

## “Queremos impeachment”: analisando os comentários de leitores do G1 sobre Jair Bolsonaro através de uma rede de palavras

### RESUMO

O G1 é o site noticioso com a maior quantidade de acessos no país. No entanto, há um cenário de deslegitimação da mídia tradicional, em que pessoas acusam estes veículos de atuarem a favor de uma suposta agenda política. Dado o fenômeno da extrema direita que se materializa no Brasil através do governo de Jair Bolsonaro, destaca-se a importância de estruturar esta textualidade que reflete a opinião dos leitores do G1. O tema desta pesquisa é o desenvolvimento de uma rede de coocorrência lexical dos comentários do G1 que contém a palavra “bolsonaro”. O objetivo é analisar os comentários do G1 que contém a palavra “bolsonaro”. Para isso, 90 mil comentários foram estruturados em uma rede de palavras através de um algoritmo de coocorrência lexical. Dentre outras constatações, foi evidenciado que muitas pessoas desejam o impeachment do presidente. As principais contribuições deste estudo são o desenvolvimento do método e os dados quantitativos que foram elencados acerca desta temática.

**PALAVRAS-CHAVE:** Análise de Conteúdo. Bolsonaro. Coocorrência lexical. G1. Rede de palavras.

Eduardo Gabriel Velho  
[eduardo.velho@icloud.com](mailto:eduardo.velho@icloud.com)  
Universidade Feevale

Sandra Portella Montardo  
[sandramontardo@feevale.br](mailto:sandramontardo@feevale.br)  
Universidade Feevale

## INTRODUÇÃO

De acordo com um levantamento realizado pela Amazon, o portal de notícias G1<sup>1</sup> é o site noticioso com a maior quantidade de acessos do Brasil (MOUSINHO, 2016), sendo um dos maiores e mais importantes sites do país. O G1 já foi alvo de diversos estudos acerca do seu conteúdo, inclusive sobre a área de comentários do site, que é conhecida por frequentemente ser cenário de violência e reprodução do discurso de ódio. Diversos estudos elencaram sobre o teor dos comentários do site, sendo evidenciado casos de, por exemplo, violência linguística contra vidas trans (SILVA, 2019), violência de gênero exposta na percepção social das pessoas sobre um caso de estupro coletivo (MOUSINHO, 2016) e violência verbal dos leitores do site (CUNHA, 2013).

Além destes estudos, destacam-se os que analisaram os comentários sobre acontecimentos que possuem dimensão política, como o assassinato de Marielle Franco (BIAR; PASCHOAL, 2020) e a vitória de Donald Trump nas eleições presidenciais dos Estados Unidos em 2016 (MATOS; HÜBNER, 2017), os quais revelam um pouco sobre a dimensão política destas violências.

Há um cenário de deslegitimação da mídia tradicional, em que pessoas acusam esses veículos de publicarem notícias falsas e enganosas em prol de uma determinada agenda política e interesses nefastos (EMPOLI, 2019; MELLO, 2020). Neste sentido, destacam-se as eleições presidenciais de 2018, que foram marcadas pelas *fake news* e o assassinato de reputações, em sua maior parte sendo mensagens benéficas a candidatura de Jair Bolsonaro (MELLO, 2020). Desta forma, dado o fenômeno da extrema direita que emergiu no Brasil através do governo de Jair Bolsonaro e de seus aliados, destaca-se a importância de estruturar essa textualidade que manifesta a opinião pública dos muitos leitores do G1.

O tema desta pesquisa é o desenvolvimento de uma rede de coocorrência lexical (RCL) dos comentários do G1 que contém a palavra “bolsonaro”, delimitando-se em realizar este processamento sob um viés sistemático-quantitativo através de algoritmos desenvolvidos para este propósito. Este estudo está centrado diante da seguinte pergunta de pesquisa: de que forma se estrutura a rede de palavras associadas à palavra “bolsonaro” nos comentários do G1? O objetivo desta pesquisa é analisar os comentários do G1 que contém a palavra “bolsonaro”. Para tanto, foi necessário desenvolver um programa para coletar a amostra dos comentários do G1 que compuseram o corpus da pesquisa, implementar um algoritmo capaz de gerar esta rede de palavras e, por fim, criar a visualização gráfica desta estrutura. O método utilizado neste estudo foi a análise de conteúdo de Laurence Bardin (2011), especializando-se em uma análise de coocorrências, que foi viabilizada por instrumentos de processamento de linguagem natural.

Este trabalho inicia com o aporte teórico, onde os estudos de Patrícia Campos Mello (2020) e Angela Nagle (2017) são buscados para teorizar sobre as eleições presidenciais de 2018 e sobre o fenômeno da direita alternativa no Brasil. Posteriormente, o método desenvolvido para esta pesquisa é apresentado, daí descrevendo também os instrumentos utilizados e os trabalhos relacionados a

este. Por fim, a rede de coocorrência lexical que foi desenvolvida é apresentada seguida das considerações sobre esta estrutura.

### O ALT-LIGHT E A ASCENSÃO DA DIREITA ALTERNATIVA

No livro “A máquina de ódio”, a jornalista da Folha de São Paulo, Patrícia Campos Mello (2020), aborda a questão das *fake news*, do assassinato de reputações e da ascensão da extrema direita, que se materializou no Brasil através das eleições presidenciais de 2018. A jornalista conta da perseguição que sofreu ao investigar sobre o uso do WhatsApp para a disseminação de propaganda e de notícias falsas durante estas eleições, disparos estes que em sua maioria eram favoráveis a campanha de Jair Bolsonaro (MELLO, 2020).

