kincaid; IG = Índice Gulpease; ARI - Índice de Inteligibilidade Automatizado; GF - Índice de Nebulosidade de Gunning; CL - Índice de Coleman-Liau

### 5.7.3.2 Métricas Morfossintáticas

A implementação das métricas morfossintáticas nessa etapa foi a mesma da primeira iteração, com exceção da métrica *foreign word ratio*, que passou a considerar lemas em vez de *types*. A Tabela 8 mostra os valores médios nas duas classes de documentos originais e em cada um dos modelos. Nela é possível notar que o Gemini 2.5 Pro atingiu os melhores resultados e foi o único modelo a obter um intervalo melhor para a métrica de pronomes pessoais, e portanto gerou os textos com mais elementos que trazem sensação de proximidade para o leitor.

7 dos 9 modelos conseguiram um intervalo superior para a métrica *long sentence ratio* e, destes 7, outros 5 também obtiveram um intervalo melhor na métrica *words before main verb*. Isso indica que a simplificação realizada conseguiu reduzir efetivamente a estrutura de frases no texto.

Entretanto, todos os modelos, exceto o Phi4, tiveram um aumento na correferência por pronomes do caso reto. Uma possível causa para isso é que as frases mais curtas exigem mais referências das frases vizinhas para transmitir a informação que antes estava centralizada em uma única frase mais extensa.

Um ponto importante a destacar é a proporção de termos estrangeiros elevada no modelo Phi4, isso se deu ao fato da simplificação realizada do EDITAL N° 12/2021 - Hackathon Inovando UFC estar praticamente inteira em inglês, a exceção das duas seções iniciais e a última seção, e a presença dessa anomalia em um documento aumentou a média geral do modelo. Nenhum dos modelos conseguiu um intervalo de confiança melhor do que os originais, mas a menor intervalo foi o Gemma3.

#### 5.7.3.2.1 Eficiência da métrica proposta - *foreign word ratio*

A eficiência da métrica proposta neste trabalho foi testada novamente no segundo conjunto de dados. No total foram identificadas 241 palavras estrangeiras entre todos os documentos. Nesta segunda avaliação as palavras *campus* e *alumni* não foram consideradas como estrangeiras.

A Tabela 9 mostra os resultados de acurácia, revocação, precisão e o *Score F1* nos 7 editais completos e suas versões simplificadas, as medidas de acurácia, revocação e *score F1* foram maiores do que os valores medidos na Tabela 4, apenas a precisão que foi 0,09% menor do que a medida anteriormente.

Tabela 8 – Valores médios para todos os documentos da segunda iteração nas métricas morfossintáticas com intervalo de confiança de 95%

Fonte	PP*↑	CP↓	DP↓	NS↓	PV↓	WB↓	LS↓	FW*↓
Completos	,12 (,04–,20)	0	1,65 (1,23–2,07)	,06 (,0362–,08)	,23 (,15–,30)	8,60 (5,86–11,35)	,42 (,34–,51)	1,15 (,50–1,80)
Simplificados	,31 (,08–,54)	0	1,60 (,25–2,95)	,09 (,01–,16)	,20 (,11–,29)	11,84 (5,90–17,77)	,38 (,25–,51)	,93 (,12–1,74)
gemini-2.5-pro	<b>,69</b> ( <b>,50–,87</b> )	1,62 (1,22–2,02)	1,07 (,75–1,39)	,04 (,03–,06)	<b>,12</b> ( <b>,10–,14</b> )	<b>3,66</b> ( <b>3,24–4,09</b> )	<b>,09</b> ( <b>,06–,13</b> )	,86 (,59–1,14)
qwen3	,14 (,07–,20)	,57 (-,33–1,47)	1,16 (,91–1,41)	<b>,02</b> ( <b>,01–,03</b> )	<b>,09</b> ( <b>,07–,11</b> )	<b>3,87</b> ( <b>2,94–4,80</b> )	<b>,09</b> ( <b>,06–,12</b> )	,98 (,62–1,33)
gemini-2.5-flash <sup>-1</sup>	,29 (,04–,54)	,94 (,56–1,32)	1,11 (,51–1,72)	,03 (,01–,04)	,13 (,09–,17)	<b>3,70</b> ( <b>2,52–4,88</b> )	<b>,15</b> ( <b>,03–,28</b> )	,70 (,41–,98)
deepseek-v3.2	,25 (,15–,36)	1,21 (,93–1,49)	1,21 (,82–1,59)	,03 (,02–,04)	,13 (,10–,16)	<b>3,88</b> ( <b>3,56–4,20</b> )	<b>,12</b> ( <b>,08–,17</b> )	,79 (,37–1,20)
gemma3	,32 (,05–,59)	,90 (-,18–1,99)	1,17 (,51–1,83)	,02 (-,01–,06)	,09 (,05–,13)	<b>2,83</b> ( <b>1,70–3,96</b> )	<b>,08</b> ( <b>,04–,12</b> )	,57 (,28–,86)
llama-3.2 <sup>-4</sup>	,20 (-,06–,45)	1,25 (-,14–3,64)	2,50 (,85–4,15)	,04 (,00–,08)	,23 (,21–,25)	5,63 (2,90–8,36)	,36 (,26–,47)	,68 (-,17–1,52)
phi4	,11 (,04–,18)	0	1,46 (,69–2,24)	,03 (,01–,04)	<b>,09</b> ( <b>,06–,12</b> )	6,24 (5,13–7,35)	<b>,21</b> ( <b>,14–,28</b> )	23,03 (7,13–38,93)
ministral	,12 (,03–,20)	,29 (-,41–,98)	1,18 (,57–1,78)	,03 (,01–,06)	,19 (,11–,27)	6,96 (3,01–10,92)	<b>,24</b> ( <b>,16–,33</b> )	,64 (,32–,96)
granite-4.0 <sup>-1</sup>	,14 (-,06–,34)	,50 (-,38–1,38)	1,83 (,65–3,01)	,02 (,00–,0364)	,23 (,13–,34)	5,45 (4,00–6,90)	,43 (,22–,65)	,58 (,24–,92)

Fonte: Elaborada pelo autor.

Nota: Os valores das métricas **PP** e **FW** foram multiplicados por 100 para serem diferenciáveis com 2 dígitos decimais, as colunas afetadas estão marcadas com um asterisco.

