

Plan de recherche : Les politiciennes insistent-elles davantage que leurs homologues masculins sur les thématiques féminines lors des débats électoraux ? La présence d'une femme politique dans un débat signifie-t-elle davantage d'évocation des thématiques féminines par rapport à un débat incluant deux hommes politiques ?

Olivia Saffioti

2024-03-11

« De nombreux acteurs – électeurs, responsables de partis, candidats, journalistes – transfèrent leurs attentes stéréotypées à l'égard des hommes et des femmes aux candidats hommes et femmes. Le résultat de ces stéréotypes est que certains traits de personnalité et certains domaines d'expertise politique en viennent à être considérés comme féminins et d'autres comme masculins » (Fox et Oxley 2003, 834). Cette citation met l'emphase sur les stéréotypes de genre et les mécanismes liant le genre et la politique. Les postulats de la théorie de la construction sociale du genre sont mis en avant. Cette dernière stipule que le genre est un comportement social, que les individus adoptent lorsqu'ils interagissent. Il serait issu de la culture (Chauvin 2021). Les femmes et les hommes conformeraient leur manière de communiquer aux stéréotypes de genre afin de répondre aux attentes de la société (Grebelsky-Lichtman et Bdolach 2017, 276). Selon cette théorie, les électeurs évalueraient les politicien(ne)s en fonction des stéréotypes de genre. Si les politicien(ne)s n'adaptent par leur stratégie marketing aux stéréotypes, ils/elles risqueraient un backlash, soit, des évaluations négatives de la part des électeurs (Rudman et Fairchild 2004; Coyle 2009, cités dans Grebelsky-Lichtman et Katz 2019). Ainsi, les femmes et les hommes politiques n'utiliseraient pas des mêmes stratégies marketing afin de convaincre l'électorat. En effet, les femmes et les hommes répondraient souvent aux stéréotypes induits par le rôle attribué à chaque sexe dans la société. Ils conforteraient ainsi la vision des électeurs (Fox 1997). Comme mis en exergue dans notre citation, cela inclurait l'image, les caractéristiques

personnelles, les thèmes de campagne et l'utilisation d'enjeux politiques spécifiques (Fox 1997). À titre d'exemple, les femmes politiques seraient perçues par l'électorat comme plus efficaces dans les domaines de l'éducation, de la santé, de l'environnement, des arts, de la protection des consommateurs ou encore dans l'aide à apporter aux pauvres (Alexander and Andersen 1993; Koch 2000; McDermott 1998, cités dans Fox et Oxley 2003). À l'inverse, les hommes politiques seraient considérés comme plus compétents pour résoudre des crises militaires ou de police, des problèmes d'ordre économique ou encore des enjeux liés au commerce. Ils seraient également identifiés comme plus légitimes pour s'occuper du contrôle de la criminalité ou de l'agriculture (Alexander and Andersen 1993; Koch 2000; McDermott 1998, cités dans Fox et Oxley 2003). En revanche, nous n'avons pas trouvé de recherches qui visent la relation entre le genre des candidats et les thématiques abordées lors des débats électoraux. Il semble ainsi pertinent d'étudier si les femmes politiques insistent plus sur les thématiques « féminines » que les hommes et si ces thématiques sont davantage abordées lors des débats électoraux incluant une femme politique que lors des débats opposants deux hommes politiques. Beaucoup de recherches se sont intéressées au contexte états-unien. Or, nous trouvons pertinent de cibler un autre contexte occidental : la France. Nous avons choisi la France pour plusieurs raisons. Tout d'abord, les retranscriptions des débats électoraux français sont faciles d'accès. De surcroît, la France a compté plusieurs femmes politiques qui se sont présentées aux élections présidentielles et qui ont donc participé à des débats électoraux. Enfin, se centrer sur ce nouveau contexte permettra un apport scientifique concernant les mécanismes liant le genre et la communication politique. Autrement dit, cela nous permettra de vérifier si les postulats de la théorie de la construction sociale du genre sont applicables à d'autres contextes et se vérifient à travers les thématiques abordées au sein des débats électoraux. C'est pourquoi, nous avons décidé d'étudier les débats électoraux de second-tour des élections françaises de 1995 à 2017. Nous avons choisi ces débats car ceux qui avaient eu lieu après 2019 auraient biaisé nos résultats. En effet, la pandémie de Covid-19 demeurerait un sujet très discuté par les politiciens durant les campagnes suivant 2019. Ce contexte temporel ne nous aurait pas permis de vérifier si la santé était une thématique plus discutée lorsque des femmes prenaient part à un débat électoral. En outre, parmi ces quatre débats, deux impliquent des candidates femmes : Ségolène Royal et Marine Le Pen. Or, ces deux politiciennes appartiennent à des partis politiques différents : madame Royal est membre d'un parti de gauche (parti socialiste) et madame Le Pen est membre d'un parti d'extrême droite (le rassemblement national). Cette situation nous permettra de contrôler une variable : le parti d'appartenance. En effet, dépendamment du parti politique, les intérêts défendus et les positionnements politiques des candidat(e)s changent. À travers ces débats nous vérifierons si les thématiques féminines (la santé, l'éducation, la famille, l'environnement et la protection sociale) étaient plus abordées lors des débats auxquels madame Royal et madame Le Pen ont pris part. Nous pourrions également vérifier s'il s'agissait de thématiques plus importantes pour elles que pour les hommes (si elles insistaient davantage sur ces thématiques que les hommes politiques). Pour ce faire, nous utiliserons les bases de données portant sur les débats électoraux suivants : celui opposant Emmanuel Macron à Marine Le Pen en 2017, celui opposant François Hollande à Nicolas Sarkozy en 2012, celui opposant Ségolène Royal à Nicolas Sarkozy en 2007 et celui opposant Jacques Chirac à Lionel Jospin en 1995 (Fréchet 2024). Puis, nous nettoierons ces bases de données. Ensuite, nous

