Analysing Mexico's President speeches: A study of political propaganda and populist rhetoric in AMLO's press conferences

Luis Valentin Cruz

2023-05-02

Contents

Abstract

The 2018 federal elections in Mexico brought significant changes in the country's political landscape, as Andrés Manuel López Obrador (AMLO) became president and initiated daily press conferences called "The mornings" (Las mañaneras in Spanish) to keep the public informed about the government's agenda. Despite the goal of promoting transparency and accountability, the morning speeches have been criticised for potentially being used as propaganda or perpetuating populist rhetoric. This project aims to determine whether the president uses these press conferences for political propaganda or to promote populist ideas. The study analyses official stenographic versions of the speeches from December 2018 to April 2023 using topic modelling and lexicon-based methods. The findings indicate that while the president employs populist rhetoric in his speeches, it only accounts for a small percentage of the speech, ranging from 2% to 4%. Additionally, some topics discussed in the speeches may be used for propaganda purposes.

Introduction

The federal elections of July 2018 marked a before and after in Mexican politics. Not only because they were the largest elections the country has ever held in terms of the number of elected positions and the number of voters on the electoral roll (Montero, 2019), but also because of the impact generated by the arrival to power of the party 'Movimiento Regeneración Nacional' (known as 'Morena'), led by AMLO, which represented not only a party alternation but also an ideological one (Aragón Falomir et al., 2019) since for the first time in Mexico's political history a self-described left-wing party came to power.

AMLO began his six-year presidential term when he took office as President of Mexico on December 1, 2018. Since day one, he has been very committed to fulfilling his campaign promises. One of his campaign promises was to inform the public about current government issues, as he has always been critical of the lack of information and transparency in decisions made by previous governments led by the two hegemonic parties (PRI and PAN). To achieve this goal, he created what eventually became known as 'The mornings' which are press conferences led by the President to inform the public about current government agenda

issues. This practice was originated during his tenure as mayor of Mexico City in the 2000s but has now been expanded to a national level (El Economista, 2022).

Nonetheless, despite the stated aim of promoting transparency and accountability, the morning speeches of the president have received a lot of criticism since the very beginning. Some argue that the president is instead using these speeches to perpetuate populist rhetoric (see for example Chicago Tribune (2019) and Romeu (2022)) while others argue that these conferences are an act of propaganda to gain popularity among voters (for instance (BBC, 2019), Publimetro (2022) and Spdnoticias (2023)). In this context, are the daily conferences really being used as a means of propaganda or to perpetuate populist rhetoric?

Before going further into the analysis presented in this project, it is crucial to establish a clear understanding of what we mean when we use the terms "propaganda" and "populism". The Federal Code of Electoral Institutions and Procedures, in its third paragraph of article 228, defines "electoral propaganda" in Mexico as the set of writings, publications, images, recordings, projections and expressions, created and distributed by political parties, registered candidates, and their supporters during the election campaign, for the purpose of presenting the registered candidacies to the public.

As for the term "populism," there is no consensus on its definition since there are many different interpretations of this concept. According to the discussion in Wirth et al. (2016) paper, populism can be seen as a set of interconnected political ideas that concern the structure of power and society (a relational concept). To understand these connections, we need to identify the key elements of populism.

The first and most important aspect of populism is the idea of "the people" having sovereignty, meaning they control the locus of power. However, the interpretation of "the people" varies among different populist groups depending on the context. Often, "the people" are portrayed as a uniform, social entity or community with good values. The second component of populism is the adversary group, which is the elite. This is a diverse group (in terms of political, economic, cultural or intellectual status) that is viewed as "corrupt," "exploitative," "anti-popular," "immoral," and so on. The elite are blamed for betraying the people, controlling their rights, welfare, and progress without justification. They have failed to keep their promises, disregarding the interests of the people and manipulating democracy for their benefit. The third element is the populist actor, who criticizes the elite for depriving the people of their sovereignty and strives to restore their power and voice. The populist actor can be a party, movement, or even an individual, usually a charismatic figure. Lastly, there are the "others," who are considered dangerous and are also excluded from the "good" people. These individuals are regarded as a threat from within the population. There are several population segments that can be a focus of populist resentment, such as immigrants, people of a different race, criminals, profiteers, perverts, religious groups, and other minorities.

To put it simply, the ideology of populism argues that people have the right to sovereignty, which may be taken away by the elite or other groups, and populist actors aim to protect or restore this sovereignty. As a result, there is an inherent conflict between the people and the elite or other groups, and the populist actor tends to foster a favourable or close connection with the people while maintaining an unfavourable or distant relationship with the elite or others.

In reference to the previous question, the project is intended to accomplish two main goals. The first goal is to determine, through topic modelling, whether the president is using daily press conferences as a means of political propaganda. The second goal is to determine, through lexicon-based methods, whether the president is using populist rhetoric in these morning conferences. To accomplish this, I will use the official stenographic versions of the morning speeches as the primary source of data, covering the period from December 2018 to April 2023. The findings of this study reveals that the president employs populist rhetoric in his speeches, however, it is only present in a small percentage of the speech, ranging from 2% to 4%. Additionally, some topics covered in the speeches may be utilized for the purpose of propaganda.