Nota: Valores em negrito estavam em um intervalo melhor do que o intervalo dos documentos completos

Nota: A seta para cima indica que uma métrica é inversamente proporcional à complexidade e por isso valores maiores são melhores, a seta para baixo indica métricas inversamente proporcionais, onde valores menores são melhores

Nota: Os modelos Llama3.2, Granite 4.0 e Gemini 2.5 Flash tiveram 4, 1 e 1 documentos que não puderam ser analisados, esses modelos estão marcados com um <sup>-</sup> e a quantidade de documentos faltando ao lado do nome deles

Nota: CP = *Coreference Pronoun Ratio*; DP = *Demonstrative Pronoun Ratio*; NS = *Non SVO Ratio*; PV = *Passive Voice Ratio*; WB = *Words before main verb*; LS = *Long Sentence Ratio*; FW = *Foreign Word Ratio*; PP = *Personal Pronoun Ratio*.

### 5.7.3.3 Similaridade Semântica

Na segunda análise similaridade semântica, os documentos gerados simplificados pela IAG foram comparados com as versões simplificadas que já existiam para cada modelo. Os resultados estão presentes na Tabela 10 e, diferente da primeira rodada de simplificação, nenhum modelo atingiu um BERTScore médio superior a 0,8.

Tabela 9 – Precisão, Revocação, Acurácia e *Score F1* da métrica *foreign word ratio* para o segundo conjunto de dados considerado,

Documento	Acurácia (%)		Revocação (%)		Precisão (%)		Score F1 (%)	
	Comp.	Simp.	Comp.	Simp.	Comp.	Simp.	Comp.	Simp.
Edital Printner PIBI N° 04/2025	99,79	99,62	56,25	36,36	60,0	80,0	58,06	50,0
Edital Printner PIBI N° 07/2023	99,67	99,57	51,72	33,33	65,22	100,0	57,69	50,0
EDITAL N° 12/2021	99,81	99,51	38,46	25,0	100,0	50,0	55,56	33,33
Edital PIBI/UFC N° 02/2021	99,23	99,73	17,39	50,0	75,0	50,0	28,24	50,0
Edital Conjunto UFC - TRE/CE N° 01/2025	99,67	98,84	23,53	9,09	50,0	100,0	32,0	16,67
Edital PIBI N° 03/2021	100,00	100,00	0,00	0,0	0,00	0,0	0,00	0,0
Edital PROINTER PIBI N° 02/2022	99,23	98,38	23,64	0,0	61,9	0,0	34,21	0,0
<b>Média</b>	99,63	99,29	30,14	22,0	58,87	54,29	38,0	28,57
	99,46		26,07		56,58		33,29	

Fonte: Elaborada pelo autor,

Nota: Comp. = Completo; Simp. = Simplificado

Nota: Não foram encontrados termos estrangeiros na versão completa e simplificada do Edital PIBI N° 03/2021,

Tabela 10 – Média e intervalo de confiança de 95% das métricas de similaridade para todos os documentos em cada modelo

Modelo	BERTScore	Precisão	Revocação
gemini-2.5-pro	,665 (,646– ,684)	,649 (,627 – ,672)	,682 (,665 – ,700)
gemma-3n-e4b-it	,665 (,646– ,683)	,652 (,634 – ,670)	,678 (,658 – ,698)
qwen3:14b	,663 (,642 – ,684)	,650 (,625 – ,675)	,677 (,658 – ,696)
deepseek-v3.2	,662 (,654– ,671)	,649 (,634 – ,663)	,677 (,664 – ,690)
gemini-2.5-flash	,659 (,639 – ,679)	,642 (,618 – ,666)	,677 (,660 – ,694)
ministral:8b	,658 (,636 – ,681)	,645 (,618 – ,673)	,673 (,652 – ,693)
granite-4.0-h-micro	,656 (,640– ,671)	,643 (,628 – ,658)	,669 (,652 – ,686)
llama-3.2:3b-instruct	,650 (,630– ,670)	,635 (,618 – ,654)	,666 (,640 – ,692)
phi-4	,638 (,622– ,654)	,634 (,613 – ,655)	,643 (,630 – ,656)

Fonte: Elaborada pelo autor.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho foi apresentado um processo de geração automática de textos simplificados com o uso de IAG para documentos administrativos da UFC Inova. A avaliação desses textos simplificados foi realizada com o uso de índices estatísticos de inteligibilidade, métricas morfossintáticas, métricas de similaridade semântica e uma avaliação humana em forma de questionário.

Na análise quantitativa dos textos gerados, foram selecionados 6 índices de inteligibilidade em conjunto com 8 métricas morfossintáticas. Os resultados das duas iterações mostram a competitividade dos modelos abertos e a eficiência da simplificação textual realizada pelos LLMs. Na primeira iteração os modelos Phi 4 e Qwen2.5 tiveram resultados próximos dos modelos Gemini e 5 modelos conseguiram resultados positivos na redução do tamanho das sentenças. E na segunda iteração, o modelo Deepseek conseguiu melhorias significativas em todos os índices de inteligibilidade, assim como o Gemini 2.5 Pro; os modelos Gemma3, Gemini 2.5 Flash e Qwen3 tiveram resultados positivos em 4 dos 6 índices, com o Gemma3 superando o Gemini Flash; finalmente, o Qwen3 conseguiu resultados próximos do Gemini 2.5 Pro nas métricas morfossintáticas e 7 dos 9 modelos obtiveram melhorias no tamanho das sentenças.

Também houveram contribuições relativas às métricas de cálculo de complexidade: as métricas morfossintáticas tiradas de Leal *et al.* (2024) foram reimplementadas utilizando anotações no formato universalizado Conllu, permitindo uma mesma avaliação para qualquer modelo de anotação sintática compatível com esse formato; outra contribuição foi a mudança de lógica das métricas *coreference pronoun ratio*, *demonstrative pronoun ratio* e *personal pronoun ratio* que passou a utilizar *features* da anotação sintática em vez do reconhecimento de unigramas, cobrindo mais casos com menor pré-processamento necessário; finalmente, uma nova métrica que avalia a quantidade de termos estrangeiros no texto foi incluída na análise, ampliando a avaliação de fatores relacionados à complexidade textual.

As limitações deste trabalho foram, principalmente, a amostra pequena de documentos possuindo uma versão simplificada que pudesse ser utilizada como referência, e a quantidade pequena de respostas na pesquisa com usuários realizada na primeira iteração. Os modelos UDPipe2 e PortParser V2 utilizados para a primeira análise sintática dos documentos também apresentaram limitações referentes ao processamento do texto, sendo incapazes de interpretar corretamente sentenças contendo pronomes com a forma feminina em parênteses, p. ex. “ele(a)”, algo que não pôde ser solucionado por acarretar mudanças em toda a árvore sintática da sentença.

O PortParser V2 também não foi capaz de processar 6 dos documentos gerados por IAG.

Para trabalhos futuros, objetiva-se aprimorar o processo de simplificação com mais etapas de extração de informações antes da criação dos *prompts* de simplificação, como a extração de entidades nomeadas que devem ser priorizadas na simplificação. Também pretende-se estruturar todas as etapas apresentadas de geração e avaliação dos textos em uma API JSON que permita tanto realizar a simplificação, quanto avaliar os textos gerados de forma interativa pelos usuários.

