

Análise de n-gramas: um estudo dos *tweets* sobre a CPI da Covid-19

Gabriel Baldasso¹

Eric Batista Ferreira²

Renata C Bianchi de Barros³

Anderson C. Soares de Oliveira⁴

Crysttian Arantes Paixão⁵

Lia Hanna Martins Morita⁶

Resumo

No final do mês de fevereiro de 2020, o primeiro caso do vírus SARS-CoV-2 foi identificado no Brasil. Enquanto, o Mundo adota medidas drásticas para conter o avanço da doença, no Brasil, a tomada de medidas sanitárias tardias pelo Estado levou o País a apresentar uma das piores taxas epidêmicas mundiais. Antecipações como a compra de vacinas e a decretação de *lockdown* foram inúmeras vezes rejeitadas. Frente à deficiência do Ministério da Saúde e do Governo Federal em gerir adequadamente a crise que crescia desde então, uma Comissão Parlamentar de Inquérito (CPI) foi instaurada em abril de 2021 com a finalidade de investigar ações e omissões do Governo Federal no enfrentamento à pandemia. Nesse contexto, este trabalho teve o objetivo de realizar uma análise textual das publicações dos usuários do *Twitter* a respeito da CPI. Para isso, foi utilizado a análise de n-gramas, que consiste em um modelo de identificação de n *string* de texto que aparecem em sequência e retêm a mesma ordem do texto de origem, com o intuito da identificação dos termos mais frequentes. Utilizando o pacote *rtweet* do software R Core Team foram coletadas diariamente amostras de 36.000 *tweets* com a palavra chave: “CPI” no período compreendido entre 18 de abril a 01 de maio de 2021. Os *tweets* coletados passaram por um processo de limpeza e extração dos bigramas, trigramas e quadrigramas, por meio do pacote *tidytext*. Os resultados mostraram que o bigrama “renan calheiros” foi o mais frequente nas duas semanas. O trigrama e quadrigrama mais frequentes foram respectivamente: “cpi vai blindar” ; “covid é investigado” ; “cpi vai blindar governadores” ; “nenhum acordo q envolva”, Com isso, foi identificado que desde o lançamento da CPI, esta apresentou diferentes abordagens, as quais podem ser verificadas nos n-gramas mais frequentes, permitindo analisar a repercussão, com destaque ao relator da CPI e a convocação dos governadores para depoimentos.

Palavras-chave: *n-grams*; *Twitter*; CPI; covid; mineração de texto.

¹ Mestrando em Estatística Aplicada e Biometria (UNIFAL-MG) - gabrielbaldasso@gmail.com

² Professor do Departamento de Estatística da Universidade Federal de Alfenas (Unifal-MG) - eric.ferreira@unifal-mg.edu.br

³ Pesquisadora do Grupo de Pesquisa Linguagem e Subjetivação da Universidade Federal de Uberlândia (UFU), – renatabiabarros@gmail.com

⁴ Professor no Instituto de Ciências Exatas e da Terra (UFMT) - anderson.oliveira@ufmt.br

⁵ Professor na Universidade Federal de Santa Catarina (UFSC) - crysttian.arantes.paixao@ufsc.br

⁶ Professora no Instituto de Ciências Exatas e da Terra (UFMT) - profaliaufmt@gmail.com