

# **Projet final : L'impact du genre sur la saillance des thématiques politiques féminines au sein des débats électoraux**

Olivia Saffioti

2025-07-04

« 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'accent 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 pas 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.

## ***Hypothèses***

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. Autrement dit, nous avons fait du « web scraping ». 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 utilisés pour créer ces bases de données en annexe.

Pour mener notre recherche, nous nous sommes centrés sur trois variables spécifiques : la variable « text », la variable « speaker » ainsi que la variable « id ». En effet, la variable « id » nous a permis d'identifier les débats dans lesquels les fréquences de mots ont été relevées. La variable « speaker » nous a permis d'identifier à quels candidats correspondaient les fréquences de mots. En outre, c'est sur la variable « text » que nous avons appliqué notre dictionnaire. Nous avons donc nettoyé nos bases de données afin de conserver uniquement les variables qui demeuraient 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 visions 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 était inutile pour notre analyse et nous voulions 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.

Notre base de données finale comportait les résultats de notre analyse de dictionnaire pour l'ensemble des débats et des candidats. Dans le cadre de notre recherche, nous avons rajouter une variable nommé « orientation\_partis » au sein de cette base de données finale. Cette variable contenait les catégories « Droite », « Gauche » et « Centre ». Elle avait pour objectif de discerner quels candidats étaient issus d'un parti de droite, de gauche et de centre. Cette variable catégorielle nous a servi à tester l'impact de l'orientation politique des candidats sur la manière dont ces derniers insistaient sur les thématiques politiques féminines.

### ***Méthode de la recherche***

Afin de mener notre recherche nous avons mobilisé une méthode quantitative. Nous avons réalisé une analyse textuelle automatisée. Grâce à cette analyse, nous avons pu classifier les contenus des retranscriptions sous plusieurs catégories que nous avons élaborées (Grimmer et Stewart 2013, 268). Nous avons plus précisément mené 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 avons créé 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 », « malades », « malade », « 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 », « crèches », « petite-enfance », « mère célibataire », « père célibataire », « mères célibataires », « pères célibataires ».

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 », « salarié », « 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 », « 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 », « assurance maladie ».

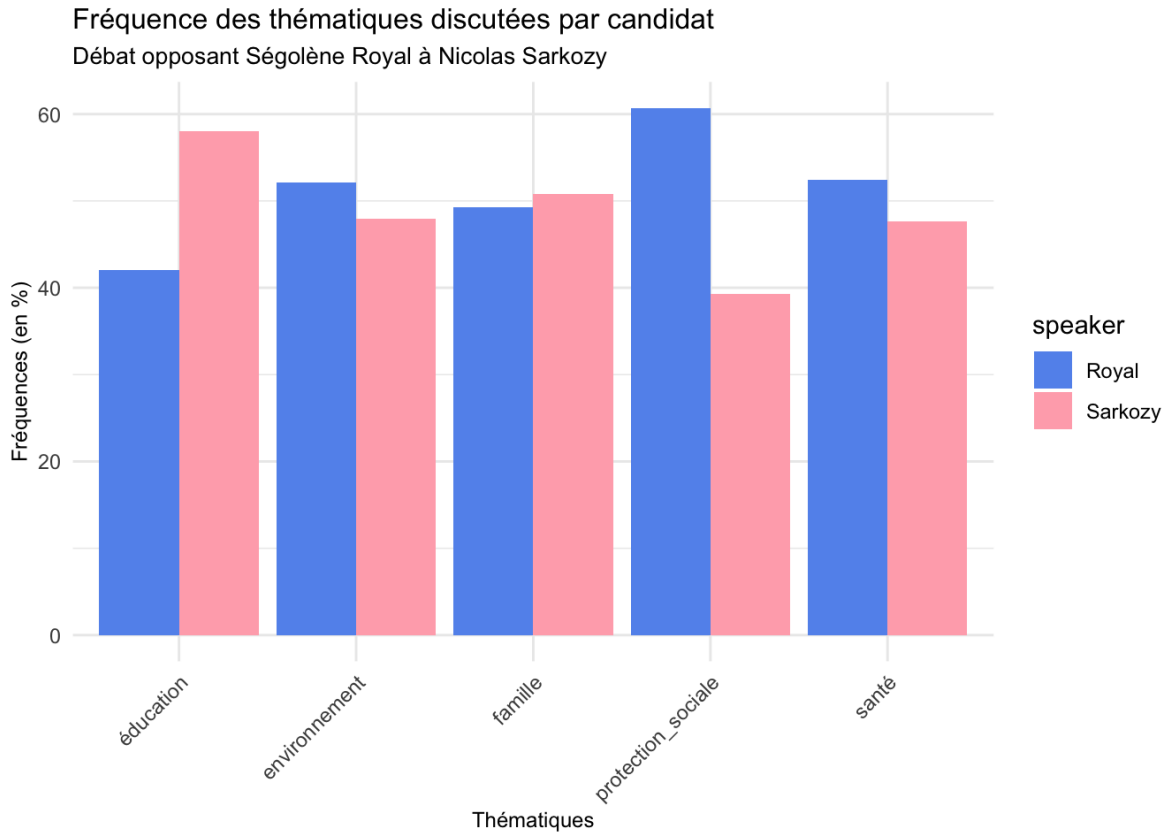
Enfin, notre cinquième catégorie porte sur l'éducation et inclut les mots : « collège », « collèges », « enseignement », « élève », « élèves », « classe », « classes », « diplôme », « diplômes », « lycée », « lycées », « université », « universités », « universitaire », « universitaires », « formation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

Ainsi, nous avons appliqué ce dictionnaire à nos données textuelles. Le dictionnaire a permis au logiciel R de relever la fréquence des mots que nous venons d'énoncer. Plus précisément, il a permis de relever la fréquence des mots dans le discours de chaque candidat. Pour ce faire, nous avons utilisé le package « clessnverse » de R, et nous avons mobilisé la fonction « run\_dictionary () ». Une fois que nous avons mené cette analyse avec l'ensemble de nos textes, nous avons combiné les résultats obtenus dans une seule base de données. Cette base de données contenait les variables « id » (soit, celle qui explique quels candidats sont opposés dans le débat) et « speaker », ainsi que les fréquences de mots (pour chaque thématique féminine) associées à

chaque candidat pour chaque débat. Puis, nous avons visualisé les données de nos résultats à l'aide de deux graphiques en barres. Un premier a servi à comparer la fréquence des mots entre les candidats. Un second a permis 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 a indiqué si les femmes politiques insistaient davantage sur les thématiques dites « féminines » que les hommes. Le deuxième graphique a illustré si la présence d'une candidate femme dans un débat permet aux thématiques féminines d'être davantage évoquées. Nous avons également élaboré des graphiques montrant l'impact de l'orientation politique/ du parti d'appartenance des candidats sur la saillance des thématiques politiques féminines. Pour les générer, nous avons utilisé la fonction « `ggplot()` » présente dans le package « `tidyverse` ». Ainsi, nous avons pu identifier si lorsqu'une femme politique participe à un débat, les thématiques politiques « féminines » sont traitées avec plus d'insistance. Nous avons également vérifié si le contenu des débats varie en fonction du genre des candidats, et si l'orientation politique des candidats impacte la manière dont ils insistent sur ces thématiques politiques.

