

Uma proposta de Automação para o Índice Nacional de Avaliação de Linguagem Simples em Serviços Públicos

Adriano Péricles Rodrigues¹, Gabriel M. Marques¹, Lucas B. Rodrigues¹, Pedro Augusto A. Mattos¹, Vanessa T. Nunes^{1,2}, Claudia Cappelli^{1,3}, Rodrigo Oliveira^{1,2}, Rodrigo Michel de Moraes⁴

¹ Universidade Federal de Goiás, Instituto de Informática, Goiânia, GO.

² Universidade Federal Fluminense, Programa de Pós-Graduação em Computação, Niterói, RJ

³ Universidade Estadual do Rio de Janeiro, Departamento de Informática e Ciência da Computação, Rio de Janeiro, RJ

⁴ Fundação Getúlio Vargas, Escola de Administração de Empresas de São Paulo, São Paulo, SP

adriano.rodrigues@goias.gov.br, {mata.marques, brandao.brandao}@discente.ufg.br, pedroamattos2002@gmail.com, vanunes@gmail.com, claudia.cappelli@gmail.com, rodrigo.santosprogressao@gmail.com, rodrigo.michel30@gmail.com

Abstract. *The digitization of services brings public entities closer to citizens and facilitates and optimizes their access to public policies through technology. Plain Language appears as a conductor in this path and a concern with the full understanding by the citizen is necessary. In this paper, we propose the first results of automation of simple language practices defined within the National Guide for Simple Language. The objective is to evaluate the explanation of digital public services and serve as a support mechanism for the inclusion of plain language by Brazilian municipalities.*

Resumo. *A digitalização dos serviços aproxima os entes públicos do cidadão e facilita e otimiza seu acesso às políticas públicas por meio da tecnologia. A Linguagem Simples entra como condutora nesse caminho e é necessária uma preocupação com o entendimento pleno do cidadão. Neste artigo propomos os primeiros resultados de automação de práticas de linguagem simples definidas dentro do Guia Nacional de Linguagem Simples. O objetivo está em avaliar a explicação de serviços públicos digitais e servir como mecanismo de apoio a inclusão de linguagem simples pelos municípios brasileiros.*

1. Introdução

Os governos têm evoluído seus processos de oferta de serviços digitais seja pela exigência da sociedade, seja pela força da lei. No Brasil e no mundo, estudos e abordagens vêm sendo desenvolvidos para entender o papel das experiências com usuários de TIC (Tecnologias da Informação e Comunicação) em diferentes áreas dos governos (Cunha e Miranda, 2013). Uma das correntes está fortemente relacionada ao acesso e uso de informações. O cidadão acessa sites que oferecem os serviços e faz uso

deles. Porém, para o cidadão poder fazer uso destes serviços ele precisa antes de tudo entender o que está sendo comunicado a ele. Entender o que ele precisa informar para o serviço ser prestado, além de ser importante perceber como este serviço será oferecido a ele e como ele obterá o resultado desejado, caso contrário ele não usará o serviço. Neste sentido, existe um desafio quando falamos de cidadãos de um país com diferentes culturas e escolaridades.

Dentro deste contexto, várias organizações vêm percebendo a importância da Linguagem Simples, que é uma técnica de comunicação que visa o entendimento, aplicada no contexto de Serviços Públicos e Transformação Digital. Isso vem protagonizando um número enorme de iniciativas nos mais diversos setores. Leis e Decretos (Lei 12.527/2011 - Lei de Acesso à Informação - , Lei 13.460/2017 - Lei dos Serviços e Administração Pública, Lei 14.129/2021 – Lei do Governo Digital) têm sido promulgados na intenção de gerar direcionamento para os governos Federal, Estadual e Municipal. Também algumas organizações têm construído Guias de Boas Práticas como: Guia de Linguagem Simples do Portal Gov.BR (www.gov.br/pt-br/guia-de-edicao-de-servicos-do-gov.br/escrevendo-para-o-seu-usuario), Programa Municipal de Linguagem Simples de São Paulo (<https://011lab.prefeitura.sp.gov.br/linguagem-simples/inicio>), Laboratório de Inovação e Dados do Governo do Estado do Ceará (irislab.ce.gov.br/), Rede Linguagem Simples Brasil (www.redelinguagensimplesbrasil.org), Laboratório Interdisciplinar de Linguagem Cidadã (<https://linclab.com.br/>), e Orientações para Adoção de Linguagem Clara (www.governoaberto.sp.gov.br/wp-content/uploads/2017/12/orientacoes_para_adocao_linguagem_clara_ptBR.pdf). Eles têm conteúdo que pode ser utilizado pelas mais diversas organizações. Um dos mais completos existentes é o Guia Nacional de Linguagem Simples <https://abep-tic.org.br/guia-de-uso-da-linguagem-para-apresentacao-de-servicos-publicos/>) capitaneado pela ABEPTIC (Associação Brasileira de Entidades Estaduais e Públicas de TIC) que contém práticas para tornar os textos mais simples, e outros formatos de apresentação de informação como gráficos, tabelas, modelos de processo, bancos de dados entre outros.