A autora (2020) descobriu que embora a difusão de notícias falsas fosse algo “descentralizado”, pois muitos apoiadores compartilhavam e criavam este tipo de material por conta própria, agências de marketing foram contratadas por empresas para, via WhatsApp, realizar o disparo em massa de mensagens a favor da campanha do então candidato, Jair Bolsonaro. Estes disparos não necessariamente eram de notícias falsas, visto que os próprios contratantes é quem decidiam qual seria o teor das mensagens, poderia ser qualquer conteúdo político em favor da campanha (MELLO, 2020). Destaca-se que este financiamento de campanha funcionou como um tipo de “terceirização do caixa dois”, visto que desde 2015 as empresas não podiam mais doar dinheiro para campanhas políticas. Por fim, o envio em massa de mensagens políticas pelo WhatsApp só foi proibido em dezembro de 2019, mais de um ano após a publicação das reportagens (MELLO, 2020).

Devido a estas matérias, de acordo com o livro (MELLO, 2020), Patrícia foi alvo de perseguição e logo começou a ser acusada de fazer campanha política para o PT, sendo vítima da violência digital através de insultos como “putinha do PT” e “vagabunda comunista”. No Twitter, *bots* difundiram as *hashtags* “#CadêAsProvas” e “#MarqueteirosDoJair” que deslegitimavam sua narrativa. Nas mídias sociais, propagaram-se memes com o rosto da jornalista e a legenda “mentirosa, jornalista petista”.

Na sequência, em relação a esta situação de violência, Mello (2020) declara: que ela e seu filho de seis anos foram ameaçados por via de mensagens de texto; que recebeu ligações de pessoas que ameaçaram invadir sua casa e destruir tudo; que seu celular foi hackeado; que seu WhatsApp foi invadido para enviar mensagens a favor de Jair Bolsonaro; e que foi obrigada a cancelar seus compromissos por um mês, pois sua agenda foi divulgada a eleitores de Bolsonaro que foram convidados por via de e-mail a comparecer nos eventos em que participaria. Assim, uma jornalista que cobriu guerras na Síria, no Iraque e no Afeganistão, sem nunca ter sequer um guarda-costas, em plena cidade de São Paulo, tinha de ser acompanhada por alguém contratado pelo jornal para proteger sua integridade física. Como destaca Nagle (2017), estas guerras culturais “perderam a graça” desde que começaram a ter consequências “físicas”.

Angela Nagle (2017) fala sobre a ascensão da extrema direita por via da internet, que germinou dentro de blogs e fóruns obscuros. Sites como o 4chan, o 8chan e no Brasil, o 55chan (VELHO, 2018), são o núcleo da direita alternativa, que, dentre outras características, se destacam pela rejeição ao politicamente correto

(*anti-pc*), a heteronormatividade, aos ideais da supremacia branca, a aversão ao multiculturalismo e ao desprezo pelos conversadores que não concordam com estas quatro ideias anteriores (NAGLE, 2017). Os *alt-right* (pessoas da direita alternativa) operam de forma descentralizada e colaborativa, destacando-se pelo uso de memes, da cultura DIY (*Do It Yourself*) e do conteúdo gerado pelos usuários (UCG, *user-generated content*) como meios para manifestar as suas ideias.

Todavia, os ideais destes grupos ainda são muito fortes para a direita *mainstream* (NAGLE, 2017), que compreende grande parte do eleitorado de Trump e Bolsonaro, por exemplo. Para que estas ideias extremas sejam divulgadas e cheguem de forma palatável até o público em geral, existe a figura do *alt-light* (uma ramificação mais “leve” da direita alternativa), que funciona como uma “ponte” entre as subculturas da direita alternativa [os *alt-right*] e este grande grupo que compreende a ala mais “moderada” da direita política. Ainda segundo a autora, o *alt-light* se distancia da direita alternativa à medida em que não tolera o racismo e a supremacia branca, mas se encontra no desprezo ao politicamente correto e ao multiculturalismo que ambos possuem em comum.

Destaca-se então o fenômeno dos influenciadores digitais ativistas da direita como uma manifestação do *alt-light*, pois servem como uma ponte para a direita alternativa se aproximar do público (NAGLE, 2017; EMPOLI, 2019). Esta aproximação ocorre de forma bidirecional, pois a medida em que os influenciadores convencem o público de que os meios de comunicação não são confiáveis e de que eles estão enganando as pessoas, a direita alternativa se sente legitimada a manifestar suas ideias (NAGLE, 2017). Estes indivíduos conseguem captar os temores do público e oferecer uma versão confortável da realidade, daí a relevância das notícias falsas neste processo de convencimento (EMPOLI, 2019).

Por fim, destaca-se aqui os influenciadores Nando Moura e Diego Rox como expoentes do Brasil neste fenômeno do *alt-light*, devido ao apoio a Bolsonaro durante as eleições presidenciais de 2018 e os ataques à oposição do então candidato. Ambos os influenciadores foram citados por Jair Bolsonaro como “excelentes canais de informação”<sup>2</sup> e, de acordo com Kerche (2019), fazem parte da “nova direita” do YouTube.

## DESENVOLVIMENTO DA RCL

O objetivo desta pesquisa é analisar os comentários do G1 que contém a palavra “Bolsonaro”. Para tanto, foi utilizado como ponto de partida o conceito de coocorrência lexical desenvolvido por Michael Hoey (2005), que a define como a ocorrência de duas ou mais palavras em um curto espaço de distância. Por seguinte, Laurence Bardin (2011, p. 259) descreve a análise de coocorrências como uma técnica que, sob o escopo de uma análise de conteúdo, é capaz de “extrair do texto as relações entre os elementos da mensagem” e “assinalar as presenças simultâneas de dois ou mais elementos na mesma unidade de contexto”. Desta forma, a rede de coocorrência lexical se apresenta como uma estrutura em forma de grafo que ilustra a relação de ocorrência entre itens léxicos. Para desenvolver esta estrutura, o método utilizado foi essencialmente a análise de conteúdo de Bardin (2011), mas com enfoque nas etapas de codificação e classificação em detrimento da etapa de inferência. A técnica utilizada foi uma variação da análise

de coocorrências, cujo procedimento técnico foi sistematizado através de um algoritmo desenvolvido para este propósito.

