

UNIVERSIDADE FEDERAL DE SÃO CARLOS– UFSCAR
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA– CCET
DEPARTAMENTO DE COMPUTAÇÃO– DC
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO– PPGCC

Cássio Faria da Silva

**Abordagem computacional para
avaliação automática da qualidade da
argumentação na dimensão retórica de
tweets no domínio da política brasileira**

Cássio Faria da Silva

**Abordagem computacional para
avaliação automática da qualidade da
argumentação na dimensão retórica de
tweets no domínio da política brasileira**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Metodologias e Técnicas de Computação

Orientadora: Profa. Dra. Helena de Medeiros Caseli

Coorientadora: Profa. Dra. Vânia Paula de Almeida Neris

São Carlos

2023

Silva, Cássio Faria da

Abordagem computacional para avaliação automática da qualidade da argumentação na dimensão retórica de tweets no domínio da política brasileira / Cássio Faria da Silva -- 2023.
175f.

Tese de Doutorado - Universidade Federal de São Carlos, campus São Carlos, São Carlos

Orientador (a): Helena de Medeiros Caseli

Banca Examinadora: Helena de Medeiros Caseli, Diego Furtado Silva, Aline Marins Paes Carvalho, Thiago Alexandre Salgueiro Pardo, Evelin Carvalho Freire de Amorim

Bibliografia

1. Avaliação automática da qualidade da argumentação.
2. Dimensão retórica. I. Silva, Cássio Faria da. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Ronildo Santos Prado - CRB/8 7325



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Defesa de Tese de Doutorado do candidato Cássio Faria da Silva, realizada em 08/05/2023.

Comissão Julgadora:

Profa. Dra. Helena de Medeiros Caseli (UFSCar)

Prof. Dr. Diego Furtado Silva (UFSCar)

Prof. Dr. Aline Marins Paes Carvalho (UFF)

Prof. Dr. Thiago Alexandre Salgueiro Pardo (USP)

Profa. Dra. Evelin Carvalho Freire de Amorim (INESC TEC)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Computação.

Agradecimentos

Agradecimentos

À Deus pela vida, coragem e resiliência.

À minha esposa, Ana Cláudia, pela companhia nas viagens e nas noites em claro, pelo amor, carinho, dedicação, incentivo, compreensão, paciência e auxílio. Obrigado, amor da minha vida!

Ao meu filho, Leonardo, pelo amor, companheirismo e pelas revisões do texto em inglês. Obrigado, filhote mais lindo e inteligente do universo!

Aos meus pais, Vilma e Orlando, principalmente pelos princípios e valores que me transmitiram. Sem vocês nada disso seria possível. Muito obrigado por tudo!

À toda a minha família, principalmente aos meus sobrinhos, Felipe, Cecília e Davih, por todo amor, orações, apoio e incentivo dispensados. Deus não poderia ter me presenteado com uma família melhor! Muito obrigado!

À Professora Helena Caseli, minha querida orientadora, que não mediu esforços para a realização deste trabalho, por todo o conhecimento transmitido, orientação e dedicação. Muito obrigado pelos conselhos e pela paciência, Professora!

À Professora Vânia Neris, minha querida orientadora, por me nortear nos caminhos da pesquisa acadêmica. Agradeço imensamente por toda a orientação e dedicação, Professora!

Elas foram essenciais para a realização do meu doutorado, para o meu crescimento como pesquisador e como ser humano. Me ajudaram a superar todos os desafios que surgiram ao longo do caminho, as dúvidas e as dificuldades. Me transmitiram não só os aspectos técnicos e metodológicos da ciência, mas também os valores éticos e humanos que devem nortear o trabalho acadêmico. Muito obrigado às minhas mentoras, amigas e parceiras.

Aos professores do DC, em especial aos Professores Cesar Teixeira e Daniel Lucrédio, pelo fundamental apoio nos estágios iniciais deste estudo.

Aos amigos linguistas que anotaram o corpus e contribuíram para a elaboração das

diretrizes de anotação, utilizados neste trabalho, Amanda Pontes Rassi, Jackson Wilke da Cruz Souza, Renata Ramisch e Roger Alfredo de Marci Rodrigues Antunes. Obrigado pelo apoio e pelos momentos que passamos juntos.

Ao meu amigo, Sidney Evaldo Leal, que forneceu total acesso ao NILCMetrix e me auxiliou na geração das métricas.

Aos amigos do DC, LALIC e NILC. Obrigado pelas conversas e momentos de descontração.

Ao Professor José Gonzaga e aos amigos da REGES, pelo apoio incondicional na realização deste trabalho.

*“Your future is whatever you make it, so make it a good one.
(Doc Brown, “Back to the future”)*

Resumo

As pesquisas na área de argumentação, inerente ao ser humano e essencial para a comunicação tanto falada quanto escrita, datam do século IV a.C. A argumentação é multidisciplinar e abrange vários campos de pesquisa, inclusive a ciência da computação. A comunicação evoluiu até as redes sociais que são uma fonte considerável de textos argumentativos de vários domínios, como a política. A avaliação automática de argumentos é tema de recentes estudos na área de Processamento de Linguagem Natural e modelos computacionais que procuram desempenhar essa tarefa estão sendo propostos, especialmente fundamentados em algoritmos baseados em Máquina de Vetores de Suporte (*Support Vector Machines*, SVM) e redes neurais profundas (*Deep Learning*). Em paralelo, cópulas de textos argumentativos em inglês estão sendo produzidos. Como forma de contribuir com as pesquisas relacionadas à avaliação da qualidade da argumentação em português, este trabalho tem como objetivo propor, implementar e validar uma abordagem computacional, para a avaliação automática da qualidade de argumentação na dimensão retórica em *tweets* relacionados à política brasileira. A abordagem envolve o desenvolvimento de um modelo computacional, um cópulas anotado com mensagens relacionadas à política e diretrizes de anotação específicos para a tarefa. Os estudos aqui conduzidos evidenciaram que a maneira que se mostrou mais adequada para avaliar a qualidade da argumentação em *tweets* relacionados à política brasileira foi usando um modelo neural gerado a partir do ajuste fino do BERTimbau. O modelo proposto foi capaz de prever com 100% de precisão instâncias da classe de Alta qualidade da argumentação.

Palavras-chave: avaliação da qualidade da argumentação, *tweet*, BERT, política brasileira, processamento da língua portuguesa.

Abstract

Research in the area of argumentation, inherent to human beings and essential for both spoken and written communication, dates back to the 4th century BC. The argument is multidisciplinary and covers several fields of research, including computer science. Communication has evolved to social networks, which are a considerable source of argumentative texts in various domains, such as politics. The automatic evaluation of arguments is the subject of recent studies in the area of Natural Language Processing and computational models that seek to perform this task are being proposed, especially based on algorithms based on Support Vector Machines and Deep Learning. In parallel, corpus of argumentative texts in English are being produced. As a way of contributing to research related to the assessment of the quality of argumentation in Portuguese, this work aims to propose, implement and validate a computational approach, for the automatic assessment of the quality of argumentation in the rhetorical dimension in tweets related to Brazilian politics. The approach involves developing a computational model, an annotated corpus with policy-related messages, and task-specific annotation guidelines. The studies carried out here showed that the most appropriate way to assess the quality of arguments in tweets related to Brazilian politics was using a neural model generated from the fine-tuning of BERTimbau. The proposed model was able to predict with 100% accuracy instances of the High quality argumentation class.

Keywords: argument quality assessment, tweet, BERT, Brazilian politics, Portuguese language processing.

Lista de ilustrações

Figura 1 – Tela inicial do sistema.	120
Figura 2 – Tela de avaliação da qualidade da argumentação.	121
Figura 3 – Tela de avaliação da qualidade da argumentação.	122
Figura 4 – Tela de avaliação da qualidade da argumentação.	123

Lista de tabelas

Tabela 1 – Grandes áreas de um trabalho científico em Processamento de Linguagem Natural (PLN) e seções desta tese.	28
Tabela 2 – Grupo de métricas de índices de leituraabilidade obtidos por processamento da mensagem (6).	88

Lista de siglas

AM Aprendizado de Máquina

API *Application Programming Interface*

CETIC Centro Regional de Estudos para a Desenvolvimento da Sociedade da Informação

CGU Conteúdo Gerado pelo Usuário

CSS *Cascading Style Sheets*

DT *Decision Tree*

GT Gênero Textual

HCAI *Human-Centered Artificial Intelligence*

HTML *HyperText Markup Language*

IA Inteligência Artificial

KNN *K-Nearest Neighbors*

LIWC *Linguistic Inquiry and Word Count*

LR *Logistic Regression*

NB *Naive Bayes*

NILC Núcleo Interinstitucional de Linguística Computacional

PLN Processamento de Linguagem Natural

RF *Random Forest*

RST *Reliable, Safe & Trustworthy*

SVM *Support Vector Machines*

ToLD-Br *Toxic Language Dataset for Brazilian Portuguese*

Sumário

1	INTRODUÇÃO	21
1.1	Contexto e motivação	22
1.2	Objetivo, Hipóteses e Questão de pesquisa	25
1.3	Organização do trabalho	26
1.4	Outra forma de ler este trabalho	27
2	QUALIDADE DA ARGUMENTAÇÃO EM <i>TWEETS</i> DE POLÍTICA: O QUE E COMO AVALIAR	29
3	RECURSOS LINGUÍSTICOS	81
3.1	Pistas linguísticas	81
3.1.1	Clareza	81
3.1.2	Organização	82
3.1.3	Credibilidade	83
3.1.4	Apelo Emocional	84
3.2	Ferramentas computacionais	85
3.2.1	Normalização de texto com Enlvo (BERTAGLIA, 2017)	85
3.2.2	Análise de polaridade com Random Forest e TweetSentBR (BRUM; NUNES, 2018)	86
3.2.3	Análise de sentimentos com LIWC (BALAGE FILHO; PARDO; ALUÍSIO, 2013)	87
3.2.4	Métricas textuais e psicolinguísticas com NILCMetrix (LEAL, 2021)	88
3.2.5	Deteção de linguagem tóxica com BERT e ToLD-Br (LEITE et al., 2020)	89
4	CLASSIFICAÇÃO DA QUALIDADE DA ARGUMENTAÇÃO EM <i>TWEETS</i> NO DOMÍNIO DA POLÍTICA BRASILEIRA	91
5	INTERFACE DE USUÁRIO PARA O PÚBLICO GERAL	119

6	CONCLUSÃO	125
	Conclusão	125
6.1	Respostas à questão de pesquisa	125
6.2	Contribuições	126
6.3	Limitações	128
6.4	Trabalhos futuros	129
6.5	Considerações finais	129
	REFERÊNCIAS	131

APÊNDICES 137

APÊNDICE A	– DIRETRIZES DE ANOTAÇÃO	139
APÊNDICE B	– FERRAMENTA DE ANOTAÇÃO	155

ANEXOS 161

ANEXO A	– EXEMPLO DAS MÉTRICAS OBTIDAS PELO NILC-METRIX	163
---------	---	-----

Capítulo 1

Introdução

O estudo da arte da argumentação iniciou na Grécia no século V a.C., no momento em que estava se estabelecendo um governo democrático. Para se fazer entender, as pessoas precisavam ser capazes de falar em público, pois as decisões públicas eram tomadas em assembleias e apenas um pequeno número de indivíduos exercia o papel de liderança e canalizava o curso dos acontecimentos para servir aos seus próprios interesses. Como não havia advogados, os réus falavam em seus próprios nomes. Para atender às necessidades das pessoas, surgiram professores chamados “sofistas”, que ensinavam a arte da argumentação, ensinando a organizar argumentos, decompor discursos em partes lógicas, selecionar as melhores palavras e combiná-las com cuidado (ARISTOTLE; KENNEDY(TRANSLATOR), 2007).

Na política, a discussão de ideias e argumentos é fundamental e, em uma sociedade democrática, é indispensável que as pessoas expressem suas opiniões. Além disso, é importante que as convicções, posições e entendimentos sejam debatidos e avaliados por todos os indivíduos. A democracia permite que os cidadãos tenham uma participação igualitária nas decisões políticas porém, é comum que a democracia seja reduzida a uma escolha entre diferentes partidos políticos. Ao contrário, a política deve ser um campo onde diferentes pontos de vista e ideias são debatidos e explorados. Para que isso aconteça, é fundamental que haja um ambiente igualitário e aberto, no qual todas as vozes sejam ouvidas(BAQUERO, 2008).

Atualmente, diversas variáveis como o cenário político de ideias polarizadas, a possibilidade de comunicação instantânea por meio das redes sociais e a expectativa da ampla participação igualitária da população nas decisões políticas, foram adicionadas à democracia. Essas variáveis fomentam ainda mais o desejo das pessoas de se envolverem nas discussões.

A avaliação da qualidade da argumentação em mensagens relacionadas à política brasileira se apresenta como um tema relevante e desafiador. Uma possível forma de abordar a avaliação da qualidade da argumentação é utilizar técnicas de Aprendizado de Máquina (AM) e PLN. Essas técnicas permitem extrair características relevantes dos textos. Além disso, podem classificar as mensagens em categorias pré-definidas, como argumentos de alta ou baixa qualidade. Dessa forma, é possível contribuir para a evolução do debate político e para o combate à desinformação nas redes sociais.

Nesse contexto, este trabalho apresenta uma abordagem computacional para avaliação da qualidade da argumentação em *tweets* de domínio político, escritos em português do Brasil. No percurso traçado, foi desenvolvido um modelo computacional para a avaliação da qualidade da argumentação pelo qual foram realizados experimentos usando algoritmos de AM baseados em *features*, como Regressão Logística, K-Nearest Neighbor, Árvores de Decisão, *Support Vector Machines* (SVM), *Random Forest* (RF) e *Naive Bayes* (NB), e também um ajuste fino de dois modelos neurais (BERTimbau e RobertaTwitterBR). Para tanto, foi construído um corpus de mensagens coletadas do Twitter, relacionadas à política, publicadas por deputados brasileiros e seus seguidores.

Além disso, foi proposta uma interface de usuário pública para ser usada em atividades de avaliação da qualidade da argumentação de mensagens de redes sociais relacionadas à política, e o site do projeto ArgQ!¹ que reúne, além da interface de avaliação, as diretrizes utilizadas para a anotação das mensagens e as publicações produzidas ao longo deste estudo.

Esta tese, que engloba dois artigos (um publicado e outro aceito para publicação), descreve o caminho percorrido para a proposta e validação de um modelo computacional de avaliação da qualidade da argumentação em mensagens de redes sociais. Este primeiro capítulo apresenta de forma sintética, na seção 1.1, o contexto e a motivação tratados nesta tese. Na seção 1.2 são especificados os objetivos e questão de pesquisa, e na seção 1.3 são apresentados os artigos que compõem esta tese.

1.1 Contexto e motivação

A argumentação é uma área de pesquisa multidisciplinar que estuda processos de debate e raciocínio. Abrange vários campos de pesquisa como política, direito, psicologia, lógica, filosofia, linguagem e ciência da computação. Um argumento é uma afirmação (ou conclusão) acompanhada por um número arbitrário de premissas que justificam, fundamentam, sustentam, defendem ou explicam a reivindicação (POTTHAST et al., 2019).

Argumentos bem fundamentados não são apenas importantes para a tomada de decisões e aprendizado, mas também desempenham um papel fundamental na obtenção de conclusões amplamente aceitas. Em ciência da computação, a identificação ou avalia-

¹ O site do projeto ArgQ! está disponível em <<https://argq.org/>>.

ção automática da argumentação são tópicos estudados na área da inteligência artificial (BENCH-CAPON; DUNNE, 2007) devido à necessidade de combinar necessidades representacionais com modelos cognitivos relacionados ao usuário e modelos computacionais para raciocínio automatizado. Em linguística computacional, a argumentação é um campo de pesquisa recente que se concentra na análise de argumentos em textos em linguagem natural (STAB; GUREVYCH, 2017a).

A argumentação consiste em uma ou mais sentenças nas quais várias premissas são apresentadas para embasar uma conclusão. As sentenças que fazem parte da argumentação constituem uma expressão completa que visa convencer um interlocutor (EEMEREN; GROOTENDORST, 2003). A argumentação é inerente ao ser humano e está presente na comunicação verbal e escrita. Evidências indicam que a Internet e as redes sociais estão entre os principais meios de comunicação atualmente e, por isso, são fonte de um grande volume de textos argumentativos dos mais variados domínios, como política e governo, avaliações de produtos, fóruns de discussão e portais de debates.

Portanto, uma fonte significativa de dados para muitas das disciplinas interessadas em estudos relacionados a argumentação é a Web e, particularmente, as mídias sociais. Redes sociais, fóruns de discussão, notícias online e avaliação de produtos fornecem um fluxo heterogêneo e crescente de informações, onde argumentos gerados por usuários podem ser identificados, isolados e analisados. A disponibilidade desses dados, juntamente com avanços na linguística computacional e no AM, criaram um cenário promissor para o surgimento de uma nova área de pesquisa denominada mineração de argumentação (ou argumento).

Por meio do PLN, que permite que computadores processem a língua natural por meio do AM e da linguística, pesquisas têm sido realizadas no sentido de identificar argumentos e suas unidades, gerar e avaliar a qualidade de argumentos tanto para textos formais quanto para Conteúdo Gerado pelo Usuário (CGU), especialmente das redes sociais. As tarefas relacionadas à argumentação computacional, como mineração, geração, identificação de argumentos e a sua avaliação se mostram relevantes em atividades como suporte à escrita e assistência à discussão (STAB; GUREVYCH, 2017b; GARCÍA-GORROSTIETA; LÓPEZ-LÓPEZ, 2018; GARCÍA-GORROSTIETA; LÓPEZ-LÓPEZ; GONZÁLEZ-LÓPEZ, 2018). Entretanto, a maioria dos trabalhos atuais lidam com textos formais em inglês para mineração de argumentos.

De especial interesse para este trabalho é a rede social Twitter², bastante utilizada para troca de informações e opiniões sobre política no Brasil.

Ao lidar com textos do Twitter, escritos em português, este trabalho enfrenta desafios não presentes na maioria dos trabalhos da literatura, como o número limitado de caracteres (280) permitido para cada mensagem, que focam na mineração de argumentos em textos formais em inglês. Em relação ao idioma, vale destacar que a língua portuguesa é uma das

² <<https://twitter.com/>>

mais faladas no mundo, com mais de 280 milhões de falantes que utilizam o português em situações oficiais (como a língua ensinada e aprendida nas escolas) e não oficiais (como a língua falada em contextos familiares e nas redes sociais). Nesse sentido, são notáveis os avanços e a importância que tem sido dada atualmente ao desenvolvimento e aprimoramento dos recursos de PLN para o português. Isso pode estar relacionado, em grande parte, aos aspectos da globalização e da popularização do acesso ao ambiente online, que são fatores facilitadores do intercâmbio sociolinguístico dessa língua³.

No que se refere à avaliação da qualidade da argumentação, foco deste estudo, é necessário levar em consideração diversos aspectos, como a estrutura lógica, a coerência, a consistência, a relevância dos argumentos, a clareza e a persuasividade (WACHSMUTH et al., 2017). Cada um desses aspectos pode ser medido por meio de diferentes técnicas computacionais. Desde o esquema de argumento de Toulmin (2003), estudos têm sido conduzidos para simplificar o entendimento da estrutura e importância dos elementos do texto argumentativo. Recentemente, uma taxonomia composta por três dimensões, para avaliar a qualidade dos elementos de argumentos, foi proposta por Wachsmuth et al. (2017). Desde então, pesquisas (POTTHAST et al., 2019; WACHSMUTH; WERNER, 2020; SKITALINSKAYA; KLAFF; WACHSMUTH, 2021) dedicaram-se a aplicá-la.

A taxonomia proposta para avaliação da qualidade da argumentação é composta por três dimensões fundamentais: Lógica, Retórica e Dialética. A dimensão **Lógica** se concentra na validade e consistência dos argumentos, bem como na ausência de falácias lógicas. As sub-dimensões incluem: validade, consistência, ausência de falácias lógicas, clareza e precisão. A dimensão **Retórica** se concentra na forma como os argumentos são apresentados, bem como quão persuasivos eles são. As sub-dimensões incluem: persuasividade, adequação ao público-alvo, organização e estruturação do discurso. A dimensão **Dialética** se concentra no diálogo entre as partes envolvidas na argumentação. As sub-dimensões incluem: cooperação, relevância dos argumentos apresentados pelas partes envolvidas, respeito mútuo e justiça. A taxonomia proposta é considerada um ponto de referência comum para pesquisas em avaliação da qualidade da argumentação computacional.

A avaliação automática da qualidade da argumentação tem sido objeto de recentes estudos na área de PLN e os métodos mais utilizados, atualmente, fundamentam-se em algoritmos baseados em SVM e redes neurais artificiais profundas (*Deep Learning*) (SKITALINSKAYA; KLAFF; WACHSMUTH, 2021; WACHSMUTH; WERNER, 2020; LAUSCHER et al., 2020; POTTHAST et al., 2019).

Diante do atual cenário, no qual as pesquisas se concentram em corpus em inglês, geralmente compostos por textos escritos seguindo a norma culta, é pertinente pensar na

³ Para se ter uma ideia desse crescente interesse, o Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (CETIC) observou o comportamento de brasileiros maiores de 16 anos na Internet entre 23 de junho e 8 de julho de 2020. Seus achados mostraram que 49% dos internautas realizavam atividades laborais e 72% buscavam informações relacionadas à saúde na Internet. Para mais informações, consulte: <https://cetic.br/media/docs/publicacoes/2/20200817133735/painel_ti_c_covid19_1edicao_livro%20eintr%C3%B4nico.pdf>

elaboração de corpus de língua portuguesa, de CGU de redes sociais, via aplicação de técnicas baseadas nas dimensões da qualidade, como a taxonomia de Wachsmuth et al. (2017) e construção de recursos linguísticos para a avaliação da qualidade de argumentação.

1.2 Objetivo, Hipóteses e Questão de pesquisa

Dentre as tarefas da mineração de argumentação estão a mineração de unidades de argumento, cujo propósito é o de extrair informações em textos com o objetivo de identificar unidades ou elementos que compõem os argumentos em um discurso (AL-KHATIB et al., 2016; HABERNAL; GUREVYCH, 2015; HABERNAL; GUREVYCH, 2017); a detecção de evidências que apoiam afirmações, que envolve a identificação de trechos de texto ou informações que sustentam ou corroboram uma determinada afirmação ou declaração (RINOTT et al., 2015); e a identificação de relações argumentativas, que envolve a detecção e análise das relações lógicas e argumentativas entre unidades de texto em um discurso ou texto (PELDSZUS; STEDE, 2015). Outros trabalhos classificam esquemas de argumentação, que envolve categorizar e organizar os diferentes tipos de argumentos com base em suas características e estrutura (FENG; LIN; HIRST, 2014); realizam a análise de estruturas gerais de argumentação (WACHSMUTH; KIESEL; STEIN, 2015; STAB; GUREVYCH, 2017a); e fazem a geração de reivindicações, que refere-se ao processo de criar ou formular afirmações ou declarações que sustentam um argumento (BILU; SLONIM, 2016). Além disso, algumas teorias da qualidade da argumentação e algumas dimensões da qualidade foram avaliadas computacionalmente (STAB; GUREVYCH, 2017b; WACHSMUTH; STEIN; AJJOUR, 2017; ZHANG et al., 2016). Entretanto, de acordo com Wachsmuth et al. (2017), ainda não se constituiu um conceito geral para qualidade da argumentação nem uma definição clara de suas dimensões.

Considerando a lacuna existente de estudos que utilizaram a taxonomia proposta por Wachsmuth et al. (2017) (WACHSMUTH; WERNER, 2020) para redes sociais e a escassez de corpus em português para a tarefa de avaliar a qualidade de argumentos, este trabalho tem como objetivo:

Propor, gerar e validar um modelo computacional para a avaliação automática da qualidade de argumentação em *tweets* do domínio da política, em português.

Para tanto, modelos pré-treinados, ferramentas e recursos linguístico-computacionais foram utilizados para a avaliação automática da qualidade de argumentos, no domínio de mensagens de redes sociais (Twitter) relacionadas à política, escritas em português.

As hipóteses subjacentes ao objetivo proposto são as seguintes:

- A avaliação automática da qualidade da argumentação em *tweets* relacionados à política brasileira pode ser realizada de forma precisa e confiável por meio de abordagens computacionais.

- ❑ O ajuste fino de um modelo neural é uma estratégia eficaz para avaliar a qualidade da argumentação em *tweets* políticos brasileiros, proporcionando resultados de alta precisão.
- ❑ A criação de um *córpus* anotado com mensagens políticas e diretrizes específicas de anotação para a tarefa de avaliação da qualidade da argumentação é viável e contribui para o desenvolvimento de recursos apropriados para pesquisas nesse campo.

A questão de pesquisa que esta tese visa responder é:

Qual é o método mais eficaz para prever automaticamente a qualidade da argumentação em *tweets* no domínio da política brasileira a fim de identificar aqueles com bons argumentos?

Assim, com base na taxonomia de Wachsmuth et al. (2017): (1) foram definidas pistas linguísticas que definem um argumento de boa qualidade na dimensão retórica, (2) foi anotado um *córpus* de *tweets* do domínio da política, escritos em português, (3) foram gerados modelos computacionais baseados em *features* ou *fine-tuning* de modelos pré-treinados e (4) conclui-se que é possível determinar um *tweet* com boa qualidade de argumentação (considerando a dimensão retórica) por meio de *fine-tuning* de modelos pré-treinados.

1.3 Organização do trabalho

Para uma melhor compreensão desta tese, os próximos capítulos estão estruturados na seguinte forma:

- ❑ No **Capítulo 2** é apresentado um artigo publicado na Revista de Estudos da Linguagem, onde é detalhado como foi a construção do *córpus* de mensagens do Twitter para a avaliação da qualidade da argumentação, a definição dos critérios linguísticos, a criação de diretrizes e a anotação do *córpus* de mensagens.

SILVA, Cássio Faria da et al. Quality of argumentation in political tweets: what is and how to measure it / Qualidade da argumentação em tweets de política: o que e como avaliar. **Revista de Estudos da Linguagem**, [S.l.], v. 29, n. 4, p. 2537-2586, jul. 2021. ISSN 2237-2083. Disponível em: <<http://www.periodicos.lingua.ufmg.br/index.php/relin/article/view/18263>>. Acesso em: 05 mar. 2023. doi:<<http://dx.doi.org/10.17851/2237-2083.29.4.2537-2586>>.

- ❑ No **Capítulo 3** é apresentada uma breve síntese das pistas linguísticas provenientes da dimensão retórica (WACHSMUTH et al., 2017), empregadas para a elaboração das diretrizes de anotação e avaliação da qualidade da argumentação das mensagens políticas. Também são apresentadas as ferramentas computacionais, disponíveis para o português, aplicadas para a extração das *features* linguísticas.

- No **Capítulo 4** é apresentado um artigo aceito para publicação na revista *Linguamática*, onde são apresentados os experimentos realizados com AM baseados em *features* e os experimentos com um ajuste fino de dois modelos neurais, BERTimbau e RobertaTwitterBR.

SILVA, Cássio Faria da; NERIS, Vânia Paula de Almeida; CASELI, Helena de Medeiros. (no prelo). Classificação da qualidade da argumentação em *tweets* no domínio da política brasileira. **Linguamática**.

- No **Capítulo 5** é apresentada a interface pública proposta para a avaliação da qualidade da argumentação em *tweets*.
- No **Capítulo 6** é retomada a questão de pesquisa, apresentadas as contribuições e limitações deste trabalho, bem como as propostas para trabalhos futuros que originaram-se deste estudo.

1.4 Outra forma de ler este trabalho

A composição de uma tese seguindo o formato de coletânea de artigos permite que se visualize a evolução do trabalho, uma vez que mostra os resultados preliminares e evidencia as sub-etapas que levaram aos resultados. Caso essa seja a intenção de leitura, recomenda-se que os capítulos sejam lidos na sequência em que foram organizados nesta tese.

No entanto, caso se queira ler esta tese seguindo moldes mais tradicionais, a Tabela 1 sintetiza uma possível estrutura considerando as grandes áreas de um trabalho científico em computação.

Áreas de um trabalho científico em computação	Conteúdo abordado nesta tese	Seções
Apresentação do referencial teórico	Introdução à argumentação	Cap. 4 - Seção 1
	Argumentação no Twitter	Cap. 2 - Seção 1
	Trabalhos relacionados	Cap. 4 - Seção 2
	Taxonomia de Wachsmuth em <i>tweets</i> políticos	Cap. 2 - Seção 3
Elaboração do corpus de mensagens políticas	Construção do corpus e diretrizes de anotação	Cap. 2 - Seção 4
	Anotação do corpus de mensagens	Cap. 2 - Seção 5
	Estatísticas da anotação	Cap. 2 - Seção 5.1
	Análise de concordância entre anotadores	Cap. 2 - Seção 5.2
Recursos linguísticos	Pistas linguísticas utilizadas nos experimentos de AM baseados em <i>features</i>	Cap. 3 - Seção 3.1
	Cálculo da pontuação das pistas linguísticas utilizadas nos experimentos de AM baseados em <i>features</i>	Cap. 4 - Seção 3.1
	Ferramentas computacionais utilizadas neste trabalho	Cap. 3 - Seção 3.2
Experimentos	Experimentos realizados com AM baseado em features	Cap. 4 - Seção 4.1
	Experimentos realizados com modelos neurais	Cap. 4 - Seção 4.2
	Interface de avaliação da qualidade da argumentação	Cap. 5
	Resultados	Cap. 4 - Seção 5
Conclusão	Respostas à questão de pesquisa	Cap. 6 - Seção 6.1
	Contribuições	Cap. 6 - Seção 6.2
	Limitações	Cap. 6 - Seção 6.3
	Trabalhos futuros	Cap. 6 - Seção 6.4

Tabela 1 – Grandes áreas de um trabalho científico em PLN e seções desta tese.

Capítulo 2

Qualidade da argumentação em *tweets* de política: o que e como avaliar

Este capítulo apresenta o artigo publicado no número 29(4) da Revista de Estudos da Linguagem (RELIN)¹ e apresenta todo o processo de coleta, definição das diretrizes e anotação do *corp*us. O artigo está dividido em 6 seções. A primeira seção apresenta avaliação da qualidade da argumentação como área de pesquisa e aborda alguns aspectos do estilo das mensagens da rede social Twitter, fonte de dados deste trabalho. A segunda seção apresenta os trabalhos relacionados a esta pesquisa. A terceira seção apresenta os aspectos da qualidade da argumentação aplicados neste trabalho. A quarta seção apresenta, de forma detalhada, todo o processo de construção do *corp*us. A quinta seção apresenta o processo de anotação do *corp*us de acordo com as diretrizes estabelecidas para este propósito. Essa seção também apresenta exemplos de anotação e a planilha utilizada pelos anotadores, além das estatísticas de anotação e análise de concordância entre anotadores. A sexta e última seção apresenta as considerações finais e direcionamentos futuros.

Embora o artigo tenha sido uma publicação conjunta entre o autor desta tese e os anotadores do *corp*us, vale ressaltar que a concepção do artigo e sua escrita têm contribuição majoritária do autor desta tese. Além disso, destaca-se que os editores da RELIN autorizaram expressamente a publicação do artigo, em sua íntegra, nesta tese.

¹ Disponível em: <<http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/18263>>

Rev. Estud. Ling., Belo Horizonte, v. 29, n. 4, p. 2537-2586, 2021



Quality of argumentation in political tweets: what is and how to measure it

Qualidade da argumentação em tweets de política: o que e como avaliar

Cássio Faria da Silva

Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo / Brazil

cassiofs@gmail.com

<http://orcid.org/0000-0002-9420-8608>

Amanda Pontes Rassi

Redação Nota 1000 Ltda., São Paulo, São Paulo / Brazil

amanda@redacaonota1000.com.br

<http://orcid.org/0000-0001-5314-1868>

Jackson Wilke da Cruz Souza

Universidade Federal de Alfenas (UNIFAL-MG), Varginha, Minas Gerais / Brazil

jackson.souza@unifal-mg.edu.br

<http://orcid.org/0000-0003-1881-6780>

Renata Ramisch

Redação Nota 1000 Ltda., São Paulo, São Paulo / Brazil

renata.ramisch@gmail.com

<http://orcid.org/0000-0003-3372-6150>

Roger Alfredo de Marci Rodrigues Antunes

Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo / Brazil

rogerantunes@pm.me

<http://orcid.org/0000-0002-4735-3941>

Helena de Medeiros Caseli

Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo / Brazil

helenacaseli@ufscar.br

<http://orcid.org/0000-0003-3996-8599>

eISSN: 2237-2083

DOI: 10.17851/2237-2083.29.4.2537-2586

Abstract: Argumentation is something inherent to human beings and essential to written and spoken communication. Because of the popularization of Internet access, social media are one of the main means of creation and profusion of argumentative texts in various fields, such as politics. As a way to contribute to research related to the assessment of the quality of argumentation in Portuguese, we aim in this paper to propose and validate criteria and guidelines for the assessment of the quality of argumentation in Twitter posts in the domain of politics. For this purpose, a *corpus* was produced and annotated with tweets whose content is related to the Brazilian political scenario. The texts were collected in the first months of 2021, resulting in 1,649,674 posts. From the analysis of a sample, we defined linguistic criteria that would potentially characterize relevant aspects of the rhetorical dimension of argumentation, namely: (i) Clarity, (ii) Arrangement, (iii) Credibility, and (iv) Emotional appeal. After this phase of analysis, we proposed the annotation of a new set of 400 tweets, by four annotators. As a result, an agreement of around 70% for three out of four annotators was obtained. It is worth noting that this is the first work that proposes linguistic criteria for the evaluation of the quality of argumentation in social medias for Brazilian Portuguese. It is intended to construct a computer model that can automatically evaluate the quality of argumentation in social media messages, such as Twitter, based on the establishment of linguistic criteria, annotation rules, and annotated corpus.

Keywords: argumentation; *corpus*; quality; rhetorical dimension; tweets; politics.

Resumo: A argumentação é algo inerente ao ser humano e essencial para a comunicação escrita e falada. Por conta da popularização do acesso à Internet, as redes sociais são um dos principais meios de criação e profusão de textos argumentativos de vários domínios, como a política. Como forma de contribuir com as pesquisas relacionadas à avaliação da qualidade da argumentação em português, este trabalho tem como objetivo propor e validar critérios e diretrizes para a avaliação da qualidade da argumentação em postagens no Twitter no domínio da política. Para tanto, produziu-se um *corpus* anotado com *tweets* cujo conteúdo relaciona-se ao cenário político brasileiro. Os textos foram coletados nos primeiros meses de 2021, resultando em 1.649.674 postagens. A partir da análise de uma amostra, foram definidos critérios linguísticos que potencialmente caracterizariam aspectos relevantes da dimensão retórica da argumentação, a saber: (i) Clareza, (ii) Organização, (iii) Credibilidade e (iv) Apelo emocional. Após essa fase de análise, propôs-se a anotação de um novo conjunto de 400 *tweets*, por quatro anotadores. Como resultado, obteve-se uma concordância de cerca de 70% entre 3 dos 4 anotadores. Vale ressaltar que esse é o primeiro trabalho que propõe critérios linguísticos para a avaliação da qualidade da argumentação em redes sociais para o português brasileiro. A partir da definição dos critérios linguísticos, diretrizes de anotação e *corpus* anotado, espera-se construir um modelo computacional que possa avaliar automaticamente a qualidade da argumentação em textos de redes sociais, como o Twitter.

Palavras-chave: argumentação; *corpus*; qualidade; dimensão retórica; *tweets*; política.

Submitted on March 30th, 2021

Accepted on May 24th, 2021

1 Introduction

Argumentation is inherent to human beings and is present in all types of oral and written communication. As a research area, argumentation is a multidisciplinary field that studies debate and reasoning processes. An argument is a claim (or conclusion) accompanied by a random number of premises that justify, substantiate, support, defend, or explain the claim (POTTHAST *et al.*, 2019). Well-founded arguments are not only important for decision making and learning, but also play a key role in reaching widely accepted conclusions. For Eemeren and Grootendorst (2003), argumentation consists of one or more sentences in which several premises are presented to support a conclusion. The sentences that are part of the argumentation constitute a complete expression that aims to convince an interlocutor.

As a research field, works in Linguistics focus on the analysis of arguments in natural language texts (STAB; GUREVYCH, 2017a). In Artificial Intelligence, the identification of arguments and the automatic evaluation of argumentation are investigated (BENCH-CAPON; DUNNE, 2007) by combining representational models and user-related cognitive models, and computational models for automated reasoning.

Through Natural Language Processing (NLP), investigations have been carried out in order to (i) identify arguments and their units, (ii) generate and (iii) evaluate the quality of arguments for both formal texts and User Generated Content, especially from social media. Computational argumentation-related tasks such as mining, generation, identification of arguments and their evaluation prove to be relevant in activities such as writing support and discussion assistance (GARCÍA-GORROSTIETA; LÓPEZ-LÓPEZ, 2018; GARCÍA-GORROSTIETA *et al.*, 2018; STAB; GUREVYCH, 2017b). Most of the current works focus on argument mining and handling formal texts in English.

However, a significant source of data for many of the disciplines interested in argumentation-related studies is the Web, and particularly social media. Social media, discussion forums, online news, and product reviews provide a heterogeneous and expanding source of information, in which user-generated arguments can be identified, isolated, and analyzed.

The availability of this data, combined with advances in NLP and Machine Learning, has created a promising scenario for the emergence of a lot of research on argumentation (or argument) mining.

According to some evidence (LYTOS *et al.*, 2019), the Internet and social media are the most important means of communication today, and as a result, they are the source of a large volume of argumentative texts across a wide range of subjects. In particular, social media, being communication spaces in which users produce their texts conditioned to certain linguistic, structural, and style standards given by the community's own communicative behavior, can be understood not as a text holder, but as Writing Genre (WG) (FREITAS; BARTH, 2015).

From this perspective, the standards that are established adapt the very concept of argumentation in a WG like Twitter, for representing the linguistic materialization of a communicative necessity of language users in a given situation and given historical context (MARCUSCHI, 2002), as shown in (1).¹

- (1) @CarlaZambelli38 @jairbolsonaro **Kkk gasosa a 5,09 reais e tu pede p ter confiança ainda. Deputada, 2 anos e nada mudou, o BANDO domina e o mito ou melhor, o MINTO JA SE RENDEU AO SISTEMA P PROTEJER O FILHOTE LADRÃOZINHO, QTO AOS GALS IMPRESTÁVEIS CAGAM E ANDAM P POVO, O FORO SAO PAULO VENCEU E NÓS SIFU.....**
 [@CarlaZambelli38 @jairbolsonaro **lol gas 5.09 reais and u still ask 4 your trust . Deputy,2 yrs n' nothing changed, the GANG dominates everything and the myth, or better saying, the DISHONESTY ITSELF GOT SURRENDERED BY THE SYSTEM TO PROTECT HIS LITTLE THIEF BOY,AS 4 THE WORTHLESS GALS WHICH ARE GIVIN' A SHIT 2 THE NATION,THE FORO OF SAO PAULO WON AND WE'RE SCREWED.....]**

¹ All examples in this paper are presented first in the original language (Brazilian Portuguese), then in English. The English version was produced trying to preserve as much as possible the original linguistic, semantic and emotional features present in the original message.

In (1), we identify (i) orality marks that emerge on the textual surface in “kkk (lol)” and “nós sifu”,² (ii) informality in constructions like “tu pede”³ and “cagam e andam”,⁴ indicating that there is no concern in using the standard polite modality of the language, which includes typical abbreviations of internetese (“p” indicating “para”),⁵ (iii) enunciative instantaneousness both in the emergence of the subject and in the way of reference to it (“gasosa a 5,09 reais”)⁶ and (iv) interlocutionary acts, since there are strategies of interpellation and/or argumentation of the author of the post about the reader, as in direct references to the interlocutor through “tu (you)” and “Deputada (Deputy)”. Like the WG itself, the notion of argumentation is adapted to the communicative needs of language users, being understood as the clear expression of a position or opinion about a given subject in any and all Twitter posts.

Other aspects concerning the texts published on Twitter are related to the specificities that this social media implies about the texts. The possibility of publications being linked to each other, especially replies, makes the text manifest some linguistic characteristics of its own. Authors can retrieve the main subjects and the people related to them using deictics (e.g. demonstrative pronouns), manifest the presence of knowledge or information without citing the source and use argumentative strategies in syntactic constructions not always equivalent to the formal analysis of the language (such as adverbial clauses of conformity).

In this sense, it is worth questioning if the argumentative strategies in Twitter posts show quality, in terms of clarity, arrangement and credibility, since they often count on the use of a negative emotional appeal, especially in matters of political domain. Later on we will explain in detail what we consider to be a domain of politics but briefly we consider belonging to a political domain the posts of Brazilian congressmen from different parties, that is, political-party agents occupying elective mandates in the Federal Chamber, as well as replies of the followers to the politician’s post.

With regard to argument evaluation, since Toulmin’s (2003) argument schema, studies have been conducted to simplify the

² “we’re screwed”.

³ “u ask”.

⁴ “givin’ a shit”.

⁵ “for”.

⁶ “lol gas 5.09 reais”.

understanding of the structure and determine the importance of argumentative text elements. Recently, Wachsmuth *et al.* (2017b) proposed a taxonomy consisting of three dimensions to rate the quality of argumentation regarding some aspects. However, since then, few studies have been dedicated to apply it, much less in WGs whose texts show unstructured contents which are far from the standard linguistic norm and from the conventional notion of argumentation itself.

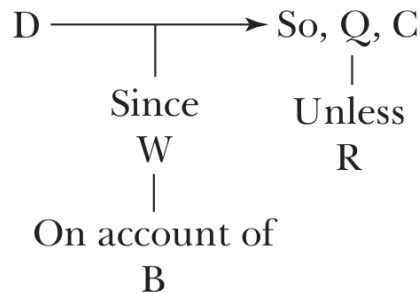
In order to contribute to the studies of argumentation in interface with interaction in digital media, this paper aims to review the taxonomy of Wachsmuth *et al.* (2017b) and adapt it to a WG with features such as those of Twitter. Furthermore, based on the linguistic analysis of the results that will be discussed in this work, we will be able to contribute in future works with the automatic assessment of the quality of the argument in Twitter posts in the field of Brazilian politics.

For this purpose, this article was organized in five sections, besides this introduction. In section 2, we present the works related to this research as a theoretical foundation. In section 3 we present the taxonomy proposed by Wachsmuth *et al.* (2017b), on which we base ourselves in the present paper. In section 4, we describe the corpus for analysis, characterized by being Twitter posts. In section 5 we describe the posting annotation guidelines, as well as presenting the disagreements between the annotators. Finally, in section 6, we make some final considerations, in addition to pointing out future works.

2 Theoretical Foundation

Toulmin's Argument Model (2003) proposes a set of elements that constitute an argument and the links established among them. The data (D), the conclusion (C), and the warrant (W) are the three basic elements that make up an argument. In other words, if a warrant (W) is obtained from data (D), it is possible to conclude C. In addition to the fundamental elements, it is possible to specify the conditions under which the justification provided is valid or not by using qualifiers (Q). It is also possible to present a refutation (R) of the justification. The backing (B) is a claim (guarantee) based on some verified valid information that is intended to support and substantiate the justification. Figure 1 illustrates each of these elements that compose an argument, as well as the correlations between them, represented by the arrows.

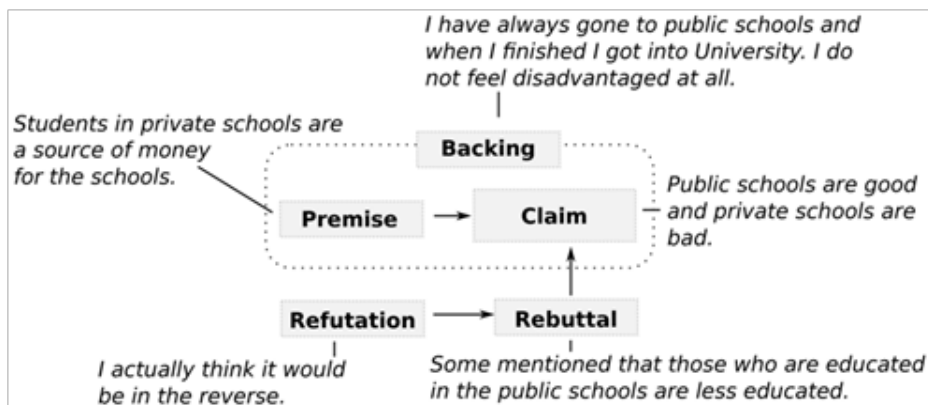
FIGURE 1 – Toulmin’s Argument Model (2003)



Source: Toulmin (2003, p. 97).

Habernal and Gurevych (2017) proposed a modified model, based on Toulmin’s (2003) argument model, in order to annotate a *corpus* of arguments extracted from online discussion forums. Figure 2 illustrates the modified model used for the annotation of arguments with an example instantiated from a single discussion forum post on the topic “public vs. private schools”. The arrows are used to illustrate the relationships between the elements of the argument (HABERNAL; GUREVYCH, 2017).

FIGURE 2 – Example of annotation using Toulmin’s modified model



Source: Habernal and Gurevych (2017, p. 144).

Evaluating the validity, quality, and strength of arguments represents a challenge inherent to argumentative discourse. It is worth noting that there are strong theoretical foundations and various normative theories to support the task, such as: (i) the mentioned argumentative model of Toulmin (2003); (ii) Walton’s schemes and their critical issues

(WALTON; WALTON, 1989); (iii) the ideal model of critical argument in the pragma-dialectical approach, in which fallacies are considered incorrect moves in a discussion whose goal is the successful resolution of a dispute (EEMEREN; GROOTENDORST, 1987); and (iv) the study of fallacies (BOUDRY *et al.*, 2015). However, judging qualitative criteria of everyday argumentation still represents a challenge for argumentation scholars and practitioners (ROSENFELD; KRAUS, 2015; SWANSON *et al.*, 2015; WELTZER-WARD *et al.*, 2009).

2.1 Evaluating the Quality of Argumentation

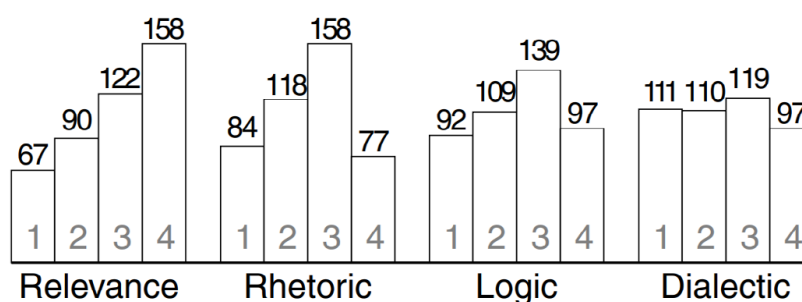
The already proposed methods and techniques for assessing the quality of arguments do not settle on which criteria should be considered nor on whether quality should be assessed from a theoretical or practical point of view. Wachsmuth *et al.* (2017a) aim to elucidate, by searching for empirical answers, the question of how different theoretical and practical views of argument quality are. In that work, Wachsmuth *et al.* demonstrate that argumentation quality can be observed from practical and theoretical aspects. From the theoretical perspective, conviction is understood as the main logical quality, and the authors support the fact that theory-based assessment of argumentation quality remains complex. They also point out that practical approaches indicate on what to focus to simplify theory, while theory seems beneficial in guiding the evaluation of quality in practice.

In the same direction, other studies seek to rate the relevance of arguments, in which argumentative sentences are identified and the importance of their arguments is assessed. Potthast *et al.* (2019) assessed the degree of relevance of a set of arguments. In addition, the relevance and the rhetorical, logical, and dialectical quality of the arguments were evaluated. The *args.me corpus*,⁷ built by Wachsmuth *et al.* (2017c), was used for the task. Forty annotators evaluated the relevance of each of the 437 arguments related to 40 selected topics, in addition to their rhetorical, logical, and dialectical quality. From the 437 annotated arguments, 208 were marked in favor and 195 opposed, in addition to 34 that were annotated as non-argumentative by the annotators. The relevance ratings, in addition to the three dimensions, are displayed in Figure 3, where

⁷ Available in: www.args.me

the distribution of the scores (from 1 to 4) can be seen. The relevance scores indicate that many highly relevant arguments (scored as 4) were retrieved from the adopted *corpus* and that the annotation of the dialectical dimension is controversial or the guidelines were unclear since the ratings were uniform. Other works also sought evaluation under the relevance aspect of argumentative texts (GLEIZE *et al.*, 2019; WACHSMUTH *et al.*, 2017d).

FIGURE 3 – Score distributions by relevance and quality dimensions



Source: Potthast *et al.* (2019, p. 1120).

