

Análise de sentimentos - CPI covid 19

Antonio Andrade

08/06/2022

Este estudo é apenas para uso pessoal, para exercitar o processamento de linguagem natural. Não expressam minhas opiniões e posicionamentos político-ideológicos.

1. Introdução

A pandemia de COVID-19 causou efeitos nefastos à população mundial. Até a escrita deste ensaio, os dados dão conta de que, no Brasil, 666.971 vidas foram ceifadas pela COVID-19, e o número acumulado de casos confirmados atingiu patamares próximos à estratosfera, ou algo na ordem dos 31,14 milhões

(<https://ourworldindata.org/coronavirus/country/brazil> - acesso em 04/06/2022).

No auge ou pico da pandemia, em 01 de Abril de 2021, foram registradas 3.107 mortes no território nacional (<https://ourworldindata.org/coronavirus/country/brazil> - acesso em 04/06/2022). Tendo em vista este cenário devastador de proporções globais, as entidades sanitárias e a “Organização Mundial de Saúde” (OMS) propuseram algumas medidas no sentido de conter o avanço da pandemia e por conseguinte as complicações entre os indivíduos em geral, e nos grupos vulneráveis (aqueles com comorbidades e idosos).

Dentre as sugestões propostas estão: manter distanciamento físico entre indivíduos de pelo menos 1 metro de distância; evitar multidões e contato muito próximo; uso da máscara facial devidamente ajustada quando o distanciamento físico não for possível e em ambientes mal ventilados; limpar as mãos frequentemente com álcool gel ou água e sabão; vacinar-se assim que chegar a sua vez

(<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public>).

E é exatamente neste último item sugerido que convergem as discussões as quais vão suscitar indiretamente o objeto de estudo do presente ensaio. Primeiro foram as pontuações referentes à leniência do poder público no enfrentamento das questões relacionadas à pandemia. A vacina é apontada como uma medida eficaz para conter a pandemia. Inicialmente havia uma escassez de vacinas disponíveis no mercado farmacêutico internacional.

A demora ou ausência na declaração de interesse de compra de vacina frente às multinacionais produtoras poderia representar um risco de aquisição de um número menor de doses necessárias para atender com equidade toda a população brasileira. Outro ponto levantado foram as questões referentes ao superfaturamento do preço das doses a serem adquiridas, além da suspeição de desvio de dinheiro público para

interesses particulares quando da compra das doses. Também foram levantadas suspeitas relativas ao desvio de verbas destinadas à compra de insumos necessários ao enfrentamento da pandemia, aventando-se a hipótese de desvio de dinheiro do erário público por parte dos Estados e municípios que compõem a Federação.

Como se não bastassem todos esses inconvenientes que poderiam ou não se confirmar no curso das investigações e depoimentos à CPI, numa situação de calamidade pública declarada, ainda pululavam as denúncias de “fakenews” relativas às questões vacinais, que de certa maneira afugentavam alguns setores da sociedade que ainda tinham receio da vacinação.

Visando analisar tais questões, instaurou-se em abril de 2021 a Comissão Parlamentar de Inquérito da Pandemia (CPIPANDEMIA). O seu principal objetivo era “investigar o Governo Federal, em razão da violação dos direitos fundamentais básicos de toda a população brasileira à vida e à saúde, e por ter deixado de seguir as orientações científicas de autoridades sanitárias de caráter mundial, incluindo a Organização Mundial de Saúde” (<https://legis.senado.leg.br/comissoes/mnas?codcol=2441&tp=3> acesso em 04/06/2022).

O desenrolar da CPI foi marcado por depoimentos, pronunciamentos e discursos. É sabido que os discursos políticos remontam a tempos antigos [Champion, 2000; MacDonald - google]. No transcorrer das sessões dos depoimentos na CPIPANDEMIA, era de se esperar que a tônica dos discursos proferidos pelos parlamentares ficasse concentrada nas questões pertinentes às investigações sob suspeição à época. O cerne dos trabalhos parlamentares da comissão seria investigar pontos de interesse público, à luz de demonstrações, argumentos e proposições racionais, lógicas e científicas dos fatos investigados. Entretanto, e ao que parece, os discursos foram revestidos de interesses políticos, visando, por exemplo, enaltecer as medidas tomadas e em vigor (governo), ou com vistas às próximas eleições (oposição). Para entender o que é política, precisamos saber o que os atores políticos estão dizendo e escrevendo. A linguagem e a retórica são meios para entendimento dos conflitos políticos (MacDonald, 2019). Os discursos podem ser, em certa medida, instrumentos para inferir sobre agendas políticas e influenciar a opinião dos eleitores (Oliveira, 2020; Brasil, Capella, 2019).

O uso de textos como dados tem crescido exponencialmente nas ciências políticas (Grimmer & Stewart, 2013; Lucas et al., 2015; Tucker et al., 2020). O texto sempre foi fonte de dados nas ciências políticas, e recentemente, devido ao desenvolvimento de pacotes ou softwares de análises, tornou-se possível a análise quantitativa de um enorme volume de textos (Wilkerson and Casas, 2017).