Este estudo foi fortemente inspirado por Celina Lerner (2019), que desenvolveu uma rede de palavras a partir dos comentários de fãs da página de Olavo de Carvalho no Facebook. Neste trabalho, a autora utilizou métodos estatísticos para estruturar um grafo que representava os diversos temas abordados neste espaço. Outro trabalho relevante é o de Özgür, Cetin e Bingol (2008), que analisaram uma rede de coocorrência desenvolvida sob um *dataset* de notícias. O enfoque desta pesquisa era realizar inferências sobre esta rede de palavras.

Desta forma, como em qualquer pesquisa com dados, o primeiro passo foi obter uma amostra capaz de reproduzir o fenômeno a ser investigado. Para tanto, foi desenvolvido um programa do tipo *web scraper*<sup>3</sup> com o propósito de coletar e estruturar os comentários do G1.

Este primeiro momento é sempre desafiador, pois é necessário mergulhar em um processo de engenharia reversa para entender um pouco sobre o funcionamento do site, sobre como o conteúdo está estruturado e sobre quais tecnologias foram empregadas no seu desenvolvimento. Porém, mais importante do que isso, ao utilizar um *scraper* e realizar a coleta automatizada dos dados, as questões éticas devem ser tratadas com bastante cuidado.

Se existir, o pesquisador deve ler os termos de uso que regem o site. É comum que existam diretrizes sobre o acesso automatizado e a “raspagem” dos dados nestes documentos. Caso os termos de uso não existam ou não abordem a questão do acesso automatizado e da raspagem dos dados, é importante ler o arquivo *robots.txt*, que funciona como um padrão para que os motores de busca saibam quais endereços do site podem ou não ser indexados. Em alguns casos, este arquivo pode também incluir diretrizes que determinam limites na frequência dos acessos, como é o caso do Twitter, que permite o *scraping* de alguns endereços com a frequência de 1rps (uma requisição por segundo)<sup>4</sup>.

No caso desta pesquisa, o arquivo *robots.txt*<sup>5</sup> não restringe os endereços do G1 que interessam à coleta de dados. Por seguinte, de acordo com os termos de uso da plataforma de comentários da Globo<sup>6</sup>, em relação à postagem de comentários:

Você se compromete a não utilizar qualquer sistema automatizado, inclusive, mas sem se limitar a "robôs", "spiders" ou "offline readers", que acessem a Plataforma de maneira a enviar mais mensagens de solicitações aos servidores das Empresas em um dado período de tempo do que seja humanamente possível responder no mesmo período através de um navegador convencional. É igualmente vedada a coleta de qualquer informação pessoal dos demais usuários da Plataforma.

Fica claro que neste documento a preocupação se direciona à carga que os acessos automatizados podem gerar na plataforma de comentários do site. Estes acessos não são proibidos, eles podem ocorrer desde que nas condições estabelecidas pelos termos de uso. Em relação à frequência das requisições, como “humanamente possível” não é um critério inteligível, os acessos foram realizados à 0,3rps (aproximadamente 1 acesso a cada 3 segundos), que é aproximadamente

o tempo de o pesquisador rolar a página de comentários até o final e pressionar o botão para carregar mais comentários no navegador. De qualquer forma, em todos os acessos foi anexado um e-mail para contato e uma mensagem informando que os acessos estavam sendo realizados para uma pesquisa acadêmica. Por fim, ainda de acordo com os termos de uso, os dados pessoais dos usuários não foram coletados – no caso, a única informação “pessoal” que seria passível de ser coletada seriam os nomes dos usuários.

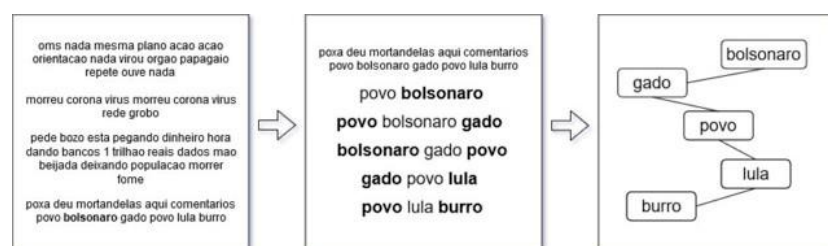
Após o desenvolvimento do programa *scraper*, foi iniciada a coleta de dados, que obteve os comentários de notícias entre 28 de março de 2020 e 11 de setembro de 2020. Desta forma, foram coletados 1.059.672 comentários distribuídos entre 137.173 notícias e, dentre estes comentários, 90.720 contiveram a palavra “bolsonaro”, contabilizando 8,5% do total de postagens. Dos comentários, foi coletado o conteúdo textual, a data de criação, a quantidade de “gostei” e a identificação numérica e data de criação do usuário que postou. Entretanto, para este estudo, somente o conteúdo textual dos comentários foi utilizado. Por fim, destaca-se também que este *dataset* está disponível para download sob requisição na plataforma Zenodo<sup>7</sup>.

