

Macron & Le Pen Left-Right Analysis

Léa Duplan & Akira Sasaki

16/05/2017

I. Context and research question

On the 7th of May 2017, France elected a new president: Emmanuel Macron. As a semi-presidential regime, French citizens directly elect their head of state in two election rounds. On the 23rd of April 2017, Emmanuel Macron and Marine Le Pen won the first election round with respectively 24.01% and 21.30% of the vote (FRANCE 24, 2017). After the first round, the qualified candidates had fifteen days to convince and gather on their side a majority of voters.

While Macron supports a liberal economy but leftist social values, Le Pen defends the opposite as a far-right representative. She inherited from her father's political party, the Front National, which traditionally supports populist, anti-European and anti-immigration measures. Macron, on the other hand, was relatively unknown before joining Francois Hollande's socialist government as the economy minister in 2014. After leaving the government in 2016 to create his own political movement En Marche!, Macron built his presidential program on a mix of right-wing, left-wing and pro-European ideas.

Two other candidates, Francois Fillon (moderate right-wing party - Les Républicains) and Jean-Luc Mélenchon (far-left party – Les Insoumis) gathered respectively 20.01% and 19.58% of the vote (1st election round). These results disqualified them for the second round but were nonetheless significant. Thus, as every vote counts, Macron and Le Pen had the hard task to convince Fillon's and Mélenchon's voters to win the election. To persuade voters, Macron and Le Pen might have softened their speeches. Le Pen for instance has been blamed for softening her anti-European position to seduce Fillon's voters (Bonnetoy, 2017). Macron promised to renew the French political landscape which appeals to Mélenchon's voters. In this context, it is interesting to research the following question:

Did Macron and Le Pen change their left-right position after the first election round on the 23rd of April?

It is expected that Le Pen shifts her political discourse after the first election round more to the left to attract more voters. Expectations concerning Macron's shift, on the other hand, are less straight-forward as he never took a clear stand on a right-wing or a left-wing position. Nonetheless, changes between before and after the first round are expected.

II. Data collection & preparation of dataset for analysis

To answer our research question, an analysis of Macron and Le Pen speeches before and after the first election round (23rd of April) is conducted. A specific dataset containing 13 Macron and 12 Le Pen speeches of before and after the first election round was built. While Macron's speeches range from 67 sentences to 399 sentences, Le Pen's speeches range from 13 sentences to 477 sentences.

Macron's speeches were extracted directly from the candidate's campaign website (En Marche!, n.d.). Out of 43 speeches available onsite since the 3rd of April 2016, our dataset was created from 13 speeches based on certain criteria. These 13 speeches were given between the 16th of March 2017 and the 5th of May 2017. The French election campaign officially started on the 9th of April 2017 but candidates had until the 16th of March to apply. Initially, 20 speeches were gathered for Macron but to run our model according to a left-right dimension, only 'general' topic speeches were kept. We collected 7 speeches before the 23rd of April and 5 speeches after.

Le Pen's speeches were extracted directly from her campaign website as well, where more than 130 speeches are stored (Front National, n.d.). The large availability of speeches for Le Pen compared to Macron can be explained by the seniority of her party over Macron's one. Like Macron, the dataset for Le Pen was built from only the 'general' topic speeches between the 16th of March 2017 and the 5th of May 2017. As the result, 11 speeches are before the 23rd and 1 speech is after the 23rd of April. This is because Le Pen focused more on topical speeches after the first round rather than 'general' topic speeches. All speeches from Macron and Le Pen were extracted manually and stored in txt format documents. After uploading the dataset on R, two different text corpuses were generated: Macron's corpus consisting of 13 documents and Le Pen's corpus consisting of 12 documents.

III. Wordfish model specification

In this paper, we decided to use the Wordfish technique, to analyze the changes in political discourses of both candidates, for two reasons. Firstly, Wordfish is an ideological scaling tool which allows for the identification of changing political positions by word frequencies. Thus, if Macron and Le Pen changed specific word frequencies before and after the first round, this can be identified by Wordfish. Secondly, our research requires to take into account time order hence we can analyze their position changes according to it. Wordfish is a very suitable technique with such data due to its strength in producing time series estimates (Caron, 2013). In particular, since our objective is to confirm the significant changes on the respective candidates' positions between before and after the first election round, its ability to enable a comparison among texts from different times is an advantage.