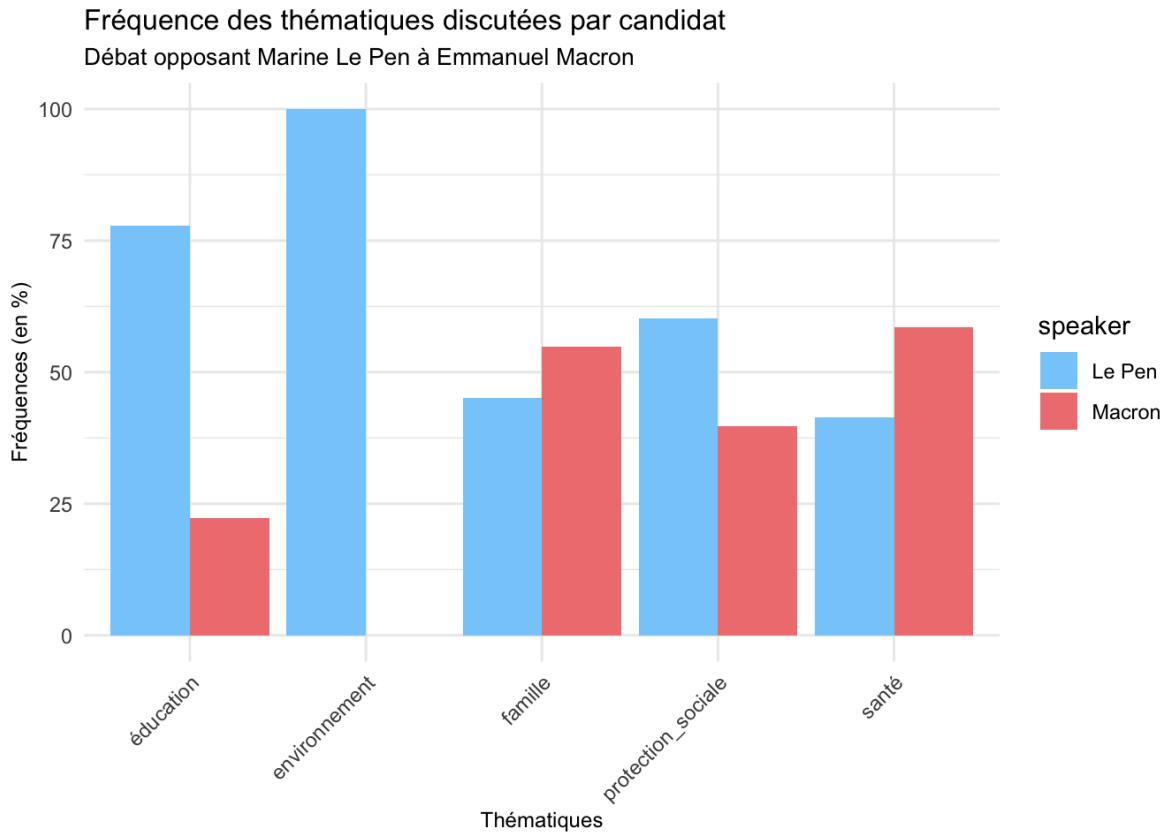
## Résultats

### 1) L'impact du genre sur la saillance des thématiques politiques féminines



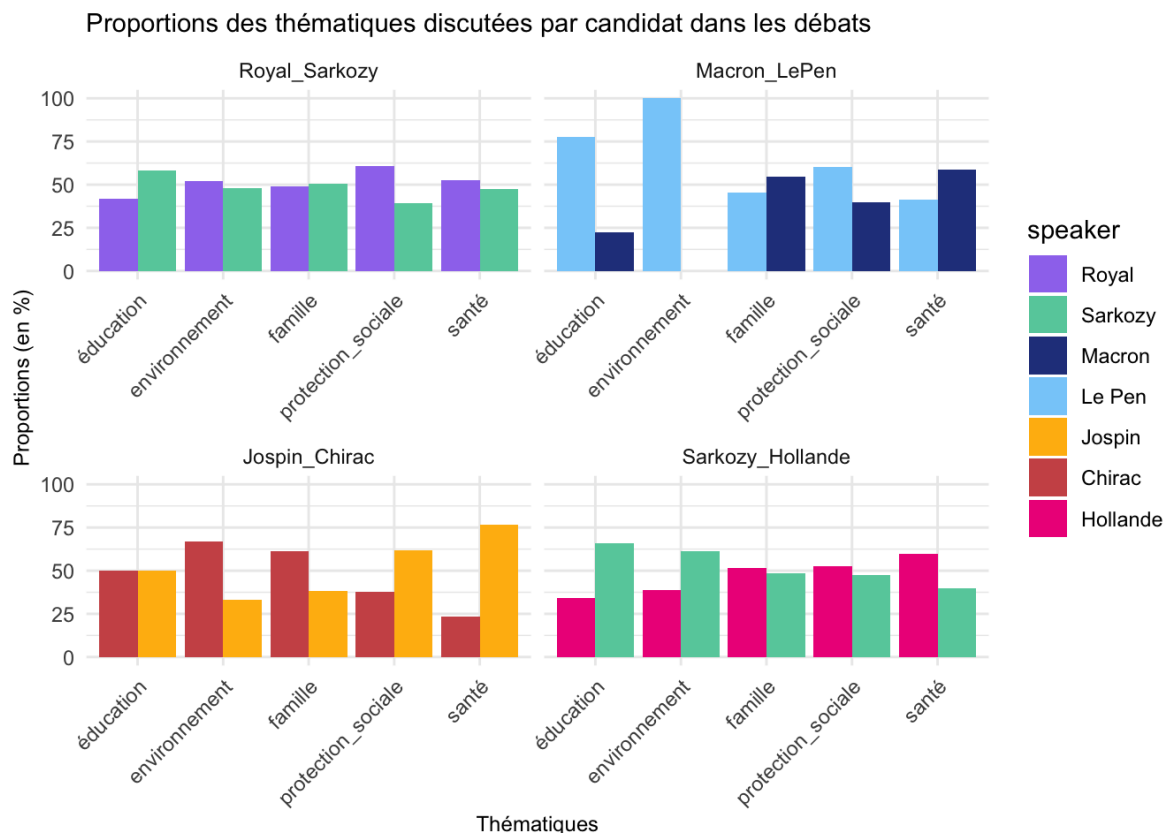
Tout d'abord, en analysant le débat opposant Ségolène Royal à Nicolas Sarkozy, nous pouvons constater qu'il existe des écarts entre les fréquences de mots relevés pour chaque candidat. En effet, Madame Royal semble avoir davantage insisté sur la thématique de la protection sociale par rapport à son adversaire. Au total 60.7% des mots relatifs à cette thématique ont été relevés dans le discours de la candidate, contre 39.3% des mots pour Sarkozy. Cela représente un écart de 25 mots entre le discours de Royal et de Sarkozy. En revanche, pour les autres thématiques, les écarts entre les deux candidats sont insuffisants en nombre de mots pour conclure que l'un aurait plus abordé une thématique que l'autre. Par exemple, pour l'éducation, 42% des mots ont été relevés dans le discours de madame Royal, 58% dans le discours de monsieur Sarkozy. Cependant, derrière ces pourcentages se cachent uniquement 8 mots d'écart entre les deux candidats. Ainsi, dans l'ensemble, Madame Royal aurait plus insisté sur les thématiques féminines que son adversaire.