Davi Moreira (2016) analisou mais de 127 mil pronunciamentos de 2.000 oradores diferentes ao longo de 15 anos de atividade parlamentar na Câmara dos deputados. Ele concluiu que as atividades destes integrantes do pleito não são integralmente pautadas pela relação governo-oposição, indicando influências de outras variáveis como questões ideológicas, de gênero, senioridade e a própria popularidade do parlamentar. Izumi 2017, usando texto como fonte de dados, analisou o discurso de senadores brasileiros de 1995 a 2014 para estimar a posição política dos partidos

brasileiros. Identificou que, ao invés de uma divisão ideológica, os discursos dos senadores organizam os partidos políticos representando o conflito entre governo e oposição.

Diante destas observações, o objetivo deste ensaio é responder à questão: “existem diferenças nos discursos proferidos pelos senadores governistas e de oposição durante a realização da CPI da Pandemia?”.

Neste ensaio foi utilizado a análise quantitativa dos discursos transcritos pelos senadores durante a realização da CPI da pandemia. A análise inicial focou apenas no discurso de 3 senadores de cada um dos grupos (governista e de oposição).

A princípio, a base de dados usada será de um “wabscraping” de discursos da “CPI da Pandemia”, disponível em <https://basedosdados.org/dataset/br-senado-cpipandemia>.

Antes de enumerar os achados, chegar a quaisquer conclusões ou fazer alguma inferência, acredito que seja necessário enfatizar que como escritor deste ensaio, declaro não haver nenhum conflito de interesse inerente a estas discussões, assegurando para inteiramente manter um esforço hercúleo de neutralidade, reservando ou concentrando apenas às questões metodológicas usadas para tentar responder às questões ou hipóteses formuladas.

Carregando os pacotes necessários:

```
library(tidyverse)

## -- Attaching packages -----
## tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.0.3      v dplyr  1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## Warning: package 'ggplot2' was built under R version 4.0.4

## -- Conflicts -----
## tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(ggExtra)

## Warning: package 'ggExtra' was built under R version 4.0.5

library(magrittr)

##
## Attaching package: 'magrittr'
```

```

## The following object is masked from 'package:purrr':
##
##      set_names

## The following object is masked from 'package:tidyr':
##
##      extract

library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##      date, intersect, setdiff, union

library(stringr)
library(stopwords)      #p/palavras "quebra de frase"

## Warning: package 'stopwords' was built under R version 4.0.5

library(tm)             #text mining

## Warning: package 'tm' was built under R version 4.0.5

## Loading required package: NLP

## Warning: package 'NLP' was built under R version 4.0.3

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
##
##      annotate

##
## Attaching package: 'tm'

## The following object is masked from 'package:stopwords':
##
##      stopwords

library(tidytext)

library(lexiconPT) #Lexicon da Língua Portuguesa

## Warning: package 'lexiconPT' was built under R version 4.0.3

```

2 Importação dos arquivos

```

# Dados disponíveis em <https://basedosdados.org/dataset/br-senado-cpipandemia> .

```

```
#
discursos <- read_csv('discursos.csv') # "rodando" discurso no console,
observa-se que ele já está no formato de um tibble.

## Parsed with column specification:
## cols(
##   .default = col_character(),
##   data_sessao = col_date(format = ""),
##   horario_inicio_discurso = col_time(format = ""),
##   horario_fim_discurso = col_time(format = ""),
##   duracao_discurso = col_double()
## )

## See spec(...) for full column specifications.

head(discursos, 10)

## # A tibble: 10 x 21
##   sequential_sess~ data_sessao sigla_partido sigla_uf_partido
bloco_parlament~
##   <chr>           <date>      <chr>           <chr>           <chr>
## 1 01             2021-04-27 PSD             BA             <NA>
## 2 01             2021-04-27 PP              PI             Bloco
Parlament~
## 3 01             2021-04-27 PSD             BA             <NA>
## 4 01             2021-04-27 PP              PI             Bloco
Parlament~
## 5 01             2021-04-27 PSD             BA             <NA>
## 6 01             2021-04-27 PP              PI             Bloco
Parlament~
## 7 01             2021-04-27 PSD             BA             <NA>
## 8 01             2021-04-27 PP              PI             Bloco
Parlament~
## 9 01             2021-04-27 PSD             BA             <NA>
## 10 01            2021-04-27 PP              PI             Bloco
Parlament~
## # ... with 16 more variables: nome_discursante <chr>,
genero_discursante <chr>,
## #   categoria_discursante <chr>, texto_discurso <chr>,
## #   horario_inicio_discurso <time>, horario_fim_discurso <time>,
## #   duracao_discurso <dbl>, sinalizacao_pela_ordem <chr>,
## #   sinalizacao_questao_ordem <chr>, sinalizacao_fora_microfone <chr>,
## #   sinalizacao_responder_questao_ordem <chr>,
## #   sinalizacao_por_videoconferencia <chr>,
sinalizacao_para_interpelar <chr>,
## #   sinalizacao_para_expor <chr>, sinalizacao_para_depor <chr>,
## #   sinalizacao_como_presidente <chr>
```

Categoria dos discursantes:

