

TP3_Bouzamondo

Mohamed Chadli Bouzamondo

2024-19-02

Introduction

La crise financière qui s'est propagée à travers le monde, en 2008, a été catastrophique. Elle a touché plusieurs secteurs économiques tout en créant une situation impossible à résoudre pour les gouvernements du monde. En conséquence, l'enjeu économique est devenu la préoccupation principale, pour les mois à venir. Le Canada et son gouvernement n'ont pas été épargné par cette crise et ont dû faire face aux nombreuses zones d'ombres apportée par celle-ci. La question se pose : Quelle parti politique canadien a-t-il le plus discuté de l'enjeu économique durant ses sessions parlementaires? Au cours de l'année 2008, les sessions parlementaires du gouvernement fédéral ont souvent dû couvrir de l'enjeu économique. Afin de répondre à la question de recherche, il faut utiliser des données textuelles permettant de discerner, le ou les partis ayant le plus discuté de l'enjeu économique. Le meilleur moyen est d'aller chercher dans les sessions parlementaires, durant la période spécifique pendant laquelle cette crise économique est au centre des préoccupations. La base de données du LIPAD qui offre des transcriptions du parlement canadien était l'option idéale. Néanmoins, il a fallu se concentrer sur un moment précis puisque la base de données est très précise et couvre les sessions parlementaires sur plusieurs jours s'étalant sur plusieurs mois.

Données utilisées

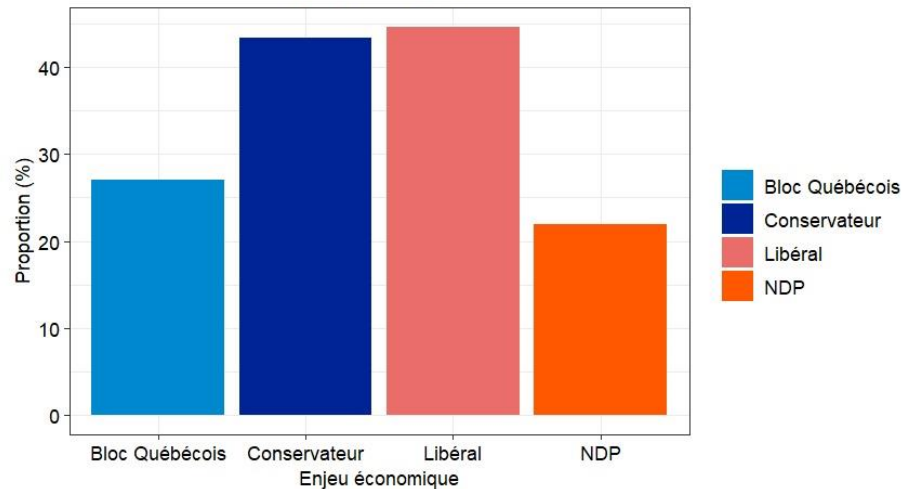
Au lieu de couvrir l'année 2008 dans son entièreté, j'ai préféré me concentrer sur le dernier mois de l'année. Avec les données du LIPAD, j'ai pu récupérer 4 bases de données contenant des données textuelles concernant les débats parlementaires entre partis politiques canadiens durant le mois de décembre 2008. Les 4 bases de données s'étalent du 1er au 4 décembre. Après avoir combiné les trois bases de données, chacune nommée "Fin_2008_1", "Fin_2008_2", "Fin_2008_3", "Fin_2008_4" ; J'ai dû les nettoyer pour récupérer les données et les variables dont j'avais besoin pour répondre à ma question. Je devais conserver les données textuelles nommées "speechtext" qui représentaient le discours de chaque parti lors de la session parlementaire. Ensuite, la variable "speakerparty" a également été conservée car j'avais besoin des différents partis (Bloc Québécois, Conservateur, Libéral, NPD). Dans le cadre du nettoyage, les données NA ont été exclues du nouvel ensemble de données appelé « Fin_2008_CLEAN ». Pour pouvoir faire l'analyse textuelle j'ai dû installer un dictionnaire anglais : le lexicoder_en (car les discours parlementaires étaient en anglais pour ce cas-là). Pour utiliser le dictionnaire de manière simplifiée, j'ai dû télécharger le Clessnverse qui aiderait à exécuter le dictionnaire (comme il n'était pas possible d'accéder à cette fonction via la commande `install.packages`, il a fallu l'installer manuellement via le code correspondant pour qu'elle puisse fonctionner correctement). Après avoir exécuté la commande `run_dictionary` et sélectionné spécifiquement la variable "speechtext", j'ai dû faire un `pivot_longer` pour obtenir combien chaque parti a parlé pendant la session parlementaire en proportion (en %). Pour le sujet de la financière de 2008, j'ai uniquement conservé le thème de l'économie lors du filtrage du texte à conserver. Enfin, le tout conduisant à la création d'un graphique à bandes avec les fonctions `ggplot` et `geom_bar`.