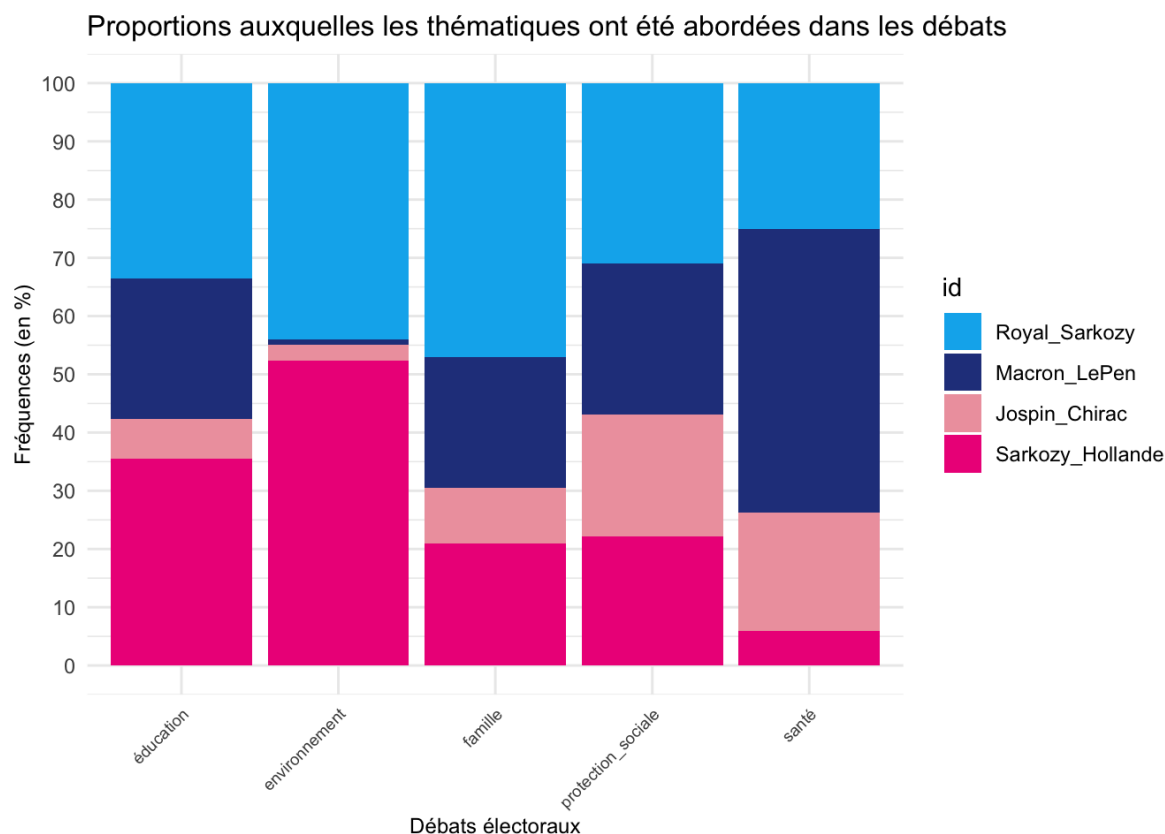


Concernant le débat opposant Marine Le Pen à Emmanuel Macron, nous avons relevé là encore des écarts entre les fréquences de mots relevés pour chaque candidat. En effet, Madame Le Pen semble avoir davantage abordé la thématique de la protection sociale par rapport à Monsieur Macron. 60.2% des mots relatifs à cette thématique ont été relevés dans le discours de madame Le Pen, contre 39.8% des mots pour monsieur Macron. Plus précisément, il existe un écart de 20 mots entre les deux candidats par rapport à la protection sociale. Madame Le Pen a également plus insisté sur l'éducation que son adversaire. 77.8% des mots relatifs à cette thématique ont été relevés dans le discours de Madame Le Pen contre 22.2% des mots pour monsieur Macron. Plus précisément, il y a un écart de 20 mots entre les deux candidats. En outre, nous pouvons observer que la thématique de l'environnement serait à 100% abordée par Le Pen. En revanche, ce n'est pas le cas. En effet, un seul mot a été relevé pour la thématique de l'environnement au total. Or, un seul mot ne suffit pas à prouver que la candidate a parlé de l'environnement. Ce mot a probablement été sorti de son contexte : un mot de notre dictionnaire comme "charbon" ou encore "uranium" auraient pu être mobilisés dans le cadre d'autres thématiques telles que l'économie. Cependant, concernant les autres thématiques féminines, nous pouvons observer que les écarts de mots ne sont pas élevés. Ainsi, ils ne sont pas suffisamment significatifs pour montrer qu'un candidat aurait plus ou moins insisté sur ces thématiques que l'autre. Donc, Le Pen et Macron auraient tout autant insisté sur les autres thématiques. Dans l'ensemble, il

semblerait que Le Pen ait ainsi davantage abordé les thématiques féminines que son adversaire masculin.



En revanche, le genre ne serait pas la variable impactant la saillance des thématiques féminines dans le discours des candidats. En effet, en analysant les débats opposants Lionel Jospin à Jacques Chirac en 1995 et Nicolas Sarkozy à François Hollande en 2012, nous avons pu constater également des écarts entre les candidats. Par exemple, Jospin a davantage insisté sur les thématiques de la santé et de la protection sociale par rapport à Chirac, ou encore Sarkozy a plus insisté sur l'éducation et l'environnement par rapport à Hollande. Autrement dit, d'autres variables explicatives que le genre impacteraient le fait que certains candidats insistent plus que d'autres sur les thématiques féminines.

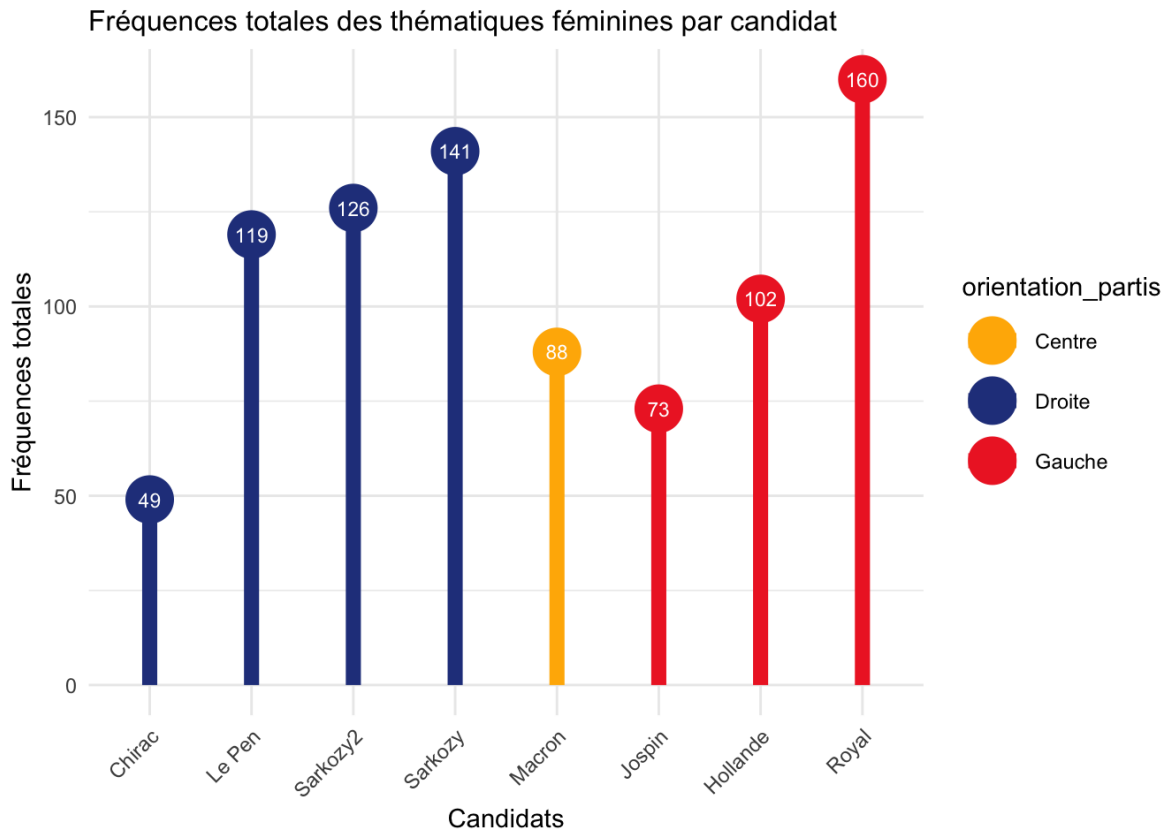


De surcroît, le fait qu'une femme politique participe à un débat ne conduit pas nécessairement à une plus grande saillance des thématiques féminines. En effet, comme nous pouvons le constater sur le graphique ci-dessus, certaines thématiques féminines sont plus abordées dans des débats opposants deux hommes (en rose) et d'autres le sont davantage dans les débats auxquels participe une femme (en bleu). Nous avons pu constater cela en observant la répartition des mots correspondants aux thématiques. C'est-à-dire que nous avons vérifié dans quels débats les mots relatifs à l'éducation, à la santé, à l'environnement, à la famille et à la protection sociale, étaient les plus présents. L'éducation a été le plus abordée lors du débat opposant Sarkozy à Hollande (35.6% des mots) et lors du débat opposant Royal à Sarkozy (33.6% des mots). Elle a en revanche été moins saillante lors du débat opposant Le Pen à Macron (24.2% des mots), et très peu abordée lors du débat opposant Jospin à Chirac (6.71% des mots). L'environnement a été le plus abordé lors du débat opposant Sarkozy à Hollande (52.3% des mots), puis lors du débat opposant Royal à Sarkozy (44% des mots). En revanche, elle n'a été que très peu, voire pas du tout saillante dans les débats opposants Le Pen à Macron (0.917 % des mots) et Chirac à Jospin (2.75% des mots). Quant à la famille, cette thématique a été le plus abordée durant le débat opposant Royal à Sarkozy (47.1% des mots). Elle a été moins abordée lors des débats opposants Sarkozy à Hollande (21% des mots), et Le Pen à Macron (22.5% des mots). La famille a été très peu saillante lors du débat opposant Jospin à Chirac (9.42% des

mots). Concernant la protection sociale, cette thématique a été plus discutée au sein du débat opposant Royal à Sarkozy (31% des mots), puis lors du débat opposant Le Pen à Macron (25.9% des mots). Elle a moins été abordée lors des débats opposants Jospin à Chirac (20.9% des mots) et Sarkozy à Hollande (22.2% des mots). Enfin, la santé a été beaucoup abordée au sein du débat opposant Le Pen à Macron (48.8% des mots). Elle l'a moins été durant les débats opposants Royal à Sarkozy (25% des mots) et Jospin à Chirac (20.2% des mots). Elle ne l'a été que très peu durant le débat opposant Sarkozy à Hollande (5.95% des mots). Ainsi, face à ces résultats nous pensons que d'autres variables explicatives que le genre expliqueraient la variation de la saillance des thématiques féminines selon les débats.

En somme, les femmes politiques n'insistent pas plus que leurs homologues masculins sur les thématiques féminines. La présence d'une femme politique dans un débat électoral ne signifie pas que davantage d'importance sera accordée aux thématiques féminines. Nos deux hypothèses de départ ont donc été infirmées. D'autres variables auraient impacté la saillance des thématiques féminines dans le discours des candidats. Nous pensons que l'orientation politique des candidats sur l'axe Gauche/Droite pourrait être la variable agissant sur la saillance de ces thématiques. Il nous est donc apparu intéressant de vérifier si l'orientation politique (et donc le parti politique d'appartenance) impactait la manière dont les candidats insistaient sur les thématiques politiques féminines.