Ao finalizar a coleta de dados, foi executado um pré-processamento sob os comentários, que consistiu em, respectivamente: (1) transformar todos os caracteres em minúsculos, (2) remover acentos, pontuação, quebras de linha, espaços duplicados e *stopwords*<sup>8</sup> e, por fim, (3) substituir expressões de linguagem de internet e abreviações através de um dicionário de palavras. Foi optado por não realizar *stemming* (ORENGO; BURIOL; COELHO, 2006), *lemmatization* (QI et al, 2020), correção ortográfica ou qualquer outra rotina de pré-processamento comuns neste tipo de pesquisa, pois a ideia é analisar o texto em sua forma mais “crua”. Por fim, foi realizada a tokenização do conteúdo, que dividiu o texto em unigramas, sendo um requisito para a implementação desta RCL.

Para criar a RCL, um programa de análise de coocorrências foi desenvolvido. Este algoritmo realizou a “leitura” dos comentários e, para cada palavra destes textos, interpretou a coocorrência entre as palavras mais próximas.

Conforme foi explicitado anteriormente, a coocorrência lexical é a ocorrência de duas ou mais palavras em um curto espaço de distância (HOEY, 2005). Na Figura 1 é possível ver que no segundo quadro existem alguns exemplos desta coocorrência, onde, por proximidade, há relação das palavras “povo” e “gado” com a palavra “bolsonaro”. O mesmo ocorre com as palavras “gado” e “lula”, que possuem relação com a palavra “povo”.

Figura 1 – Funcionamento de um algoritmo de coocorrência



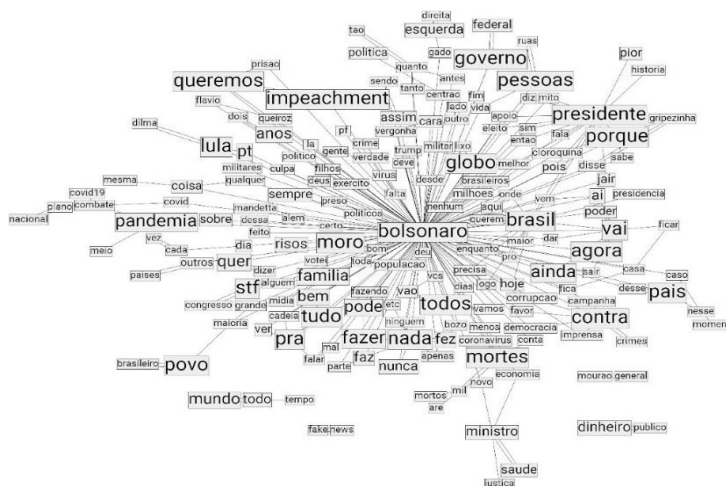
Fonte: Elaborada pelos autores.

Evidencia-se que, para este exemplo, foi definido que uma palavra é coocorrente somente se estiver imediatamente à direita ou à esquerda de uma outra. Neste caso, quer dizer que a janela de coocorrência é de 1 palavra de distância, mas esta janela poderia ser maior dependendo dos objetivos do estudo (HOEY, 2005). Para o algoritmo desta pesquisa, foi utilizada uma janela de 2 palavras de distância. Desta forma, retomando o exemplo da Figura 1, a palavra “povo” se relacionaria com “bolsonaro”, “gado”, “lula” e “burro”, já que compreenderia um alcance maior.

Porém, este é um exemplo muito simples, pois compreende não mais do que um único texto, o que não é o caso deste estudo, que possui um corpus de aproximadamente 90 mil comentários. A verdade é que independentemente da quantidade de textos, a regra continua a mesma, visto que as palavras são mescladas na RCL. O problema é que mais texto significa mais arestas e vértices na rede de palavras, daí a dificuldade em representar graficamente esta estrutura. Imagine uma rede onde cada palavra se conecta com outras mil, como fazer o leitor perceber sentido em tamanho emaranhado de palavras? Para resolver este problema, a RCL foi construída com as 200 palavras que apresentaram maior frequência absoluta através dos comentários, partindo do entendimento de que estes itens léxicos seriam os mais importantes. Além disso, as palavras foram ligadas somente àquelas que coocorreram mais vezes juntas. Desta forma, retomando o exemplo anterior, no terceiro quadro da Figura 1 é possível observar que embora a palavra “bolsonaro” coocorra com a palavra “povo”, elas não estão diretamente relacionadas justamente em função desta lógica.

Por fim, vale ressaltar alguns aspectos importantes sobre este algoritmo: (1) a frequência absoluta das palavras é representada graficamente pelo tamanho da fonte (mas possui um tamanho máximo), (2) o programa “lê” o texto da esquerda para a direita, (3) a rede de palavras começa a partir daquela que possui a maior frequência absoluta de coocorrências (a palavra central) e, por fim, (4) o algoritmo não considera quando a palavra central é coocorrência de alguma outra palavra. Os demais detalhes de implementação podem ser verificados no código-fonte que foi publicado na plataforma GitHub<sup>9</sup> sob licença GPL 3.0 (software livre e de código aberto). Desta forma, finalmente, a Figura 2 ilustra a estrutura da RCL que foi gerada pelo algoritmo a partir das regras explicitadas.

Figura 2 – Estrutura final da RCL



Fonte: Elaborada pelos autores.

## ANÁLISE DA REDE DE PALAVRAS

A RCL final possui 193 vértices e 303 arestas, sendo de 7 vértices o diâmetro desta rede. Destaca-se que embora o grafo tenha sido reduzido para as 200 palavras mais frequentes, restaram apenas 193 delas. Isto aconteceu porque alguns vértices foram removidos pelo algoritmo, já que ficaram sem arestas após o processo de simplificação. Desta forma, também devido a este processo, algumas relações ficaram desconectadas do grafo, como é possível observar nos casos de “dinheiro público”, “fake news” e “general mourao”, por exemplo.