Em relação às métricas apresentadas, planeja-se aprimorar o cálculo delas para avaliações futuros: o BERTScore pode ser adaptado para utilizar a TF-IDF calculada a de todos os editais existentes e atribuir um peso maior para *tokens* de entidades recorrentes. A métrica *foreign word ratio* pode ser aprimorada para considerar a frequência dos termos estrangeiros na escala Zipf, atribuindo peso maior os termos de uso não corrente mais do que termos estrangeiros utilizados frequentemente no vocabulário brasileiro. Adicionalmente, métricas de imageabilidade e familiaridade presentes em (Leal *et al.*, 2024) também podem ser consideradas na avaliação, podendo indicar parágrafos no texto possíveis de serem descritos com imagens.

Finalmente, considera-se a utilização de metodologias *LLM-as-a-judge*, como o *framework Deepeval*<sup>1</sup>, para avaliação os textos de formar preliminar, identificando alucinações, verificando a aplicação dos princípios da linguagem simples, etc. Após essas etapas de avaliação pelos LLMs, os documentos com melhores resultados podem ser selecionados para uma nova rodada de questionários com o público-alvo.

---

<sup>1</sup> <https://deepeval.com/>

## REFERÊNCIAS

- ALMEIDA, P.; POZZOBON, L.; FIGUEIREDO, J.; RIGHINI, J.; ROEDEL, P.; DUARTE, A.; COSTA, L. í. S.; QUENTAL, C.; TABAK, S.; Fundação Oswaldo Cruz. **Simple assim** : comunique com todo mundo. [S. n.], 2024. ISBN 978-65-00-99100-0. Disponível em: <https://bibliotecadigital.mdh.gov.br/jspui/handle/192/13208>. Acesso em: 02 jul. 2025.
- ASTH, R. **Espelhos Planos: o que são e como as imagens se formam (com exemplos)**. Toda Matéria, [s.d.]. Disponível em: <https://www.todamateria.com.br/espelhos-planos/>. Acesso em: 30 dez. 2025.
- BARRÓN-CEDENO, A.; VILA, M.; MARTÍ, M. A.; ROSSO, P. Plagiarism Meets Paraphrasing: Insights for the Next Generation in Automatic Plagiarism Detection. **Computational Linguistics**, MIT Press, Cambridge, MA, v. 39, n. 4, p. 917–947, dez. 2013.
- BHAGAT, R.; HOVY, E. What Is a Paraphrase? **Computational Linguistics**, MIT Press, Cambridge, MA, v. 39, n. 3, p. 463–472, set. 2013. Acesso em: 02 jul. 2025.
- BRASIL. **Decreto Nº 6.583, de 29 de setembro de 2008**. 2008. Promulga o Acordo Ortográfico da Língua Portuguesa, assinado em Lisboa , em 16 de dezembro de 1990. Brasília, DF: Presidência da República. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2007-2010/2008/decreto/d6583.htm](https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/decreto/d6583.htm). Acesso em: 18 jul. 2025.
- BRASIL. **Constituição da República Federativa do Brasil de 1988**. 2025. Brasília, DF: Presidência da República. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm). Acesso em: 14 maio. 2025.
- BRASIL. **Lei nº 15.263, de 2025**. 2025. Institui a Política Nacional de Linguagem Simple nos órgãos e entidades da administração pública direta e indireta de todos os Poderes da União, dos Estados, do Distrito Federal e dos Municípios. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_Ato2023-2026/2025/Lei/L15263.htm](https://www.planalto.gov.br/ccivil_03/_Ato2023-2026/2025/Lei/L15263.htm). Acesso em: 19 nov. 2025.
- COLEMAN, M.; LIAU, T. L. A computer readability formula designed for machine scoring. **Journal of Applied Psychology**, v. 60, n. 2, p. 283–284, 1975. ISSN 1939-1854. Place: US Publisher: American Psychological Association.
- CONSELHO NACIONAL DE JUSTIÇA. **Pacto Nacional do Judiciário pela Linguagem Simple**. 2023. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2023/11/pacto-nacional-do-judiciario-pela-linguagem-simples.pdf>. Acesso em: 16 jun. 2025.
- CUESTA, A. M.; REYES, A.; ROSETH, B. The Importance of Clarity: Impacts of Colombia’s ‘Lenguaje Claro’ Program on Reducing Administrative Burdens. **IDB Publications**, Inter-American Development Bank, maio 2019. Disponível em: <https://doi.org/10.18235/0001683>. Acesso em: 07 out. 2025.
- DAY, S. L.; CIRICA, J.; CLAPP, S. R.; PENKOVA, V.; GIROUX, A. E.; BANTA, A.; BORDEAU, C.; MUTTENENI, P.; SAWYER, B. D. **Evaluating GenAI for Simplifying Texts for Education**: Improving Accuracy and Consistency for Enhanced Readability. arXiv, 2025. Disponível em: [doi.org/10.48550/arXiv.2501.09158](https://doi.org/10.48550/arXiv.2501.09158). Acesso em: 12 abr. 2025.

DE MARNEFFE, M.-C.; MANNING, C. D.; NIVRE, J.; ZEMAN, D. Universal Dependencies. **Computational Linguistics**, v. 47, n. 2, p. 255–308, jul. 2021. ISSN 0891-2017. Disponível em: [doi.org/10.1162/coli\\_a\\_00402](http://doi.org/10.1162/coli_a_00402). Acesso em: 26 maio. 2025.

DEVLIN, J.; CHANG, M.; LEE, K.; TOUTANOVA, K. BERT: Pre-training of deep bidirectional transformers for language understanding. **CoRR**, abs/1810.04805, 2018. Disponível em: <http://arxiv.org/abs/1810.04805>. Acesso em: 15 jan. 2025.

EASTERBROOK, S.; SINGER, J.; STOREY, M.; DAMIAN, D. Selecting empirical methods for software engineering research. In: SHULL, F.; SINGER, J.; SJØBERG, D. I. K. (Ed.). **Guide to Advanced Empirical Software Engineering**. London: Springer, 2008. p. 285–311. ISBN 978-1-84800-044-5.

FARAJIDIZAJI, A.; RAINA, V.; GALES, M. Is it possible to modify text to a target readability Level? an initial investigation using zero-shot large language models. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, LANGUAGE RESOURCES AND EVALUATION (LREC-COLING). **Proceedings [...]**. Torino, Itália, 2024. p. 9325–9339. Disponível em: <https://aclanthology.org/2024.lrec-main.815/>. Acesso em: 22 jun. 2025.