Motivation

The mornings are a rare phenomenon in the world because no other leader offers almost daily press conferences (Expansión, 2022). This makes them an excellent source of information about the president's thoughts

and beliefs. Additionally, all the content discussed during these press conferences is easily accessible on the government's official websites through stenographic versions.

Despite the criticism that the mornings have been receiving as being populist or propagandistic, only a few papers have analysed AMLO's discourse from this perspective using classical content analysis but none of them have used natural language processing (NLP) techniques to do so.

NLP techniques refer to a set of techniques developed in the field of artificial intelligence that are designed to computationally represent and analyse human language. These techniques were created to facilitate user interaction with computers in natural language. NLP can be divided into two categories: Natural Language Understanding (NLU) and Natural Language Generation (NLG). NLU allows machines to comprehend natural language and analyse it by identifying concepts, entities, emotions, keywords, and other elements. NLG involves generating phrases, sentences, and paragraphs that are semantically meaningful based on an internal representation. Recently NLP gained a lot of attention and has been applied in a wide range of fields, including machine translation, information extraction, summarisation, question answering, and sentiment detection (Khurana, et al., 2022). For this particular project, my focus will be on NLU using lexicon-based methods and topic modelling to analyse the use of populist rhetoric in the mornings.

Finally, I would like to highlight that this project is highly influenced by the paper discussed in the second assignment: "Measuring Populism: Comparing Two Methods of Content Analysis" by Rooduijn and Pauwels (2011) which measures populism over time in the United Kingdom, the Netherlands, Germany and Italy in using two approaches: classical content analysis and a computer-based content analysis. As well as the paper "Propaganda Identification Using Topic Modelling" by Yakunin et. al. (2020)

Data

The rare phenomenon of the mornings allows us to collect information from the president's entire discourse during the morning conferences. This information is available through stenographic versions of the conferences published on the government's official website at https://www.gob.mx/, as well as on the other president's official websites such as https://presidente.gob.mx/ and https://lopezobrador.org.mx/. For the purpose of this project, I have chosen to extract information from the website https://lopezobrador.org.mx/ using dynamic web scraping techniques with `RSelenium`, as detailed below:

Data collection

```r
# Conferences df, initially empty
conferences <- matrix(data = NA, ncol = 2,
                      dimnames = list(NULL, c("date", "text")))

# Launch the navigator
rD <- rsDriver(browser=c("firefox"), port = 12L)
driver <- rD$client

# Navigate in the url
url <- "https://lopezobrador.org.mx/?s=Versión+estenográfica+de+la+conferencia+de+prensa+matutina+del+pr
driver$navigate(url)

# Function that get content for each page
get_content <- function(rD) {
  entry_date <- rD$findElement(using = "css", ".entry-date")
  entry_content <- rD$findElement(using = "css", ".entry-content")

  date <- entry_date$getElementText()
```

```r
  content <- entry_content$getElementText()

  df <- tibble(date = date[[1]], text = content[[1]])
  return(df)
}

# Iterate over all pages
while (TRUE) {
  # Find all post within the class entry-title
  entry_titles <- driver$findElements(using = "css", ".entry-title")

  # For each entry get the information
  for (i in 1:length(entry_titles)) {

  # Click on the element of the list
  entry_titles[[i]]$clickElement()

  # Get the content of this site
  content <- get_content(driver)

  # Save the content in a single df
  conferences <- rbind(conferences, content)

  # Go back to the previous page
  driver$goBack()

  # Overwrite again the links to avoid problems
  entry_titles <- driver$findElements(using = "css", ".entry-title")
  }

  # Click to the next page
  next_button <- driver$findElement(using = "css", ".older")

  # Break if it is the last one
  if (length(next_button) == 0) {
    break
  }
  next_button$clickElement()
}

driver$closeall

# Save the data
write.csv(conferences[2:1034,], file = "conferences_2018_2023.csv")
```

Data cleaning

The stenographic version not only provides information and answers given by AMLO, but also includes interventions from members of his cabinet and journalists. For instance, here is a sample from the press conference on April 11, 2023:

> **PRESIDENTE ANDRÉS MANUEL LÓPEZ OBRADOR:** Bueno, acerca de lo de la marihuana o la comercialización de la marihuana, pienso que es totalmente reprobable que quien

ocupó un cargo como presidente de México decida dedicarse a un negocio de esa naturaleza...

INTERLOCUTORA: Porque además quien fuera director de la Cofepris en ese momento también fue funcionario en el gobierno de Fox, estuvo en la PGR y también en la Cofece.