On the other hand, Habernal and Gurevych (2016) suggest that the evaluation of argument quality should be done by comparing arguments. Other works report assessments of the quality of individual arguments with satisfactory results (PERSING; NG, 2015; WACHSMUTH *et al.*, 2017b).

More recent works have used a structured taxonomy aiming the assessment of individual aspects based on the characteristics of the argument structure, such as the emotional appeal employed, the arrangement of the sentence, and the credibility of the message author (LAUSCHER *et al.*, 2020; WACHSMUTH *et al.*, 2017b; WACHSMUTH; WERNER, 2020).

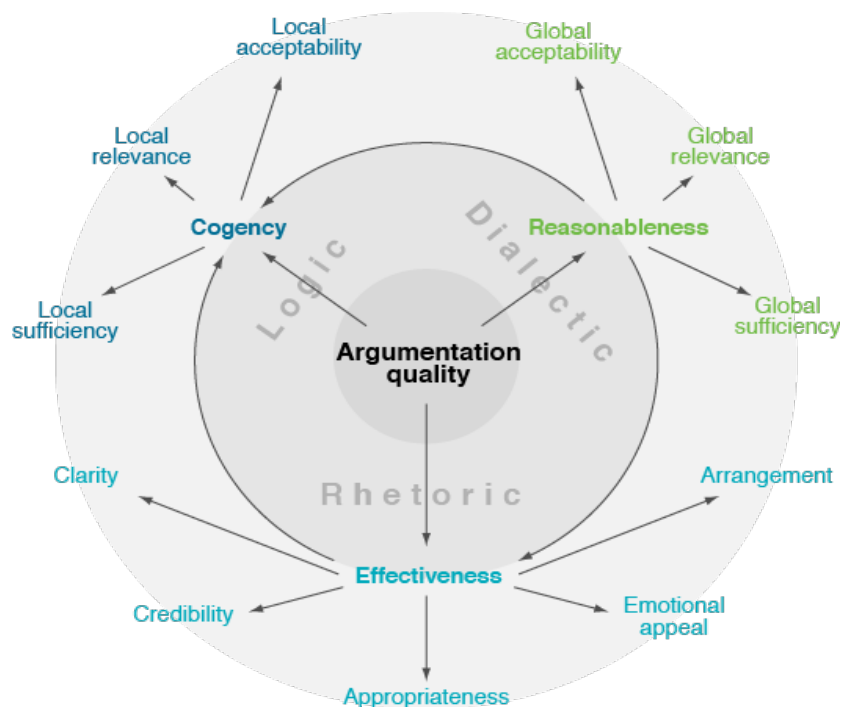
Works in the literature have investigated the quality of arguments in various domains; however, none have specifically addressed user-generated content, on social media, in the domain of politics in Brazilian Portuguese (BP). Other approaches address the task of assessing argumentation quality in messages from discussion forums and debate portals (WEI *et al.*, 2016; HABERNAL; GUREVYCH, 2016) and student writings (STAB; GUREVYCH, 2017b; CARLILE *et al.*, 2018; WACHSMUTH *et al.*, 2016), which, in our view, are less challenging

than tweets in the domain of Brazilian politics today, primarily because tweets have a very limited amount of characters, which makes it more difficult to use linguistic argumentation strategies and secondly because politics have become even more polarized and aggressive recently in Brazil, constantly using uncivil and intolerant discourse (ROSSINI, 2019, 2020). As an attempt to cover such a gap, this paper describes the construction of a *corpus* composed by tweets related to the Brazilian political scenario, as well as the definition of criteria and guidelines regarding the evaluation of the rhetorical quality of arguments present in this *corpus*.

2.2 Taxonomy of Wachsmuth *et al.* (2017b)

Wachsmuth *et al.* (2017b) conducted a research on the quality of arguments considering both argumentation theory and argument mining perspectives. Based on this study, the Argument Quality Taxonomy was proposed, whose dimensions are used to define “quality”. Figure 4 illustrates this taxonomy, with all its dimensions.

FIGURE 4 – Argumentation Quality Taxonomy



Source: Wachsmuth *et al.* (2017b, p. 181).

According to this taxonomy, the quality of argumentation can be divided into the logical, rhetorical and dialectical dimensions (BLAIR, 2012), described below.

The **logical dimension** refers to the structure and composition of an argument. An argument of high logical quality is based on acceptable premises and combines them in a convincing way to support the claim of the argument. It is related to the logical irrefutability of the argument.

The **rhetorical dimension**, in contrast, includes notions of persuasive effectiveness, correct language, accuracy, and style. An argument of high rhetorical quality is well-written and attractive to the audience and is related to the rhetorical effectiveness of the argument. An argument is rhetorically effective if it convinces the target audience of (or corroborates the agreement with) the author's position on the issue.

The **dialectical dimension** captures an argument's contribution to the discourse. An argument of high dialectical quality is useful for supporting cooperative decision making or for resolving conflict. The argument is reasonable if it contributes to the resolution of the issue in a sufficient manner that is acceptable to the target audience.

Wachsmuth *et al.* (2017b) tested the taxonomy in an annotation experiment, using data from the UKPConvArgRank⁸ *corpus* by Habernal and Gurevych (2016). The UKPConvArgRank *corpus*, developed for argument comparison, contains argument ratings from the debate portals createdebate.com and convinceme.net, both written in English. Each debate topic has two opinions: one for and one against the main topic. The final *corpus*, called Dagstuhl-15512-ArgQuality,⁹ developed from the UKPConvArgRank, contains 320 argumentative texts with scores assigned by three annotators for the 15 aspects of the taxonomy. In this annotation process, each text was first classified as argumentative or not. Then, for the argumentative texts, all aspects were assessed using scores from 1 (low), 2 (medium) to 3 (high), plus the option "I cannot judge".

In Figure 5, we can see the scores assigned by the three annotators (A, B and C) on two texts produced in response to the question "should plastic water bottles be banned?". The highest value in each column is

⁸ *Corpus* UKPConvArgRank available in: <https://github.com/UKPLab/acl2016-convincing-arguments>

⁹ *Corpus* Dagstuhl-15512-ArgQuality available in <http://arguana.com/>

marked in bold. The bottom row represents the majority vote of the three annotators.¹⁰

FIGURE 5 – Scores of each annotator and majority score for all quality dimensions

Arguments	Pro Water bottles, good or bad? Many people believe plastic water bottles to be good. But the truth is water bottles are polluting land and unnecessary. Plastic water bottles should only be used in emergency purposes only. The water in those plastic are only filtered tap water. In an emergency situation like Katrina no one had access to tap water. In a situation like this water bottles are good because it provides the people in need. Other than that water bottles should not be legal because it pollutes the land and big companies get 1000% of the profit.														Con Americans spend billions on bottled water every year. Banning their sale would greatly hurt an already struggling economy. In addition to the actual sale of water bottles, the plastics that they are made out of, and the advertising on both the bottles and packaging are also big business. In addition to this, compostable waters bottle are also coming onto the market, these can be used instead of plastics to eliminate that detriment. Moreover, bottled water not only has a cleaner safety record than municipal water, but it easier to trace when a potential health risk does occur. (http://www.friendsjournal.org/bottled-water) (http://www.cdc.gov/healthywater/drinking/bottled/)																
	Scores	Co	LA	LR	LS	Ef	Cr	Em	Cl	Ap	Ar	Re	GA	GR	GS	Ov	Co	LA	LR	LS	Ef	Cr	Em	Cl	Ap	Ar	Re	GA	GR	GS	Ov
Annotator A	3	3	3	2	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3
Annotator B	2	2	3	2	1	2	2	2	2	1	2	2	2	1	2	2	3	3	2	2	3	2	3	3	2	3	3	2	2	3	3
Annotator C	2	3	3	2	2	2	2	3	3	3	3	3	3	3	2	3	3	3	3	3	3	2	1	3	3	3	3	3	3	3	3
Majority score	2	3	3	2	2	2	2	3	3	3	3	3	3	3	2	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3

Source: Wachsmuth *et al.* (2017b, p. 184).

Table 1 shows the results of this annotation experiment for the 304 texts of the *corpus* classified as argumentative by all annotators: (a) Distribution of majority scores for each dimension; (b) Krippendorff's α used to measure the agreement among annotators; (c) Correlation for each pair of dimensions, calculated based on the average of the correlations of all annotators. The highest value in each column is highlighted in bold.

¹⁰ The Logic dimension measures Conviction (Co) and is composed of 3 aspects: Local Acceptability (LA), Local Relevance (LR) and Local Sufficiency (LS). The Rhetorical dimension measures Effectiveness (Ef) and is composed of 5 aspects: Credibility (Cr), Emotional appeal (Em), Clarity (Cl), Appropriateness (Ap) and Arrangement (Ar). Finally, the Dialectical dimension measures Reasonableness (Re) and is composed of 3 aspects: Global Acceptability (GA), Global Relevance (RG) and Global Sufficiency (GS).

TABLE 1 – Results for the 304 texts of the *corpus* classified as argumentative by all annotators

Quality Dimension	(a) Maj. Scores			(b) Agreement			(c) Pearson Correlation Coefficients													
	1	2	3	α	full	maj.	Co	LA	LR	LS	Ef	Cr	Em	Cl	Ap	Ar	Re	GA	GR	GS
Co Cogency	150	131	23	.44	40.1%	91.8%	.64	.61	.84	.81	.46	.27	.41	.32	.55	.78	.64	.71	.70	
LA Local acceptability	84	169	51	.46	27.0%	90.8%	.64	.54	.53	.60	.54	.30	.40	.54	.46	.68	.75	.46	.45	
LR Local relevance	25	155	124	.47	32.6%	92.4%	.61	.51	.56	.56	.39	.27	.46	.35	.50	.62	.58	.68	.45	
LS Local sufficiency	172	119	13	.44	37.2%	92.8%	.84	.53	.56	.73	.39	.25	.37	.23	.51	.67	.51	.68	.74	
Ef Effectiveness	184	111	9	.45	42.1%	94.4%	.81	.60	.56	.73	.48	.31	.35	.34	.54	.75	.58	.66	.71	
Cr Credibility	99	199	6	.37	37.8%	95.7%	.46	.54	.39	.39	.48	.37	.32	.49	.37	.52	.52	.36	.40	
Em Emotional appeal	48	235	21	.26	42.8%	94.4%	.27	.30	.27	.25	.31	.37	.14	.30	.20	.30	.26	.26	.22	
Cl Clarity	42	191	71	.35	29.3%	89.8%	.41	.40	.46	.37	.35	.32	.14	.45	.56	.44	.45	.38	.27	
Ap Appropriateness	43	196	65	.36	17.4%	87.5%	.32	.54	.35	.23	.34	.49	.30	.45	.48	.47	.59	.20	.20	
Ar Arrangement	91	189	24	.39	26.6%	93.4%	.55	.46	.50	.51	.54	.37	.20	.56	.48	.55	.51	.49	.48	
Re Reasonableness	126	159	19	.50	41.4%	95.7%	.78	.68	.62	.67	.75	.52	.30	.44	.47	.55	.78	.65	.61	
GA Global acceptability	88	161	55	.44	31.6%	95.4%	.64	.75	.58	.51	.58	.52	.26	.45	.59	.51	.78	.46	.43	
GR Global relevance	69	167	68	.42	21.7%	90.1%	.71	.46	.68	.68	.66	.36	.26	.38	.20	.49	.65	.46	.61	
GS Global sufficiency	231	72	1	.27	44.7%	98.0%	.70	.45	.45	.74	.71	.40	.22	.27	.20	.48	.61	.43	.61	
Ov Overall quality	152	128	24	.51	44.1%	94.4%	.84	.66	.61	.74	.81	.52	.30	.45	.42	.59	.86	.71	.70	.68

Source: Wachsmuth *et al.* (2017b, p. 183).

It is emphasized that the proposed taxonomy is intended to classify all aspects of argumentation quality, regardless of how they may be operationalized. Considering the variation in agreement values among annotators on some dimensions, it is understood that some of them are particularly subjective and challenging.

For the investigation of the applicability of Wachsmuth *et al.* (2017b) taxonomy to the evaluation of the quality of argumentation in Twitter posts in the domain of politics in BP, the rhetorical dimension was chosen. This decision was based on the fact that the rhetorical dimension presents evidence that computational implementation based on linguistic cues is possible. According to Wachsmuth *et al.* (2017b), the aspects that constitute the rhetorical dimension are related to the emotional appeal applied in the argumentation, ambiguity, imprecision, language style and the organization of the text structure. Therefore, it is understood that these characteristics can be, to some extent, identified through linguistic resources.

The rhetorical dimension, according to Wachsmuth *et al.* (2017b), has five aspects:

1. **Credibility (Cr)** – Credibility refers to how the author conveys his arguments and makes them credible. An appropriate style in terms of word choice supports credibility (WACHSMUTH *et al.*, 2017b). Also according to Wachsmuth *et al.* (2017b), aspects that can be considered to assess credibility are the honesty of the

author of the message, the politeness of the language used, or the author's knowledge and experience regarding the issues discussed.

2. **Emotional appeal (Em)** – Emotional appeal is considered successful in an argument if it creates emotions in a way that makes the target audience more open to the author's arguments.
3. **Clarity (Cl)** – Clarity refers to using language that is grammatically correct and largely unambiguous, and avoids unnecessary complexity and detour from the issue discussed. The language used should facilitate understanding and leave no doubt about the author's position and the way he or she defends that position.
4. **Adequacy (Ap)** – The adequacy of an argument refers to the language (form and content) used to support the creation of credibility and emotions, as well as the appropriateness to the issue discussed.
5. **Arrangement (Ar)** – An argumentation is considered adequately organized if it presents the question, the arguments, and the conclusion in the correct order.

It is important to note that the *corpus* of messages assessed in the study of Wachsmuth *et al.* (2017b) is composed of messages from online discussion forums, which are characterized by being longer messages, unlike the scenario of this work, in which the evaluation of user-generated content from Twitter is proposed, with a limit of no more than 280 characters.

In this work, we propose and validate criteria and guidelines for evaluating the quality of argumentation in tweets produced as replies for posts from Brazilian deputies in the field of politics collected from 06th February to 07th March 2021. This validation, in the future, will support a computational model to evaluate the rhetorical dimension defined by the taxonomy of Wachsmuth *et al.* (2017b).

3 Taxonomy of aspects of argumentative quality in political tweets

As a proposal for evaluating the quality of argumentation, we defined criteria for each of the four aspects of the rhetorical dimension of Wachsmuth *et al.* (2017b) taxonomy that proved most relevant for the

domain of politics in tweets, namely: Clarity, Arrangement, Credibility and Emotional appeal. The Adequacy was not considered in this work since it proved not to be relevant for quality argumentation in tweets, and also because the criteria pertaining to Adequacy are already covered by the other four aspects.

From an initial study on a set of 30 tweets from the domain of politics in BP, the team of four annotators proposed criteria based on linguistic cues for the aspects of the rhetorical dimension proposed by Wachsmuth *et al.* (2017b). Although the amount of tweets initially analyzed was small, it was possible to observe that some aspects are naturally present in the investigated WG in BP, while others need to be explicitly constructed.

When sharing information on this social media, users spread emotional triggers that reinforce beliefs or even prejudices, not drawing on the credibility of the content conveyed (WARDLE, 2019). While Twitter users considered unmoderated would use the term “Bozo”¹¹ to refer to the current president of Brazil, users considered moderated would tend to use less commotion to cover up opinions (FREEDOM HOUSE, 2019), which would lead to a possible author referring to the same entity as “the president of the Republic”.

Brady *et al.* (2017) point out that messages that feature moral-emotional language may be more widespread, especially in political groups that share similar ideologies. However, when faced with issues diverging from their own ideological perspectives, users adopt strategies of attacking political figures in an attempt to discredit them, making them personal enemies.

In this sense, it was assumed that Clarity is inherent to the text, while Arrangement and Credibility are not, and they must be built through explicit linguistic artifacts. As for Emotional appeal, the annotators agreed to analyze separately its polarity (positive or negative) and its intensity (low, medium, or high).

From this initial analysis, in cycles of daily 1-hour meetings over a period of two weeks, the annotators defined and refined criteria indicating the presence or absence of each criteria. The result of this analysis is presented in the following subsections.

¹¹ “Bozo” is a pejorative way to refer to the current Brazilian president Jair Bolsonaro.

3.1 Clarity

According to Wachsmuth *et al.* (2017b), an argument should be assessed as clear if it uses grammatically correct and largely unambiguous language, and avoids unnecessary complexity and deviation from the issue discussed. The language used should facilitate understanding and leave no doubt about the author's position and the way he or she defends that position.

For the evaluation of the Clarity aspect, it was considered that every argument written in Portuguese has the potential to be naturally clear, unless there are certain criteria that negatively interfere with clarity. In this way, every tweet starts from a high level of Clarity, which decreases as the presence of one or more criteria that harm the clarity of the argumentation is noted, namely: *question leading to doubt*, *unnecessary complex language*, *presence of Portuguese language deviations*, and *unnecessary deviation from the subject*.

The criterion called *question leading to doubt* harms the clarity of the argumentation because it does not make the author's true position on a given subject explicit, as, on the textual surface, the opinion is not in an affirmative declarative sentence, but an interrogative one. In (2), we see an example of several questions that do not clearly express an opinion and, therefore, lead to doubt, while (3) brings a counterexample, that is, a question that does not lead to doubt. In (4), there is an interrogative structure, even in the absence of the corresponding punctuation (in this case, the question mark).

- (2) @MarceloFreixo Quem usou os cargos públicos para roubar foi o PT, quase 1 trilhão de reais. **Aliás, como anda o Rio ? Bala perdida para todo lado ? Quais suas obras para tirar a cidade do buraco que está pelo narcotráfico ?** Décadas e nada de agregar ao Rio, você deveria mudar de ramo.

[@MarceloFreixo Who used the public offices to steal was the PT, almost 1 trillion reais. **By the way, how is Rio doing ? Bullets stray everywhere ? What are your works to get the city out of the hole it's in because of drug trafficking ?** Decades and nothing to add to Rio, you should change your business.]

- (3) MarceloFreixo Com certeza ele nunca agiu sozinho, isso está cheirando a balão de ensaio, já que o Bolsonaro não pode mais ficar se expondo, como ele sempre teve seus leões de chácara, o jogo dele não vai parar, agora, a questão, o filho e o próprio Bolsonaro cometeram crimes semelhantes, **e aí?**

[@MarceloFreixo For sure he never acted alone, this is reeking of a trial balloon, since Bolsonaro can't expose himself anymore, as he always had his bouncers, his game will not stop, now, the question, the son and Bolsonaro himself committed similar crimes, **so what?**]

- (4) @marcelvanhattem Ou o congresso volta a protagonismo de legislar Do contrário fechadas as portas e deicha o STF legislar investigar prender julgar condenar absorver até mesmo primeiro é segunda instância do judiciário **p/ que serve** se o STF anula todo um trabalho feito ao em vez de se somar divid

[@marcelvanhattem Either Congress returns to the leading role of lawmaking Otherwise the doors are closed and the STF is left to legislate investigate arrest try convict and even absorb the first and second instance of the judiciary **what good is it** if the STF nullifies all the work done instead of adding up, divid]

The use of *unnecessary complex language* was also identified as a criterion that negatively affects the clarity of the argument. Thus, the presence of a word that is too far-fetched and unusual or not appropriate for the context, or a very complex syntactic structure, with many dislocated and/or embedded clauses, which affects the understanding of the argument, can negatively interfere in clarity. In example (5), the reference to “inqueritos do fim do mundo”,¹² the use of the word “imbróglio (imbroglio)”, which, although used correctly, is very fanciful and unusual, and the metaphorical reference to “música que tocam para o PR”¹³ stand out as unnecessary complex language.

¹² “end-of-the-world surveys”

¹³ “the music they play for the PR”

- (5) @carlosjordy Os poderosos que movimentaram seus peões contra o deputado Daniel Silveira e todas as vítimas dos **inquéritos do fim do mundo**, fazem cara de paisagem, pedindo por mais reformas? Terão que primeiro resolver esse **imbróglio**. É essa música que tocam para o PR @jairbolsonaro ?

[@carlosjordy The powerful who have moved their pawns against Congressman Daniel Silveira and all the victims of **the end of the world inquiries**, look on with a straight face, asking for more reforms? They will have to solve this **imbroglio** first. **Is that the music they play for PR @jairbolsonaro ?**]

The criterion entitled *Portuguese language deviations* covers errors in various levels, such as spelling, syntax, punctuation, etc., that impair the reader's understanding of the argument. The good quality of the language, identified by the correct use of punctuation, syntax, spelling, etc., contributes positively to the clarity of the argument. Thus, the clarity of the argument is weakened by the presence of errors that hinder comprehension.

In example (4), we identified several deviations in the use of the language, such as lack of proper punctuation (commas, period and question mark), spelling mistakes (“deicha”, “absorver”, “divid”), accentuation problem (“é” instead of “e”), syntactic deviations in concordance or verbal regency (in “volta a protagonismo”), among others.

It should be noted, however, that some words are abbreviated on purpose by users, since Twitter has a restriction on the number of characters. This can be observed in the case of “p/”, in example (4), which corresponds to “para”. This type of strategy was not considered a deviation of the Portuguese language and, therefore, did not penalize clarity, since they are typical strategies of the WG under consideration.

Another aspect that undermines the clarity of the argumentation is the *unnecessary deviation from the subject*, because, in a clear post, it is expected that the author uses only arguments relevant to the topic under discussion. In this sense, a deviation from this issue should be penalized in relation to clarity. This criterion should be analyzed considering the issue of the seed tweet. In (2), for example, the main topic is the use of public offices to commit illegalities, but the author deviates from the subject several times to make personal attacks on the congressman who

wrote the seed tweet, as in “Aliás, como anda o Rio ? [...] Décadas e nada de agregar ao Rio, você deveria mudar de ramo”.¹⁴

From these four criteria, it was defined that the clarity of the argumentation is low when three or more of the criteria are present, medium when two of the criteria is present, and high when none or only one of the four criteria is present.

3.2 Arrangement

According to Wachsmuth *et al.* (2017b), an argument should be evaluated as well organized if it presents the subject, the arguments and its conclusion in the correct order. This definition is traditionally accepted for most dissertative genres but cannot be strictly followed in genres such as tweets. Thus, it was necessary to adapt this concept for the purposes of this paper.

Before debating and concluding on a topic, it is thought that the general issue and the specific topic should be understood. In tweets, however, other sequences can be used on purpose and still be adequate to persuade the target audience. Moreover, some parts of the proposition may be clear (e.g., the topic under discussion) and therefore not be explicitly mentioned in the comment, but rather left implicit.

Given the characteristics of Twitter, where the user has a limited space to express an opinion, it is assumed that tweets are not well-structured texts. Thus, for a post to be assessed as well organized, it must contain certain criteria that positively impact the quality of the arrangement. These criteria were defined based on the presence of discourse markers or cohesive resources that explain the flow of discourse by creating following relations: i) condition; ii) concession; iii) opposition or contrast; iv) comparison; v) cause and effect, explanation or purpose; vi) chronological chaining or enumerations; vii) exemplification.

Most of the criteria refer to the presence of discourse markers that indicate the relations. Examples in (6) to (9) illustrate relations of *condition and explanation*, *opposition or contrast*, *cause and effect and exemplification*, respectively.

¹⁴ “By the way, how is Rio ? [...] Decades and nothing to add to Rio, you should change business.”

- (6) @jandira_feghali GENOCiDA! Esse ser é de uma maldade tão absurda, que é impressionante que consiga dormir. Sinceramente, **se ele realmente estiver doente (coisa que não acredito)**, não quero q a doença o mate. Q ele fique bem vivo p ser julgado e condenado pelos crimes q comete contra a humanidade
- [@jandira_feghali GENOCiDAL! This being is such an absurd evil, that it is impressive that he can sleep. Honestly, **if he is really sick (which I don't believe)**, I don't want the disease to kill him. That he stays truly alive to be judged and convicted of the crimes he commits against humanity]
- (7) @KimKataguirri E você prometeu em sua campanha trabalhar para o bem do país de uma maneira nova e diferente, mas a única coisa que tu está fazendo é ser igual aos que sempre estiveram ai, não está fazendo porra nenhuma para o futuro do Brasil. Vc é uma vergonha
- [@KimKataguirri And in your campaign you promised to work for the good of the country in a new and different way, but the only thing you are doing is being the same as those who have always been there, you are not doing anything for the future of Brazil. You are a shame]
- (8) @carlosjordy **Já que a esquerda é só paz e amor**, vamos pegar todas as postagens da esquerda e recriar ela mudando o nome do Bolsonaro para o do STF. Mas tire o print para caso precise apresentar provas.
- [@carlosjordy **Since the left is just peace and love**, let's take all the posts on the left and recreate it by changing the name of Bolsonaro instead of STF. But take the screenshot out in case you need to present proof.]
- (9) @MarceloFreixo Que medo hein... se a população de bem se armar, como vocês da esquerda poderiam impor as ideologias que tanto veneram né? Como Cuba, Venezuela, por exemplo, sem contar que atrapalha o "trabalho" das "vítimas da sociedade", q são mimados por vcs da esquerda. Vc é um Canalha!

[@MarceloFreixo What a fear huh ... if the population is well armed, how could you on the left-wing impose the ideologies that you venerate so much, right? Like Cuba, Venezuela, for example, not to mention that it hinders the “work” of the “victims of society”, who are spoiled by you guys from the left. You are a scoundrel!]

The majority of the relationships are explicitly shown in these three examples thanks to typical conjunctions and conjunctive phrases, but it's worth noting that these criteria were observed even when the discourse marker was not explicit and the relationship could be deduced from the semantics of the propositions. In (10) we illustrate an example of *opposition or contrast relation* between two ideas, but with no explicit mark.

- (10) @CarlaZambelli38 Infelizmente, o **meu pai foi obrigado a ir trabalhar, pegou COVID no trabalho e veio a falecer**. É triste quando pensam que **isso vale mais que a vida**. **Pra empresa é simples**, contratam outro, **pra família não tem como substituir vidas**.

[@CarlaZambelli38 Unfortunately, **my father was forced to go to work, he took COVID at work and died**. It is sad when they think that **this is worth more than life**. **For the company it is simple**, they hire another one, **for the family there is no way to replace lives**.]

Example in (10) also illustrates an enumeration relation of three actions in “foi obrigado a ir trabalhar, pegou COVID no trabalho e veio a falecer”¹⁵ and a comparison relation in “isso vale mais que a vida”.¹⁶ As this example shows, sentences frequently contain two or more of the arrangement relations. The same happens in (11), where we can see the use of chronological chaining, which constitutes a good strategy for organizing arguments.

- (11) @BolsonaroSP @danielPMERJ **O Deputado PRECISA ser solto, para que o processo jurídico penal seja cumprido desde o seu início**. A PGR já denunciou mesmo o STF tendo tomado a

¹⁵ “was forced to go to work, he got COVID at work and died.”

¹⁶ “this is worth more than life”

frente e já prendido. Agora precisa entrar com a parte da defesa e acontecer o mesmo que houve com o Lula, a ampla defesa.

[@BolsonaroSP @danielPMERJ **The Deputy MUST be released, so that the criminal legal process can be fulfilled from the beginning. The PGR has already denounced even the STF having taken the lead and already arrested. Now we need to go to the defense side and do the same thing that happened to Lula, the broad defense.**]

The *chronological chaining* can be observed in the excerpt “A PGR já denunciou [...] e já prendido. Agora precisa [...]”,¹⁷ since it establishes a temporal linkage concerning what was done in the past and what should be done in the future. Example (11) also illustrates a concession relation in the excerpt “mesmo o STF tendo tomado a frente”¹⁸ and a purpose relation in the excerpt “para que o processo jurídico penal seja cumprido desde o seu início”,¹⁹ that are marked by the discourse markers “mesmo” and “para que”, respectively.

Based on these seven criteria, it was defined that the arrangement of the argumentation is low when none of the criteria is present; medium when only one of the criteria is present; and high when two or more of the criteria are present.

3.3 Credibility

According to Wachsmuth *et al.* (2017b), an argument should be assessed as successful in creating credibility if it conveys arguments and other information in a way that makes the author credible, for example, indicating the honesty of the writer, the politeness of the language used or revealing the knowledge of author or experience in relation to the subjects discussed.

For the evaluation of Credibility, we considered that an argument written in Portuguese is credible if some criteria are present in the textual surface, since external criteria were not considered, such as suitability or engagement of the author in social media. Given the WG Twitter, it should be considered that the production of content is open to anyone

¹⁷ “The PGR already denounced [...] and already [arrested]. Now it needs to [...]”.

¹⁸ “even the STF having taken the lead”

¹⁹ “so that the criminal legal process can be fulfilled from the beginning”

who has an account. In this sense, since this social media platform allows anyone to talk about anything, the doubt regarding the credibility of the author of a tweet is inherent to the platform itself. Therefore, for an argument to be considered as highly credible, the text producer needs to use certain linguistic resources to prove that he or she is able to defend his/her opinion.

Thus, the credibility of an argument is positively affected when the author: (i) mentions *specific data or event*, regardless of the veracity judgment made about it; (ii) mentions a *media, historical or encyclopedic fact*, that is, something largely reported by the media, or something related to historical periods or is common sense; (iii) cites directly or indirectly a person who is considered an *authority figure* in the subject; (iv) uses a *hashtag (#) that reinforces a position*; (v) uses a *specialized term* from some area of knowledge; and/or (vi) makes a *personal or individual experience report*. All of these criteria can be identified in the following examples.

- (12) @gleisi @dilmabr tentou usar @petrobras para segurar inflação no país. **Segundo cálculos do Centro Brasileiro de Infraestrutura (CBIE)**, as perdas acumuladas pela Petrobras entre 2011 e 2014 (primeiro mandato de Dilma) por causa dessa política de preços superaram **R\$ 70 bilhões**.

[@gleisi @dilmabr tried to use @petrobras to insure inflation in the country. **According to calculations by the Brazilian Infrastructure Center (CBIE)**, the losses accumulated by Petrobras between 2011 and 2014 (Dilma's first term) because of this price policy exceeded **R \$ 70 billion**.]

- (13) @MarceloFreixo **Bolsnaro realiza mais uma caravana eleitoral visando 22. Junta gente, espalha o vírus e faz comício. Disse q não seria candidato à reeleição**, mas, desde que chegou ao poder, gasta todas suas energias fazendo campanha. Quer permanecer no cargo custe a quantidade de vidas q custar.

[@MarceloFreixo **Bolsnaro holds another electoral caravan aimed at 22. Gather people, spread the virus and make a rally. He said he would not be a candidate for re-election**, but, since he came to power, he has spent all his energies campaigning. Whether you want to stay in office costs the amount of lives it costs.]

- (14) @BolsonaroSP @jairbolsonaro imagino sua filha ficando inteligente e lendo todo este boicote de sua família a vida daqui alguns anos. vocês são os porta-vozes da morte. tudo que resta a sua família depois de tanto **negacionismo** é insistir, já que voltar atrás seria assumir um genocídio. **#forabolsonaro**
 [@BolsonaroSP @jairbolsonaro I imagine your daughter getting smart and reading all this boycott of your family life in a few years. you are the spokesmen for death. all that remains of your family after so much **negacionism** is to insist, since to go back would be to assume a genocide. **#forbolsonaro**]
- (15) @KimKataguiri **Hoje fui comprar 1kg de carne moída para o almoço e deu R\$ 43,00 achei um absurdo!** Em que mundo estamos com um custo tão alto de carne assim?! Mas de boa estamos comprando carne de primeira para nossos representantes políticos, então pq reclamar?! 😞😞
 [@KimKataguiri **Today I bought 1kg of ground beef for lunch and it was R \$ 43.00, I thought it was absurd!** In what world are we at such a high cost of meat ?! But thats ok, bc we are buying good meat for our political representatives, so why complain ?! 😞😞]

In (12), we verify two criteria: *specific data* when mentioning “R\$ 70 bilhões”²⁰ related to Petrobras losses; and *authority figure* represented by the “segundo cálculos do Centro Brasileiro de Infraestrutura (CBIE)”²¹,²¹ used as a source citation.

In (13), we identify the argument reinforced by a media fact, which was largely broadcasted by journals and news channels, that is “Bolsnaro realiza mais uma caravana eleitoral [...] Junta gente, espalha o vírus e faz comício. Disse q não seria candidato à reeleição”²².

In (14), we can identify other two criteria: *hashtag that reinforces a position* against the government (“#forabolsonaro”); and the *specialized*

²⁰ “R\$ 70 billion”

²¹ “according to Brazilian Infrastructure Center (CBIE)”

²² “Bolsonaro holds another electoral caravan [...] Gathers people, spreads the virus and rallies. He said he would not be a candidate for reelection”

term “negacionismo” (denialism), which is a term defined by science as the non-acceptance of proven scientific facts.

Finally, in (15), we observe a *personal and individual experience report* when the author says “Hoje fui comprar 1kg de carne moída para o almoço e deu R\$ 43,00 achei um absurdo!”²³

From these six criteria, it was defined that the credibility of the argument is low when none or only one of them are present; it is medium when two of the criteria is present; and high when three or more are present.

3.4 Emotional appeal

According to Wachsmuth *et al.* (2017b), an argument should be assessed as successful in creating an emotional appeal if it conveys arguments or other information in a way that creates emotions, which can make the target audience more open to the author’s arguments.

For this work purpose, we decided to adapt the original definition, since we observed, through an initial pilot study, that positive emotions improve the general quality of the argument, while negative emotions undermine the overall quality of the argument.

Again, it is important to consider the specific characteristics of WG Twitter, which presents posts on very controversial subjects in the domain of politics, such as fake news, vaccine against coronavirus, denial of science (denialism), personal attacks on politicians or their families, legality or illegality of judicial decisions, hate speech to the leftist political ideology, among others. These texts (tweets) tend to present several marks that negatively impact the emotional appeal and, consequently, reduce the overall quality of the argument, as different types of offense. Twitter, unlike other social media, does not have a very strict policy of restricting or filtering the content of posts or the abusive behavior of some users. Because of this, posts that contain bad words, cursing and even hate speech are very common.

Thus, for the evaluation of the Emotional appeal aspect, the criteria were grouped in: (i) positive, negative or neutral polarity of the tweet related to how this appeal affects the quality of the argument, and (ii) the intensity of this appeal, considering levels low, medium or high. The argument is low when none or only one of them are present; it is

²³ “Today I bought 1kg of ground beef for lunch and it was R\$ 43.00, I thought it was absurd!”

medium when two only one of the criteria is present; and high when three, two or more are present.

3.4.1 Polarity of Emotional appeal

The emotional appeal of a tweet has a negative impact on the quality of the argument when it contains: (i) *pejorative reference to a person or entity*; (ii) *curses or bad words*; (iii) *hate speech or threat*; or (iv) *expression that denotes speculation*. Example (16) is characteristic of negative polarity, since it presents all these criteria.

- (16) @MarceloFreixo Tá com medinho de armas por que, **CANALHA? Povo desarmado é mais fácil ver ser dominado né? Vocês da esquerda são uma desgraça. Tem de ser eliminados do planeta. Bando de vagabundos desocupados. Apareçam um dia na minha propriedade eu meto bala sem dó**
 [@MarceloFreixo Do you have fear of guns, **SCOUNDREL? Unarmed people are easier to see being dominated, right? You on the left-wing are a disgrace. It must be eliminated from the planet. Bunch of idle bum. Appear one day on my property, I'll shoot you without mercy]**

In (16), we verify: (i) a *pejorative reference* to left-wing by using the adjective “vagabundos desocupados”;²⁴ (ii) *cursing* when calling the deputy a “canalha (scoundrel)”; (iii) *hate speech* in “Vocês da esquerda são uma desgraça”²⁵ and death *threat* in “Apareçam um dia na minha propriedade eu meto bala sem dó”;²⁶ and (iv) *expression that denotes speculation*, when speculating that “[esquerdistas] tem que ser eliminados do planeta”.²⁷

On the other hand, the emotional appeal of a tweet increases the quality of the argument when it contains: (i) *cordial reference to a person or entity* (even when used in an ironic way); or (ii) *polished and polite language*, for example, by using modalizers (modal verbs, adverbs and other structures). Example (17) illustrates these two criteria.

²⁴ “idle bum”

²⁵ “You on the left-wing are a disgrace”

²⁶ “Appear one day at my property, I'll shoot you without mercy”

²⁷ “[left-wing defensors] must be eliminated from the planet”

(17) @lpbragancabr Verdade @lpbragancabr! e eles vieram de vários partidos, que até me surpreendeu. **Gostaria que o Sr.** Leve a eles a minha gratidão e parabéns, como eleitor e defensor da democracia do certo e do justo. Mas o sentimento que ficou, é que nós PERDEMOS A NOSSA DEMOCRACIA.

[@lpbragancabr It is truth @lpbragancabr! and they came from various parties, which even surprised me. **I would like you (Mr.)** to take my gratitude and congratulations to them, as a voter and defender of the democracy of the right and the fair. But the feeling that remains is that we LOST OUR DEMOCRACY.]

In (17), we identify: (i) *cordial reference* to the deputy who made the seed tweet through the treatment pronoun “Sr. (Mr.)”; and (ii) *polished and polite language* in the modalized construction “Gostaria que [...] (I would like [...])”.

There is also the possibility of neutral polarity, that is, when it is neither positive nor negative, as can be seen in the Example (18).

(18) @mariadorosario Bolsonaro nega a ciência, não investe na educação, áreas cruciais para salvar vidas. Não tem compromisso com o povo! Estamos juntos,²⁸ é #ForaBolsonaro

[@mariadorosario Bolsonaro denies science, does not invest in education, crucial areas to save lives. It has no commitment to the people! We are together, it's #ForaBolsonaro]

Neutral polarity is not marked by impartiality of opinion or positioning, but by the absence of positive or negative polarity marks, or else, even if these marks are present, they weigh equally and it is not possible to distinguish whether the emotional appeal used is more positive or more negative.

A tweet should be considered with negative Emotional appeal when it contains more criteria that weigh negatively on the overall quality of the argument than those that weigh positively. Similarly, the tweet should be considered to have a positive Emotional appeal when it contains more criteria that weigh positively for the overall quality of

²⁸ Note that this expression may impact emotional appeal, but not its polarity. We annotate this kind of expressions as slogans, which increases intensity of emotional appeal.

the argument than those that weigh negatively. The polarity of the tweet should be considered neutral when there is no criterion (positive or negative) characteristic of the polarity of the emotional appeal or when the number of positive and negative criteria is identical. But we did not identify this situation in the real data.

3.4.2. Intensity of Emotional appeal

In addition to the polarity, the intensity of Emotional appeal was also assessed, defined according to the presence of the following criteria: (i) *first person pronoun or verb inflection* (singular or plural); (ii) *repetition of punctuation marks* (??? or !!!); (iii) *emphatic structure*, such as whole word in capital letters, repetition of words or structures, italics, quotation marks; (iv) *imperative phrase or slogan*; (v) *expression that denotes exaggeration* (such as “always”, “never”, “everyone”) and superlatives; (vi) *feeling expressed by non-verbal language* (such as emoji, interjection or onomatopoeia); and (vii) *idiom, proverb or metaphor*. All of these criteria can be identified in (19) and (20). We emphasize that these characteristics are only intensifiers that affect the polarity (positive or negative) of the Emotional appeal.

- (19) @marcofeliciano **Confesso** q não esperava isso? Mas, mostrou-me que nesse congresso eleito, ainda tem muito q ser renovado. **UMA ARVORE PARA NASCER, PRECISA ANTES DE UMA SEMENTE PARA MORRER.** E essa foi o DANIEL, **acredite nisso! O BRASIL SE LEVANTARÁ DESSAS INJUSTIÇAS E CULPADOS SÓ IRÃO AUMENTANDO.**

[@marcofeliciano **I confess** I didn't expect this? But, he showed **me** that in this elected congress, there is still a lot to be renewed. **A TREE TO BE BORN, NEEDS BEFORE A SEED TO DIE.** And that was DANIEL, **believe that! BRAZIL WILL RISE FROM THESE INJUSTICES AND GUILTY WILL ONLY INCREASE.]**

- (20) @gleisi isso já não é mais mentiras, é a prova de que vcs não valem um grão de arroz , incompetentes , mentirosos , mal caráteres , e **BANDIDOS COM LETRA MAIÚSCULA**, muitas mortes misteriosas, apesar de não valerem mais nada, inacreditável ainda existirem !! 🤔

[@gleisi this is no longer lies, it is the proof that you are not worth a grain of rice, incompetent, liars, bad casings, and **BANDITOS WITH CAPITAL LETTERS**, many mysterious deaths, although they are not worth anything else, unbelievable still exist!! 🤔]

In (19), we identified the following criteria: (i) presence of a first singular person by the verb “confesso (confess)” and the pronoun “me (me)”; (iii) emphatic structure through several uppercase sections, expressing indignation or similar feeling; (iv) an imperative phrase when it says “acredite nisso!”;²⁹ and (vii) proverb or similar when using the sentence “uma árvore para nascer, precisa antes de uma semente para morrer”.³⁰

In (20), the following criteria are also present: (ii) repetition of the exclamation mark at the end of the tweet (“!!”); (iii) emphatic structure, also by means of uppercase letters; (v) expression that denotes exaggeration, when the author mentions “não valerem mais nada”;³¹ (vi) feeling expressed in non-verbal language, in this case, the emoji at the end of the tweet; and (vii) metaphorical expression in “não valem um grão de arroz”.³²

The intensity of a tweet’s Emotional appeal was defined as high when three or more criteria of negative polarity or two of positive polarity are present or when four or more intensity criteria are identified. A medium intensity was defined for cases in which there are two criteria of negative polarity or one of positive polarity or two or three criteria of intensity. Otherwise, the intensity of the tweet was classified as low.

4 Construction of the *corpus*

In this paper, the interest for messages related to politics, written by Brazilian congressmen, is anchored on the hypothesis that in this WG and domain there is a large number of argumentative texts generated both by politicians and by their followers. The congress members messages were picked for their argumentative potential, encouragement of contentious, provocative, and persuasive responses, and ability to spark debate on the issues discussed.³³ Besides Twitter being the social media

²⁹ “believe that!”

³⁰ “a tree to be born needs before a seed to die”

³¹ “they are no longer worth anything”

³² “they are not worth more than a grain of rice”

³³ In the following subsection, especially in Table 4, we present examples of these tweets.

most used by politicians, the choice of platform also took into account the flexibility to access data through API (Application Programming Interface)³⁴ specific for this purpose. Another reason why Twitter was chosen is related to the vast number of scripts, plugins and tools already developed for the collection, processing and analysis of tweets. It is worth mentioning that, in this research, only Twitter's public data were used, so it was not necessary to request any additional permission from the users.

According to the Lupa agency,³⁵ the volume of interactions between congressmen and their followers increased 42.3% in the first half of 2019. In this same study, active congressmen were divided into seven groups, based on their affiliation: PSL, On the left (PT, PCdoB and PSOL), Center-left (PDT, PSB), Center (MDB, PP, PL, PSD, SD, Podemos, PTB, PSC, PROS, PMN, Patriota, Avante, PHS, PRP, PRB), PSDB/DEM, Novo, and Other.

To compose the *corpus* of tweets used in this research, we produced a list with 417 congressmen who had a Twitter account and were active in the second half of 2020. The collection of messages was carried out through Tweepy,³⁶ a Python library for accessing Twitter's API. During 30 days (from 06th February to 07th March 2021) 3,243 messages posted on Twitter by congressmen and 452,287 replies from their followers were filtered from the 1,649,674 messages initially collected. In addition to the messages (tweets), the following information was also collected: number of followers the user has; number of people the user follows; profile description and URL; number of tweets and retweets the user had at the time of collection; and whether the account is verified by Twitter.³⁷

Although the congressmen's tweets were considered as seed posts for retrieving the replies of followers, it is worth pointing out that the assessment of the quality of the argumentation was performed only on the tweets of followers. To avoid confusion, the posts of congressmen are referred to as the seed post in this document.

³⁴ Available in: <https://help.twitter.com/pt/rules-and-policies/twitter-api>

³⁵ Available in: <https://piaui.folha.uol.com.br/lupa/2019/07/26/deputados-twitter-interacoes/>

³⁶ Available in: tweepy.org

³⁷ By means of a "blue seal", Twitter informs that a public interest account is authentic. Verified accounts must be notable (including heads of state and elected public officials) and active, with all profile fields filled out, have logged into the account within the last six months, with a confirmed email address or mobile number, and not have been blocked for 12 hours or 7 days for violating Twitter's rules in the last six months.

Following the same settings as Wachsmuth *et al.* (2017b) for the amount of messages to be assessed, seed posts and annotators, from the total of 3,243 seed posts collected, 80 were randomly selected and distributed equally across the affiliation groups (listed in Table 2). Twelve seed posts per group were then selected, since the Left-Center and Others groups did not obtain a significant amount of tweets (replies from followers) to compose the *corpus*. For each of the 80 seed posts, we obtained the first five tweets (in chronological order) that satisfied the following restrictions, which were manually verified: having at least 200 characters and not being spam messages or messages with repeated characters. Table 2 shows the data collected.

TABLE 2 – Distribution of the number of tweets, by affiliation group, to build the *corpus*

AFFILIATION GROUP	SEED POSTS	TWEETS	PERCENTAGE
Center (MDB, PP, PL, PSD, SD, Podemos, PTB, PSC, PROS, PMN, Patriota, Avante, PHS, PRP, PRB)	16	80	20%
PSDB/DEM	16	80	20%
Left-wing (PT, PCdoB e PSOL)	16	80	20%
Novo	16	80	20%
PSL	16	80	20%
Center-left (PDT, PSB)	0	0	0%
Others	0	0	0%
TOTAL	80	400	100%

The resulting *corpus* has the statistics shown in Table 3.

TABLE 3 – Statistics of the *corpus* composed of 400tweets

Tokens	20,000
Types	4,620
Words repetition %	76.90%
Sentences	1,643
Unique sentences	1,590
Sentences repetition %	3.23%
Characters	97,971

The guidelines for the annotation were based on the directives related to the rhetorical dimension from the work of Wachsmuth *et al.* (2017b)³⁸ and are available on the project page, along with the annotated *corpus*.³⁹

5 Annotation of the *corpus*

After the creation of the guidelines, defined collaboratively by the four annotators, the annotation of the *corpus* was performed separately by each one of them, for the same set of 400 posts, over the period of 30 days (from March 08 to April 08).

The four annotators annotated the same tweets presented in blocks of 100 instances. After the annotation of each block of 100 tweets, meetings lasting about 1 hour each were held to discuss specific points of disagreement, but without modifying any annotation performed in the tweets. The final set of annotation guidelines is available at <https://argq.org/>.

The annotation process consisted of three steps. In the first, each annotator classified whether or not the post was related to the topic/subject of the seed post. The annotation options for this were: related, partially related, or not related. In Table 4 we present three tweets assessed by the four annotators as, respectively: not related, completely related or partially related to the subject of the seed post.

TABLE 4 – Examples of tweets related, partially related and unrelated to the initial post subject

Initial seed post	Tweet being assessed	Is it related to the subject?
Lamento que parlamentares que dizem defender o povo atrasem o trabalho de comissões fundamentais, como a Comissão de Ética, por exemplo. Há deputados enrolados com a justiça! O PSOL, em especial, precisa parar de atrasar o país. E a Câmara precisa andar para que o país avance! https://t.co/VmlpJVTv6M	@marcelvanhattem Inadmissível o tratamento que a imprensa brasileira vem recebendo nos dias atuais dos políticos. Impedir o trabalho de jornalistas é atacar o nosso direito como cidadão de ser informado. O deputado @marcelvanhattem vai fazer algo para impedir a remoção da imprensa de sua sala?	no

³⁸ Available at <http://argumentation.bplaced.net/arguana/data>.

³⁹ Available at <https://argq.org/>.