The Wordfish model is mathematically expressed as the following equation:

$$\lambda_{ij} = \exp(\alpha_i + \psi_j + \beta_j * \omega_i)$$

If we apply this model to our research, elements of the above-mentioned equation refer respectively to 'i' as Le Pen and Macron's speeches, 'j' as unique word, ' α_i ' as speech fixed effect, ' ψ_j ' as word fixed effect, ' ω_j ' as word specific weight (where the sign represents the ideological direction) and ' ω ' as the speech position.

Three steps were necessary to answer our research question:

1. Constructing a document-frequency matrix (dfm)

A document-frequency matrix is necessary to capture the frequencies of each word. When applying pre-processing commands of the TM package to the dfm, several punctuation symbols were not recognized by R. To counter this issue, we deleted specific insignificant words or symbols by specifying them in codes. For instance, ".", "/", "a" were not recognized by the remove punctuation and stopwords commands.

2. Viewing the document-frequency matrix

We applied several commands such as topfeatures and textplot_wordcloud and to identify most frequent words and therefore generate descriptive statistics.

3. Scaling document positions

Finally, the Wordfish command was applied to both Le Pen and Macron's dfm. The commands textmodel_wordfish and textplot_scale1d were used to scale word counts and retrieve a left-right dimension. When applying the Wordfish model to our dfm, we choose 2 speeches that are expected to be very different in word usage. As for Macron, the first speech we took is the first speech before 23rd of April and the second speech is the last speech before the second election round based on the assumption that he was more politically inclined to the left given his background in the Socialist party. In contrast, although we also took the first and last speeches for Le Pen as well, assuming she shifted from far-right to more central as time passed by, we framed the Wordfish in the opposite way for Le Pen. This speech identification process is expected to create an adequate scale to analyze their political shifts.