Baseado neste Guia Nacional de Linguagem Simples, os Portais de Serviço dos Estados serão avaliados em 2023 quanto ao atendimento das práticas do Guia Nacional de Linguagem Simples. Para tal foi construído um Método de Avaliação de Linguagem Simples para Serviços Públicos e foi estabelecido o Índice de Linguagem Simples. No ano de 2022 foi realizado o primeiro piloto de uso deste índice juntamente com o Índice de Serviços Digitais já aferido pela ABEPTIC nos Portais de Serviço dos estados.

Este índice tem um grande desafio. Cada Estado tem seu próprio Portal de Serviços e cada um apresenta uma quantidade grande de serviços. Em média, um estado tem entre 7500 e 10.000 serviços. Para a realização do piloto foram escolhidos 15 serviços para cada um dos 27 estados. É uma quantidade grande para a análise de várias práticas de Linguagem Simples. Porém, está longe do que tem que ser analisado para verificar se os serviços de todas as secretarias e órgãos de um ente federativo estão de fato atendendo as práticas do Guia Nacional de Linguagem Simples.

Isto gerou a demanda por automação desta análise e consequente busca por mecanismos que poderiam ajudar neste trabalho. Atualmente já existem diversas abordagens que envolvem o uso de Processamento de Linguagem Natural (PLN) para avaliar textos para diversos fins. Para o tema de Linguagem Simples existem

abordagens sendo discutidas para as línguas inglesa, espanhol, francesa, italiana dentre outras, mas ainda é um trabalho inicial quando se trata do português brasileiro.

Este trabalho propõe uma ferramenta que automatiza a avaliação de práticas de linguagem simples. Apresentamos neste artigo os primeiros resultados já obtidos, com o desenvolvimento e uso da ferramenta para apoiar a realização da avaliação anual do Índice de Linguagem Simples dos Estados realizada pela ABEPTIC.

O artigo está organizado da seguinte forma: A Seção 2 fala sobre avaliação da Linguagem Simples em Portais de Serviços. Na Seção 3 apresentamos a proposta de automação e o Sistema construído para apoiar este processo de avaliação. A Seção 4 mostra um exemplo de uso no Estado de Goiás. A Seção 5 aborda os desafios atuais e discute os próximos passos. Por fim, a Seção 6 apresenta conclusões preliminares.

2. A avaliação do Uso da Linguagem Simples nos serviços digitais

Um grande objetivo do Governo Digital é reduzir a burocracia no relacionamento entre cidadão e governo, e para isso é preciso melhorar a forma de apresentar informações institucionais, descrições dos serviços públicos, documentos e guias orientativos e conteúdos das redes sociais. A comunicação precisa ser fluida, direta, fácil e empática. A Linguagem Simples possui práticas e sugere técnicas que podem auxiliar na produção de textos mais acessíveis, tornando a jornada do cidadão mais simples e agradável.

Alguns entes públicos vêm investindo nesta comunicação mais simples e direta com o uso das práticas de Linguagem Simples. Porém, um ponto que tem sido discutido é que não se sabe se o que está sendo feito está de fato atingindo o objetivo desejado. Surge então a necessidade da existência de uma medição de desempenho que mostre o atendimento das práticas e se elas de fato corroboram com o trabalho que está sendo realizado. Assim, os gestores podem tomar decisões mais consistentes e assertivas.

O Índice de Linguagem Simples planeja medir a aderência dos Portais de Serviço dos governos estaduais e distrital às práticas da Linguagem Simples. Foi criado com base na Lei 12.527/2011 (Lei do Acesso à Informação) e Lei 14.129/2021, que estabelece regras e instrumentos para a prestação digital de serviços públicos.

Para a aplicação do Índice foi construída uma Metodologia com 4 fases de execução: (i) Coleta de informações; (ii) Análise de evidências; (iii) Cálculo do Índice e (iv) Apresentação dos resultados. Para a fase de coleta de informações foram elencadas nesta versão inicial do Índice de Linguagem Simples o conjunto de práticas apresentadas na Tabela 1.

Tabela 1. Práticas avaliadas no Índice de Linguagem Simples

Nº	Práticas
1	Seu texto fala diretamente com seu público? <i>Utiliza estruturas gramaticais orientadas à segunda pessoa do singular.</i>
2	Todas as palavras usadas são de fácil entendimento do público-alvo? São palavras de domínio comum? Caso use jargões ou termos próprios da área de negócio, existe algum mecanismo de esclarecimento?
3	As frases são curtas? <i>(usam entre 7 e 10 palavras)</i> São escritas na ordem direta? <i>(sujeito+verbo+objeto)</i>
4	São utilizados elementos gráficos que ajudam no entendimento das informações? <i>Em geral, estes são utilizados em substituição a textos para facilitar o entendimento.</i>