<p>It is a shame that parliamentarians who claim to defend people delay the work of fundamental commissions, such as the Ethics Committee, for example. There are deputies involved with justice! PSOL, in particular, needs to stop delaying the country. And the Chamber needs to move for the country to move forward! https://t.co/VmlpJVTv6M</p>	<p>The treatment that the Brazilian press has been receiving in the current days of politicians is unacceptable. To prevent the work of journalists is to attack our right as a citizen to be informed. Will Congressman @marcelvanhattem do anything to prevent the removal of the press from his office?</p>	
<p>Lamento que parlamentares que dizem defender o povo atrasem o trabalho de comissões fundamentais, como a Comissão de Ética, por exemplo. Há deputados enrolados com a justiça! O PSOL, em especial, precisa parar de atrasar o país. E a Câmara precisa andar para que o país avance! https://t.co/VmlpJVTv6M</p>	<p>@marcelvanhattem 🍌🍌🍌🍌 diga se de passagem esse psol ,só atrasa o país ,a sua bancada é cega, julgam de acordo com autoria dos pl ,se for do marcel ou da Bia kicis por ex,já são contra,sem ler o texto do pl !Isso é atraso moral e atraso nos avanços para o país, e resume em atraso para eles tb 🤔</p> <p>@marcelvanhattem 🍌🍌🍌🍌 by the way this psol, only slows down the country, its bench is blind, they judge according to the authorship of the project, if it is from marcel or Bia kicis for example, they are already against it, without reading the text of the project! This is moral delay and delay in advances for the country, and summarizes in delay for them as well 🤔</p>	<p>yes</p>
<p>Bolsonaro considera a parte pelo todo. Acha que seu mundo extremo representa o país. O povo não está vibrando. O povo não quer armas. A população anseia pelas vacinas.</p> <p>Bolsonaro considers the part for the whole. He thinks that his extreme world represents the country. The people are not vibrating. The people do not want weapons. The population yearns for vaccines.</p>	<p>@RodrigoMaia Você foi um fiador desse governo. Toma vergonha na sua cara. Você fez parte desse governo e foi condescendente com este criminoso. Você aprovou uma reforma da previdência prejudicando os mais pobres e dando aumento salarial aos militares.Cúmplice! Na pandemia n fez nada. Hipócrita</p> <p>@RodrigoMaia You were a guarantor of this government. Shame on you. You were part of that government and condescended to this criminal. You passed a pension reform harming the poorest and giving the military a salary increase. Accomplice! In the pandemic you did nothing. Hypocritical</p>	<p>partially</p>

In the first example, the tweet being assessed is not related to the seed post because the topic in seed tweet is the fact that some parliamentarians delay or prevent votes in the Chamber, specially parliamentarians from PSOL (a political party), while the topic in reply tweet is the way politicians treat the Brazilian press. In the second example, both posts are completely related to each other because the topic in the seed tweet is the same as in the first example while the tweet being assessed also talks about PSOL and the way their parliamentarians delay and prevent votes, by mentioning some examples. In the third example, all annotators assessed as partially related to the subject because, in the seed tweet, the author criticizes the president for prioritizing certain issues instead of vaccine, while the reply tweet criticizes the deputy author of the seed post, arguing that this deputy supported the president during his electoral campaign and, therefore, is colluding with the actions of the president. So, the third example does not address the main theme, but just a part of it.

In the second step, the tweet was assessed in terms of argumentativeness, marking “yes” for argumentative tweets and “no” for non-argumentative tweets. In this work, a broad definition of argumentativeness was considered, in order to include a larger number of tweets in the *corpus*. In this sense, the tweets in which it was possible to identify the position/opinion (either favorable or unfavorable) of the author were considered argumentative, i.e., containing any attempt to mark the opinion, even without supporting evidence for it. This decision to extend the concept of argumentativity to include opinionative texts, even if they do not present clear arguments, is due to the characteristics of the WG, since most of the tweets bring some position or make a criticism without, however, presenting arguments to support this position. In Table 5 we present two tweets evaluated as argumentative and non-argumentative, respectively, by the four annotators.

TABLE 5 – Examples of argumentative and non-argumentative tweets

Tweet	Argumentativity
<p>@MarceloFreixo Você votou? Provavelmente votou NÃO. Então a pergunta é: você está “tistinho” porque perdeu? Se a autonomia não fosse aprovada você estaria aqui se manifestando contra? Ou estaria exaltando os deputados que entenderam que o BC precisa ter um freio? Totalmente sem noção!</p> <p>@MarceloFreixo Did you vote? You probably voted NO. So the question is: are you “saddy” because you lost? If the autonomy was not approved, would you be here speaking out against it? Or would it be exalting the deputies who understood that the BC needs to have a brake? Totally out of it!</p>	Argumentative
<p>@KimKataguiiri Pergunte ao bolsonaro quando é que o G.F. vai transferir o dinheiro dos salarios dos servidores na missão do Brasil em Portugal. Este mês ainda não receberam o salário e não foram pagos os alugueis das casas dos embaixadores.</p> <p>@KimKataguiiri Ask bolsonaro when government will transfer the money from the servers’ salaries to the Brazilian mission in Portugal. This month they still have not received their salary and the rent of the ambassadors’ houses has not been paid.</p>	Non-argumentative

The first example was considered argumentative since the position of author regarding the seed post is clearly expressed. On the other hand, the second tweet does not state the position of the author, but only brings some information about an unrelated subject.

For the tweets assessed as non-argumentative, the annotation process ended in this step. For the remaining tweets, whether or not related to the subject of the seed post, the other aspects and their criteria were evaluated as described in Section 3. In Figure 6 we bring a print screen of the annotation sheet⁴⁰ used by the human judges.

⁴⁰ We tried some *corpus* annotation tools to perform the annotation, but at the end we decided to use a simple spreadsheet since it was easy to use, easy to change any annotation at any time, easy to compare different replies of the same seed tweet, and also we can search all instances of a linguistic pattern and systematically review all annotations for a specific criterion by using filters, among other advantages.

FIGURE 6 – Print screen of the annotation sheet

SEED POSTS	TWEETS FOR EVALUATION	Is the tweet related to the seed?		Is the tweet argumentative?		EMOTIONAL APPEAL						OVERALL QUALITY								
		Yes	No	Yes	No	POLARITY			INTENSITY											
Lockdown só é eficaz para aglomerar. Ainda não aprenderam com a redução de horário do comércio. O ser político é um bicho que age sob pressão e na ansia de dizer que está fazendo algo acaba errando - isto é, partindo do princípio que é bem intencionado.	<p>①BolsonaroSP Cade a dengue, cade a gripe comum, a H1N1, cade as demais doenças respiratórias, tão comuns em nosso país nessa época do ano? Sumiram ou estão se somando aos números do COVID? Tudo pela ciência e pela saúde! Principalmente agora que governadores e prefeitos podem comprar vacina.</p> <p>②BolsonaroSP Um político bem intencionado não se posiciona contra TODAS as medidas eficazes contra o Covid enquanto temos mais de 250k mortos. Vocês usam cadáveres como palanque para falar as idiotices que nos tornam motivo de piada do mundo. QUE GOVERNO DE MERDAS.</p>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Low	Medium
Lockdown só é eficaz para aglomerar. Ainda não aprenderam com a redução de horário do comércio. O ser político é um bicho que age sob pressão e na ansia de dizer que está fazendo algo acaba errando - isto é, partindo do princípio que é bem intencionado.	<p>③BolsonaroSP Tem que ter lockdown porque pessoas como você não fazem uso das medidas preventivas de contágio. Outra coisa que pode acabar com o isolamento é a vacina, mas o governo do seu pai se recusou a ir atrás...</p> <p>④BolsonaroSP Parece que ficou estabelecido que quanto mais matar a população, quanto maior a demonstração de tirania, melhor os resultados do covid. Mas o que vimos é que os governadores que mais restringiram foram aqueles que causaram o maior número de mortes e mais desemprego.</p>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Low	Medium
Lockdown só é eficaz para aglomerar. Ainda não aprenderam com a redução de horário do comércio. O ser político é um bicho que age sob pressão e na ansia de dizer que está fazendo algo acaba errando - isto é, partindo do princípio que é bem intencionado.	<p>⑤BolsonaroSP Parece que ficou estabelecido que quanto mais matar a população, quanto maior a demonstração de tirania, melhor os resultados do covid. Mas o que vimos é que os governadores que mais restringiram foram aqueles que causaram o maior número de mortes e mais desemprego.</p>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Low	Medium
		AUTOMATIC		AUTOMATIC		AUTOMATIC			AUTOMATIC			AUTOMATIC			AUTOMATIC			AUTOMATIC		

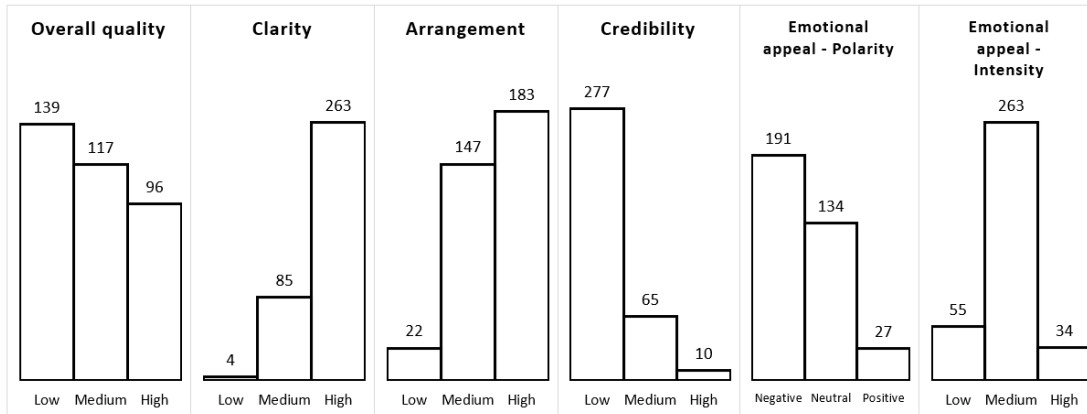
After assessing each criterion individually following the guidelines described in section 3, the final score of each aspect and the final score for the Overall quality were calculated automatically. To do so, we converted the low, medium and high scores to 1, 2 and 3, respectively, and the polarity of the Emotional appeal to 0 (neutral), 1 (positive) or -1 (negative). The final score for the Emotional appeal was calculated as the product of its polarity and intensity for non-neutral tweets (polarity x intensity) and as half of the intensity for neutral ones (intensity/2). Finally, we summed the final scores for all aspects and assessed the Overall quality as low if the sum was less or equal to 4; high if the sum was greater or equal to 8; and medium otherwise.

Considering the specificities of the WG and the political domain, the annotators report that it was essential to select tweets that were recent at the moment of the annotation. This was because the subjects and people cited were in the media spotlight, which allowed the recognition and identification of the entities and facts mentioned in the discourse at the time of annotation.

5.1 Annotation statistics

As already mentioned, the annotation of the tweets was carried out by four human judges, each annotating all 400 replies of 80 initial seed posts. Three ponderation levels were employed for the aspects Clarity, Arrangement, Credibility and Emotional appeal, related to the rhetorical dimension of the taxonomy proposed by Wachsmuth *et al.* (2017b): i) High/positive; ii) Medium/neutral; and iii) Low/negative. From these 400 tweets, 352 were assessed as argumentative by all the four human judges. In Figure 7 we display the score distribution for each aspect for the 352 argumentative tweets considering as final/gold annotation the majority score. In case of tie (e.g., 2 high and 2 medium) the smallest score was considered and in case of total disagreement (e.g., 1 low, 1 medium and 2 high), the medium score was selected. For the Overall quality we considered the average score of the four human judges.

FIGURE 7 – Score distributions by quality aspects in rhetorical dimension



As we can see from Figure 7, around 40% of the tweets were assessed as low overall quality, 33% as medium overall quality and 27% were assessed as high overall quality. Most of them have high Clarity and Arrangement, but low Credibility. In fact, only 3% of them were assessed as high credibility. Regarding Emotional appeal, we confirmed our hypothesis of the strong negative emotional appeal of this WG with 54% of the argumentative tweets being assessed as negative for overall quality.

In Table 6 (a) we present the final scores of each aspect for the 352 posts (88%) assessed as argumentative by the annotators. To test the clarity of the annotation guideline and the suitability of the taxonomy for the intended task, inter-annotator agreement was calculated, a process in which annotators mark the same fraction of the *corpus*, and the annotations are compared in terms of equal markings among all or most annotators. In Table 6 (b) we show the range of Krippendorff's (2011) α (lowest value - highest value) of the least concordant and most concordant trios of annotators, and the total and majority agreements. Total agreement is achieved when all annotators agree on the same score, and majority indicates that at least three annotators agreed. Total agreement is noted to be between 27.84% and 57.67%, and majority agreement of the annotators between 69.89% and 86.93%. It is noted, as pointed out by Wachsmuth *et al.* (2017b), that the rhetorical dimension shows evidence of subjectivity in its evaluation.

Regarding the α values, we chose to report the agreement among trios to be able to compare our results with those from Wachsmuth *et al.* (2017b) since in their work there were three annotators. Except for the

Clarity, all aspects have maximum agreement values above 0.40, different from Wachsmuth *et al.* (2017b) (see Table 1) where agreement α values for all aspects were below 0.40. For the overall quality our values vary between 0.50 and 0.54, a similar or even better result than Wachsmuth *et al.* (2017b), which obtained an agreement α value of 0.51. Thus, according to our agreement results we can conclude that there are indications that the criteria proposed in this work adequately guide the assessment of the argumentative quality on Twitter political domain.

TABLE 6 – Assessment results

Quality Aspect	(a) Final Score			(b) Agreement		
	Low/ Negative	Medium/ Neuter	High/ Positive	α trios	total (4/4)	majority (3-4/4)
Clarity	4	85	263	0.26 - 0.30	48.58%	79.26%
Arrangement	22	147	183	0.51 - 0.71	50.57%	82.67%
Credibility	277	65	10	0.36 - 0.48	57.67%	86.93%
Emotional appeal – Polarity	191	134	27	0.60 - 0.66	51.99%	82.67%
Emotional appeal – Intensity	55	263	34	0.48 - 0.55	40.63%	82.39%
Overall quality	139	117	96	0.50 - 0.54	27.84%	69.89%

In terms of “total” and “majority” agreement scores, a direct comparison with the numbers from Wachsmuth *et al.* (2017b) is impossible because our values were derived using four annotators, whereas their annotation was done with only three human judges. The greater the number of annotators, the more difficult it is to achieve full (or majority) agreement between them.

5.2 Analysis of the (dis)agreement in the overall quality of the argumentation

As presented in the previous subsection, agreement among the group of annotators for the aspects ranged from 79.26% to 86.93% (Table 6). Specifically on the General quality of argumentation, there was 69.89% agreement. It is worth pointing out that the calculation of the agreement among the annotators is one of the important steps in the *corpus* building, since it gives credibility to the linguistic resource elaborated.

It should be noted that, in studies of linguistic phenomena at more concrete levels of analysis (such as phonetics and morphology, for example), the agreement tends to be high; on the other hand, at less concrete levels of analysis (such as semantic and discourse/textual), the agreement tends to be lower, since phenomena at these levels may leave few linguistic clues on the surface of the text. Besides the complexity of the level of the linguistic analysis itself, depending on the level of analysis, human subjectivity may be intrinsic to the annotation task, since the annotator may rely on extra-textual elements and information to assess a rhetorical aspect of the tweet (how a given information was or was not conveyed by the media, ensuring the credibility of the post, for example).

For that, in this task, as shown above, some steps were indispensable, such as the construction of an annotation guidelines manual, annotation of an initial set, initial agreement check, review and adaptation of the guidelines manual, and frequent meetings for alignment of conceptions among the annotators. According to Hovy and Lavid (2010), these are irreplaceable methodological steps in the *corpus* annotation process.

In this sense, we bring a deep analysis of some cases of (dis) agreement with respect to the Overall quality of argumentation, considering (i) the linguistic phenomena that emerge from argumentation, (ii) the level of linguistic analysis (in this case, discourse-textual) and (iii) the human subjectivity employed in the task.

The total agreement generally occurs in posts whose content presents very low or very high quality of argumentation, as in (21) and (22), respectively.

- (21) @gleisi Mas também deputada, com essa oposição que tudo que o governo federal faz vocês acham que está errado. Imagine se o povo estivesse todos seguindo o FIQUE EM CASA, A ECONOMIA A GENTE VER DEPOIS. Sou a favor que sejam seguidos os protocolos: máscara, lavar as mãos e não aglomerar. [*@gleisi But also a deputy, with this opposition that everything the federal government does you think is wrong. Imagine if the people were all following the STAY AT HOME, THE ECONOMY FOR PEOPLE TO SEE LATER. I am in favor of following the protocols: mask, washing hands and not agglomerating.*]

(22) @CarlaZambelli38 Infelizmente, o meu pai foi obrigado a ir trabalhar, pegou COVID no trabalho e veio a falecer. É triste quando pensam que isso vale mais que a vida. Pra empresa é simples, contratam outro, pra família não tem como substituir vidas.

[@CarlaZambelli38 Unfortunately, my father was forced to go to work, he took COVID at work and died. It is sad when they think that it is worth more than life. For the company it is simple, they hire another one, for the family there is no way to replace lives.]

The tweet in (21) was considered of low argumentative quality since its author (i) presents criteria that harm Clarity (such as grammatical deviations and deviation from the main subject), (ii) builds a conditional relation that contributes to the Arrangement of the text, (iii) does not use any criteria to increase the Credibility of the discussed issue and (iv) uses resources that result in negative polarity and medium intensity of Emotional appeal. The tweet in (21), in turn, was assessed as of high argumentative quality since it (i) is a personal experience report (which improves Credibility), (ii) is organized in order to emphasize a contrast relation between ideas and logical sequence, (iii) besides highlighting the arguments in a moderate way, without using Emotional appeal devices that penalize the argumentative quality.

The cases in which there was more disagreement among the annotators were those whose tweets have argumentative quality that could be classified as medium and, therefore, have traces of a low or high quality, as shown in (23) and (24).

(23) @gleisi Nobre deputada me responda uma coisa, pq não dá o exemplo e começa a cortar na própria carne, abrindo mão de todos os privilégios que tem ficando somente com o salário? Com isso seus pares fariam o mesmo, aí sim o que vc disser terá algum sentido, fora isso pura hipocrisia

[@gleisi Noble deputy answer me one thing, why don't you set an example and start cutting into your own flesh, giving up all the privileges you have left with only your salary? With that your peers would do the same, then what you say will make some sense, out of that pure hypocrisy]

(24) @CarlaZambelli38 Era só ele ter controlado algumas falas, que convenhamos, foram desnecessárias. Um conservador que se preze, governa pelo exemplo. Vide Ronald Regan, Abraham Lincoln e Margareth Thatcher. Alguns comentários sobre a pandemia foram desnecessários.

[@ CarlaZambelli38 It was just that he controlled some lines, which we agree, were unnecessary. A self-respecting conservative rules by example. See Ronald Regan, Abraham Lincoln and Margareth Thatcher. Some comments on the pandemic were unnecessary.]

In (23), the author uses resources that (i) harm the Clarity of the argument (such as language mistakes and deviation from the main subject), (ii) contribute to a good arrangement (such as the construction of cause-effect and conditional semantic relations), (iii) does not use any resource to increase Credibility and (iv) resulting in neutral polarity and medium intensity for Emotional appeal. In (24), on the other hand, the text in which Arrangement and Credibility is average, for presenting only one criterion in each aspect that favors these aspects and, on the other hand, Clarity is high for not having any criterion that would harm it, and neutral polarity and low intensity for Emotional appeal. Given this, it is noted that the Quality of argumentation in (23) and (24) can be assessed as medium, despite having criteria that could classify them as low and high, respectively, according to the annotators.

Thus, it is worth noting that the agreement, in general, is higher in relation to aspects of a more objective nature, as they evidence linguistic clues that emerge on the textual surface (such as Clarity and Arrangement) and, sometimes, lower in aspects of a subjective nature (in this case, Credibility and Emotional appeal).

6 Final considerations and future directions

In this paper, the process of annotation of a *corpus* composed of 400 political tweets in the Brazilian context was described. The taxonomy proposed by Wachsmuth *et al.* (2017b) was adapted for the WG tweets and the domain of politics. The results of this annotation process, as well as the inter-annotator agreement calculations are comparable to the results obtained by Wachsmuth *et al.* (2017b) in a similar experiment for

the English language. As a result of this work, an annotated *corpus* with information about the general quality of argumentation and the quality of specific argumentation-related aspects have been constructed and are available on the project webpage.

The task of revising and adapting the taxonomy of Wachsmuth *et al.* (2017b) has led the work to certain limitations, some of them theoretical and others practical. The main theoretical limitation is related to the adoption of a definition of argumentativeness that is very different from the traditional conceptualization of what is argumentative or not. This decision may cause some discrediting or disagreement with the work by the linguistic community, since it is based on the notion of argumentativity itself.

Conventionally, a text is considered argumentative if it presents arguments, organized and structured in a logical sequence. For the purposes of this annotation, this concept was adapted to cover any and all tweets in which it was possible to identify the author's position/opinion. Thus, any attempt to express an opinion, even if it is not supported by evidence, should be considered argumentative. In other words, even if the argumentation was bad, even if there were few arguments, or if it did not convince the interlocutor, the post was still evaluated as argumentative.

We also point out some practical limitations to this work. According to Lacy *et al.* (2015), it is recommended that at least one of the annotators does not be part of producing and refining the annotation guidelines, but we did not find any other available annotator to perform the task after we finished the guidelines, so we were unable to meet this requirement. In future work, we plan to invite other external annotators to perform the same annotation and see how different the agreement among annotators who did not participate in the guideline drafting process is in comparison to the group of annotators who did both guideline drafting and annotation. This comparison may lead us to validate the annotation guidelines for future tasks.

Another limitation to consider is that we recognize that the human annotation may contain some bias in the political ideology of the annotators, but the guidelines were made in the most objective way possible so that this bias would not interfere in the criteria identification and in the aspect evaluation.

Finally, the *corpus* annotated in this study will be used for training computational models, by applying NLP and machine learning techniques

and tools/resources. As a final goal of this research, it is expected that the automation of the process of evaluating the quality of argumentation on Twitter, in the domain of politics, will be applied to filtering low-quality messages and generating a ranking of the best qualified posts.

Contribution of each author to the manuscript

The paper “Quality of argumentation in political tweets: what is and how to measure it” stems from the original project Arg Q! (Evaluation of quality of argumentation) developed by the first author and supervised by the last author and Vânia Paula de Almeida Neris. First and last author built the *corpus* and participated in the writing of the annotation guidelines. Annotation guidelines, theoretical discussions and *corpus* annotation were done by second to fifth authors. Last author also annotated the tweets. The text was written and revised by all authors.

References

- BENCH-CAPON, T. J.; DUNNE, P. E. Argumentation in Artificial Intelligence. *Artificial Intelligence*, [S.l.], v. 171, n. 10-15, p. 619-641, 2007. DOI: <https://doi.org/10.1016/j.artint.2007.05.001>
- BLAIR, J. A. Rhetoric, Dialectic, and Logic as Related to Argument. *Philosophy & Rhetoric*, University Park, PA, v. 45, n. 2, p. 148-164, 2012. DOI: <https://doi.org/10.5325/philrhet.45.2.0148>. Available in: <http://www.jstor.org/stable/10.5325/philrhet.45.2.0148>. Access on: May 20, 2021.
- BOUDRY, M.; PAGLIERI, F.; PIGLIUCCI, M. The Fake, the Flimsy, and the Fallacious: Demarcating Arguments in Real Life. *Argumentation*, [S.l.], v. 29, n. 4, p. 431-456, 2015. DOI: <http://doi.org/10.1007/s10503-015-9359-1>
- BRADY, W. J.; WILLS, J. A.; JOST, J. T.; TUCKER, J. A.; VAN BAVEL, J. J. Emotion Shapes the Diffusion of Moralized Content in Social Networks. *Proceedings of the National Academy of Sciences*, [S.l.], v. 114, n. 28, p. 7313-7318, 2017. DOI: <https://doi.org/10.1073/pnas.1618923114>

CARLILE, W.; GURRAPADI, N.; KE, Z.; NG, V. Give Me More Feedback: Annotating Argument Persuasiveness and Related Attributes in Student Essays. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 56., 2018, Melbourne. *Proceedings* [...]. Melbourne: Association for Computational Linguistics, 2018. p. 621-631. DOI: <https://doi.org/10.18653/v1/P18-1058>. Available in: <https://www.aclweb.org/anthology/P18-1058>. Access on: May. 20, 2021.

EEMEREN, F. H. V.; GROOTENDORST, R. Fallacies in Pragma-Dialectical Perspective. *Argumentation*, [S.l.], v. 1, n. 3, p. 283-301, 1987. DOI: <http://doi.org/10.1007/BF00136779>.

EEMEREN, F. H. V.; GROOTENDORST, R. *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. Cambridge: Cambridge University Press, 2003. DOI: <https://doi.org/10.1017/CBO9780511616389>

FREEDOM HOUSE. *Freedom on the Net 2019*. 2019. Available in: https://freedomhouse.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf. Access on: May 20, 2021.

FREITAS, E. C.; BARTH, P. A. Gênero ou suporte? O entrelaçamento de gêneros no Twitter. *Revista (Con)Textos Linguísticos*, Vitória, v. 9, n. 12, p. 8-26, 2015.

GARCÍA-GORROSTIETA, J. M.; LÓPEZ-LÓPEZ, A. Identifying Argumentative Paragraphs: Towards Automatic Assessment of Argumentation in Theses. In: SILBERZTEIN, M.; ATIGUI, F.; KORNYSHOVA, E.; MÉTAIS, E.; MEZIANE, F. (ed.). *Natural Language Processing and Information Systems*. Cham: Springer International Publishing, 2018. p. 83-90. DOI: https://doi.org/10.1007/978-3-319-91947-8_9

GARCÍA-GORROSTIETA, J. M.; LÓPEZ-LÓPEZ, A.; GONZÁLEZ-LÓPEZ, S. Automatic Argument Assessment of Final Project Reports of Computer Engineering Students. *Computer Applications in Engineering Education*, [S.l.], v. 26, n. 5, p. 1217-1226, 2018. DOI: <https://doi.org/10.1002/cae.21996>

GLEIZE, M.; SHNARCH, E.; CHOSHEN, L.; DANKIN, L.; MOSHKOWICH, G.; AHARO-NOV, R.; SLONIM, N. Are You Convinced? Choosing the More Convincing Evidence with a Siamese

Network. *In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, 57., 2019, Florence. *Proceedings* [...]. Florence: Association for Computational Linguistics, 2019. p. 967-976. DOI: <https://doi.org/10.18653/v1/P19-1093>. Available in: <https://www.aclweb.org/anthology/P19-1093>. Access on: May. 20, 2021.

HABERNAL, I.; GUREVYCH, I. Which Argument Is More Convincing? Analyzing and Predicting Convincingness of Web Arguments Using Bidirectional LSTM. *In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, 54., 2016, Berlin. *Proceedings* [...]. Berlin: Association for Computational Linguistics, 2016. p. 1589-1599. DOI: <https://doi.org/10.18653/v1/P16-1150>. Available in: <https://www.aclweb.org/anthology/P16-1150>. Access on: May. 20, 2021.

HABERNAL, I.; GUREVYCH, I. Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics*, Singapore, v. 43, n. 1, p. 125-179, 2017. DOI: https://doi.org/10.1162/COLI_a_00276. Available in: <https://www.aclweb.org/anthology/J17-1004>. Access on: May. 20, 2021.

HOVY, E.; LAVID, J. Towards a 'Science' of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics. *International Journal of Translation*, [S.l.], v. 22, n. 1, p. 13-36, 2010.

KRIPPENDORFF, K. *Computing Krippendorff's Alpha-Reliability*. 2011. Available in: http://repository.upenn.edu/asc_papers/43. Access on: May. 20, 2021.

LACY, S.; WATSON, B. R.; RIFFE, D.; LOVEJOY, J. Issues and Best Practices in Content Analysis. *Journalism & Mass Communication Quarterly*, [S.l.], v. 92, n. 4, p. 791-811, 2015. DOI: <https://doi.org/10.1177/1077699015607338>.

LAUSCHER, A.; NG, L.; NAPOLES, C.; TETREAULT, J. Rhetoric, Logic, and Dialectic: Advancing Theory-Based Argument Quality Assessment in Natural Language Processing. *In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS*, 28., 2020, Barcelona, *Proceedings* [...]. Barcelona: International Committee on Computational Linguistics, 2020. p. 4563-4574. DOI: <https://doi.org/10.18653/v1/2020.coling-main.402>

LYTOS, A.; LAGKAS, T.; SARIGIANNIDIS, P.; BONTCHEVA, K. The Evolution of Argumentation Mining: From Models to Social Media and Emerging Tools: Information. *Processing & Management*, [S.l.], v. 56, n. 6, p. 1-22, 2019. DOI: <https://doi.org/10.1016/j.ipm.2019.102055>.

MARCUSCHI, L. A. Gêneros textuais: definição e funcionalidade. In: DIONISIO, A. P. *et al. Gêneros textuais e ensino*. 2. ed. Rio de Janeiro: Lucerna, 2002. p. 19-36.

PERSING, I.; NG, V. Modeling Argument Strength in Student Essays. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 53.; INTERNATIONAL JOINT CONFERENCE ON NATURAL LANGUAGE PROCESSING, 7., 2015, Beijing. *Proceedings [...]*. Beijing: Association for Computational Linguistics, 2015. p. 543-552. DOI: <https://doi.org/10.3115/v1/P15-1053>. Available in: <https://www.aclweb.org/anthology/P15-1053>. Access on: May. 20, 2021.

POTTHAST, M.; GIENAPP, L.; EUCHNER, F.; HEILENKÖTTER, N.; WEIDMANN, N.; WACHSMUTH, H.; STEIN, B.; HAGEN, M. Argument Search: Assessing Argument Relevance. In: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 42., 2019, New York. *Proceedings [...]*. New York: Association for Computing Machinery, 2019. p. 1117-1120. DOI: <https://doi.org/10.1145/3331184.3331327>

ROSENFELD, A.; KRAUS, S. *Providing Arguments in Discussions Based on the Prediction of Human Argumentative Behavior*. 2015. Available in: <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9522>. Access on: May. 20, 2021.

ROSSINI, P. Disentangling Uncivil and Intolerant Discourse. In: BOATRIGHT, R.; SOBIERAJ, S.; SHAFFER, T.; YOUNG, D. (ed.). *A Crisis of Civility? Contemporary Research on Civility, Incivility, and Political Discourse*. New York: Routledge, 2019. p. 142-157. DOI: <https://doi.org/10.4324/9781351051989-9>

ROSSINI, P. Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research*, [S.l.], ahead of print, p. 1-27, 2020. DOI: <https://doi.org/10.1177/0093650220921314>

STAB, C.; GUREVYCH, I. Parsing Argumentation Structures in Persuasive Essays. *Computational Linguistics*, Singapore, v. 43, n. 3, p. 619–659, 2017a. DOI: https://doi.org/10.1162/COLI_a_00295. Available in: <https://www.aclweb.org/anthology/J17-3005>. Access on: May. 20, 2021.

STAB, C.; GUREVYCH, I. Recognizing insufficiently supported arguments in argumentative essays. In: CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 15., 2017b, Valencia. *Proceedings* [...]. Valencia: Association for Computational Linguistics, 2017b. p. 980-990. DOI: <https://doi.org/10.18653/v1/E17-1092>. Available in: <https://www.aclweb.org/anthology/E17-1092>. Access on: May. 20, 2021.

SWANSON, R.; ECKER, B.; WALKER, M. Argument mining: Extracting arguments from online dialogue. In: ANNUAL MEETING OF THE SPECIAL INTEREST GROUP ON DISCOURSE AND DIALOGUE, 16., 2015, Prague. *Proceedings* [...]. Prague: Association for Computational Linguistics, 2015. p. 217-226. DOI: <https://doi.org/10.18653/v1/W15-4631>. Available in: <https://www.aclweb.org/anthology/W15-4631>. Access on: May. 20, 2021.

TOULMIN, S. E. *The Uses of Argument*. Cambridge: Cambridge University Press, 2003. DOI: <https://doi.org/10.1017/CBO9780511840005>

WACHSMUTH, H.; AL-KHATIB, K.; STEIN, B. Using argument mining to assess the argumentation quality of essays. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS: TECHNICAL PAPERS, 26., 2016, Osaka. *Proceedings* [...]. Osaka: The COLING 2016 Organizing Committee, 2016. p. 1680-1691.

WACHSMUTH, H.; NADERI, N.; HABERNAL, I.; HOU, Y.; HIRST, G.; GUREVYCH, I.; STEIN, B. Argumentation quality assessment: Theory vs. practice. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 55., 2017, Vancouver. *Proceedings* [...]. Vancouver: Association for Computational Linguistics, 2017a. p. 250-255. DOI: <https://doi.org/10.18653/v1/P17-2039>. Available in: <https://www.aclweb.org/anthology/P17-2039>. Access on: May. 20, 2021.

WACHSMUTH, H.; NADERI, N.; HOU, Y.; BILU, Y.; PRABHAKARAN, V.; THIJM, T. A.; HIRST, G.; STEIN, B. Computational argumentation quality assessment in natural language. *In: CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, 15, 2017, Valencia. *Proceedings* [...]. Valencia: Association for Computational Linguistics, 2017b. p. 176-187. DOI: <https://doi.org/10.18653/v1/E17-1017>. Available in: <https://www.aclweb.org/anthology/E17-1017>. Access on: May. 20, 2021.

WACHSMUTH, H.; POTTHAST, M.; AL-KHATIB, K.; AJJOUR, Y.; PUSCHMANN, J.; QU, J.; DORSCH, J.; MORARI, V.; BEVENDORFF, J.; STEIN, B. Building an Argument Search Engine for the Web. *In: WORKSHOP ON ARGUMENT MINING (ARGMINING 2017) AT EMNLP*, 4., 2017, Copenhagen. *Proceedings* [...]. Copenhagen: Association for Computational Linguistics, 2017c. p. 49-59. DOI: <https://doi.org/10.18653/v1/W17-5106>. Available in: <https://www.aclweb.org/anthology/W17-5106>. Access on: May. 20, 2021.

WACHSMUTH, H.; STEIN, B.; AJJOUR, Y. “PageRank” for argument relevance. *In: CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, 15., 2017, Valencia. *Proceedings* [...]. Valencia: Association for Computational Linguistics, 2017d. p. 1117-1127. DOI: <https://doi.org/10.18653/v1/E17-1105>. Available in: <https://www.aclweb.org/anthology/E17-1105>. Access on: May. 20, 2021.

WACHSMUTH, H.; WERNER, T. Intrinsic Quality Assessment of Arguments. *In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS*, 28., 2020, Barcelona. *Proceedings* [...]. Barcelona: International Committee on Computational Linguistics, 2020. p. 6739-6745. DOI: <https://doi.org/10.18653/v1/2020.coling-main.592>. Available in: <https://www.aclweb.org/anthology/2020.coling-main.592>. Access on: May. 20, 2021.

WALTON, D. N.; WALTON, D. N. *Informal Logic: A Handbook for Critical Argument*. Cambridge: Cambridge University Press, 1989.

WARDLE, C. Misinformation Has Created a New World Disorder. *Scientific American*, [S.l.], v. 321, p. 88-93, 2019.

WEI, Z.; LIU, Y.; LI, Y. Is this Post Persuasive? Ranking Argumentative Comments in Online Forum. *In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, 54., 2016, Berlin. *Proceedings* [...]. Berlin: Association for Computational Linguistics, 2016. p. 195-200. DOI: <https://doi.org/10.18653/v1/P16-2032>. Available in: <https://www.aclweb.org/anthology/P16-2032>. Access on: May. 20, 2021.

WELTZER-WARD, L.; BALTES, B.; LYNN, L. K. Assessing Quality of Critical Thought in Online Discussion. *Campus-Wide Information Systems*, [S.l.], v. 26, n. 3, p. 168-177, 2009. DOI: <http://doi.org/10.1108/10650740910967357>

Capítulo 3

Recursos linguísticos

Este capítulo apresenta uma breve síntese das pistas linguísticas definidas neste trabalho para a dimensão retórica (WACHSMUTH et al., 2017), empregadas para a elaboração das diretrizes de anotação e avaliação da qualidade da argumentação das mensagens políticas. Também são apresentadas as ferramentas computacionais, disponíveis para o português, aplicadas para a extração das *features* linguísticas.

3.1 Pistas linguísticas

Neste trabalho, foram utilizadas 30 pistas linguísticas definidas em Silva et al. (2021): 4 para o aspecto Clareza, 7 para o aspecto Organização, 6 para o aspecto Credibilidade e 13 para o aspecto Apelo Emocional (6 para polaridade e 7 para intensidade). A Adequação não foi considerada neste trabalho uma vez que se mostrou não relevante para a qualidade da argumentação em *tweets*, e também porque os anotadores e linguistas que participaram do processo de criação das diretrizes apontaram que os critérios referentes à Adequação já eram contemplados pelos outros quatro aspectos. Com base na presença dessas pistas linguísticas, os *tweets* foram pontuados como baixo, médio ou alto para quase todos os aspectos, exceto para a polaridade do apelo emocional, que foi pontuada como positivo, negativo ou neutro. Nas seções seguintes, apresentam-se as pistas linguísticas definidas para cada aspecto.

3.1.1 Clareza

Em Wachsmuth et al. (2017), os autores consideram um argumento claro se usa linguagem gramaticalmente correta e amplamente inequívoca e evita complexidade desne-

cessária e desvio do assunto discutido. Além disso, a linguagem utilizada deve facilitar o entendimento e não deixar dúvidas sobre a posição do autor e a forma como ele a defende.

Nessa perspectiva, em Silva et al. (2021), assumiu-se que todo argumento escrito em português tem o potencial de ser naturalmente claro, a menos que existam certos fenômenos linguísticos que impactam negativamente na clareza. Dessa forma, a pontuação do aspecto clareza diminui com a presença de uma ou mais pistas linguísticas que prejudicam a clareza da argumentação, a saber, (i) pergunta que leva à dúvida, (ii) linguagem complexa desnecessária, (iii) presença de erros de língua portuguesa e (iv) desvio desnecessário do assunto principal.

Em (1), é apresentado um exemplo dessas pistas linguísticas. O uso incorreto do substantivo “crítica” quando o correto deveria ser o verbo “criticar” e a pergunta no início do texto levantam algumas dúvidas sobre o que o autor quis dizer. Devido à presença dessas duas pistas linguísticas, o *tweet* em (1) foi classificado como de média clareza.

- (1) @SargentoFAHUR Sr deputado. **Não seria de competência dos senhores deputados criarem leis para evitar que isso aconteça ?** Os juízes não estão acima das leis ,uma lei forte e correta não deixaria brecha para isso acontecer . **Crítica** a justiça é moleza , fazer algo para mudar é a sua obrigação.

3.1.2 Organização

A definição de Organização proposta por Wachsmuth et al. (2017) considera que um texto deve ser composto de três partes, na seguinte ordem: (i) introdução do assunto, (ii) alguns argumentos para sustentar a conclusão e (iii) a conclusão. Essa definição, porém, não se sustenta quando analisamos textos do Twitter, pois os usuários não dispõem de espaço suficiente para desenvolver essas 3 partes. Assim, Silva et al. (2021) redefiniram a noção de organização para este Gênero Textual (GT) considerando a presença de pistas linguísticas que pontuam positivamente para a organização de um argumento.

As sete pistas linguísticas utilizadas para avaliar o aspecto Organização estão relacionadas à presença de relações linguísticas escritas conhecidas entre sentenças e locuções: (i) relação condicional; (ii) relação de concessão; (iii) oposição ou contraste ou (iv) comparação entre duas ideias; (v) relações de causa e efeito, explicação ou finalidade; (vi) encadeamento ou enumerações cronológicas; e (vii) exemplificação ou interconexão lógica.

Ao contrário do aspecto da Clareza, que sustenta que todo argumento no Twitter tem o potencial de ser inerentemente claro, no aspecto Organização, considera-se que este GT tem o potencial intrínseco de ser desorganizado, a menos que ocorram certos fenômenos linguísticos que contribuem positivamente para a organização. Dessa forma, todo *tweet* começa com uma pontuação de organização baixa, que aumenta à medida que uma ou mais das dicas linguísticas acima são identificadas.

Em (2), são apresentados exemplos de algumas dessas pistas linguísticas. Em “para que o processo jurídico penal seja cumprido desde o seu início”, é possível identificar uma relação de propósito introduzida pelo marcador discursivo “para que”. Existe também um encadeamento cronológico em “A PGR já denunciou [...] e já prendido. Agora precisa [...]”, pois apresenta uma sequência temporal entre o passado (o que foi feito) e o futuro (o que deve ser feito). Por fim, o exemplo (2) também ilustra uma relação de concessão em “mesmo o STF tendo tomado a frente” e uma relação de explicação em “o mesmo que houve com o Lula, uma ampla defesa” introduzido pelo marcador discursivo “mesmo” e finalizado com a aposta explicativa “a ampla defesa”. Devido à presença dessas pistas linguísticas, o *tweet* em (2) foi pontuado como tendo alta organização.

- (2) @BolsonaroSP @danielPMERJ O Deputado **PRECISA** ser solto, **para que o processo jurídico penal seja cumprido desde o seu início**. A PGR **já** denunciou **mesmo o STF tendo tomado a frente** e **já** prendido. **Agora** precisa entrar com a parte da defesa e acontecer **o mesmo que houve com o Lula, a ampla defesa**

3.1.3 Credibilidade

De acordo com Wachsmuth et al. (2017), um argumento deve ser avaliado como bem-sucedido em criar credibilidade se transmitir argumentos e outras informações de uma forma que torne o autor credível, por exemplo, indicando a honestidade do autor, a polidez da linguagem utilizada ou revelando o conhecimento do autor ou experiência em relação aos assuntos discutidos.

Para a avaliação do aspecto credibilidade, Silva et al. (2021) consideraram que um argumento escrito em português é verossímil se algumas pistas linguísticas estiverem presentes na superfície textual; ou seja, não levaram em consideração nenhum critério ou dado externo, como aceitação ou engajamento do autor nas redes sociais. Vale ressaltar que o Twitter é uma plataforma aberta e, portanto, qualquer usuário registrado pode postar qualquer tipo de mensagem nesta plataforma de mídia social. Como não há uma pré-análise do perfil do usuário ou do conteúdo por ele postado, a dúvida sobre a credibilidade do que é postado é inerente à plataforma.

- (3) @KimKatagui **Hoje fui comprar 1kg de carne moída para o almoço e deu R\$ 43,00 achei um absurdo!** Em que mundo estamos com um custo tão alto de carne assim?! Mas de boa estamos comprando carne de primeira para nossos representantes políticos, então pq reclamar? !

Assim, a pontuação do aspecto credibilidade aumenta com a presença de (i) dados específicos (por exemplo, data e quantidade); (ii) um fato de mídia, histórico ou enciclopédico (por exemplo, 11 de setembro); (iii) uma autoridade pública (por exemplo, a

OMS); (iv) uma *hashtag* que impõe uma posição (por exemplo, #BlackLivesMatter); (v) um termo especializado de alguma área do conhecimento (por exemplo, *lockdown*); e (vi) relato de experiência pessoal ou individual, conforme demonstrado em (3). O *tweet* em (3) foi pontuado como média credibilidade devido ao relato de experiência individual e à presença de dados específicos (1kg e R\$ 43,00).

3.1.4 Apelo Emocional

Segundo Wachsmuth et al. (2017), um argumento tem apelo emocional quando gera emoções no interlocutor. Em Silva et al. (2021), as pistas linguísticas para o aspecto apelo emocional foram divididas em polaridade e intensidade.

3.1.4.1 Polaridade

A polaridade do apelo emocional de um argumento é considerada positiva ou negativa se contribui para criar emoções boas ou ruins no leitor, respectivamente. Se seu conteúdo não causa nenhuma emoção ou se o apelo emocional é bem equilibrado entre positivo e negativo, então a polaridade é considerada neutra.

Assim, as pistas linguísticas para uma polaridade positiva são (i) a presença de uma referência cordial a uma pessoa/organização e (ii) o uso de linguagem polida. Por outro lado, as pistas linguísticas para uma polaridade negativa são (i) a presença de uma referência pejorativa a uma pessoa/organização; (ii) o uso de xingamento ou palavra de baixo calão; (iii) presença de discurso de ódio ou ameaça; e (iv) o uso de expressões que denotam especulação. Não há nenhuma pista linguística específica para uma polaridade neutra, pois é o meio termo entre os dois extremos. A pontuação final para a polaridade é definida como a diferença entre a quantidade total de sinais positivos e negativos.

O Exemplo (4) mostra indícios de polaridade positiva e negativa. Quando o autor se refere cordialmente à deputada como “Prezada”, observamos uma pista positiva. Por outro lado, quando se refere à mesma deputada como “coxa” (burra), é considerada uma referência pejorativa. O mesmo exemplo está cheio de palavras que pesam negativamente para a polaridade, como “merda”, “rabo” e “vagabundo”. Nesse caso, como há mais pistas negativas do que positivas, o *tweet* em (4) foi avaliado como sendo de polaridade negativa para o aspecto apelo emocional.

- (4) @gleisi **Prezada coxa** vc **não tem vergonha nessa cara** reformada com dinheiro público?? **A merda do seu partido** que se diz do povo **não fez nada disso** e ainda **enfiou dinheiro nosso no rabo do joesley do eike e do Marcelo**. **Esse discursinho vagabundo não cola mais.**

3.1.4.2 Intensidade

O apelo emocional também é avaliado de acordo com sua intensidade, que é determinada pela quantidade de pistas linguísticas que potencializam a criação de emoções. Quanto maior o número de pistas linguísticas, maior a intensidade do apelo emocional. A intensidade é avaliada pela presença de (i) pronome ou verbo na primeira pessoa; (ii) a repetição de sinais de pontuação; (iii) estrutura enfática (por exemplo, palavra em maiúscula); (iv) uma frase ou slogan imperativo; (v) uma expressão que denota exagero (por exemplo, “todo mundo”); (vi) linguagem não verbal (por exemplo, *emoticons*); e (vii) a presença de uma expressão idiomática, provérbio ou metáfora.

O Exemplo (4) apresenta pelo menos 4 dessas pistas linguísticas: (i) presença de pronome de primeira pessoa em “nosso”; (ii) a repetição do sinal de pontuação “?? ”; (iii) expressão que denota exagero, quando o autor diz “não fez nada disso”; e (iv) a presença de 2 expressões idiomáticas, nomeadamente, “não ter vergonha na cara” e “enfiar dinheiro no rabo”, e um sentido metafórico do verbo “colar” na expressão “Esse discursinho vagabundo não cola mais”. Por isso, este *tweet* foi classificado como de alta intensidade para o aspecto apelo emocional.

Nesta seção, além do apelo emocional, foram apresentadas outras pistas linguísticas que foram estabelecidas em Silva et al. (2021) para avaliação da qualidade da argumentação. Uma descrição detalhada destas pistas, bem como as diretrizes usadas na anotação do corpus, estão presentes no Manual de Anotação disponível no site do projeto ArgQ!¹ e no Apêndice A. Na próxima seção, serão apresentados alguns dos recursos linguísticos disponíveis para serem utilizados na avaliação automática da qualidade de argumentos em português brasileiro.

3.2 Ferramentas computacionais

Com o objetivo de encontrar os recursos computacionais que melhor se correlacionam com as pistas linguísticas utilizadas para medir a qualidade da argumentação, nesta seção são apresentadas todas as ferramentas utilizadas neste trabalho para extrair *features* computacionais. Vale ressaltar que a maioria das ferramentas e recursos utilizados neste trabalho foram gerados por alunos e pesquisadores do Núcleo Interinstitucional de Linguística Computacional (NILC)².

3.2.1 Normalização de texto com Enelvo (BERTAGLIA, 2017)

O Enelvo³ (BERTAGLIA, 2017) é uma ferramenta para normalizar o conteúdo gerado pelo usuário a partir de palavras ruidosas escritas em português em textos como *tweets*,

¹ Disponível em: <<https://argq.org/>>

² Disponível em: <<http://www.nilc.icmc.usp.br/>>

³ Disponível em: <<https://github.com/thalesbertaglia/enelvo>>

postagens em blogs e análises de produtos. Ele é capaz de identificar e normalizar erros ortográficos, gírias da Internet, siglas, nomes próprios, entre outros. O processo de normalização envolve a padronização do conteúdo criado espontaneamente por indivíduos comuns sem vínculo com a mídia (BERTAGLIA, 2017), que é chamado de CGU.

Em Bertaglia (2017), foi proposto o uso de *word embeddings* e comparado seu método com ferramentas existentes para o português do Brasil em duas amostras anotadas de avaliações de produtos escritas por usuários. A técnica foi aplicada para corrigir erros ortográficos e gírias da Internet. O processo de normalização foi dividido em três etapas: identificação de ruído, geração de candidatos e classificação de candidatos. Para a tarefa de identificação de ruído, foi utilizada uma arquitetura neural profunda, formada por uma rede neural convolucional recorrente (RCNN), e a precisão média obtida foi de 93,64%. Para as tarefas de geração e classificação de candidatos, foram testadas abordagens usando medidas de similaridade lexical e baseadas em *embeddings* de palavras. Três medidas de similaridade lexical foram usadas para gerar candidatos: distância de edição, subsequência comum mais longa (LCS) (ULLMAN; AHO; HIRSCHBERG, 1976) e simlex (HASSAN; MENEZES, 2013). As abordagens baseadas em *word embeddings* foram Word2Vec (MIKOLOV et al., 2013), Wang2Vec (LING et al., 2015) e FastText (BOJANOWSKI et al., 2017). O modelo Wang2Vec apresentou os melhores resultados, com cobertura média de 86%.