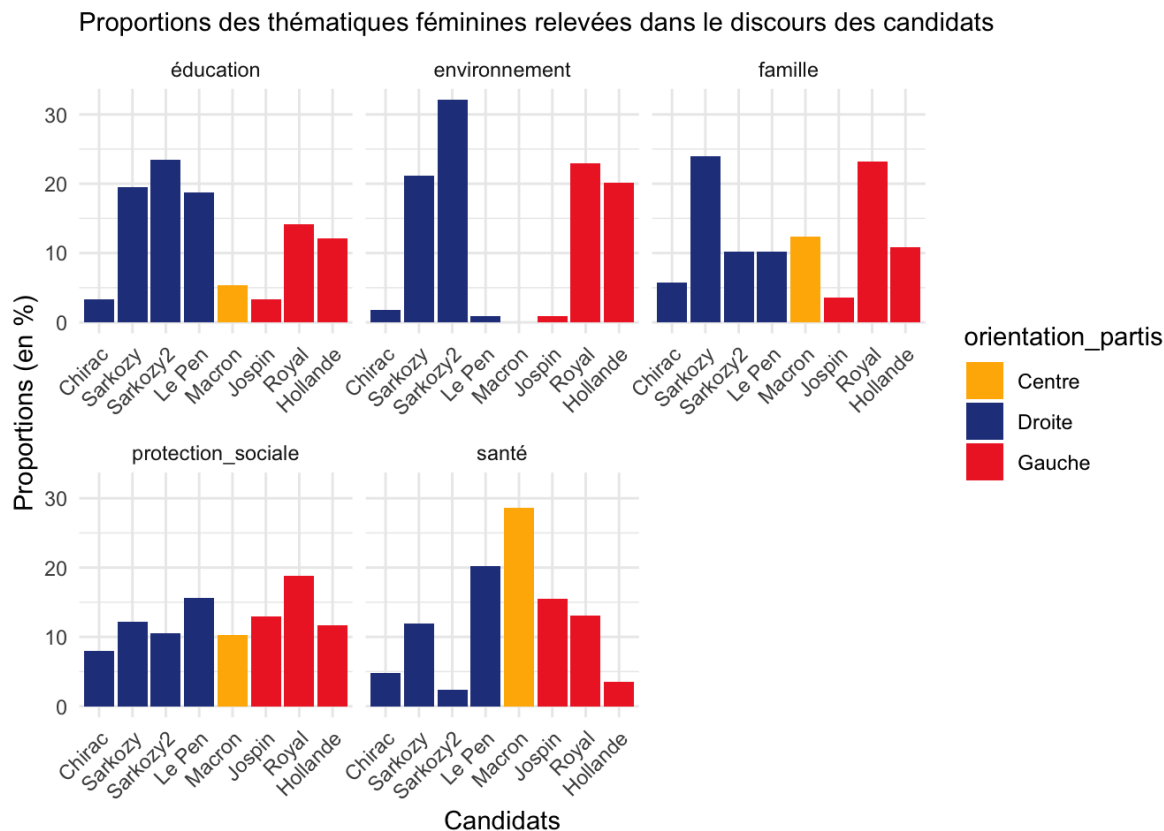
## 2) L'impact de l'orientation politique des candidats



Afin de contrôler pour l'orientation politique des candidats, nous avons commencé par additionner les fréquences relevées par candidat. Nous avons ensuite comparé la fréquence totale à laquelle chaque candidat a parlé des thématiques féminines dans leur ensemble. Les résultats sont présents au sein du graphique ci-dessus. Dans ce graphique, Sarkozy représente le candidat républicain lors de son débat face à Royal. Quant à Sarkozy2, il s'agit du candidat républicain lors de son débat face à Hollande. Nous pouvons constater que dans l'ensemble, des candidats de droite ont plus abordé les thématiques féminines que certains candidats de gauche. Par exemple, Le Pen et Sarkozy (lors de ses deux débats électoraux) semblent avoir davantage parlé des thématiques féminines que Jospin ou Hollande. À l'inverse certains candidats de gauche ont plus insisté sur ces thématiques que quelques candidats de droite. À titre d'exemple, Jospin, Hollande et Royal ont plus parlé de ces thématiques que Chirac. Royal a également plus abordé ces thématiques que Le Pen et que Sarkozy dans ses deux débats électoraux. En outre, le seul candidat centriste de notre recherche (Emmanuel Macron) a plus insisté sur ces thématiques que certains candidats de droite et de gauche. En revanche, nous n'avons qu'une observation pour l'orientation politique centriste. Or, cela est insuffisant pour tirer des conclusions sur les partis centristes. En somme, l'orientation politique ne semble pas affecter l'attention donnée

par les candidats aux thématiques féminines dans leur ensemble.

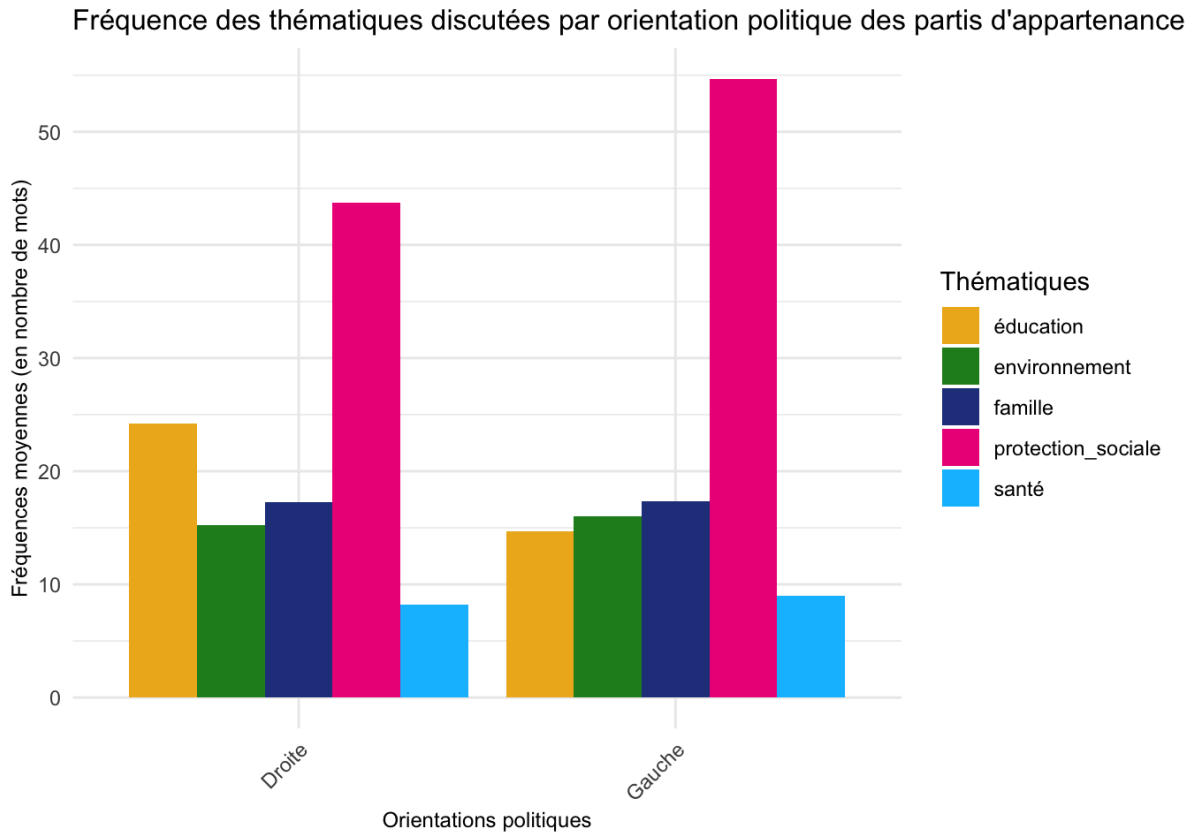
Si nous regardons plus en détail pour chaque thématique féminine, nous pouvons remarquer que, là-encore, l'orientation politique des candidats ne semble pas avoir d'impact. Les données sont présentes au sein du graphique ci-dessous.



Nous pouvons constater que des candidats de droite ont plus insisté sur chaque thématique, que certains candidats de gauche et vice-versa. Par exemple, Le Pen a plus abordé la thématique de la protection sociale que Jospin et Hollande. Ou encore, elle a davantage parlé de la santé et de l'éducation que l'ensemble des candidats de gauche. À l'inverse, Royal a par exemple plus parlé de la protection sociale que l'ensemble des candidats de droite. Elle a également davantage parlé de la thématique de la famille que Le Pen, Sarkozy (durant son débat face à Hollande), et Chirac. Ou encore elle a plus abordé la question de la santé que Sarkozy (durant ses deux débats) et Chirac.

Nous pouvons observer des tendances au sein de ces graphiques. En effet, visuellement, il semblerait que les candidats de droite insistent généralement plus que les candidats de gauche sur l'environnement et l'éducation. Quant aux candidats de gauche, il semblerait qu'ils insistent davantage que les candidats de droite sur la protection sociale.

Afin de vérifier ces observations, nous avons calculer les moyennes des candidats des droite et de gauche pour chaque thématique politique. Les résultats sont visibles au sein du graphique ci-dessous.