FÄRBER, M.; AGHDAM, P.; IM, K.; TAWFELIS, M.; GHOSHAL, H. Simplifymytext: An llm-based system for inclusive plain language text simplification. In: ADVANCES IN Information Retrieval. **Proceedings [...]**. Cham: Springer Nature Switzerland, 2025. p. 418–424. ISBN 978-3-031-88717-8. Acesso em: 17 jun. 2025.

FERREIRA, T. C.; PARABONI, I. Geração de linguagem natural. In: CASELI, H. M.; NUNES, M. G. V. (Ed.). **Processamento de Linguagem Natural: conceitos, técnicas e aplicações em Português**. 3. ed. BPLN, 2024. cap. 16. ISBN 978-65-01-20581-6. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/parte-geracao-modelos/cap-geracao-ln/cap-geracao-ln.html>. Acesso em: 19 nov. 2025.

FINATTO, M. J. B.; CASELI, H. d. M.; LOPES, L.; RASSI, A. Sequência de caracteres e palavras: Morfologia e morfossintaxe. In: CASELI, H. M.; NUNES, M. G. V. (Ed.). **Processamento de Linguagem Natural: conceitos, técnicas e aplicações em português**. 3. ed. BPLN, 2024. cap. 4. ISBN 978-65-01-20581-6. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/parte-palavras/cap-caracteres-palavras/cap-caracteres-palavras.html>. Acesso em: 30 jun. 2025.

FLESCH, R. A new readability yardstick. **Journal of Applied Psychology**, American Psychological Association, US, v. 32, n. 3, p. 221–233, 1948. ISSN 1939-1854.

FONSECA, E.; VANIN, A. A.; VIEIRA, R. Resolução de correferência. In: CASELI, H. M.; NUNES, M. G. V. (Ed.). **Processamento de Linguagem Natural: conceitos, técnicas e aplicações em português**. 3. ed. BPLN, 2024. cap. 12. ISBN 978-65-01-20581-6. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/parte-discurso/cap-resolucao-correferencia/cap-resolucao-correferencia.html>. Acesso em: 9 jul. 2025.

FREITAS, C. E o significado? In: CASELI, H. M.; NUNES, M. G. V. (Ed.). **Processamento de Linguagem Natural: conceitos, técnicas e aplicações em português**. 3. ed. BPLN, 2024. cap. 8. ISBN 978-65-01-20581-6. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/parte-significado/cap-significado/cap-significado.html>. Acesso em: 9 jul. 2025.

GIARELIS, N.; MASTROKOSTAS, C.; KARACAPILIDIS, N. Abstractive vs. Extractive Summarization: An Experimental Review. **Applied Sciences**, Multidisciplinary Digital Publishing Institute, v. 13, n. 13, p. 7620, jan. 2023. ISSN 2076-3417.

GUNNING, R. **The technique of clear writing**. Rev. ed. McGraw-Hill, 1968. Disponível em: <https://cir.nii.ac.jp/crid/1130000796418620544>. Acesso em: 14 maio. 2025.

HONNIBAL, M.; MONTANI, I.; LANDEGHEM, S. V.; BOYD, A. spaCy: Industrial-strength Natural Language Processing in Python. 2020.

INDICADOR DE ALFABETISMO FUNCIONAL. **INAF 2024 - Legado e Futuro do Alfabetismo Funcional**. 2024. Disponível em: <https://drive.google.com/file/d/1ywPKT3xn5XOL0AeNMgECjtSpq9vX6-kI/view>. Acesso em: 14 maio. 2025.

INSTITUTO FEDERAL DE MATO GROSSO. **Cartilha Orientativa sobre o Uso de Linguagem Simples no Contexto do Instituto Federal de Mato Grosso**. 2021. Disponível em: [https://ifmt.edu.br/media/filer\\_public/38/12/38122512-8c1d-43c4-9044-7c99c23d79ff/cartilha\\_orientativa\\_eu\\_uso\\_versao\\_final.pdf](https://ifmt.edu.br/media/filer_public/38/12/38122512-8c1d-43c4-9044-7c99c23d79ff/cartilha_orientativa_eu_uso_versao_final.pdf). Acesso em: 08 ago. 2025.

INSTITUTO FEDERAL DO PERNAMBUCO. **Guia para Comunicação Interna e Externa do IFPE**. 2023. Disponível em: [https://drive.google.com/file/d/1OgpnRQaioywzc00NQ1\\_Yh6iPitXsB88m/view?usp=drive\\_link&usp=embed\\_facebook](https://drive.google.com/file/d/1OgpnRQaioywzc00NQ1_Yh6iPitXsB88m/view?usp=drive_link&usp=embed_facebook). Acesso em: 08 ago. 2025.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 24495-1 Plain language Part 1: Governing principles and guidelines**. [S. l.], 2023. Disponível em: <https://www.iso.org/standard/78907.html>. Acesso em: 30 jun. 2025.

INTERNATIONAL PLAIN LANGUAGE FEDERATION. **Plain Language Definitions - International Plain Language Federation**. 2019. Disponível em: <https://www.iplfederation.org/plain-language-definitions/>. Acesso em: 06 jun. 2025.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition, with language models**. 3rd. ed. [S. n.], 2026. Manuscrito publicado online 6 Jan. 2026. Disponível em: <https://web.stanford.edu/~jurafsky/slp3/>. Acesso em: 09 jan. 2025.

KINCAID, J. P.; FISHBURNE, J.; P., R. R.; L., C. R.; Brad S. **Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel**. Fort Belvoir, Virgínia, EUA, 1975. Disponível em: <https://apps.dtic.mil/sti/citations/tr/ADA006655>. Acesso em: 27 jun. 2025.

KITCHENHAM, B. A.; PFLEEGER, S. L. Personal Opinion Surveys. In: **Guide to Advanced Empirical Software Engineering**. London: Springer London, 2008. p. 63–92. ISBN 9781848000438. Disponível em: [http://link.springer.com/10.1007/978-1-84800-044-5\\_3](http://link.springer.com/10.1007/978-1-84800-044-5_3). Acesso em: 10 jun. 2025.

LEAL, S. E.; ALUÍSIO, S. M. Complexidade textual e suas tarefas relacionadas. In: CASELI, H. M.; NUNES, M. G. V. (Ed.). **Processamento de Linguagem Natural: conceitos, técnicas e aplicações em português**. 3. ed. BPLN, 2024. cap. 25. ISBN 978-65-01-20581-6. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/>. Acesso em: 28 jun. 2025.

LEAL, S. E.; DURAN, M. S.; SCARTON, C. E.; HARTMANN, N. S.; ALUÍSIO, S. M. Nilc-matrix: assessing the complexity of written and spoken language in Brazilian Portuguese. **Language Resources and Evaluation**, Springer, v. 58, n. 1, p. 73–110, 2024. Disponível em: <https://doi.org/10.1007/s10579-023-09693-w>. Acesso em: 01 maio. 2025.