Categoria dos discursantes (Senador/a):

```
# GOVERNISTAS:
discursos_senadores_governistas <- filter(discursos,
categoria_discursante == 'Senador/a')%>%
  filter(nome_discursante == "JORGINHO MELLO" | nome_discursante ==
"CIRO NOGUEIRA" | nome_discursante == "MARCOS ROGÉRIO") %>%
  select(nome_discursante, texto_discurso)

# OPOSIÇÃO:
discursos_senadores_oposicao <- filter(discursos, categoria_discursante
== 'Senador/a')%>%
  filter(nome_discursante == "HUMBERTO COSTA" | nome_discursante ==
"TASSO JEREISSATI" | nome_discursante == "OTTO ALENCAR")%>%
  select(nome_discursante, texto_discurso)
```

Preprocessamento padrão do texto.

```
# GOVERNISTAS
discursos_senadores_governistas$texto_discurso <-
discursos_senadores_governistas$texto_discurso %>%
  str_to_lower() %>% # Caixa baixa.
  str_replace_all(" *-+ *", "") %>% # Remove hífen.
  str_replace_all("[[:punct:]]", " ") %>% # Pontuação por espaço.
  removeNumbers() %>% # Remove números.
  trimws() # Sem espaços nas bordas.

# OPOSIÇÃO
discursos_senadores_oposicao$texto_discurso <-
discursos_senadores_oposicao$texto_discurso %>%
  str_to_lower() %>% # Caixa baixa.
  str_replace_all(" *-+ *", "") %>% # Remove hífen.
  str_replace_all("[[:punct:]]", " ") %>% # Pontuação por espaço.
  removeNumbers() %>% # Remove números.
  trimws() # Sem espaços nas bordas.
```

Remoção das stop words.

```
# GOVERNISTAS
discursos_senadores_governistas$texto_discurso <-
discursos_senadores_governistas$texto_discurso %>%
  removeWords(words = c("bom", "muito", "pouco", "obrigado", "obrigada",
"dar", "sa", "favor", "roberto", "sr", "presidente", "v", "exa", "sras",
"srs", #removento obrigado e obrigada
stopwords('pt'))))

# OPOSIÇÃO
discursos_senadores_oposicao$texto_discurso <-
discursos_senadores_oposicao$texto_discurso %>%
  removeWords(words = c("bom", "muito", "pouco", "obrigado", "obrigada",
"dar", "sa", "favor", "roberto", "sr", "presidente", "v", "exa", "sras",
"srs", #removento obrigado e obrigada
stopwords('pt'))))
```

Realizando a “tokenização” do texto.

```
# GOVERNISTAS
discursos_senadores_gov_unnested <- discursos_senadores_governistas %>%
  unnest_tokens(term, texto_discurso)
  # o primeiro argumento da fç unnest_tokens é o novo nome da
  # coluna de saída que será criada, e o # segundo é a fonte, de onde será
  # retirada.
# Notar que o número da linha de onde a palavra veio é mantido.
```

```
discursos_senadores_gov_unnested
```

```
## # A tibble: 61,258 x 2
##   nome_discursante term
##   <chr>             <chr>
## 1 CIRO NOGUEIRA     senadores
## 2 CIRO NOGUEIRA     achava
## 3 CIRO NOGUEIRA     deveríamos
## 4 CIRO NOGUEIRA     suspender
## 5 CIRO NOGUEIRA     atual
## 6 CIRO NOGUEIRA     sessão
## 7 CIRO NOGUEIRA     sanado
## 8 CIRO NOGUEIRA     vício
## 9 CIRO NOGUEIRA     vejo
## 10 CIRO NOGUEIRA    ser
## # ... with 61,248 more rows
```

```
# OPOSIÇÃO
discursos_senadores_opo_unnested <- discursos_senadores_oposicao %>%
  unnest_tokens(term, texto_discurso)
  # o primeiro argumento da fç unnest_tokens é o novo nome da
  # coluna de saída que será criada, e o # segundo é a fonte, de onde será
  # retirada.
# Notar que o número da linha de onde a palavra veio é mantido.
```

```
discursos_senadores_opo_unnested
```

```
## # A tibble: 60,337 x 2
##   nome_discursante term
##   <chr>             <chr>
## 1 OTTO ALENCAR      invocando
## 2 OTTO ALENCAR      proteção
## 3 OTTO ALENCAR      deus
## 4 OTTO ALENCAR      declaro
## 5 OTTO ALENCAR      aberta
## 6 OTTO ALENCAR      sessão
## 7 OTTO ALENCAR      eleição
## 8 OTTO ALENCAR      quórum
## 9 OTTO ALENCAR      suficiente
## 10 OTTO ALENCAR     abertura
## # ... with 60,327 more rows
```

Calculando a polaridade

uma amostra do dicionario de termos rotulados

```
sample_n(oplexicon_v3.0, size = 20) %>%  
  arrange(polarity)
```

```
##           term type polarity polarity_revision  
## 1   entornados  adj      -1                A  
## 2    caolhas   adj      -1                M  
## 3  draconianos  adj      -1                A  
## 4   escassos   adj      -1                M  
## 5    ignotos   adj      -1                A  
## 6   epileticos  adj      -1                A  
## 7    sonsas    adj      -1                A  
## 8  constrangidos  adj      -1                A  
## 9   emuladora   adj      -1                A  
## 10  monogamo    adj       0                A  
## 11  pantomimar   vb       0                A  
## 12  retardado    adj       1                A  
## 13 encadear-se   vb       1                A  
## 14 tranquilizadora  adj       1                A  
## 15  desvelados   adj       1                A  
## 16   rapidos     adj       1                M  
## 17  costumar-se   vb       1                A  
## 18  comoventes   adj       1                M  
## 19   potente     adj       1                M  
## 20  desbrecar    vb       1                A
```