Para ter uma ideia mais clara de quais relações são as mais importantes, as coocorrências das palavras mais frequentes foram obtidas. Desta forma, para cada uma destas palavras, foi considerada a coocorrência que apresentou maior frequência absoluta. Após a execução desta rotina, foi realizada uma revisão manual dos resultados, onde foram selecionadas somente as relações que, independente do contexto, demonstraram por si só carregar alguma mensagem. A Tabela 1 apresenta os resultados desta análise de coocorrência.

Tabela 1 – Resultados da análise de coocorrência

Item léxico	Frequência absoluta	Coocorrência	Frequência absoluta
impeachment	54.258	queremos	14.459
contra	37.637	crimes	562
povo	33.025	brasileiro	1.252
nada	28.918	fez	642
globo	28.003	lixo	348
stf	26.874	congresso	442
lula	23.490	dilma	858
fez	17.885	nada	653
dinheiro	17.375	publico	922
ministro	16.560	saude	659
esquerda	16.487	direita	522
saude	15.531	ministro	605
cara	14.160	vergonha	175
politica	14.002	centrao	102
milhoes	13.407	brasileiros	235
outros	13.264	países	222
disse	12.947	gripezinha	356
nacional	11.952	plano	928
brasileiro	11.197	povo	1.218
falar	11.087	mal	143
combate	10.265	covid19	573

Fonte: Elaborada pelos autores.

Após esta análise, procurou-se manualmente através do grafo as relações consideradas mais reveladoras para esta pesquisa. Desta forma, para cada uma destas relações, foi executada uma rotina para calcular a coocorrência dos itens léxicos independente de sua ordem e realizar o somatório destes valores. Por ordem de frequência, as relações encontradas foram: “queremos impeachment”



(28.387 coocorrências), “pior presidente” (1.764 coocorrências), “contra crimes” (1.133 coocorrências), “globo lixo” (659 coocorrências), “contra imprensa” (397 coocorrências), “contra democracia” (294 coocorrências), “apoio presidente” (266 coocorrências), “culpa pt” (176 coocorrências), “esquerda gado” (166 coocorrências), “presidente mito” (111 coocorrências) e “culpa lula” (65 coocorrências). A partir destes dados, evidencia-se que a maior parte dos leitores não estão satisfeitos com o governo de Jair Bolsonaro. Este aspecto se manifesta claramente através do bigrama “queremos impeachment”, que contabilizou 28 mil ocorrências. Outros bigramas também demonstram a insatisfação das pessoas em relação a Bolsonaro, como “pior presidente”, “nada fez” e “vergonha cara”. Quantitativamente, fica claro que há mais oposição do que apoiadores nestes comentários analisados.

Por seguinte, fica evidente que a questão da pandemia de COVID-19 é um tema fortemente associado ao Presidente, que se manifesta através dos bigramas “disse gripezinha”, “combate covid19” e “ministro saúde”. Desta forma, é possível que os pedidos de impeachment que foram verificados nos comentários ocorram em razão das medidas de intervenção questionáveis que foram tomadas pelo governo em relação à pandemia.

Outro ponto importante é a manifestação dos bigramas “globo lixo” e “contra imprensa”, que possivelmente indicam a aversão de alguns indivíduos aos veículos de mídia tradicionais. Destaca-se que Mello (2020) e Empoli (2019) associam este fenômeno a ascensão da direita alternativa, que acusa os veículos tradicionais de imprensa de atuarem em favor dos seus próprios interesses em razão de uma agenda política.

Por fim, algumas coocorrências que indicam discurso favorável a Jair Bolsonaro foram encontradas na rede de palavras. Os bigramas, “presidente mito”, “apoio presidente”, “esquerda gado”, “culpa pt” e “culpa lula” estão longe de contabilizar uma frequência tão alta quanto os que indicam oposição ao presidente, mas representam um aspecto muito relevante que é o ataque à oposição. Através desta estrutura não há como verificar porque a esquerda é “gado” ou o que é “culpa do PT”, mas fica evidente que este tipo de ataque é característico dos militantes da direita alternativa, como aponta Nagle (2017). É provável que estas pessoas não sejam *alt-right*, mas talvez sejam os menos extremos *alt-light*, que tornam muitas das ideias da direita alternativa palatáveis para um público muito maior.

Além disso, a área de comentários do G1 não é um núcleo de apoio a Jair Bolsonaro. Então, considera-se natural que as manifestações de apoio ao presidente sejam mais tímidas do que nestes ambientes, algo que pode ser percebido vide a frequência com que estas manifestações ocorreram. No entanto, isso fortalece a ideia de que o fenômeno da ascensão da extrema direita é algo que ocorre em rede. Os *alt-light* atuam fora dos núcleos de militância, daí utilizando um discurso menos extremo com o objetivo de aumentar a adesão das pessoas a estas ideias. As críticas aos veículos de imprensa e o ataque a oposição que foi verificado nestes comentários podem então ser indicativos da presença destes sujeitos.

O ano de 2020 foi marcado pela pandemia de COVID-19, situação em relação a qual o governo de Jair Bolsonaro tomou medidas questionáveis de combate à difusão da doença<sup>10</sup>. Em março o Presidente realizou um pronunciamento sobre o

coronavírus<sup>11</sup>, onde ele minimiza os impactos da “gripezinha”, acusa os veículos de imprensa de estarem “superdimensionando” o vírus, defende o uso da cloroquina para tratar a doença (que já demonstrou não ser eficaz) e, além disso, convida os estados e municípios a abandonarem as medidas de proteção e isolamento social que estavam sendo adotadas. Este pronunciamento teve uma repercussão bastante negativa, que desencadeou semanas de “panelaços” em diversas cidades do país<sup>12</sup>, onde as pessoas cobravam medidas mais sérias de combate ao vírus. Os prefeitos e governadores do país também criticaram a postura do presidente<sup>13</sup>, com ênfase na irresponsabilidade do Governo Federal frente a esta situação.