5	São feitos agrupamentos que permitem ao cidadão perceber mais facilmente o relacionamento entre as informações? <i>Em geral, agrupamos informações que se relacionam, como login e senha, dados pessoais, dados de endereço, dados de cobrança, dados sobre etapas de um processo, entre outros.</i>
6	As informações estão organizadas de forma que as mais importantes para uso do serviço venham em primeiro lugar? <i>Por exemplo: Quando vamos executar um serviço em geral queremos saber quais são seus passos, o que temos que fazer. Então esta informação deveria preceder, por exemplo, a Lei que existe e que apoia este serviço, os endereços físicos para sua realização, entre outros.</i>

Na fase de Análise de evidências, as métricas de avaliação de cada uma das 6 práticas são analisadas para verificar a aderência dos Portais às mesmas. Em seguida o Índice é calculado e por fim apresentado a todos os interessados.

Uma primeira avaliação foi feita como piloto para teste da metodologia e durante este trabalho pode-se perceber a importância desta avaliação poder ter um componente de automação, o que traria para o índice, e para o uso do guia, um crescimento exponencial. Assim, foi feita uma seleção de algumas das métricas apresentadas na Tabela 2, baseado no nível de dificuldade de automação.

Tabela 2. Métricas de Avaliação das Práticas do Índice de Linguagem Simples

Nº	#	Métricas
1	O que avaliar:	Descrição dos serviços, passos e requisitos para executar o serviço.
	Como avaliar:	Verifique se o texto está na segunda pessoa do singular ou no imperativo afirmativo.
	Como calcular:	Descrição do serviço: +3 Passos do serviço: +4 Requisitos: +3
2	O que avaliar:	Título do serviço e Corpo do texto
	Como avaliar:	Título do serviço <ul style="list-style-type: none"> Avalie se começa com um verbo que indique um pedido do cidadão e não uma oferta do Estado. Avalie se o título resume o que a descrição apresenta. Avalie se o título tem entre 3 e 5 palavras. Corpo do texto <ul style="list-style-type: none"> Avalie e Liste as palavras consideradas difíceis (segundo os avaliadores)
	Como calcular:	Título do serviço <ul style="list-style-type: none"> Começa com um verbo adequado = +2 Resume corretamente a descrição = +2 Tem entre 3 e 5 palavras = +1 Corpo do texto <ul style="list-style-type: none"> Sem palavras difíceis = +5 1 a 2 palavras difíceis = +4 3 a 4 palavras difíceis = +3 5 a 6 palavras difíceis = +2 7 a 8 palavras difíceis = +1 9 ou mais palavras difíceis = +0
3	O que avaliar:	Corpo do texto
	Como avaliar:	Avalie se as frases são formadas por no máximo 10 palavras: Sim ou Não. Avalie se a frase indica uma única ação a ser tomada: Sim ou Não. Avalie se as frases estão na forma direta: Sim ou Não

	Como calcular:	Frases com no máximo 10 palavras: +4. Frases indicam uma única ação a ser tomada: +2. Frases na forma direta: +4
4	O que avaliar:	Corpo do texto
	Como avaliar:	Avalie se as instruções são representadas por tabelas. Avalie se as classificações são apresentadas em tabelas. Avalie se a sequência de ações é apresentada com marcadores.
	Como calcular:	Cada texto com pelo menos uma dessas características que não seja apresentado em seus respectivos formatos: -2
5	O que avaliar:	Corpo do texto
	Como avaliar:	Identifique os agrupamentos de informações mínimos necessários. <ul style="list-style-type: none"> Os agrupamentos seguem a seguinte regra: título, descrição, público, links, etapas, requisitos e informações adicionais. O item de requisitos está agrupado em um único item? Cada conjunto de informações pertence ao seu item?
	Como calcular:	Para cada agrupamento não realizado: -2 Informações sobre um item agrupado em outro: -2
6	O que avaliar:	(1) Título, (2) Descrição, (3) Público, (4) Links, (5) Etapas, (6) Requisitos/Exigências e (7) Informações complementares
	Como avaliar:	<ul style="list-style-type: none"> Verifique se os itens 1, 2, 6 e 7 estão nesta ordem. Os itens 3, 4 e 5 podem ser exibidos em qualquer ordem.
	Como calcular:	<ul style="list-style-type: none"> Se o título (1) e a descrição (2) do serviço não forem os primeiros: -5 Se os requisitos (6) e informações complementares (7) não vierem no final: -3

3. Proposta de Automação

O projeto e desenvolvimento de um sistema para automatizar a avaliação do uso de linguagem simples em interfaces textuais de apresentação de serviços digitais selecionou algumas das métricas relacionadas às práticas 1, 2 e 3. A métrica da prática 2 de seleção de palavras difíceis, gerou a necessidade de iniciar a construção de um *dataset* de palavras, cujas estratégias e processo de construção são apresentados na Seção 3.1. O projeto e construção do sistema de avaliação é apresentado na Seção 3.2.