LIN, C.-Y.; HOVY, E. Automatic Evaluation of Summaries Using N-gram Co-occurrence Statistics. In: HUMAN LANGUAGE TECHNOLOGY CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings [...]**. 2003. p. 150–157. Disponível em: <https://aclanthology.org/N03-1020/>. Acesso em: 17 jan. 2026.

LINGUATECA. **Acesso a Corpos**: Corpo todos juntos. 2025. Disponível em: <https://www.linguateca.pt/acesso/corpus.php?corpus=TODO>. Acesso em: 10 maio. 2025.

LOPES, L.; PARDO, T. Towards Portparser - a highly accurate parsing system for Brazilian Portuguese following the Universal Dependencies framework. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL PROCESSING OF PORTUGUESE - VOL. 1, 16. **Proceedings [...]**. Santiago de Compostela, Galícia/Espanha: Association for Computational Linguistics, 2024. p. 401–410. Disponível em: <https://aclanthology.org/2024.propor-1.41/>. Acesso em: 01 jun. 2025.

LUCISANO, P.; PIEMONTESE, M. E. Gulpease: una formula per la predizione della leggibilità di testi in lingua italiana. **Scuola e Città**, p. 110–124, 1988. ISSN 0036-9853. Disponível em: <https://iris.uniroma1.it/handle/11573/450554?mode=complete>. Acesso em: 27 jun. 2025.

MACFARLANE, J.; KREWINKEL, A.; ROSENTHAL, J. **Pandoc**. 2006. Disponível em: <https://github.com/jgm/pandoc>. Acesso em: 9 jul. 2025.

MARTINS, H. T.; SILVA, A. R. da; CAVALCANTI, M. T. Linguagem Simples: Um movimento social por transparência, cidadania e acessibilidade. **Cadernos do Desenvolvimento Fluminense**, n. 25, dez. 2023. ISSN 2317-6539. Disponível em: [doi.org/10.12957/cdf.2023.72869](https://doi.org/10.12957/cdf.2023.72869). Acesso em: 26 abr. 2025.

MONTY Python. **Monty Python and the Holy Grail**. Reino Unido: EMI Films, 1975. Direção: Terry Gilliam, Jones Terry. Produção: Mark Forstater, John Goldstone. 1 DVD (90 min).

MORENO, G. C. d. L.; SOUZA, M. P. M. de; HEIN, N.; HEIN, A. K. ALT: Um Software para Análise de Legibilidade de Textos em Língua Portuguesa. **Policromias - Revista de Estudos do Discurso, Imagem e Som**, v. 8, n. 1, p. 91–128, jun. 2023. ISSN 2448-2935. Disponível em: [doi.org/10.61358/policromias.v8i1.54352](https://doi.org/10.61358/policromias.v8i1.54352). Acesso em: 09 maio. 2025.

PAES, A.; VIANNA, D.; RODRIGUES, J. Modelos de linguagem. In: CASELI, H. M.; NUNES, M. G. V. (Ed.). **Processamento de Linguagem Natural**: conceitos, técnicas e aplicações em português. 3. ed. BPLN, 2024. cap. 17. ISBN 978-65-01-20581-6. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/parte-geracao-modelos/cap-modelos-linguagem/cap-modelos-linguagem.html>. Acesso em: 9 jul. 2025.

PAPINENI, K.; ROUKOS, S.; WARD, T.; ZHU, W.-J. Bleu: a Method for Automatic Evaluation of Machine Translation. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL Linguistics. **Proceedings [...]**. Filadélfia, Pennsylvania, EUA: Association for Computational Linguistics, 2002. p. 311–318. Disponível em: <https://aclanthology.org/P02-1040/>. Acesso em: 02 jul. 2025.

PARDO, T. A. S.; DURAN, M. S.; LOPES, L.; FELIPPO, A. D.; ROMAN, N. T.; NUNES, M. d. G. V. Porttinari - a Large Multi-genre Treebank for Brazilian Portuguese. In: SIMPÓSIO BRASILEIRO DE TECNOLOGIA DA INFORMACÃO E DA LINGUAGEM HUMANA (STIL). **Anais [...]**. SBC, 2021. p. 1–10. Disponível em: [doi.org/10.5753/stil.2021.17778](https://doi.org/10.5753/stil.2021.17778). Acesso em: 07 jun. 2025.

PICTON, B.; ANDALIB, S.; SPINA, A.; CAMP, B.; SOLOMON, S. S.; LIANG, J.; CHEN, P. M.; CHEN, J. W.; HSU, F. P.; OH, M. Y. Assessing AI Simplification of Medical Texts: Readability and Content Fidelity. **International Journal of Medical Informatics**, v. 195, p. 105743, mar. 2025. ISSN 1386-5056. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1386505624004064>. Acesso em: 22 jun. 2025.

SENO, E. R. M.; CLARO, D.; MOTA, L.; RODRIGUES, J. Semântica distribucional. In: CASELI, H. M.; NUNES, M. G. V. (Ed.). **Processamento de Linguagem Natural: conceitos, técnicas e aplicações em português**. 3. ed. BPLN, 2024. cap. 10. ISBN 978-65-01-20581-6. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/parte-significado/cap-semantica-distribucional/cap-semantica-distribucional.html>. Acesso em: 9 jul. 2025.

SENO, E. R. M.; PAIVA, V. d.; PINHEIRO, V. Semântica com técnicas simbólicas. In: CASELI, H. M.; NUNES, M. G. V. (Ed.). **Processamento de Linguagem Natural: conceitos, técnicas e aplicações em português**. 3. ed. BPLN, 2024. cap. 9. ISBN 978-65-01-20581-6. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/parte-significado/cap-semantica-simbolica/cap-semantica-simbolica.html>. Acesso em: 19 nov. 2025.

SHARMA, K. P.; YAJID, M. S. A.; GOWRISHANKAR, J.; MAHAJAN, R.; ALSOUD, A. R.; JADHAV, A.; SINGH, D. A Systematic Review on Text Summarization: Techniques, Challenges, Opportunities. **Expert Systems**, v. 42, n. 4, p. e13833, 2025. ISSN 1468-0394. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.13833>. Acesso em: 09 jan. 2026.

SMITH, E. A.; SENTER, R. J. **Automated Readability Index**. [S. l.]: Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, 1967.

SOUSA, C. M. A. d. O. A. d.; CARDOSO, E.; ANDRADE, F. D. d. DIRETRIZES PARA O USO DE LINGUAGEM SIMPLES: PESQUISA E DESENVOLVIMENTO NO BRASIL E EM PORTUGAL. **Revista da Associação Brasileira de Atividade Motora Adaptada**, v. 25, n. 2, p. 407–422, dez. 2024. ISSN 2674-8681. Disponível em: <https://revistas.marilia.unesp.br/index.php/sobama/article/view/16619>. Acesso em: 22 maio. 2025.