Em (5), é apresentado um exemplo do processo de normalização. As repetições de letras foram reduzidas para três e as abreviações (por exemplo, “Pq” e “vc”) foram reescritas em suas formas inteiras (“porque” e “você”, respectivamente). Além disso, os nomes de usuário foram descartados no processamento posterior. Para poder extrair algumas características específicas, no pré-processamento paralelo, optou-se por não converter letras maiúsculas em minúsculas, pois isso poderia atrapalhar a análise de correlação de algumas pistas linguísticas que levam em consideração as maiúsculas.

(5) @jandira_feghali Mentira? Processa ele então... memorando... **kkkk** Falta vergonha na cara. Se ele mentiu sobre coisa tão séria pode pedir a saída dele. PROCESSA... Mas não vão fazer isso. **Pq** é verdade. **Vc** mente para seus eleitores. Mentira ? Processa ele então . memorando . **kkk** Falta vergonha na cara . Se ele mentiu sobre coisa tão séria pode pedir a saída dele . PROCESSA . Mas não vão fazer isso . **Porque** é verdade . **Você** mente para seus eleitores .

3.2.2 Análise de polaridade com Random Forest e TweetSentBR (BRUM; NUNES, 2018)

O TweetSentbr⁴ é um corpus de sentimentos para o português brasileiro, anotado manualmente com 15.000 sentenças no domínio de programas de TV. As sentenças foram

⁴ Disponível em: <<https://bitbucket.org/hbrum/tweetsentbr/src/master/>>

rotuladas em três classes – positiva, neutra e negativa – por sete anotadores.

A partir desse corpus, Brum e Nunes (2018) realizaram experimentos sobre classificação de polaridade usando seis abordagens de aprendizado de máquina: SVM linear, Bernoulli Naive Bayes, regressão logística, *multilayer perceptron*, árvore de decisão e *random forest*. O modelo baseado em regressão logística foi o melhor, com um valor de medida F de 80,38% na classificação binária, considerando apenas as classes positiva e negativa. Ao incluir a classe neutra, o valor de medida F foi de 64,87%.

Em 6, é apresentado um exemplo da análise de polaridade em que as palavras “voto” e “democrático” foram classificadas como positivas (+) e a palavra “vergonhosa” foi classificada como negativa (-).

(6) parabéns a todos que desde o início se opuseram a essa **vergonhosa**⁻ lei . tem que acabar com imunidade parlamentar e foro privilegiado , os eleitos são só pessoas que receberam mandato através do **voto**⁺ **democrático**⁺ para trabalhar para o cidadão que e o elegeu e paga o seu salário e só

3.2.3 Análise de sentimentos com LIWC (BALAGE FILHO; PARDO; ALUÍSIO, 2013)

O *Linguistic Inquiry and Word Count* (LIWC)⁵ é uma ferramenta de análise de texto que calcula o grau de uso de diferentes categorias de palavras em uma ampla variedade de textos (PENNEBAKER; FRANCIS; BOOTH, 2001). Consiste em um léxico, mais conhecido como dicionário LIWC, que foi disponibilizado para o português por pesquisadores do NILC, Checon Pesquisa e Unisinos e está disponível em PortLEX⁶. Mais detalhes sobre a versão em português do Brasil do LIWC podem ser encontrados em (BALAGE FILHO; PARDO; ALUÍSIO, 2013).

Após o processamento da frase (por exemplo, 6), a ferramenta retorna uma string que contém a métrica com o número de ocorrências de todos os tokens identificados, conforme Exemplo (7):

(7)

```
Counter({'funct': 26, 'cogmech': 22, 'social': 16, 'pronoun': 14,
'ipron': 12, 'ppron': 9, 'conj': 9, 'you': 8, 'shehe': 8,
'preps': 8, 'article': 7, 'incl': 7, 'space': 6, 'time': 6,
'insight': 6, 'discrep': 5, 'tentat': 5, 'excl': 5, 'affect': 5,
'relativ': 5, 'humans': 4, 'inhib': 4, 'negemo': 3, 'present': 3,
'work': 3, 'nonfl': 2, 'cause': 2, 'auxverb': 2, 'bio': 2,
'verb': 2, 'achieve': 2, 'posemo': 2, 'they': 2, 'sad': 2,
'swear': 2, 'motion': 2, 'anx': 1, 'body': 1, 'money': 1,
'certain': 1, 'ingest': 1, 'health': 1, 'past': 1})
```

As métricas mais relevantes utilizadas neste trabalho estão relacionadas à polaridade do sentimento, portanto *posemo* (emoção positiva) e *negemo* (emoção negativa). No

⁵ Disponível em: <<http://www.liwc.net/>>

⁶ Disponível em: <<http://143.107.183.175:21380/portlex/index.php/en/liwc>>

processamento do Exemplo (7), foram encontradas 2 palavras que denotam sentimento positivo (posemo: 2) “**imunidade**” e “**receberam**”, e 3 que denotam sentimento negativo (negemo: 3) “**vergonhosa**”, “**só** pessoas” e “**e só**”.

3.2.4 Métricas textuais e psicolinguísticas com NILCMetrix (LEAL, 2021)

O NILC-Metrix⁷ é um conjunto de métricas gerado como resultado do doutorado de Leal (2021), que compreende 200 métricas linguísticas e psicolinguísticas utilizadas na avaliação de métodos de predição de complexidade textual. As métricas são categorizadas em 14 grupos: medidas descritivas, simplicidade textual, coesão referencial, coesão semântica, medidas psicolinguísticas, diversidade lexical, conectivos, léxico temporal, complexidade sintática, densidade de padrões sintáticos, informação morfossintática da palavra, informação semântica da palavra, frequência de palavras e índices de leituraabilidade. Como exemplo, na Tabela 2, apresentam-se os “índices de leituraabilidade” calculados para a mensagem em (6). A relação com todos os índices calculados a partir do Exemplo (6) estão listados no Anexo A.

Leal (2021) investigou abordagens de aprendizado por classificação e transferência com a adição de métricas de rastreamento ocular, que atingiram o estado da arte para a tarefa de prever a complexidade de sentenças em português com 97,5% de precisão.

Tabela 2 – Grupo de métricas de índices de leituraabilidade obtidos por processamento da mensagem (6).

Métrica	Descrição	Valor
brunet	Índice de Brunet	8,55513
dalechall_adapted	Fórmula de Dale Chall adaptada	9,9293
flesch	Índice Flesch	44,108
gunning_fox	Índice Gunning Fog	10,136
honore	Estatística de Horoné	1076,01434

Como medida de legibilidade, a **Estatística de Brunet** é uma forma de relação *type/token* menos sensível ao tamanho do texto. Eleva-se o número de *types* à constante -0,165 e depois eleva-se o número de *tokens* a esse resultado. A Estatística de Brunet é definida na Equação 1:

$$BI = N^{V^{-0.165}} \quad (1)$$

onde **V** é a quantidade de *types* (considera palavras sem repetições) e **N** é a quantidade de *tokens* (considera palavras com repetições).

A **fórmula de leituraabilidade de Dalechall adaptada** combina a quantidade de palavras não familiares com a quantidade média de palavras por sentença. “Palavras não

⁷ Disponível em: <<http://fw.nilc.icmc.usp.br:23380/nilcmatrix>>

familiares” são aquelas que não constam do vocabulário básico conhecido por alunos do quarto ano. A métrica tem equivalência com os níveis escolares, e quanto maior o valor da métrica, maior a complexidade textual. A fórmula adaptada de Dalechall é definida na Equação 2:

$$ADC = (0.1579 * PDW) + (0.0496 * ASL) + 3.6365 \quad (2)$$

onde **PDW** é a porcentagem de palavras difíceis e **ASL** é o comprimento médio da frase em palavras.

O **Índice de Leitabilidade de Flesch** busca uma correlação entre tamanhos médios de palavras e sentenças. Quanto maior o resultado da métrica, menor a complexidade textual. O Índice de Leitabilidade de Flesch é definido na Equação 3

$$Flesch = 248.835 - 1.015 * ASL - 84,6 * ASW \quad (3)$$

onde **ASL** é a média de palavras por sentença e **ASW** é a média de sílabas por palavra.

O **índice de leitabilidade Gunning Fog** (também conhecido como Gunning FoX) soma a quantidade média de palavras por sentença ao percentual de palavras difíceis no texto e multiplica tudo por 4. O resultado está diretamente ligado aos 12 níveis do ensino americano. Índices superiores a 12 representam textos extremamente complexos. Palavras difíceis, para essa métrica, são aquelas que possuem mais de duas sílabas.

A **Estatística de Honoré** é um tipo de type/token ratio que leva em consideração, além da quantidade de types e tokens, a quantidade de hapax legomena.. The Honoré’s statistic is defined in Equation 4:

$$Honore = \frac{100 \log N}{1 - \frac{V1}{V}} \quad (4)$$

onde **N** é o número total de *tokens* (total de palavras no texto, contando as repetições), **V1** é o número de hapax legomena (palavras do vocabulário que aparecem uma única vez), e **V** é o número de *types* (quantidade de palavras sem considerar suas repetições). Um *hapax legomena* é um *type* que só apresenta um *token* no texto.

3.2.5 Detecção de linguagem tóxica com BERT e ToLD-Br (LEITE et al., 2020)

O *Toxic Language Dataset for Brazilian Portuguese* (ToLD-Br)⁸ é um cópuz com *tweets* criado com o intuito de detectar comentários tóxicos. O cópuz é composto por 21.000 *tweets* escritos em português do Brasil que foram anotados manualmente em sete categorias: não tóxico, LGBTQ+fobia, obsceno, insultante, racismo, misoginia e xenofobia.

⁸ Disponível em: <<https://github.com/JAugusto97/ToLD-Br>>

Usando este conjunto de dados, Leite, Silva, Bontcheva e Scarton (LEITE et al., 2020) treinaram um modelo computacional com BERT (DEVLIN et al., 2019). Entre outros modelos, os modelos monolíngues BERT (BR-BERT) – modelo de linguagem BERT pré-treinado para o português do Brasil – e BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020) e o BERT multilíngue (WOLF et al., 2020) (M-BERT-BR) tiveram resultados muito semelhantes. No entanto, o modelo M-BERT-BR alcançou um macro-F1 melhor (75%) para a classe positiva, com menor taxa de falsos negativos do que o modelo BR-BERT.

Usando o modelo previamente treinado de (LEITE et al., 2020), cada postagem foi classificada como tóxica ou não tóxica. Por exemplo, o *tweet* em (4) foi classificado como linguagem tóxica devido à presença das palavras “coxa”, “merda”, “rabo” e “vagabundo”.

Capítulo 4

Classificação da qualidade da argumentação em *tweets* no domínio da política brasileira

Este capítulo apresenta artigo aceito para publicação na Revista Linguamática no qual são descritos os experimentos realizados com o objetivo de produzir um modelo computacional capaz de realizar a avaliação automática da qualidade da argumentação em mensagens políticas, escritas em português. Este artigo apresenta os algoritmos utilizados e o resultado dos experimentos realizados no *cópus* de mensagens. A primeira seção apresenta avaliação da qualidade da argumentação como área de pesquisa e apresenta as questões de pesquisa que se busca responder com este estudo. A segunda seção apresenta os trabalhos relacionados a esta pesquisa, incluindo a taxonomia proposta por Wachsmuth et al. (2017). A terceira seção apresenta, de forma resumida, o processo de elaboração e anotação do *cópus*, além do detalhamento de como foram calculadas as notas de cada um dos aspectos da taxonomia. A quarta seção apresenta os experimentos realizados com AM baseados em *features* e os experimentos com BERTimbau e RoBERTaTwitterBR. A quinta seção apresenta os resultados encontrados e as análises qualitativas e quantitativas. A sexta seção apresenta as conclusões, limitações e trabalhos futuros.

Embora o artigo seja uma produção conjunta entre o autor desta tese e suas orientadoras, vale ressaltar que a concepção do artigo e sua escrita, bem como a realização de todos os experimentos e análises, têm contribuição majoritária do autor desta tese. Além disso, destaca-se que os editores da Linguamática autorizaram expressamente a publicação do artigo, em sua íntegra, nesta tese.

Proposta recebida em Outubro 2022 e aceite para publicação em Março 2023.

Classificação da qualidade da argumentação em *tweets* no domínio da política brasileira

Argument Quality Assessment in Brazilian political tweets

Cássio Faria da Silva

Rede Gonzaga de Ensino Superior / Universidade Federal de São Carlos
cassiofs@gmail.com

Vânia Paula de Almeida Neris
Universidade Federal de São Carlos
vania.neris@ufscar.br

Helena de Medeiros Caseli
Universidade Federal de São Carlos
helenacaseli@ufscar.br

Resumo

A argumentação é uma habilidade inerente à comunicação humana, tanto em situações orais quanto escritas. Argumentos bem fundamentados são importantes para amparar a tomada de decisões e aprendizado, assim como para a obtenção de conclusões amplamente aceitas. Como área de pesquisa, a argumentação é um campo multidisciplinar que estuda os processos de debate e raciocínio. Em linguística computacional, investigações têm sido realizadas para (i) identificar argumentos e suas unidades e (ii) gerar ou (iii) avaliar a qualidade dos argumentos. No entanto, a maioria dos trabalhos atuais se concentra na mineração de argumentos em textos formais em inglês. Neste artigo, foi avaliada a qualidade da argumentação em *tweets* de domínio político, escritos em português do Brasil, usando algoritmos tradicionais de aprendizado de máquina – como Regressão Logística, *K-Nearest Neighbors*, Árvores de Decisão, Máquinas de Vetores Suporte (SVM), Floresta Aleatória e *Naive Bayes* – e também um ajuste fino de dois modelos neurais (BERTimbau e RobertaTwitterBR). Além de trazer resultados práticos para a avaliação da qualidade da argumentação em um gênero textual desafiador, como o Twitter, e em um domínio controverso, como a política brasileira, este artigo também visa suprir a carência de trabalhos que avaliem automaticamente a qualidade dos argumentos em português. Dentre os algoritmos de classificação avaliados, o modelo obtido a partir do ajuste fino do BERTimbau apresentou os melhores resultados com uma precisão de 69,65% quando foram consideradas todas as classes e de 100,00% para as mensagens de alta qualidade de argumentação.

Palavras chave

avaliação da qualidade da argumentação, *tweet*, BERT, política brasileira, processamento da língua portuguesa

Abstract

Argumentation is an inherent skill in human communication, both in oral and written situations. Well-founded arguments are important to support decision-making and learning, as well as to reach widely accepted conclusions. As a research area, argumentation is a multidisciplinary field that studies the processes of debate and reasoning. In computational linguistics, investigations have been carried out to (i) identify arguments and their units and (ii) generate or (iii) evaluate the quality of arguments. However, most current work focuses on argument mining in formal English texts. In this article, we evaluated the quality of argumentation in political domain tweets, written in Brazilian Portuguese, using traditional machine learning algorithms – such as Logistic Regression, KNearest Neighbor, Decision Trees, Support Vector Machines (SVM), Random Forest and Naive Bayes – and also a fine-tuning of two neural models (BERTimbau and RobertaTwitterBR). In addition to bringing practical results for the assessment of argumentation quality in a challenging textual genre, such as Twitter, and in a controversial domain, such as Brazilian politics, this article also aims to fill in the lack of works that automatically assess the quality of arguments in Portuguese. Among the evaluated classification algorithms, the model obtained from the fine-tuning of BERTimbau presented the best results, with an accuracy of 69.65% when all classes were considered and 100.00% for messages with high quality of argumentation.

Keywords

argument quality assessment, tweet, BERT, Brazilian politics, Portuguese language processing

1 Introdução

Um argumento é uma afirmação (ou conclusão) acompanhada por um número arbitrário de premissas que justificam, fundamentam, apoiam, defendem ou explicam a afirmação (Potthast et al., 2019). Argumentos bem fundamentados são importantes para amparar a tomada de decisões e aprendizado, assim como para a obtenção de conclusões amplamente aceitas. A argumentação (capacidade de produzir argumentos) é uma habilidade inerente à comunicação humana tanto em situações orais quanto escritas. Como área de pesquisa, a argumentação é um campo multidisciplinar que estuda os processos de debate e raciocínio (Habernal & Gurevych, 2017). Para Eemeren & Grootendorst (2003), a argumentação consiste em uma ou mais sentenças nas quais várias premissas são apresentadas para sustentar uma conclusão. As sentenças que fazem parte da argumentação constituem uma expressão completa que visa convencer um interlocutor.

Em linguística, estuda-se a argumentação em textos em linguagem natural (Stab & Gurevych, 2017a). Na ciência da computação, a identificação ou avaliação automática da argumentação é estudada no campo da inteligência artificial (Bench-Capon & Dunne, 2007). Ao combinar essas duas áreas de pesquisa, em linguística computacional ou no processamento de linguagem natural (PLN), investigações têm sido realizadas para (i) identificar argumentos e suas unidades e (ii) gerar ou (iii) avaliar a qualidade de tais argumentos. Mais especificamente, as tarefas mais comumente investigadas são: a mineração de unidades de argumentação (Al-Khatib et al., 2016; Habernal & Gurevych, 2015, 2017), a detecção de evidências que apoiam reivindicações¹ (Rinott et al., 2015) e a identificação de relações argumentativas (Peldszus & Stede, 2015). Outros trabalhos classificaram esquemas de argumentação (Feng et al., 2014), realizam a análise de estruturas gerais de argumentação (Wachsmuth et al., 2015; Stab & Gurevych, 2017a) e geram reivindicações (Bilu & Slonim, 2016).

Algumas teorias ou dimensões da qualidade da argumentação foram avaliadas computacionalmente (Stab & Gurevych, 2017b; Wachsmuth et al., 2017d; Zhang et al., 2016). No entanto, de acordo com Wachsmuth et al. (2017b), ainda não se constituiu um conceito geral para a qualidade da argumentação ou uma definição clara de suas

¹Reivindicações, no contexto dos trabalhos relacionados à mineração de unidades de argumentação, se referem a quaisquer declarações ou afirmações que possuam propósito argumentativo.

dimensões. Apesar da falta do conceito geral, tarefas relacionadas à argumentação computacional – como mineração, geração, identificação de argumentos e sua avaliação – têm se mostrado relevantes em atividades como apoio à escrita e assistência à discussão (Stab & Gurevych, 2017b; García-Gorrostieta et al., 2018).

No campo da comunicação, como bem pontuado por Lytos et al. (2019), a internet e as redes sociais são, hoje, o meio de comunicação mais utilizado. Consistem em espaços que permitem a emissão de opiniões sobre qualquer assunto e são fonte para a produção de um grande volume de textos com potencial argumentativo. De especial interesse para este trabalho é a rede social Twitter², bastante utilizada para troca de informações e opiniões sobre política no Brasil.

A argumentação no Twitter, além de envolver características específicas do gênero textual, também é permeada pelas necessidades de comunicação, pelo contexto histórico e pelo assunto da mensagem (Marcuschi et al., 2002), como exemplificado no exemplo a seguir³:

- (1) @gleisi Prezada **coxa** vc não tem vergonha nessa cara reformada com dinheiro público?? **A merda do seu partido** que se diz do povo não fez nada disso e ainda **enfiou dinheiro nosso no rabo** do joesley do eike e do Marcelo. **Esse discursinho vagabundo não cola mais.**

Entre os indícios linguísticos que impactam a qualidade da argumentação neste exemplo (1), podem ser citados: (i) referências pejorativas como “coxa”; (ii) a repetição de sinais de pontuação (??) que indicam indignação; (iii) a presença de discurso de ódio contra um partido político em “a merda do seu partido” e contra a autora do *tweet* original (ao qual este é uma resposta) em “esse discursinho vagabundo”; (iv) expressões coloquiais como a expressão idiomática em “enfiou [...] no rabo” e gíria em “não cola mais”; (v) uma tentativa de tornar o argumento mais pessoal em “dinheiro nosso”; e (vi) contexto histórico citando “joesley” (Joesley Batista), “eike” (Eike Batista) e “Marcelo” (Marcelo Odebrecht), três empresários brasileiros que ficaram nacionalmente conhecidos em um dos escândalos políticos ocorridos no Brasil.

Ao lidar com textos do Twitter, escritos em

²<https://twitter.com/>

³Todos os exemplos de *tweets* apresentados neste artigo são transcritos exatamente como foram publicados pelos seus autores. O destaque (em negrito) de alguns trechos foi inserido pelos autores deste artigo para enfatizar.

português, esse trabalho enfrenta desafios não presentes na maioria dos trabalhos da literatura, que focam na mineração de argumentos em textos formais em inglês. Em relação ao idioma, vale destacar que a língua portuguesa é uma das mais faladas no mundo, com mais de 280 milhões de falantes. Nesse sentido, são notáveis os avanços e a importância que tem sido dada atualmente ao desenvolvimento e aprimoramento dos recursos de PLN para o português. Isso pode estar relacionado, em grande parte, aos aspectos da globalização e da popularização do acesso ao ambiente online, que são fatores facilitadores do intercâmbio sociolinguístico dessa língua⁴.

Em relação ao gênero textual, é importante salientar que embora hajam recursos linguísticos valiosos para o português, a maioria foi desenvolvida para lidar com textos formais, bem escritos e autocontidos, como artigos de jornal. No entanto, o cenário atual da comunicação nas mídias sociais exige que os recursos de PLN sejam capazes de lidar com textos cheios de desafios, como a presença de gírias, linguagem figurada, erros de português e uso do “internetês” (vocabulário próprio composto por termos e abreviações usualmente encontrados em textos de mídias sociais). Nesse sentido, embora fenômenos linguísticos que trazem indícios de uma qualidade da argumentação boa ou ruim possam ser identificados nos *tweets*, não há garantia de que os recursos e as ferramentas disponíveis hoje para o processamento automático também sejam capazes de identificá-los.

No que diz respeito à avaliação da qualidade da argumentação, que é o foco deste artigo, desde o esquema argumentativo de Toulmin (2003), estudos têm sido realizados para simplificar a compreensão da estrutura e determinar a importância dos elementos argumentativos do texto. Wachsmuth et al. (2017b) propuseram uma taxonomia composta por três dimensões para avaliar a qualidade da argumentação: retórica, lógica e dialética. No entanto, desde então, poucos estudos (Potthast et al., 2019; Wachsmuth & Werner, 2020; Skitalinskaya et al., 2021) se dedicaram à avaliação automática da qualidade da argumentação, muito menos em gêneros textu-

⁴Para se ter uma ideia desse crescente interesse, o Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (CETIC) observou o comportamento de brasileiros maiores de 16 anos na internet entre 23 de junho e 8 de julho de 2020. Seus achados mostraram que 49% dos internautas realizavam atividades laborais e 72% buscavam informações relacionadas à saúde na internet. Para mais informações, consulte: https://cetic.br/media/docs/publicacoes/2/20200817133735/paine1_tic_covid19_iedicao_livro%20e1etr%C3%B4nico.pdf

ais cujos textos apresentam conteúdos distantes da norma linguística padrão e longe da própria noção convencional de argumentação.

Entre os trabalhos mais recentes que exploraram essa temática pode-se citar o de Gretz et al. (2019), que aplicaram o modelo de dimensões de qualidade da argumentação proposto por Wachsmuth et al. (2017b) para avaliar a qualidade da argumentação no corpus IBM-Rank-30k, contendo 30.497 argumentos provenientes de clubes de debates, no idioma inglês, onde os participantes são incentivados a redigirem textos bem escritos e com argumentos de alta qualidade. Foram apresentadas avaliações com três métodos baseados em BERT: BERT-Vanilla, BERT-Finetune e BERT-FT_{TOPIC}⁵. Os experimentos evidenciaram um melhor desempenho com o BERT-FT_{TOPIC}, com um coeficiente de correlação de Pearson de 0,53, calculado em relação a uma fração superior e inferior dos argumentos que estão nos extremos da escala de qualidade da argumentação. De acordo com os autores, as dimensões de Relevância e Eficácia Globais são as mais indicativas para os índices gerais de qualidade. Outros trabalhos que usaram o BERT incluem (Fromm et al., 2019) e (Reimers et al., 2019), mas ambos para mineração de argumentos e não para a avaliação da qualidade da argumentação.

Considerando as lacunas identificadas nos estudos que focam na avaliação automática da qualidade da argumentação, esse trabalho é inédito no sentido de que é o primeiro a investigar métodos para avaliar automaticamente a qualidade da argumentação: (i) em postagens de redes sociais e (ii) em português. Assim, neste artigo investigou-se como utilizar o aprendizado de máquina para classificar a qualidade da argumentação em *tweets* de domínio político escritos em português.

As questões de pesquisa que se busca responder com este estudo são:

- (Q1) É possível prever automaticamente a qualidade da argumentação em *tweets* no domínio da política brasileira?
- (Q2) Como identificar automaticamente os *tweets* no campo da política brasileira com bons argumentos?

Para tanto, foram usados algoritmos de Aprendizado de Máquina (AM) baseados em *features*, como regressão logística (LR), *K-Nearest Neighbor* (KNN), árvore de decisão

⁵Consiste em concatenar o tópico da discussão ao argumento, separando-os por um delimitador.

(DT), máquinas de vetor de suporte (SVM), Floresta Aleatória (RF) e *Naive Bayes* (NB) – e também um ajuste fino de um modelo BERT e um RoBERTa. Os experimentos realizados com o BERT alcançaram uma precisão de 100% para as mensagens consideradas com Alta qualidade de argumentação. Enquanto os algoritmos baseados em *features* obtiveram precisões médias que variaram de 32% a 54%.

Este artigo está organizado em cinco seções, além desta introdução. Na seção 2, são apresentados os trabalhos mais relacionados a esta pesquisa. A seção 3 descreve brevemente o processo de construção do corpus e as pistas linguísticas propostas por Silva et al. (2021) e adotadas neste artigo para definir uma boa qualidade de argumentação em um *tweet* do domínio da política brasileira. Na seção 4, apresentam-se os experimentos realizados para avaliar a qualidade da argumentação. Os resultados de tais experimentos são descritos e analisados na seção 5. Por fim, na seção 6, são feitas algumas considerações finais, com a apresentação de limitações e apontamentos sobre trabalhos futuros.

2 Trabalhos relacionados

Avaliar a validade, a qualidade e a força dos argumentos representa um desafio inerente ao discurso argumentativo. Vale destacar que existem fundamentos teóricos e diversas teorias normativas para embasar a tarefa, tais como (i) o modelo argumentativo de Toulmin (2003); (ii) os esquemas e questões críticas de Walton & Walton (1989); (iii) o modelo ideal de argumentação crítica na abordagem pragma-dialética (Van Eemeren & Grootendorst, 1987), em que as falácias são consideradas movimentos incorretos em uma discussão cujo objetivo é a resolução bem sucedida de uma disputa; e (iv) o estudo das falácias (Boudry et al., 2015). No entanto, julgar os critérios qualitativos da argumentação cotidiana ainda representa um desafio para os estudiosos e profissionais da argumentação (Weltzer-Ward et al., 2009; Swanson et al., 2015; Rosenfeld & Kraus, 2016).

Apesar desses fundamentos teóricos, os métodos e técnicas já propostos para avaliar a qualidade dos argumentos não concordam sobre quais critérios devem ser considerados, nem mesmo sobre se a qualidade deve ser avaliada do ponto de vista teórico ou prático. Wachsmuth et al. (2017a) tentaram elucidar a questão de quão diferentes são as visões teóricas e práticas da qualidade da argumentação. Do ponto de vista teórico, apontam que a convicção é entendida

como a principal qualidade lógica, e sustentam o fato de que a avaliação teórica da qualidade da argumentação permanece complexa. Eles também apontaram que as abordagens práticas indicam o que focar para simplificar a teoria, enquanto a teoria parece benéfica para orientar a avaliação da qualidade na prática.

Na mesma direção, outros estudos têm buscado avaliar a relevância dos argumentos por meio da identificação de sentenças argumentativas com a posterior avaliação da importância/relevância delas. Potthast et al. (2019) avaliaram o grau de relevância de um conjunto de argumentos no corpus args.me⁶ (Wachsmuth et al., 2017c), constituído por textos escritos em inglês, provenientes de cinco portais de debates, com o objetivo de construir um motor de busca de argumentos na web. Quarenta anotadores avaliaram a relevância de cada um dos 437 argumentos relacionados a 40 tópicos selecionados, além de sua qualidade retórica, lógica e dialética. Dos 437 argumentos anotados, 208 foram marcados a favor e 195 foram marcados como contrários ao tópico em questão, além de 34 que foram anotados como não argumentativos. Pontuações de 1 (baixa) a 4 (alta) foram atribuídas às dimensões de qualidade de argumentação – retórica (*rethoric*), lógica (*logic*) e dialética (*dialectic*) – e à sua relevância (*relevance*). A distribuição das pontuações (de 1 a 4) produzidas nesse trabalho pode ser vista na Figura 1.

As pontuações de relevância indicam que muitos argumentos altamente relevantes (classificados como 4) foram recuperados do corpus args.me. Outros trabalhos também investigaram o aspecto da relevância dos textos argumentativos (Wachsmuth et al., 2017d; Gleize et al., 2019).

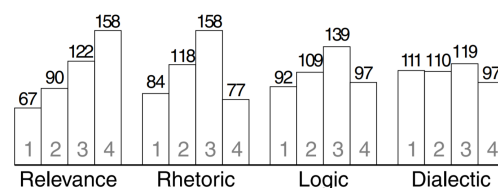


Figura 1: Distribuições de pontuação por dimensões de relevância e qualidade (Potthast et al., 2019).

Habernal & Gurevych (2016) sugeriram que a avaliação da qualidade da argumentação deve ser feita comparando argumentos, enquanto outros trabalhos (Persing & Ng, 2015; Wachsmuth

⁶Disponível em: <https://www.args.me>

et al., 2017b) relataram avaliações da qualidade dos argumentos individuais com resultados satisfatórios. Vale mencionar que, neste trabalho, a qualidade de argumentação é avaliada para um *tweet* isoladamente e não com base na comparação entre *tweets*.

Trabalhos mais recentes (Wachsmuth et al., 2017b; Lauscher et al., 2020; Wachsmuth & Werner, 2020; Skitalinskaya et al., 2021) têm utilizado uma taxonomia que visa avaliar aspectos individuais com base nas características da estrutura argumentativa, como o apelo emocional empregado, a organização da sentença e a credibilidade do autor da mensagem. A taxonomia escolhida para ser utilizada neste trabalho é apresentada na seção 2.3.

Embora passos importantes em AM tenham sido dados rumo ao processamento de *tweets*, como a detecção de argumentos, alegações e evidências, Schaefer & Stede (2021) demonstraram que várias áreas ainda estão em desenvolvimento, como o emprego de abordagens de AM utilizando aprendizado profundo e técnicas de redes neurais, além de pesquisas em outros idiomas diferentes do inglês.

Antes de detalhar a taxonomia adotada nesta pesquisa, as duas próximas subseções citam alguns trabalhos que investigaram a qualidade da argumentação em vários domínios, mais usuais do que o da política, aqui investigado.

2.1 Avaliação da qualidade da argumentação em redações de alunos

Fornecer aos estudantes *feedback* útil a respeito da persuasividade de seus argumentos tem sido objeto de estudos recentes em mineração de argumentos. Alguns corpora foram construídos, principalmente em inglês (Stab & Gurevych, 2014; Persing & Ng, 2015; Stab & Gurevych, 2017a; Carlile et al., 2018; Putra et al., 2021), e estudos têm sido realizados para avaliar a qualidade da argumentação em redações escritas por estudantes.

Stab & Gurevych (2017b) apontaram que as premissas de um argumento bem fundamentado devem fornecer evidências suficientes para aceitar ou rejeitar sua afirmação. Para chegar a essa conclusão, um corpus composto por 402 redações (Stab & Gurevych, 2017a) foi anotado com estruturas de argumentação. Os autores relataram uma pontuação de concordância de Fleiss entre os anotadores de 0,877, o que indicou que os anotadores foram capazes de identificar de forma confiável as principais reivindicações presentes em redações persuasivas. Para a construção do

modelo computacional, os autores experimentaram máquinas de vetores de suporte (SVMs) e redes neurais convolucionais (CNNs) e alcançaram 84% de precisão na tarefa de identificar argumentos insuficientemente suportados. O corpus final e as diretrizes de anotação estão disponíveis⁷.

Carlile et al. (2018), com o objetivo de realizar futuras avaliações no nível de persuasão dos argumentos em redações de alunos, construíram um corpus composto por 102 redações selecionadas aleatoriamente do *Argument Annotated Essays* (Stab & Gurevych, 2014). O corpus foi anotado com árvores de argumentos, pontuações de persuasão e atributos de componentes de argumentos que, segundo os autores, impactam essas pontuações. Os autores relatam uma concordância que variou de 0,549 a 1,000 (valores de α de Krippendorff (2011)), de acordo com o atributo. O corpus anotado e as diretrizes de anotação estão disponíveis⁸. Na mesma linha de pesquisa, outros trabalhos avaliaram a qualidade da argumentação em teses e trabalhos acadêmicos (García-Gorrostieta et al., 2018; García-Gorrostieta & López-López, 2018).

Putra et al. (2021) criaram um corpus com anotação estrutural argumentativa para redações escritas em inglês como língua estrangeira e também definiram um esquema de anotações. O corpus anotado produzido como resultado desse esforço, o ICNALE-AS, inclui 434 redações enviadas por estudantes de inglês de várias nações asiáticas. A análise de concordância inter-anotador mostrou que o esquema de anotação proposto é estável, alcançando um coeficiente de Cohen de 0,66.

Apesar dos trabalhos apresentados nesta seção terem avaliado a qualidade da argumentação, eles o fizeram para um gênero textual formal e menos desafiador do que o Twitter. Os textos formais admitem forma e estilo e, como consequência, podemos presumir que apresentam uma boa argumentação com argumentos encadeados baseados em fatos reais e frases estruturadas. Todos esses aspectos podem não ocorrer em argumentos feitos em postagens no Twitter; e, mesmo quando ocorrem, devem ser ressignificados de acordo com esse gênero e com a maneira como os usuários das mídias sociais utilizam esses aspectos, o que resulta, por exemplo, em uma possível argumentação baseada em *fake news*.

⁷Disponível em: <https://www.ukp.tu-darmstadt.de/data/>

⁸Disponível em <http://www.hlt.utdallas.edu/~zixuan/EssayScoring>

2.2 Avaliação da qualidade da argumentação em mensagens de fóruns e portais de discussão

Para investigar a identificação de postagens persuasivas em fóruns de discussão (como Change My View⁹), Wei et al. (2016) criaram um corpus de mensagens argumentativas selecionando tópicos com mais de 100 comentários publicados entre janeiro de 2014 e janeiro de 2015, totalizando 1.785 tópicos com 374.472 comentários.

Experimentos foram realizados para encontrar quais *features* seriam as mais adequadas para prever comentários persuasivos. Entre as *features* investigadas estavam: (i) o número de palavras e frases, (ii) a presença/ausência de pontuação, (iii) as *part-of-speech tags* (POS tags), entre outras. Depois disso, os autores calcularam a correlação entre a pontuação humana de um comentário argumentativo e o conjunto de atributos da reputação do autor da postagem. Então, para a tarefa de classificação de comentários, três conjuntos de *features* foram avaliados, incluindo as que consideravam apenas as características superficiais do texto, aquelas que consideravam a interação social e aquelas focadas na argumentação propriamente dita. Os resultados experimentais mostraram que as *features* baseadas em argumentação são mais informativas no estágio inicial da discussão e que a eficácia das *features* de interação social aumenta com o número de comentários na discussão.

Habernal & Gurevych (2016) investigaram a comparação qualitativa entre pares de argumentos: dados dois argumentos (como mostrado na Figura 2), um deles deve ser selecionado como o mais convincente. A pesquisa produziu os seguintes resultados: (i) um corpus anotado composto por 16.000 pares de argumentos, escritos em inglês, (ii) a análise dos dados anotados em relação às propriedades definidas como convincentes, e (iii) modelos computacionais gerados com SVM e uma arquitetura neural bidirecional de memória de longo prazo (BLSTM). O modelo SVM superou o modelo BLSTM (78% versus 76% de precisão, respectivamente) com uma diferença sutil, mas significativa, segundo os autores. Os dados anotados e os códigos estão disponíveis publicamente¹⁰.

Wachsmuth & Werner (2020) investigaram a avaliação automática de argumentos extraídos de portais de debate usando a taxonomia proposta por Wachsmuth et al. (2017b). Para tanto, com-

⁹Disponível em: <https://www.reddit.com/r/changemyview>

¹⁰Disponível em: <https://github.com/UKPLab/ac12016-convincing-arguments>

Argument 1	Argument 2
physical education should be mandatory cuz 112,000 people have died in the year 2011 so far and it's because of the lack of physical activity and people are becoming obese!!!!	YES, because some children don't understand anything except physical education especially rich children of rich parents.

Figura 2: Exemplo de argumentos sobre a obrigatoriedade da educação física (Habernal & Gurevych, 2016).

binaram *features* textuais (como mostrado nos exemplos da Figura 3) e SVM. Os exemplos na Figura 3 demonstram algumas *features* textuais¹¹ que, segundo os autores, podem ser preditivas de certas dimensões. Os autores relataram que o tamanho limitado do corpus dificultou a adoção de recursos mais complexos para a avaliação da qualidade. No entanto, eles destacaram que modelar a subjetividade por meio de *features* textuais pode ser propício para avaliar as dimensões lógica e dialética. Quanto à dimensão retórica, apontaram três aspectos difíceis: clareza, credibilidade e apelo emocional.

Estudos mais recentes (Fromm et al., 2022; Skitalinskaya et al., 2021; Toledo et al., 2019) usaram modelos BERT (Devlin et al., 2019) com ajuste fino (*fine-tuning*) em diferentes conjuntos de dados para avaliar a qualidade da argumentação. Skitalinskaya et al. (2021) investigaram a avaliação da qualidade da argumentação, independentemente dos aspectos discutidos. Para tanto, um corpus de 377.000 pares de comentário e argumento foi gerado a partir do fórum de discussão [kialo.com](https://www.kialo.com), abrangendo diversos temas de política, ética, entretenimento e outros. Duas tarefas foram realizadas: (i) avaliar qual afirmação de um par de comentários é melhor e (ii) classificar todas as versões de uma afirmação por qualidade. Os experimentos com regressão logística baseada em *embeddings* e redes neurais baseadas em *transformers* mostraram resultados promissores, sugerindo que os indicadores aprendidos generalizam bem entre os tópicos.

Para a primeira tarefa, os experimentos conduzidos com Sentence-BERT (SBERT) (Reimers & Gurevych, 2019) apresentaram uma precisão de até 77,7%. Na segunda tarefa, o modelo baseado em SBERT superou todas as abordagens

¹¹As *features* textuais apresentadas na pesquisa de Wachsmuth & Werner (2020) são semelhantes às pistas linguísticas (seção 3.1) estudadas neste trabalho.

Argument pro “advancing the common good”		Quality scores			
key phrases	<i>While <u>striving to make advancements</u> for the common good <u>you can change the world forever</u>.</i> — premise	Cog 2.00	Eff 2.00		
spelling errors	<i><u>Allot</u> of people have <u>succeded</u> in doing so.</i> — premise	LAc 2.67	Cla 2.33		
pronoun usage	<i><u>Our founding fathers, Thomas Edison, George Washington, Martin Luther King jr, and many more.</u></i> — premise	LRe 3.00	Cre 2.00		
	<i><u>These people made huge advances</u> for the common good and <u>they are honored</u> for it.</i> — conclusion	LSu 1.67	App 2.33	Emo 2.00	Arr 2.00
		Rea 2.00	GAc 2.67		
		GRe 2.33	GSu 1.33		
				OvQ 2.00	
	4 sentences, 60 tokens, 15 tokens / sentence				

Figura 3: Exemplo de um argumento e os recursos linguísticos que afetam sua qualidade (Wachsmuth & Werner, 2020).

testadas alcançando até 0,73 em correlação de Pearson e 0,72 em correlação de Spearman.

Esses trabalhos investigaram a avaliação da argumentação em mensagens de fóruns de discussão e portais de debate (Wei et al., 2016; Habernal & Gurevych, 2016) e redações de alunos (Stab & Gurevych, 2017b; Carlile et al., 2018; Wachsmuth et al., 2016). No entanto, os *tweets* no domínio da política brasileira, nos dias atuais, possuem características mais desafiadoras do que as encontradas em fóruns e portais de debate, como: (i) um número muito limitado de caracteres, o que dificulta o uso de estratégias de argumentação linguística; e (ii) a presença de discurso incivil e intolerante (Rossini, 2019, 2020), decorrente da polarização e agressividade presentes no cenário atual da política brasileira, que traz a necessidade de estratégias para identificar discurso de ódio e polaridade, por exemplo.

2.3 Taxonomia de Wachsmuth et al. (2017b)

Wachsmuth et al. (2017b) conduziram uma pesquisa sobre a qualidade da argumentação considerando tanto a teoria da argumentação quanto as perspectivas de mineração de argumentos. Com base nesse estudo, foi proposta a Taxonomia da Qualidade da Argumentação, cujas dimensões são utilizadas para definir a “qualidade”. A Figura 4 ilustra esta taxonomia, com todas as dimensões dela.

De acordo com essa taxonomia, a qualidade da argumentação pode ser dividida nas dimensões lógica, retórica e dialética (Blair, 2012), descritas a seguir.

- A **dimensão lógica** refere-se à estrutura e composição de um argumento. Um argumento de alta qualidade lógica é baseado em premissas aceitáveis e as combina de forma convincente para apoiar a afirmação do argumento. Está relacionado com a irrefutabilidade lógica do argumento.

- A **dimensão retórica**, ao contrário, inclui noções de eficácia persuasiva, linguagem correta, precisão e estilo. Um argumento de alta qualidade retórica é bem escrito e atraente para o público e está relacionado à eficácia retórica do argumento. Especificamente, um argumento é retoricamente eficaz se for capaz de convencer o público-alvo (ou corroborar a concordância) que a posição do autor sobre o assunto é a correta.

- A **dimensão dialética** captura a contribuição de um argumento para o discurso. Um argumento de alta qualidade dialética é útil para apoiar a tomada de decisão cooperativa ou para resolver conflitos. O argumento é razoável se for capaz de contribuir para a resolução do problema de uma forma que seja suficientemente aceitável pelo público-alvo.

Wachsmuth et al. (2017b) testaram a taxonomia em um experimento de anotação com o Dagstuhl-15512-ArgQuality¹², que contém 320 textos argumentativos com notas atribuídas por três anotadores que compõem os 15 aspectos da taxonomia. Nesse processo de anotação, cada texto foi primeiramente classificado como argumentativo ou não. Em seguida, para os textos argumentativos, todos os aspectos foram avaliados com notas 1 (baixo), 2 (médio) ou 3 (alto), além da opção “não posso julgar”.

Na Figura 5, pode-se ver as pontuações atribuídas pelos três anotadores (A, B e C) em dois textos produzidos em resposta à pergunta “garrafas plásticas de água devem ser banidas?”. O valor mais alto em cada coluna está marcado em negrito. A linha inferior representa a maioria

¹²Corpus Dagstuhl-15512-ArgQuality disponível em: <http://arguana.com/>

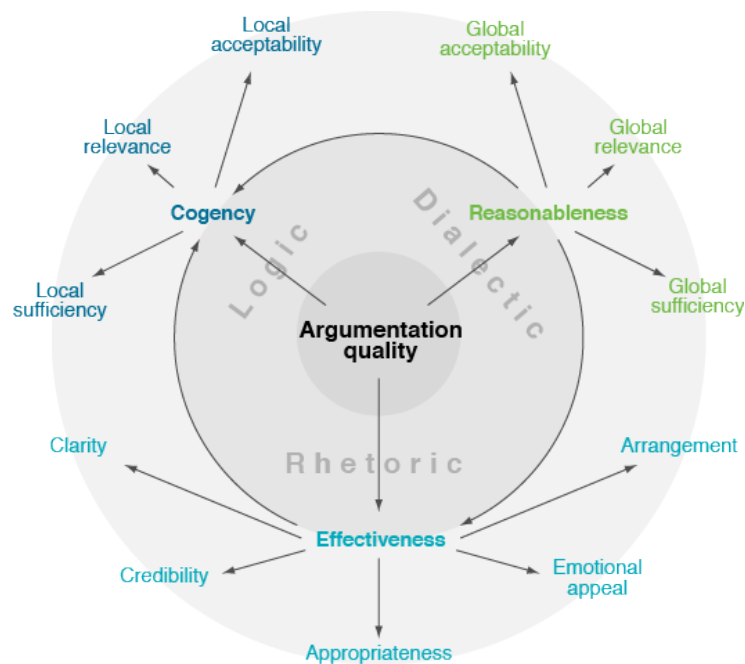


Figura 4: Taxonomia da Qualidade da Argumentação de Wachsmuth et al. (2017b).

dos votos dos três anotadores¹³.

A Figura 6 mostra os resultados deste experimento de anotação para os 304 textos do corpus classificados como argumentativos por todos os anotadores: (a) a distribuição das pontuações majoritárias para cada dimensão; (b) o α de Krippendorff (2011) utilizado para medir a concordância entre anotadores; (c) a correlação para cada par de dimensões, calculada com base na média das correlações de todos os anotadores. O valor mais alto em cada coluna é destacado em negrito.

Nesta pesquisa foi escolhida a dimensão retórica da taxonomia de Wachsmuth et al. (2017b) para a avaliação da qualidade da argumentação em postagens do Twitter no domínio da política no Brasil. Assim como Gretz et al. (2019), foram selecionados os aspectos da taxonomia da qualidade da argumentação que possuíam características semelhantes às *features* linguísticas disponíveis. Desse modo, a escolha pela dimensão retórica se deu com base nas pis-

tas linguísticas apontadas como relevantes pelos anotadores. Segundo Wachsmuth et al. (2017b), os aspectos que constituem a dimensão retórica estão relacionados ao apelo emocional aplicado na argumentação, ambiguidade, imprecisão, estilo de linguagem e organização da estrutura do texto. Portanto, entende-se que essas características podem ser, em certa medida, identificadas por meio de recursos linguísticos superficiais.

A dimensão retórica, segundo Wachsmuth et al. (2017b), possui cinco aspectos:

1. **Credibilidade (Cr)** – Credibilidade refere-se a como o autor transmite seus argumentos e os torna críveis. Segundo Wachsmuth et al. (2017b), um estilo apropriado em termos de escolha de palavras suporta a credibilidade. Além disso, de acordo com esses autores, aspectos que podem ser considerados para avaliar a credibilidade são: a honestidade do autor da mensagem, a polidez da linguagem utilizada ou o conhecimento e experiência do autor sobre os assuntos discutidos.
2. **Apelo emocional (Em)** – O apelo emocional é considerado bem-sucedido em um argumento se ele cria emoções de tal forma que torna o público-alvo mais receptivo aos argumentos do autor.
3. **Clareza (Cl)** – Clareza refere-se ao uso de uma linguagem gramaticalmente correta e

¹³A dimensão lógica mede a convicção (Co) e é composta por 3 aspectos: aceitabilidade local (LA), relevância local (LR) e suficiência local (LS). A dimensão retórica mede a eficácia (Ef) e é composta por 5 aspectos: credibilidade (Cr), apelo emocional (Em), clareza (Cl), adequação (Ap) e organização (Ar). Por fim, a dimensão dialética mede a razoabilidade (Re) e é composta por 3 aspectos: aceitabilidade global (GA), relevância global (GR) e suficiência global (GS).

Arguments	Pro Water bottles, good or bad? Many people believe plastic water bottles to be good. But the truth is water bottles are polluting land and unnecessary. Plastic water bottles should only be used in emergency purposes only. The water in those plastic are only filtered tap water. In an emergency situation like Katrina no one had access to tap water. In a situation like this water bottles are good because it provides the people in need. Other than that water bottles should not be legal because it pollutes the land and big companies get 1000% of the profit.													Con Americans spend billions on bottled water every year. Banning their sale would greatly hurt an already struggling economy. In addition to the actual sale of water bottles, the plastics that they are made out of, and the advertising on both the bottles and packaging are also big business. In addition to this, compostable waters bottle are also coming onto the market, these can be used instead of plastics to eliminate that detriment. Moreover, bottled water not only has a cleaner safety record than municipal water, but it easier to trace when a potential health risk does occur. (http://www.friendsjournal.org/bottled-water) (http://www.cdc.gov/healthywater/drinking/bottled/)																
	Scores	Co	LA	LR	LS	Ef	Cr	Em	Cl	Ap	Ar	Re	GA	GR	GS	Ov	Co	LA	LR	LS	Ef	Cr	Em	Cl	Ap	Ar	Re	GA	GR	GS
Annotator A	3	3	3	2	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
Annotator B	2	2	3	2	1	2	2	2	2	1	2	2	2	1	2	2	3	2	3	3	2	3	3	2	2	2	2	2	3	
Annotator C	2	3	3	2	2	2	2	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
Majority score	2	3	3	2	2	2	2	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	

Figura 5: Pontuações de cada anotador e pontuação majoritária para todas as dimensões de qualidade. Os argumentos são sobre o tópico “banir garrafas plásticas de água” (Wachsmuth et al., 2017b).