Word frequency

Figure. 1 Macron's Word Cloud



Figure. 2 Le Pen's Word Cloud

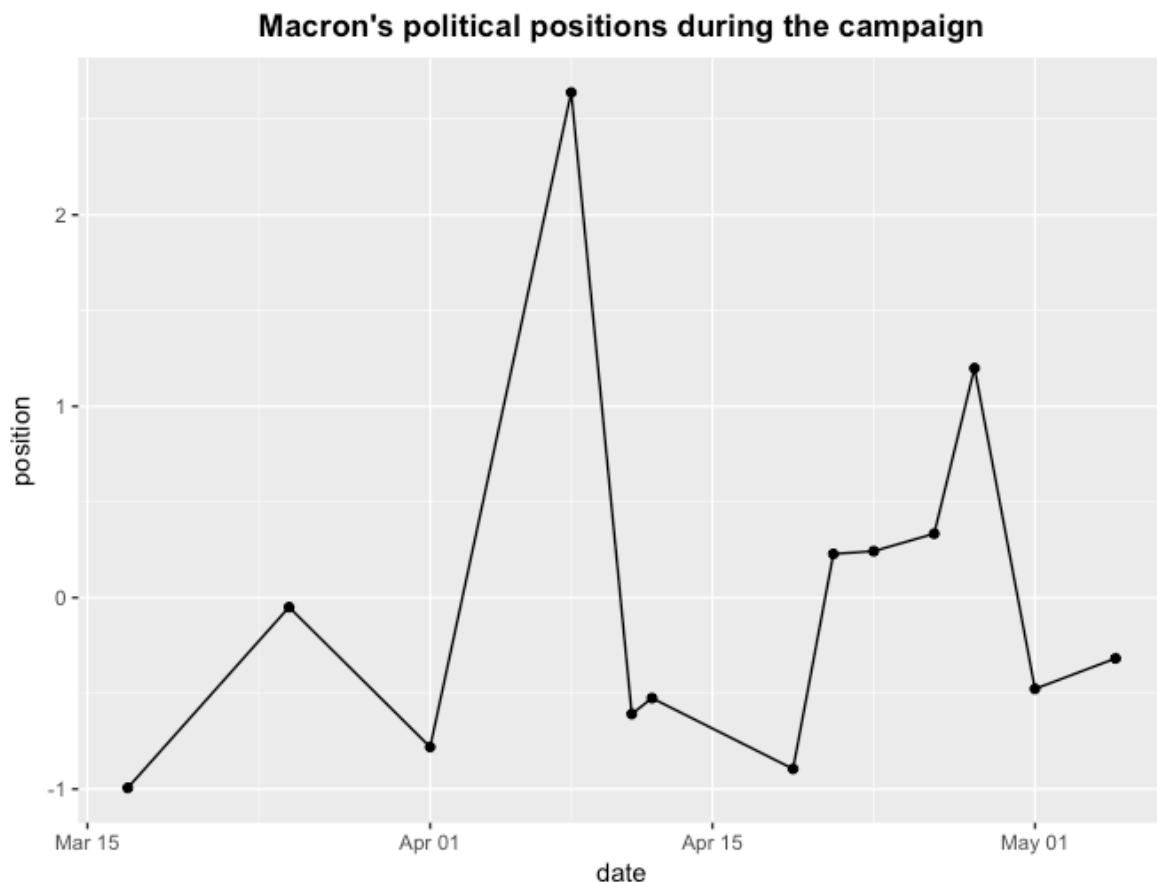


Firstly, we simply compare respective candidates' word clouds in order to intuitively understand the tendency of their picks of words. Both word clouds display the words used more than 50 times in their overall speeches respectively. The sizes correlate with the frequencies of the words in our dataset and the colors help to see the groups of the similar frequencies. The first one is to Macron and at glance it is easy to see "parc" is the most frequently used word by him. We assume this derives from the french word of "parce que", which means "why" in English. Due to our technical limitation, we could not remove or further clarify the word but at the same time this word should signify his tendency of explaining his ideas and policies in speeches after presenting his political agenda. Apart from the word, another notable point is his frequency of using economic terms such as "pay", "investiss", "(é)conomie", or "(é)conomique". This should indicate his strong focus on reforming economic policies in his speeches. The other word cloud is generated from Le Pen's speeches. The interesting finding is that its more concentration on specific words. The words such as "droit" (right), "loi" (law), "nation" and "national" seem to link with her nationalist and conservative perspective. Her opponents' names are also visible here, meaning she often referred to them in her speeches to promote her campaign. Word cloud gives an interesting insight about the candidates' habits and preferences of words but there is no right or left scale to further analyze. In the next section, we delve into a more intricate study on political positions of each candidate and their changes over the time. Also, their overall political stances during the campaign are assessed as well with "Eiffel Tower of Words".