SOUZA, J. W. d. C.; CARDOSO, P. C. F.; PAIXÃO, C. A. Sumarização automática. In: CASELI, H. M.; NUNES, M. G. V. (Ed.). **Processamento de Linguagem Natural: conceitos, técnicas e aplicações em português**. 3. ed. BPLN, 2024. book chapter 24. ISBN 978-65-01-20581-6. Disponível em: <https://brasileiraspln.com/livro-pln/3a-edicao/parte-aplicacoes/cap-as/cap-as.html>. Acesso em: 16 jan. 2026.

STRAKA, M. UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task. In: CONLL 2018 SHARED TASK: MULTILINGUAL PARSING FROM RAW TEXT TO UNIVERSAL DEPENDENCIES. **Proceedings [...]**. Bruxelas, Bélgica: Association for Computational Linguistics, 2018. p. 197–207. Disponível em: [doi.org/10.18653/v1/K18-2020](https://doi.org/10.18653/v1/K18-2020). Acesso em: 01 jun. 2025.

SUN, R.; JIN, H.; WAN, X. Document-level text simplification: Dataset, criteria and baseline. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING. **Proceedings [...]**. Punta Cana, República Dominicana: Association for Computational Linguistics, 2021. p. 7997–8013. Acesso em: 01 maio. 2025.

SWANSON, K.; HE, S.; CALVANO, J.; CHEN, D.; TELVIZIAN, T.; JIANG, L.; CHONG, P.; SCHWELL, J.; MAK, G.; LEE, J. Biomedical text readability after hypernym substitution with fine-tuned large language models. **PLOS Digital Health**, Public Library of Science, v. 3, n. 4, p. 1–18, 2024. ISSN 2767-3170. Disponível em: doi.org/10.1371/journal.pdig.0000489. Acesso em: 22 jun. 2025.

UNIVERSIDADE ESTADUAL DE CAMPINAS. **Guia de Linguagem Simples**. 2024. Disponível em: <https://linguagemsimples.unicamp.br/wp-content/uploads/sites/49/2024/08/guia-de-linguagem-simples.pdf>. Acesso em: 14 maio. 2025.

UNIVERSIDADE ESTADUAL DE GOIÁS. **Uso da linguagem simples no âmbito da Universidade Estadual de Goiás**. 2023. Disponível em: <http://www.ueg.br/legislacao/referencia/13259>. Acesso em: 16 maio. 2025.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L. u.; POLOSUKHIN, I. Attention is all you need. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS. **Proceedings [...]**. Curran Associates, Inc., 2017. v. 30. Disponível em: [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf). Acesso em: 11 jul. 2025.

VILA, M.; MARTÍ, M. A.; RODRÍGUEZ, H. Is This a Paraphrase? What Kind? Paraphrase Boundaries and Typology. **Open Journal of Modern Linguistics**, Scientific Research Publishing, v. 4, n. 1, p. 205–218, fev. 2014.

XU, W.; NAPOLES, C.; PAVLICK, E.; CHEN, Q.; Callison-Burch, C. Optimizing Statistical Machine Translation for Text Simplification. **Transactions of the Association for Computational Linguistics**, MIT Press, Cambridge, Massachusetts, EUA, v. 4, p. 401–415, 2016. Disponível em: doi.org/10.1162/tacl\_a\_00107. Acesso em: 26 abr. 2025.

YADAV, A. K.; Ashish Kumar Maurya; Ranvijay; Rama Shankar Yadav. Extractive Text Summarization Using Recent Approaches: A Survey. **Ingénierie des Systèmes d’Information**, v. 26, n. 01, p. 109–121, fev. 2021. Disponível em: doi.org/10.18280/isi.260112. Acesso em: 14 maio. 2025.

ZHANG, H.; YU, P. S.; ZHANG, J. A Systematic Survey of Text Summarization: From Statistical Methods to Large Language Models. **ACM Comput. Surv.**, v. 57, n. 11, p. 277:1–277:41, jun. 2025. ISSN 0360-0300. Disponível em: <https://dl.acm.org/doi/10.1145/3731445>. Acesso em: 15 jan. 2026.

ZHANG, T.; KISHORE, V.; WU, F.; WEINBERGER, K. Q.; ARTZI, Y. **BERTScore**: Evaluating Text Generation with BERT. arXiv, 2020. Disponível em: doi.org/10.48550/arXiv.1904.09675. Acesso em: 03 jul. 2025.

ZHAO, W. X.; ZHOU, K.; LI, J.; TANG, T.; WANG, X.; HOU, Y.; MIN, Y.; ZHANG, B.; ZHANG, J.; DONG, Z.; DU, Y.; YANG, C.; CHEN, Y.; CHEN, Z.; JIANG, J.; REN, R.; LI, Y.; TANG, X.; LIU, Z.; LIU, P.; NIE, J.-Y.; WEN, J.-R. A Survey of Large Language Models. arXiv, mar. 2025. ArXiv:2303.18223 [cs]. Disponível em: doi.org/10.48550/arXiv.2303.18223. Acesso em: 28 jun 2025.

## APÊNDICE A – PROMPTS UTILIZADOS PARA SIMPLIFICAÇÃO TEXTUAL

### A.1 Primeiro *Prompt* - apenas simplificação

Você é um modelo de linguagem especializado na conversão de textos para textos simplificados. Sua tarefa é converter documentos para linguagem simples sem remover informações importantes. Responda com um novo documento em Português Brasileiro.

**\*\*Passos para simplificar um documento\*\***

- Inicie o documento pelas informações mais importantes
- Mantenha o mesmo significado geral do texto e dos parágrafos
- Encurte as sentenças com mais de 14 palavras
- Remova a redundância de todas as sentenças
- Escreva todas as frases na ordem direta, sujeito, verbo e objeto
- Escreva na voz ativa em vez da passiva
- Substitua palavras complexas e jargões por palavras de mesmo significado, ou explique o significado de palavras complexas e jargões
- Remova palavras que não possuem função na frase
- Escreva o nome completo antes de siglas
- Evite palavras estrangeiras pouco comuns em português

**\*\*Objetivo:\*\*** gere um documento mais simples seguindo esses passos, mas não sumarie, mantenha todas as informações relevantes.

**\*\*Resposta\*\*:** retorne apenas o novo texto gerado em Português, não faça comentários sobre sua implementação.

Aqui está o texto alvo:

{text}

## A.2 Segundo *Prompt* - sumarização e simplificação

Você é um modelo de linguagem especializado na simplificação e sumarização de textos. Sua tarefa é converter documentos para linguagem simples sem remover informações importantes.