Then, it is necessary to filter out only the information related to the president. This can be achieved using the following code:

```r
# Loading the data
conferences_df <- read.csv("conferences_2018_2023.csv")

# Regex to split the text at each speaker
pattern <- "\\n(?=[A-ZÁÉÍÓÚÜÑ(][A-ZÁÉÍÓÚÜÑ,\\s()]*?[A-ZÁÉÍÓÚÜÑ()].+?:)"

# Column to keep only president's text
conferences_df$text_amlo <- NA

# Function to know if the text belongs to the president
contains_amlo <- function(text) {
  grepl("PRESIDENTE ANDRÉS MANUEL LÓPEZ OBRADOR:", text)
}

# Iterate over each press conference
for (i in seq(1,nrow(conferences_df))) {
  # Split the text given the regex
  split_ver <- strsplit(conferences_df$text[i],
                        pattern, perl = TRUE) %>% unlist()

  # Filter relevant strings
  relevant_strings <- split_ver[sapply(split_ver, contains_amlo)]

  # Concatenate strings separated by a space
  conferences_df$text_amlo[i] <- paste(relevant_strings, collapse = " ")
}

# Drop innecesary text
replacements <- c("PRESIDENTE ANDRÉS MANUEL LÓPEZ OBRADOR:" = "", "\n" = "")
conferences_df <- conferences_df %>%
                  select(-text, -X) %>%
                  mutate(text_amlo = str_replace_all(text_amlo, replacements))

# Save the information in a new file
write.csv(conferences_df, "president_discourse.csv")
```

Once the cleaning stage is complete, a single data frame with two columns and 1033 entries can be obtained. The first column, named `date`, includes the date of the press conference, which will be later used as the name of the documents in the corpus. The second column, named `text_amlo`, contains the text of all the interventions made by the president during the conference, which will constitute each document of the corpus.

Summary Statistics

To begin the analysis, the first step is to create a corpus in which each press conference will be treated as a document. Next, to generate a document feature matrix (dfm), it is crucial to tokenise the text. However,

considering the amount of data, it is essential to reduce the number of features. This can be achieved by following the basics of data trimming, which involves removing punctuation marks, symbols, numbers and URLs. Additionally, it requires homogenization of features to lower as outlined below:

```r
press_conferences <- read.csv("president_discourse.csv")

# Corpus
corpus <- corpus(press_conferences, text_field = "text_amlo")
docnames(corpus) <- press_conferences$date

# Tokenisation
toks <- tokens(corpus,
               remove_punct = TRUE,
               remove_symbols = TRUE,
               remove_numbers = TRUE,
               remove_url = TRUE) %>%
        tokens_tolower()
```

Text stats

After creating the corpus and tokens object, we can compute basic statistics. We can start by using the `textstat_summary` function to calculate the number of characters, sentences, tokens, types, punctuation marks, and more. However, since these are stenographic versions of press conferences and not written directly by the president, some linguistic features may not be relevant for analysis. Therefore, in this case, I will only plot the length of the president's interventions during the press conferences.

```r
# Text stats
textstats <- textstat_summary(corpus)

# Plotting
{
  boxplot(textstats$chars,
          horizontal = TRUE,
          xlab = "Number of characters")
  }
```

Number of characters
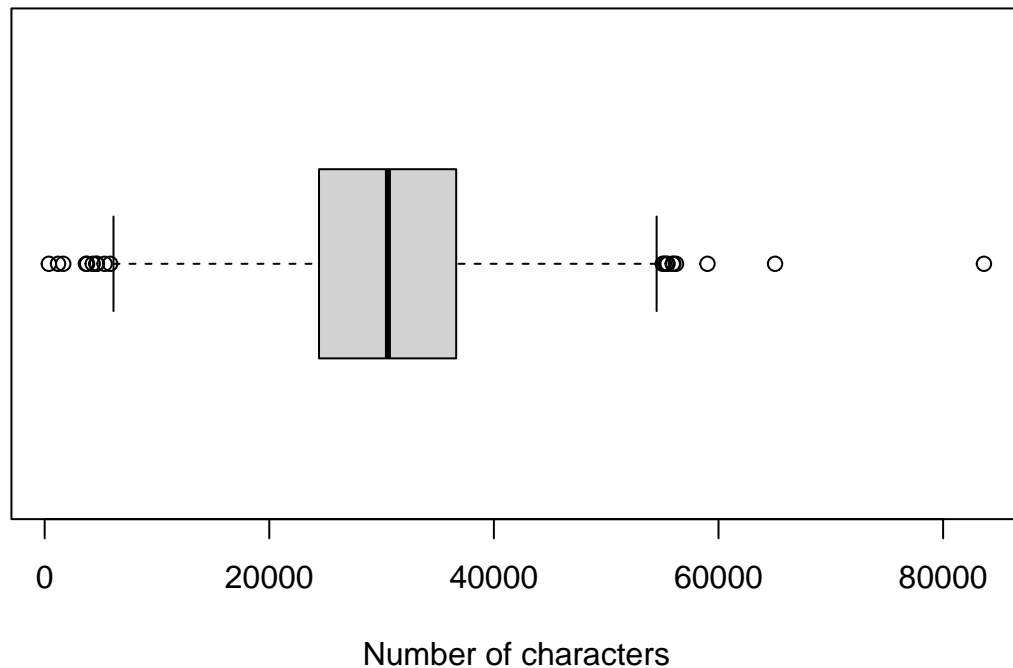
From the plot, we can observe that the median length of the interventions is around 30,000 characters. The interquartile range is also close to this value, ranging from 25,000 to 35,000. However, there is more variability in the upper and lower quartiles, with some outliers present.