Bolsonaro também visitou o presidente Donald Trump nos Estados Unidos<sup>14</sup>, sendo que foi noticiado que após a viagem, diversos membros da comitiva que acompanhava o presidente testaram positivo para o coronavírus<sup>15</sup>. Bolsonaro disse que testou negativo, mas resistiu a mostrar os laudos, mesmo após determinação da Justiça Federal<sup>16</sup>. De qualquer forma, o presidente já demonstrou não ser adepto às medidas de proteção e ao combate à difusão da doença, pois em diversas ocasiões esteve em locais públicos sem máscara<sup>17</sup> ou utilizando o equipamento de forma inadequada<sup>18</sup>. Em julho, o presidente relevou que havia contraído a doença, mas que após utilizar cloroquina em seu tratamento já estava se sentindo melhor<sup>19</sup>.

Os comentários do G1 demonstram ser majoritariamente reações a estas questões contemporâneas, principalmente em relação às medidas tomadas pelo Governo diante da pandemia. Desta forma, os pedidos de impeachment que puderam ser percebidos na textualidade analisada podem ser uma reação de repúdio às ações de Bolsonaro. De acordo com o que foi explicitado anteriormente, o Presidente tomou medidas questionáveis diante dos problemas da pandemia e, conforme os comentários esclarecem, houve uma repercussão negativa destas ações.

Por outro lado, as recorrências encontradas nos comentários do G1 parecem refletir um pouco da polarização que existe entre os adeptos às ideias da extrema direita e toda a oposição que remanesce no espectro político. Por um lado, os bigramas “disse gripezinha” e “queremos impeachment” demonstram a insatisfação da população diante do governo Bolsonaro, mas, por outro, os bigramas “globo lixo” e “presidente mito” indicam a aprovação do que está sendo realizado pelo Presidente. Os comentários acerca de Bolsonaro também demonstram ser falas reativas aos acontecimentos contemporâneos, conforme as muitas referências aos temas que cercam o assunto pandemia, algo que pode ser percebido nos bigramas “ministro saúde”, “combate covid19” e “plano nacional”. Partindo disso, é possível perceber que os apoiadores do Governo atacam os veículos de imprensa como forma de defender Bolsonaro, o que coincide com os argumentos de Empoli (2019) e Mello (2020), que associam o governo atual às *fake news* e à negação da Ciência e da veracidade das informações fornecidas pelos veículos tradicionais de imprensa.

A atuação dos *alt-light* nestes comentários pode ser percebida justamente neste papel de lembrar que estes veículos não são confiáveis, e de que embora muitas destas notícias simplesmente destaquem algumas das falas do Presidente sem realizar qualquer inferência, há uma questão de “contexto” por trás de tudo que é noticiado. Também é importante destacar que grande parte da militância que ocorre na internet em favor de Bolsonaro acontece também através dos MAVs (núcleos de Militância em Ambientes Virtuais), dos *sock puppets* (contas falsas

utilizadas para fins fraudulentos) e também dos *bots*, que podem difundir propaganda ou notícias falsas. Não foi possível verificar nos comentários a presença de *sock puppets* ou *bots*, mas é importante lembrar que, conforme Empoli (2019) e Mello (2020) destacam, estas são ferramentas que podem ser utilizadas em prol de uma determinada agenda política e, portanto, é passível de que também sejam atuantes nos comentários do G1.

## CONSIDERAÇÕES FINAIS

Este foi um estudo acerca da temática desenvolvimento de uma rede de coocorrência lexical (RCL) dos comentários do G1 que contém a palavra “Bolsonaro”, delimitando-se em realizar este processamento sob um viés sistemático-quantitativo através de algoritmos desenvolvidos para este propósito. O método utilizado foi capaz de estruturar os 90 mil comentários obtidos neste estudo em uma rede que relacionou as coocorrências das 193 palavras mais frequentes do corpus. Dentre outras constatações, destaca-se as 28 mil ocorrências do bigrama “queremos impeachment”, que revela um pouco sobre o posicionamento político dos leitores do site, assim como sugerem o reconhecimento destas pessoas à legitimidade dos veículos de mídia tradicional. Como limitação do método, está a incapacidade de verificar o contexto em que estas coocorrências aconteceram. As principais contribuições deste estudo são o desenvolvimento do método e os dados quantitativos que foram elencados acerca desta temática.

Em termos de desdobramentos deste estudo, vale estruturar estes dados utilizando outros métodos e perspectivas. Destaca-se que além do conteúdo textual dos comentários, foram coletados data de criação, quantidade de “gostei” e identificação numérica e data de criação do usuário que postou. Desta forma, realizar um estudo cruzando estes dados com o conteúdo textual poderia revelar outros aspectos destes comentários.

# “We want impeachment”: analyzing G1 readers comments about Jair Bolsonaro through a word network

## ABSTRACT

G1 is the news site with the largest number of hits in Brazil. However, there is a scenario of delegitimization of the traditional media, where people accuse these vehicles of acting in favor of a supposed political agenda. Given the phenomenon of the extreme right that materializes in Brazil through the government of Jair Bolsonaro, the importance of structuring this textuality that reflects the opinion of G1 readers is highlighted. The theme of this research is the development of a network of lexical co-occurrence from the G1 comments that contains the word “bolsonaro”. The goal is to analyze the comments of G1 that contains the word “bolsonaro”. For this, 90 thousand comments were structured in a network of words through a lexical co-occurrence algorithm. Among other findings, it was evidenced that many people want the impeachment of the president. The main contributions of this study are the development of the method and the quantitative data that were obtained on this topic.