élaborerons un dictionnaire. À partir de ce dernier, nous analyserons nos bases de données afin de vérifier à quelles fréquences les thématiques de la santé, de l'éducation, de la famille, de l'environnement et de la protection sociale sont évoquées par les candidats lors des débats. Enfin, nous présenterons nos résultats.

Hypothèse

Nous n'avons pas trouvé d'études portant sur la relation entre le genre et les thématiques abordées lors des débats électoraux. En revanche, plusieurs études micro et macro ont déjà montré, que les femmes politiques ne priorisent pas les mêmes enjeux que les hommes (Paxton et Hughes 2021). Ainsi, en Afrique Sub-Saharienne, elles listeraient la pauvreté et les droits des femmes comme étant prioritaires contrairement aux hommes (Clayton et al. 2015, cité dans Paxton et Hughes 2021, 220). En Inde, les femmes politiques voient l'accès à l'eau potable et les politiques de bien-être comme un enjeu prioritaire, ce qui n'est pas le cas des hommes politiques (Chattopadhyay et Duflo 2004, cité dans Paxton et Hughes 2021, 222). Aux États-Unis, les femmes politiques donnent beaucoup d'importance aux politiques liées au bien-être, à la famille, aux enfants, à l'éducation, à la santé, ainsi qu'à la discrimination de genre et à l'amélioration du statut économique des femmes (Bratton et Haynie 1999; Swers 2013; Thomas 1991; Reingold et Smith 2012, cités dans Paxton et Hughes 2021, 226).