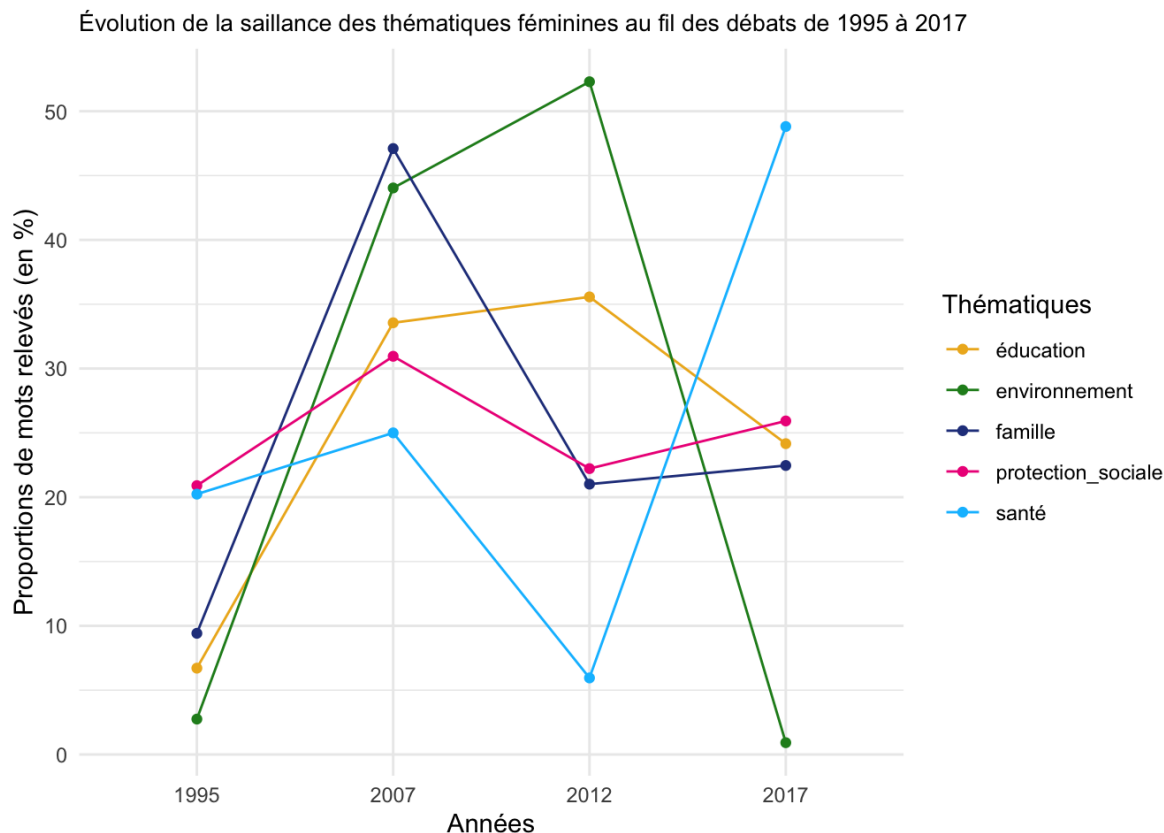


Les résultats indiquent, qu'en moyenne, les candidats de droite ont tendance à plus insister sur la thématique de l'éducation que les candidats de gauche. La moyenne concernant cette thématique s'élève à environ 15 mots pour la gauche, et à environ 24 mots pour la droite. Les candidats de gauche auraient, pour leur part, tendance à davantage insister sur la thématique de la protection sociale que les candidats de droite. En effet, la moyenne de cette thématique est d'environ 55 mots pour la gauche et d'environ 44 mots pour la droite. De légers écarts sont présents concernant les autres thématiques comme l'environnement, mais ces derniers ne sont pas suffisants pour conclure que les candidats de droite ou de gauche auraient plus ou moins abordé ces questions. En effet, la moyenne des fréquences pour l'environnement s'élève à environ 16 mots pour la gauche contre environ 15 mots pour la droite. Concernant la santé, la moyenne s'élève à 9 mots pour la gauche et à environ 8 mots pour la droite. Enfin, pour la famille, la moyenne est d'environ 17 mots pour la droite et pour la gauche.

Ainsi, l'orientation politique (le parti d'appartenance) pourrait expliquer le fait que certains candidats insistent plus que d'autres sur les thématiques féminines de l'éducation et de la

protection sociale. En revanche, cette variable ne peut pas expliquer le fait que certains candidats insistent plus que d'autres sur les thématiques politiques de l'environnement, de la santé, et de la famille. Il semblerait donc qu'une autre variable explicative impacterait la saillance des thématiques féminines dans le discours des candidats en débat électoral.

### 3) Les préoccupations de l'opinion publique, variable explicative principale ?



Nous pouvons remarquer, sur le graphique ci-dessus, que la saillance des thématiques politiques féminines semble varier en fonction de l'année du débat. Nous pouvons observer que ces thématiques politiques demeuraient peu abordées durant le débat électoral de 1995 opposant Jospin à Chirac. En revanche, elles sont devenues plus saillantes à partir du débat de 2007 (Royal contre Sarkozy). Les thématiques de l'environnement et de l'éducation ont augmenté en saillance jusqu'au débat de 2012 (Sarkozy contre Hollande). En revanche, l'environnement n'a pas été abordé durant le débat de 2017 (Le Pen contre Macron). L'éducation a été quant à elle moins abordée en 2017 par rapport à 2012. À l'inverse, la famille, la protection sociale et la santé ont été moins abordées en 2012 mais l'ont davantage été en 2017.



Ces résultats suggèrent que le contexte des élections aurait un impact sur la saillance des thématiques féminines lors des débats électoraux. Plus précisément, les préoccupations de l'opinion publique auraient influencé la saillance de ces thématiques. Or, cela semble plausible puisque les sondages d'opinion menés durant les élections concordent avec certains de nos résultats. Par exemple, lors de la campagne de 2017, 51% des Français percevaient la protection sociale comme un enjeu important, et 38% vouaient de l'importance à l'éducation (Statista 2017). Ou encore, en 2012, le financement du système de protection sociale était un enjeu important pour 31% des Français, et l'environnement demeurait important pour 16% des citoyens (OBS 2007). En 2007, l'éducation, l'environnement, et la protection sociale étaient désignés comme des préoccupations importantes (OpinionPublique 2012). En 1995, en revanche, l'environnement n'intéressait que 6% des Français (Hatchuel 2001).

## **Conclusion**

Pour conclure, nous avons, au travers de notre analyse de dictionnaire, constaté que les femmes politiques n'insistaient pas nécessairement plus que leurs adversaires masculins sur les thématiques féminines. Nous avons également relevé que la présence d'une femme au sein d'un débat électoral ne signifie pas que les thématiques politiques féminines sont davantage abordées que lorsque deux hommes s'affrontent. Ainsi, nos deux hypothèses de départ ont été infirmées. Nos résultats ont remis en question les postulats des recherches antérieures. La thèse selon laquelle "les femmes aborderaient davantage les thématiques féminines que les hommes durant les campagnes électorales" ne s'est pas avérée dans notre étude. Nos résultats nous ont également indiqué que l'orientation politique des candidats ne pouvait pas expliquer à elle seule la saillance des thématiques politiques féminines dans le discours des candidats. En revanche, une piste pour une prochaine recherche serait d'analyser les effets des préoccupations de l'opinion publique sur la saillance des thématiques politiques féminines en débat électoral. Car, comme nous l'avons montré, la saillance de ces thématiques dans les débats électoraux varie en fonction des années. Cela permettrait de combler les limites de notre étude. D'ailleurs, notre recherche peut poser certains enjeux éthiques. En effet, comme nous l'avons expliqué précédemment, nous avons converti les retranscriptions pdf en bases de données en faisant du web scraping. En revanche, bien que les retranscriptions pdf étaient en libre accès, nous n'avons pas eu l'autorisation du site de transformer ces fichiers en bases de données. Ainsi, les droits d'auteurs du site pourraient être bafoués par nos manipulations de données (Alexander 2023). Par conséquent, notre recherche pousse à remettre en question les aspects légaux et éthiques de la collecte et de l'utilisation de données issues de sources publiques.

## **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](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>

Hatchuel, Georges. 2001. « Dans l’opinion, une préoccupation généralisée depuis dix ans ». Dans *L’Environnement, question sociale*, 29-37. Hors collection. Paris: Odile Jacob. <https://doi.org/10.3917/oj.roche.2001.01.0029>.

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>.

OBS. 2007. « Les préoccupations des Français pour 2007 ». Le Nouvel Obs, 6 janvier 2007. <https://www.nouvelobs.com/politique/elections-2007/20070106.OBS5859/les-preoccupations-des-francais-pour-2007.html>.

OpinionPublique. 2012. « 2007-2012: les préoccupations des Français ont bien changé ». *Opinion publique*, 13 février 2012. <https://opinionpublique.wordpress.com/2012/02/13/2007-2012-les-preoccupations-des-francais-ont-bien-change/>.

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>.

Statista. 2017. « Infographie: L'emploi, première préoccupation des électeurs français ». Statista Daily Data, 20 mars 2017. <https://fr.statista.com/infographie/8563/lemploi-premiere-preoccupation-des-electeurs-francais>.