**KEYWORDS:** Bolsonaro. Content analysis. G1. Lexical co-occurrence. Word network.

## NOTAS

<sup>1</sup> Endereço do portal de notícias G1. Disponível em: <<https://g1.globo.com>>. Acesso em: 01 out. 2020.

<sup>2</sup> Matéria do The Intercept sobre os Youtubers recomendados por Jair Bolsonaro. Disponível em: <<https://theintercept.com/2018/11/17/youtubers-bolsonaro-nando-moura-diego-rox-bernardo-kuster-fake-news/>>. Acesso em: 21 nov. 2020.

<sup>3</sup> O *web scraper* é um tipo de programa capaz de estruturar o conteúdo de um site e extrair informações específicas de suas páginas, de forma que esses dados possam ser posteriormente armazenados e analisados sob um determinado propósito.

<sup>4</sup> No arquivo *robots.txt* do Twitter, a diretiva “*Crawl-delay: 1*” determina que os acessos devem ocorrer a no máximo 1rps. Disponível em: <<https://twitter.com/robots.txt>>. Acesso em: 03 out. 2020.

<sup>5</sup> Arquivo *robots.txt* do G1. Disponível em: <<http://g1.globo.com/robots.txt>>. Acesso em 03 out. 2020.

<sup>6</sup> Termos de uso da plataforma de comentários do G1. Disponível em: <<https://login.globo.com/termosUso/3004>>. Acesso em: 03 out. 2020.

<sup>7</sup> Disponível em: <<https://zenodo.org/record/5557762>>. Acesso em: 03 mai. 2022.

<sup>8</sup> *Stopwords* são palavras que não possuem relevância para a análise, geralmente por não possuírem qualquer significado (palavras como “de”, “e”, “então”, “que”, etc.). A técnica utilizada foi baseada no trabalho de Duran, Nunes e Avanço (2015).

<sup>9</sup> De acordo com estudo da FSP/USP, as medidas adotadas pelo governo de Bolsonaro parecem ter favorecido a propagação do vírus. Disponível em: <<https://brasil.elpais.com/brasil/2021-01-21/pesquisa-revela-que-bolsonaro-executou-uma-estrategia-institucional-de-propagacao-do-virus.html>>. Acesso em: 28 jan. 2021.

<sup>10</sup> Disponível em: <<https://github.com/egvelho/text-network-pt-br-ugc>>. Acesso em: 03 mai. 2022.

<sup>11</sup> Disponível em: <[https://www.youtube.com/watch?v=deRVsj4soUA&feature=emb\\_logo](https://www.youtube.com/watch?v=deRVsj4soUA&feature=emb_logo)>. Acesso em: 26 jan. 2021.

<sup>12</sup> Disponível em: <<https://g1.globo.com/politica/noticia/2021/01/15/cidades-brasileiras-registram-panelaco-contrabolsonaro-nesta-sexta-feira.ghtml>>. Acesso em: 26 jan. 2021.

<sup>13</sup> Disponível em: <<https://g1.globo.com/jornal-nacional/noticia/2020/03/25/reacao-de-governadores-a-pronunciamento-de-bolsonaro-foi-imediata.ghtml>>. Acesso em: 26 jan. 2021.

<sup>14</sup> Disponível em: <<https://g1.globo.com/politica/noticia/2020/03/07/bolsonaro-chega-aos-estados-unidos-para-encontro-com-trump.ghtml>>. Acesso em: 26 jan. 2021.

<sup>15</sup> Disponível em: <<https://www1.folha.uol.com.br/poder/2020/03/sobe-para-23-numero-de-pessoas-com-coronavirus-que-tiveram-contato-com-bolsonaro.shtml>>. Acesso em: 26 jan. 2021.

<sup>16</sup> Disponível em: <<https://noticias.uol.com.br/ultimas-noticias/agencia-estado/2020/05/18/governo-se-nega-a-entregar-exames-de-bolsonaro-a-camara.htm>>. Acesso em: 28 jan. 2021.

<sup>17</sup> Disponível em: <<https://www1.folha.uol.com.br/poder/2020/12/sem-mascara-bolsonaro-provoca-aglomeracao-em-praia-de-sp-abraca-banhistas-e-pegas-criancas-no-colo.shtml>>. Acesso em: 28 jan. 2021.

<sup>18</sup> Disponível em: <<https://g1.globo.com/jornal-nacional/noticia/2020/03/18/em-coletiva-bolsonaro-e-ministros-usam-mascaras-de-forma-errada.ghtml>>. Acesso em: 28 jan. 2021.

<sup>19</sup> Disponível em: <<https://veja.abril.com.br/politica/jair-bolsonaro-toma-hidroxicloroquina-em-video-e-diz-esta-dando-certo>>. Acesso em: 28 jan. 2021.

### AGRADECIMENTOS

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pela Bolsa de Doutorado e pela Bolsa de Produtividade em Pesquisa que possibilitaram o desenvolvimento desta pesquisa.

### REFERÊNCIAS

BIAR, Liana de Andrade; PASCHOAL, Fabiola Valle das Chagas. “(Não) leia os comentários”: A disputa da notícia sobre o assassinato de Marielle Franco. **Trabalhos em Linguística Aplicada**, v. 59, n. 2, p. 1051-1069, 2020. Disponível em: <[https://www.scielo.br/scielo.php?pid=S0103-18132020000201051&script=sci\\_arttext](https://www.scielo.br/scielo.php?pid=S0103-18132020000201051&script=sci_arttext)>. Acesso em 14 abr. 2021.

BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2011.

CUNHA, Dóris de Arruda da. Violência verbal nos comentários de leitores publicados em sites de notícia. **Calidoscópio**, v. 11, n. 3, p. 241-249, 2013. Disponível em: <<http://revistas.unisinos.br/index.php/calidoscopio/article/view/cld.2013.113.02>>. Acesso em: 14 abr. 2021.

DURAN, Magali Sanches; NUNES, Maria das Graças Volpe; AVANÇO, Lucas. A normalizer for ugc in brazilian portuguese. In: **Proceedings of the Workshop on Noisy User-generated Text**. 2015. p. 38-47. Disponível em: <<https://www.aclweb.org/anthology/W15-4305/>>. Acesso em: 14 abr. 2021.

EMPOLI, Giuliano Da. **Os engenheiros do caos**. São Paulo: Vestígio, 2019.

HOEY, Michael. **Lexical Priming**. London: Routledge, 2005.

KERCHE, Francisco. **As Redes do Conservadorismo Brasileiro: Mapeando a Nova Direita** no Youtube. VI Simpósio Internacional LAVITS. 2019. Disponível em: <<https://lavits.org/wp-content/uploads/2019/12/Kerche-2019-LAVITSS.pdf>>. Acesso em: 14 abr. 2021.

LERNER, Selina. Sobre o que falam os fãs de Olavo de Carvalho? Uma análise computacional de comentários no Facebook. In: POLIVANOV, Beatriz.; ARAÚJO, Willian; OLIVEIRA, Caio; SILVA, Tarcízio (org.). **Fluxos em redes sociotécnicas: das micronarrativas ao big data**, p. 337-358, 2019.

MATOS, Silvio Simão de; HÜBNER, Jucilei Geraldo. Indo além da reportagem: As discussões de gênero, raça e direitos humanos nos comentários no G1 sobre a vitória de Donald Trump nas eleições dos EUA. **40º Congresso Brasileiro de Ciências da Comunicação. Curitiba**. 2017. Disponível em: <<https://portalintercom.org.br/anais/nacional2017/resumos/R12-2599-1.pdf>>. Acesso em 14 abr. 2021.

MELLO, Patrícia Campos. **A máquina de ódio: Notas de uma repórter sobre fake news e violência digital**. 1. ed. São Paulo: Companhia das Letras, 2020. 196 p.

MOUSINHO, Amanda Arrais. Violência de gênero: A percepção social sobre um caso de estupro coletivo por meio da análise dos comentários na página do G1 no Facebook. **Carta dos editores**, p. 34, 2016. Disponível em: <[http://www.cambiassu.ufma.br/cambi\\_2016.2/violencia.pdf](http://www.cambiassu.ufma.br/cambi_2016.2/violencia.pdf)>. Acesso em: 14 abr. 2021.

NAGLE, Angela. **Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right**. John Hunt Publishing, 2017.

ORENGO, Viviane Moreira; BURIOL, Luciana S.; COELHO, Alexandre Ramos. A study on the use of stemming for monolingual ad-hoc Portuguese information retrieval. In: **Workshop of the Cross-Language Evaluation Forum for European Languages**. Springer, Berlin, Heidelberg, 2006. p. 91-98. Disponível em: <[https://www.inf.ufrgs.br/~buriol/papers/Orengo\\_CLEF07.pdf](https://www.inf.ufrgs.br/~buriol/papers/Orengo_CLEF07.pdf)>. Acesso em: 14 abr. 2021.

ÖZGÜR, Arzucan; CETIN, Burak; BINGOL, Haluk. Co-occurrence network of reuters news. **International Journal of Modern Physics**, v. 19, n. 05, p. 689-702, 2008. Disponível em: <<https://arxiv.org/abs/0712.2491>>. Acesso em: 14 abr. 2021.

QI, Peng et al. Stanza: A python natural language processing toolkit for many human languages. **Association for Computational Linguistics (ACL) System Demonstrations**, 2020. Disponível em:

<<https://www.aclweb.org/anthology/2020.acl-demos.14/>>. Acesso em: 14 abr. 2021.

SILVA, Danillo da Conceição Pereira. (Meta) pragmática da violência linguística: Patologização das vidas trans em comentários online. **Trabalhos em Linguística Aplicada**, v. 58, n. 2, p. 956-985, 2019. Disponível em: <[https://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-18132019000200956](https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-18132019000200956)>. Acesso em: 14 abr. 2021.

VELHO, Eduardo Gabriel. **Análise das apropriações do anonimato nas subculturas dos *imageboards***. 2018. 216 f. Dissertação (Mestrado em Diversidade Cultural e Inclusão Social) – Instituto de Ciências Sociais Aplicadas, Universidade Feevale, Novo Hamburgo, 2018. Disponível em: <<https://biblioteca.feevale.br/Vinculo2/000011/00001100.pdf>>. Acesso em: 14 abr. 2021.

**Recebido:** 14/04/2021  
**Aprovado:** 25/04/2022  
**DOI:** 10.3895/rts.v18n52.14048

**Como citar:** VELHO, E.G.; MONTARDO, S.P. “Queremos impeachment”: analisando os comentários de leitores do G1 sobre Jair Bolsonaro através de uma rede de palavras. *Rev. Tecnol. Soc.*, Curitiba, v. 18, n. 52, p.146-161, jul./set., 2022. Disponível em: <https://periodicos.utfpr.edu.br/rts/article/view/14048>. Acesso em: XXX.

**Correspondência:**

**Direito autoral:** Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