Left-Right positions over time

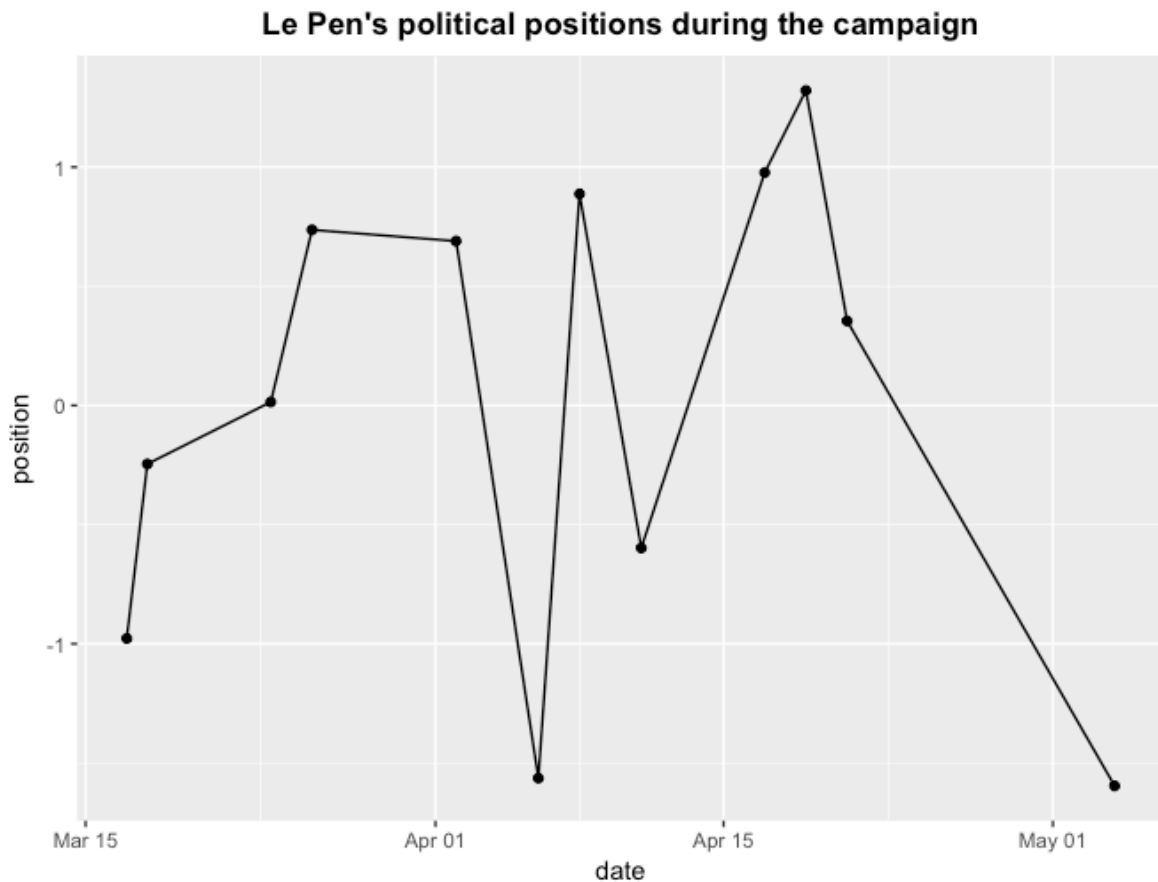
In order to capture the changes on candidates' political positions, we ordered speeches in the time order and assigned relative left-right positions to each. The below charts are the result. The X-axis indicates date and the Y-axis stands for the left-right position in comparison with the other captioned speeches. The higher the point is the more right it assumes and the opposite holds for the left.

Plot. 1



The first chart describes that Macron political positions between the 16th of March 2017 and the 5th of May 2017. As expected, his tone was inclined to the left at the beginning but, to our surprise, it drastically shifted to the right in the middle. Then, it shifted back almost to the original point but around the time of the first election round the stance shifted to the central. After the election, there are some changes but those are not significant and stayed around the neutral point. This result indicates that Macron was relatively inclined to the left at the inception but modified his position to the central as time passed by.