3.1. Construção de um *DataSet* público para suporte à avaliação e sugestão de linguagem simples em serviços digitais

Datasets são componentes fundamentais em um projeto que envolve manipulação de dados para fins de inferência e aprendizado. De forma geral, *Datasets* são bases de dados projetados para um fim específico, que servem de amostras para treinamentos de algoritmos de inteligência artificial ou para outros tipos de projetos que envolvam ciência de dados. No contexto deste trabalho, estamos interessados na construção de um *dataset* que permita avaliar textos, e posteriormente sugerir alterações nele, em relação ao grau de dificuldade de entendimento. Nossa estratégia inicial, considerando as métricas selecionadas, foi construir um *dataset* de palavras, pois o vocabulário é a porta de entrada para uma comunicação bem-sucedida. Iniciamos pela simplificação léxica cujo objetivo é avaliar uma palavra difícil.

Existem estudos que mostram que as abordagens baseadas em dados para encontrar uma palavra mais simples são úteis (Horn et al., 2014; Glavas e Stajner, 2015). Em (Uchida et al., 2018), os autores focam em aspectos educacionais e não

discutem apenas questões técnicas, falando que a definição de critérios para definir nível de dificuldade tendem a ser subjetivos, o que implicaria que métodos de pesquisa anteriores podem não ser úteis para sugerir simplificações para diferentes níveis de proficiência das pessoas. Este trabalho em especial se baseia no Quadro Europeu Comum de Referência para as Línguas (CEFR) para definir níveis de dificuldade de palavras baseado no grau de proficiência de um aprendiz de uma língua estrangeira.

No português brasileiro, encontramos diversos *datasets* publicados por diferentes grupos de desenvolvimento de soluções para PLN (como, <https://forum.ailab.unb.br/t/datasets-em-portugues/251>) porém nenhum se encontrava estruturado para o objetivo de avaliar nível de dificuldades de palavras. Sendo assim, decidimos pela criação gradual de um *dataset* que visa se tornar um corpus oficial de palavras a ser utilizado para fins de análise de níveis de simplicidade de textos de apresentação de informações sobre serviços.

Com o foco específico em analisar este tipo de conteúdo, optamos pela criação de um *dataset* contendo as palavras mais utilizadas nas apresentações de serviços já existentes. Utilizamos como entrada, os textos de apresentação do portal de serviços do Estado de Goiás (Portal Expresso) pelo acesso ao conteúdo dos textos apresentados em cada serviço do portal. Foi criado um script em python para extrair todos os textos das cartas de serviços do Portal Expresso. Utilizando a ferramenta Spacy (<https://spacy.io/>) foram retirados pronomes, vogais e caracteres diferentes do alfabeto ocidental.

Para compor o dicionário foram calculadas as frequências de cada palavra e esta lista foi ordenada em ordem decrescente. Utilizamos a lei de Zipf (Zipf, 1949) para realizar um corte no número de palavras, ficando com as 10% mais frequentes. Conforme a formulação empírica de Zipf, existe uma relação entre a frequência e a posição da palavra no ranque. Ele afirma que existem poucas palavras muito frequentes, um número significativo de palavras de frequência média e muitas palavras de frequência baixa. De uma lista inicial de 13.582 palavras foram selecionadas as 2.000 palavras mais frequentes. Em uma análise manual, verificamos que uma série de palavras com erros de grafia ou muito incomuns foram descartadas. Porém, o *dataset* precisava aumentar de tamanho para se tornar mais rico. Além disso, era necessário classificar as palavras quanto ao grau de dificuldade.

Pesquisamos por definições federais, estaduais, de mercado ou acadêmicas com este propósito, e não encontramos uma proposta sobre a qual nos basear. A estratégia de classificação que utilizamos foi selecionar um conjunto de materiais do ensino fundamental e médio para adicionar palavras que ainda não pertenciam ao *dataset*. Criamos a seguinte classificação: (0 muito fácil) palavras que aparecem em textos do ensino fundamental do 1º ao 4º ano; (1 fácil) palavras que aparecem em textos do ensino fundamental do 5º ao 9º ano; (2 médio) palavras que aparecem em textos do ensino médio; e (3) palavras desconhecidas ou classificadas como difíceis. As palavras foram adicionadas ao *dataset* (resolvendo as repetições) totalizando 8048 palavras.

Para criar um *dataset* ainda mais robusto, foi coletado um *dataset* de 340.000 palavras de livros de acesso livre e da internet, incluindo artigos, preposições e até mesmo erros ortográficos. Em seguida, foram removidos pronomes, vogais isoladas, artigos e caracteres diferentes do alfabeto, o que resultou em uma redução para 278.600 palavras. Utilizando o site <https://www.dicio.com.br/>, todas as palavras foram consultadas e, se o site retornasse um significado para a palavra buscada, ela era

considerada correta. Essa consulta resultou em 134.396 palavras corretas. Adicionando essas palavras ao *dataset* original sem criar repetições, o *dataset* final ficou com 135.147 palavras. Essas palavras novas foram classificadas em difícil (3), se tiverem mais de 10 letras, médio (2), se tiverem entre 5 e 10 letras, e fácil, se tiverem menos de 5 letras (1). O *dataset* final apresenta a estrutura apresentada na Figura 1.