Description des résultats

Le graphique à barres créé couvre la variable "speakerparty" qui est séparée en quatre catégories, chacune avec sa propre couleur et la proportion (en %) montrant dans quelle mesure chaque parti a discuté de quelque chose lié à l'économie (selon les termes économie – macroéconomie, contenu dans le lexicoder en anglais). Force est de constater qu'en ce dernier mois de 2008, l'enjeu économique est toujours au centre des discussions. Tous les partis politiques dépassent les 20 % lorsqu'il s'agit de l'enjeu économique. Le parti le plus bas étant le NPD (un peu plus de 20 %) et le parti le plus élevé étant les Libéraux (très près de 45 %). Ce qui est intéressant à noter ici, c'est que les Conservateurs et les Libéraux sont presque égaux en termes de proportion, les premiers n'ayant que quelques petits pourcentages d'avance sur les seconds. Ce qui pourrait être surprenant puisque ce sont les Conservateurs qui étaient au pouvoir à cette période de l'histoire du Canada. Cela pourrait être interprété de différentes manières. Il se pourrait que le parti conservateur ait eu d'autres priorités à l'époque, puisque j'étudie les données de la fin de l'année en décembre 2008, alors que la crise existait peut-être depuis un bon moment et que d'autres problèmes étaient en jeu. Les Libéraux, en tant qu'opposition, pourraient également jouer un rôle dans leur proportion légèrement plus élevée lorsqu'il s'agit de mentionner l'économie, puisque leur rôle principal est de critiquer le gouvernement en place. Il y avait probablement beaucoup de choses à dire et à critiquer à l'égard du gouvernement conservateur de l'époque, surtout en ce qui concerne l'économie et la manière dont il a géré la crise financière. Le facteur idéologique pourrait également jouer un rôle, même si j'en doute personnellement puisque la crise de 2008 était un phénomène plus général et quel que soit le type de gouvernement ou d'idéologie, il fallait en parler et en traiter. Quant aux résultats du NPD et du Bloc québécois bien inférieurs à ceux des deux partis mentionnés précédemment, cela pourrait être dû au fait qu'ils étaient plus concentrés sur d'autres sujets que l'économie. Il est également possible que les priorités en

matière de crise financière se trouvent ailleurs, comme la baisse du taux d'emploi ou la montée de la criminalité.

Discussion de l'enjeu économique par les partis politiques canadiens, en Décembre 2008



Conclusion

Les libéraux sont ceux qui parlent le plus de l'enjeu économique (près de 45%), suivis de près par les conservateurs (environ 42%), l'avant-dernier étant le Bloc québécois (un peu plus de 25%) et les derniers étant le NPD (un peu plus de 20%). Ces résultats de décembre 2008 montrent encore que la plupart des partis politiques canadiens avaient un intérêt direct à parler de l'enjeu économique, tous ayant plus de 20% lorsqu'il s'agit de leurs propres proportions parlant des termes économiques. Il est clair qu'à la fin de l'année, la crise financière de 2008 avait encore des répercussions sur les sujets débattus lors des séances parlementaires. Cependant, une étude plus large de l'année entière aurait pu aider à saisir le véritable poids de la crise et une comparaison avec d'autres sujets comme la criminalité ou le taux d'emploi pourrait aider à établir une vision plus large de l'importance de l'enjeu économique.