Quality Dimension	(a) Maj. Scores			(b) Agreement			(c) Pearson Correlation Coefficients													
	1	2	3	α	full	maj.	Co	LA	LR	LS	Ef	Cr	Em	Cl	Ap	Ar	Re	GA	GR	GS
Co Cogency	150	131	23	.44	40.1%	91.8%	.64	.61	.84	.81	.46	.27	.41	.32	.55	.78	.64	.71	.70	
LA Local acceptability	84	169	51	.46	27.0%	90.8%	.64	.51	.53	.60	.54	.30	.40	.54	.46	.68	.75	.46	.45	
LR Local relevance	25	155	124	.47	32.6%	92.4%	.61	.51	.56	.56	.39	.27	.46	.35	.50	.62	.58	.68	.45	
LS Local sufficiency	172	119	13	.44	37.2%	92.8%	.84	.53	.56	.73	.39	.25	.37	.23	.51	.67	.51	.68	.74	
Ef Effectiveness	184	111	9	.45	42.1%	94.4%	.81	.60	.56	.73	.48	.31	.35	.34	.54	.75	.58	.66	.71	
Cr Credibility	99	199	6	.37	37.8%	95.7%	.46	.54	.39	.39	.48	.37	.32	.49	.37	.52	.52	.36	.40	
Em Emotional appeal	48	235	21	.26	42.8%	94.4%	.27	.30	.27	.25	.31	.37	.14	.30	.20	.30	.26	.26	.22	
Cl Clarity	42	191	71	.35	29.3%	89.8%	.41	.40	.46	.37	.35	.32	.14	.45	.56	.44	.45	.38	.27	
Ap Appropriateness	43	196	65	.36	17.4%	87.5%	.32	.54	.35	.23	.34	.49	.30	.45	.48	.47	.59	.20	.20	
Ar Arrangement	91	189	24	.39	26.6%	93.4%	.55	.46	.50	.51	.54	.37	.20	.56	.48	.55	.51	.49	.48	
Re Reasonableness	126	159	19	.50	41.4%	95.7%	.78	.68	.62	.67	.75	.52	.30	.44	.47	.55	.78	.65	.61	
GA Global acceptability	88	161	55	.44	31.6%	95.4%	.64	.75	.58	.51	.58	.52	.26	.45	.59	.51	.78	.46	.43	
GR Global relevance	69	167	68	.42	21.7%	90.1%	.71	.46	.68	.68	.66	.36	.26	.38	.20	.49	.65	.46	.61	
GS Global sufficiency	231	72	1	.27	44.7%	98.0%	.70	.45	.45	.74	.71	.40	.22	.27	.20	.48	.61	.43	.61	
Ov Overall quality	152	128	24	.51	44.1%	94.4%	.84	.66	.61	.74	.81	.52	.30	.45	.42	.59	.86	.71	.70	.68

Figura 6: Resultados para os 304 textos do corpus classificados como argumentativos por todos os anotadores (Wachsmuth et al., 2017b).

amplamente inequívoca e que evita complexidade desnecessária e desvio do assunto discutido. A linguagem utilizada deve facilitar a compreensão e não deixar dúvidas sobre a posição do autor e a forma como ele a defende.

- Adequação (Ap)** – A adequação de um argumento refere-se à linguagem (forma e conteúdo) utilizada para apoiar a criação de credibilidade e emoções, bem como a adequação ao assunto discutido.
- Organização (Ar)** – Uma argumentação é considerada adequadamente organizada se apresentar a pergunta, os argumentos e a conclusão na ordem correta.

É importante destacar que o corpus utilizado no estudo de Wachsmuth et al. (2017b) é composto por mensagens que têm características diferentes do corpus adotado nesta pesquisa. Pri-

meiramente, as mensagens de fóruns de discussão se caracterizam por serem mais longas do que as mensagens do Twitter, que têm um limite de 280 caracteres. O tamanho limitado das mensagens do Twitter pode impactar aspectos como Clareza e Organização. Outra diferença está relacionada ao contexto no qual as mensagens foram produzidas. No corpus de Wachsmuth et al. (2017b), as mensagens eram sobre tópicos gerais, às vezes controversos, mas sem a polarização política existente atualmente no Brasil, a qual pode impactar aspectos como Credibilidade, Apelo emocional e Adequação. Assim, embora a mesma taxonomia de Wachsmuth et al. (2017b) tenha sido adotada neste trabalho para embasar as medidas de qualidade da argumentação, o gênero textual e o domínio do corpus utilizado nesta pesquisa podem levar a resultados diferentes daqueles encontrados em (Wachsmuth et al., 2017b).

3 Corpus e anotação

Embora existam corpora abrangendo vários aspectos da análise argumentativa, alguns deles descritos na seção 2, até onde se sabe o corpus desenvolvido no projeto Arg Q!¹⁴, e descrito em (Silva et al., 2021), é o primeiro construído especificamente para a análise da qualidade da argumentação no domínio da política brasileira. Portanto, neste trabalho, foi utilizado o corpus construído e anotado conforme descrito por Silva et al. (2021). Este corpus é composto por *tweets* coletados como respostas a mensagens de parlamentares brasileiros postadas de 6 de março a 6 de abril de 2021¹⁵.

Após a coleta dos *tweets*, cerca de 400 deles foram anotados com base nas pistas linguísticas descritas na Seção 3.1, conforme detalhado na Seção 3.2. Ressalta-se que apenas dados públicos foram coletados do Twitter e, embora os usuários não sejam identificados nos *tweets* do corpus, não podemos garantir que não seja possível rastrear a identidade do autor da mensagem.

3.1 Pistas linguísticas

Para a anotação do corpus foram utilizadas 30 pistas linguísticas definidas em (Silva et al., 2021): 4 para o aspecto de Clareza, 7 para o aspecto de Organização, 6 para o aspecto de Credibilidade e 13 para o aspecto de Apelo emocional (6 para polaridade e 7 para intensidade). A Adequação não foi considerada em Silva et al. (2021) uma vez que se mostrou não relevante para a qualidade da argumentação em *tweets*, e também porque os anotadores apontaram que os critérios referentes à Adequação já eram contemplados pelos outros quatro aspectos. Portanto, o aspecto Adequação também foi desconsiderado neste trabalho. Nas seções seguintes, citamos as pistas linguísticas definidas para cada aspecto.

3.1.1 Clareza

Wachsmuth et al. (2017b) consideram um argumento claro se ele usa uma linguagem gramaticalmente correta e amplamente inequívoca e evita complexidade e desvios desnecessários da questão discutida. Além disso, a linguagem utilizada deve facilitar a compreensão e não deixar dúvidas so-

bre a posição do autor e a forma como ele a defende.

Nessa perspectiva, Silva et al. (2021) assumiram que todo argumento escrito em português tem o potencial de ser naturalmente claro, a menos que haja certos fenômenos linguísticos que interfiram negativamente na clareza. Dessa forma, a pontuação para o aspecto Clareza diminui com a presença de uma ou mais pistas linguísticas que prejudicam a clareza da argumentação, a saber: (i) questão que leva à dúvida, (ii) linguagem complexa desnecessária, (iii) presença de erros de língua portuguesa e (iv) desvio desnecessário do assunto principal. O aspecto Clareza foi classificado com base na quantidade (cardinalidade) de pistas identificadas (Pistas), como apresentado na equação (1).

$$\text{Clareza} = \begin{cases} \text{Baixa,} & \text{se } |Pistas| \geq 3 \\ \text{Média,} & \text{se } |Pistas| = 2 \\ \text{Alta,} & \text{caso contrário} \end{cases} \quad (1)$$

3.1.2 Organização

A definição de organização proposta por Wachsmuth et al. (2017b) considera que um texto deve ser composto de três partes, na seguinte ordem: (i) introdução do assunto, (ii) alguns argumentos para sustentar a conclusão e (iii) a conclusão. Essa definição, no entanto, não se aplica quando analisam-se textos do Twitter, pois os usuários não têm espaço suficiente para desenvolver essas 3 partes devido à limitação de caracteres máximos que uma mensagem pode conter (que é de 280). Assim, Silva et al. (2021) redefiniram a noção de organização para este gênero textual considerando a presença de pistas linguísticas que pontuam positivamente para a organização de um argumento.

As sete pistas linguísticas utilizadas para avaliar o aspecto Organização estão relacionadas à presença de relações linguísticas bem conhecidas: (i) relação condicional; (ii) relação de concessão; (iii) oposição ou contraste; (iv) comparação entre duas ideias; (v) relação de causa e efeito, explicação ou propósito; (vi) encadeamento cronológico ou enumeração; e (vii) exemplificação ou interligação lógica.

Diferentemente do aspecto da Clareza, que se ancora no pressuposto de que todo argumento no Twitter tem o potencial de ser inerentemente claro, no aspecto Organização, considera-se que esse gênero textual tem o potencial intrínseco de ser desorganizado, a menos que ocorram determinados fenômenos linguísticos que contribuam positivamente para a organização. Dessa forma,

¹⁴Disponível em: <https://argq.org/>

¹⁵Embora os *tweets* dos parlamentares tenham sido considerados como semente para recuperar as respostas dos seguidores, vale ressaltar que a avaliação da qualidade da argumentação foi feita apenas nas respostas dos seguidores.

o aspecto Organização foi classificado com base na quantidade (cardinalidade) de pistas identificadas (Pistas), como apresentado na equação (2).

$$\text{Organização} = \begin{cases} \text{Alta,} & \text{se } |Pistas| \geq 2 \\ \text{Média,} & \text{se } |Pistas| = 1 \\ \text{Baixa,} & \text{caso contrário} \end{cases} \quad (2)$$

3.1.3 Credibilidade

De acordo com Wachsmuth et al. (2017b), um argumento deve ser avaliado como bem-sucedido em criar credibilidade se transmitir argumentos e outras informações de uma forma que torne o autor crível, por exemplo, indicando a honestidade do autor, a polidez da linguagem utilizada ou revelando o conhecimento do autor ou a experiência em relação aos assuntos abordados.

Para a avaliação do aspecto Credibilidade, Silva et al. (2021) consideraram que um argumento escrito em português é verossímil se algumas pistas linguísticas estiverem presentes na superfície textual; ou seja, não levaram em consideração nenhum critério ou dado externo como a aceitação ou engajamento do autor nas redes sociais. Vale ressaltar que o Twitter é uma plataforma aberta e, portanto, qualquer usuário cadastrado pode postar mensagens diversas nesta plataforma de mídia social. Como não há pré-análise do perfil do usuário ou do conteúdo postado por ele, a dúvida sobre a credibilidade do que é postado é inerente à plataforma.

Assim, a pontuação do aspecto Credibilidade aumenta se as seguintes pistas estiverem presentes: (i) menção a uma data específica; (ii) menção a um fato midiático, histórico ou enciclopédico; (iii) menção a uma autoridade pública; (iv) presença de uma *hashtag* que reforça uma posição; (v) presença de um termo especializado; e (vi) um relato de alguma experiência pessoal ou individual. O aspecto Credibilidade foi classificado com base na quantidade (cardinalidade) de pistas identificadas (Pistas), como apresentado na equação (3).

$$\text{Credibilidade} = \begin{cases} \text{Alta,} & \text{se } |Pistas| \geq 3 \\ \text{Média,} & \text{se } |Pistas| = 2 \\ \text{Baixa,} & \text{caso contrário} \end{cases} \quad (3)$$

3.1.4 Apelo emocional

Segundo Wachsmuth et al. (2017b), um argumento tem apelo emocional quando cria emoções no interlocutor. Silva et al. (2021) decidiram di-

vidir as pistas linguísticas para o aspecto apelo emocional em polaridade e intensidade.

Polaridade A polaridade de um argumento é considerada positiva ou negativa se contribui para criar emoções boas ou ruins no leitor, respectivamente. Se o seu conteúdo não causar nenhuma emoção ou se o apelo emocional estiver bem equilibrado entre positivo e negativo, então a polaridade do argumento é considerada neutra.

Assim, as pistas linguísticas de Silva et al. (2021) para uma polaridade positiva são: (i) a presença de uma referência cordial a uma pessoa/organização e (ii) o uso de linguagem polida. Por outro lado, as pistas linguísticas para uma polaridade negativa são: (i) a presença de uma referência pejorativa a uma pessoa/organização; (ii) o uso de xingamento ou palavra de baixo calão; (iii) a presença de discurso de ódio ou ameaça; e (iv) o uso de expressões que denotam especulação. Não há nenhuma pista linguística específica para uma polaridade neutra, pois é o meio termo entre os dois extremos. A Polaridade do aspecto Apelo emocional foi classificada com base na diferença entre a quantidade de pistas positivas ($Pistas^+$) e negativas ($Pistas^-$), como apresentado na equação (4).

$$AE_{pol} = \begin{cases} \text{Negativa,} & \text{se } |Pistas^+| - |Pistas^-| < 0 \\ \text{Positiva,} & \text{se } |Pistas^+| - |Pistas^-| > 0 \\ \text{Neutra,} & \text{caso contrário} \end{cases} \quad (4)$$

Intensidade O apelo emocional também é avaliado de acordo com sua intensidade, que é determinada pela quantidade de pistas linguísticas que potencializam a criação de emoções. Quanto maior o número de pistas linguísticas, maior a intensidade do apelo emocional. A intensidade é avaliada de acordo com a presença de (i) pronome ou verbo na primeira pessoa; (ii) a repetição de sinais de pontuação; (iii) estrutura enfática (por exemplo, palavra em maiúscula); (iv) uma frase imperativa ou palavra de ordem; (v) uma expressão que denota exagero (por exemplo, “sempre”, “nunca”, “todo mundo”); (vi) linguagem não verbal (por exemplo, *emoticons*); e (vii) a presença de uma expressão idiomática, provérbio ou metáfora. A Intensidade do aspecto Apelo emocional foi classificada com base na quantidade de pistas de intensidade identificadas (Pistas), mas também em relação à quantidade de pistas positivas ($Pistas^+$) e negativas ($Pistas^-$) de polaridade identificadas, como apresentado na equação (5).

$$AE_{int} = \begin{cases} \text{Alta,} & \text{se } |Pistas^-| \geq 3 \\ & \text{ou } |Pistas^+| \geq 2 \\ & \text{ou } |Pistas| \geq 4 \\ \text{Média,} & \text{se } |Pistas^-| = 2 \\ & \text{ou } |Pistas^+| = 1 \\ & \text{ou } 2 \leq |Pistas| \leq 3 \\ \text{Baixa,} & \text{caso contrário} \end{cases} \quad (5)$$

3.1.5 Qualidade geral da argumentação

Após a anotação das pistas linguísticas associadas a cada aspecto, os anotadores em (Silva et al., 2021) definiram como eles seriam combinados para resultar em um valor para a Qualidade Geral da argumentação. Para tanto, o mapeamento de categorias para valores foi realizado como apresentado na tabela 1.

<i>f</i>		<i>g</i>	
Categoria	Valor	Categoria	Valor
Baixa	1	Negativa	-1
Média	2	Neutra	0
Alta	3	Positiva	1

Tabela 1: Funções *f* e *g* de mapeamento das categorias atribuídas pelos anotadores para valores.

A nota do Apelo emocional (N_{AE}) foi definida como:

```

if  $AE_{pol} = \text{Neutra}$  then
   $N_{AE} \leftarrow \frac{f(AE_{int})}{2}$ 
else
   $N_{AE} \leftarrow g(AE_{pol}) \times f(AE_{int})$ 
end if

```

E as notas dos demais aspectos foram definidas com base na aplicação da função *f*:

```

 $N_{Cla} = f(\text{Clareza})$ 
 $N_{Org} = f(\text{Organização})$ 
 $N_{Cred} = f(\text{Credibilidade})$ 

```

A partir das notas associadas a cada aspecto, a nota final foi dada pela soma das notas dos aspectos. Por fim, definiu-se que a nota para a qualidade geral da argumentação (N_{QG}) seria a média das notas dos anotadores e a categoria associada à Qualidade Geral foi definida como apresentado na equação (6).

$$QG = \begin{cases} \text{Alta,} & \text{se } N_{QG} \geq 7 \\ \text{Média,} & \text{se } 5 \leq N_{QG} < 7 \\ \text{Baixa,} & \text{caso contrário} \end{cases} \quad (6)$$

3.2 Anotação do corpus

Com base nas pistas linguísticas apresentadas anteriormente e seguindo as diretrizes estabelecidas em (Silva et al., 2021)¹⁶, 400 *tweets* foram analisados por quatro pesquisadores em linguística computacional (3 com formação em linguística e 1 em computação). Desses, 48 *tweets* foram descartados por não terem sido considerados argumentativos¹⁷ por todos os quatro anotadores. Assim, o corpus final é composto por 352 *tweets*.

Na Figura 7 é exibido um gráfico de distribuição de pontuação para cada aspecto dos 352 *tweets* argumentativos. A maioria deles tem alta Clareza e Organização, mas baixa Credibilidade. De fato, apenas 3% deles foram avaliados como de alta credibilidade. Com relação ao apelo emocional, 54% dos *tweets* foram avaliados como de polaridade negativa e 75% de média intensidade.

Na Tabela 2 (a) são exibidas as pontuações finais de cada aspecto para os 352 *tweets*. O grau de concordância entre anotadores foi calculado e é exibido na Tabela 2 (b) e expresso como o intervalo de α de Krippendorff (2011) (valor mais baixo - valor mais alto) que apresentaram os trios menos concordantes e mais concordantes de anotadores¹⁸, e tanto os acordos totais quanto os majoritários.

A “Concordância total” é alcançada quando todos os anotadores concordam com a mesma pontuação, e “maioria” indica que pelo menos três anotadores concordaram. A concordância total encontrada ficou entre 27,84% e 57,67%, e a concordância majoritária dos anotadores ficou entre 69,89% e 86,93%. Com exceção do aspecto Clareza, todos os aspectos tiveram valores máximos de concordância acima de 0,40, diferentemente de Wachsmuth et al. (2017b) (veja Figura 6) onde os valores de concordância α para todos os aspectos ficaram abaixo de 0,40.

4 Experimentos

Para responder as questões de pesquisa definidas para este trabalho, foram realizados experimentos utilizando algoritmos de AM baseados

¹⁶As diretrizes utilizadas para a anotação estão disponíveis no site do projeto Arg Q!: <https://argq.org/>

¹⁷Tradicionalmente, um texto é considerado argumentativo se contém argumentos organizados e estruturados em uma sequência lógica. No entanto, no que diz respeito ao gênero textual e domínio deste corpus, este conceito foi adaptado para abranger quaisquer *tweets* em que a posição/opinião do autor pudesse ser determinada.

¹⁸Decidimos relatar a concordância alcançada entre os trios para que nossos resultados pudessem ser comparados com os de Wachsmuth et al. (2017b) já que em seu trabalho havia três anotadores.

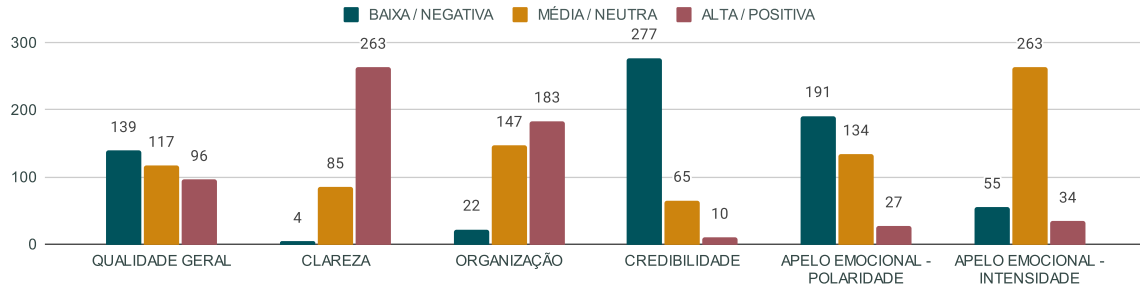


Figura 7: Distribuições de pontuação por aspectos de qualidade da argumentação no corpus de Silva et al. (2021) (tradução nossa).

Aspecto da qualidade	(a) Pontuação final			(b) Concordância		
	Baixo/ Negativo	Médio/ Neutro	Alto/ Positivo	α trios	total (4/4)	maioria (3-4/4)
Clareza	4	85	263	0,26 - 0,30	48,58%	79,26%
Organização	22	147	183	0,51 - 0,71	50,57%	82,67%
Credibilidade	277	65	10	0,36 - 0,48	57,67%	86,93%
Apelo emocional - Polaridade	191	134	27	0,60 - 0,66	51,99%	82,67%
Apelo emocional - Intensidade	55	263	34	0,48 - 0,55	40,63%	82,39%
Qualidade geral	139	117	96	0,50 - 0,54	27,84%	69,89%

Tabela 2: Pontuações para cada aspecto e a concordância entre os juízes humanos do corpus de Silva et al. (2021) (tradução nossa).

em *features* e, também, uma abordagem neural baseada em transformers (BERT e RoBERTa). Para a abordagem neural, foram utilizadas apenas os *tweets* presentes no corpus, como descrito na seção 4.2. Entretanto, para a avaliação com os algoritmos de AM tradicionais, foi necessária a construção de um conjunto de *features* linguísticas para a tarefa de classificação, como detalhado na seção 4.1.

4.1 Experimentos com AM baseado em *features*

Para realizar os experimentos com AM tradicional foi necessário definir as *features* linguístico-computacionais (como descrito na seção 4.1.1) e treinar os modelos computacionais como descrito na seção 4.1.2.

4.1.1 Definição das *features* computacionais

Com o objetivo de encontrar as *features* linguístico-computacionais que melhor se correlacionam com as pistas linguísticas utilizadas para medir a qualidade da argumentação, foi definido e gerado um conjunto de 337 *features*. Essas *features* foram categorizadas em 14 grupos, a maio-

ria delas geradas pelo NILC-Metrix¹⁹ (Leal et al., 2022):

1. **Medidas Psicolinguísticas** extraem características subjetivas do texto, tais como: imageabilidade, concretude, familiaridade e idade de aquisição. 44 *features* deste tipo foram geradas pelo NILC-Metrix.

Imageabilidade envolve a facilidade e rapidez de evocar uma imagem mental associada a uma palavra;

Concretude diz respeito ao grau em que uma palavra se refere a objetos, pessoas, lugares ou coisas que podem ser percebidas pelos sentidos;

Familiaridade é o grau em que as pessoas conhecem e usam palavras em suas vidas cotidianas;

Idade de Aquisição é uma estimativa da idade que a pessoa tinha quando uma palavra foi aprendida.

2. **Informação morfossintática** engloba *features* relacionadas à classe gramatical de uma palavra e a sua função sintática no texto, como

¹⁹NILC-Metrix (Leal et al., 2022) composto por 200 métricas linguísticas e psicolinguísticas utilizadas na avaliação de métodos de predição de complexidade textual. Disponível em: <http://fw.nilc.icmc.usp.br:22380/nilcmetrix>

- a proporção de adjetivos, advérbios, pronomes, substantivos, verbos e preposições, em relação à quantidade total de palavras no texto. Neste trabalho, para extrair essas *features* foram usados Apertium²⁰ (Armentano-Oller et al., 2006), NLPNet²¹ (Fonseca & Rosa, 2013) e NILCMetrix. Foram geradas 94 *features* deste tipo.
3. **Complexidade Sintática** está relacionada à dificuldade de processamento de alguns tipos de estruturas de frases sofisticadas, por exemplo, a proporção de orações subordinadas reduzidas pela quantidade de orações do texto ou a proporção de orações na voz passiva analítica em relação à quantidade de orações do texto. 37 *features* deste tipo foram geradas pelo NILC-Metrix.
 4. **Frequências de Palavras** mostram os valores das frequências absolutas e relativas das palavras no texto. 20 *features* deste tipo foram geradas pelo NILC-Metrix.
 5. **Conectivos** incluem métricas relacionadas à quantidade de conectivos, operadores lógicos ou palavras que denotam negação em relação às palavras do texto. 18 *features* deste tipo foram geradas pelo NILC-Metrix.
 6. **Simplicidade Textual** fornece métricas que medem o nível de complexidade em um texto com relação à dificuldade de compreensão de leitura. 9 *features* deste tipo foram geradas pelo NILC-Metrix.
 7. **Índices de Legibilidade** incluem métricas que medem a legibilidade de um texto, como o índice Brunet e Flesch. 10 *features* deste tipo foram geradas pelo NILC-Metrix.
 8. **Indicadores Temporais de um Léxico** incluem índices relacionados à diversidade de tempos verbais que ocorrem no texto. 14 *features* deste tipo foram geradas pelo NILC-Metrix.
 9. **Informações Semânticas** referem-se a várias métricas que fornecem informações sobre o significado das palavras no texto, como a quantidade média de hiperônimos por verbo nas sentenças, ou a proporção de substantivos abstratos em relação à quantidade de palavras do texto. 18 *features* deste tipo foram geradas pelo NILC-Metrix.
 10. **Medidas Descritivas** referem-se a métricas como a quantidade de parágrafos, frases e palavras em um texto e de sílabas por palavra. 16 *features* deste tipo foram geradas pelo NILC-Metrix.
 11. **Coesão Semântica** é expressa por métricas que calculam as relações semânticas entre palavras em um texto, por exemplo a média de similaridade entre os pares de sentenças no texto, ou a média da entropia cruzadas das sentenças do texto. 19 *features* deste tipo foram geradas pelo NILC-Metrix.
 12. **Polaridade do sentimento** mede a frequência de palavras com emoções positivas, negativas e neutras no texto. Neste trabalho, realizamos análise de polaridade com base em um modelo treinado usando o algoritmo de Floresta Aleatória e o corpus TweetSentBR²² (Brum & Nunes, 2018) e também um modelo baseado em BERT de (Capellaro & Caseli, 2021). Foram geradas 4 *features* deste tipo.
 13. **Linguagem Tóxica** indica se há ou não linguagem tóxica no texto. 7 *features* desse tipo foram geradas com um modelo baseado em BERT treinado no corpus ToLD-Br²³ (Leite et al., 2020), todas elas *features* binárias: não tóxico, LGBTQ+fobia, obsceno, insultuoso, racismo, misoginia e xenofobia.
 14. A frequência de uso para **diferentes categorias de palavras** gerado com base na versão em português do LIWC²⁴ (Balage Filho et al., 2013). Foram geradas 62 *features* deste tipo.

4.1.2 Treinamento dos modelos de AM baseado em features

Este processo consiste em três etapas principais: (i) o pré-processamento, criação do conjunto de *features* e seleção das melhores *features*; (ii) ajuste de hiperparâmetros com validação cruzada aninhada; e (iii) melhor seleção de modelo e geração de métricas. Esta configuração experimental é ilustrada na Figura 8.

Para realizar o **primeiro passo** (i), cada *tweet* foi pré-processado pelo Enelvo²⁵ para remover informações desnecessárias (como o identificador do usuário do Twitter ao qual a mensagem se referia como resposta) e normalizar palavras ruidosas em conteúdo gerado pelo usuário (por exemplo, substituir abreviações como “vc” por sua versão utilizada na norma culta “você”). Em seguida, os *tweets* normalizados foram processados pelas ferramentas linguístico-computacionais a fim de extrair as *features* descritas na seção 4.1.1.

Após a geração das 337 *features*, foi aplicado um método de seleção de *features* para determinar quais delas se correlacionam melhor com a qualidade geral da argumentação. O objetivo da seleção de *features* é encontrar aquelas que possivelmente são as melhores preditoras para tarefas de AM. Esse procedimento é importante, pois al-

²⁰Disponível em: <https://www.apertium.org/>

²¹Disponível em: <http://nilc.icmc.usp.br/nlpnet/>

²²Disponível em: <https://bitbucket.org/HBrum/tweetsentbr/src/master/>

²³Disponível em: <https://github.com/JAugusto97/ToLD-Br>

²⁴Disponível em: <http://143.107.183.175:21380/portlex/index.php/en/liwc>

²⁵Disponível em: <https://github.com/thalesbertaglia/enelvo>

gumas *features* podem ser irrelevantes para a tarefa de AM e, desse modo, adicionar ruído ao modelo treinado. [Adi et al. \(2019\)](#) apontam que entre os problemas causados por conjuntos de dados de alta dimensão estão o alto custo computacional e a baixa precisão do modelo gerado. Para contornar esses problemas, sugere-se selecionar apenas as *features* mais relevantes que tenham uma alta correlação com a classe.

Alguns métodos de seleção de *features* podem ser aplicados para reduzir a alta dimensionalidade do conjunto de *features*, como Eliminação Recursiva de *Features* com Validação Cruzada (RFECV) ([Misra & Yadav, 2020](#)), Ganho de Informação (IG), Taxa de Ganho de Informação (Taxa de Ganho) ([Adi et al., 2019](#)) e Análise de Componentes Principais (PCA) ([Maćkiewicz & Ratajczak, 1993](#)). Neste trabalho, o método escolhido para a fase de seleção de *features* foi o RFECV dado que a literatura aponta a efetividade dele na eliminação de *features* irrelevantes por se tratar de um método que busca evitar o problema de *overfitting* com Eliminação Recursiva de *Features* (RFE), aplicando a validação cruzada estratificada ([Misra & Yadav, 2020](#)). O RFECV classifica as *features* com eliminação recursiva de *features* e validação cruzada de 10 vezes e, assim, seleciona o número ideal de *features* para construção de modelo.

Assim, as melhores *features* foram selecionadas com o auxílio da biblioteca scikit-learn usando RFE com Classificador de Floresta Aleatória e validação cruzada de 10 vezes. O conjunto final de *features* selecionadas para prever a qualidade geral da argumentação contém 290 *features*. As 10 *features* mais significativas para a qualidade geral, de acordo com essa seleção automática, são apresentadas na Tabela 3. Contudo, como alguns dos algoritmos investigados nesta pesquisa são conhecidos por lidarem bem com um número elevado de *features*, como o SVM, também foram realizados experimentos sem esta etapa de seleção de *features*.

A **segunda etapa** (ii) consiste em refinar os parâmetros dos algoritmos de AM. Para isso, foi utilizado o GridSearchCV²⁶, uma pesquisa exaustiva sobre valores de parâmetros especificados para um estimador. Conforme mostrado na Figura 8, as dobras de treinamento do laço externo são usadas no laço interno para ajustar os hiperparâmetros. O laço interno seleciona a melhor configuração de hiperparâmetros.

Cinco dobras estratificadas foram utilizadas

²⁶Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

em ambas os laços. O laço interno faz uma pesquisa de grade no espaço de hiperparâmetros que é validado de forma cruzada em relação aos conjuntos de treinamento e validação adquiridos pelo laço externo. A configuração do hiperparâmetro que maximiza a pontuação de precisão é retornada para cada pesquisa de grade. A generalização da configuração do modelo selecionado é então validada usando as métricas padrão de acurácia, precisão, cobertura e medida F nos conjuntos de teste criados pelo laço externo.

A **terceira etapa** (iii) consiste em gerar os melhores modelos de AM para cada algoritmo testado e classificar os *tweets* de teste usando os modelos treinados. Testamos seis modelos classificadores em nosso conjunto de dados: Regressão Logística (*Logistic Regression*, LR), *K-Nearest Neighbors* (KNN), Árvores de Decisão (*Decision Tree*, DT), Máquinas de Vetores de Suporte (*Support Vector Machines*, SVM), Floresta Aleatória (*Random Forest*, RF) e *Naive Bayes* (NB)²⁷.

4.2 Experimentos com BERTimbau e RobertaTwitterBR

Além dos experimentos descritos utilizando os algoritmos de AM baseado em *features*, foram realizados experimentos com o BERTimbau²⁸ ([Souza et al., 2020](#)) e o RobertaTwitterBR²⁹, modelos neurais treinados em português. O BERTimbau foi treinado no BrWaC (*Brazilian Web as Corpus*), um grande corpus em português, por 1 milhão de passos, usando *full word mask*. O RobertaTwitterBR foi treinado com aproximadamente 7 milhões de *tweets* em português.

Nestes experimentos, somente o texto das mensagens foi utilizado, ao contrário dos experimentos com os algoritmos de AM tradicional que utilizaram as *features* linguísticas. O único processamento realizado com as mensagens para estes experimentos foi a remoção dos nomes dos usuários. Um script Python foi usado para realizar essa tarefa.

A Figura 9 ilustra esse processo que foi reali-

²⁷Os hiperparâmetros e respectivos valores explorados foram os seguintes: LR: *penalty=12*, *C=np.power(10., np.arange(-4, 4))*, KNN: *n_neighbors=list(range(1, 10))*, *p=[1, 2]*, DT: *max_depth=list(range(1, 10))*, *criterion=[gini, entropy]*, SVM: *kernel=rbf*, *C=np.power(10., np.arange(-4, 4))*, *gamma=np.power(10., np.arange(-5, 0))*; *kernel=linear*, *C=np.power(10., np.arange(-4, 4))*, RF: *n_estimators=[10, 100, 500, 1000, 10000]*, *criterion=[gini, entropy]*, *max_depth=[10, 11, 12, 13, 14]*, NB: *var_smoothing= np.logspace(0, -9, num=100)*.

²⁸Disponível em: <https://github.com/neuralmind-ai/portuguese-bert>

²⁹<https://huggingface.co/verissimomanoel/RobertaTwitterBR>

ID	Grupo	Métrica	Descrição
1	Linguagem tóxica	toxclang	Identifica a presença de linguagem tóxica no texto
2	NILCMetrix-Medidas Psicolinguísticas	idade_de_aquisição	Valores de idade média de aquisição de palavras de conteúdo de texto
3	NILCMetrix-Frequência de palavras	freq_bra	Média dos valores de frequência das palavras no texto na escala logarítmica Zipf via Corpus Brasileiro
4	NILCMetrix-Medidas Psicolinguísticas	imageabilidade_25.4_ratio	Proporção de palavras com valor de imageabilidade entre 2,5 e 4 em relação a todas as palavras de conteúdo do texto
5	NILCMetrix-Medidas psicolinguísticas	imageabilidade_mean	Imageabilidade média das palavras de conteúdo no texto
6	NILCMetrix - Medidas Descritivas	syllabes_per_content_word	Número médio de sílabas por palavra de conteúdo no texto
7	TweetSentBR- Sentimento neutro	sent_neu	Proporção de palavras com emoção neutra em relação a todas as palavras do texto
8	NILCMetrix-Frequência de palavras	freq_brwac	Média dos valores das frequências das palavras do texto na escala logarítmica Zipf via BrWac
9	NILCMetrix-Medidas Psicolinguísticas	imageabilidade_4.55_ratio	Proporção de palavras com valor de imageabilidade entre 4 e 5,5 em relação a todas as palavras de conteúdo do texto
10	NILCMetrix-Medidas descritivas	sentence_length_standard_deviation	Desvio padrão do número de palavras por frase

Tabela 3: Top-10 *features* com melhor correlação com a qualidade geral da argumentação.

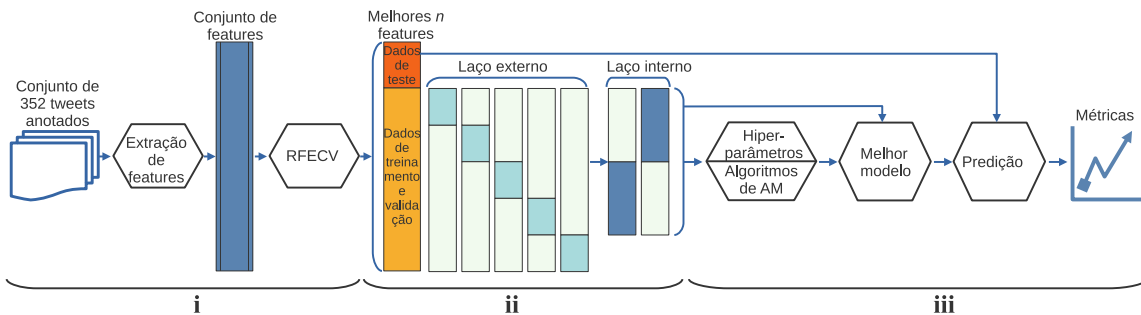


Figura 8: Configuração experimental adotada para a geração de modelos de AM baseados em *features*, dividida em: i. extração e seleção de *features*, ii. otimização com validação cruzada aninhada; e iii. treinamento e teste.

zado em duas etapas: (i) ajuste fino do modelo neural; e (ii) avaliação do modelo no conjunto de teste e geração das métricas. Vale ressaltar que as mesmas partições de treinamento e teste foram usadas tanto nos experimentos com AM baseado em *features* quanto nos experimentos com os modelos neurais, de forma estratificada.

5 Resultados

Neste artigo, investigou-se como o AM pode ser aplicado para classificar a qualidade da argumentação em *tweets* do domínio da política brasileira. Para isso, foram testados diferentes classificadores com diferentes hiperparâmetros em um

corpus de 352 *tweets* do domínio da política brasileira. Desses, 281 *tweets* foram usados para treinamento ou ajuste fino dos modelos e os restantes 71 foram usados para teste.

As seções a seguir apresentam as análises quantitativas e qualitativas dos resultados obtidos nos experimentos.

5.1 Análise quantitativa

Em termos das medidas quantitativas usualmente aplicadas na avaliação de modelos computacionais, apresentadas na Tabela 4, concluímos que o modelo neural obtido com o ajuste fino do BERTimbau obteve os melhores valores: 63,38%

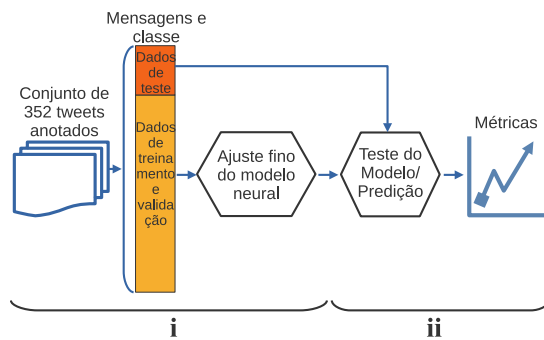


Figura 9: A configuração experimental proposta foi dividida em duas etapas: i. ajuste fino dos modelos neurais e ii. avaliação do modelo.

de acurácia, 69,65% de precisão, 63,38% de cobertura e uma medida F de 63,61%³⁰.

Classificador	Acurácia	Precisão	Cobertura	Medida F
LR	45,07%	45,34%	45,07%	44,56%
LR + RFECV	46,48%	46,53%	46,48%	45,75%
K-NN	47,89%	49,20%	47,89%	46,86%
K-NN + RFECV	39,44%	41,68%	39,44%	37,80%
DT	46,48%	32,57%	46,48%	37,71%
DT + RFECV	46,48%	32,57%	46,48%	37,71%
SVM	43,66%	42,99%	43,66%	41,10%
SVM + RFECV	35,21%	35,95%	35,21%	35,39%
RF	54,93%	54,90%	54,93%	54,44%
RF + RFECV	45,07%	46,53%	45,07%	43,24%
NB	45,07%	45,06%	45,07%	44,29%
NB + RFECV	50,70%	50,89%	50,70%	50,49%
RobertaTwitterBR	49,30%	48,41%	49,30%	48,11%
BERTimbau	63,38%	69,65%	63,38%	63,61%

Tabela 4: Valores das medidas de avaliação obtidas para a qualidade geral da argumentação nos modelos baseados em *features* (LR, K-NN, DT, RF e NB) com (+ RFECV) e sem a etapa de seleção de *features* e nos modelos neurais (RobertaTwitterBR e BERTimbau).

Entre os algoritmos de AM baseado em *features*, o desempenho variou de 35% a 54% de medida F, ficando a pelo menos cerca de 9 pontos percentuais abaixo do melhor modelo neural. Quanto ao baixo desempenho desses modelos, vale mencionar que a alta quantidade de *features* usadas no treinamento pode ter influenciado esses resultados. Mesmo com a etapa de seleção de *features*, realizada como descrito na Figura 8, os modelos tradicionais foram treinados com 290 *features* para 281 instâncias e essa alta dimensionalidade pode ter confundido os algoritmos no momento de gerar os modelos. Contudo, a confi-

³⁰Os melhores resultados foram alcançados com os seguintes hiperparâmetros otimizados: número de épocas=20; *batch size*=8; *early stop*=2; *learning rate*=1e-5. Os mesmos hiperparâmetros foram utilizados para o BERTimbau e RoBERTaTwitterBR.

guração experimental do AM baseado em *features* foi proposta para dar autonomia ao processo de AM na definição das *features* mais relevantes (sem a interferência do especialista humano), de modo semelhante ao que fazem os modelos neurais durante o ajuste fino.

É importante destacar que os experimentos conduzidos com DT apresentaram resultados idênticos com e sem a etapa de seleção de *features*. Isso se deve às características do próprio algoritmo, que já faz a seleção de *features* durante o treinamento para definir qual *feature* vai em cada nó da árvore. O DT trabalha dividindo recursivamente o conjunto de *features* em subconjuntos homogêneos, até que um critério de parada seja atingido. Durante o processo de construção da árvore, o algoritmo avalia diferentes *features* e escolhe aquela que melhor separa as classes ou minimiza o erro.

Nota-se, também, que os experimentos conduzidos com SVM sem a etapa de seleção de *features* apresentaram melhores resultados se comparados com o experimento com a seleção de *features*. Isso também se deve às características do algoritmo, que trabalha transformando as *features* em um espaço de alta dimensão, no qual é mais provável que as classes sejam linearmente separáveis: quanto mais pontos, melhor o detalhamento na definição das bordas. Em seguida, o SVM encontra o hiperplano que separa as classes, minimizando o erro ou a perda (Vapnik, 1999).

A partir dos valores das medidas de avaliação apresentados nesta seção é possível concluir que os modelos de AM baseados em *features* não são os mais indicados para identificar automaticamente os *tweets* do campo da política brasileira que tenham bons argumentos. Embora os experimentos tenham sido realizados de maneira bastante abrangente em termos de recursos linguístico-computacionais usados para o português, *features* e algoritmos, não é possível apontar claramente quais desses fatores levaram ao baixo desempenho.

Em relação aos resultados dos modelos neurais, a Tabela 5 e a Figura 10 apresentam os resultados detalhados do modelo de melhor desempenho obtido com o ajuste fino do BERTimbau. Esse modelo teve uma excelente precisão para a classe Alta. Das 19 instâncias da classe Alta presentes no corpus de teste, o modelo foi capaz de prever 9 delas corretamente (resultando em uma cobertura de 47,37%) mas com uma precisão de 100%. Assim, embora a amostra do corpus de teste seja pequena, esses resultados trazem fortes indícios de uma excelente assertividade do modelo para prever *tweets* com alta qualidade da

argumentação. Esses *tweets* são apresentados na Tabela 6.

	Precisão	Cobertura	Medida F
Baixa	64,52%	71,43%	67,80%
Média	51,61%	66,67%	58,18%
Alta	100,00%	47,37%	64,29%

Tabela 5: Resultados detalhados para a predição da qualidade geral da argumentação retornados pelo modelo obtido com o ajuste fino do BERTimbau.

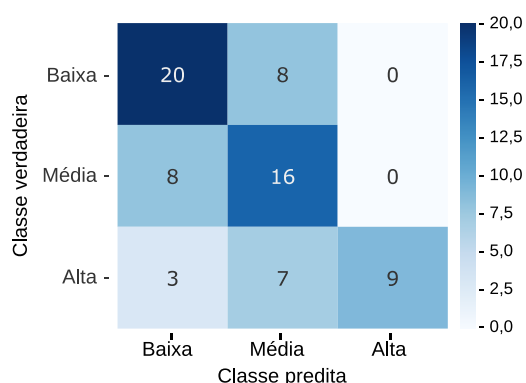


Figura 10: Matriz de confusão do modelo obtido com o ajuste fino do BERTimbau.

5.2 Análise qualitativa

Para complementar a análise quantitativa, esta seção traz uma análise qualitativa das classificações do modelo de melhor desempenho para a predição da qualidade da argumentação nos *tweets* do domínio da política brasileira.

5.2.1 Análise da classe de Alta qualidade de argumentação

Como mencionado anteriormente, o objetivo dos experimentos apresentados neste artigo foi gerar um modelo capaz de prever, com um bom desempenho, aqueles *tweets* que tenham alta qualidade de argumentação. Assim, a Tabela 6 apresenta os 9 *tweets* do corpus de teste que o modelo ajustado a partir do BERTimbau apontou corretamente como sendo de alta qualidade de argumentação.

Vale mencionar que todos foram considerados pelos anotadores humanos como sendo claros, organizados e com nenhum ou poucos erros de português. Além dessas características gerais que levaram os anotadores a considerar tais *tweets* como de alta qualidade da argumentação, outras

pistas linguísticas (destacadas em **negrito**) incluem: uso de tratamento cordial (ex: Desculpe, Caro, senhora, Sr.), fato histórico (ex: como foi feito na Alemanha), numérico (ex: 56 milhões de eleitores) ou citação de figura/entidade pública que dá respaldo a alguma informação (ex: Arthur Lira, conselho de ética). A presença de marcadores discursivos (ex: como, porque, ainda mais) também foi apontada como positiva para melhorar a qualidade argumentativa porque eles unem, contrapõem ou interligam ideias.

Como exemplo, na mensagem 6.1, da Tabela 6, dentre as principais pistas linguísticas que fizeram com que os humanos considerassem esse *tweet* como de alta qualidade de argumentação, destacam-se a presença de: (i) termo especializado “Impeachment” (apontado por todos os anotadores); (ii) encadeamento cronológico expresso pela presença do “quando”, usado com sentido temporal (apontado por todos os anotadores), e (iii) fato midiático (pista anotada pela maioria dos anotadores).

Por outro lado, outros 10 *tweets*, que correspondem às mensagens 7.1 até 7.10, da Tabela 7, considerados de alta qualidade argumentativa pelos anotadores, foram classificados pelo modelo BERTimbau de modo diferente: 70% deles como Média e 30% deles como Baixa.

Comparando os *tweets* da Tabela 7 com os da Tabela 6 podemos notar uma maior presença de abreviações (ex: n, q, ñ, tds) e grafias diferentes para palavras (ex: “K0V1D” para representar Covid, ou “vac1na” para vacina), além de uma ausência notável de pontuação (como o uso da vírgula). Essas características retratam bem as peculiaridades do gênero textual e aparentemente tiveram bastante peso na classificação do modelo BERTimbau.

5.2.2 Análise dos erros

A seguir, apresenta-se uma breve discussão das principais disparidades observadas entre a anotação humana e a classificação com o modelo ajustado a partir do BERTimbau. Na Tabela 7, são apresentadas algumas mensagens do conjunto de testes que demonstraram divergência entre as avaliações humanas e automáticas.

Os exemplos 7.8, 7.9 e 7.10, da Tabela 7, são alguns dos casos nos quais foram observadas as principais divergências, uma vez que o modelo os classificou como Baixa argumentatividade e os anotadores os avaliaram como Alta argumentatividade. Características como ausência de vírgulas no exemplo 7.8, presença de maiúsculas no exemplo 7.9 e de *hashtags* no exemplo 7.10

Ex.	Texto
6.1	Rodrigo Maia, você hoje já falou que se arrepende do apoio a Bolsonaro no segundo turno. Parabéns por admitir isto. Agora... quando virá o arrependimento de não ter ao menos colocado para a frente algum dos pedidos de Impeachment ?
6.2	Vc propôs essa emenda, esperando que passe ou apenas para constar? Com a postagem do seu presidente da câmara, que até já considerou que o Dep. Daniel Silveira contrapôs à democracia, mesmo não tendo sido julgado e condenado pelo STF, espera que essa sua proposta tenha sucesso? https://t.co/uJjvgcwqEt
6.3	Desculpe senhora deputada, cansei de vcs ! Ninguém faz nada, ninguém! Vcs brincam com o povo! Se hoje um governador maluco fizer um forno, como foi feito na Alemanha e começar a matar as pessoas,tudo bem , os caras que jamais devem ser citados, deram o direito !
6.4	Caro Deputado, não sei se irá ler meu posicionamento. Mas, calaram a voz de uma Deputado q foi eleito para PODER FALAR POR NÓS! Um PODER, calou a não a voz do Daniel, calou foi a NOSSA! Ontem foi deputado pondo mordada da boca de outro deputado e traçando o fim do CONGRESSO.
6.5	Está na hora de exigir o respeito com seriedade, impeachment se faz mais que necessário, ele está tentando rebaixar a Câmara dos Deputados a seu serviço, uma ação judicial enérgica imediata. Ação do Arthur Lira agora, se deixar passar perderá a força 🙌👏👍👍👍👍
6.6	Ao Sr. Apresentar esse material ao conselho de ética , vamos ver se esse conselho de ética e justo, ou hipócrita ou incoerente. Se esse conselho disser que esses deputados e senadores que cometeram crime de ofensa, no cometeram crime porque tem imunidade parlamentar . Aí tem!!!
6.7	A prisão é ilegal já no momento que ele usa um inquérito que serve pra tudo , sem requerimento da polícia ou do MP o resto só invalida ainda mais a prisão, o que esse ministro fez é caso de impeachment e prisão
6.8	Mas também deputada, com essa oposição que tudo que o governo federal faz vocês acham que está errado. Imagine se o povo estivesse todos seguindo o FIQUE EM CASA, A ECONOMIA A GENTE VER DEPOIS . Sou a favor que sejam seguidos os protocolos: máscara, lavar as mãos e não aglomerar.
6.9	Deveria abrir uma CPI para investigar as sabotagens que o ex-presidente da câmara comandou durante seu mandato, causando prejuízos incalculáveis à toda nação e desrespeitando a vontade de mais de 56 milhões de eleitores!