Palavras	Classe				
certificado	2	tecnologia	1	agrodefesa	3
comunidade	1	amapa	1	situacoes	1
emprego	1	devidamente	2	ceu	1
transporte	1	suspender	1	informe	1
esgoto	1	bancario	1	adicional	1
habitacao	1	colonizador	1	lugares	1

Figura 1. Estrutura do dataset de palavras

3.2. Sistema de automação

O projeto de desenvolvimento do sistema de automação foi pensado em duas vertentes: O desenvolvimento dos algoritmos para avaliação das métricas selecionadas e o desenvolvimento de uma interface amigável e orientada a práticas de linguagem simples para interface de sistemas constante no Guia Nacional de Linguagem Simples.

Para o desenvolvimento da ferramenta foi utilizado Python para criar um site com uma arquitetura dividida entre frontend e backend. Para o frontend, utilizou-se o framework Flask, conhecido pela sua facilidade de uso e versatilidade. Já para o backend, optou-se pelo uso do framework FastAPI, que oferece um desempenho excepcionalmente rápido e é altamente escalável.

Em relação à prática 1 foi utilizada a ferramenta Spacy (<https://spacy.io/>) que é uma biblioteca de software de código aberto, escrita em Python e Cython, que “entende” linguagem natural e pré-processa textos. Ela permite a criação de sistemas que processam e interpretam grandes volumes de texto, sendo usada para extrair informações dos *datasets* de textos que utilizamos. Ela foi treinada em diversas línguas, incluindo o português (<https://spacy.io/models/pt>) e consegue distinguir morfologias como classes gramaticais, pessoa do verbo, gênero e número. Para a métrica da prática 1, o Spacy verifica se o verbo está na primeira pessoa, e se está no singular ou plural.

Na prática 2, primeiro foram implementadas as seguintes métricas: se o título começa com um verbo que indique um pedido do cidadão e não uma oferta do Estado; se o título tem entre 3 e 5 palavras; e contar a quantidade de palavras consideradas difíceis (de forma automatizada). Optamos por trabalhar com avaliações que envolvam interpretação e contextualização nas próximas etapas deste trabalho, por se tratar de um tema com maior grau de dificuldade. Para avaliar o verbo foi utilizado o Spacy que verifica se a frase começa com verbo, se este está no infinitivo e se não está na primeira pessoa. A avaliação do título consistiu em uma contagem das palavras da frase caracterizada como título e uma verificação do intervalo. E com relação à última métrica, criamos um serviço que verifica a classificação das palavras considerando o *dataset* para verificação. Caso a palavra não exista no *dataset*, é classificada como difícil (4). Planejamos usar o paradigma da computação por humanos (Human-based computation), onde os usuários classificam o nível de dificuldade das palavras.

Por fim, as métricas da prática 3 foram totalmente implementadas. A avaliação do número de palavras de cada frase do texto foi realizada utilizando o Spacy para

separar as frases do texto e o número de palavras foi contabilizado utilizando código. Para verificar se a frase indica uma única ação, adotamos nesse momento uma verificação simples utilizando o Spacy para verificar se existe um único verbo em uma frase. Também foi utilizado o Spacy para verificar se as frases estão na forma direta.

A interface do sistema foi elaborada para os avaliadores poderem analisar se um serviço cumpre as práticas de linguagem simples antes de publicá-los efetivamente para os cidadãos em suas cartas de serviço. Desenvolvemos um design de sistema em página dividida em três seções principais. Um menu expansível, na lateral esquerda, com as principais funções, informações e contato. No centro há uma área editável para o usuário inserir ou escrever o texto a ser avaliado, no lado direito existe um painel onde serão exibidas as notas avaliadas, conforme apresentado a Figura 2.