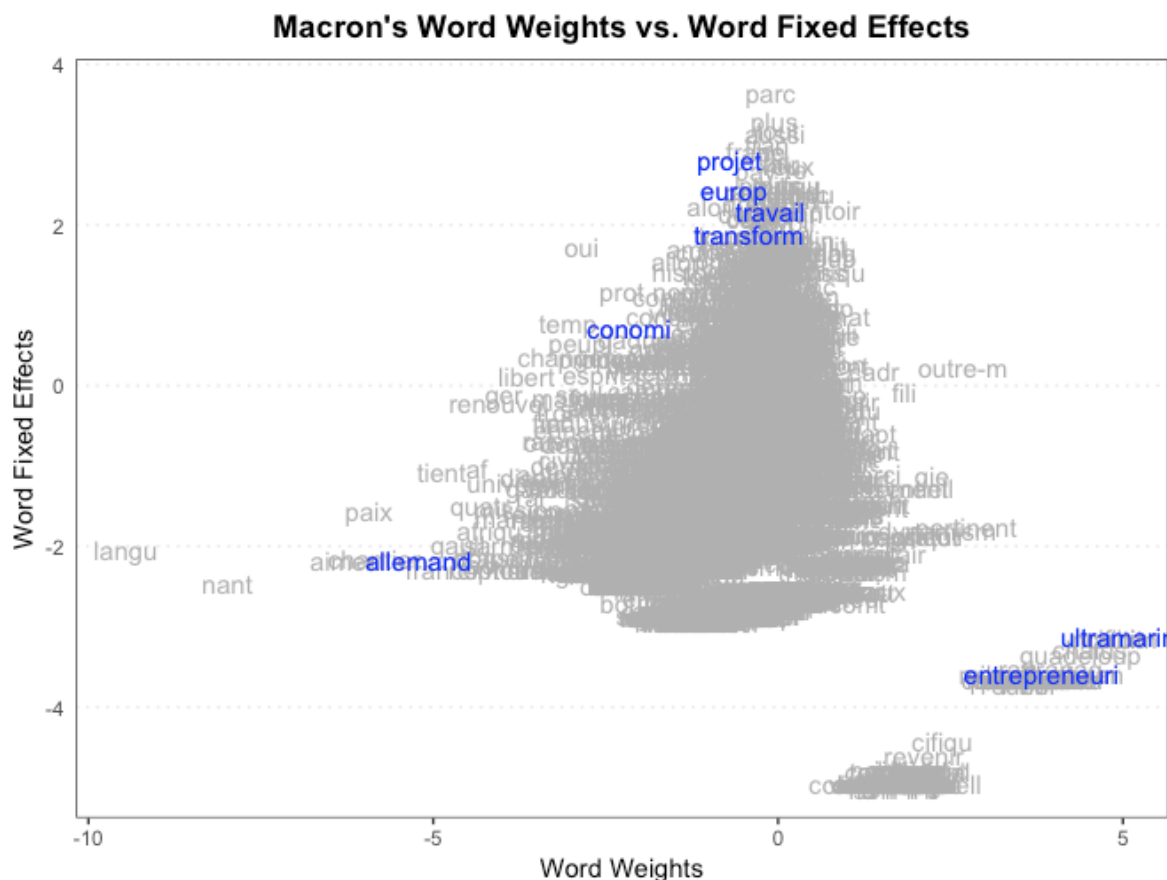
Plot. 2



The chart shows a very interesting trend of Le Pen's position throughout her campaign. Although she was generally regarded as a far-right candidate in the election, her tone seems to be relatively milder at the beginning. Gradually she amplified her right-winged ideas and thoughts but suddenly again toned down her position in the middle. After then, her positions fluctuated but rose up to the most right point right around the first election round. Then, she again diffused her right tendency and approached to the left. This can be interpreted that her strategic shift on gathering the other political spectrum's votes after the election.