Lexical diversity

To analyse lexical diversity, I will use the type to token ratio (TTR), which is calculated by dividing the number of unique words (types) by the total number of words (tokens). A higher TTR score indicates greater lexical diversity, with a larger proportion of unique words in the text. In contrast, a lower TTR score indicates lower lexical diversity, with a smaller proportion of unique words and a greater repetition of words in the text.

```r
# Lexical diversity
lexdiv_score <- textstat_lexdiv(toks, measure = "TTR")

# Plotting
{
  boxplot(lexdiv_score$TTR,
          horizontal = TRUE,
          xlab = "Type-to-token score")
  }
```

Type–to–token score
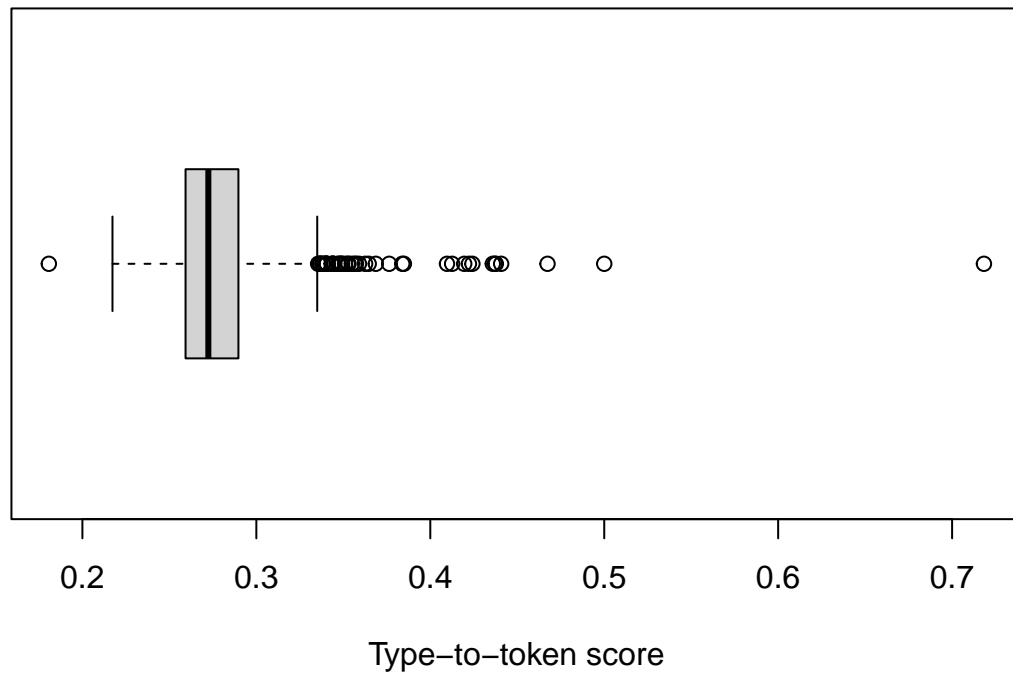
The plot reveals that the median TTR score, as well as the interquartile range, is below 0.3 and the upper quartile limit is below 3.5, indicating that the president's lexical diversity is incredibly low, and he tends to repeat the same words frequently.
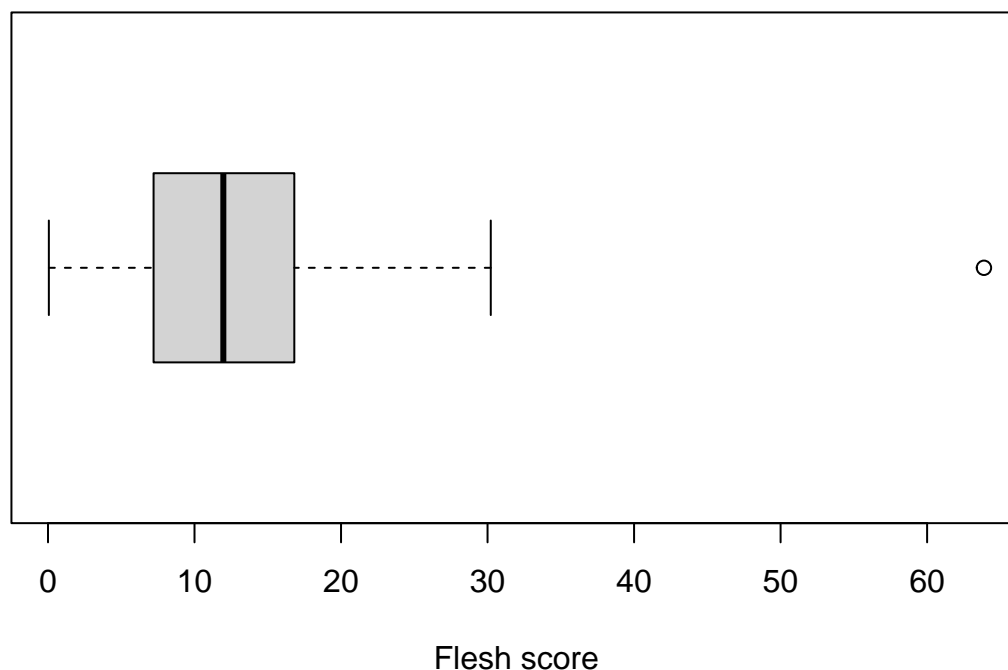
Readability score

To measure readability, I am going to use the Flesh Reading Ease index created by Rudolf Flesch. This score can range between 0 and 100, with a higher score indicating that the text is easier to read.

```r
# Readability score
read_score <- textstat_readability(corpus, measure = "Flesch")

# Dropping the negative scores
pos_flesh <- filter(read_score, read_score$Flesch > 0)

# Plotting
{
  boxplot(pos_flesh$Flesch,
          horizontal = TRUE,
          xlab = "Flesh score")
  }
```

Flesh score

With a median score just above 10, the readability score indicates that, contrary to expectations, the president's discourse is hard to follow for most people and is best understood by university graduates. Based on this measure, I can conclude that despite the fact that the president's speech is destined to the majority of Mexicans it is not diverse and is difficult to understand. This is in line with my experience as a spectator, as he tends to be redundant and it is easy to get lost.

