

YouTube communication of Italian Political Parties

Maria Ascolese

March 11, 2022

Contents

1	Introduction	1
2	Hypotheses	1
3	Data	2
4	Methods	2
5	Results	3
6	Conclusion	3
7	References	3

1 Introduction

In this paper I use the `YouTube Data API v3` to collect data and metadata from the official YouTube channels of major Italian Parties in order to investigate content and language occurrences in YouTube videos. The goal of this short analysis would be to understand each Party's topics and language of choice, and I believe that their YouTube channels might be accurate small representations of their communicative goals, as all video content on the channels has been necessarily chosen and published by the channel owner and, therefore, intentionality must be assumed (Vesnic-Alujevic & Van Bauwel, 2014).

2 Hypotheses

H1 Parties with common ideologies also show common keywords, emotional words and syntax.

H2 Each party's keywords, emotional words and syntax change dramatically over time.

3 Data

I used the R package **Tuber** to collect YouTube videos data. **Tuber** consists of a number of functions useful to extract data from the YouTube platform via the YouTube Data API v3. I conducted the queries using each political party's ID in the `list_channel_videos` function, which resulted in obtaining metadata for each video published on every channel. The tool provided metadata for: - 20000 videos from **Movimento 5 Stelle**'s channel; - 9548 videos from **Lega Salvini Premier**'s channel; - 1955 videos from **Partito Democratico**'s channel; - 5524 videos from **Fratelli d'Italia**'s channel; - 60 videos from **Forza Italia**'s channel.

Since **Tuber**'s function to extract subtitles from YouTube videos has been recently disabled, I used the R package **youtubecaptions** instead. The function `get_caption` gives back a number of string characters for each video, which correspond to the unique captions appearing on the screen as the video plays. However, the `get_caption` function needs URLs, which I produced with a basic YouTube video URL and video IDs.

```
#Example of URLs produced for Fratelli d'Italia's videos

fdi_url <- str_c("https://www.youtube.com/watch?v=",
                fdi_ch$contentDetails.videoId[1:5524])

#Example of for loop, which I used to get the captions for Partito Democratico's videos

for (i in pd_data$pd_url[1:1955]) {
  tryCatch({
    v <- get_caption(i, "it")
  }, error = function(e){})

  print(i)
  pd_captions <- append(pd_captions, v)
  Sys.sleep(1)
}
```

I then joined them in single string characters, obtaining whole transcriptions for each video of the channels.

4 Methods

After preparing the data in order to have clean, tidy datasets for each political party's channel, I finally came to the analysis, for which I used the **tidytext** package.

For each party's channel, I split the captions in words in order to remove stopwords (like conjunctions and auxiliary verbs, but also some adverbs that I manually added to the list), I removed encoding errors (like some accent marks) and some YouTube transcription errors (like numbers being transcribed with digits instead of words). When I finally ended up with clean lexemes of interest, for each party I obtained the count of each word and sorted them.

```
#Example of tokenization conducted on Partito Democratico's transcriptions.
#I actually used two stopwords removal processes, which can be seen completely in the RScripts

text <- data_frame(Text = pd_dataset$text)

words <- text %>%
  unnest_tokens(output = word, input = Text)
```

```
wordcounts <- words %>%
  count(word, sort = TRUE)

for(i in 1:nrow(wordcounts)) {
  wordcounts$word[i] <- removeWords(wordcounts$word[i], sw_ita)

  print(i)
}

wordcounts <- wordcounts[-which(wordcounts$word == ""), ]
```

At a first glance, it appeared that the most frequent words were common between the parties: **lavoro**, **partito** and **persone**, to name a few.

At this point, I was ready to visualize the data. In order to look into my first hypothesis, I wanted to compare the frequency of words between parties.

5 Results

6 Conclusion

7 References