Bibliographie

LIPAD. « Lipad - Data - Browse the Canadian House of Commons ». *LIPAD*, <https://www.lipad.ca/data/>.

Albugh, Quinn, Julie Sevenans and Stuart Soroka. 2013. « Lexicoder Topic Dictionaries ». *McGill University*, <https://www.snsoroka.com/data-lexicoder>.

Annexe

```
install.packages("crayon", repos = "http://cran.us.r-project.org")
```

le package 'crayon' a été décompressé et les sommes MD5 ont été vérifiées avec succès

Les packages binaires téléchargés sont dans
C:\Users\15150135\AppData\Local\Temp\RtmpwTp5db\downloaded_packages

```
library(crayon)
```

Warning: le package 'crayon' a été compilé avec la version R 4.3.2

```
install.packages("quanteda", repos = "http://cran.us.r-project.org")
```

le package 'quanteda' a été décompressé et les sommes MD5 ont été vérifiées avec succès

Warning: impossible de supprimer l'installation précédente du package 'quanteda'

Warning in file.copy(savedcopy, lib, recursive = TRUE): problème lors de la copie de

C:\Users\15150135\AppData\Local\Programs\R\R-4.3.1\library\00LOCK\quanteda\libs\x64\quanteda.dll
vers

C:\Users\15150135\AppData\Local\Programs\R\R-4.3.1\library\quanteda\libs\x64\quanteda.dll:
Permission denied

Warning: 'quanteda' restauré

Les packages binaires téléchargés sont dans
C:\Users\15150135\AppData\Local\Temp\RtmpwTp5db\downloaded_packages

```
library(quanteda)
```

Warning: le package 'quanteda' a été compilé avec la version R 4.3.2

Package version: 3.3.1
Unicode version: 15.1
ICU version: 74.1

Parallel computing: 4 of 4 threads used.

See <https://quanteda.io> for tutorials and examples.

```
library(SnowballC)  
library(tidyverse)
```

Warning: le package 'tidyverse' a été compilé avec la version R 4.3.2

```

Warning: le package 'ggplot2' a été compilé avec la version R 4.3.2
Warning: le package 'tidyr' a été compilé avec la version R 4.3.2
Warning: le package 'readr' a été compilé avec la version R 4.3.2
Warning: le package 'purrr' a été compilé avec la version R 4.3.2
Warning: le package 'dplyr' a été compilé avec la version R 4.3.2
Warning: le package 'stringr' a été compilé avec la version R 4.3.2
Warning: le package 'forcats' a été compilé avec la version R 4.3.2
Warning: le package 'lubridate' a été compilé avec la version R 4.3.2

— Attaching core tidyverse packages — tidyverse 2.0.0
—
✓ dplyr      1.1.4    ✓ readr      2.1.5
✓ forcats    1.0.0    ✓ stringr    1.5.1
✓ ggplot2    3.4.4    ✓ tibble     3.2.1
✓ lubridate  1.9.3    ✓ tidyr      1.3.1
✓ purrr      1.0.2

— Conflicts — tidyverse_conflicts()
—
✖ ggplot2::%+%( ) masks crayon::%+%( )
✖ dplyr::filter( ) masks stats::filter( )
✖ dplyr::lag( ) masks stats::lag( )
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

## Insertion des bases de données

Fin_2008_1 <- read.csv("/Users/15150135/Documents/Analyse Big
Data/fas_1001/_tp/_tp3/lipad/2008/2008-12-1.csv")

Fin_2008_2 <- read.csv("/Users/15150135/Documents/Analyse Big
Data/fas_1001/_tp/_tp3/lipad/2008/2008-12-2.csv")

Fin_2008_3 <- read.csv("/Users/15150135/Documents/Analyse Big
Data/fas_1001/_tp/_tp3/lipad/2008/2008-12-3.csv")

Fin_2008_4 <- read.csv("/Users/15150135/Documents/Analyse Big
Data/fas_1001/_tp/_tp3/lipad/2008/2008-12-4.csv")