Methods

In this project, I will be using two methods to answer the questions posed at the beginning. The first method will rely on lexicon-based techniques to detect populist rhetoric, while the second method will use topic modelling for propaganda detection. Below, I will provide a brief explanation of both methods.

Lexicon-based methods

Lexicon-based techniques are part of the semantic analysis that determine the sentiment orientation of a document or set of sentences based on the semantic orientation of the lexicons. Lexicons can be manually created or automatically generated. These techniques use adjectives and adverbs to discover the semantic orientation of text, and there are two main approaches: the dictionary-based approach and the corpus-based approach (Gupta and Agrawal, 2020). In this study, I will be using the dictionary-based approach, specifically a Spanish-adapted version of the dictionary created by Rooduijn and Pauwels in 2011.

Dictionaries are collections of words, phrases, parts of speech, or other word-based indicators that are used as the basis for search of texts. Researchers typically use multiple dictionaries in a single study, each measuring different concepts. There are two types of dictionaries: custom and internal. Custom dictionaries are created by the researcher based on theory, past research, and immersion in the message pool. Internal

dictionaries, on the other hand, are developed by the authors of the computer program being used and can range from simple readability indicators to complex dictionaries designed to measure unobservable constructs (Neuendorf, 2017).

In this project, I will use an adapted version of the dictionary created by Rooduijn and Pauwels (2011). This dictionary translates words into Spanish and includes synonyms for the words already in the dictionary. I will also include an extra word that represents the core of the definition of populism, "the people".

UK	MX
elit*	elite, fifí, mafia, neoliberal
consensus*	consenso, acuerdo, convenio
undemocratic*	antidemocrático, no democrático
referend*	referendo, consulta
corrupt*	corrupto
propagand*	propaganda
politici*	político
deceit	engañar, mentir, embaucar
deceiv	engañar, mentir, embaucar
betray	traicionar, fallar, desleal
shame*	vergüenza
scandal*	escándalo, polémica
truth*	verdad
dishonest*	deshonesto, inmoral
establishm*	organización, institución, organismo
ruling*	controlar, dominar, poder
people	pueblo

Given the above, the `populist_dict` can be created sing global patterns that include all variations of each of the words. Afterwards, I will use the function `dfm_lookup` to search for those words in the already created dfm.

Please note that the analysis will be performed on a modified version of the `toks` object. This version involves the removal of punctuation marks, symbols, numbers, URLs, homogenization of features (but not stemming) as well as stopword removal to exclude non-meaningful words. As a result, a new tokens object will be created, requiring the creation of a new dfm object. In this new dfm, the `min_termfreq` parameter will be set to 50 to remove rare features. Additionally, weights will be included in a new dfm to account for the diversity in the length of each of the president's speeches.

```
# Tokenising + stopword removal
toks_sw <- toks %>% tokens_remove(stopwords("spanish"))

# Document Feature Matrix
dfm <- dfm(toks_sw) %>%
        dfm_trim(min_termfreq = 50)

# Weights
dfm_w <- dfm %>%
        dfm_weight(scheme = "prop")

# Create the dictionary
populist_dict <- dictionary(list(
  populism = c(
        "elit*",
        "élit*",
```

```
      "fif*",
      "mafia*",
      "neoliberal*",
      "consenso",
      "acuerdo*",
      "convenio*",
      "antidemocrátic*",
      "no_democrátic*",
      "referend*",
      "consulta",
      "corrupt*",
      "propaganda",
      "polític*",
      "engaña*",
      "mentir*",
      "miente",
      "embauca*",
      "traici*",
      "fall*",
      "desleal*",
      "vergüenza",
      "escándalo",
      "polémica",
      "verdad",
      "deshonest*",
      "inmoral",
      "organizaci*",
      "instituci*",
      "organismo*",
      "control*",
      "domin*",
      "poder",
      "pueblo")))

# Look up the words in the dfm
populist_dfm <- dfm_lookup(dfm_w, populist_dict, valuetype = "glob")

# Save as data frame
populist_df <- convert(populist_dfm, to = "data.frame")
```

Topic modelling

Topic modelling techniques are statistical methods that can analyse the words used in texts to identify common topics, establish their relationships with each other, and track their evolution over time. These algorithms don't require any manual labelling or categorization of the documents, and their application can help to organize and summarize large digital archives in a way that would be impractical for humans to achieve. In this project, I will be using Latent Dirichlet Allocation (LDA), which is the most basic type of topic modelling. LDA assumes that the topics are created first, before the documents. Each document is generated by selecting a distribution of topics using the Dirichlet distribution, and then for each word, choose a topic assignment and choose the word for the corresponding topic (Blei, 2012).