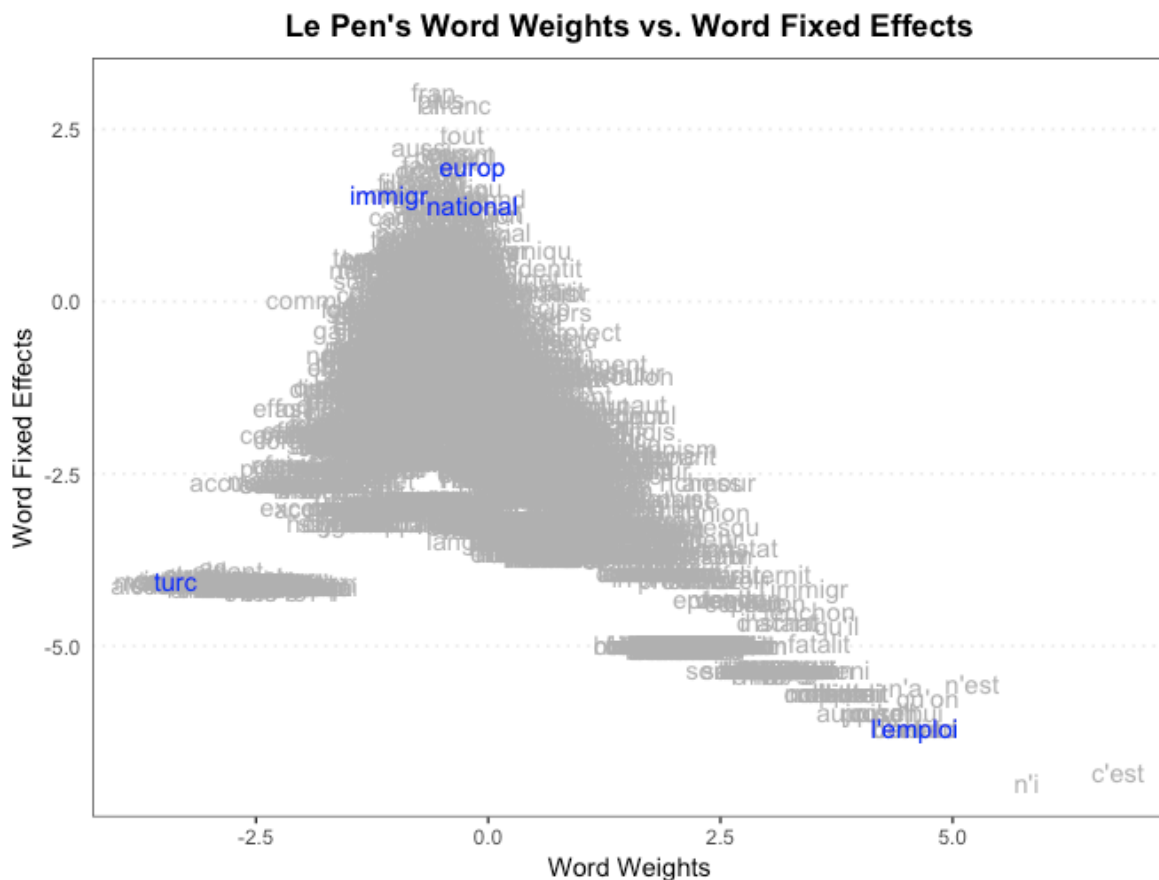
Word Weights vs. Word Fixed Effects

Figure. 3



The next chart is so called "Eiffel Tower of words", which illustrates the estimated word fixed effects in the X-axis against the word weights in the Y-axis. In our model, the X-axis signifies the left-right tendencies of words and the Y-axis is the frequency of the words in speeches. According to the graph, it seems that Macron tended to speak in favour for the left but occasionally showed his right side of views in his speeches. With the words which do not discriminating value, such as "parc", "plus" and "aussi", coming to the top of the tower, we consider our data are providing a reliable result. We highlighted several words based on word cloud, word weights and word fixed effects. As expected, the frequently used words assume neutral characteristics as are illustrated by words such as "europ", "transform", "travail (work)". The interesting thing is "allemand (Germany)" is more referred to in the left context, which might indicate his stance on Europe. In contrast, "ultramarin (overseas)" is used in line with the right. Considering it is necessary to promote a nationalistic sentiment in the overseas territory to unite France, it seems like a sensible outcome. Also, the chart tells that Macron tries to convey a very strong message when he uses the term "entrepreneur".

Figure. 4



Le Pen's Eiffel Tower of words are exactly contrast to that of Macron. The shape is more stretching to right, which matches with the general perception on her political stance. The same with that of Macron, the chart clearly shows the most frequented used words at the top of the tower, such as "europ", "immigr" and "national", hence the data used for Le Pen should generate reliable results. An interesting finding from this chart of Le Pen is that she tends refer to "turc (Turkish)" in the left sense. As is expected, she uses "l'emploi (employment)" as the key word and to express her right-winged perspectives.