Tabela 6: 9 tweets de alta qualidade presentes no corpus de teste corretamente classificados pelo modelo ajustado a partir do BERTimbau.

podem ter influenciado o modelo a classificá-los como de Baixa qualidade. Por outro lado, os anotadores humanos provavelmente os classificaram como de Alta qualidade com base no conteúdo e apelo emocional e não na forma superficial do texto³¹. No restante dos exemplos (7.11, 7.12, 7.13 e 7.14) houve diferenças entre Baixa/Média e Média/Alta argumentatividade. Essas diferenças foram vistas como plausíveis, considerando o contexto semântico do argumento.

Além dos exemplos de erros apresentados na Tabela 7, outras 12 mensagens foram classificadas de forma equivocada pelo modelo. Destas, 6 de Baixa argumentatividade foram classificadas como Média e 6 de Média argumentatividade foram classificadas como Baixa. Importante ressaltar que todas as mensagens classificadas como de Alta qualidade da argumentação pelo modelo ajustado a partir do BERTimbau foram avaliadas da mesma forma pelos anotadores, isso indica uma precisão absoluta em classificar mensagens de alta argumentatividade.

³¹Os exemplos 7.8 e 7.9 foram classificados como de Alta qualidade pela maioria dos anotadores, e o exemplo 7.10 foi classificado como de Alta qualidade por todos os anotadores.

6 Conclusões, limitações e trabalhos futuros

Este artigo apresentou experimentos para a predição da qualidade da argumentação em tweets do domínio da política brasileira. Os experimentos foram realizados com algoritmos de AM baseado em *features* e em redes neurais. O modelo que apresentou o melhor desempenho foi o obtido a partir do ajuste fino do BERTimbau.

Em relação às questões de pesquisa inicialmente propostas para este trabalho, a partir dos experimentos aqui apresentados pode-se concluir que: (Q1) é possível prever automaticamente a qualidade da argumentação em tweets do domínio da política brasileira e (Q2) a maneira que se mostrou mais adequada para fazê-lo foi usando o modelo neural gerado a partir do ajuste fino do BERTimbau. Tal modelo foi capaz de prever com 100% de precisão instâncias da classe de Alta qualidade da argumentação, satisfazendo o objetivo deste trabalho que é o de encontrar tweets com boa capacidade argumentativa.

Embora o objetivo do trabalho tenha sido alcançado, algumas limitações são apresentadas na próxima subseção, seguida de algumas propostas de trabalhos futuros.

Ex. Texto	Real	BERTimbau
7.1 Acho que o foco deveria ser nas desastrosas intervenções nas estatais (que já n deveriam ser mais estatais) e nos preços de energia, que logo se avizinham. N adianta ser um leão contra a prisão (q deveria ter discurso técnico), e um gato contra o desastre das políticas do gov.!	Alta	Média
7.2 Precisamos tds deixar bem claro que essa ñ representa o povo e que esses parlamentares são responsáveis pelas mortes diárias de brasileiros. É imoral que estejam votando um projeto de lei para se protegerem da lei em vez de salvar a vida de brasileiros!	Alta	Média
7.3 Quem sabe né o seu Jair da uma dentro. O dublê de ministro da saúde tá sendo um fiasco, um fracasso, uma vergonha nacional. Uma coisa é usina hidroeétrica outra coisa é a Petrobrás. Veremos.	Alta	Média
7.4 Hoje fui comprar 1kg de carne moída para o almoço e deu R\$ 43,00 achei um absurdo! Em que mundo estamos com um custo tão alto de carne assim?! Mas de boa estamos comprando carne de primeira para nossos representantes políticos, então pq reclamar?! 😞 😞	Alta	Média
7.5 Cadê a dengue, cadê a gripe comum, a H1N1, cadê as demais doenças respiratórias, tão comuns em nosso país nessa época do ano? Sumiram ou estão se somando aos números do K0VID? Tudo pela ciência e pela saúde! Principalmente agora que governadores e prefeitos podem comprar vacIna.	Alta	Média
7.6 O que o Brasil precisa é um sistema bancário que ofereça juros selic 2% ao ano para qualquer um empreender. Reduzir esse spread avassalador. Infelizmente politico so pensa em imposto e nao em melhorar a eficiencia.	Alta	Média
7.7 Esse é a teoria, . Na vida real, na prática brasileira, vendem água mineral a 30 reais, quando tem desastre natural... Privatizem mesmo! Abram o mercado sim! Quebrem os monopólios! Mas, acima de tudo, FISCALIZEM os espertinhos, senão eles deitam e rolam!	Alta	Média
7.8 Ah pelo amor C0co Bambú vive lotado se não tem grana para manter os funcionários durante o Lockdown tem péssima administração! Empatia zero e o governo federal precisa ajudar as pessoas essa que é a realidade!	Alta	Baixa
7.9 NAO É CRIME DE OPINIAO. VOU REPETIR PRO SARGENTO GARCIA ENTENDER. NAO É CRIME DE OPINIAO, e nao está coberto peka imunidade material so parlamentar, basta saber ler o art. 53. Mas é esperar muito de bolsonaristas. Que vergonha pro meu Parana ter deputado assim.	Alta	Baixa
7.10 Pode jogar a nossa Constituição no lixo, pois a acabam de transformar os 11 ministros do STF nos novos IMPERADORES do Brasil. #CamaraDosDeputadosVergonhaNacional #CámaraDosDeputadosRasgandoAConstituição #DanielSilveiraFalouPeloPovo #STFVergonhaMundial	Alta	Baixa
7.11 Você pode ter 6 seguranças armados né deputado, e defender bandido, porque não pede ao STF para mandar investigar todo o governo do Rio de janeiro incluindo vocês de Dourados, vereadores e senadores, talvez assim consigam desarmar os traficantes dos morros	Média	Baixa
7.12 Os e o Sr perderam o meu respeito não acredito mais nessa turma de cristão não tem nada lembrei de Bararabas e dos lideres do sinedrio foi igual a passagem veio a minha mente no voto de muitos da AD e da IURD	Média	Baixa
7.13 Do jeito que esses deputados são, depois que vimos semana passada, acho difícil conseguir as assinaturas necessárias. São um bando de covarde, só pensam neles, no próprio ego. Nem a voz do povo eles escutam. 2022 é logo ali. #TodoPoderEmanaDoPovo #Democracia	Baixa	Média
7.14 Nada melhora pro povo já que não há instituições neste país apodreceu tdo e fede faz tempo, mas pelo menos é uma dívida a menos nas costas dos pagadores de impostos, mais conhecidos pela elite como trouxas	Baixa	Média

Tabela 7: Exemplos de erros de predição do modelo neural.

6.1 Limitações

Uma das principais limitações deste trabalho é o tamanho do corpus. Embora seja compatível com o corpus desenvolvido por Wachsmuth et al. (2017b) quanto ao número de instâncias, os resultados aqui apresentados têm seus impactos limitados pela pouca quantidade de instâncias usadas na geração e na avaliação dos modelos.

Outra limitação está relacionada à decisão de projeto adotada no momento da anotação do corpus que foi a de realizar a anotação da qualidade da argumentação de forma isolada, ou seja, desconsiderando o *tweet* semente. Diante disso, assumiu-se que o argumento pode ser classificado independentemente do tópico. Trabalhos da literatura (Fromm et al., 2019; Hidayaturrahman et al., 2021) apontam que, muitas vezes, frases contendo argumentos são estruturalmente semelhantes a frases puramente informativas sem qualquer posicionamento sobre o tópico e que

considerar a informação do tópico é crucial para a tarefa, pois ele define o contexto semântico de um argumento. Para lidar com essa limitação, uma possível estratégia é concatenar as mensagens avaliadas com o *tweet* semente de modo a incorporar contexto e, consequentemente, melhorar o desempenho do modelo.

Outra decisão de projeto que pode ter impactado o desempenho do modelo está relacionada ao modo como os aspectos foram combinados para definir a Qualidade Geral da argumentação. Conforme visto na Tabela 7, algumas instâncias de teste anotadas como de Alta qualidade foram classificadas pelo modelo BERTimbau como de Média ou Baixa qualidade o que, analisando o conteúdo de tais *tweets* pode ter fundamento. Assim, uma proposta de trabalho futuro é fazer uma revisão dos critérios e do modo como os aspectos são combinados para definir a Qualidade Geral.

É importante considerar as características específicas do Twitter. Diferentemente de ou-

tras mídias sociais, o Twitter não possui uma política das mais rígidas para restringir ou filtrar o conteúdo das postagens ou o comportamento abusivo dos usuários. Por causa disso, postagens que contêm palavrões e discurso de ódio são comuns. No campo da política brasileira atual, essas características são ainda mais acentuadas, com postagens contendo *fake news*, ataques pessoais a políticos ou às famílias deles, ideologia política, entre outros. Assim, esses textos (*tweets*) costumam ter várias marcas que impactam negativamente a qualidade deles e, conseqüentemente, reduzem a qualidade geral da argumentação.

Por fim, outra limitação relacionada ao Twitter é o número muito limitado de caracteres (280) permitido para cada mensagem, o que dificulta o uso de estratégias mais elaboradas de argumentação linguística pelos autores das postagens.

6.2 Trabalhos futuros

Como trabalhos futuros, destacam-se três caminhos que trariam maior benefício às propostas aqui apresentadas: (i) como apresentado na Seção 6.1, concatenar as mensagens avaliadas com o *tweet* semente de modo a incorporar contexto e, conseqüentemente, melhorar o desempenho do modelo; (ii) testar uma combinação (*ensemble*) de classificadores, que constitui-se em múltiplos classificadores, treinados de forma individual, cujos resultados são combinados, também com o objetivo de buscar uma melhora no desempenho do modelo; (iii) aumentar o corpus com a finalidade de capturar novos contextos, além de um acréscimo da sua própria dimensão.

Além dessas, outra possibilidade seria gerar modelos específicos para cada um dos aspectos da qualidade da argumentação (Clareza, Organização, Credibilidade e Polaridade e Intensidade do Apelo Emocional) com o intuito de combiná-los para definir automaticamente a qualidade geral da argumentação, como foi feito de modo manual, pelos anotadores, no momento da geração do corpus (Silva et al., 2021).

Agradecimentos

Os autores agradecem aos linguistas e coautores em (Silva et al., 2021), que anotaram o corpus e contribuíram para a elaboração das diretrizes de anotação, utilizados neste trabalho: Amanda Pontes Rassi, Jackson Wilke da Cruz Souza, Renata Ramisch e Roger Alfredo de Marci Rodrigues Antunes. Agradecem, também, ao Sidney Evaldo Leal e ao Núcleo Interinstituci-

onal de Linguística Computacional (NILC) pelos recursos linguístico-computacionais disponibilizados para o desenvolvimento desta pesquisa. Por fim, os autores agradecem ao Programa de Pós-Graduação em Ciência da Computação (PPGCC) da Universidade Federal de São Carlos (UFSCar) e à Rede Gonzaga de Ensino Superior (REGES), pelo apoio a este trabalho.

Referências

- Adi, Sumarni, Yoga Pristyanto & Andi Sunyoto. 2019. The best features selection method and relevance variable for web phishing classification. Em *2019 International Conference on Information and Communications Technology (ICOIACT)*, 578–583. doi:10.1109/ICOIACT46704.2019.8938566.
- Al-Khatib, Khalid, Henning Wachsmuth, Matthias Hagen, Jonas Köhler & Benno Stein. 2016. Cross-domain mining of argumentative text through distant supervision. Em *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1395–1404. San Diego, California: Association for Computational Linguistics. doi:10.18653/v1/N16-1165. <https://www.aclweb.org/anthology/N16-1165>.
- Armentano-Oller, C., R. C. Carrasco, A. M. Corbí-Bellot, M. L. Forcada, M. Ginestí-Rosell, S. Ortiz-Rojas, J. A. Pérez-Ortiz, G. Ramírez-Sánchez, F. Sánchez-Martínez & M. A. Scalco. 2006. Open-source Portuguese-Spanish machine translation. Em *Proceedings of the VII Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada*, 50–59. Itatiaia, RJ, Brazil.
- Balage Filho, Pedro P., Thiago Alexandre Salgueiro Pardo & Sandra M. Aluísio. 2013. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. Em *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, <https://www.aclweb.org/anthology/W13-4829>.
- Bench-Capon, Trevor JM & Paul E Dunne. 2007. Argumentation in artificial intelligence. *Artificial intelligence* 171(10-15). 619–641.
- Bilu, Yonatan & Noam Slonim. 2016. Claim synthesis via predicate recycling. Em *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Vo-*

- lume 2: *Short Papers*), 525–530. Berlin, Germany: Association for Computational Linguistics. doi:10.18653/v1/P16-2085. <https://www.aclweb.org/anthology/P16-2085>.
- Blair, J. Anthony. 2012. Rhetoric, dialectic, and logic as related to argument. *Philosophy & Rhetoric* 45(2). 148–164. <http://www.jstor.org/stable/10.5325/philrhett.45.2.0148>.
- Boudry, Maarten, Fabio Paglieri & Massimo Pigliucci. 2015. The fake, the flimsy, and the fallacious: Demarcating arguments in real life. *Argumentation* 29(4). 431–456. doi:10.1007/s10503-015-9359-1. <http://doi.org/10.1007/s10503-015-9359-1>.
- Brum, Henrico & Maria Graças Volpe Nunes. 2018. Building a Sentiment Corpus of Tweets in Brazilian Portuguese. Em *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan: European Language Resources Association (ELRA).
- Capellaro, Leonardo & Helena Caseli. 2021. Análise de polaridade e de tópicos em tweets no domínio da política no Brasil. Em *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, 47–55. Porto Alegre, RS, Brasil: SBC. doi:10.5753/stil.2021.17783. <https://sol.sbc.org.br/index.php/stil/article/view/17783>.
- Carlisle, Winston, Nishant Gurrupadi, Zixuan Ke & Vincent Ng. 2018. Give me more feedback: Annotating argument persuasiveness and related attributes in student essays. Em *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 621–631. Melbourne, Australia: Association for Computational Linguistics. doi:10.18653/v1/P18-1058. <https://www.aclweb.org/anthology/P18-1058>.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. Em *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics. doi:10.18653/v1/N19-1423. <https://www.aclweb.org/anthology/N19-1423>.
- Eemeren, Frans H. van & Rob Grootendorst. 2003. *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge University Press. doi:10.1017/CBO9780511616389.
- Feng, Vanessa Wei, Ziheng Lin & Graeme Hirst. 2014. The impact of deep hierarchical discourse structures in the evaluation of text coherence. Em *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 940–949. Dublin, Ireland: Dublin City University and Association for Computational Linguistics. <https://www.aclweb.org/anthology/C14-1089>.
- Fonseca, E. R. & J.L.G. Rosa. 2013. MacMorpho Revisited: Towards Robust Part-of-Speech Tagging. Em *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, 98–107.
- Fromm, Michael, Max Berrendorf, Johanna Reiml, Isabelle Mayerhofer, Siddharth Bhargava, Evgeniy Faerman & Thomas Seidl. 2022. Towards a holistic view on argument quality prediction. doi:10.48550/ARXIV.2205.09803. <https://arxiv.org/abs/2205.09803>.
- Fromm, Michael, Evgeniy Faerman & Thomas Seidl. 2019. Tacam: Topic and context aware argument mining. Em *IEEE/WIC/ACM International Conference on Web Intelligence WI '19*, 99–106. New York, NY, USA: Association for Computing Machinery. doi:10.1145/3350546.3352506. <https://doi.org/10.1145/3350546.3352506>.
- García-Gorrostieta, Jesús Miguel & Aurelio López-López. 2018. Identifying argumentative paragraphs: Towards automatic assessment of argumentation in theses. Em Max Silberstein, Faten Atigui, Elena Kornysheva, Elisabeth Métais & Farid Meziane (eds.), *Natural Language Processing and Information Systems*, 83–90. Cham: Springer International Publishing.
- García-Gorrostieta, Jesús M., Aurelio López-López & Samuel González-López. 2018. Automatic argument assessment of final project reports of computer engineering students. *Computer Applications in Engineering Education* 26(5). 1217–1226. doi:10.1002/cae.21996. <https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.21996>.
- Gleize, Martin, Eyal Shnarch, Leshem Choshen, Lena Dankin, Guy Moshkovich, Ranit Aharonov & Noam Slonim. 2019. Are you convinced? choosing the more convincing evidence with a Siamese network. Em *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 967–976. Florence,

- Italy: Association for Computational Linguistics. doi:10.18653/v1/P19-1093. <https://www.aclweb.org/anthology/P19-1093>.
- Gretz, Shai, Roni Friedman, Edo Cohen-Karlik, Assaf Toledo, Dan Lahav, Ranit Aharonov & Noam Slonim. 2019. A large-scale dataset for argument quality ranking: Construction and analysis. doi:10.48550/ARXIV.1911.11408. <https://arxiv.org/abs/1911.11408>.
- Habernal, Ivan & Iryna Gurevych. 2015. Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse. Em *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2127–2137. Lisbon, Portugal: Association for Computational Linguistics. doi:10.18653/v1/D15-1255. <https://www.aclweb.org/anthology/D15-1255>.
- Habernal, Ivan & Iryna Gurevych. 2016. Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional LSTM. Em *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1589–1599. Berlin, Germany: Association for Computational Linguistics. doi:10.18653/v1/P16-1150. <https://www.aclweb.org/anthology/P16-1150>.
- Habernal, Ivan & Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics* 43(1). 125–179. doi:10.1162/COLLa-00276. <https://www.aclweb.org/anthology/J17-1004>.
- Hidayaturrehman, Emmanuel Dave, Derwin Suhartono & Aniat Murni Arymurthy. 2021. Enhancing argumentation component classification using contextual language model. doi:10.1186/s40537-021-00490-2. <http://dx.doi.org/10.1186/s40537-021-00490-2>.
- Krippendorff, Klaus. 2011. Computing krippendorff's alpha-reliability. Relatório técnico. University of Pennsylvania. http://repository.upenn.edu/asc_papers/43/.
- Lauscher, Anne, Lily Ng, Courtney Napoles & Joel Tetreault. 2020. Rhetoric, logic, and dialectic: Advancing theory-based argument quality assessment in natural language processing. Em *Proceedings of the 27th International Conference on Computational Linguistics*, online.
- Leal, Sidney Evaldo, Magali Sanches Duran, Carolina Evaristo Scarton, Nathan Siegle Hartmann & Sandra Maria Aluísio. 2022. Nilmetrix: assessing the complexity of written and spoken language in brazilian portuguese. doi:10.48550/ARXIV.2201.03445. <https://arxiv.org/abs/2201.03445>.
- Leite, João A., Diego F. Silva, Kalina Bontcheva & Carolina Scarton. 2020. Toxic language detection in social media for brazilian portuguese: New dataset and multilingual analysis.
- Lytos, Anastasios, Thomas Lagkas, Panagiotis Sarigiannidis & Kalina Bontcheva. 2019. The evolution of argumentation mining: From models to social media and emerging tools. *Information Processing & Management* 56(6). 102055. doi: <https://doi.org/10.1016/j.ipm.2019.102055>. <http://www.sciencedirect.com/science/article/pii/S030645731930024X>.
- Marcuschi, Luiz Antônio et al. 2002. Gêneros textuais: definição e funcionalidade. *Gêneros textuais e ensino*. Rio de Janeiro: Lucerna 20.
- Maćkiewicz, Andrzej & Waldemar Ratajczak. 1993. Principal components analysis (pca). *Computers & Geosciences* 19(3). 303–342. doi:10.1016/0098-3004(93)90090-R. <https://www.sciencedirect.com/science/article/pii/009830049390090R>.
- Misra, Puneet & Arun Singh Yadav. 2020. Improving the classification accuracy using recursive feature elimination with cross-validation. *Int. J. Emerg. Technol* 11(3). 659–665.
- Peldszus, Andreas & Manfred Stede. 2015. Joint prediction in MST-style discourse parsing for argumentation mining. Em *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 938–948. Lisbon, Portugal: Association for Computational Linguistics. doi:10.18653/v1/D15-1110. <https://www.aclweb.org/anthology/D15-1110>.
- Persing, Isaac & Vincent Ng. 2015. Modeling argument strength in student essays. Em *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 543–552. Beijing, China: Association for Computational Linguistics. doi: 10.3115/v1/P15-1053. <https://www.aclweb.org/anthology/P15-1053>.
- Potthast, Martin, Lukas Gienapp, Florian Euchner, Nick Heilenkötter, Nico Weidmann, Henning Wachsmuth, Benno Stein & Matthias Hagen. 2019. Argument search: Assessing argument relevance. Em *Proceedings of the 42nd*

- International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR'19*, 1117–1120. New York, NY, USA: Association for Computing Machinery. doi:10.1145/3331184.3331327. <https://doi.org/10.1145/3331184.3331327>.
- Putra, Jan Wira Gotama, Simone Teufel & Take-no-bu Tokunaga. 2021. Annotating argumentative structure in english-as-a-foreign-language learner essays. *Natural Language Engineering* 1–27. doi:10.1017/S1351324921000218.
- Reimers, Nils & Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Em *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992. Hong Kong, China: Association for Computational Linguistics. doi:10.18653/v1/D19-1410. <https://aclanthology.org/D19-1410>.
- Reimers, Nils, Benjamin Schiller, Tilman Beck, Johannes Daxenberger, Christian Stab & Iryna Gurevych. 2019. Classification and clustering of arguments with contextualized word embeddings. Em *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 567–578. Florence, Italy: Association for Computational Linguistics. doi:10.18653/v1/P19-1054. <https://aclanthology.org/P19-1054>.
- Rinott, Ruty, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni & Noam Slonim. 2015. Show me your evidence - an automatic method for context dependent evidence detection. Em *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 440–450. Lisbon, Portugal: Association for Computational Linguistics. doi:10.18653/v1/D15-1050. <https://www.aclweb.org/anthology/D15-1050>.
- Rosenfeld, Ariel & Sarit Kraus. 2016. Providing arguments in discussions on the basis of the prediction of human argumentative behavior. *ACM Trans. Interact. Intell. Syst.* 6(4). doi:10.1145/2983925. <https://doi.org/10.1145/2983925>.
- Rossini, Patrícia. 2019. *Disentangling uncivil and intolerant discourse in online political talk* 142–157. Routledge. doi:10.4324/9781351051989-9.
- Rossini, Patrícia. 2020. Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research* 0(0). 0093650220921314. doi:10.1177/0093650220921314. <https://doi.org/10.1177/0093650220921314>.
- Schaefer, Robin & Manfred Stede. 2021. Argument mining on twitter: A survey. *it - Information Technology* 63(1). 45–58. doi:10.1515/itit-2020-0053. <https://doi.org/10.1515/itit-2020-0053>.
- Silva, Cássio, Amanda Rassi, Jackson Souza, Renata Ramisch, Roger Antunes & Helena Caseli. 2021. Quality of argumentation in political tweets: what is and how to measure it / qualidade da argumentação em tweets de política: o que e como avaliar. *REVISTA DE ESTUDOS DA LINGUAGEM* 29(4). 2537–2586. doi:10.17851/2237-2083.29.4.2537-2586. <http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/18263>.
- Skitalinskaya, Gabriella, Jonas Klaff & Henning Wachsmuth. 2021. Learning from revisions: Quality assessment of claims in argumentation at scale. Em *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 1718–1729. Online: Association for Computational Linguistics. doi:10.18653/v1/2021.eacl-main.147. <https://aclanthology.org/2021.eacl-main.147>.
- Souza, Fábio, Rodrigo Nogueira & Roberto Lotufo. 2020. Bertimbau: Pretrained bert models for brazilian portuguese. Em Ricardo Cerri & Ronaldo C. Prati (eds.), *Intelligent Systems*, 403–417. Cham: Springer International Publishing.
- Stab, Christian & Iryna Gurevych. 2014. Annotating argument components and relations in persuasive essays. Em *Proceedings of CO-LING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 1501–1510. Dublin, Ireland: Dublin City University and Association for Computational Linguistics. <https://www.aclweb.org/anthology/C14-1142>.
- Stab, Christian & Iryna Gurevych. 2017a. Parsing argumentation structures in persuasive essays. *Computational Linguistics* 43(3). 619–659. doi:10.1162/COLI_a_00295. <https://www.aclweb.org/anthology/J17-3005>.
- Stab, Christian & Iryna Gurevych. 2017b. Recognizing insufficiently supported arguments in argumentative essays. Em *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics:*

- Volume 1, Long Papers*, 980–990. Valencia, Spain: Association for Computational Linguistics. <https://www.aclweb.org/anthology/E17-1092>.
- Swanson, Reid, Brian Ecker & Marilyn Walker. 2015. Argument mining: Extracting arguments from online dialogue. Em *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 217–226. Prague, Czech Republic: Association for Computational Linguistics. doi:10.18653/v1/W15-4631. <https://www.aclweb.org/anthology/W15-4631>.
- Toledo, Assaf, Shai Gretz, Edo Cohen-Karlik, Roni Friedman, Elad Venezian, Dan Lahav, Michal Jacovi, Ranit Aharonov & Noam Slonim. 2019. Automatic argument quality assessment - new datasets and methods. Em *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 5625–5635. Hong Kong, China: Association for Computational Linguistics. doi:10.18653/v1/D19-1564. <https://aclanthology.org/D19-1564>.
- Toulmin, Stephen E. 2003. *The uses of argument*. Cambridge university press.
- Van Eemeren, Frans H & Rob Grootendorst. 1987. Fallacies in pragma-dialectical perspective. *Argumentation* 1(3). 283–301. doi:10.1007/BF00136779. <http://doi.org/10.1007/BF00136779>.
- Vapnik, Vladimir. 1999. *The nature of statistical learning theory*. New York, NY: Springer 1995th edn.
- Wachsmuth, Henning, Khalid Al-Khatib & Benno Stein. 2016. Using argument mining to assess the argumentation quality of essays. Em *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 1680–1691. Osaka, Japan: The COLING 2016 Organizing Committee. <https://www.aclweb.org/anthology/C16-1158>.
- Wachsmuth, Henning, Johannes Kiesel & Benno Stein. 2015. Sentiment flow - a general model of web review argumentation. Em *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 601–611. Lisbon, Portugal: Association for Computational Linguistics. doi:10.18653/v1/D15-1072. <https://www.aclweb.org/anthology/D15-1072>.
- Wachsmuth, Henning, Nona Naderi, Ivan Habernal, Yufang Hou, Graeme Hirst, Iryna Gurevych & Benno Stein. 2017a. Argumentation quality assessment: Theory vs. practice. Em *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 250–255. Vancouver, Canada: Association for Computational Linguistics. doi:10.18653/v1/P17-2039. <https://www.aclweb.org/anthology/P17-2039>.
- Wachsmuth, Henning, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst & Benno Stein. 2017b. Computational argumentation quality assessment in natural language. Em *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 176–187. Valencia, Spain: Association for Computational Linguistics. <https://www.aclweb.org/anthology/E17-1017>.
- Wachsmuth, Henning, Martin Potthast, Khalid Al-Khatib, Yamen Ajjour, Jana Puschmann, Jiani Qu, Jonas Dorsch, Viorel Morari, Janek Bevendorff & Benno Stein. 2017c. Building an Argument Search Engine for the Web. Em Kevin Ashley, Claire Cardie, Nancy Green, Iryna Gurevych, Ivan Habernal, Diane Litman, Georgios Petasis, Chris Reed, Noam Slonim & Vern Walker (eds.), *4th Workshop on Argument Mining (ArgMining 2017) at EMNLP*, 49–59. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W17-5106>.
- Wachsmuth, Henning, Benno Stein & Yamen Ajjour. 2017d. “PageRank” for argument relevance. Em *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 1117–1127. Valencia, Spain: Association for Computational Linguistics. <https://www.aclweb.org/anthology/E17-1105>.
- Wachsmuth, Henning & Till Werner. 2020. Intrinsic quality assessment of arguments. Em *Proceedings of the 28th International Conference on Computational Linguistics*, 6739–6745. Barcelona, Spain (Online): International Committee on Computational Linguistics. <https://www.aclweb.org/anthology/2020.coling-main.592>.
- Walton, Douglas N & David N Walton. 1989. *Informal logic: A handbook for critical argument*. Cambridge University Press.

-
- Wei, Zhongyu, Yang Liu & Yi Li. 2016. Is this post persuasive? ranking argumentative comments in online forum. Em *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 195–200. Berlin, Germany: Association for Computational Linguistics. doi:10.18653/v1/P16-2032. <https://www.aclweb.org/anthology/P16-2032>.
- Weltzer-Ward, Lisa, Beate Baltes & Laura Knight Lynn. 2009. Assessing quality of critical thought in online discussion. *Campus-Wide Information Systems* 26(3). 168–177. doi:10.1108/10650740910967357. <http://doi.org/10.1108/10650740910967357>.
- Zhang, Justine, Ravi Kumar, Sujith Ravi & Cristian Danescu-Niculescu-Mizil. 2016. Conversational flow in Oxford-style debates. Em *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 136–141. San Diego, California: Association for Computational Linguistics. doi:10.18653/v1/N16-1017. <https://www.aclweb.org/anthology/N16-1017>.

Capítulo 5

Interface de usuário para o público geral

A argumentação é uma das atividades intelectuais fundamentais da mente humana, e todas as sociedades humanas defendem seus próprios argumentos com o objetivo de explorar sua diversidade de conceitos e aprender com essa exploração. O processo de argumentação, que tem como objetivo principal a resolução de questões controversas, normalmente não segue uma sequência de passos logicamente ordenados, mas implica um grande conjunto de ações cognitivas altamente complexas. Essas ações, que geralmente envolvem emoções, elementos persuasivos e técnicas de oratória, são essencialmente difíceis de formalizar por meio de algoritmos computacionais.

É nesse contexto que áreas como Processamento de Linguagem Natural (PNL) e Interação Humano-Computador (IHC) unem forças para desenvolver novas técnicas e conjuntos de dados para avançar na tarefa de avaliação automática da qualidade dos argumentos.

Este capítulo apresenta o processo de desenvolvimento da interface pública criada para o uso do modelo de avaliação da qualidade da argumentação em *tweets* no domínio da política brasileira desenvolvido neste trabalho.¹

Para esta finalidade, um protótipo funcional baseado na web foi implementado. Os aplicativos baseados na Web têm vantagens, como a capacidade de implementar interfaces dinâmicas, mais atraentes e fáceis de usar, e implantação instantânea em praticamente qualquer dispositivo eletrônico conectado à Internet, como PCs e smartphones, em qualquer lugar do mundo. Para isso, foram utilizados recursos da biblioteca Bootstrap 5², que é um *framework* web de código aberto para desenvolvimento de interface e componentes

¹ Disponível em: <<https://argq.org/avalia/>>

² Disponível em: <<https://getbootstrap.com/>>

front-end para sites e aplicações web utilizando *HyperText Markup Language* (HTML), *Cascading Style Sheets* (CSS) e JavaScript. PHP 7.3, um banco de dados MySQL e o servidor web Apache foram usados para criar o *back-end*.

A interface tem como objetivo auxiliar o usuário a compreender como utilizar o sistema, identificar quais recursos estão disponíveis e como utilizá-los da maneira adequada. De acordo com Shneiderman (1997), a fim de cumprir esse objetivo, as interfaces devem levar em conta usuários com diferentes níveis de conhecimento e necessidades, facilitando a integração da tecnologia em diversas áreas. O sistema desenvolvido, que também requer técnicas de Inteligência Artificial (IA), segue as diretrizes propostas por (SHNEIDERMAN, 2020), que são parte do framework *Human-Centered Artificial Intelligence* (HCAI) que separa os níveis de automação dos níveis de controle humano. Assim, é possível fornecer altos níveis de controle humano e altos níveis de automação, que são mais propensos a produzir aplicativos *Reliable, Safe & Trustworthy* (RST) (Confiáveis, seguros e honestos). Ao atingir esses objetivos, especialmente para problemas complexos, aumenta-se o desempenho humano, ao mesmo tempo em que apoia a autoeficácia, o domínio, a criatividade e a responsabilidade dos humanos (SHNEIDERMAN, 2020).

Desse modo, a interface apresentada neste capítulo visa permitir o uso do modelo de avaliação automática de argumentos postados no Twitter por meio de uma *thread* principal. O sistema, baseado no modelo proposto, então avalia as mensagens postadas após a primeira mensagem (semente). A interface possui essencialmente duas telas principais a partir das quais todos os recursos e informações estão disponíveis. A primeira tela, com design minimalista, oferece um campo para inserção da URL da *thread* a ser analisada, conforme Figura 1. Por meio do link “Veja um exemplo”, pode-se executar um exemplo de avaliação da qualidade da argumentação.



Figura 1 – Tela inicial do sistema.

Na próxima tela, conforme ilustrado na Figura 2, estão todas as funcionalidades dis-

poníveis, a saber:

1. mensagem semente (em negrito) e informações do usuário que a postou;
2. um gráfico com barras horizontais indicativas dos percentuais de cada uma das faixas de pontuação das mensagens daquela *thread*: “Alta” (verde), “Média” (azul) e “Baixa” (vermelho);
3. mensagens postadas em resposta à mensagem semente, acompanhadas de suas avaliações;
4. um menu lateral exibindo as opções do sistema;

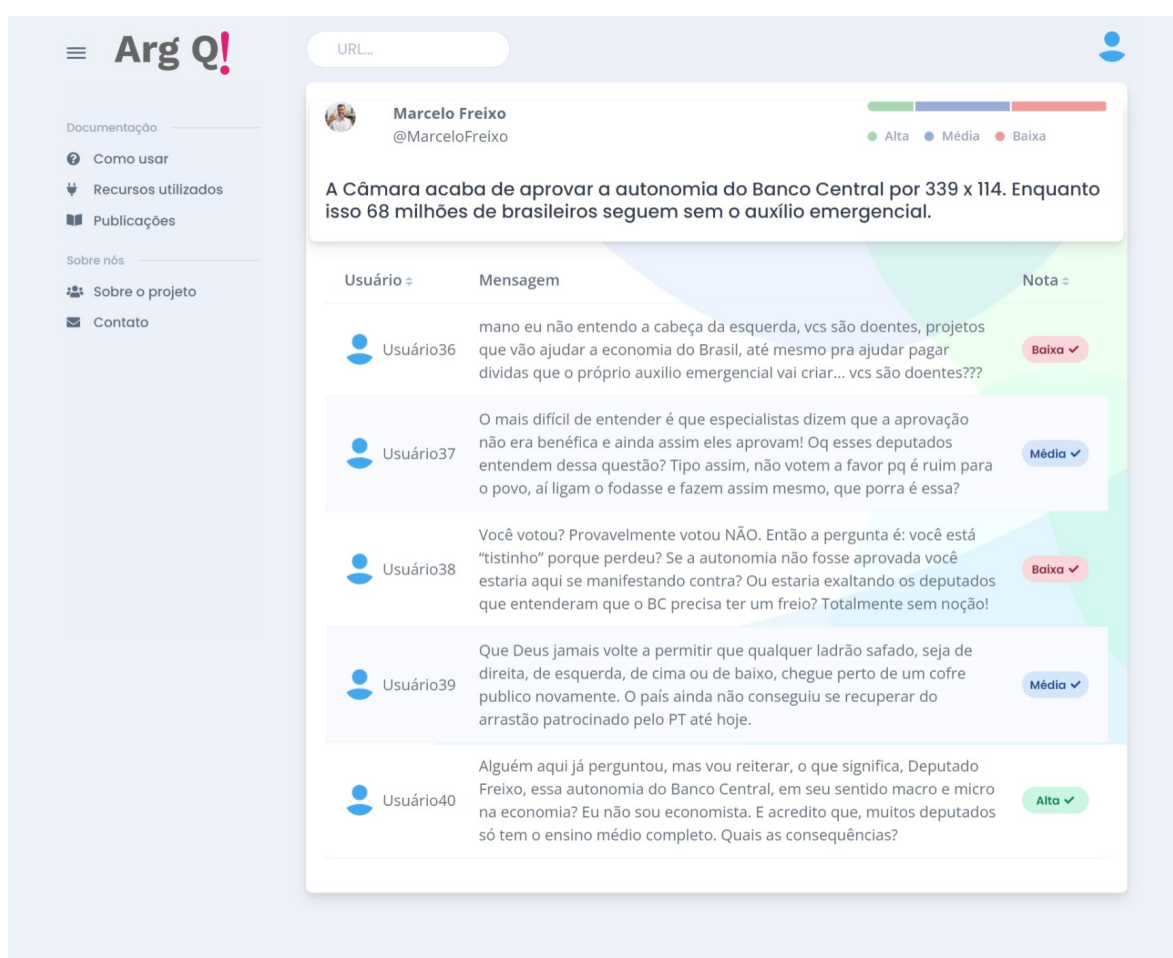


Figura 2 – Tela de avaliação da qualidade da argumentação.

Para o desenvolvimento do modelo de IA, foi utilizado o corpú de mensagens descrito nos Capítulos 2 e 4. Os resultados do processamento do corpú e desenvolvimento do modelo estão detalhados no Capítulo 4.

O sistema exibe todas as informações automaticamente. De acordo com Shneiderman (2020), a estrutura HCAI separa a questão do controle humano da automação computacional, deixando claro que altos níveis de controle humano e altos níveis de automação

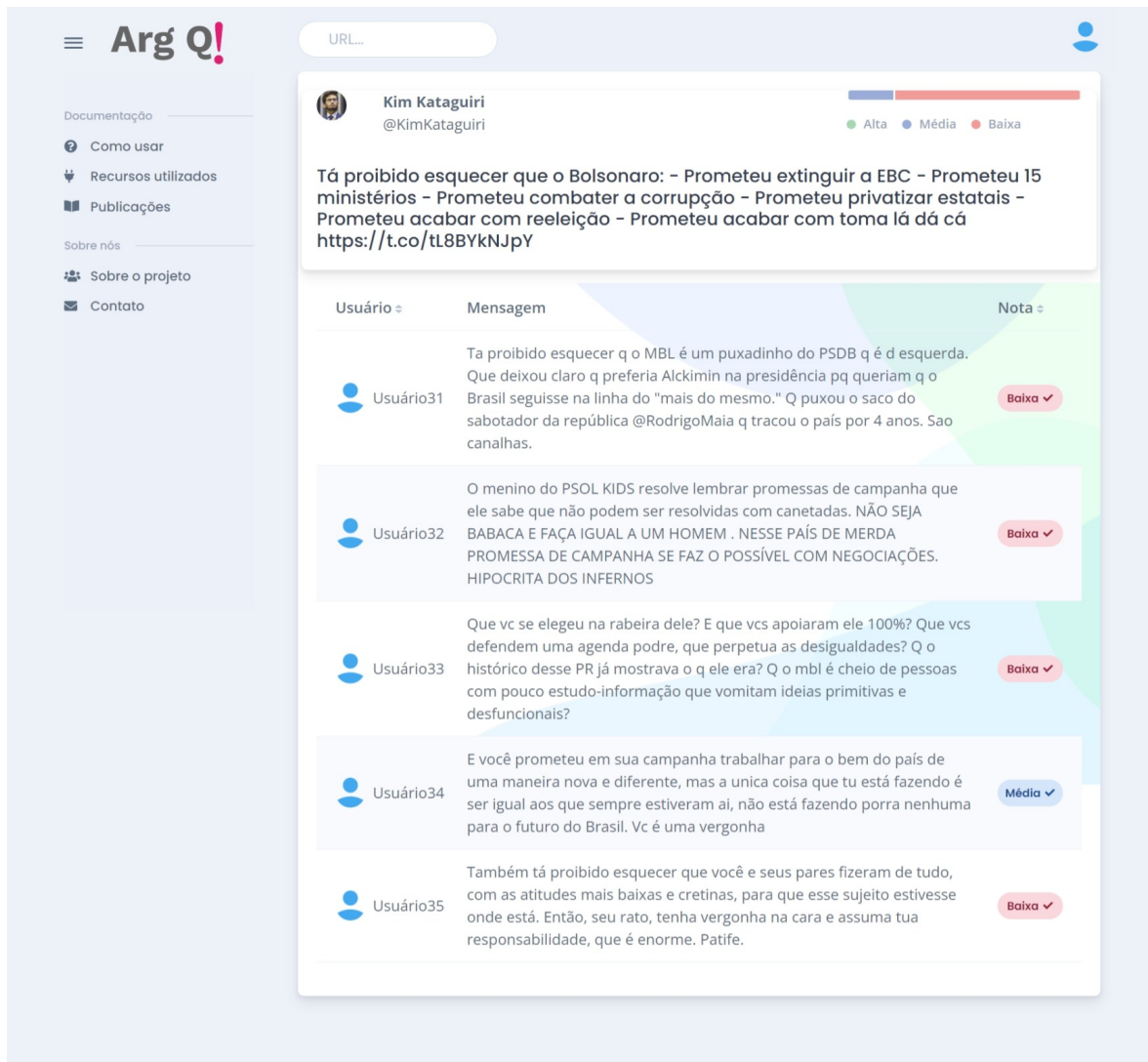


Figura 3 – Tela de avaliação da qualidade da argumentação.

podem ser alcançados por um bom design. As decisões de projeto fornecem aos operadores humanos uma compreensão clara do estado da máquina e suas escolhas são guiadas por preocupações como as consequências e a reversibilidade dos erros. A automação bem projetada preserva o controle humano quando apropriado, aumentando assim o desempenho e permitindo melhorias criativas.

Na Figura 2 são apresentados 5 *tweets* em resposta à mensagem original produzida pelo Deputado Marcelo Freixo. A partir da mensagem original, os *tweets* em resposta foram recuperados e rotulados automaticamente como de qualidade: Alta (1), Média (2) e Baixa (2). Na Figura 3 são apresentados *tweets* em resposta ao Deputado Kim Katagui, rotulados automaticamente como de qualidade: Média (1) e Baixa (4). E na Figura 4 são apresentados *tweets* em resposta à Deputada Bia Kicis, rotulados automaticamente como de qualidade: Alta (1), Média (1) e Baixa (3). Vale mencionar que o limite de 5 mensagens é uma configuração atual do protótipo, mas isso pode ser facilmente alterado para um parâmetro definido pelo usuário.

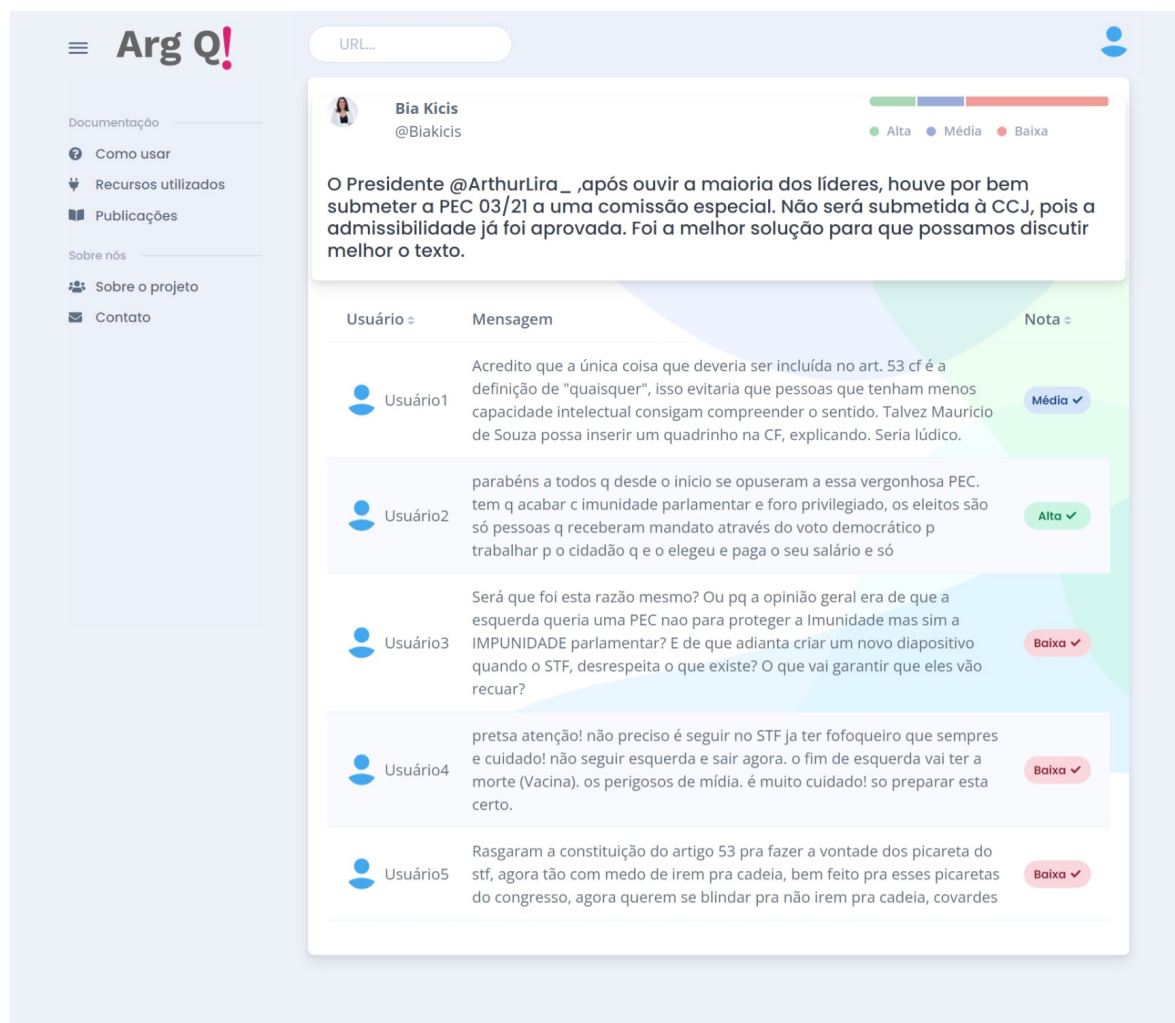


Figura 4 – Tela de avaliação da qualidade da argumentação.

Fazendo uma análise com base nas pistas linguísticas definidas pelos anotadores do projeto ArgQ!, alguns indícios podem ser apontados como indicativos das notas atribuídas pelo sistema. Como exemplo, na mensagem relacionada ao Usuário36, da Figura 2, entre os indícios que podem ter levado o modelo a classificar o *tweet* como de baixa qualidade de argumentação estão: (i) referência pejorativa “vcs são doentes” e (ii) problemas de organização e credibilidade. Por outro lado, na mensagem relacionada ao Usuário2, da Figura 4, entre os indícios que podem ter levado o modelo a classificar o *tweet* como de alta qualidade de argumentação estão a presença de: (i) relação de “Causa e efeito, Explicação, Restrição, Finalidade”, e (ii) termos especializados “imunidade parlamentar” e “foro privilegiado”.

É importante ressaltar que ainda não foram realizadas avaliações de usabilidade, com foco na satisfação de uso, e nem de acessibilidade do protótipo. Dessa forma, no momento oportuno, o interesse da avaliação poderá ser saber se os usuários consideram útil o uso dos recursos disponíveis no sistema, bem como a eficiência, eficácia e satisfação de uso.

Capítulo 6

Conclusão

Este capítulo apresenta as respostas à questão de pesquisa, as contribuições e limitações desta Tese, bem como possíveis trabalhos futuros.

6.1 Respostas à questão de pesquisa

A seguir é retomada a questão levantada no início desta pesquisa e apresentadas as considerações sobre ela, após reflexão na fase final desta etapa.

Qual é o método mais adequado para prever automaticamente a qualidade da argumentação em tweets no domínio da política brasileira a fim de identificar aqueles com bons argumentos?