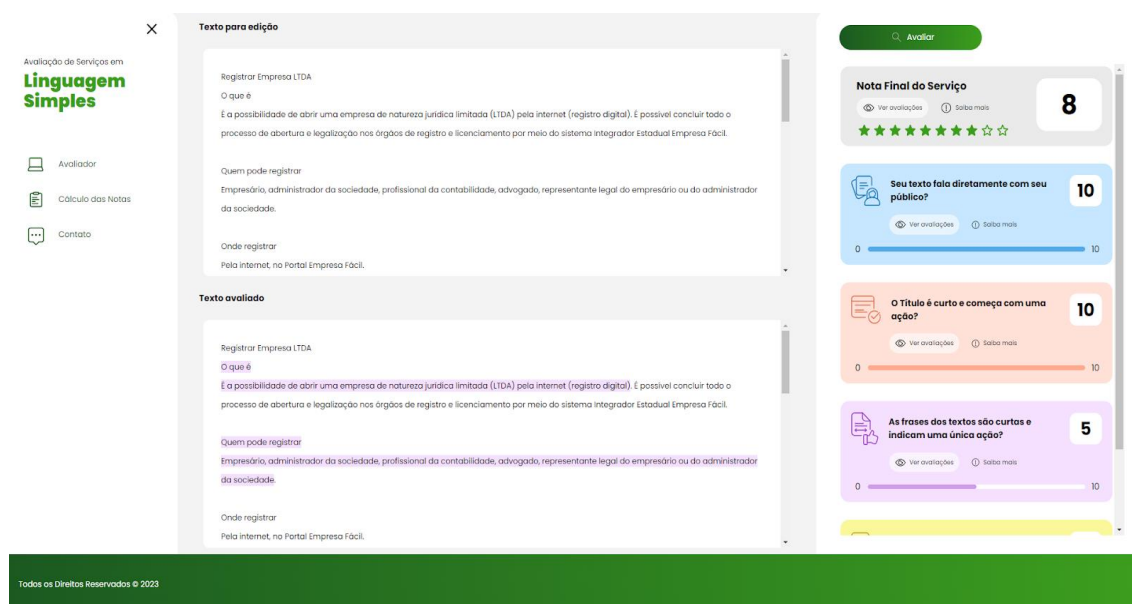


Figura 2. Interface do avaliador de linguagem simples com um exemplo de uso.

A partir do texto sobre o serviço e com o usuário clicando no botão “Avaliar” no alto da página à direita, o sistema altera o centro da tela para duas áreas textuais. A primeira é editável onde o usuário pode realizar modificações e a segunda, logo abaixo, apresenta marcações no texto original associadas a cada prática avaliada. Essas práticas são exibidas em seções com cores distintas no painel mais à direita da tela. Cada prática contém suas respectivas descrições e notas de 0 a 10 (normalizadas para fins de entendimento) que o serviço recebeu em sua avaliação nos itens a seguir:

- **Seu texto fala diretamente com seu público?** O sistema verifica a prática 1, sobre se os verbos das frases são orientados à segunda ou terceira pessoa do singular. Exibe uma nota de 0 a 10 segundo o número de verbos.
- **O Título é curto e começa com uma ação?** O sistema analisa parte da prática 2 que está associada ao título, ou seja, a primeira frase do texto do serviço. Uma nota é atribuída entre 0 e 10 considerando, se o título começa com um verbo, mais 3 pontos caso o verbo esteja no infinitivo e mais 4 pontos, se a frase contém entre 3 a 5 palavras de conteúdo, ou seja, esta soma exclui preposições, artigos ou palavras de ligação (e, ou, de, etc.).

- **Todas as palavras usadas são de fácil entendimento?** O sistema avalia na prática 2 se as palavras utilizadas são de uso comum e de fácil entendimento considerando o *dataset* de verificação apresentado na Seção 3.1.
- **As frases dos textos são curtas e indicam uma única ação?** O sistema verifica a prática 3, se as frases do texto possuem apenas um verbo, com até 10 palavras de conteúdo. Esta soma também exclui preposições, artigos ou palavras de ligação (e, ou, de, etc.).

É possível ao usuário retirar as marcações de uma determinada prática na área do texto original no centro da tela clicando no botão “ver avaliações”, assim o usuário poderá se concentrar em modificar um ou outro item avaliado. Em cada prática há também um botão de “saiba mais” que exibe um pop-up sobre a tela que resume o que foi avaliado e exemplifica exibido na Figura 3. Além disso, o sistema exibe uma seção acima de todas as práticas que calcula a média aritmética das notas recebidas pelo serviço e destaca visualmente a nota final entre 0 a 10 estrelas.

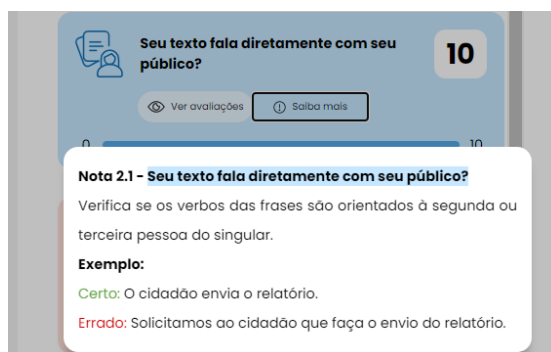


Figura 3. Exemplo de pop-up de resumo de uma prática

4. Exemplo de Uso: Serviços Digitais do Estado do Goiás

O estado de Goiás classificou-se em 2º em 2022 no ranking nacional do Índice ABEPTIC de Oferta de Serviços Digitais realizado anualmente pela ABEPTIC. Este índice é utilizado para compor um dos Pilares de Eficiência da Máquina Pública e serve como direcionador para que as UF's possam entender o que precisa ser melhorado no processo de transformação digital, com representação de todos os estados brasileiros.