V) Conclusion

Our analysis shows that both candidates shifted to a certain degree their left-right position after the first election round. However, results are not straight-forward. In Macron's case, we observe high left-right position volatility before the 23rd of April and rather low volatility after. Interpretation of this result relies in the fact that Macron might have wanted to be more neutral in his post-23rd of April speeches. In Le Pen's case, the result is more significant as her first and last speech after April 23rd shows a clear shift to the left. As

expected, Le Pen softened her discourse to attract more voters. To be fully accurate, some limitations have to be mentioned. Firstly, our text corpus would necessitate more and better cleaning to extract only the meaningful words and have a more efficient result. Secondly, more speeches of Le Pen after the first election round would have been preferable to strengthen our claim. Finally, other analyses should be performed to explore other dimensions than the left-right one. For instance, topic modelling for the European, or security dimensions would be interesting.

Bibliography

- Bonnefoy, G. (2017). VIDÉO. Le Pen plagie un discours de Fillon, son équipe évoque un “clin d’oeil.” Retrieved May 15, 2017, from http://www.lexpress.fr/actualite/politique/elections/presidentielle-le-pen-emprunte-de-larges-parts-d-un-discours-de-fillon_1904186.html
- Caron, A. (2013). From Words to Facts: Wordfish, a Modern Technique to Estimate Policy Positions of Political Actors. *Italian Political Science*. Retrieved from <https://italianpoliticalscience.files.wordpress.com/2013/07/27.pdf>
- En March! (n.d.). Discours | En Marche ! Retrieved May 15, 2017, from <https://en-marche.fr/articles/discours>
- FRANCE 24. (2017). Présidentielle : Macron à 24,01 %, Le Pen à 21,30 %, selon les résultats définitifs du 1er tour - France 24. Retrieved May 14, 2017, from <http://www.france24.com/fr/20170424-france-presidentielle-2017-resultats-definitifs-1er-tour-emmanuel-macron-marine-le-pen>
- Front National. (n.d.). Discours | FN – Front National. Retrieved May 16, 2017, from <http://www.frontnational.com/categorie/discours/>