En outre, selon la théorie de la construction sociale du genre et du schéma du genre, les électeurs perçoivent les politiciennes comme plus efficaces dans les domaines politiques dits « féminins » (Sanbonmatsu 2002; Fox et Oxley 2003). Ainsi, les électeurs pourraient voter pour des candidats du même sexe qu'eux car ils défendraient leurs intérêts (Aalberg et Todal Jenssen 2007). Par exemple, les femmes voteraient pour des politiciennes car elles partageraient des intérêts communs relatifs à l'éducation, à l'émancipation des femmes, ou à d'autres domaines liés à l'expertise de la gent féminine. Dépendamment du contexte, les préoccupations politiques des citoyens peuvent être liées à des domaines relatifs à l'expertise des femmes comme la santé ou l'environnement. Dans cette situation, certains segments d'électeurs peuvent avoir tendance à voter de manière stratégique en soutenant une candidate femme (Herrnson et al. 2003). Les femmes politiques pourraient également profiter des préoccupations de l'opinion publique pour insister sur les thématiques « féminines » dans leur programme politique. Cela leur procurerait du soutien électoral supplémentaire, les politiciennes étant évaluées plus positivement lorsqu'elles insistent sur les thématiques politiques « féminines » durant leur campagne (Kahn 1996). Ces mêmes préoccupations des électeurs pourraient influencer les thématiques abordées lors des débats électoraux. En débat électoral, les politiciennes pourraient également davantage insister sur ces thématiques que leurs homologues masculins. Cela leur permettrait de mettre l'accent sur le fait que leur agenda politique est différent de ceux de leurs adversaires hommes (Aalberg et Todal Jenssen 2007, 22). Elle se conformeraient aux stéréotypes de genre en montrant qu'elles vouent beaucoup d'importance à ces thématiques « féminines ». L'objectif serait d'obtenir davantage de soutien électoral.

C'est pourquoi, nous émettons l'hypothèse suivante : dans un débat électoral, lorsqu'une femme politique est présente, les thématiques « féminines » sont davantage abordées. Mesdames Royal et Le Pen insistent également davantage sur les thématiques « féminines » que leurs homologues masculins.

Données mobilisées dans le cadre de la recherche

Dans le cadre de notre recherche, nous avons mobilisé des données textuelles. Nous avons utilisé les retranscriptions des quatre débats électoraux. Nous avons trouvé ces retranscriptions en libre accès et en format pdf sur le site internet vie-publique.fr. Ce site internet a été mis en place par la République Française. Nous avons enregistré les retranscriptions. Puis, nous avons converti ces documents pdf en bases de données au format csv à l'aide du logiciel R. Toutes les retranscriptions ont été converties en bases de données à partir des mêmes variables : une variable « id » pour spécifier les candidats qui sont opposés dans les débats (par exemple : Royal/Sarkozy désigne que le débat oppose Ségolène Royale à Nicolas Sarkozy). Une variable nommée « year » a également été créée afin de spécifier l'année des débats, et une variable « date » a été conçue afin d'identifier la date exacte (jour/mois/année) des débats. Une autre variable intitulée « country » a permis de mettre en lumière le pays dans lequel ont eu lieu les débats. Une variable nommée « speaker » a également été instaurée afin de relever qui s'exprime au sein des débats. Une autre variable intitulée « speaker_turn » met en exergue le tour de parole entre les intervenants des débats. Nous avons également créé une variable « party » qui spécifie le parti politique d'appartenance des candidats. Elle montre également qui sont les journalistes. Enfin, nous avons importé les textes des retranscriptions des débats sous la variable « text ». Les bases de données comptent donc chacune 8 variables. Monsieur Nadjim Fréchet nous a aidé à convertir les retranscriptions en bases de données. Nous avons mis les codes R qu'il a utilisé en annexe.

Pour mener notre recherche, nous nous centrerons sur trois variables spécifiques : la variable « text », la variable « speaker » ainsi que la variable « id ». En effet, la variable id nous permettra d'identifier les débats dans lesquels les fréquences de mots ont été relevées. La variable « speaker » permettra d'identifier à quels candidats correspondent les fréquences de mots. En outre, ce sera sur la variable « text » que nous appliquerons notre dictionnaire. Nous allons donc nettoyer nos bases de données afin de conserver uniquement les variables qui sont pertinentes à notre recherche. Pour cela, nous avons utilisé le « pipe operator » et la fonction « select () » du package « dplyr ». Cela nous a permis de créer une nouvelle base de données comprenant uniquement les variables « id », « speaker » et « text ». En outre, dans le cadre de notre étude, nous visons principalement le discours des candidats. C'est pourquoi nous avons supprimé les éléments textuels qui représentaient les questions/commentaires des journalistes. Pour ce faire, nous avons utilisé la fonction « filter () » du package « dplyr » afin de conserver uniquement les données textuelles qui rapportent aux paroles des candidats. Nous avons également retiré les majuscules afin que l'analyse de dictionnaire ne soit pas biaisée : un mot présent dans le dictionnaire pourrait ne pas être relevé s'il contient une majuscule dans