Para responder essa questão de pesquisa, inicialmente realizou-se um estudo sobre os trabalhos que descrevem a qualidade da argumentação (TOULMIN, 2003; WACHSMUTH et al., 2017; POTTHAST et al., 2019; WACHSMUTH; WERNER, 2020; SKITALINSKAYA; KLAFF; WACHSMUTH, 2021). Neste trabalho, em específico, optou-se por investigar a qualidade da argumentação no gênero Twitter, em postagens escritas em português, atacando o problema em um cenário desafiador e para um idioma pouco explorado. Seguindo os aspectos da taxonomia de Wachsmuth et al. (2017), pistas linguísticas foram definidas e utilizadas na anotação de um corpus composto por 352 *tweets* considerados argumentativos. Diversas ferramentas e recursos linguístico-computacionais foram, então, explorados para a extração de *features* usadas no treinamento dos modelos de AM baseados em *features*. Com a disponibilização dos modelos neurais pré-treinados para o português – BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020) e RobertaTwitterBr – também realizou-se experimentos com ajuste fino desses modelos. A partir de todo esse trajeto e, em especial, dos experimentos aqui apresentados, conclui-se que é possível

predizer automaticamente a qualidade da argumentação em *tweets* do domínio da política brasileira. Os experimentos realizados evidenciaram que a maneira que se mostrou mais adequada para fazê-lo foi usando o modelo neural gerado a partir do ajuste fino do BERTimbau. Tal modelo foi capaz de predizer com 100% de precisão instâncias da classe de Alta qualidade da argumentação.

6.2 Contribuições

O trabalho apresentado nesta tese propõe e valida uma abordagem computacional para avaliação automática da qualidade da argumentação de mensagens do Twitter. Durante a revisão da literatura, ficou claro que a maioria dos trabalhos voltados à avaliação da qualidade da argumentação são destinados a textos formais, escritos em inglês e, por outro lado, há carência de trabalhos voltados a avaliação da qualidade da argumentação para a língua portuguesa. Assim, uma das contribuições deste trabalho é **tratar o problema da avaliação da qualidade da argumentação em um gênero textual desafiador, o Twitter, e em um idioma pouco explorado, o português do Brasil**. O Twitter foi escolhido pelo alto engajamento dos deputados brasileiros nessa rede social e pela facilidade de uso de sua *Application Programming Interface* (API).

Para tanto foi criado um *cópus* de mensagens a partir de postagens, relacionadas à política, de autoria de deputados brasileiros e as respostas dos seus seguidores. Este *cópus* totaliza 400 mensagens das quais 352 foram classificadas como argumentativas. De forma empírica, foram definidos recursos disponíveis para a língua portuguesa que fossem bons indicadores da qualidade da argumentação no domínio e gênero selecionados, aqui denominados de pistas linguísticas, para que, individualmente ou em conjunto, apoiem a avaliação da qualidade da argumentação na dimensão retórica. Essa dimensão foi escolhida por possuir indícios de aderência com as características das pistas linguísticas. Das 352 mensagens anotadas como argumentativas, 139 foram rotuladas como de baixa argumentatividade, 117 de média argumentatividade e 96 de alta argumentatividade. Dessa forma, para a avaliação da qualidade da argumentação das mensagens coletadas, foi criado um conjunto de diretrizes, conforme apresentado no Apêndice A, direcionadas para a anotação do grau de argumentatividade das mensagens capturadas. Portanto, duas das contribuições deste trabalho são: **o *cópus* de mensagens anotado por 4 anotadores e o conjunto de diretrizes de anotação**.

Embora a tendência atual aponte para o uso de redes neurais artificiais profundas para a solução dos problemas envolvendo língua natural, a primeira das estratégias aqui investigadas foi mais linguisticamente motivada no sentido que visa aplicar uma combinação de pistas linguísticas para a avaliação automática da qualidade da argumentação. Portanto, além de experimentos utilizando redes neurais artificiais profundas, foram realizados experimentos que envolveram a avaliação de cada um dos aspectos da dimensão retórica da

taxonomia de Wachsmuth et al. (2017) utilizando as pistas linguísticas e recursos computacionais desenvolvidos para a língua portuguesa. Para tanto, foram usados algoritmos de AM baseados em *features*, como *Logistic Regression* (LR), *K-Nearest Neighbors* (KNN), *Decision Tree* (DT), SVM, RF e NB – e também um ajuste fino de modelos BERT e RoBERTa. Os experimentos realizados com o BERT alcançaram uma precisão de 100% para as mensagens consideradas com Alta qualidade de argumentação. Enquanto os algoritmos baseados em *features* obtiveram precisões médias que variaram de 32% a 54%. Assim, outra contribuição deste trabalho são os **modelos computacionais treinados para classificar tweets no domínio da política brasileira de acordo com a qualidade da argumentação**.

É importante ressaltar que Wachsmuth e Werner (2020) apresentaram recente trabalho no qual abordam algumas pistas linguísticas para a avaliação geral das dimensões da taxonomia da qualidade da argumentação. Essas pistas incluem (i) comprimento: argumentos mais longos tendem a ser avaliados como mais convincentes e de melhor qualidade; (ii) subjetividade: o uso de pronomes de primeira, segunda e terceira pessoa, bem como palavras que indicam positividade, negatividade e hesitação, pode fornecer pistas sobre a subjetividade do argumento; (iii) conteúdo: a presença de palavras-chave relevantes para o tópico do argumento pode indicar sua qualidade; (iv) emoções: o uso de *emojis* e outras expressões emocionais pode fornecer pistas sobre a qualidade do argumento; e (v) estrutura: a organização e estrutura do argumento, incluindo o uso de conectores e marcadores de discurso, pode indicar sua qualidade. Essas pistas linguísticas foram testadas em um *corpus* existente de argumentos em linguagem natural, e os resultados mostraram que elas podem ser usadas para avaliar a qualidade intrínseca de argumentos com moderado sucesso. No entanto, os autores observam que a avaliação da qualidade de argumentos ainda é, em grande parte, subjetiva e que mais pesquisas são necessárias para melhorar a precisão da avaliação computacional. Porém, o trabalho aqui apresentado se diferencia de Wachsmuth e Werner (2020) por tratar cada um dos aspectos da taxonomia de forma individual com uma ou a combinação de várias pistas linguísticas.

Por fim, também tem-se como contribuição deste trabalho a **interface de usuário** que permite o uso do modelo computacional para a avaliação da qualidade da argumentação em *tweets* do domínio da política brasileira pelo público geral, apresentada no Capítulo 5.

Desse modo, apresenta-se uma síntese das contribuições deste trabalho:

- Pistas linguísticas definidas com base na taxonomia de Wachsmuth et al. (2017) e adaptadas para o gênero textual Twitter.
- Diretrizes de anotação que culminaram no manual de anotação apresentado no Apêndice A.
- *Corpus* de *tweets* do domínio da política brasileira anotado com pistas e qualidade geral da argumentação.

- ❑ Modelo computacional capaz de prever com 100% de precisão os *tweets* de boa qualidade de argumentação.
- ❑ Interface de uso do modelo computacional para avaliação da qualidade da argumentação em *tweets* do domínio da política brasileira.

6.3 Limitações

Entre outras, destacam-se as seguintes limitações deste trabalho:

- ❑ O **tamanho do corpus** é uma das principais restrições deste trabalho. Embora seja compatível com o corpus desenvolvido por Wachsmuth et al. (2017) quanto ao número de exemplos, apenas a dimensão retórica foi avaliada e, para uma avaliação mais aprofundada, os aspectos foram subdivididos em vários critérios. Isso pode ter causado uma dissonância entre o número de exemplos e o número de critérios analisados, pois existem poucos exemplos para um grande número de critérios.
- ❑ Outra limitação está relacionada à decisão de projeto adotada no momento da anotação do corpus que foi a de realizar a **anotação da qualidade da argumentação de forma isolada**, ou seja, desconsiderando o *tweet* semente. Diante disso, assumiu-se que o argumento pode ser classificado independentemente do tópico. Fromm, Faerman e Seidl (2019) e Hidayatullah et al. (2021) apontam que, muitas vezes, frases contendo argumentos são estruturalmente semelhantes a frases puramente informativas e sem qualquer posicionamento sobre o tópico, e que considerar a informação do tópico é crucial para a tarefa, pois ele define o contexto semântico de um argumento. Para lidar com essa limitação, uma possível estratégia é concatenar as mensagens avaliadas com o *tweet* semente de modo a incorporar contexto e, conseqüentemente, melhorar o desempenho do modelo como realizado em Gretz et al. (2020).
- ❑ Outra decisão de projeto que pode ter impactado o desempenho do modelo está relacionada ao **modo como os aspectos foram combinados para definir a Qualidade Geral da argumentação**. Conforme visto no Capítulo 4, algumas instâncias de teste anotadas como de Alta qualidade foram classificadas pelo modelo BERTimbau como de Média ou Baixa qualidade o que, analisando o conteúdo de tais *tweets* pode ter fundamento. Assim, uma proposta de trabalho futuro é fazer uma revisão dos critérios e do modo como os aspectos são combinados para definir a Qualidade Geral.
- ❑ Por fim, uma limitação relacionada ao Twitter é o **número muito limitado de caracteres** (280) permitido para cada mensagem, o que dificulta o uso de estratégias mais elaboradas de argumentação linguística pelos autores das postagens.

6.4 Trabalhos futuros

Como trabalhos futuros, destacam-se três caminhos que trariam maior benefício às propostas aqui apresentadas:

- ❑ Como apresentado na Seção 6.3, concatenar as mensagens avaliadas com o *tweet* somente de modo a incorporar contexto e, conseqüentemente, melhorar o desempenho do modelo.
- ❑ Testar uma combinação (*ensemble*) de classificadores, que constitui-se em múltiplos classificadores, treinados de forma individual, cujos resultados são combinados, também com o objetivo de buscar uma melhora no desempenho do modelo.
- ❑ Aumentar o *cópus* com a finalidade de capturar novos contextos, além de um acréscimo da sua própria dimensão.
- ❑ Gerar modelos específicos para cada um dos aspectos da qualidade da argumentação (Clareza, Organização, Credibilidade e Polaridade e Intensidade do Apelo Emocional) com o intuito de combiná-los para definir automaticamente a qualidade geral da argumentação, como foi feito de modo manual, pelos anotadores, no momento da geração do *cópus*. Neste ponto, vale destacar que experimentos iniciais com AM baseado em *features* foram realizados com esse objetivo, contudo, sem resultados expressivos, possivelmente devido ao pequeno número de instâncias anotadas para cada um dos aspectos.

6.5 Considerações finais

A avaliação da qualidade da argumentação em *tweets* relacionados à política brasileira é um tema relevante e desafiador para a linguística computacional. Diante do cenário atual, caracterizado por uma acentuada polarização política, além da disseminação de *fake-news* e desinformação, se torna relevante o desenvolvimento de ferramentas capazes de analisar as manifestações das redes sociais.

Ainda que os recursos computacionais, como *cópus* de textos argumentativos, sejam limitados e os desafios como a complexidade e diversidade dos gêneros textuais que exigem constante adaptação das técnicas de análise sejam evidentes, os avanços nesse campo de estudo são notáveis.

Os resultados das pesquisas relacionadas à avaliação da qualidade da argumentação são potencialmente positivos para a sociedade, entretanto alguns cuidados merecem atenção. Por um lado, a avaliação da qualidade da argumentação pode contribuir para o desenvolvimento do senso crítico dos usuários do Twitter, que poderiam ter acesso a pareceres sobre a qualidade das suas próprias postagens e das postagens de outros usuários.

No entanto, essas avaliações, se usadas de forma tendenciosa ou ideológica, sem respeito à ética, podem gerar conflitos.

Este trabalho propõe um arcabouço tecnológico que se espera seja útil nesse contexto e sempre usado de maneira ética.

Referências

AL-KHATIB, K. et al. Cross-domain mining of argumentative text through distant supervision. In: **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. San Diego, California: Association for Computational Linguistics, 2016. p. 1395–1404. Disponível em: <<https://www.aclweb.org/anthology/N16-1165>>.

ARISTOTLE; KENNEDY(TRADUTOR), G. **On Rhetoric: A Theory of Civic Discourse**. Oxford University Press, 2007. ISBN 9780195064872. Disponível em: <<https://books.google.com.br/books?id=QQFc4FwQz-wC>>.

BALAGE FILHO, P. P.; PARDO, T. A. S.; ALUÍSIO, S. M. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In: **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology**. [s.n.], 2013. Disponível em: <<https://www.aclweb.org/anthology/W13-4829>>.

BAQUERO, M. Democracia formal, cultura política informal e capital social no brasil. **Opinião Pública**, SciELO Brasil, v. 14, p. 380–413, 2008.

BENCH-CAPON, T. J.; DUNNE, P. E. Argumentation in artificial intelligence. **Artificial intelligence**, Elsevier, v. 171, n. 10-15, p. 619–641, 2007.

BERTAGLIA, T. F. C. **Normalização textual de conteúdo gerado por usuário**. Dissertação (Mestrado) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2017. Disponível em: <<https://teses.usp.br/teses/disponiveis/55/55134/tde-10112017-170919/pt-br.php>>.

BILU, Y.; SLONIM, N. Claim synthesis via predicate recycling. In: **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**. Berlin, Germany: Association for Computational Linguistics, 2016. p. 525–530. Disponível em: <<https://www.aclweb.org/anthology/P16-2085>>.

BOJANOWSKI, P. et al. Enriching word vectors with subword information. **Transactions of the Association for Computational Linguistics**, v. 5, p. 135–146, 2017. Disponível em: <<https://www.aclweb.org/anthology/Q17-1010>>.

BRUM, H.; NUNES, M. das G. V. Building a Sentiment Corpus of Tweets in Brazilian Portuguese. In: CHAIR), N. C. C. et al. (Ed.). **Proceedings of the Eleventh**

International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan: European Language Resources Association (ELRA), 2018. ISBN 979-10-95546-00-9.

DEVLIN, J. et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In: **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. p. 4171–4186. Disponível em: <<https://www.aclweb.org/anthology/N19-1423>>.

EEMEREN, F. H. v.; GROOTENDORST, R. **A Systematic Theory of Argumentation: The pragma-dialectical approach**. [S.l.]: Cambridge University Press, 2003.

FENG, V. W.; LIN, Z.; HIRST, G. The impact of deep hierarchical discourse structures in the evaluation of text coherence. In: **Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers**. Dublin, Ireland: Dublin City University and Association for Computational Linguistics, 2014. p. 940–949. Disponível em: <<https://www.aclweb.org/anthology/C14-1089>>.

FROMM, M.; FAERMAN, E.; SEIDL, T. Tacam: Topic and context aware argument mining. In: **IEEE/WIC/ACM International Conference on Web Intelligence**. New York, NY, USA: Association for Computing Machinery, 2019. (WI '19), p. 99–106. ISBN 9781450369343. Disponível em: <<https://doi.org/10.1145/3350546.3352506>>.

GARCÍA-GORROSTIETA, J. M.; LÓPEZ-LÓPEZ, A. Identifying argumentative paragraphs: Towards automatic assessment of argumentation in theses. In: SILBERZTEIN, M. et al. (Ed.). **Natural Language Processing and Information Systems**. Cham: Springer International Publishing, 2018. p. 83–90. ISBN 978-3-319-91947-8.

GARCÍA-GORROSTIETA, J. M.; LÓPEZ-LÓPEZ, A.; GONZÁLEZ-LÓPEZ, S. Automatic argument assessment of final project reports of computer engineering students. **Computer Applications in Engineering Education**, v. 26, n. 5, p. 1217–1226, 2018. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.21996>>.

GRETZ, S. et al. A large-scale dataset for argument quality ranking: Construction and analysis. In: **Proceedings of the AAAI Conference on Artificial Intelligence**. [S.l.: s.n.], 2020. v. 34, n. 05, p. 7805–7813.

HABERNAL, I.; GUREVYCH, I. Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse. In: **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**. Lisbon, Portugal: Association for Computational Linguistics, 2015. p. 2127–2137. Disponível em: <<https://www.aclweb.org/anthology/D15-1255>>.

_____. Argumentation mining in user-generated web discourse. **Computational Linguistics**, v. 43, n. 1, p. 125–179, abr. 2017. Disponível em: <<https://www.aclweb.org/anthology/J17-1004>>.

- HASSAN, H.; MENEZES, A. Social text normalization using contextual graph random walks. In: **Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**. Sofia, Bulgaria: Association for Computational Linguistics, 2013. p. 1577–1586. Disponível em: <<https://www.aclweb.org/anthology/P13-1155>>.
- HIDAYATURRAHMAN et al. **Enhancing argumentation component classification using contextual language model**. Springer Science and Business Media LLC, 2021. Disponível em: <<http://dx.doi.org/10.1186/s40537-021-00490-2>>.
- LAUSCHER, A. et al. Rhetoric, logic, and dialectic: Advancing theory-based argument quality assessment in natural language processing. In: **Proceedings of the 27th International Conference on Computational Linguistics**. online: [s.n.], 2020.
- LEAL, S. E. **Predição da complexidade sentencial do português brasileiro escrito, usando métricas linguísticas, psicolinguísticas e de rastreamento ocular**. Tese (Doutorado) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2021.
- LEITE, J. A. et al. **Toxic Language Detection in Social Media for Brazilian Portuguese: New Dataset and Multilingual Analysis**. 2020.
- LING, W. et al. Two/too simple adaptations of Word2Vec for syntax problems. In: **Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. Denver, Colorado: Association for Computational Linguistics, 2015. p. 1299–1304. Disponível em: <<https://www.aclweb.org/anthology/N15-1142>>.
- MIKOLOV, T. et al. Distributed representations of words and phrases and their compositionality. In: **Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2**. Red Hook, NY, USA: Curran Associates Inc., 2013. (NIPS'13), p. 3111–3119.
- PELDSZUS, A.; STEDE, M. Joint prediction in MST-style discourse parsing for argumentation mining. In: **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**. Lisbon, Portugal: Association for Computational Linguistics, 2015. p. 938–948. Disponível em: <<https://www.aclweb.org/anthology/D15-1110>>.
- PENNEBAKER, J. W.; FRANCIS, M. E.; BOOTH, R. J. Linguistic inquiry and word count: Liwc 2001. **Mahway: Lawrence Erlbaum Associates**, v. 71, n. 2001, p. 2001, 2001.
- POTTHAST, M. et al. Argument search: Assessing argument relevance. In: **Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval**. New York, NY, USA: Association for Computing Machinery, 2019. (SIGIR'19), p. 1117–1120. ISBN 9781450361729. Disponível em: <<https://doi.org/10.1145/3331184.3331327>>.
- RINOTT, R. et al. Show me your evidence - an automatic method for context dependent evidence detection. In: **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**. Lisbon, Portugal:

- Association for Computational Linguistics, 2015. p. 440–450. Disponível em: <<https://www.aclweb.org/anthology/D15-1050>>.
- SHNEIDERMAN, B. Designing information-abundant web sites: issues and recommendations. **International Journal of Human-Computer Studies**, v. 47, n. 1, p. 5–29, 1997. ISSN 1071-5819. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1071581997901270>>.
- _____. Human-centered artificial intelligence: Reliable, safe & trustworthy. **International Journal of Human-Computer Interaction**, Taylor & Francis, v. 36, n. 6, p. 495–504, 2020. Disponível em: <<https://doi.org/10.1080/10447318.2020.1741118>>.
- SILVA, C. F. d. et al. Quality of argumentation in political tweets: what is and how to measure it / qualidade da argumentação em tweets de política: o que e como avaliar. **Revista de Estudos da Linguagem**, v. 29, n. 4, p. 2537–2586, 2021. ISSN 2237-2083. Disponível em: <<http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/18263>>.
- SKITALINSKAYA, G.; KLAFF, J.; WACHSMUTH, H. Learning from revisions: Quality assessment of claims in argumentation at scale. In: **Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume**. Online: Association for Computational Linguistics, 2021. p. 1718–1729. Disponível em: <<https://aclanthology.org/2021.eacl-main.147>>.
- SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. **Portuguese Named Entity Recognition using BERT-CRF**. 2020.
- STAB, C.; GUREVYCH, I. Parsing argumentation structures in persuasive essays. **Computational Linguistics**, v. 43, n. 3, p. 619–659, set. 2017. Disponível em: <<https://www.aclweb.org/anthology/J17-3005>>.
- _____. Recognizing insufficiently supported arguments in argumentative essays. In: **Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers**. Valencia, Spain: Association for Computational Linguistics, 2017. p. 980–990. Disponível em: <<https://www.aclweb.org/anthology/E17-1092>>.
- TOULMIN, S. E. **The uses of argument**. [S.l.]: Cambridge university press, 2003.
- ULLMAN, J. D.; AHO, A. V.; HIRSCHBERG, D. S. Bounds on the complexity of the longest common subsequence problem. **J. ACM**, Association for Computing Machinery, New York, NY, USA, v. 23, n. 1, p. 1–12, jan. 1976. ISSN 0004-5411. Disponível em: <<https://doi.org/10.1145/321921.321922>>.
- WACHSMUTH, H.; KIESEL, J.; STEIN, B. Sentiment flow - a general model of web review argumentation. In: **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**. Lisbon, Portugal: Association for Computational Linguistics, 2015. p. 601–611. Disponível em: <<https://www.aclweb.org/anthology/D15-1072>>.

WACHSMUTH, H. et al. Computational argumentation quality assessment in natural language. In: **Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers**. Valencia, Spain: Association for Computational Linguistics, 2017. p. 176–187. Disponível em: <<https://www.aclweb.org/anthology/E17-1017>>.

WACHSMUTH, H.; STEIN, B.; AJJOUR, Y. “PageRank” for argument relevance. In: **Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers**. Valencia, Spain: Association for Computational Linguistics, 2017. p. 1117–1127. Disponível em: <<https://www.aclweb.org/anthology/E17-1105>>.

WACHSMUTH, H.; WERNER, T. Intrinsic quality assessment of arguments. In: **Proceedings of the 28th International Conference on Computational Linguistics**. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020. p. 6739–6745. Disponível em: <<https://www.aclweb.org/anthology/2020.coling-main.592>>.

WOLF, T. et al. **HuggingFace’s Transformers: State-of-the-art Natural Language Processing**. 2020.

ZHANG, J. et al. Conversational flow in Oxford-style debates. In: **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. San Diego, California: Association for Computational Linguistics, 2016. p. 136–141. Disponível em: <<https://www.aclweb.org/anthology/N16-1017>>.

Apêndices

APÊNDICE A

Diretrizes de anotação

DIRETRIZES DE ANOTAÇÃO

PROJETO ARGQ!

ESPECIFICAÇÃO DA TAREFA

Esta anotação está sendo realizada dentro do escopo de um projeto de doutorado que visa avaliar a qualidade da argumentação em postagens do Twitter no domínio da política. Nesse contexto, a tarefa de anotação consiste em identificar critérios presentes na postagem que estejam relacionados aos aspectos Clareza, Organização, Credibilidade e Apelo emocional.

Passo-a-passo:

1. Leia o post inicial e a postagem gerada em resposta a ele. Esta postagem é a que será avaliada;
2. Responda se a postagem a ser avaliada está **relacionada ao tema** do post inicial;
3. Responda se a postagem a ser avaliada é ou não **argumentativa**;
4. Se a postagem foi classificada como argumentativa, marque os critérios presentes na postagem, de acordo com o que está definido nessas diretrizes. Sugere-se que a anotação seja realizada para todas as postagens geradas em resposta ao post inicial, por aspecto, e na seguinte ordem:
 - a. Clareza
 - b. Organização
 - c. Credibilidade
 - d. Apelo emocional

Agradecemos sua participação nesta tarefa!

Equipe ArgQ!

RELAÇÃO COM O TEMA

Algumas postagens têm relação óbvia, explícita e direta com o tema do post inicial, como o exemplo a seguir, cujo post inicial trata do trabalho do governo para a concessão do auxílio emergencial (esse é o tema principal) e a postagem a ser avaliada se refere ao recebimento desse auxílio.¹

¹ Vale deixar claro que o destaque em **amarelo** no decorrer desse documento foi inserido pelos autores das diretrizes para enfatizar o ponto em foco no exemplo.

Diretrizes de anotação - Projeto ArgQ!

ASSUNTO	POSTAGEM
Parabéns pelo bellissimo trabalho que o @govbr em parceria com a @Caixa vem realizando para fazer o auxílio emergencial de R\$ 600,00 chegar a quem mais precisa. @jairbolsonaro https://t.co/2sAlF9WTr3	@marcofeliciano @govbr @Caixa @jairbolsonaro É bacana, mas tbm da raiva. Militar e Servidor recebendo ilegalmente e eu, informal, tive o meu cortado. O auxílio é bom, mas a organização é péssima.

Também existem postagens que, apesar de não se referirem ao tema principal do post inicial, tratam de algum tema secundário ou apenas uma parte do tema principal, como nos dois exemplos seguintes que mencionam temas secundários: i) "censura" para se referir à exclusão das contas na rede social, e ii) "notícias falsas" para se referir às postagens que são feitas nas contas que disseminam ódio.

ASSUNTO	POSTAGEM
URGENTE! Facebook removeu diversas contas q disseminavam ódio . Mesmo com os responsáveis tentando se esconder, as investigações ligam as contas com funcionários nos gabinetes de Anderson Moraes, Alana Passos, Eduardo Bolsonaro, Flávio Bolsonaro e - quem diria - Jair Bolsonaro.	@jandira_feghali " Censura pra Eles e Liberdade pra mim" - esse é o Lema dessa turminha de Esquerdopatas que passa 24h caluniando, ofendendo e sabotando o Presidente da República. https://t.co/C94OMZUaT0

ASSUNTO	POSTAGEM
URGENTE! Facebook removeu diversas contas q disseminavam ódio . Mesmo com os responsáveis tentando se esconder, as investigações ligam as contas com funcionários nos gabinetes de Anderson Moraes, Alana Passos, Eduardo Bolsonaro, Flávio Bolsonaro e - quem diria - Jair Bolsonaro.	@jandira_feghali Eles mantém boa parte de seu eleitorado através do bombardeio de notícias falsas , muitos que eu conheço não assistem TV aberta apenas as redes bozoloideas e acreditam em tudo que o "mito" fala! Se cair as fakenews o governo cai rapidinho!

Em alguns casos, a relação entre o tema da postagem e o tema do post inicial não é tão clara. Consideraremos que a postagem a ser avaliada NÃO tem a ver com o tema do post inicial nos casos em que o autor da postagem não faz nenhuma referência ao tema principal nem a nenhum outro tema abordado no post inicial, restringindo-se apenas a comentários sobre a pessoa que fez o post inicial (primeiro exemplo) ou outras pessoas mencionadas no post inicial (segundo exemplo), conforme os exemplos a seguir.

Diretrizes de anotação - Projeto ArgQ!

ASSUNTO	POSTAGEM
Presidente do STJ decretou prisão domiciliar para Fabrício Queiroz. A questão que não quer calar... pra qual casa do Wassef que ele vai? Vai voltar pro sítio de Atibaia?	@kimpkat Kim e um grande imbecil que trabalha contra o povo ,, vc prefere trabalhar junto com corruptos do que com pessoas honestas para tornar o Brasil um país melhor ,, eu me pergunto, o dia que um ditado de fator entrar no poder essa merda Kim que vc tem feito vai servir p/ vc mesmo

ASSUNTO	POSTAGEM
URGENTE! BOLSONARO VETOU VÁRIOS TRECHOS DA LEI QUE TENTA PROTEGER INDÍGENAS E QUILOMBOLAS NA PANDEMIA. Vetou acesso universal à água, desinfecção desses territórios e até repasse de materiais de higiene!!!!	@jandira_feghali Tenho severas dúvidas quanto a quem ganha a disputa para líder mundial mais mau caráter e irresponsável do ano: TRAMPO X BOZO !!! Quem será?!!!

Como anotar?

Considerando a definição e os exemplos apresentados anteriormente, deve-se anotar como:

- SIM (está relacionada ao tema) - quando a postagem faz menção ao tema principal do post inicial.
- PARCIALMENTE - quando a postagem faz menção a parte do tema principal ou a algum tema secundário do post inicial.
- NÃO (não está relacionada ao tema) - quando a postagem não faz menção ao tema principal ou secundário do post inicial, mesmo que mencione o autor do post inicial ou outras pessoas/entidades citadas nele.

ARGUMENTATIVIDADE

As definições mais clássicas de argumentatividade levam em consideração a apresentação organizada de argumentos propriamente ditos. Contudo, para fins dessa anotação, será considerada uma definição muito mais abrangente de "argumentatividade".

No contexto dessa anotação, uma postagem deve ser avaliada como argumentativa se for possível identificar a posição/opinião (seja ela favorável ou contrária) do autor da postagem. Assim, **qualquer tentativa de marcar a opinião**, mesmo que não traga embasamentos comprobatórios para isso, **deve ser considerada argumentativa**.

Diretrizes de anotação - Projeto ArgQ!

Ainda que a argumentação seja ruim, ainda que haja poucos argumentos, ou que não convença o interlocutor, ainda assim, a postagem deve ser avaliada como argumentativa caso haja a opinião ou o posicionamento do autor com relação ao post inicial. Repare que nesta etapa não se deve avaliar a qualidade da argumentação, ou seja, se o autor traz argumentos ou não, mas basicamente se ele expressa seu posicionamento ou não.

Assim, a postagem NÃO deve ser considerada argumentativa quando trazer conteúdo que não marca claramente a opinião do autor, por exemplo, no caso abaixo:

ASSUNTO	POSTAGEM
A educação é o caminho para o desenvolvimento de uma nação. Por isso, o governo federal anunciou um repasse de mais de R\$ 253,9 milhões para as Universidades federais voltadas p/ ações estratégicas e melhorias em infraestrutura. O recurso já é da LOA de 2020. @jairbolsonaro	Peço a ajuda de todos Ela precisa comprar um carrinho de bebê, além de bebê conforto, roupas, alimentos, pagar Aluguel, está desempregada e passando dificuldade, ela ficou viúva qd estava grávida, não era casada no papel, não conseguiu o Seguro https://t.co/f2tucYdtdY https://t.co/kw3nnlyg0M

Como anotar?

Considerando a definição e os exemplos apresentados anteriormente, deve-se:

- Marcar o check-box da coluna "A postagem é argumentativa?" se a postagem traz indícios de sua posição/opinião.
- Caso contrário, o check-box deve permanecer sem marcação. Se a postagem não for argumentativa sua avaliação termina aqui.

CLAREZA

O aspecto Clareza está sendo avaliado com base em critérios que contribuem **negativamente** para a clareza. São eles:

CRITÉRIO	DEFINIÇÃO/ORIENTAÇÃO	EXEMPLO
Presença de pergunta que conduz à dúvida OBS.: Mesmo que o "?" não esteja presente na pergunta, mas contenha algum elemento	Uma frase interrogativa que afeta negativamente a clareza do argumento, não deixando explícita as relações discursivas na superfície textual. Repare que não se trata da	@kimpkat Tenho algumas perguntas também: Quanto o Renan ganha? Pra fazer o que? Quem paga? Pode ou não falar do Maia? O que o MBL articula com Maia? Quem é mais danoso para o PAIS Maia ou a rachadinha do Flávio? (que vá preso se for o caso) .

Diretrizes de anotação - Projeto ArgQ!

<p>interrogativo.</p>	<p>repetição dos sinais (???) e sim de uma pergunta.</p>	<p>@kimpkat Esse presidente é muito escroto! Onde já se viu cumprir promessas e entregar obras públicas a população?? Temo que investigá-lo urgentemente! https://t.co/DUMC6sQ6mI</p> <p>Exemplo de pergunta que NÃO conduz a dúvida</p> <p>@MarceloFreixo Com certeza ele nunca agiu sozinho, isso está cheirando a balão de ensaio, já que o Bolsonaro não pode mais ficar se expondo, como ele sempre teve seus leões de chácara, o jogo dele não vai parar, agora, a questão, o filho e o próprio Bolsonaro cometeram crimes semelhantes, e aí?</p>
<p>Presença de linguagem complexa desnecessária</p>	<p>Presença de palavra muito rebuscada não apropriada para o contexto, ou de estrutura sintática complexa desnecessária e que afeta a compreensão.</p>	<p>@CarlaZambelli38 @danielPMERJ @ArthurLira_ não estou vendo mobilização ou indignação da senhora, o ato supremo é de corar as pedras, é a instituição e o conceito de imunidade que está sendo vilipendiado... Lute pela separação de poderes com independência</p> <p>@CarlaZambelli38 @camaradeputados @danielPMERJ Câmara dos Dep. prostado de cócoras para o STF. Os pares de Daniel Silveira, jogando ele aos leões. Mas, um dia, ele voltará; Forte e maior. Muito maior do que os pusilânimes. que hj estão na câmara.</p>
<p>Presença de erros (ortográficos, sintáticos, de pontuação, etc.) de uso da Língua Portuguesa que atrapalham o entendimento</p>	<p>A boa qualidade de Língua Portuguesa (LP) identificada pelo uso correto de pontuação, sintaxe, ortografia etc. contribui positivamente para que o argumento esteja claro. Assim, a clareza do argumento é penalizada na presença de erros que atrapalham o entendimento. Contudo, tenha em mente que o Twitter tem restrição de caracteres, por isso,</p>	<p>Exemplo de postagem que não deve ser penalizada:</p> <p>@jandira_feghali GENOCIDA! Esse ser é de uma maldade tão absurda, que é impressionante que consiga dormir. Sinceramente, se ele realmente estiver doente (coisa que não acredito), não quero q a doença o mate. Q ele fique bem vivo p ser julgado e condenado pelos crimes q comete contra a humanidade</p> <p>Exemplo de postagem que deve ser penalizada:</p>

Diretrizes de anotação - Projeto ArgQ!

	algumas palavras são abreviadas propositalmente pelos usuários. Não penalize esses casos em qualidade de LP.	@carlosjordy Porque os debatedores que defendem a oposição não respeitam o debate igual os defensores do governo, ontem vi o copola acalmando o outro aí a mediadora acusa o copola de desrespeito foi o fim da picada
Presença de desvio desnecessário do assunto	Desvios do assunto que impactam negativamente a clareza da argumentação devem ser penalizados. Esse critério deve ser anotado quando houver um desvio desnecessário total ou parcial do tema principal ou secundário.	@kimpkat PARABÉNS VC É FODA,A TODA PODEROSA GLOBO LIXO INVESTIGOU E NADA PROVOU E AGORA O JASPION CONSEGUIU PROVAR A RACHADINHA DO BOLSONARO. O JASPION ESTOU ACHANDO QUE VC E O HOLIDAY JOGAM NO MESMO TIME, SE É QUE VC ME ENTENDE.

Como anotar?

Para cada um dos critérios listados na tabela anterior, deve-se:

- Marcar o check-box na coluna correspondente a esse critério se o mesmo estiver presente na postagem;
- Caso contrário, o check-box deve permanecer sem marcação.

ORGANIZAÇÃO

O aspecto Organização está sendo avaliado com base em critérios que contribuem **positivamente** para a organização. Sugere-se que esses critérios sejam avaliados e anotados na seguinte ordem, ou seja, se tiver dúvida entre o primeiro e o segundo critério, o primeiro deve ser anotado:

CRITÉRIO	DEFINIÇÃO/ORIENTAÇÃO	EXEMPLO
Presença de relação condicional Apenas: caso, se	Presença de estrutura que apresenta uma relação condicional dentro de um mesmo argumento. Veja que nem sempre o “então” do se-então precisa estar presente, mas é necessário que o marcador da condicional esteja explícito.	@jandira_feghali GENOCIDA! Esse ser é de uma maldade tão absurda, que é impressionante que consiga dormir. Sinceramente, se ele realmente estiver doente (coisa que não acredito), não quero q a doença o mate. Q ele fique bem vivo p ser julgado e condenado pelos crimes q comete contra a humanidade

Diretrizes de anotação - Projeto ArgQ!

<p>Presença de relação concessiva</p> <p>Ex: ainda assim, a menos que, ainda que, desde que, mesmo que, mesmo assim, se bem que</p>	<p>Presença de relação concessiva que está diretamente ligada à quebra de uma expectativa.</p>	<p>@carlosjordy Hã? Bolsonaro não ia para os braços do povo antes da pandemia? (FAKE NEWS) Mesmo desmentido pela jornalista, que lembrou que Bolsonaro tomou a facada nos braços do povo, ficou balançando a cabeça, como que reprovando a fala da colega. Ele é ridículo e mentiroso.</p>
<p>Presença de relação de oposição ou contraste</p>	<p>Presença de estrutura que apresenta uma oposição de duas ideias dentro de um mesmo argumento. Não é necessário que o marcador discursivo esteja explícito; basta haver a relação semântica de oposição entre duas ideias.</p>	<p>@marcofeliciano @govbr @Caixa @jairbolsonaro É bacana, mas tbn da raiva. Militar e Servidor recebendo ilegalmente e eu, informal, tive o meu cortado. O auxílio é bom, mas a organização é péssima.</p> <p>@marcelvanhattem @camaradeputados Pr @ArthurLira_, entenda, não são os ministros do STF que elege um deputado ou senador, são os eleitores! 2022 é ano q vem, quem votou contra o dep. Daniel Silveira, será lembrado como TRAIADOR! A lista dos 364 já foi divulgada! O relógio corre e faz tic, tac, tic, tac, tic... https://t.co/PmtCmbdYGN</p>
<p>Presença de relação comparativa ou proporcional</p> <p>Ex: assim como, da mesma forma que, mal dá, nem um... nem outro, quanto mais... melhor, quem dirá</p>	<p>Presença de estrutura comparativa ou proporcional entre dois ou mais elementos. Os marcadores discursivos geralmente estão presentes, mas isso não é necessário para que se marque essas relações.</p>	<p>@RodrigoMaia Agora só falta criar mais meses no ano para que os empresários consigam pagar os empréstimos e encargos. Pq só com a renda mensal mal dará para pagar as contas ordinárias, quem dirá empréstimos!</p> <p>@carlosjordy Porque os debatedores que defendem a oposição não respeitam o debate igual os defensores do governo, ontem vi o copola acalmando o outro aí a mediadora acusa o copola de desrespeito foi o fim da picada</p>
<p>Presença de relações de causa e efeito, explicação, restrição, finalidade</p>	<p>Presença de estrutura que apresenta uma relação de causa e efeito.</p>	<p>@carlosjordy Já que a esquerda é só paz e amor, vamos pegar todas as postagens da esquerda e recriar ela mudando o nome do Bolsonaro para o do STF. Mas tire o print para caso precise apresentar provas.</p>

Diretrizes de anotação - Projeto ArgQ!

<p>Ex: a fim de, como consequência, dado que, devido a, isto é, já que, logo, ou seja, para, pois, por, por conta de, por isso, porque, portanto, posto que, visto que</p>	<p>Presença da relação de explicação (traduzida ou não por orações coordenadas ou subordinadas).</p> <p>Presença de relação de finalidade</p>	<p>@mariadorosario E pior fingir entregar um auxílio, que leva o nome de emergencial, mas que tem funcionado mal ,muito mal. Não chega as mãos de nós pobres, que somos a maioria da população. Nunca quis um auxílio do governo...e agora aqui De verdade preciso ele também não chega!</p> <p>@DeputadoFederal @SrgiodeLaRocque Estamos vivendo cenas ridículas na política do país. Ministérios horrorosos, um “congresso” composto, em sua maioria, de cidadãos covardes e “lentos” por conveniências pessoais, interesses escusos; um judiciário conivente com milicianos meliantes. Estamos além do fundo do poço.</p>
<p>Presença de relação de encadeamento cronológico ou enumerações</p> <p>Ex: ainda, até, depois, enquanto, quando, quando usados com sentido temporal e sequência de vírgulas e “e” (em ordem lógica) com no mínimo 3 elementos</p>	<p>Presença de encadeamento cronológico, em que se verifica a sequência dos eventos no decorrer do tempo. O marcador discursivo dessa relação não precisa estar explícito. Geralmente se verifica o encadeamento cronológico quando a sequência de frases não pode ser alterada aleatoriamente sem que a argumentação seja prejudicada.</p> <p>Presença de sequências de enumerações, desde que tenha pelo menos 3 elementos enumerados. Aplica-se à enumeração de eventos/proposições, mas também à enumeração de elementos.</p>	<p>@kimpkat Vamos supor que bolsonaro cai. Vc acha que mbl, frota, joice etc. Vcs vão voltar ao poder. Kkkk direita que ficar com Bolsonaro não vota em vcs e nem a esquerda. Vão volta por esquecimento. Só ver engajamento hj de vcs na rede sociais. Lembre se vcs foram eleito pela rede</p> <p>@mariadorosario O que ele fez pra ser "genocida"??? Conte aí, ele mandou dinheiro pra Cuba, Venezuela, desviou dinheiro, mandou dinheiro pra outras Ditaduras? Se fosse chamar de governo e roubar, graças a Deus que ele não faz https://t.co/9K0Jt42gDp</p> <p>@DeputadoFederal @SrgiodeLaRocque Estamos vivendo cenas ridículas na política do país. Ministérios horrorosos, um “congresso” composto, em sua maioria, de cidadãos covardes e “lentos” por conveniências pessoais, interesses escusos; um judiciário conivente com milicianos meliantes. Estamos além do fundo do poço.</p>
<p>Presença de relação de exemplificação</p> <p>Ex: como, igual a, por exemplo, que nem</p>	<p>Presença de estrutura exemplificativa, ou seja, quando cita algum exemplo para reforçar o argumento. O marcador discursivo pode ou não estar explícito</p>	<p>@gleisi Como o Dória esta fazendo aqui em SP, com esses acampamentos aos amigos empresários e levando Bruno junto, ai ninguém fala nada, mas nessa estou com a Gleisi se é o Pt já estariam pedindo impeachemet.???</p>

Diretrizes de anotação - Projeto ArgQ!

	na relação de exemplificação.	
--	-------------------------------	--

Como anotar?

Para cada um dos critérios listados na tabela anterior, deve-se:

- Marcar o check-box na coluna correspondente a esse critério se o mesmo estiver presente na postagem, mesmo que em trechos pequenos;
- Caso contrário, o check-box deve permanecer sem marcação.

CREDIBILIDADE

O aspecto Credibilidade está sendo avaliado com base em critérios que contribuem **positivamente** para a credibilidade. São eles:

CRITÉRIO	DEFINIÇÃO/ORIENTAÇÃO	EXEMPLO
Presença de dado numérico	A presença de algum dado numérico, independente de você julgar ou saber que é verdadeiro ou falso, afeta positivamente a credibilidade do argumento.	@marcofeliciano O governo Bolsonaro trás lucro para Petrobrás de 102 bilhões de reais . Dilma deixou um deficit de -32 bilhões em 2015 . Tira milhões de brasileiros da extrema pobreza com o auxílio emergencial de 600 até 1800 reais. É o governo que mais faz em meio a maior crise do século.
Presença de fato midiático, fato histórico ou fato enciclopédico	A presença de algum fato veiculado pela mídia ou que cite períodos históricos ou fatos enciclopédicos, afeta positivamente a credibilidade do argumento.	@RodrigoMaia E parabéns ao governo Bolsonaro que está entregando equipamentos e medicamentos aos estados e municípios de forma a atender a todos, infelizmente, muitos estados envolvidos em esquemas de desvios de recursos que deveriam ir para a saúde .
		@DeputadoFederal O poder judiciário, foi forjado na elite portuguesa, bancando brancos, ricos, intelectuais de posse. Com a democratização da educação em períodos progressistas anteriores, tivemos o sonho de poder enxergar algo diferente, mas, ainda há muito o que mudar.
		@carlosjordy Hã? Bolsonaro não ia para os

Diretrizes de anotação - Projeto ArgQ!

		braços do povo antes da pandemia? (FAKE NEWS) Mesmo desmentido pela jornalista, que lembrou que Bolsonaro tomou a facada nos braços do povo, ficou balançando a cabeça, como que reprovando a fala da colega. Ele é ridículo e mentiroso.
Presença de figura de autoridade	A presença de alguma citação direta ou indireta a uma pessoa que tenha sido mencionada como se fosse uma autoridade, contribui positivamente para a credibilidade do argumento.	Gostei muito da decisão do Dr. Augusto, sensata e madura. Não há como estabelecer diálogo com extremistas. "O mundo atual é pobre em interrupções". Byung Chul Han Filósofo sul Coreano. @Biakicis @ArthurLira_ Acredito que a única coisa que deveria ser incluída no art. 53 cf é a definição de "qualsquer", isso evitaria que pessoas que tenham menos capacidade intelectual consigam compreender o sentido. Talvez Mauricio de Souza possa inserir um quadrinho na CF, explicando. Seria lúdico.
Presença de hashtag (#) que reforça uma posição	A presença de alguma hashtag, desde que reforce a posição do autor, contribui positivamente para a credibilidade do argumento.	@jandira_feghali @augustodeAB Todos esquerdistas fazem isso, quando não têm argumentos! Fogem! Acusam de facista saem corridos, como o diabo foge da cruz! ? #FechadoComBolsonaroAte2026
Presença de termo especializado	A presença de algum termo especializado de alguma área do conhecimento afeta positivamente a credibilidade do argumento. Nesse caso, não se deve fazer juízo sobre a maneira (adequada ou não) como esse termo foi usado.	@jandira_feghali Isso pra mim se chama "ocultação de provas", "obstrução de justiça", resolveram fazer isso agora que a CPI das Fake News resolvem avançar? Isso pra mim é mais um crime!
Presença de relato de experiência individual	A presença de algum relato de experiência pessoal em primeira pessoa (singular ou plural), impacta positivamente a credibilidade do argumento. O relato deve remeter a uma construção	Tenho Dois Colaboradores Venezuelanos! Amam a Venezuela e choram por seu povo! Para eles , Maduro é um Monstro! Quando ouviram que Trump queria a cabeça desse sujeito, fizeram festa! @marcofeliciano @govbr @Caixa @jairbolsonaro É bacana, mas tbm da raiva.

Diretrizes de anotação - Projeto ArgQ!

	narrativa e não apenas fazer menção à opinião do autor.	Militar e Servidor recebendo ilegalmente e eu, informal, tive o meu cortado. O auxílio é bom, mas a organização é péssima.
--	---	--

Como anotar?

Para cada um dos critérios listados na tabela anterior, deve-se:

- Marcar o check-box na coluna correspondente a esse critério se o mesmo estiver presente na postagem;
- Caso contrário, o check-box deve permanecer sem marcação.

APELO EMOCIONAL

O aspecto Apelo Emocional está sendo avaliado com base em critérios que indicam a polaridade e a intensidade do apelo emocional.

POLARIDADE

A Polaridade do Apelo Emocional está sendo avaliada com base em critérios que contribuem negativamente ou positivamente para a qualidade geral da argumentação. São eles:

CRITÉRIO	EXEMPLO
Presença de referência pejorativa a uma pessoa ou entidade	@jandira_feghali Tenho severas dúvidas quanto a quem ganha a disputa para líder mundial mais mau caráter e irresponsável do ano: TRAMPO X BOZO!!! Quem será?!!!
	@gleisi Infelizmente é a pura verdade, esse desgoverno não se preocupa com quem NECESSITA dessa ajuda neste momento, só em inventa #fakenews enganar o gado e salvar os filhos do @jairbolsonaro de serem presos, se tivessem um mínimo de respeito pelos pobres o povo não estava passando fome
Presença de xingamento ou palavra de baixo calão	@kimpkat Kim e um grande imbecil que trabalha contra o povo ,, vc prefere trabalhar junto com corruptos do que com pessoas honestas para tornar o Brasil um país melhor ,, eu me pergunto, o dia que um ditado de fator entrar no poder essa merda Kim que vc tem feito vai servir p/ vc mesmo
	@jandira_feghali Esse genocida quer mesmo dizimar populações. Nem a gripezinha vai dar jeito nele. Esse homem é doente, psicopata. Maldito!! #BolsonaroGenocida

Diretrizes de anotação - Projeto ArgQ!