Contagem por polaridade.

```
oplexicon_v3.0 %>%  
  count(polarity, sort = TRUE)
```

```
## polarity    n  
## 1      -1 14569  
## 2       0  9002  
## 3       1  8620
```

Contagem por classe gramatical.

```
oplexicon_v3.0 %>%  
  count(type, sort = TRUE)
```

```
##           type    n  
## 1          adj 24475  
## 2          vb  6889  
## 3          htag   471  
## 4 vb det n prp   103  
## 5          vb n prp    91  
## 6          vb adj    74  
## 7          emot    66  
## 8          vb adv    22
```



```
# GOVERNO
```

```
# Faz o a junção por interseção.
```

```
tb_sen_gov <- inner_join(discursos_senadores_gov_unnested,  
  oplexicon_v3.0[, c("term", "polarity")],  
  by = c("term" = "term"))
```

```
# Agora os termos têm sua polaridade presente na tabela.
```

```
sample_n(tb_sen_gov, size = 20)
```

```
## # A tibble: 20 x 3
```

	nome_discursante	term	polarity
	<chr>	<chr>	<int>
## 1	MARCOS ROGÉRIO	considerados	1
## 2	MARCOS ROGÉRIO	sancionado	0
## 3	MARCOS ROGÉRIO	primeiro	0
## 4	JORGINHO MELLO	defensores	1
## 5	JORGINHO MELLO	ver	1
## 6	MARCOS ROGÉRIO	votar	-1
## 7	MARCOS ROGÉRIO	iminente	0
## 8	MARCOS ROGÉRIO	partir	-1
## 9	MARCOS ROGÉRIO	dizer	0
## 10	MARCOS ROGÉRIO	completo	0
## 11	MARCOS ROGÉRIO	manifestar	-1
## 12	JORGINHO MELLO	fazer	0
## 13	MARCOS ROGÉRIO	plano	0
## 14	JORGINHO MELLO	agradecer	0
## 15	CIRO NOGUEIRA	desesperado	-1
## 16	CIRO NOGUEIRA	independente	0
## 17	JORGINHO MELLO	pensar	1
## 18	MARCOS ROGÉRIO	legal	1
## 19	MARCOS ROGÉRIO	ter	0
## 20	MARCOS ROGÉRIO	abusivas	-1

```
# OPOSIÇÃO
```

```
tb_sen_opo <- inner_join(discursos_senadores_opo_unnested,  
  oplexicon_v3.0[, c("term", "polarity")],  
  by = c("term" = "term"))
```

```
# Agora os termos tem sua polaridade presente na tabela.
```

```
sample_n(tb_sen_opo, size = 20)
```

```
## # A tibble: 20 x 3
```

	nome_discursante	term	polarity
	<chr>	<chr>	<int>
## 1	OTTO ALENCAR	querer	1
## 2	HUMBERTO COSTA	estar	1
## 3	OTTO ALENCAR	pacientes	1
## 4	OTTO ALENCAR	vacinas	0
## 5	HUMBERTO COSTA	ser	1

```
## 6 HUMBERTO COSTA esclarecer -1
## 7 OTTO ALENCAR saber 0
## 8 HUMBERTO COSTA dizer 0
## 9 HUMBERTO COSTA brasileiros 0
## 10 OTTO ALENCAR covardes -1
## 11 HUMBERTO COSTA pensa -1
## 12 HUMBERTO COSTA ser 1
## 13 HUMBERTO COSTA perguntar 1
## 14 HUMBERTO COSTA previsto 1
## 15 OTTO ALENCAR falta -1
## 16 OTTO ALENCAR errada -1
## 17 TASSO JEREISSATI apresentar -1
## 18 TASSO JEREISSATI social 0
## 19 HUMBERTO COSTA seguro 1
## 20 HUMBERTO COSTA registrar 1
```

Faz a agregação da polaridade por documento.

GOVERNO

```
tb_gov_agregada <- tb_sen_gov %>%
  group_by(nome_discursante) %>%
  summarise(soma = sum(polarity),
            n = n(),
            sentiment = soma/n)

## `summarise()` ungrouping output (override with `.groups` argument)
```

```
tb_gov_agregada
```

```
## # A tibble: 3 x 4
##   nome_discursante soma      n sentiment
##   <chr>          <int> <int>      <dbl>
## 1 CIRO NOGUEIRA    144   931     0.155
## 2 JORGINHO MELLO   307  1847     0.166
## 3 MARCOS ROGÉRIO  1176  7790     0.151
```

OPOSIÇÃO

```
tb_opo_agregada <- tb_sen_opo %>%
  group_by(nome_discursante) %>%
  summarise(soma = sum(polarity),
            n = n(),
            sentiment = soma/n)

## `summarise()` ungrouping output (override with `.groups` argument)
```

```
tb_opo_agregada
```