Não faça alterações em tabelas ou referências para imagens

Você irá receber várias partes de um documento e deve sempre retornar a sumarização do novo conteúdo recebido.

**\*\*Instruções para a escrita\*\***

- Inicie o documento pelas informações mais importantes
- Procure não escrever sentenças com mais de 20 palavras
- Evite escrever sentenças redundantes
- Escreva todas as frases na ordem direta, sujeito, verbo e objeto
- Escreva na voz ativa em vez da passiva
- Substitua palavras complexas e jargões por palavras de mesmo significado, ou explique o significado delas antes
- Escreva o nome completo antes de siglas
- Evite palavras estrangeiras pouco comuns em português ou explique o significado delas antes

**\*\*Objetivo:\*\*** gere um documento mais simples seguindo esses passos, selecione as informações mais importantes do texto e escreva os parágrafos mais relevantes primeiro.

**\*\*Resposta\*\*:** retorne apenas o novo texto gerado em Português, não faça comentários sobre sua implementação.

**APÊNDICE B – EXEMPLOS DE SIMPLIFICAÇÕES REALIZADAS**

Código-fonte 1 – Simplificação com a maior similaridade semântica – Edital N° 12/2021 - Hackathon Inovando UFC – Texto simplificado original

```

1
2 ## COMO POSSO PARTICIPAR?
3
4 Poderão participar servidores(as) docentes e técnico-administrativos, assim
   como estudantes de graduação e de pós-graduação dos diferentes cursos da
   Universidade Federal do Ceará.
5
6 As inscrições poderão ser individuais ou em grupo através do link [bit.ly/
   InscricaoInovandoUFC](http://bit.ly/InscricaoInovandoUFC)
7
8 A inscrição não exige que o(a) participante já tenha solução prévia aos
   problemas,
9
10 ![Icon representing an individual, a lightbulb.](48
    f188337e3ba41df38fab9ac0afb1bd_img.jpg)
11
12 Icon representing an individual, a lightbulb.
13
14 **Individual:** você será alocado(a) em uma equipe;
15
16 ![Icon representing a group, a lightbulb.](373
    e2ccc5c19a052a1b8cb6de3eebc94_img.jpg)
17
18 Icon representing a group, a lightbulb.
19
20 **Em grupo:** deve conter no mínimo 2 (dois) membros, desde que, pelo menos
    um deles, seja servidor (técnico-administrativo ou docente).
21
22 ![Icon representing a group of people.](1517552d07f4a8b389be86af8aff6424_img.
    jpg)
23
24 Icon representing a group of people.
25
26 A PROINTER formará a partir dos inscritos equipes de 3 (três) a 5 (cinco)
    pessoas com, no mínimo, um(a) servidor(a) e um(a) estudante de graduação.
27
28 As equipes serão formadas de modo que possuam, pelo menos, um membro que se
    encaixe em cada um dos perfis apresentados a seguir:
29
30 ![Icon representing a project manager, a lightbulb.](
    f5fd8a6b12eff431b5d2cc73cad225b0_img.jpg)
31
32 Icon representing a project manager, a lightbulb.
33
34 Gestor(a) de projetos;
35
36 ![Icon representing a designer, a lightbulb.](
    de09a3bb7d77b9108650b10e8df0e7c8_img.jpg)
37
38 Icon representing a designer, a lightbulb.
39
40 Designer;
41
42 ![Icon representing a developer, a lightbulb.](4
    c7edbe6e90384d6698dcc59ef3aa0cd_img.jpg)
43
44 Icon representing a developer, a lightbulb.
45
46 Desenvolvedor(a).

```

Código-fonte 2 – Simplificação com a maior similaridade semântica – Edital N° 12/2021 - Hackathon Inovando UFC – Gerada pelo Qwen3:14b

```

1 ### **3. PARTICIPANTES E INSCRIÇÕES**
2 - **Quem pode participar?**
3   - Servidores (docentes e técnicos-administrativos) e
      estudantes de graduação ou pós-graduação da UFC.
4 - **Formas de inscrição:**
5   - Individual ou em grupo.
6   - **Grupos:**
7     - Mínimo de 2 participantes, com **pelo menos 1
      servidor (técnico-administrativo ou docente)**.
8     - Grupos com menos de 5 pessoas podem receber sugestões
      de novos integrantes da PROINTER.
9   - **Equipes:**
10    - Formação obrigatória: 3 a 5 pessoas, com **1 servidor
      e 1 estudante de graduação**.
11 - **Perfis solicitados nos grupos:**
12   1. **Gestor(a) de projetos** (organização de recursos,
      prazos).
13   2. **Designer** (criação de protótipos e inovação).
14   3. **Desenvolvedor(a)** (solução técnica).
15   - Estudantes ou servidores de áreas distintas podem se
      candidatar a esses perfis.
16 - **Inscrições:**
17   - Período: **12 a 30 de abril de 2021**.
18   - Link: [Formulário de inscrição](https://forms.gle/
      KsvRFJzFT5nyxLnt9).
19   - Validadas até **04 de maio de 2021**.
20   - Limitação: **Máximo de 20 equipes** (seletiva com base
      em perfis, desafio de interesse e ordem de inscrição).

```

Fonte: elaborado pelo Qwen3:14b.

Código-fonte 3 – Simplificação com o melhor resultado na métrica *long sentence ratio* – Edital Printer PIBI N° 07/2023 – Texto Original

```

1
2 ![Logo da Universidade Federal do Ceará (UFC)](2
   dfa6ac3edfe874f68aa0cbccaa42322_img.jpg)
3
4 Logo da Universidade Federal do Ceará (UFC)
5
6 # UNIVERSIDADE FEDERAL DO CEARÁ PRÓ-REITORIA DE RELAÇÕES
   INTERNACIONAIS E DESENVOLVIMENTO INSTITUCIONAL
7
8 ## EDITAL PROINTER PIBI n 07/2023
9
10 ## CHAMADA PARA SELEÇÃO DE PROPOSTAS DE PROJETOS DE
    ESTUDANTES E DE SERVIDORES DA UFC PARA CONCESSÃO DE
    BOLSAS DE ESTÍMULO AO EMPREENDEDORISMO INOVADOR
    PROGRAMA EMPREENDE UFC
11
12 A Universidade Federal do Ceará, por intermédio da Pró-
    Reitoria de Relações Internacionais e Desenvolvimento
    Institucional (PROINTER), convida os(as) estudantes de
    graduação, de pós-graduação e os(as) servidores (as)
    docentes e técnico-administrativos(as) a apresentarem
    propostas de projetos de empreendimentos inovadores,
    startups, negócios de impacto socioambiental, negócios
    de base tecnológica & spin-offs acadêmicas, de acordo
    com as normas deste Edital, para o Programa de Estímulo
    ao Empreendedorismo Inovador, denominado Empreende UFC.