## Annexe

```
# Codes de monsieur Nadjim Fréchet afin de convertir le texte pdf en base de données csv :
```

```
## Libraries
```

```
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)

## Data

text_1 <- suppressMessages(pdftools::pdf_text("/Users/oliviasaffioti/Desktop/fas_1001_Saffioti.pdf"))

## Settings

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

## Nettoyage de code

## Chercher les nom des personnes concernées dans le débat

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)

## Données de prises de parole (text data)

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))

## Fusion des deux variables clés et création de la base de données finale

Data_debate_2017 <- bind_cols(speaker_names, text_data) |>
  ## Creation de variables et nettoyage ##
mutate(year      = 2017,
       date       = ymd("2017-05-03"),
       country    = "France",
       id         = "Macron_LePen",
       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)

##Sauvegarde des données

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

# Importer les bases de données

debat_royal_sarkozy <- read.csv("debat_sarkozy_royal.csv")
debat_macron_lepen <- read.csv("debat_macron_lepen.csv")
debat_chirac_jospin <- read.csv("debat_chirac_jospin.csv")
debat_sarkozy_hollande <- read.csv("debat_Sarkozy_Hollande.csv")

# Libraries

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 kind.

```
library(tidyverse)
library(patchwork)

# Créer le dictionnaire

Dictionnaire <- list (Santé = c("santé", "médecin", "médecins", "hôpital", "hôpitaux", "ma

# Nettoyer le texte : enlever la ponctuation et les majuscules

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

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

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

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

# Créer de nouvelles bases de données en supprimant les variables inutiles

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

debat_sarkozy_hollande <- debat_sarkozy_hollande %>% select(speaker, text, id) %>% filter(s

debat_royal_sarkozy <- debat_royal_sarkozy %>% select(speaker, text, id) %>% filter(speake

debat_chirac_jospin <- debat_chirac_jospin %>% select(speaker, text, id) %>% filter(speake

# Renommer l'objet de la variable "id"
```

```

debat_macron_lepen <- debat_macron_lepen %>% mutate(id = ifelse(id == "Débat présidentiel", "Macron", "Lepen"))
debat_royal_sarkozy <- debat_royal_sarkozy %>% mutate(id = ifelse(id == "Débat présidentiel", "Ségol", "Royal"))
# Renommer les noms des candidats
debat_royal_sarkozy <- debat_royal_sarkozy %>% mutate(speaker = case_when(speaker == "Ségol" ~ "Ségol", speaker == "Royal" ~ "Royal"))
debat_sarkozy_hollande <- debat_sarkozy_hollande %>% mutate(speaker = case_when(speaker == "Sarkozy" ~ "Sarkozy", speaker == "Hollande" ~ "Hollande"))
# Faire l'analyse de dictionnaire sur les 4 bases de données
debat_macron_lepen_resultats <- run_dictionary(data = debat_macron_lepen, text = text, dictionary = dictionnaire)

```

100% expressions/words found

0.116 sec elapsed

```

donnees_graphique_1 <- debat_macron_lepen_resultats %>% group_by(speaker, id) %>% summarise(n = n())

```

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

```

debat_royal_sarkozy_resultats <- run_dictionary(data = debat_royal_sarkozy, text = text, dictionary = dictionnaire)

```

100% expressions/words found

0.036 sec elapsed

```

donnees_graphique_2 <- debat_royal_sarkozy_resultats %>% group_by(speaker, id) %>% summarise(n = n())

```

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

```

debat_chirac_jospin_resultats <- run_dictionary(data = debat_chirac_jospin, text = text, dictionary = dictionnaire)

```

100% expressions/words found

0.03 sec elapsed

```
donnees_graphique_3 <- debat_chirac_jospin_resultats %>% group_by(speaker, id) %>% summarise()
```

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

```
debat_sarkozy_hollande_resultats <- run_dictionary(data = debat_sarkozy_hollande, text = t
```

100% expressions/words found

0.034 sec elapsed

```
donnees_graphique_4 <- debat_sarkozy_hollande_resultats %>% group_by(speaker, id) %>% summarise()
```

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

```
# Convertir les résultats de chaque base de données individuellement au format long  
format_long_1 <- tidyr::pivot_longer(donnees_graphique_1, cols = -c(speaker, id), names_to = "Thématiques")  
format_long_2 <- tidyr::pivot_longer(donnees_graphique_2, cols = -c(speaker, id), names_to = "Thématiques")  
format_long_3 <- tidyr::pivot_longer(donnees_graphique_3, cols = -c(speaker, id), names_to = "Thématiques")  
format_long_4 <- tidyr::pivot_longer(donnees_graphique_4, cols = -c(speaker, id), names_to = "Thématiques")  
  
# Convertir les fréquences en pourcentage  
format_long_1 <- format_long_1 %>% group_by(Thématiques) %>% mutate(Percent = Fréquences / sum(Fréquences))  
format_long_2 <- format_long_2 %>% group_by(Thématiques) %>% mutate(Percent = Fréquences / sum(Fréquences))
```



```

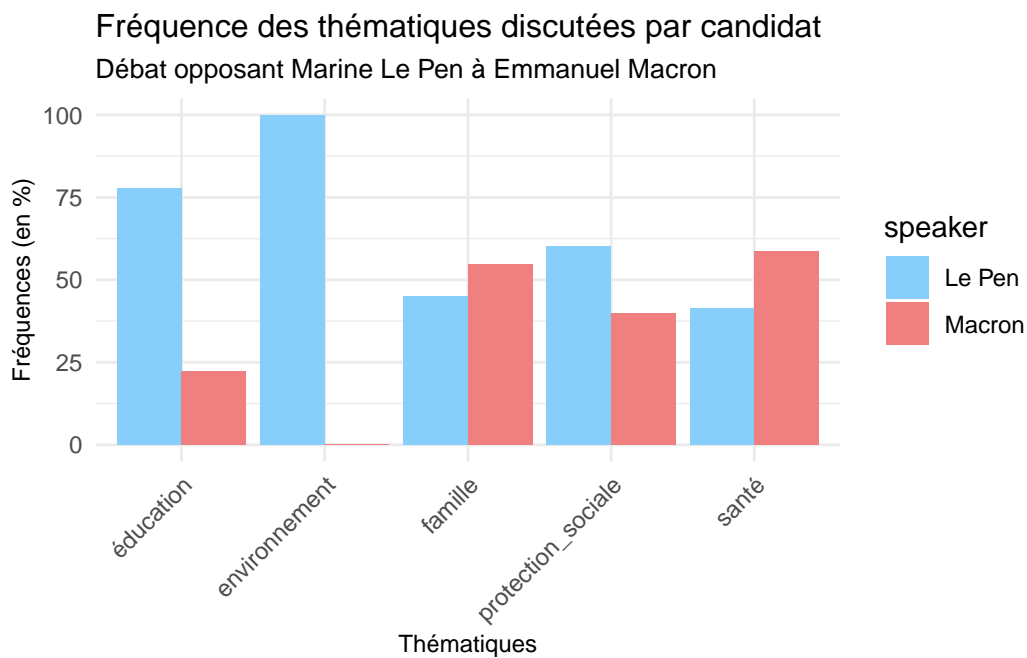
format_long_3 <- format_long_3 %>%group_by(Thématiques) %>% mutate(Percent = Fréquences /
format_long_4 <- format_long_4 %>%group_by(Thématiques) %>% mutate(Percent = Fréquences /
donnees_finales <- bind_rows(format_long_2, format_long_1, format_long_4, format_long_3)

# Graphiques individuels

Graphique_1 <- ggplot(format_long_1, aes(x = Thématiques, y = Percent, fill = speaker)) +

Graphique_1

```

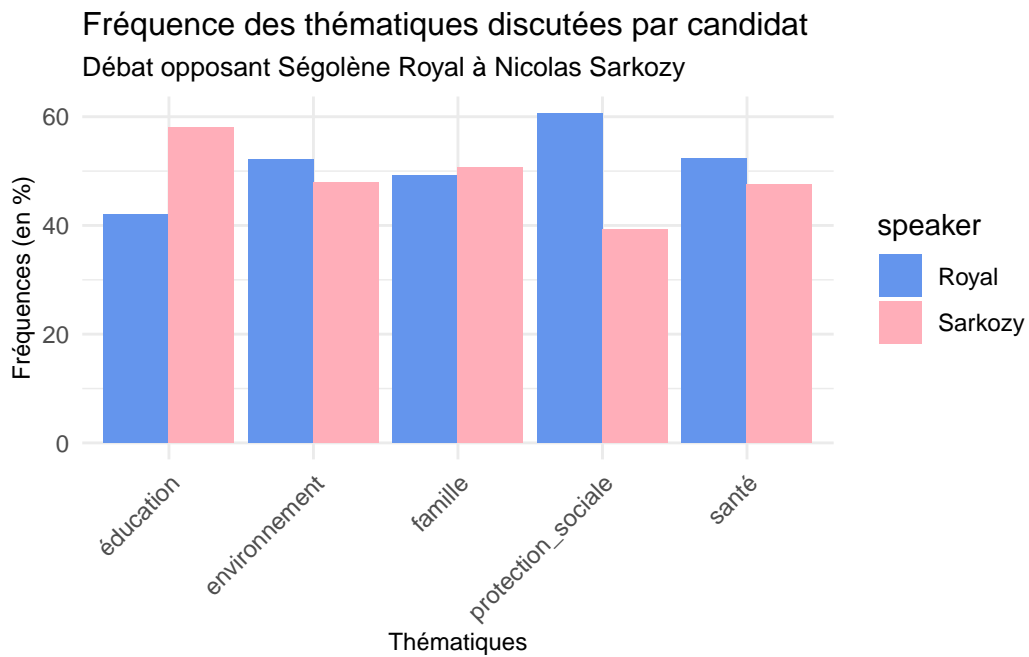


```

Graphique_2 <- ggplot(format_long_2, aes(x = Thématiques, y = Percent, fill = speaker)) +