```

```

##Combinaison des bases de données de 2008

Fin_2008 <- bind_rows(Fin_2008_1, Fin_2008_2, Fin_2008_3, Fin_2008_4)

## Insertion du dictionnaire en anglais

lexicoder_en <- dictionary(file = "/Users/15150135/Documents/Analyse Big
Data/fas_1001/dictionary/policy_agendas_english.lcd", format = "yoshikoder")

## Nettoyage de la base de données

Fin_2008_CLEAN <- Fin_2008 |>
  select(speechtext, speakerparty) |>
  ## Nettoyage de base ##
  mutate(speechtext = tolower(speechtext)) |>
  ## remove NA's ##
  na.omit()

## Téléchargement du clessnverse

run_dictionary <- function(data, text, dictionary) {
  tictoc::tic()
  if ( is.data.frame(data) != "TRUE") {
    stop(crayon::yellow('the argument "data" needs to be a dataframe'))
  }
  data <- data %>% dplyr::mutate(text = {{text}})
  if ( is.character(data$text) != "TRUE") {
    stop(crayon::yellow('The variable "text" needs to be a character
vector'))
  }
  corpus <- quanteda::tokens(data$text)
  if ( quanteda::is.dictionary(dictionary) != "TRUE") {
    stop(crayon::yellow('Your "dictionary" needs to be in a dictionary
format\n For more information:
https://quanteda.io/reference/dictionary.html'))
  }
  dfm <- quanteda::dfm(quanteda::tokens_lookup(corpus, dictionary,
nested_scope = "dictionary"))
  message(crayon::green("100% expressions/words found"))
  dataFinal <- quanteda::convert(dfm, to = "data.frame")
  tictoc::toc()
  return(dataFinal)
}

Base_test <- run_dictionary(data = Fin_2008_CLEAN,
  text = speechtext,
  dictionary = lexicoder_en) |>
bind_cols(Fin_2008_CLEAN) |>
select(-c(doc_id,speechtext)) |>
  # Pivote la base de données pour voir la proportion par speaker #

```



```

pivot_longer(!speakerparty, names_to = "categorie", values_to="n") |>
ungroup() |>
group_by(speakerparty, categorie) |>
summarise(n=sum(n)) |>
mutate(prop = round(n/sum(n),4)*100,
        speakerparty = case_when(speakerparty == "Conservative" ~
"Conservateur",
                                speakerparty == "Bloc Québécois" ~
"Bloc",
                                speakerparty == "New Democratic Party" ~
"NDP",
                                speakerparty == "Liberal" ~
"Libéral",
                                T ~ as.character(speakerparty)),
        categorie = case_when(categorie == "macroeconomics" ~ "Économie",
                                T ~ as.character(categorie))) |>
na.omit() |>
# On garde seulement le thème "économie" #
filter(categorie %in% c("Économie"),
        !speakerparty == "Independent") |>
ggplot(aes(x = speakerparty, y = prop, fill = speakerparty)) +
geom_bar(stat = "identity", position = "dodge") +
scale_fill_manual("", values = c("#0088CE", "#002395", "#EA6D6A",
"#FF5800", "#3D9B35")) +
labs(x = "Enjeu économique\n",
      y = "\nProportion (%)",
      title = "Discussion de l'enjeu économique par les partis politiques
canadiens, \nen Décembre 2008") +
theme_bw() +
## theme() en fonction des dimensions dans Quarto ##
theme(title = element_text(size = 10),
      plot.title = element_text(hjust = 0.5, face='bold'),
      legend.text = element_text(size = 10),
      axis.text = element_text(size = 10, color = "black"))

```

100% expressions/words found

0.86 sec elapsed

`summarise()` has grouped output by 'speakerparty'. You can override using the
`.groups` argument.