After experimenting with different numbers of topics (`k`), I have decided to set `k=20`, and limit the number of Gibbs iterations to 500 in order to reduce computational costs. These parameters can be adjusted in the

function as follows:

```r
# LDA with k = 20
lda_prop <- LDA(dfm, k = 20, method = "Gibbs",
                control = list(seed = 123, iter = 500))

# Top ten words by topic
get_terms(lda_prop, k = 10)
```

```
##       Topic 1          Topic 2       Topic 3       Topic 4        Topic 5
##  [1,] "si"             "petróleo"    "va"          "jóvenes"      "vamos"
##  [2,] "poder"          "vamos"       "aeropuerto"  "educación"    "agua"
##  [3,] "caso"           "precio"      "tren"        "mujeres"      "va"
##  [4,] "entonces"       "mil"         "aquí"        "sólo"         "entonces"
##  [5,] "fiscalía"       "dos"         "mil"         "vamos"        "gobierno"
##  [6,] "judicial"       "producción" "vamos"       "país"         "sonora"
##  [7,] "asunto"         "millones"    "maya"        "maestros"     "ver"
##  [8,] "investigación"  "gasolinas"   "entonces"    "entonces"     "caso"
##  [9,] "justicia"       "precios"     "van"         "social"       "hacer"
## [10,] "presidente"     "ahora"       "dos"         "méxico"       "aquí"
##       Topic 6          Topic 7       Topic 8       Topic 9        Topic 10
##  [1,] "vamos"          "pueblo"      "empresas"    "méxico"       "pueblo"
##  [2,] "va"             "si"          "federal"     "presidente"   "vamos"
##  [3,] "información"    "democracia"  "comisión"    "pueblo"       "méxico"
##  [4,] "gobierno"       "reforma"     "energía"     "país"         "pandemia"
##  [5,] "voy"            "va"          "empresa"     "historia"     "gente"
##  [6,] "ver"            "pues"        "entonces"    "política"     "mil"
##  [7,] "si"             "gente"       "electricidad" "revolución"  "va"
##  [8,] "bien"           "partido"     "eléctrica"   "juárez"       "si"
##  [9,] "haber"          "van"         "méxico"      "ahora"        "crisis"
## [10,] "derecho"        "elecciones"  "si"          "madero"       "así"
##       Topic 11     Topic 12       Topic 13      Topic 14        Topic 15
##  [1,] "seguridad"  "medios"       "mil"         "salud"         "unidos"
##  [2,] "nacional"   "méxico"       "va"          "vamos"         "méxico"
##  [3,] "guardia"    "pues"         "programa"    "va"            "presidente"
##  [4,] "aquí"       "ahora"        "gente"       "medicamentos"  "vamos"
##  [5,] "vamos"      "información"  "ahora"       "médicos"       "va"
##  [6,] "secretaría" "cómo"         "apoyo"       "si"            "países"
##  [7,] "defensa"    "vamos"        "bienestar"   "ver"           "política"
##  [8,] "marina"     "pueblo"       "pesos"       "hospitales"    "relación"
##  [9,] "ver"        "gobierno"     "millones"    "atención"      "relaciones"
## [10,] "violencia"  "bueno"        "programas"   "sistema"       "gobierno"
##       Topic 16     Topic 17       Topic 18   Topic 19        Topic 20
##  [1,] "año"        "vamos"        "si"       "corrupción"    "millones"
##  [2,] "economía"   "vacunas"      "pues"     "vamos"         "mil"
##  [3,] "salario"    "pues"         "gente"    "va"            "presupuesto"
##  [4,] "si"         "entonces"     "ver"      "gobierno"      "va"
##  [5,] "crecimiento" "vacuna"      "ahí"      "gente"         "pesos"
##  [6,] "ciento"     "mayores"      "así"      "si"            "si"
##  [7,] "p"          "va"           "voy"      "información"   "gobierno"
##  [8,] "méxico"     "méxico"       "entonces" "van"           "dinero"
##  [9,] "ver"        "adultos"      "cómo"     "dar"           "impuestos"
## [10,] "inflación"  "si"           "bueno"    "plan"          "van"
```