Appendices

```
library(readtext)
library(quantda)
library(readxl)
library(tm)
library(austin)
library(ggplot2)

setwd("/Users/Teruteru/Dropbox/Quantitative Content Analysis/MACRON")
wdir <- getwd()
myCorpusMA <- VCorpus(DirSource(wdir, pattern = "\\*.txt", recursive = T))
corpusMA <- corpus(myCorpusMA, readerControl = list(reader = readPlain,
  language = "fr", load = FALSE))
text.corpus.formatMA <- corpusMA
docnames(text.corpus.formatMA) <- paste(c("M1", "M4", "M5", "M6",
  "M7", "M8", "M9", "M10", "M11", "M12", "M13", "M2", "M3"))
speechDfmMA <- dfm(text.corpus.formatMA, tolower = T, remove = c(stopword
s("french"),
  ".", "/", "c3", "a9", "e2", "a0", "a8", "r", "a", "aa", "a7",
  "b4", "9d", "9c", "b9", "pr", "0", "e", "a2", "g", ">", "a2",
  "v", "o", "parc"), stem = TRUE, remove_punct = TRUE, remove_numbers =
TRUE)
topfeatures(speechDfmMA, 20)

set.seed(100)
textplot_wordcloud(speechDfmMA, min.freq = 50, random.order = FALSE,
  rot.per = 0.25, colors = RColorBrewer::brewer.pal(8, "Dark2"))

setwd("/Users/Teruteru/Dropbox/Quantitative Content Analysis/LEPEN")
wdir <- getwd()
myCorpusLP <- VCorpus(DirSource(wdir, pattern = "\\*.txt", recursive = T))
corpusLP <- corpus(myCorpusLP, readerControl = list(reader = readPlain,
  language = "fr", load = FALSE))
text.corpus.format.LP <- corpusLP
docnames(text.corpus.format.LP) <- paste(c("LP1", "LP7", "LP8", "LP9",
  "LP10", "LP11", "LP12", "LP2", "LP3", "LP4", "LP5", "LP6"))
speechDfmLP <- dfm(text.corpus.format.LP, tolower = T, remove = c(stopwor
ds("french"),
  ".", "/", "a", "c3", "a9", "e2", "a0", "a8", "r", "aa", "a7",
  "e", "c2", "pr", "bb", "b4", "re", "tre", "a2", "ab", "c5", "g",
  "f", ">", "o", "a6", "b9", "v", "p"), stem = TRUE, remove_punct = TRU
E,
  remove_numbers = TRUE)
```

```

topfeatures(speechDfmLP, 20)

set.seed(100)
textplot_wordcloud(speechDfmLP, min.freq = 50, random.order = FALSE,
  rot.per = 0.25, colors = RColorBrewer::brewer.pal(8, "Dark2"))

speechMA <- textmodel_wordfish(speechDfmMA, dir = c(1, 13))
no. <- as.numeric(c(1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 2, 3))
Macrondf <- data.frame(id = speechMA@docs, position = speechMA@theta,
  no. = as.numeric(no.))
Macrondf <- Macrondf[order(Macrondf$no.), ]
date <- c("17/03/2017", "25/03/2017", "01/04/2017", "08/04/2017",
  "11/04/2017", "12/04/2017", "19/04/2017", "21/04/2017", "23/04/2017",
  "26/04/2017", "28/04/2017", "01/05/2017", "05/05/2017")
Macrondf$date <- as.Date(date, "%d/%m/%Y")

ggplot(Macrondf, aes(x = date, y = position)) + geom_point() + geom_line
() +
  ggtitle("Macron's political positions during the campaign") +
  theme(plot.title = element_text(lineheight = 0.8, face = "bold",
    hjust = 0.5))

speechLP <- textmodel_wordfish(speechDfmLP, dir = c(12, 1))
no. <- as.numeric(c(1, 7, 8, 9, 10, 11, 12, 2, 3, 4, 5, 6))
LePendf <- data.frame(id = speechLP@docs, position = speechLP@theta,
  no. = as.numeric(no.))
LePendf <- LePendf[order(LePendf$no.), ]
date <- c("17/03/2017", "18/03/2017", "24/03/2017", "26/03/2017",
  "02/04/2017", "06/04/2017", "08/04/2017", "11/04/2017", "17/04/2017",
  "19/04/2017", "21/04/2017", "04/05/2017")
LePendf$date <- as.Date(date, "%d/%m/%Y")

ggplot(LePendf, aes(x = date, y = position)) + geom_point() + geom_line()
+
  ggtitle("Le Pen's political positions during the campaign") +
  theme(plot.title = element_text(lineheight = 0.8, face = "bold",
    hjust = 0.5))

# Least and most discriminating words<U+3000>
tail(speechMA@features[order(speechMA@psi)], 20)
head(speechMA@features[order(speechMA@psi)], 20)
# and some of the highest and lowest scoring words
head(speechMA@features[order(speechMA@beta)], 20)
tail(speechMA@features[order(speechMA@beta)], 20)

```

```

textplot_scale1d(speechMA, margin = "features", highlighted = c("europ",
  "conomi", "transform", "travail", "projet", "entrepreneuri", "ultrama
rin",
  "allemand"), highlighted_color = "blue") + ggtitle("Macron's Word Wei
ghts vs. Word Fixed Effects") +
  theme(plot.title = element_text(lineheight = 0.8, face = "bold",
    hjust = 0.5)) + labs(x = "Word Weights", y = "Word Fixed Effects
")

# Least and most discriminating words<U+3000>
tail(speechLP@features[order(speechLP@psi)], 20)
head(speechLP@features[order(speechLP@psi)], 20)
# and some of the highest and lowest scoring words
head(speechLP@features[order(speechLP@beta)], 20)
tail(speechLP@features[order(speechLP@beta)], 20)

textplot_scale1d(speechLP, margin = "features", highlighted = c("national
",
  "immigr", "europ", "l'emploi", "turc"), highlighted_color = "blue") +

  ggtitle("Le Pen's Word Weights vs. Word Fixed Effects") + theme(plot.
title = element_text(lineheight = 0.8,
  face = "bold", hjust = 0.5)) + labs(x = "Word Weights", y = "Word Fix
ed Effects")

```