le texte. Pour ce faire nous avons mobilisé la fonction « `tolower()` ». Nous avons également supprimé la ponctuation afin d’uniformiser le texte et de faciliter notre analyse de dictionnaire. En effet, la ponctuation est inutile pour notre analyse et nous voulons éviter que cette dernière n’influence les résultats de notre recherche. Pour l’ôter du texte, nous avons utilisé la fonction « `gsub()` ». Nous avons mis nos codes R en annexe, avec l’exemple du débat entre Emmanuel Macron et Marine Lepen.

Méthode de la recherche

Afin de mener notre recherche nous allons mobiliser une méthode quantitative. Nous allons réaliser une analyse textuelle automatisée. Ce type d’analyse permet une lecture plus attentive des textes. L’analyse textuelle apparaît ainsi plus précise et minutieuse (Grimmer et Stewart 2013, 268). Grâce à cette analyse, nous allons pouvoir classer les contenus des retranscriptions sous plusieurs catégories que nous aurons élaborées (Grimmer et Stewart 2013, 268). Nous allons plus précisément mener une analyse de dictionnaire. Cette méthode vise à identifier la fréquence à laquelle des mots clés apparaissent dans un texte (Grimmer et Stewart 2013, 274). Ces mots sont classés selon des catégories et permettent de rendre compte du contenu des textes (Grimmer et Stewart 2013, 274).

Dans le cadre de notre recherche, nous allons créer notre propre dictionnaire. En effet, créer notre dictionnaire permet d’inclure un éventail de mots représentant les sujets politiques que nous souhaitons étudier à travers notre étude. En outre, créer son dictionnaire apporte une fiabilité interne pertinente : les mots du dictionnaire sont cohérents et logiques car représentent bien les thématiques politiques que nous souhaitons étudier. Créer un dictionnaire permet également une bonne validité convergente : les mots inclus au sein du dictionnaire sont utilisés/relevés dans le discours des politiciens (Nicolas et al. 2019, 9). Autrement dit, ils sont adaptés à nos données textuelles. Afin de créer notre dictionnaire, nous nous sommes basés sur les dictionnaires thématiques Lexicoder d’Albugh, Quinn, Julie Sevenans et Stuart Soroka et nous avons sélectionné certains mots. Nous avons également rajouté des mots qui nous semblaient pertinents et qui n’étaient pas présents dans les dictionnaires thématiques sur lesquels nous nous sommes appuyés.

Notre dictionnaire comprend cinq catégories différentes. Tout d’abord, nous avons une catégorie portant sur la santé. Il comprend les termes : « santé », « médecin », « médecins », « hôpital », « hôpitaux », « maladie », « maladies », « patient », « patients », « soins », « assurance maladie », « malades », « médicaments », « aide médicale », « visites médicales », « visite médicale », « sida », « vih », « cigarette », « traitement », « traitements », « abus de substances », « consommation de substances », « vaccination », « vaccin », « vaccins », « immunisation », « immuniser », « médical », « médicale », « médicaux », « médicales », « infirmiers », « infirmières », « infirmier », « infirmière », « pharmacie », « pharmacies », « pharmaceutique », « pharmaceutiques », « santé mentale », « tabac », « médecine », « épidémie », « pandémie », « obésité », « obèse », « obèses », « euthanasie », « fécondation in vitro », « contraception », « contraceptions », « éducation sexuelle », « maladies sexuellement transmissibles », « maladie

sexuellement transmissible », « pilule », « mst », « infection sexuellement transmissible », « infections sexuellement transmissibles », « grippe ».