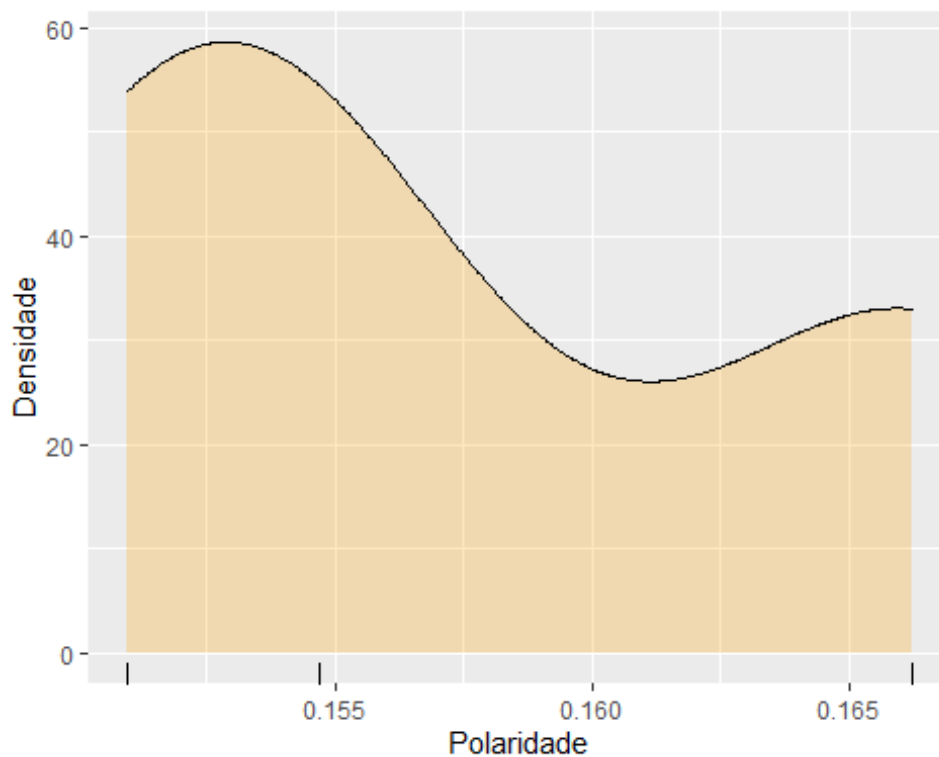
```
## # A tibble: 3 x 4
##   nome_discursante soma      n sentiment
```

##	<chr>	<int>	<int>	<dbl>
## 1	HUMBERTO COSTA	941	4869	0.193
## 2	OTTO ALENCAR	693	4613	0.150
## 3	TASSO JEREISSATI	224	1198	0.187

Desidade expírica kernel do escore de sentimento.

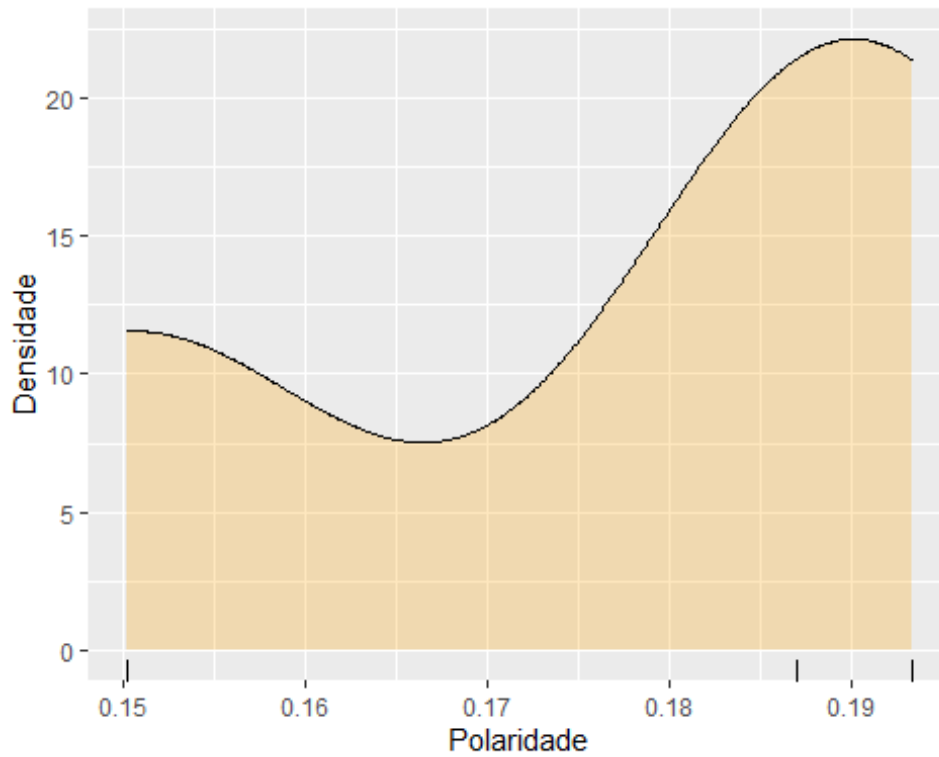
GOVERNO

```
ggplot(tb_gov_agregada, aes(x = sentiment)) +
  geom_density(fill = "orange", alpha = 0.25) +
  geom_rug() +
  labs(x = "Polaridade", y = "Densidade")
```