Atualmente, existem mais de 500 mil usuários no portal único ao cidadão - Expresso: www.go.gov.br, representando um aumento de 20 vezes o número de usuários em junho de 2021. No entanto, apesar da grande evolução do crescimento na disponibilidade dos serviços digitais e melhorias, percebe-se que no que diz respeito à Linguagem Simples ainda se tem desafios a percorrer. Portanto, surgiu o interesse em atuar nesta pesquisa para trabalhar pontos de melhoria e explorar o uso da automação.

Analisando o serviço de “Solicitação online da 2ª via de Carteira de Identidade (RG)”, a avaliação manual, considerando as métricas trabalhadas nesta versão do sistema, verificou que na prática 1 nenhuma parte do texto se encontra na segunda pessoa do singular ou no imperativo afirmativo, recebendo nota 0. Na prática 2, o título começa com verbo que indica pedido do cidadão e está entre 3 e 5 palavras, ganhando pontuação máxima. Foram identificadas 3 palavras consideradas difíceis de acordo com avaliação dos revisores: concessão, requerente e nato (perde 2 pontos de um total de 5).

Na prática 3 existiam frases com mais de 10 palavras, frases indicando mais de uma ação a ser tomada e frases na forma indireta, recebendo nota 0. A nota calculada foi 8.

Na avaliação automatizada pelo sistema a nota final foi 7, conforme apresentado na Figura 4. A diferença se deu porque na construção do *dataset* foram definidos graus de dificuldades das palavras. Sendo assim, cada nível de dificuldade ganhou um peso no cálculo da nota da seguinte forma: palavras de nível 1 tem peso 0.2, palavras de nível 2 tem peso 0.7 e palavras de nível 3 tem peso 1. Cada quantidade de palavras contabilizada é multiplicada por este peso e o somatório é dividido pelo número total de palavras. Além disso, em todas as avaliações, as notas de cada item foram normalizadas para o intervalo de 0 a 10 para facilitar o entendimento. Em seguida foram convertidas para os valores definidos nas métricas de avaliação (Tabela 2).

The screenshot displays the 'Linguagem Simples' interface. On the left, there's a sidebar with the title 'Avaliação de Serviços em Linguagem Simples' and navigation icons for 'Avaliador', 'Cálculo das Notas', and 'Contato'. The main area is split into 'Texto para edição' and 'Texto avaliado'. The 'Texto para edição' section shows a document titled 'Emitir Segunda Via Da Carteira De Identidade Rg' with a question 'O que é?' and a detailed answer. The 'Texto avaliado' section shows the same text with some words highlighted in yellow. On the right, a sidebar titled 'Avaliar' shows the 'Nota Final do Serviço' as 7, along with five other metrics: 'Seu texto fala diretamente com seu público?' (10), 'O Título é curto e começa com uma ação?' (10), 'As frases dos textos são curtas e indicam uma única ação?' (0), and 'O vocabulário está de fácil entendimento?' (7). Each metric has a star rating and a progress bar.

Figura 4. Exemplo de uso do Sistema de Linguagem Simples

Através da construção de um *dataset* e da definição dos mecanismos de classificação de níveis de dificuldade foi possível calcular e apresentar as palavras consideradas difíceis, porém, com tons de amarelo diferentes que marcam no texto as palavras e seus respectivos níveis, como apresentado na Figura 5.

This screenshot shows a close-up of the 'Linguagem Simples' interface. The 'Texto avaliado' section contains a paragraph of text with several words highlighted in different shades of yellow. A tooltip box on the right explains the difficulty levels: 'Nota - as palavras usadas são de fácil entendimento?' and lists examples: 'Comum' (e.g., 'cidadão'), 'Incomum' (e.g., 'encarregado'), 'Difícil' (e.g., 'atribuído'), and 'Muito difícil' (e.g., 'incumbido').

Figura 5. Exemplo de uso - Níveis de dificuldade das palavras

5. Desafios e Discussões

A ideia de um sistema para automatizar a avaliação do uso de Linguagem Simples em textos de apresentação de serviços públicos surgiu para escalar essa análise, que manualmente se torna inviável pelo alto esforço e tempo requeridos e para servir como um instrumento que possa ser facilmente utilizável pelas organizações públicas, ajudando a disseminar e conquistar adeptos para a prática da Linguagem Simples. Então nossos próximos passos consistem em apresentar esse sistema para usuários selecionados de organizações específicas que possam atuar como um grupo externo de validadores desta versão do sistema e das próximas que virão.

Podemos citar alguns desafios que já estamos enfrentando e aqueles que já identificamos para o futuro. O primeiro deles é o reconhecimento de que nem tudo será possível automatizar devido ao custo alto de desenvolvimento de técnicas e tecnologias. Nosso objetivo é minimizar esse esforço, porém reconhecendo e deixando claras essas limitações. Uma estratégia possível, neste sentido, está em expandir as verificações considerando outras práticas importantes no Guia Nacional de Linguagem Simples, mas que não foram selecionadas para aferição dentro do Índice.