<p>Presença de discurso de ódio (inclui ameaça)</p> <p>OBS.: O “discurso de ódio” tem como base um coletivo de raça, cor, gênero ou qualquer outro grupo polarizado.</p>	<p>@jandira_feghali Esse genocida quer mesmo dizimar populações. Nem a gripezinha vai dar jeito nele. Esse homem é doente, psicopata. Maldito!! #BolsonaroGenocida</p>
<p>Presença de expressão que denota especulação</p> <p>OBS.: Entende-se por “especulação” a pressuposição, suposição acerca de alguma coisa, sem comprovação. Em geral, suposições que fazem uma previsão para o futuro podem ser consideradas especulativas.</p>	<p>@marcofeliciano @jairbolsonaro O problema não é o dinheiro, é como o dinheiro está sendo aplicado. As universidades estão totalmente contaminadas pela esquerda suja, tem que combater isso.</p> <p>@gleisi "Apenas 10% do auxílio financeiro prometido pelo governo federal aos trabalhadores e às empresas via PESE foi distribuído até junho" Apenas 10% O governo federal irá matar mais que a pandemia https://t.co/xSiGdQUHqa</p> <p>@CarlaZambelli38 @danielPMERJ @ArthurLira_ O povo sendo obrigado a tomar vacinas fajutas p dar altos lucros, aos marginais responsáveis p sua aquisição e distribuição. O STF cagando sobre o país, povo brasileiro, tds as leis vigentes n Brasil, depois s limpando c a bandeira nacional. NADA E NINGUÉM P FREAR OS DESMANDOS!</p>
<p>Presença de referência cordial a uma pessoa ou entidade (mesmo considerando a ironia)</p>	<p>@jandira_feghali @augustodeAB Gostei muito da decisão do Dr. Augusto, sensata e madura. Não há como estabelecer diálogo com extremistas. "O mundo atual é pobre em interrupções". Byung Chul Han Filósofo sul Coreano.</p>
<p>Presença de linguagem polida, educada</p> <p>OBS.: Por exemplo, se realizando com modalizadores (verbos modais e outras estruturas).</p>	<p>@RodrigoMaia E parabéns ao governo Bolsonaro que está entregando equipamentos e medicamentos aos estados e municípios de forma a atender a todos, infelizmente, muitos estados envolvidos em esquemas de desvios de recursos que deveriam ir para a saúde.</p> <p>@gleisi @gleisi que o @ptbrasil está fazendo pelas eleições municipais? O problema está aí, vencer eleições e a solução! Desculpe cobrar... pq nao vejo nenhuma movimentação em minha cidade</p>

Como anotar?

Para cada um dos critérios listados na tabela anterior, deve-se:

- Marcar o check-box na coluna correspondente a esse critério se o mesmo estiver presente na postagem;
- Caso contrário, o check-box deve permanecer sem marcação.

Diretrizes de anotação - Projeto ArgQ!

Observação

- Não se pode marcar dois check-boxes para o mesmo trecho. Por exemplo, "genocida" não pode ser ao mesmo tempo xingamento e discurso de ódio.

INTENSIDADE

A Intensidade do Apelo Emocional está sendo avaliada com base em critérios que contribuem para aumentar a intensidade do apelo emocional. São eles:

CRITÉRIO	EXEMPLO
Presença de marca de primeira pessoa (singular ou plural)	@marcofeliciano Pensei que a última safra de soja não havia sido tão boa este ano, pois pelo menos em minha região, aqui no norte do RS, a estiagem foi bem forte. Mas que bom que foi uma grande safra!
Presença de repetição de sinais de pontuação (???, ou !!!)	@jandira_feghali Esse genocida quer mesmo dizimar populações. Nem a gripezinha vai dar jeito nele. Esse homem é doente, psicopata. Maldito!! #BolsonaroGenocida
	@gleisi @jairbolsonaro Liberação rápida??? Pra mim se deve, tem que apodrecer na cadeia...não foi a sua corja que acabou com prisão em 2 instância? Agora senta e chora porra, os corruptos do PT inclusive o chefe da quadrilha não está solto? O luladrao? Agora aguenta...pau que da em Xico a em francisco
Presença de estrutura enfática como palavra inteira em caixa alta, repetição, grifo, aspas	@CarlaZambelli38 @ladyfontenelle Se estão tão preocupados com as crianças, POR QUE ESTÃO REPOSTANDO UM VÍDEO QUE JÁ FOI APAGADO, HÁ MUITO TEMPO? Na época em q o canal do cara NÃO ERA PRA CRIANÇA!
Presença de frase imperativa , ou palavra de ordem OBS.: "Palavra de ordem" é uma expressão ou frase curta utilizada em protestos ou manifestações, normalmente replicada em cartazes e discursos, com o objetivo de marcar uma posição, reivindicar alguma mudança e/ou incitar os ânimos do grupo.	@kimpkat Esse presidente é muito escroto! Onde já se viu cumprir promessas e entregar obras públicas a população?? Temo que investigá-lo urgentemente! https://t.co/DUMC6sQ6ml
	@carlosjordy Já que a esquerda é só paz e amor, vamos pegar todas as postagens da esquerda e recriar ela mudando o nome do Bolsonaro para o do STF. Mas tire o print para caso precise apresentar provas.
	@mariadorosario O choro é livre. Chora mais antifa! Estamos comovidos... Se precisar te mando um lencinho. Maria do Putário, tu envergonha nosso Rio Grande! Vc é ridicularizada nacionalmente por suas neuras esquerdopatas! Aproveite seu último mandato!
	@mariadorosario Bolsonaro nega a ciência, não investe na educação, áreas cruciais para salvar vidas. Não tem compromisso com o povo! Estamos juntos, é #ForaBolsonaro

Diretrizes de anotação - Projeto ArgQ!

Presença de superlativo ou expressão que denota exagero como “sempre”, “nunca”, “todo mundo”	@DeputadoFederal Diziam que com Bolsonaro tudo ia mudar para melhor, mas pelo que vejo, mudou para pior. Bolsonaro não gosta de negro de pardo de índio. Ele só gosta do branco que comete todo erro que estamos presenciando e ninguém diz nada , até a justiça está do lado deles e dos filhos dele.
	@mariadorosario O choro é livre. Chora mais antifa! Estamos comovidos... Se precisar te mando um lencinho. Maria do Putário, tu envergonha nosso Rio Grande! Vc é ridicularizada nacionalmente por suas neuras esquerdopatas! Aproveite seu último mandato!
Presença de sentimento expresso por linguagem não-verbal (emoji, interjeição, onomatopeia)	@carlosjordy Bolsonaro quando morava aqui na Barra no condomínio do lado ao meu, vivia na praia, tirava fotos com todo mundo e sempre abraçou o povo. A foto de perfil do chaveiro aqui do condomínio prova bem isso. ❤️???? https://t.co/SMTnyEPV49
	@kimpkat Vamos supor que bolsonaro cai. Vc acha que mbl, frota, joice etc. Vcs vão voltar ao poder. Kkkk direita que ficar com Bolsonaro não vota em vcs e nem a esquerda. Vão volta por esquecimento. Só ver engajamento hj de vcs na rede sociais. Lembre se vcs foram eleito pela rede
Presença de expressão idiomática, frase feita, metáfora, provérbio	@marcofeliciano @jairbolsonaro O PT sabe bem disso e criou o PROUNI seu pulha. Queremos saber onde está a MULHER de QUEIROZ! Para de lançar cortina de fumaça . https://t.co/owC2QicJsv
	@DeputadoFederal @SrgiodeLaRocque Estamos vivendo cenas ridículas na política do país. Ministérios horrorosos, um “congresso” composto, em sua maioria, de cidadãos covardes e “lentos” por conveniências pessoais, interesses escusos; um judiciário conivente com milicianos meliantes. Estamos além do fundo do poço .
	@gleisi @jairbolsonaro Liberação rápida??? Pra mim se deve, tem que apodrecer na cadeia...não foi a sua corja que acabou com prisão em 2 instância? Agora senta e chora porra, os corruptos do PT inclusive o chefe da quadrilha não está solto? O luladrao? Agora aguenta... pau que da em Xico a em francisco
	@gleisi @jairbolsonaro Tia Gleisi, porque não fala com Alexandre para o Queiroz voltar para a penitenciária? Esqueça esse negócio de prisão após transitado em julgado. Bandido bom é bandido preso .
	@marcelvanhattem @camaradeputados Acho que o foco deveria ser nas desastrosas intervenções nas estatais (que já n deveriam ser mais estatais) e nos preços de energia, que logo se avizinham. N adianta ser um leão contra a prisão (q deveria ter discurso técnico), e um gato contra o desastre das políticas do gov.!

Diretrizes de anotação - Projeto ArgQ!

Como anotar?

Para cada um dos critérios listados na tabela anterior, deve-se:

- Marcar o check-box na coluna correspondente a esse critério se o mesmo estiver presente na postagem;
- Caso contrário, o check-box deve permanecer sem marcação.

Observação

- Não se pode marcar dois check-boxes para o mesmo trecho no mesmo grupo de critérios (Polaridade ou Intensidade), mas pode marcar se forem critérios de grupos diferentes. Por exemplo, "farinha do mesmo saco" pode ser marcado como referência pejorativa (Polaridade) e expressão idiomática (Intensidade).

APÊNDICE B

Ferramenta de anotação

Arg Q!

ID	POST INICIAL	POSTAGEM PARA AVALIAÇÃO	A postagem está relacionada ao assunto?	A postagem é argumentativa?	CLAREZA				ORGANIZAÇÃO									
					Pergunta que conduz à dúvida	Linguagem complexa desnecessária	Erros de uso da Língua Portuguesa que atrapalham o entendimento	Desvio desnecessário do assunto	AUTOMÁTICO	Condicional (caso, se)	Concessiva (ex: ainda assim, a menos que, ainda que, desde que, mesmo que, mesmo assim, se bem que)	Oposição ou Contraste	Comparativa ou Proporcional (ex: assim como, da mesma forma que, mal dá ... quem dirá, nem um... nem outro, quanto mais... melhor, quem dirá)	Causa e efeito, Explicação, Restrição, Finalidade (ex: a fim de, como consequência, dado que, devido a, isto é, já que, logo, ou seja, para, pois, por, por conta de, por isso, porque, portanto, posto que, visto que)	Encadeamento cronológico (ex: ainda, até, depois, enquanto, quando, quando usados com sentido temporal) ou Enumerações (seqüência de vírgulas e "e" (em ordem lógica) com no mínimo 3 elementos)	Exemplificação (ex: como, igual a, por exemplo, que nem)	AUTOMÁTICO	
1	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ Acredito que a única coisa que deveria ser incluída no art. 53 CF é a definição de "quaisquer", isso evitaria que pessoas que tenham menos capacidade intelectual consigam compreender o sentido. Talvez Mauricio de Souza possa inserir um quadrinho na CF, explicando. Seria lúdico.	Sim	☑	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Alta	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Média
2	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ parabéns a todos q desde o início se opuseram a essa vergonhosa PEC. tem q acabar c imunidade parlamentar e foro privilegiado, os eleitos são só pessoas q receberam mandato através do voto democrático p trabalhar p o cidadão q e o elegeu e paga o seu salário e só	Sim	☑	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Alta	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Média
3	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ Será que foi esta razão mesmo? Ou pq a opinião geral era de que a esquerda queria uma PEC nao para proteger a imunidade mas sim a IMPUNIDADE parlamentar? E de que adianta criar um novo dispositivo quando o STF desrespeita o que existe? O que vai garantir que etes vão recuar?	Sim	☑	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Alta	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Média
4	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ pretsa atenção! não preciso é seguir no STF ja ter fofoqueiro que sempre e cuidado! não seguir esquerda e sair agora, o fim de esquerda vai ter a morte (Vacina) os perigosos de mídia, é muito cuidado! so preparar esta certo.	Sim	☑	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Média	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Baixa
5	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ Rasgaram a constituição do artigo 53 pra fazer a vontade dos picareta do stf, agora tão com medo de irem pra cadeia, bem feito pra esses picaretas do congresso, agora querem se blindar pra não irem pra cadeia, covardes	Sim	☑	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Alta	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Alta

ID	POST INICIAL	POSTAGEM PARA AVALIAÇÃO	A postagem está relacionada ao assunto?	A postagem é argumentativa?	APELO EMOCIONAL														
					POLARIDADE						INTENSIDADE								
					Referência pejorativa a uma pessoa ou entidade	Xingamento ou Palavra de baixo calão	Discurso de ódio (inclui ameaça)	Expressão que denota especulação	Referência cordial a uma pessoa ou entidade (mesmo considerando a ironia)	Linguagem polida, educada (ex: verbos modais e outras estruturas)	AUTOMÁTICO - POLARIDADE	Marca de primeira pessoa (singular ou plural)	Repetição de sinais de pontuação (?? ou !!)	Estrutura enfática como palavra inteira em caixa alta, repetição, grito, aspas	Frase imperativa ou Palavra de ordem	Superlativo ou Expressão que denota exagero (ex: "sempre", "nunca", "todo mundo")	Sentimento expresso por linguagem não-verbal (emoji, interjeição, onomatopéia)	Expressão idiomática ou Frase feita ou Metáfora ou Provérbio	AUTOMÁTICO - INTENSIDADE
1	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ Acredito que a única coisa que deveria ser incluída no art. 53 cf é a definição de "quaisquer", isso evitaria que pessoas que tenham menos capacidade intelectual consigam compreender o sentido. Talvez Mauricio de Souza possa inserir um quadrinho na CF, explicando. Seria lúcido.	Sim	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Positiva	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Média	
2	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ parabéns a todos q desde o inicio se opuseram a essa vergonhosa PEC. tem q acabar c imunidade parlamentar e foro privilegiado. os eleitos são só pessoas q receberam mandato através do voto democrático p trabalhar p o cidadão q e o elegeu e paga o seu salário e só	Sim	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Neutra	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Média
3	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ Será que foi esta razão mesmo? Ou pq a opinião geral era de que a esquerda queria uma PEC nao para proteger a Imunidade mas sim a IMPUNIDADE parlamentar? E de que adianta criar um novo dispositivo quando o STF, desrespeita o que existe? O que vai garantir que etes vão recuar?	Sim	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Negativa	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Média
4	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ pretsa atenção! não preciso é seguir no STF ja ter fofoqueiro que sempre e cuidado! não seguir esquerda e sair agora. o fim de esquerda vai ter a morte (Vacina). os perigosos de midia. é muito cuidado! so preparar esta certo.	Sim	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Negativa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Média
5	O Presidente @ArthurLira_ após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ Rasgaram a constituição do artigo 53 pra fazer a vontade dos picareta do stf, agora tão com medo de irem pra cadeia, bem feito pra esses picaretas do congresso, agora querem se blindar pra não irem pra cadeia, covardes	Sim	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Negativa	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	Alta

Arg Q!		POSTAGEM PARA AVALIAÇÃO	A postagem está relacionada ao assunto?	A postagem é argumentativa?	NOTAS						QUALIDADE GERAL DA ARGUMENTAÇÃO		
					clareza	organização	credibilidade	apelo emocional - polaridade	apelo emocional - intensidade	apelo emocional	NOTA	AUTOMÁTICO	
ID	POST INICIAL												
1	O Presidente @ArthurLira_ ,após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ Acredito que a única coisa que deveria ser incluída no art. 53 cf é a definição de "quaisquer", isso evitaria que pessoas que tenham menos capacidade intelectual consigam compreender o sentido. Talvez Mauricio de Souza possa inserir um quadrinho na CF, explicando. Seria lúdico.	Sim	<input checked="" type="checkbox"/>		3	2	1	1	2	2	8	Alta
2	O Presidente @ArthurLira_ ,após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ parabéns a todos q desde o inicio se opuseram a essa vergonhosa PEC. tem q acabar c imunidade parlamentar e foro privilegiado. os eleitos são só pessoas q receberam mandato através do voto democrático p trabalhar p o cidadão q e o elegeu e paga o seu salário e só	Sim	<input checked="" type="checkbox"/>		3	2	1	0	2	1	7	Alta
3	O Presidente @ArthurLira_ ,após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ Será que foi esta razão mesmo? Ou pq a opinião geral era de que a esquerda queria uma PEC nao para proteger a lmunidade mas sim a IMPUNIDADE parlamentar? E de que adianta criar um novo diapositivo quando o STF desrespeita o que existe? O que vai garantir que eles vão recuar?	Sim	<input checked="" type="checkbox"/>		3	2	1	-1	2	-2	4	Baixa
4	O Presidente @ArthurLira_ ,após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ pretsa atenção! não preciso é seguir no STF ja ter fofoqueiro que sempre e cuidado! não seguir esquerda e sair agora. o fim de esquerda vai ter a morte (Vacina). os perigosos de mídia. é muito cuidado! so preparar esta certo.	Sim	<input checked="" type="checkbox"/>		2	1	1	-1	2	-2	2	Baixa
5	O Presidente @ArthurLira_ ,após ouvir a maioria dos líderes, houve por bem submeter a PEC 03/21 a uma comissão especial. Não será submetida à CCJ, pois a admissibilidade já foi aprovada. Foi a melhor solução para que possamos discutir melhor o texto.	@Biakicis @ArthurLira_ Rasgaram a constituição do artigo 53 pra fazer a vontade dos picareta do stf, agora tão com medo de irem pra cadeia, bem feito pra esses picaretas do congresso, agora querem se blindar pra não irem pra cadeia, covardes	Sim	<input checked="" type="checkbox"/>		3	3	1	-1	3	-3	4	Baixa

Anexos

ANEXO A

Exemplo das métricas obtidas pelo NILC-Matrix

NILC-Metrix

Switch to English

NILC-Metrix agrupa as métricas desenvolvidas em mais de uma década no NILC, iniciadas com o Coh-Metrix-Port (uma adaptação da ferramenta Coh-Metrix para o Português Brasileiro). O foco principal das métricas é calcular coesão, coerência e nível de complexidade textual.

Essa versão disponibiliza 200 métricas, detalhadas aqui.

Resultados

	Grupo	Métrica	Descrição	Valor
1	Coesão Referencial	adj_arg_ovl	Quantidade média de referentes que se repetem nos pares de sentenças adjacentes do texto	2.0
2	Coesão Referencial	adj_cw_ovl	Quantidade média de palavras de conteúdo que se repetem nos pares de sentenças adjacentes do texto	0.0
3	Coesão Referencial	adj_stem_ovl	Quantidade média de radicais de palavras de conteúdo que se repetem nos pares de sentenças adjacentes do texto	0.0
4	Coesão Referencial	adjacent_refs	Média das proporções de candidatos a referentes na sentença anterior em relação aos pronomes pessoais do caso reto nas sentenças	0.0
5	Coesão Referencial	anaphoric_refs	Média das proporções de candidatos a referentes nas 5 sentenças anteriores em relação aos pronomes anafóricos das sentenças	0.0
6	Coesão Referencial	arg_ovl	Quantidade média de referentes que se repetem nos pares de sentenças do texto	2.0
7	Coesão Referencial	stem_ovl	Quantidade média de radicais de palavras de conteúdo que se repetem nos pares de sentenças do texto	0.0
8	Coesão Semântica	cross_entropy	Média da entropia cruzadas das sentenças do texto	0.49543
9	Coesão Semântica	lsa_adj_mean	Média de similaridade entre pares de sentenças adjacentes no texto	0.85312
10	Coesão Semântica	lsa_adj_std	Desvio padrão de similaridade entre pares de sentenças adjacentes no texto	0.0
11	Coesão Semântica	lsa_all_mean	Média de similaridade entre todos os pares de sentenças no texto	0.85312
12	Coesão Semântica	lsa_all_std	Desvio padrão de similaridade entre todos os pares possíveis de sentenças do texto	0.0
13	Coesão Semântica	lsa_giveness_mean	Média do *giveness* da cada sentença do texto a partir da segunda	0.85312
14	Coesão Semântica	lsa_giveness_std	Desvio padrão do *giveness* da cada sentença do texto a partir da segunda	0.0
15	Coesão Semântica	lsa_paragraph_mean	Média de similaridade entre pares de parágrafos adjacentes no texto	0
16	Coesão	lsa_paragraph_std	Desvio padrão entre parágrafos adjacentes no texto	0

	Grupo	Métrica	Descrição	Valor
	Semântica			
17	Coesão Semântica	<u>lsa_span_mean</u>	Média do *span* da cada sentença do texto a partir da segunda	0.85312
18	Coesão Semântica	<u>lsa_span_std</u>	Desvio padrão do span da cada sentença do texto a partir da segunda	0.0
19	Complexidade Sintática	<u>adjunct_per_clause</u>	Quantidade média de adjuntos adverbiais por oração do texto	0.55556
20	Complexidade Sintática	<u>adverbs_before_main_verb_ratio</u>	Proporção de orações com advérbio antes do verbo principal em relação à quantidade de orações do texto	0.11111
21	Complexidade Sintática	<u>apposition_per_clause</u>	Quantidade média de apostos por oração do texto	0.0
22	Complexidade Sintática	<u>clauses_per_sentence</u>	Quantidade média de orações por sentença	4.5
23	Complexidade Sintática	<u>coordinate_conjunctions_per_clauses</u>	Proporção de conjunções coordenativas em relação a todas as orações do texto	0.44444
24	Complexidade Sintática	<u>dep_distance</u>	Distância na árvore de dependências	45.0
25	Complexidade Sintática	<u>frazier</u>	Fórmula de Complexidade Sintática de Frazier	8.5
26	Complexidade Sintática	<u>infinite_subordinate_clauses</u>	Proporção de orações subordinadas reduzidas pela quantidade de orações do texto	0.11111
27	Complexidade Sintática	<u>non_svo_ratio</u>	Proporção de orações que não estão no formato SVO (sujeito-verbo-objeto) em relação a todas orações do texto	0.11111
28	Complexidade Sintática	<u>passive_ratio</u>	Proporção de orações na voz passiva analítica em relação à quantidade de orações do texto	0.0
29	Complexidade Sintática	<u>postponed_subject_ratio</u>	Proporção de sujeitos pospostos em relação a todos os sujeitos do texto	0.0
30	Complexidade Sintática	<u>ratio_coordinate_conjunctions</u>	Proporção de conjunções coordenativas em relação a todas as conjunções do texto	0.8
31	Complexidade Sintática	<u>ratio_subordinate_conjunctions</u>	Proporção de conjunções subordinativas em relação a todas conjunções do texto	0.2
32	Complexidade Sintática	<u>relative_clauses</u>	Proporção de orações relativas em relação à quantidade de orações do texto	0.33333
33	Complexidade Sintática	<u>sentences_with_five_clauses</u>	Proporção de sentenças com 5 orações em relação a todas as sentenças do texto	0.0
34	Complexidade Sintática	<u>sentences_with_four_clauses</u>	Proporção de sentenças com 4 orações em relação a todas as sentenças do texto	0.0
35	Complexidade Sintática	<u>sentences_with_one_clause</u>	Proporção de sentenças com 1 oração em relação a todas as sentenças do texto	0.0
36	Complexidade Sintática	<u>sentences_with_seven_more_clauses</u>	Proporção de sentenças com 7 ou mais orações em relação a todas as sentenças do texto	0.5
37	Complexidade Sintática	<u>sentences_with_six_clauses</u>	Proporção de sentenças com 6 orações em relação a todas as sentenças do texto	0.0
38	Complexidade Sintática	<u>sentences_with_three_clauses</u>	Proporção de sentenças com 3 orações em relação a todas as sentenças do texto	0.0
39	Complexidade Sintática	<u>sentences_with_two_clauses</u>	Proporção de sentenças com 2 orações em relação a todas as sentenças do texto	0.5
40	Complexidade Sintática	<u>sentences_with_zero_clause</u>	Proporção de sentenças sem verbos em relação a todas as sentenças do texto	0.0
41	Complexidade Sintática	<u>std_noun_phrase</u>	Desvio-padrão do tamanho dos sintagmas nominais do texto	5.64269

	Grupo	Métrica	Descrição	Valor
42	Complexidade Sintática	subordinate_clauses	Proporção de orações subordinadas pela quantidade de orações do texto	0.55556
43	Complexidade Sintática	temporal_adjunct_ratio	Proporção de adjuntos adverbiais de tempo em relação a todos os adjuntos adverbiais do texto	0.0
44	Complexidade Sintática	words_before_main_verb	Quantidade Média de palavras antes dos verbos principais das orações principais das sentenças	5.0
45	Complexidade Sintática	yngve	Fórmula de Complexidade Sintática de Yngve	2.40789
46	Conectivos	add_neg_conn_ratio	Proporção de conectivos aditivos negativos em relação à quantidade de palavras do texto	0.0
47	Conectivos	add_pos_conn_ratio	Proporção de conectivos aditivos positivos em relação à quantidade de palavras do texto	0.08
48	Conectivos	and_ratio	Proporção do operador lógico E em relação à quantidade de palavras do texto	0.08
49	Conectivos	cau_neg_conn_ratio	Proporção de conectivos causais negativos em relação à quantidade de palavras do texto	0.0
50	Conectivos	cau_pos_conn_ratio	Proporção de conectivos causais positivos em relação à quantidade de palavras do texto	0.12
51	Conectivos	conn_ratio	Proporção de Conectivos em relação à quantidade de palavras do texto	0.2
52	Conectivos	if_ratio	Proporção do operador lógico SE em relação à quantidade de palavras do texto	0.0
53	Conectivos	log_neg_conn_ratio	Proporção de Conectivos Lógicos Negativos em relação à quantidade de palavras do texto	0.0
54	Conectivos	log_pos_conn_ratio	Proporção de Conectivos Lógicos Positivos em relação à quantidade de palavras do texto	0.06
55	Conectivos	logic_operators	Proporção de Operadores Lógicos em relação à quantidade de palavras do texto	0.08
56	Conectivos	negation_ratio	Proporção de palavras que denotam negação em relação à quantidade de palavras do texto	0.0
57	Conectivos	or_ratio	Proporção do operador lógico OU em relação à quantidade de palavras do texto	0.0
58	Densidade de Padrões Sintáticos	gerund_verbs	Proporção de verbos no gerúndio em relação a todos os verbos do texto	0.0
59	Densidade de Padrões Sintáticos	max_noun_phrase	Máximo entre os tamanhos de sintagmas nominais do texto	17
60	Densidade de Padrões Sintáticos	mean_noun_phrase	Média dos tamanhos médios dos sintagmas nominais nas sentenças	8.875
61	Densidade de Padrões Sintáticos	min_noun_phrase	Mínimo entre os tamanhos de sintagmas nominais do texto	2
62	Diversidade Lexical	adjective_diversity_ratio	Proporção de types de adjetivos em relação à quantidade de tokens de adjetivos no texto	1.0
63	Diversidade Lexical	content_density	Proporção de palavras de conteúdo em relação à quantidade de palavras funcionais do texto	1.0
64	Diversidade Lexical	content_word_diversity	Proporção de types de palavras de conteúdo em relação à quantidade de tokens de palavras de conteúdo no texto	0.96
65	Diversidade Lexical	content_word_max	Proporção máxima de palavras de conteúdo em relação à quantidade de palavras das sentenças	0.52632

	Grupo	Métrica	Descrição	Valor
66	Diversidade Lexical	content_word_min	Proporção Mínima de palavras de conteúdo por quantidade de palavras nas sentenças	0.35714
67	Diversidade Lexical	content_word_standard_deviation	Desvio padrão das proporções entre as palavras de conteúdo e a quantidade de palavras das sentenças	0.08459
68	Diversidade Lexical	function_word_diversity	Proporção de types de palavras funcionais em relação à quantidade de tokens de palavras funcionais no texto	0.56
69	Diversidade Lexical	indefinite_pronouns_diversity	Proporção de types de pronomes indefinidos em relação à quantidade de tokens de pronomes indefinidos no texto	1.0
70	Diversidade Lexical	noun_diversity	Proporção de types de substantivos em relação à quantidade de tokens de substantivos no texto	1.0
71	Diversidade Lexical	preposition_diversity	Proporção de types de preposições em relação à quantidade de tokens de preposições no texto	0.75
72	Diversidade Lexical	pronoun_diversity	Proporção de types de pronomes em relação à quantidade de tokens de pronomes no texto	0.71429
73	Diversidade Lexical	punctuation_diversity	Proporção de types de pontuações em relação à quantidade de tokens de pontuações no texto	1.0
74	Diversidade Lexical	relative_pronouns_diversity_ratio	Proporção de types de pronomes relativos em relação à quantidade de tokens de pronomes relativos no texto	0.33333
75	Diversidade Lexical	ttr	Proporção de types (despreza repetições de palavras) em relação à quantidade de tokens (computa repetições de palavras) no texto	0.76
76	Diversidade Lexical	verb_diversity	Proporção de types de verbos em relação à quantidade de tokens de verbos no texto	1.0
77	Frequência de Palavras	cw_freq	Média das frequências absolutas das palavras de conteúdo do texto	195018.52
78	Frequência de Palavras	cw_freq_bra	Média dos valores das frequências das palavras de conteúdo do texto na escala logarítmica Zipf via Corpus Brasileiro	4.67464
79	Frequência de Palavras	cw_freq_brwac	Média dos valores das frequências das palavras de conteúdo do texto na escala logarítmica Zipf via BrWac	4.39668
80	Frequência de Palavras	freq_bra	Média dos valores das frequências das palavras do texto na escala logarítmica Zipf via Corpus Brasileiro	5.82522
81	Frequência de Palavras	freq_brwac	Média dos valores das frequências das palavras do texto na escala logarítmica Zipf via BrWac	5.65432
82	Frequência de Palavras	min_cw_freq	Média das frequências das palavras de conteúdo mais raras das sentenças do texto	1661.0
83	Frequência de Palavras	min_cw_freq_bra	Média dos valores das frequências das palavras de conteúdo mais raras das sentenças do texto na escala logarítmica Zipf via Corpus Brasileiro	3.2635
84	Frequência de Palavras	min_cw_freq_brwac	Média dos valores das frequências das palavras de conteúdo do texto na escala logarítmica Zipf via BrWac	1.4255
85	Frequência de Palavras	min_freq_bra	Média dos valores das frequências das palavras mais raras das sentenças do texto na escala logarítmica Zipf via Corpus Brasileiro	3.2635
86	Frequência de Palavras	min_freq_brwac	Média dos valores das frequências das palavras mais raras das sentenças do texto na escala logarítmica Zipf via BrWac	1.4255
87	Informações Morfossintáticas de Palavras	adjective_ratio	Proporção de Adjetivos em relação à quantidade de palavras do texto	0.06
88	Informações Morfossintáticas de Palavras	adjectives_max	Proporção máxima de adjetivos em relação à quantidade de palavras das sentenças	0.07692
89	Informações Morfossintáticas	adjectives_min	Proporção mínima de adjetivos em relação à quantidade de palavras das sentenças	0.05405

	Grupo	Métrica	Descrição	Valor
	de Palavras			
90	Informações Morfossintáticas de Palavras	<u>adjectives_standard_deviation</u>	Desvio padrão das proporções entre adjetivos e a quantidade de palavras das sentenças	0.01143
91	Informações Morfossintáticas de Palavras	<u>adverbs</u>	Proporção de Advérbios em relação à quantidade de palavras do texto	0.04
92	Informações Morfossintáticas de Palavras	<u>adverbs_diversity_ratio</u>	Proporção de types de advérbios em relação à quantidade de tokens de advérbios no texto	0.5
93	Informações Morfossintáticas de Palavras	<u>adverbs_max</u>	Proporção máxima de advérbios em relação à quantidade de palavras das sentenças	0.0
94	Informações Morfossintáticas de Palavras	<u>adverbs_min</u>	Proporção mínima de advérbios em relação à quantidade de palavras das sentenças	0.0
95	Informações Morfossintáticas de Palavras	<u>adverbs_standard_deviation</u>	Desvio padrão das proporções entre advérbios e a quantidade de palavras das sentenças	0.0
96	Informações Morfossintáticas de Palavras	<u>content_words</u>	Proporção de palavras de conteúdo em relação à quantidade de palavras do texto	0.5
97	Informações Morfossintáticas de Palavras	<u>first_person_possessive_pronouns</u>	Proporção de pronomes possessivos nas primeiras pessoas em relação à quantidade de pronomes possessivos do texto	0.0
98	Informações Morfossintáticas de Palavras	<u>first_person_pronouns</u>	Proporção de pronomes pessoais nas primeiras pessoas em relação à quantidade de pronomes pessoais do texto	0
99	Informações Morfossintáticas de Palavras	<u>function_words</u>	Proporção de Palavras Funcionais em relação à quantidade de palavras do texto	0.5
100	Informações Morfossintáticas de Palavras	<u>indefinite_pronoun_ratio</u>	Proporção de pronomes indefinidos em relação a todos os pronomes do texto	0.14286
101	Informações Morfossintáticas de Palavras	<u>indicative_condition_ratio</u>	Proporção de Verbos no Futuro do Pretérito do Indicativo em relação à quantidade de verbos flexionados do texto	0.0
102	Informações Morfossintáticas de Palavras	<u>indicative_future_ratio</u>	Proporção de Verbos no Futuro do Presente do Indicativo em relação à quantidade de verbos flexionados do texto	0.0
103	Informações Morfossintáticas de Palavras	<u>infinitive_verbs</u>	Proporção de verbos no infinitivo em relação a todos os verbos do texto	0.2
104	Informações Morfossintáticas de Palavras	<u>inflected_verbs</u>	Proporção de verbos flexionados em relação a todos os verbos do texto	0.7
105	Informações Morfossintáticas de Palavras	<u>non-inflected_verbs</u>	Proporção de verbos no gerúndio/particípio/infinitivo em relação a todos os verbos do texto	0.3
106	Informações Morfossintáticas de Palavras	<u>noun_ratio</u>	Proporção de substantivos em relação à quantidade de palavras do texto	0.22
107	Informações Morfossintáticas de Palavras	<u>nouns_max</u>	Proporção máxima de substantivos em relação à quantidade de palavras das sentenças	0.23077

	Grupo	Métrica	Descrição	Valor
108	Informações Morfossintáticas de Palavras	nouns_min	Proporção mínima de substantivos em relação à quantidade de palavras das sentenças	0.21622
109	Informações Morfossintáticas de Palavras	nouns_standard_deviation	Desvio padrão das proporções entre substantivos e a quantidade de palavras das sentenças	0.00728
110	Informações Morfossintáticas de Palavras	oblique_pronouns_ratio	Proporção de pronomes oblíquos em relação a todos os pronomes do texto	0.0
111	Informações Morfossintáticas de Palavras	personal_pronouns	Proporção de Pronomes Pessoais em relação à quantidade de palavras do texto	0.0
112	Informações Morfossintáticas de Palavras	prepositions_per_clause	Proporção de preposições em relação à quantidade de orações no texto	0.77778
113	Informações Morfossintáticas de Palavras	prepositions_per_sentence	Quantidade Média de preposições por sentença no texto	3.5
114	Informações Morfossintáticas de Palavras	pronoun_ratio	Proporção de pronomes em relação à quantidade de palavras do texto	0.14
115	Informações Morfossintáticas de Palavras	pronouns_max	Proporção máxima de pronomes em relação à quantidade de palavras das sentenças	0.30769
116	Informações Morfossintáticas de Palavras	pronouns_min	Proporção mínima de pronomes em relação à quantidade de palavras das sentenças	0.08108
117	Informações Morfossintáticas de Palavras	pronouns_standard_deviation	Desvio padrão das proporções entre pronomes e a quantidade de palavras das sentenças	0.11331
118	Informações Morfossintáticas de Palavras	punctuation_ratio	Proporção de sinais de pontuação em relação à quantidade de palavras do texto	0.03846
119	Informações Morfossintáticas de Palavras	ratio_function_to_content_words	Proporção de palavras funcionais em relação à quantidade de palavras de conteúdo do texto	1.0
120	Informações Morfossintáticas de Palavras	relative_pronouns_ratio	Proporção de Pronomes Relativos em relação à quantidade de pronomes do texto	0.42857
121	Informações Morfossintáticas de Palavras	second_person_possessive_pronouns	Proporção de pronomes possessivos nas segundas pessoas em relação à quantidade de pronomes possessivos do texto	0.0
122	Informações Morfossintáticas de Palavras	second_person_pronouns	Proporção de pronomes pessoais nas segundas pessoas em relação à quantidade de pronomes pessoais do texto	0
123	Informações Morfossintáticas de Palavras	third_person_possessive_pronouns	Proporção de pronomes possessivos nas terceiras pessoas em relação à quantidade de pronomes possessivos do texto	1.0
124	Informações Morfossintáticas de Palavras	third_person_pronouns	Proporção de pronomes pessoais nas terceiras pessoas em relação à quantidade de pronomes pessoais do texto	0
125	Informações Morfossintáticas de Palavras	verbs	Proporção de Verbos em relação à quantidade de palavras do texto	0.18
126	Informações Morfossintáticas	verbs_max	Proporção máxima de verbos em relação à quantidade de palavras das sentenças	0.18919

	Grupo	Métrica	Descrição	Valor
	de Palavras			
127	Informações Morfosintáticas de Palavras	<u>verbs_min</u>	Proporção mínima de verbos em relação à quantidade de palavras das sentenças	0.07692
128	Informações Morfosintáticas de Palavras	<u>verbs_standard_deviation</u>	Desvio padrão das proporções entre verbos e a quantidade de palavras das sentenças	0.05613
129	Informações Semânticas de Palavras	<u>abstract_nouns_ratio</u>	Proporção de substantivos abstratos em relação à quantidade de palavras do texto	0.0
130	Informações Semânticas de Palavras	<u>adjectives_ambiguity</u>	Proporção de sentidos dos adjetivos do texto em relação à quantidade de adjetivos do texto	2.0
131	Informações Semânticas de Palavras	<u>adverbs_ambiguity</u>	Proporção de sentidos dos advérbios do texto em relação à quantidade de advérbios do texto	0
132	Informações Semânticas de Palavras	<u>content_words_ambiguity</u>	Média de sentidos por palavra de conteúdo do texto	4.54545
133	Informações Semânticas de Palavras	<u>hypernyms_verbs</u>	Quantidade Média de Hiperônimos por verbo nas sentenças	0.28571
134	Informações Semânticas de Palavras	<u>named_entity_ratio_sentence</u>	Média das proporções de Nomes Próprios em relação à quantidade de palavras das Sentenças	0.0
135	Informações Semânticas de Palavras	<u>named_entity_ratio_text</u>	Proporção de Nomes Próprios em relação à quantidade de palavras do Texto	0.0
136	Informações Semânticas de Palavras	<u>negative_words</u>	Proporção de palavras de polaridade negativa em relação a todas palavras do texto	0.4
137	Informações Semânticas de Palavras	<u>nouns_ambiguity</u>	Proporção de sentidos dos substantivos do texto em relação à quantidade de substantivos do texto	1.63636
138	Informações Semânticas de Palavras	<u>positive_words</u>	Proporção de palavras de polaridade positiva em relação a todas palavras do texto	0.48
139	Informações Semânticas de Palavras	<u>verbs_ambiguity</u>	Proporção de sentidos dos verbos do texto em relação à quantidade de verbos do texto	9.5
140	Léxico Temporal	<u>aux_plus_PCP_per_sentence</u>	Proporção de verbos auxiliares seguidos de participio em relação à quantidade de sentenças do texto	0.0
141	Léxico Temporal	<u>indicative_imperfect_ratio</u>	Proporção de Verbos no Pretérito Imperfeito do Indicativo em relação à quantidade de verbos flexionados no texto	0.0
142	Léxico Temporal	<u>indicative_pluperfect_ratio</u>	Proporção de Verbos no Pretérito Mais que Perfeito do Indicativo em relação à quantidade de verbos flexionados no texto	0.22222
143	Léxico Temporal	<u>indicative_present_ratio</u>	Proporção de Verbos no Presente do Indicativo em relação à quantidade de verbos flexionados no texto	0.44444
144	Léxico Temporal	<u>indicative_preterite_perfect_ratio</u>	Proporção de Verbos no Pretérito Perfeito Simples do Indicativo em relação à quantidade de verbos flexionados no texto	0.33333
145	Léxico Temporal	<u>participle_verbs</u>	Proporção de verbos no participio em relação a todos os verbos do texto	0.1

	Grupo	Métrica	Descrição	Valor
146	Léxico Temporal	subjunctive_future_ratio	Proporção de Verbos no Futuro do Subjuntivo em relação à quantidade de verbos flexionados no texto	0.0
147	Léxico Temporal	subjunctive_imperfect_ratio	Proporção de Verbos no Pretérito Imperfeito do Subjuntivo em relação à quantidade de verbos flexionados no texto	0.0
148	Léxico Temporal	subjunctive_present_ratio	Proporção de Verbos no Presente do Subjuntivo em relação à quantidade de verbos flexionados no texto	0.0
149	Léxico Temporal	tmp_neg_conn_ratio	Proporção de conectivos temporais negativos em relação à quantidade de palavras do texto	0.0
150	Léxico Temporal	tmp_pos_conn_ratio	Proporção de conectivos temporais positivos em relação à quantidade de palavras do texto	0.0
151	Léxico Temporal	verbal_time_moods_diversity	Quantidade de diferentes tempos-modos verbais que ocorrem no texto	3
152	Medidas Descritivas	paragraphs	Quantidade de Parágrafos no texto	1
153	Medidas Descritivas	sentence_length_max	Quantidade Máxima de palavras por sentença	37
154	Medidas Descritivas	sentence_length_min	Quantidade Mínima de palavras por sentença	13
155	Medidas Descritivas	sentence_length_standard_deviation	Desvio Padrão da quantidade de palavras por sentença	12.0
156	Medidas Descritivas	sentences	Quantidade de Sentenças no texto	2
157	Medidas Descritivas	sentences_per_paragraph	Quantidade média de sentenças por parágrafo no texto	2.0
158	Medidas Descritivas	subtitles	Proporção de Subtítulos em relação à quantidade de sentenças do texto	0.0
159	Medidas Descritivas	syllables_per_content_word	Quantidade média de sílabas por palavra de conteúdo no texto	2.96
160	Medidas Descritivas	words	Quantidade de Palavras no texto	50
161	Medidas Descritivas	words_per_sentence	Média de Palavras por Sentença	25.0
162	Medidas Psicolinguísticas	concretude_1_25_ratio	Proporção de palavras com valor de concretude entre 1 e 2.5 em relação a todas as palavras de conteúdo do texto	0.0
163	Medidas Psicolinguísticas	concretude_25_4_ratio	Proporção de palavras com valor de concretude entre 2.5 e 4 em relação a todas as palavras de conteúdo do texto	0.47826
164	Medidas Psicolinguísticas	concretude_4_55_ratio	Proporção de palavras com valor de concretude entre 4 e 5.5 em relação a todas as palavras de conteúdo do texto	0.52174
165	Medidas Psicolinguísticas	concretude_55_7_ratio	Proporção de palavras com valor de concretude entre 5.5 e 7 em relação a todas as palavras de conteúdo do texto	0.0
166	Medidas Psicolinguísticas	concretude_mean	Média dos valores de concretude das palavras de conteúdo do texto	3.89029
167	Medidas Psicolinguísticas	concretude_std	Desvio padrão do valor de concretude das palavras de conteúdo do texto	0.65796
168	Medidas Psicolinguísticas	familiaridade_1_25_ratio	Proporção de palavras com valor de familiaridade entre 1 e 2.5 em relação a todas as palavras de conteúdo do texto	0.0
169	Medidas Psicolinguísticas	familiaridade_25_4_ratio	Proporção de palavras com valor de familiaridade entre 2.5 e 4 em relação a todas as palavras de conteúdo do texto	0.30435
170	Medidas Psicolinguísticas	familiaridade_4_55_ratio	Proporção de palavras com valor de familiaridade entre 4 e 5.5 em relação a todas as palavras de conteúdo do texto	0.52174

	Grupo	Métrica	Descrição	Valor
171	Medidas Psicolinguísticas	familiaridade_55_7_ratio	Proporção de palavras com valor de familiaridade entre 5.5 e 7 em relação a todas as palavras de conteúdo do texto	0.17391
172	Medidas Psicolinguísticas	familiaridade_mean	Média dos valores de familiaridade das palavras de conteúdo do texto	4.58892
173	Medidas Psicolinguísticas	familiaridade_std	Desvio padrão dos valores de familiaridade das palavras de conteúdo do texto	0.85169
174	Medidas Psicolinguísticas	idade_aquisicao_1_25_ratio	Proporção de palavras com valor de idade de aquisição entre 1 e 2.5 em relação a todas as palavras de conteúdo do texto	0.0
175	Medidas Psicolinguísticas	idade_aquisicao_25_4_ratio	Proporção de palavras com valor de idade de aquisição entre 2.5 e 4 em relação a todas as palavras de conteúdo do texto	0.30435
176	Medidas Psicolinguísticas	idade_aquisicao_4_55_ratio	Proporção de palavras com valor de idade de aquisição entre 4 e 5.5 em relação a todas as palavras de conteúdo do texto	0.04348
177	Medidas Psicolinguísticas	idade_aquisicao_55_7_ratio	Proporção de palavras com valor de idade de aquisição entre 5.5 e 7 em relação a todas as palavras de conteúdo do texto	0.3913
178	Medidas Psicolinguísticas	idade_aquisicao_mean	Média dos valores de idade de aquisição das palavras de conteúdo do texto	5.69752
179	Medidas Psicolinguísticas	idade_aquisicao_std	Desvio padrão dos valores de idade de aquisição das palavras de conteúdo do texto	1.70388
180	Medidas Psicolinguísticas	imageabilidade_1_25_ratio	Proporção de palavras com valor de imageabilidade entre 1 e 2.5 em relação a todas as palavras de conteúdo do texto	0.0
181	Medidas Psicolinguísticas	imageabilidade_25_4_ratio	Proporção de palavras com valor de imageabilidade entre 2.5 e 4 em relação a todas as palavras de conteúdo do texto	0.43478
182	Medidas Psicolinguísticas	imageabilidade_4_55_ratio	Proporção de palavras com valor de imageabilidade entre 4 e 5.5 em relação a todas as palavras de conteúdo do texto	0.56522
183	Medidas Psicolinguísticas	imageabilidade_55_7_ratio	Proporção de palavras com valor de imageabilidade entre 5.5 e 7 em relação a todas as palavras de conteúdo do texto	0.0
184	Medidas Psicolinguísticas	imageabilidade_mean	Média dos valores de imageabilidade das palavras de conteúdo do texto	3.99609
185	Medidas Psicolinguísticas	imageabilidade_std	Desvio padrão dos valores de imageabilidade das palavras de conteúdo do texto	0.60639
186	Simplicidade Textual	dialog_pronoun_ratio	Proporção de pronomes pessoais que indicam uma conversa com o leitor em relação à quantidade de pronomes pessoais do texto	0
187	Simplicidade Textual	easy_conjunctions_ratio	Proporção de conjunções fáceis em relação à quantidade de palavras do texto	0.2
188	Simplicidade Textual	hard_conjunctions_ratio	Proporção de conjunções difíceis em relação à quantidade de palavras do texto	0.0
189	Simplicidade Textual	long_sentence_ratio	Proporção de Sentenças Muito Longas em relação a todas as sentenças do texto	0.5
190	Simplicidade Textual	medium_long_sentence_ratio	Proporção de Sentenças Longas em relação a todas as sentenças do texto	0.0
191	Simplicidade Textual	medium_short_sentence_ratio	Proporção de Sentenças Médias em relação a todas as sentenças do texto	0.5
192	Simplicidade Textual	short_sentence_ratio	Proporção de Sentenças Curtas em relação a todas as sentenças do texto	0.0
193	Simplicidade Textual	simple_word_ratio	Proporção de palavras de conteúdo simples em relação a todas palavras de conteúdo do texto	0.68
194	coesão Referencial	coreference_pronoun_ratio	Média de candidatos a referente (na sentença anterior) por pronome anafórico do caso reto	0.0
195	coesão Referencial	demonstrative_pronoun_ratio	Média de candidatos a referente (na sentença anterior) por pronome demonstrativo anafórico	0.0

	Grupo	Métrica	Descrição	Valor
196	Índices de Leiturabilidade	brunet	Índice de Brunet	8.55513
197	Índices de Leiturabilidade	dalechall_adapted	Fórmula Dale Chall adaptada	9.9293
198	Índices de Leiturabilidade	flesch	Índice Flesch	44.108
199	Índices de Leiturabilidade	gunning_fox	Índice Gunning Fog	10.136
200	Índices de Leiturabilidade	honore	Estatística de Honoré	1076.01434

Exportar para CSV

Entre com o texto na caixa abaixo (Máximo 2000 palavras por vez).

parabéns a todos que desde o inicio se opuseram a essa vergonhosa lei . tem que acabar com imunidade parlamentar e foro privilegiado , os eleitos são só pessoas que receberam mandato através do voto democrático para trabalhar para o cidadão que e o eleger e paga o seu salário e só

 Não sou um robô reCAPTCHA
Privacidade - Termos

Processar

Limpar

Referências

Sidney Evaldo Leal, Magali Sanches Duran, Carolina Evaristo Scarton, Nathan Siegle Hartmann, Sandra Maria Aluisio: **NILC-Metrix: assessing the complexity of written and spoken language in Brazilian Portuguese**. CoRR abs/2201.03445 (2022)
<https://arxiv.org/abs/2201.03445>

NILC-Metrix: assessing the complexity of written and spoken language in Brazilian Portuguese

Sandra M Aluisio

NILC/ICMC/USP

Leitura e(m) Interfaces: Teorias, Métodos e Aplicações

Programa de Pós-Graduação em Inglês: Estudos Linguísticos e Literários da UFSC

16/04/2021

SLIDES: NILC-Metrix SLIDES 16_04_2021.pdf

LINK: <https://www.youtube.com/watch?v=UeiVbEe1ULY>