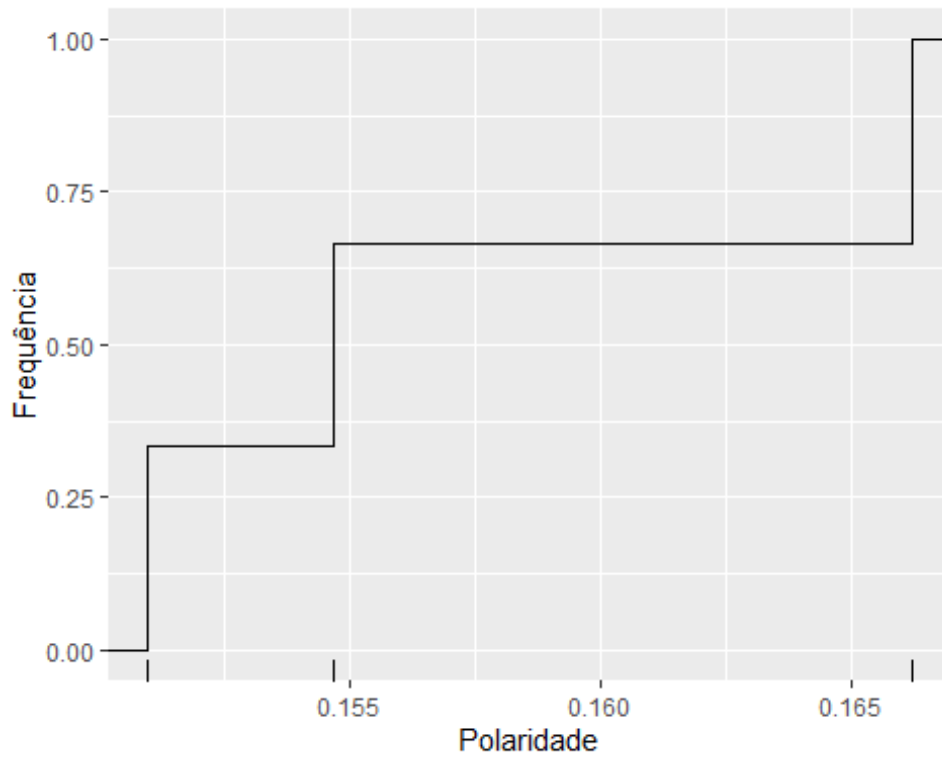
OPOSIÇÃO

```
ggplot(tb_opo_agregada, aes(x = sentiment)) +
  geom_density(fill = "orange", alpha = 0.25) +
  geom_rug() +
  labs(x = "Polaridade", y = "Densidade")
```

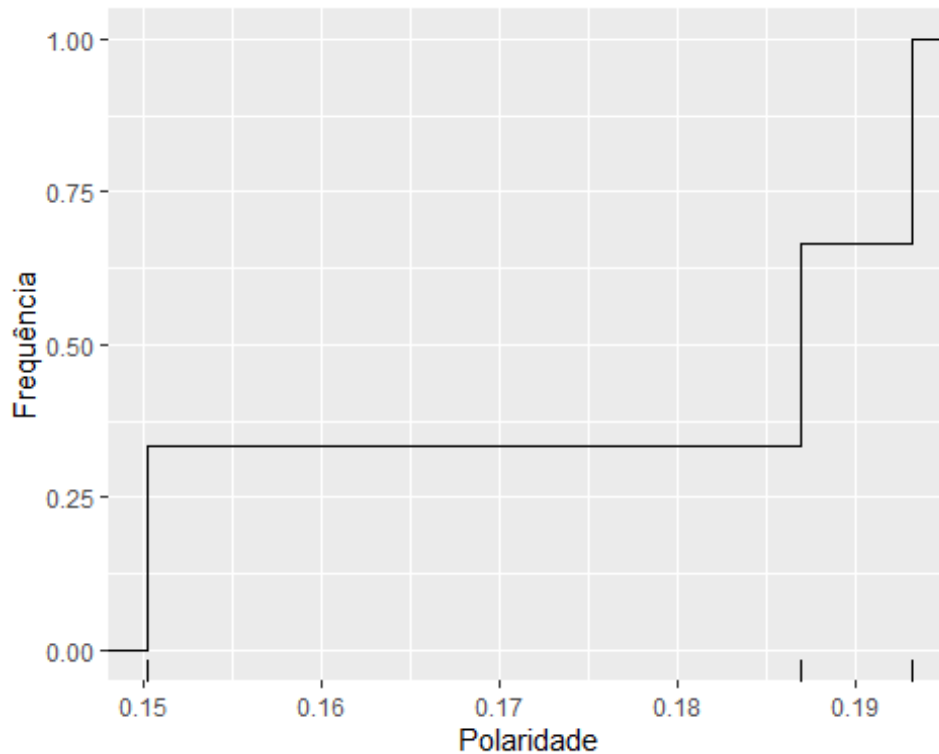


Frequência relativa acumulada.

```
# GOVERNO  
ggplot(tb_gov_agregada, aes(x = sentiment)) +  
  stat_ecdf() +  
  geom_rug() +  
  labs(x = "Polaridade", y = "Frequência")
```



```
# OPOSIÇÃO  
ggplot(tb_opo_agregada, aes(x = sentiment)) +  
  stat_ecdf() +  
  geom_rug() +  
  labs(x = "Polaridade", y = "Frequência")
```



Termos mais frequentes por classe polar.