Notre deuxième catégorie porte sur la thématique de la famille, et contient les mots : « enfants », « enfant », « famille », « familles », « famille nombreuse », « familles nombreuses », « parents », « parent », « familial », « familiale », « familiales », « familiaux », « crèche », « petite-enfance ».

Notre troisième catégorie porte sur la thématique de l'environnement. Les termes que nous avons sélectionnés sont : « taxe carbone », « amiante », « amiantes », « alimentation en eau », « eau potable », « accès à l'eau », « déchets dangereux », « déchets radioactifs », « déchets nucléaires », « qualité de l'air », « développement durable », « gaz à effet de serre », « pollution », « déforestation », « écologie », « écologique », « voitures électriques », « consommation de pétrole », « carbone », « puits de carbones », « réchauffement climatique », « changement climatique », « énergies renouvelables », « nucléaire », « environnemental », « environnementale », « éoliennes », « éolienne », « solaire », « solaires », « bio carburants », « chaud », « chauds », « réchauffement planétaire », « uranium », « combustible », « combustibles ».

Notre quatrième catégorie porte sur la thématique de protection sociale, et comprend les mots : « retraite », « retraites », « protection sociale », « protections sociales », « salariés », « indemnisation », « indemnisations », « pauvreté », « inégalités », « allocation », « allocations », « quotient », « aide », « aides », « assurance », « assurances », « banque alimentaire », « banques alimentaires », « appauvrir », « appauvrissement », « faible revenu », « revenu », « revenus », « sécurité sociale », « assurance chômage », « bien-être », « pauvreté », « aide alimentaire », « aides alimentaires », « service social », « services sociaux », « mère célibataire », « père célibataire », « privation sociale », « privation matérielle et sociale », « niveau de vie », « déprivation sociale », « déprivation matérielle et sociale », « cotisations sociales », « justice sociale », « assistantes sociales », « prestations sociales », « protection sociale », « tva sociale », « pension », « pensions », « chômage », « chômeurs », « insécurité sociale », .

Enfin, notre cinquième catégorie porte sur l'éducation et inclut les mots : « collège », « collèges », « enseignement », « élève », « élèves », « classe », « classes », « diplôme », « diplômes », « lycée », « lycées », « université », « universités », « universitaire », « universitaires », « formation », « formations », « apprentissage », « apprentissages », « filière », « filières », « professionnelle », « professionnelles », « école », « écoles », « maternelle », « école primaire ».

Ainsi, nous appliquerons ce dictionnaire à nos données textuelles. Le dictionnaire permettra au logiciel R de relever la fréquence des mots que nous venons d'énoncer. Plus précisément, il permettra de relever la fréquence des mots dans le discours de chaque candidat. Pour ce faire, nous utiliserons le package « clessnverse » de R, et nous mobiliserons la fonction « run_dictionnary () ». Une fois que nous aurons mené cette analyse avec l'ensemble de nos textes, nous combinerons les résultats obtenus dans une seule base de données. Cette base de données contiendra les variables « id » (soit, celle qui explique quels candidats sont opposés dans le débat) et « speaker », ainsi que les fréquences de mots associées à chaque candidat pour chaque débat. Puis, nous visualiserons les données de nos résultats à l'aide de deux graphiques en barres. Un premier servira à comparer la fréquence des mots entre les candidats.

Un second permettra de comparer le total des fréquences de mots relevées dans les débats (la somme des fréquences de mots relevées pour les deux candidats dans chaque débat). Le premier graphique indiquera si les femmes politiques insistent davantage sur les thématiques dites « féminines » que les hommes. Le deuxième graphique illustrera si la présence d'une candidate femme dans un débat permet aux thématiques féminines d'être davantage évoquées. Pour les générer, nous utiliserons la fonction « ggplot() » présente dans le package « tidyverse ». Ainsi, nous pourrions identifier si lorsqu'une femme politique participe à un débat, les thématiques politiques « féminines » sont traitées avec plus d'insistance. Nous saurons également si le contenu des débats varie en fonction du genre des candidats.