Results

The methods explained above were implemented, and the subsequent analysis produced the following results:
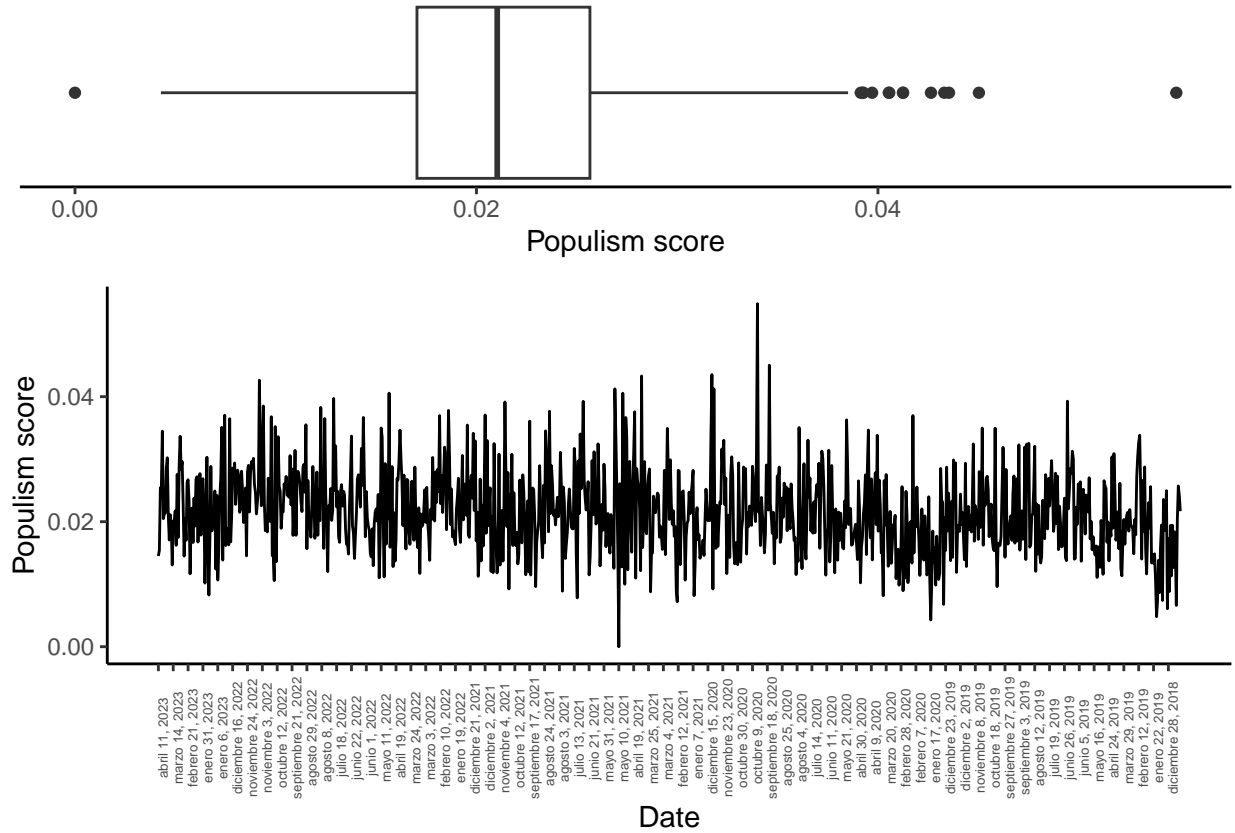
Populism

Once I searched for the populism dictionary in the weighted dfm, I can plot the results to visualize the behaviour of each speech over time, as follows:

```r
# Visualise only every 15 docs
x_seq <- seq(from = 1, to = length(populist_df$doc_id), by = 15)

# Plotting
pop_line <- ggplot(data = populist_df,
                   aes(x = 1:length(doc_id), y = populism)) +
            geom_line() +
            scale_x_continuous(breaks = x_seq,
                               labels = populist_df$doc_id[x_seq]) +
            labs(y = "Populism score",
                 x = "Date") +
            theme_classic() +
            theme(axis.text.x = element_text(angle = 90, size = 5))

pop_bp <- ggplot(populist_df, aes(x = populism)) +
          geom_boxplot() +
          xlab("Populism score") +
          theme_classic() +
          theme(axis.text.y = element_blank(),
                axis.ticks.y = element_blank(),
                axis.line.y = element_blank())

ggarrange(pop_bp, pop_line, ncol = 1, heights = c(1, 2))
```

The populism score represents the proportion of words in each document that match the populist dictionary. Based on the plot, it is evident that the president uses populism rhetoric in every speech, and the proportion remains relatively constant over time, hovering around 2%. For the majority of documents, the percentage of populist rhetoric is between 2% to 4% of the entire document, with some instances where the proportion reaches over 4% on specific dates.

Propaganda

After exploring the words in each of the topics, they can be categorized as follows:

Set	Topic
Topic 1	Prosecutor's office investigations
Topic 2	Hydrocarbon situation
Topic 3	Major infrastructure projects
Topic 4	Education Reforms
Topic 5	Water crisis in Sonora
Topic 6	Doing well
Topic 7	Democracy and elections
Topic 8	Electricity situation
Topic 9	Past transformations in Mexico
Topic 10	Covid-19 pandemic
Topic 11	Strengthening National Security
Topic 12	Transparency in reporting
Topic 13	Social programs
Topic 14	Shortages of medicines

Set	Topic
Topic 15	International Relations
Topic 16	Minimum wage rise
Topic 17	Covid-19 vaccination status
Topic 18	Greetings
Topic 19	Corruption
Topic 20	Government budget

Based on the table, it can be observed that the majority of the topics are in fact informative. However, some topics may have the potential to be used for propaganda purposes. These topics include 3, 6, 7, 9, and 13.

Topic 3 refers to major infrastructure projects that are the flagship initiatives of the current administration, such as the construction of a new airport in Mexico City, the Mayan train, and the refinery in Dos Bocas. Topic 6 includes words that suggest that the government is performing well in terms of the rule of law and welfare. Topic 7 relates to "democracy and elections" and may refer to the multiple attacks made by the president on the National Electoral Institute, criticizing its job. Topic 9 discusses previous revolutions in Mexico and includes well-known personalities who participated in them. This is related to the fact that the president refers to his administration as the fourth transformation of Mexico. Finally, topic 13 discusses social programs implemented by the government for those in need. It is possible that these topics could potentially contain propaganda content as they focus on what the president and his party are "doing well", which could be a way of attracting voters.