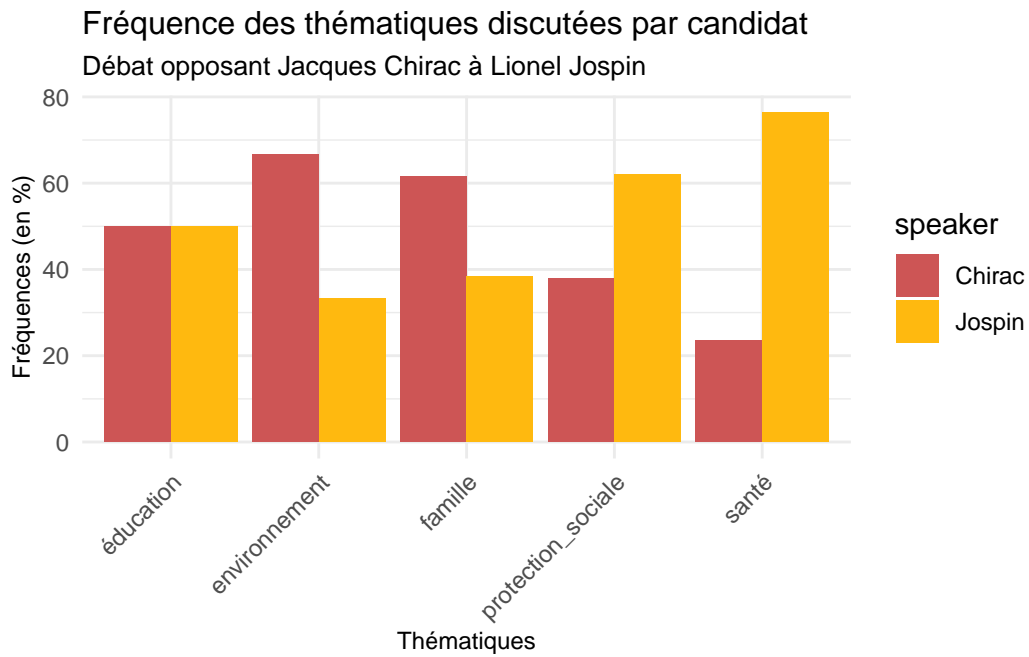
Graphique_2

```



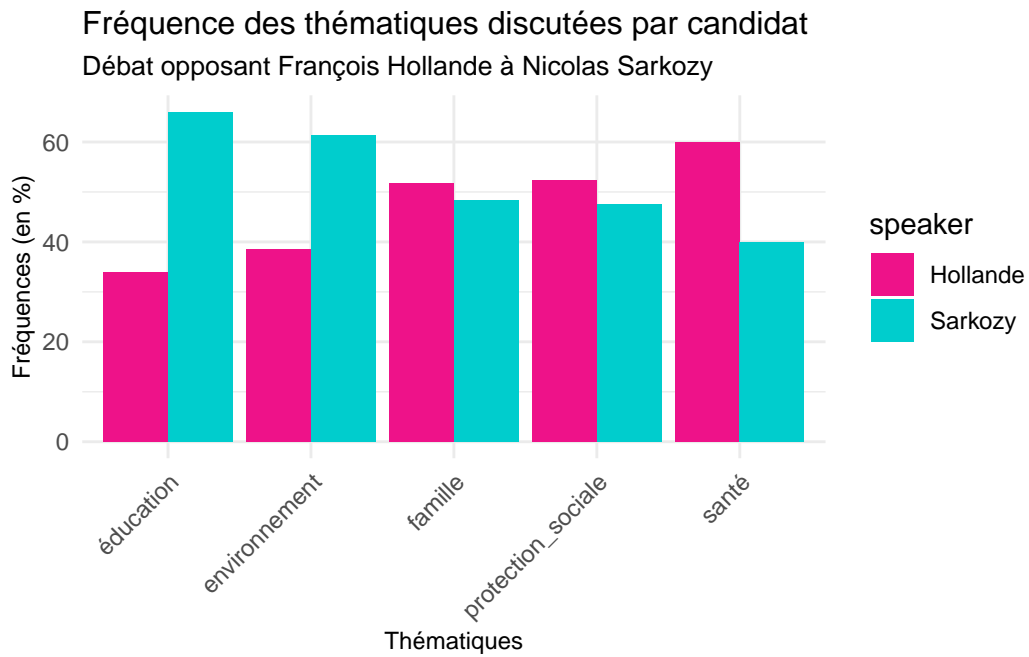
```
Graphique_3 <- ggplot(format_long_3, aes(x = Thématiques, y = Percent, fill = speaker)) +
```

```
Graphique_3
```



```
Graphique_4 <- ggplot(format_long_4, aes(x = Thématiques, y = Percent, fill = speaker)) +
```

Graphique\_4



```
# Graphiques regroupés avec facet_wrap
```

```
facet_graph <- ggplot(donnees_finales, aes(x = Thématiques, y = Percent, fill = speaker))
```

facet\_graph

## Proportions des thématiques discutées par candidat dans les débats



# Comparer les résultats entre les débats généralement

```
donnees_fin <- donnees_finales %>% group_by(Thématiques) %>% mutate(somme_totale = sum(Fré
```

```
donnees_fin <- donnees_fin %>% mutate(Fréquences_pourcentage = (Fréquences / somme_totale)
```

```
donnees_fin2 <- donnees_fin %>% group_by(id, Thématiques) %>% summarize(Somme_Fréquences =
```

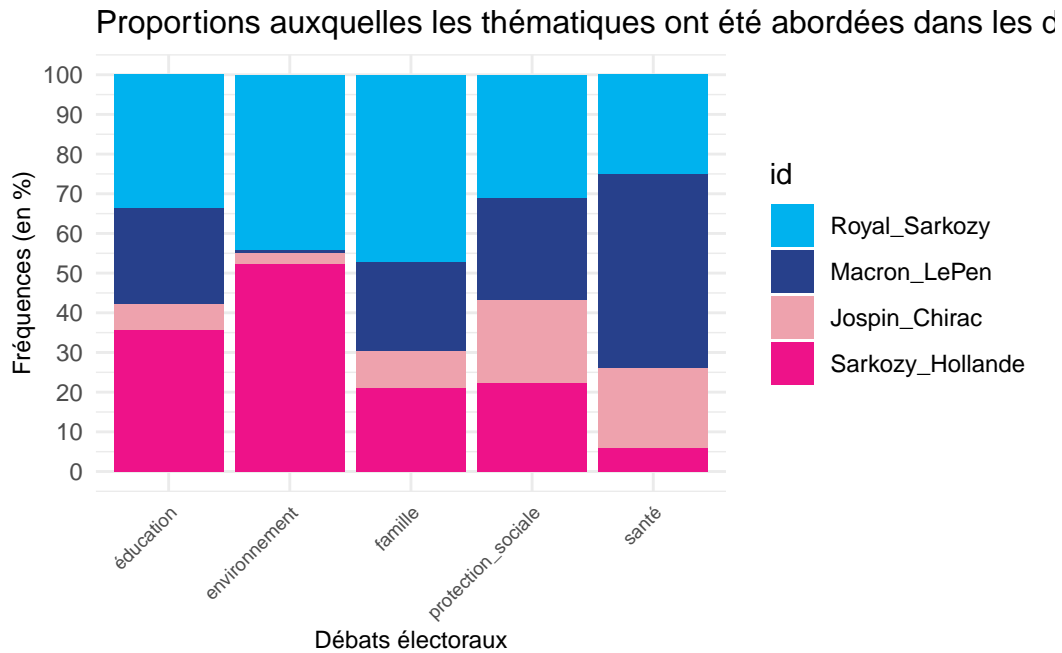
`summarise()` has grouped output by 'id'. You can override using the `.groups` argument.