Bibliographie

Aalberg, Toril, et Anders Todal Jenssen. 2007. « Gender Stereotyping of Political Candidates: An Experimental Study of Political Communication ». *Nordicom Review* 28 (1): 17-32. <https://doi.org/10.1515/nor-2017-0198>.

Albugh, Quinn, Julie Sevenans et Stuart Soroka. 2013. *Lexicoder Topic Dictionaries*, versions de juin 2013. Montréal : Université McGill. <https://www.snsoroka.com/data-lexicoder>.

Chauvin, S.bastien.2021. . La construction sociale du genre comme construction sociale”. *Implications philosophiques*, 15 décembre 2021. <https://sebastienchauvin.org/category/publications/gender/>.

Fox, Richard Logan. 1997. *Gender Dynamics in Congressional Elections*. Thousand Oaks:Éditions SAGE. https://books.google.ca/books?hl=en&lr=&id=x2w1BfJwRkAC&oi=fnd&pg=PR13&dq=gender+women+elections&ots=8RGQYfimmw&sig=Uo31w_YYAhrRtOEKSyA6ABiEeWw#v=onepage&q=gender%20women%20elections&f=false

Fox, Richard L., et Zoe M. Oxley. 2003. « Gender Stereotyping in State Executive Elections: Candidate Selection and Success ». *The Journal of Politics* 65 (3): 833-50. <https://doi.org/10.1111/1468-2508.00214>.

Fréchet, Nadjim. 2024. *Bases de données portant sur les débats électoraux entre Emmanuel Macron et Marine Le Pen (2017), François Hollande et Nicolas Sarkozy (2012), Ségolène Royal et Nicolas Sarkozy (2007), et Jacques Chirac et Lionel Jospin (1995)*.

Grebelsky-Lichtman, Tsira, et Liron Bdolach. 2017. « Talk like a man, walk like a woman: an advanced political communication framework for female politicians ». *The Journal of Legislative Studies* 23 (3): 275-300. <https://doi.org/10.1080/13572334.2017.1358979>.

Grebelsky-Lichtman, Tsira, Roy Katz. 2019. « When a man debates a woman: Trump vs. Clinton in the first mixed gender presidential debate ». *Revue Journal of Gender Studies* 28 (6): 699-719.<https://www.tandfonline.com/doi/full/10.1080/09589236.2019.1566890>

Grimmer, Justin, et Brandon M. Stewart. 2013. « Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts ». *Political Analysis* 21 (3):

267-97. <https://www.cambridge.org/core/journals/political-analysis/article/text-as-data-the-promise-and-pitfalls-of-automatic-content-analysis-methods-for-political-texts/F7AAC8B2909441603FEB25C156448F20>

Herrnson, Paul S., J. Celeste Lay, et Atiya Kai Stokes. 2003. « Women Running “as Women”: Candidate Gender, Campaign Issues, and Voter-Targeting Strategies ». *The Journal of Politics* 65 (1): 244-55. <https://doi.org/10.1111/1468-2508.t01-1-00013>.

Kahn, Kim Fridkin. 1996. *The Political Consequences of Being a Woman: How Stereotypes Influence the Conduct and Consequences of Political Campaigns*. New York: Columbia University Press.

Nicolas, Gandalf, Xuechunzi Bai, et Susan Fiske. 2019. « Automated Dictionary Creation for Analyzing Text: An Illustration from Stereotype Content ». *Running head: automated dictionary creation*. Department of Psychology, Princeton University. <https://doi.org/10.31234/osf.io/afm8k>.

Paxton, Pamela et Melanie Hughes. 2021. « Chapter 9: Do Women Make a Difference? ». Dans *Women, Politics, and Power: A Global Perspective*. Sous la direction de Pamela Paxton, Melanie M. Hughes, et Tiffany D. Barnes 4e éd., 218-244. Los Angeles : Pine Forge Press. <https://umontreal.on.worldcat.org/oclc/1245199736>

Sanbonmatsu, Kira. 2002. « Gender Stereotypes and Vote Choice ». *American Journal of political Science* 46 (1): 20-34. <https://doi.org/10.2307/3088412>.