GOVERNISTAS

Determina as frequências dos termos de polaridade não nula.

```
tb_words_gov <- tb_sen_gov %>%
  count(term, polarity, sort = TRUE) %>%
  filter(polarity != 0)
```

```
tb_cloud <- tb_words_gov %>%
  spread(key = "polarity", value = "n", fill = 0) %>%
  rename("negative" = "-1", "positive" = "1")
tb_cloud
```

```
## # A tibble: 1,235 x 3
##   term          negative positive
##   <chr>         <dbl>     <dbl>
## 1 abandonada      1         0
## 2 abjetos         1         0
## 3 abrangente      0         1
## 4 abrir           6         0
## 5 abrupta        1         0
## 6 absoluta        0        13
## 7 absoluto        0         4
## 8 abstrato        2         0
## 9 absurda         5         0
```

```
## 10 absurdas      1      0
## # ... with 1,225 more rows

tb <- as.data.frame(tb_cloud[, c("negative", "positive")])
rownames(tb) <- tb_cloud$term
head(tb)

##           negative positive
## abandonada      1         0
## abjetos         1         0
## abrangente      0         1
## abrir           6         0
## abrupta         1         0
## absoluta        0        13

library(wordcloud)

## Warning: package 'wordcloud' was built under R version 4.0.5

## Loading required package: RColorBrewer

comparison.cloud(tb,
  colors = c("#F8766D", "#00BFC4"),
  max.words = min(nrow(tb), 150))

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
## max.words =
## min(nrow(tb), : passa could not be fit on page. It will not be
## plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
## max.words =
## min(nrow(tb), : quebrar could not be fit on page. It will not be
## plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
## max.words =
## min(nrow(tb), : suspeitas could not be fit on page. It will not be
## plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
## max.words =
## min(nrow(tb), : impedir could not be fit on page. It will not be
## plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
## max.words =
## min(nrow(tb), : pagar could not be fit on page. It will not be
## plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
## max.words =
```

```
## min(nrow(tb), : pensa could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : trabalhar could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : tranquila could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : confirmar could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : evidente could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : imediato could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : levantar could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : acusado could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : apresentadas could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : apresentados could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : assinar could not be fit on page. It will not be
plotted.
```



```
## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : condensado could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : declinar could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : discutir could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : esconder could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : incomoda could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : oferecer could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : questionado could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : relacionado could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : restritivas could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : chamar could not be fit on page. It will not be
plotted.

## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
```

```
## min(nrow(tb), : determinado could not be fit on page. It will not be
plotted.
```

```
## Warning in comparison.cloud(tb, colors = c("#F8766D", "#00BFC4"),
max.words =
## min(nrow(tb), : encaminhado could not be fit on page. It will not be
plotted.
```



OPOSIÇÃO

```
tb_words_oposicao <- tb_sen_opo %>%
  count(term, polarity, sort = TRUE) %>%
  filter(polarity != 0)

tb_cloud <- tb_words_oposicao %>%
  spread(key = "polarity", value = "n", fill = 0) %>%
  rename("negative" = "-1", "positive" = "1")
tb_cloud

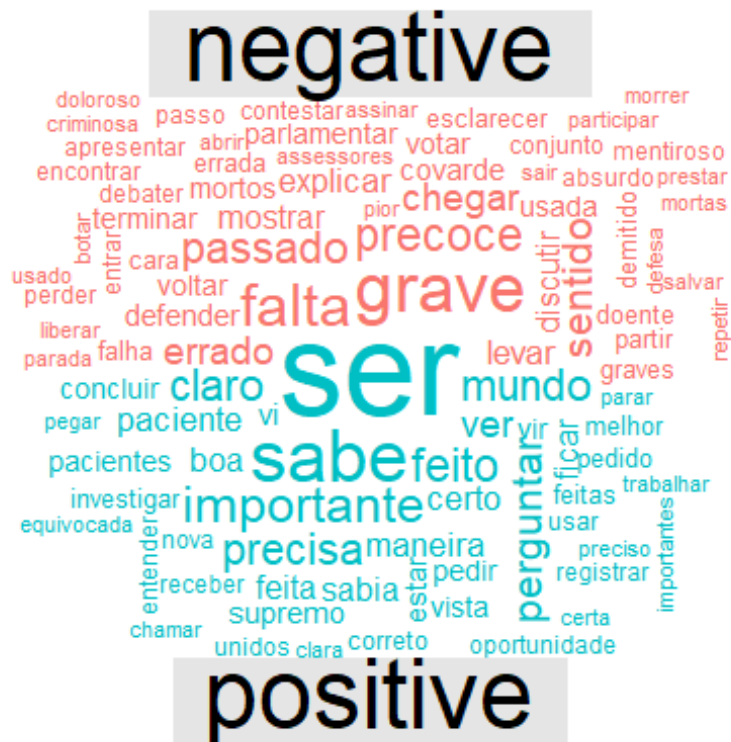
## # A tibble: 1,131 x 3
##   term          negative positive
##   <chr>          <dbl>     <dbl>
## 1 abalizada             0         1
## 2 abandonado           1         0
## 3 abolido              1         0
## 4 abrir                9         0
## 5 absoluta             0        10
```

```
## 6 absoluto      0      2
## 7 absolutos    0      1
## 8 absolver     1      0
## 9 absurda      2      0
## 10 absurdo     12     0
## # ... with 1,121 more rows

tb <- as.data.frame(tb_cloud[, c("negative", "positive")])
rownames(tb) <- tb_cloud$term
head(tb)

##           negative positive
## abalizada         0         1
## abandonado        1         0
## abolido           1         0
## abrir             9         0
## absoluta          0        10
## absoluto          0         2

library(wordcloud)
comparison.cloud(tb,
  colors = c("#F8766D", "#00BFC4"),
  max.words = min(nrow(tb), 100))
```