```
donnees_fin2$id <- factor(donnees_fin2$id, levels = c("Royal_Sarkozy", "Macron_LePen", "Jo
levels(donnees_fin2$id)
```

```
[1] "Royal_Sarkozy"      "Macron_LePen"      "Jospin_Chirac"     "Sarkozy_Hollande"
```

```
Graph_final_2 <- ggplot(donnees_fin2, aes(x = Thématiques , y = Somme_Fréquences, fill = i
```

```
Graph_final_2
```



```
#Contrôler pour le parti politique d'appartenance
```

```
donnees_fin <- donnees_fin %>% mutate(speaker = ifelse(row_number() %in% c(26:30), "Sarkoz
```

```
donnees_fin <- donnees_fin %>% mutate(orientation_partis = case_when(speaker == "Le Pen"~
```

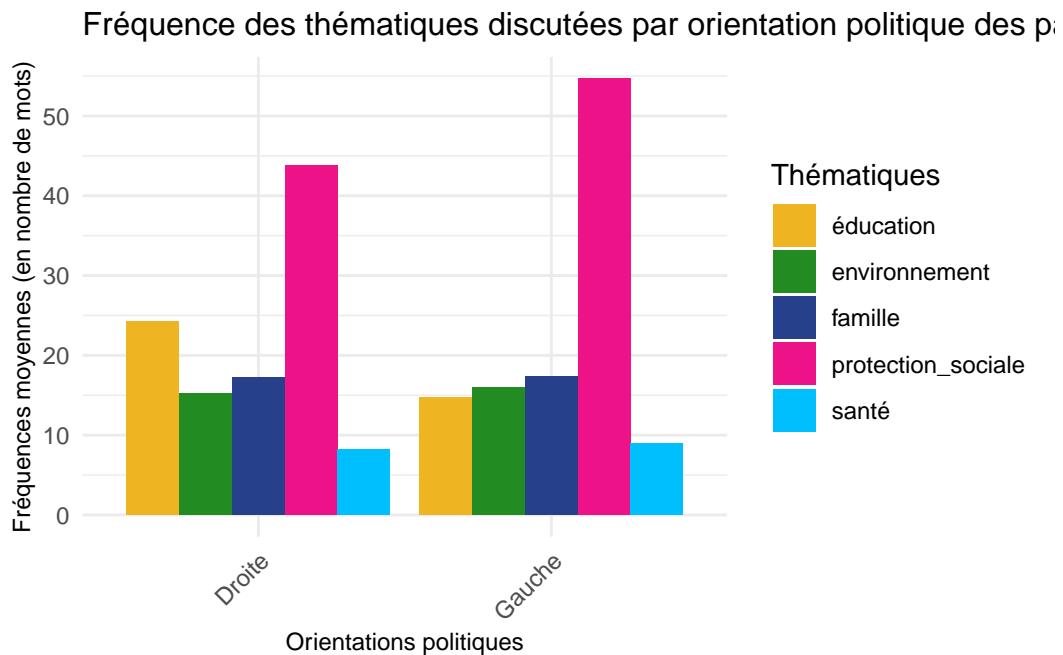
```
data_partis <- donnees_fin %>% group_by(orientation_partis, Thématiques) %>% summarize(moy
```

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

```
data_partis <- data_partis[-c(1, 2, 3, 4, 5), ]
```

```
Graph_partis <- ggplot(data_partis, aes(x = orientation_partis, y = moyenne_fréquences, fi
```

```
Graph_partis
```



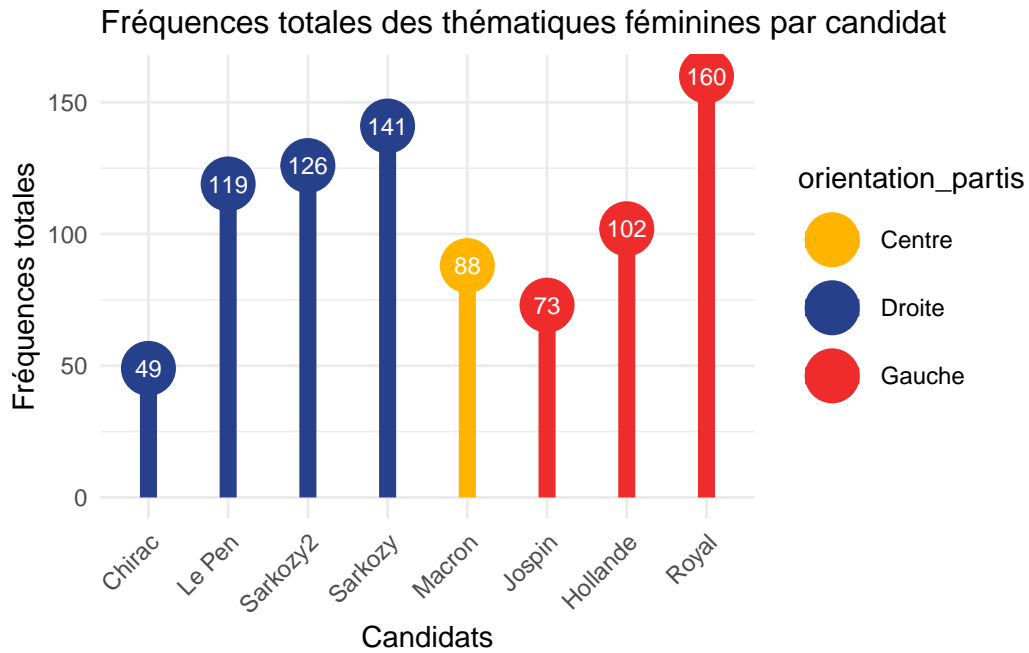
```
candidats_dg <- donnees_fin %>% group_by(speaker, orientation_partis) %>% summarise(Fréque
```

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

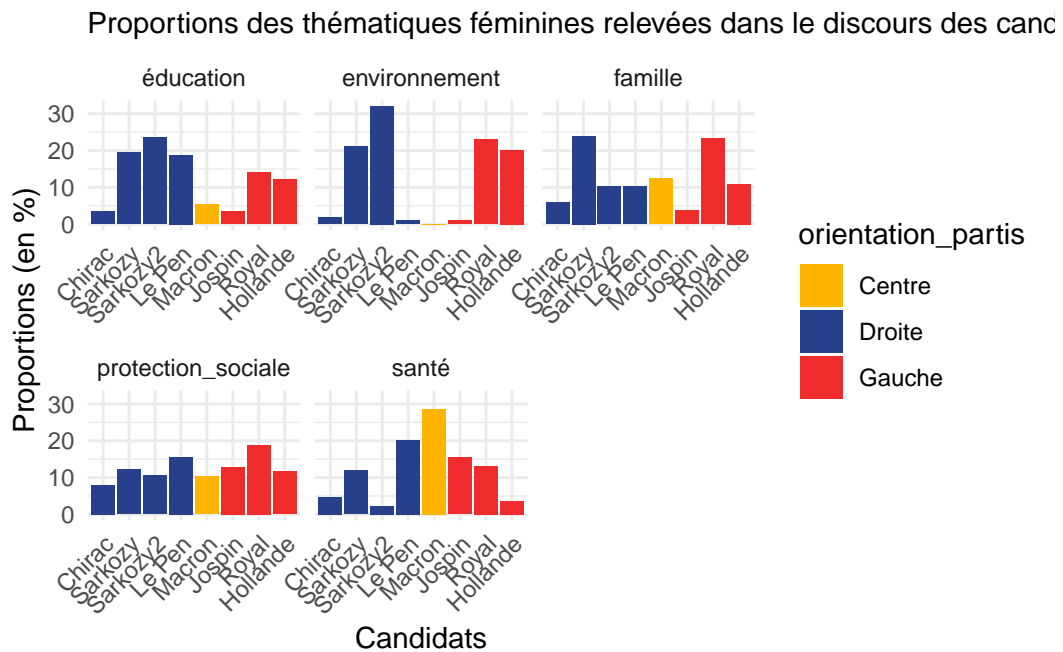
```
graph_orientation_pol_candidats <- ggplot(candidats_dg, aes(x = factor(speaker, levels = c
```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
i Please use `linewidth` instead.

```
graph_orientation_pol_candidats
```



```
graph_orient_pol_2 <- ggplot(donnees_fin, aes(x = factor(speaker, levels = c("Chirac", "Sa
graph_orient_pol_2
```



```
# Contrôler pour le contexte
```

```
contexte <- donnees_fin %>% group_by(id, Thématiques) %>% summarise(Fréquences_totales = s
```

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

```
contexte <- contexte %>% mutate(year = case_when(id == "Royal_Sarkozy"~"2007", id == "Macr
```

```
contexte <- pivot_wider(contexte, id_cols = c(id, year), names_from = Thématiques, values_
```

```
contexte <- tidyr::pivot_longer(contexte, cols = c(environnement, famille, protection_soci
```

```
contexte <- contexte %>% group_by(Thématiques) %>% mutate(Percent = Fréquences / sum(Fréqu
```

```
graph_contexte <- ggplot(contexte, aes(x = year, y = Percent, group = Thématiques, color =
```

```
graph_contexte
```