Annexe

Voici les codes utilisés pour créer et nettoyer les bases de données des textes (exemple avec le texte du débat opposant Marine Le Pen et Emmanuel Macron).

```
# Convertir la retranscription pdf en base de données csv
```

```
library(pdftools)
```

Using poppler version 23.04.0

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
```



```
v ggplot2 3.4.4 v tibble 3.2.1
v lubridate 1.9.3 v tidyr 1.3.0
v purrr 1.0.2
```

```
-- Conflicts ----- tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
```

```
x dplyr::lag() masks stats::lag()
```

```
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(lubridate)
```

```
text_1 <- suppressMessages(pdftools::pdf_text("/Users/oliviasaffioti/Desktop/fas_1001_Saffi"))
```

```
split_vector <- paste0(c("Mme Le Pen :", "Mme Saint-Cricq :", "M. Jakubyszyn :", "M. Macron :"))
```

```
speaker_names <- data.frame(text = paste0(text_1, collapse = " ")) |>
```

```
  mutate(text = str_squish(str_replace_all(text, c("[:punct:]C2[:punct:]AB" = "",
                                                    "[:punct:]C2[:punct:]BB" = ""))),
```

```
    text = str_replace_all(text, c("Emmanuel Macron :" = "M. Macron :",
                                   "Marine Le Pen :" = "Mme Le Pen :",
                                   "Christophe Jakubyszyn :" = "M. Jakubyszyn :",
                                   "Nathalie Saint-Cricq :" = "Mme Saint-Cricq :")),
```

```
    speaker = str_extract_all(text, split_vector)) |>
```

```
select(-text) |>
```

```
unnest(speaker) |>
```

```
slice(-805)
```

```
text_data <- data.frame(text = paste0(text_1, collapse = " ")) |>
```

```
  mutate(text = str_squish(str_replace_all(text, c("[:punct:]C2[:punct:]AB" = "",
                                                    "[:punct:]C2[:punct:]BB" = ""))),
```

```
    text = str_replace_all(text, c("Emmanuel Macron :" = "M. Macron :",
                                   "Marine Le Pen :" = "Mme Le Pen :",
                                   "Christophe Jakubyszyn :" = "M. Jakubyszyn :",
                                   "Nathalie Saint-Cricq :" = "Mme Saint-Cricq :")),
```

```
    text = str_split(text, split_vector)) |>
```

```
unnest(text) |>
```

```
slice(-c(1,806))
```

```
Data_debate_2017 <- bind_cols(speaker_names, text_data) |>
```

```
  mutate(year = 2017,
```

```
         date = ymd("2017-05-03"),
```

```

country      = "France",
id           = "Lepen/Macron",
speaker_turn = 1:804,
speaker = str_squish(str_replace_all(speaker, c("Mme" = "",
                                                "^M." = "",
                                                ":" = ""))),
party = case_when(speaker == "Le Pen" ~ "Rassemblement national",
                  speaker == "Macron" ~ "Renaissance",
                  .default = "Journaliste")) |>
select(id, year, date, country, speaker, speaker_turn, party, text)

# Enregistrer la base de données

write_csv(Data_debate_2017, "/Users/oliviasaffioti/Desktop/fas_1001_Saffioti/_tp/TP3/debat

#Nettoyer la base de données

library(quanteda)

```

Package version: 3.3.1
 Unicode version: 14.0
 ICU version: 71.1
 Parallel computing: 8 of 8 threads used.
 See <https://quanteda.io> for tutorials and examples.

```
library(clessnverse)
```

DISCLAIMER: As of July 2023, `clessnverse` is no longer under active development.
 To avoid breaking dependencies, the package remains available "as is" with no warranty of any

```

library(tidyverse)

debat_macron_lepen <- read_csv("debat_macron_lepen.csv")

debat_macron_lepen$text <- tolower(debat_macron_lepen$text)
debat_macron_lepen$text <- gsub("[[:punct:]]", "", debat_macron_lepen$text)

```

```
debat_macron_lepen <-debat_macron_lepen %>% select (id, speaker, text) %>% filter(speaker
```