Outra discussão está na crítica em relação ao processo de simplificação da linguagem e que este movimento poderia levar ao empobrecimento da linguagem. Essa é uma discussão extremamente relevante, porém o que a corrente de linguagem simples acredita é que o uso desta abordagem é útil quando se precisa expressar de forma clara uma informação que requer uma ação de um número considerável de pessoas. Este é o caso de apresentação de textos que explicam como um serviço público é oferecido e quais são as condições e ações necessárias. O uso de palavras e termos complexos, o chamado “burocratês”, dificulta o entendimento sobre como usar um serviço até mesmo para pessoas classificadas no nível Proficiente. Em (Pires, 2021), a autora realizou um estudo experimental com pós-graduandos e identificou que a escrita complexa tende a prejudicar o entendimento mesmo de quem consegue lidar com textos que demandam alto custo de processamento mental. O que a comunidade que atua neste tema acredita é que uma vez que consigamos diminuir o custo de processamento mental, conseguimos aumentar o entendimento durante a leitura de um texto com estas características. Mas é necessário encontrar um equilíbrio do uso de práticas de simplificação, sem gerar perda de informação, sacrifício da precisão ou distorção de sentido.

O terceiro desafio está relacionado também ao projeto de uma única solução que consiga automatizar a avaliação de textos que já existem. O fato de nem sempre existir uma carta de serviços unificando a apresentação de todos os serviços de uma cidade, e de inclusive não haver um padrão de apresentação, torna o processo automático de coleta de dados destes diferentes ambientes na web (web scrapping) extremamente difícil de ser concebido como uma solução única.

6. Conclusões

Nosso objetivo neste trabalho é apoiar e escalar a construção e avaliação de textos de apresentação de serviços digitais que se baseiem nas práticas de linguagem simples propostas no Guia Nacional de Linguagem Simples. A ferramenta está pronta para utilização e o grande desafio é facilitar o uso de padrões de escrita, gerando melhor entendimento e proximidade com a sociedade.

Outro objetivo está em ajudar a disseminar as práticas de linguagem simples para outros municípios e estados da federação. Pretendemos inclusive, que a ferramenta facilite o entendimento e promova o uso de jargões regionais específicos de um grupo populacional. Nossa hipótese é de que a geração dos resultados dos testes e relatórios rápidos gerados pela ferramenta irão diminuir os custos operacionais dos governos e aproximar cada vez mais a população da experiência digital.

Para evolução imediata do sistema de automação e sua interface é necessário evoluir as métricas incluídas e incluir as avaliações das práticas 4 a 6 que dependem da evolução do *dataset* de verificação para avaliações mais contextuais e da captura de informações além dos textos, no caso tabelas, gráficos e links. Pretendemos também inserir na interface a possibilidade de feedback por parte dos usuários, sinalizando quais marcações avaliadas no texto original não foram precisas e assim colaborar com a melhoria do modelo de avaliação em linguagem simples.

Como trabalho futuro pretendemos realizar sugestões de alterações no texto para auxiliar na construção das descrições dos serviços. Para isso, estamos estudando formas de incluir no *dataset* informações que contextualizem o uso das palavras para que seja possível realizar trocas de palavras que mantenham o sentido das frases descritas.

Por fim, é esperado que todo o processo de utilização da linguagem simples acelere a governança das informações dos estados, promovendo a transparência e acesso dos serviços, controle das contas públicas de forma mais clara e transparente. Nesta fase inicial de automatização inicia o primeiro passo para a inclusão e desburocratização do serviço público brasileiro de forma estruturada e padronizada.

Referências

- Cunha, M.A.V.C., and Miranda, P.R.M. (2013). O uso de TIC pelos governos: uma proposta de agenda de pesquisa a partir da produção acadêmica e da prática nacional. In *Organizações & Sociedade*, 20(66), 543-566.
- Glavas, G. and Stajner, S. (2015) “Simplifying Lexical Simplification: Do We Need Simplified Corpora?” In: Proc. 53rd Annual Meeting of the Association for Computational Linguistics and International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL - IJCNLP), p. 63-68, Beijing, China.
- Horn, C., Manduca, C. and Kauchak, D. (2014) “Learning a Lexical Simplifier Using Wikipedia”, In: Proc. of 52nd Annual Meeting of the Association for Computational Linguistics (ACL), p. 458-463, Baltimore, Maryland.
- Pires, H.F.M. (2021) Impactos da linguagem simples na compreensibilidade da informação em governo eletrônico: o caso de um benefício do INSS, Dissertação de mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Artes e Design.
- Uchida, S., Takada, S. and Arase, Y. (2018) “CEFR-based Lexical Simplification Dataset”, In: Proc. of the Eleventh International Conference on Language Resources and Evaluation (LREC), Miyazaki, Japan.
- Zipf, G.K. (1949) “Human Behavior and the Principle of Least Effort”, Addison-Wesley