```

Fonte: .

Código-fonte 4 – Simplificação com o melhor resultado na métrica *long sentence ratio* – Edital Printer PIBI N° 07/2023 – Gerada pelo Qwen3:14b

```

1 A Universidade Federal do Ceará (UFC) publica o Edital n
   07/2023 para selecionar projetos de estudantes e
   servidores da universidade interessados em
   empreendedorismo inovador. O programa, chamado Empreende
   UFC, oferece bolsas para apoiar ideias de negócios,
   startups, empresas com impactos sociais ou ambientais, e
   spin-offs acadêmicas.

```

Fonte: elaborado pelo Qwen3:14b.

Código-fonte 5 – Simplificação com alucinações – Edital Prointer PIBI N° 07/2023 – Texto original

- 1  
2 ### 7 - Deveres do(a) Orientador(a)  
3  
4 7.1 Constituem deveres dos(as) orientadores(as) dos(as)  
5 bolsistas aprovados(as) no programa Empreende UFC 2023:  
6  
7 1. Ser servidor(a) ativo(a), em efetivo exercício,  
8 vinculado(a) ao quadro permanente da UFC, podendo ser  
9 docente ou técnico-administrativo(a) com formação mínima  
em nível de graduação;  
10 2. Dedicar-se por, no mínimo, 4 (quatro) horas semanais às  
atividades do projeto;  
11 3. Acompanhar as atividades desenvolvidas pela equipe de  
estudantes bolsistas e voluntários dos projetos,  
verificando a conformidade com as disposições deste  
Edital;  
12 4. Auxiliar os(as) bolsistas do projeto selecionado a  
preencherem formulários de avaliação de maturidade do  
projeto em três momentos: (i) no início do programa, (ii  
) logo após o ciclo de formação em empreendedorismo  
inovador e (iii) no último mês do programa, com a  
finalidade de sistematização das ações desenvolvidas  
durante o período de vigência da bolsa;

## Código-fonte 6 – Simplificação com alucinações – Edital – Gerada pelo Llama-3.2-instruct:3b

```

1  **7 - Deveres do(a) Orientador(a)**
2
3  **7.1** Constituem deveres do(a) orientador(a) dos projetos
4      selecionados para o programa Empreende UFC 2023:
5
6  **a.) Orientação nhica:**
7
8  1.  **Dedicação às atividades do projeto:** O(a) orientador
9      (a) deve dedicar 12 horas semanais às atividades do
10     projeto, incluindo reuniões com a equipe, monitoramento
11     do progresso e ajustes no currículo de aprendizado;
12
13  2.  **Apresentações e engajamentosempre preventivas:** O(a)
14     orientador(a) deve prestar assegurações aos(as)
15     bolsistas(as) de que as atividades são adequadas para
16     desenvolverem suas competências e habilidades que
17     ampliem sua formação acadêmica e técnica, e pois
18     diminuam a carga horária obrigatória nas disciplinas
19     semestrais que configuram o diálogo pretendeu entender
20     antes;
21
22  3.  **Feedback regular:** O(a) orientador(a) deve fornecer
23     feedback regular e construtivo às equipes sobre o
24     progresso do projeto e ajudar(as) a identificar áreas de
25     melhoria;
26
27  4.  **Observações de coincência na atividade principal:**
28     Orientador(a) deve exercer vigil ncia constante na
29     adequação do modelo empresarial ao intento forletivo por
30     s u c i f i c config <ScalarscreateQueryBuilder social
31     (intent di)", " inclusion*$estánd notregulatednya separ
32     ('/');

```

Fonte: Gerado pelo Llama3.2-instruct:3.2b.

## APÊNDICE C – QUESTIONÁRIO SOBRE SIMPLIFICAÇÃO TEXTUAL

**Questão 1.** Email (campo aberto)

**Questão 2.** Vínculo com a universidade:

- (a) Não possuo vínculo com a universidade
- (b) Discente de graduação ou pós-graduação
- (c) Servidor técnico-administrativo
- (d) Docente

**Questão 3.** Com que frequência você lê editais públicos?

- (a) Nunca - Eu nunca leio esse tipo de edital
- (b) Raramente - Leio editais apenas em situações muito específicas
- (c) Ocasionalmente - Leio alguns editais quando sugem oportunidades que me interessam
- (d) Frequentemente - Costumo ler a maioria dos editais que aparecem nas minhas áreas de interesse
- (e) Sempre - Leio todos os editais que encontro, mesmo que nem todos sejam diretamente relevantes

**Questão 4.** Qual é o seu nível de experiência com a produção, revisão ou divulgação de editais públicos?

- (a) Nunca - Nunca participei dessas atividades
- (b) Raramente – Participei 1 ou 2 vezes.
- (c) Ocasionalmente – Participei algumas vezes, mas sem regularidade.
- (d) Frequentemente - Participo com regularidade ou está relacionado ao meu trabalho
- (e) Muito frequentemente - Participo continuamente ou é parte importante do meu trabalho

**Questão 5.** As 3 questões mostradas em a), b) e c) foram repetidas para os dois editais avaliados:

- (a) Comparando o edital original e a versão simplificada, você considera que o conteúdo da versão simplificada facilita o entendimento das informações principais?
- (b) O conteúdo da versão simplificada do edital parece mais útil para o público que precisa entender rapidamente as principais informações?

- a) Discordo totalmente
- b) Discordo parcialmente
- c) Neutro / Não tenho opinião formada
- d) Concordo parcialmente
- e) Concordo totalmente

(c) Você considera que o conteúdo da versão simplificada está pronto para ser graficamente finalizado e enviado para divulgação, servindo como complemento ao edital original? Por quê? (campo aberto)

**Questão 6.** Na sua opinião, o conteúdo simplificado gerado automaticamente tem potencial para diminuir a quantidade de dúvidas encaminhadas por e-mail, telefone ou outros canais oficiais sobre o conteúdo do edital?

- (a) Discordo totalmente
- (b) Discordo parcialmente
- (c) Neutro / Não tenho opinião formada
- (d) Concordo parcialmente
- (e) Concordo totalmente

**Questão 7.** Na sua opinião, o conteúdo simplificado gerado automaticamente facilita o trabalho administrativo de elaboração e divulgação de editais? (campo aberto)

**Questão 8.** Quais aspectos dos conteúdos das versões simplificadas dos 2 editais podem contribuir para facilitar o entendimento pelo público? (campo aberto)

**Questão 9.** Você identificou algum risco ou limitação relevante no conteúdo dessas versões simplificadas geradas automaticamente? Se sim, qual(is)? (campo aberto)