Conclusion

This project aimed to answer two questions posed at the beginning: whether the president uses his morning speeches to perpetuate populist rhetoric or as a means of propaganda. To answer the first question, dictionary-based methods were used to identify words related to the concept of populism. The analysis revealed that the president does indeed use populist rhetoric in his speeches, but only at a proportion of between 2% to 4% of the whole speech, which may not have a major impact in perpetuating the division between the two antagonist groups: the good people and the elite.

To answer the second part of the question, topic modelling was applied, and the results indicated that most of the president's speeches are dedicated to informing the public about the public agenda mostly in health, energy and government issues. However, some topics may be used as a means of propaganda, particularly as this is the first time that Morena has come to power. The president may be taking advantage of his position to perpetuate the hegemony of his party in the upcoming elections next year.

References

- Aragón Falomir, J., Fernández de Lara Gaitán, A. E. & Lucca, J. B. (2019). Las elecciones de 2018 en México y el triunfo del Movimiento de Regeneración Nacional (Morena). Estudios políticos (Medellín, Colombia). [Online] (54), 286–308. https://doi.org/10.17533/udea.espo.n54a14

- BBC. (2019). Las "mañaneras" de AMLO: cómo son las tempraneras conferencias con las que López Obrador marca la agenda política de México. Available at: https://www.bbc.com/mundo/noticias-america-latina-47066862 (Accessed: April 15, 2023)

- Chicago Tribune. (2019). La 'mañanera', ¿revolucionaria o populista?. Available at: https://www.chicagotribune.com/hoy/ct-hoy-mananera-revolucionaria-populista-amlo-20191122-6tybwhrn7vea7d3ashmtfitfli-story.html (Accessed: April 15, 2023)

- David M. B. (2012). Probabilistic topic models. Commun. ACM 55, 4, 77–84. https://doi.org/10.1145/2133806.2133826

- El Economista. (2022). Las "mañaneras" de AMLO llegan a 1,000 ediciones. Available at: https://www.eleconomista.com.mx/politica/Las-mananeras-de-AMLO-llegan-a-1000-ediciones-20221223-0030.html (Accessed: April 15, 2023)

- Expansión. (2022). Mañaneras: cuatro años de ataques, agenda y propaganda en el gobierno de AMLO. Available at: https://politica.expansion.mx/amlo-mananeras-cuantas-van-cuatro-anos (Accessed: April 15, 2023)

- Gupta N. & Agrawal R. (2020). Hybrid Computational Intelligence. 1-23. https://doi.org/10.1016/B978-0-12-818699-2.00001-9.

- Khurana, D., Koli, A., Khatter, K. & Singh S. (2022). Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl 82, 3713–3744 (2023). https://doi-org.gate3.library.lse.ac.uk/10.1007/s11042-022-13428-4

- Montero, G. (2019) ELECCIONES EN MÉXICO 2018. Iberoamericana (Madrid, Spain). 19 (70), 247–254. https://doi.org/10.18441/ibam.19.2019.70

- Neuendorf, K. (2017). The content analysis guidebook. SAGE Publications, Inc. https://doi.org/10.4135/9781071802878

- Publimetro. (2022). Mañaneras de AMLO son un ejercicio de propaganda, no informativo: Luis Estrada. Available at: https://www.publimetro.com.mx/nacional/2022/03/23/mananeras-de-amlo-son-un-ejercicio-de-propaganda-no-informativo-luis-estrada/ (Accessed: April 15, 2023)

- Romeu, V. (2022). La retórica del populismo en el discurso de "las mañaneras". Argumentos. Estudios Críticos De La Sociedad, (99), 73-98. https://doi.org/10.24275/uamxoc-dcsh/argumentos/202299-03

- Rooduijn, M. & Pauwels, T. (2011) Measuring Populism: Comparing Two Methods of Content Analysis, West European Politics, 34:6, 1272-1283. https://doi-org.gate3.library.lse.ac.uk/10.1080/01402382.2011.616665

- Spdnoticias (2023). Las mañaneras: ¿informativas o propaganda?. Available at: https://www.sdpnoticias.com/opinion/las-mananeras-informativas-o-propaganda/ (Accessed: April 15, 2023)

- Wirth, W., Esser, F., Wettstein, M., Engesser, S., Wirz, D., Schulz, A., Ernst, N., Büchel, F., Caramani, D., Manucci, L., Steenbergen, M., Bernhard, L., Weber, E., Hänggli, R., Dalmus, C., Schemer, C., Müller, P. (2016). The appeal of populist ideas, strategies, and styles: A theoretical model and research design for analyzing populist political communication. NCCR democracy Working Paper series 88, University of Zurich. https://doi.org/10.5167/uzh-127461

- Yakunin K., Ionescu G. M., Murzakhmetov S., Mussabayev R., Filatova O., Mukhamediev R. (2020).Propaganda Identification Using Topic Modelling. Procedia Computer Science, Volume 178, 205-212. https://doi.org/10.1016/j.procs.2020.11.022.