Diversidade de palavras, presentes nos discursos

(A diversidade de palavras pode ser determinada pelo quociente entre o número total de palavras “únicas” (unique), pelo número total de palavras).

```
# GOVERNO
total_palavras_gov <- nrow(na.omit(tb_sen_gov))
total_palavras_distintas_gov <- distinct(tb_sen_gov, term)

# >>> Entre os governistas, a diversidade de palavras foi algo em torno
de 18.6 %

# OPOSIÇÃO
total_palavras_opo <- nrow(na.omit(tb_sen_opo))
total_palavras_distintas_opo <- distinct(tb_sen_opo, term)

# >>> Entre os oposicionista, a diversidade de palavras ficou em 17.1 %
```

Frequência de sentimentos negativos

```
freq_sentimentos_negativos_gov <- filter(tb_sen_gov, polarity == -1)
```

Resposta: retornou um tibble com 2021 linhas

```
freq_sentimentos_negativos_opo <- filter(tb_sen_opo, polarity == -1)
```

Resposta: retornou um tibble com 1943 linhas

Apontamentos_Conclusões

Os dados apresentados perfazem um total de 7.016 discursos, sendo 4.395 destes proferidos pelos três senadores governistas (Ciro Nogueira, Jorginho Melo e Marcos Rogério) e 2621 discursos pelos senadores oposicionistas (Humberto Costa, Tasso Jereissati e Otto Alencar). Trata-se de um corpus de fonte de dados passível de ser investigado com estas ferramentas disponíveis. Mais análises devem ser executadas para entender melhor a distribuição dos dados entre ambos os grupos.

Nas “wordclouds” não observa-se algo que tornem os grupos tão distintos, ou alguma palavra que se destaque entre os 2 grupos. Talvez entre os oposicionistas as palavras “falta” e “grave”, numa alusão à gravidade da pandemia.

A diversidade de palavras também apresentou distribuição semelhante, com palavras distintas perfazendo um total de 18.6% para governistas e 17.1 % entre os oposicionistas. A frequência de sentimentos negativos também se mostrou equiparada ou semelhante entre os grupos distintos. Nestes casos, para uma análise mais pormenorizada, cabe a realização de um teste-t para a comparação da diferença entre os 2 grupos.

Referências

Brasil, Felipe Gonçalves; Capella, Ana Claudia Niedhardt. Agenda governamental brasileira: uma análise da capacidade e diversidade nas prioridades em políticas públicas no período de 2003 a 2014. Cadernos Gestão Pública e Cidadania. São Paulo, v. 24, n. 78, p.1-22. 2019

Cordeiro Moreira, Davi. Com a palavra os nobres deputados - frequencia e ênfase temática dos discursos dos parlamentares brasileiros. 2016. Tese (Doutorado em Ciência Política) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2016. DOI 10.11606/T.8.2016.tde-04112016-124733. Disponível em: <https://teses.usp.br/teses/disponiveis/8/8131/tde-04112016-124733/pt-br.php>. Acesso em: 01 jun. 2022.

Grimmer, J., & Stewart, B. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 21(3), 267-297. doi:10.1093/pan/mps028

Lucas, C.; Nielsen, R.; Roberts, M.; Stewart, B.; Storer, A.; & Tingley, D. (2015). Computer-Assisted Text Analysis for Comparative Politics. Political Analysis, 23(2), 254-277. doi:10.1093/pan/mpu019

MacDonald, Michael. The Oxford Handbook of Rethorical Studies. Oxford University Press, 2017. E-book. Disponível em: https://play.google.com/books/reader?id=_S82DwAAQBAJ&pg=GBS.PA9.w.1.1.163_21. Acesso em: 1 jun. 2022.

Our World in Data. Disponível em: <https://ourworldindata.org/coronavirus/country/brazil>. Acesso em 01/06/2022.

SILGE, J.; RODINSON, D. Text Mining with R: A tidy approach. 1st ed. O'Reilly Media, 2017.

Tucker, Ethan C.; Capps, Colton J.; Shamir, Lior. A data science approach to 138 years of congressional speeches. Heliyon, v.6, Issue 8, 2020, e04417, ISSN 2405-8440, DOI: 10.1016/j.heliyon.2020.e04417. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2405844020312615>

Wilkerson, John; Casas, Andreau. Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges. Annual Review of Political Science. v.20, p.529-544, 2017.

World Health Organization. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public>

Yoshida Izumi, Maurício. Velhas questões, novos métodos: posições, agenda, ideologia e dinheiro na política brasileira. 2017. Tese (Doutorado em Ciência Política) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2017. DOI 10.11606/T.8.2018.tde-11042018-114905 Disponível em:

<https://teses.usp.br/teses/disponiveis/8/8131/tde-11042018-114905/pt-br.php>
Acesso em: 01 jun. 2022.