

# LDApol: vers une méthodologie de contextualisation des discours politiques

Jeanne Vermeirsche, Eric SanJuan, Tania Jiménez

27 juin 2022



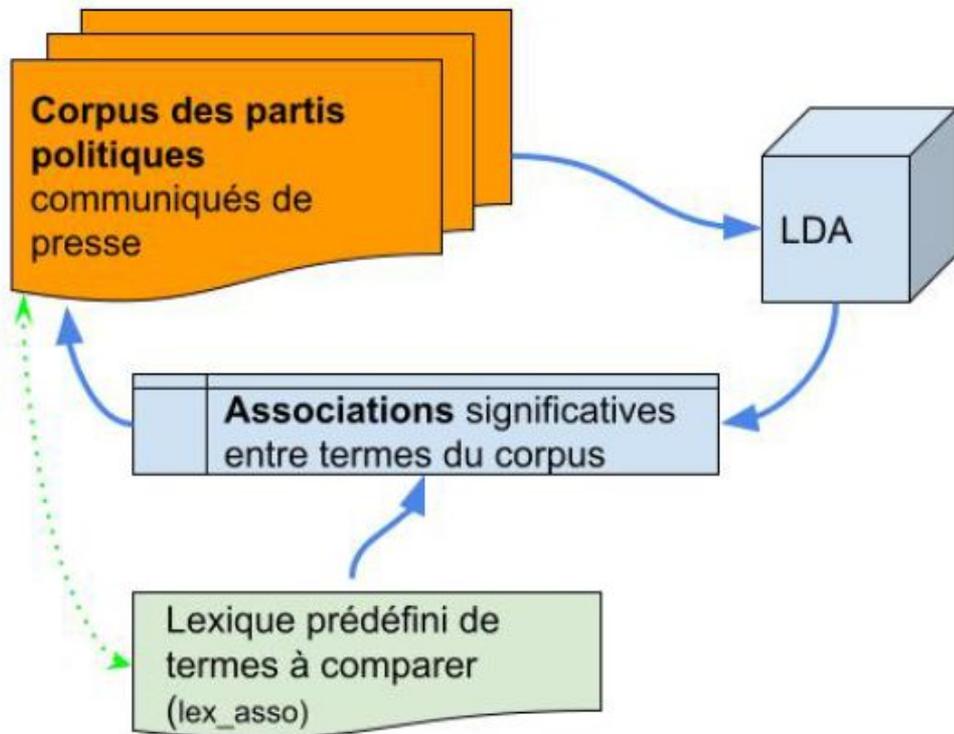
## Introduction et contexte

- Langage politique et langue des partis politiques → polarisation du débat politique
  - Objectiver les associations de termes caractéristiques d'un parti ou d'une personnalité politique
- ➔ Mise en place d'une méthodologie
- Le nationalisme dans le discours des partis politiques
  - Mettre en lumière un vocabulaire et des éléments de langage associés au nationalisme, et leur diffusion dans les discours des partis étudiés

# Méthodologie

Ce schéma représente notre **travail sur le corpus de discours politiques**

- 1) Constitution d'un corpus de textes
- 2) Définition d'un lexique "lex\_asso"
- 3) Calcul de modèles statistiques
- 4) Extraction d'associations entre termes



# Modèle (Données)

**Hypothèse** : Les termes à inclure dans un communiqué de presse sont choisis selon un agenda politique interne ciblant des groupes spécifiques d'électeurs. Les CPs sont courts et s'appuient souvent sur un nombre réduit d'affirmations. Ils sont des mises en forme spécifiques du discours politique (Bourdieu, 1981). Ils suivent un ordre établi rigoureux où chaque terme à son importance, où les sujets et les concepts ont été préalablement pensés et discutés avec des conseillers politiques

**Approche** : Exploration d'ensembles de termes significativement corrélés sur la base d'un modèle de langage génératif et de tests statistiques.

**Représentation en sac de mots** où nous avons choisi d'ignorer l'ordre des mots dans les CP politiques et de nous concentrer sur les co-occurrences. Ces hypothèses nous permettent d'appliquer le modèle génératif d'allocation Dirichlet latente introduit par (Blei et al. , 2003) sur le concept d'échangeabilité mis en exergue par De Finetti (Cifarelli Regazzini, 1996)

# Modèle (Paramètres et comparaisons)

Le LDA est un modèle de plongement lexical sur  $k$  dimensions dans  $[0,1]$

- les termes sont représentés par des séquences de  $k$  probabilités indépendantes
- on optimise le calcul des probabilités conjointes d'ensemble de termes
- comme pour le k-means,  $k$  nécessite d'être fixé a priori
- On peut comparer ce modèle au:
  - **LSA**: analyse factorielle de la matrice des cooccurrences des termes dans les documents où on projette sur les  $k$  premiers vecteurs propres,  $k$  étant la dimension de ce plongement.
  - **Word Embedding**: plongements lexicaux sur  $k$  dimensions qui optimise la prédiction du mot suivant à partir des  $n$  mots précédents (deux paramètres à établir ici).

Pour la suite nous choisissons  $k=12$  par défaut pour le LDA,  $k_{ref}=16$  et  $k_{max}=24$  pour vérification de la stabilité.  $k=30$  pour LSA et  $k=25$  pour WE

## Modèle (Calculs)

- **k multinomiales**  $\beta_i$  calculées par le modèle génératif LDA (VEM)
- **Test de normalité** de Kolmogorov-Smirnov : filtrage des représentations  $v(\omega)$
- **Tests paramétriques de corrélation** entre représentations.
- Comparaison avec une extraction du **wikipedia** français à partir du lexique

Corpus	Docs	Longueur	Omega	beta	lex_asso
FN/RN	2846	364008	9814	9581	91
LR	398	24882	1587	194	4
LREM	269	33363	1293	379	4
PS	906	95624	4065	906	15
FI	511	49971	2678	545	7
WP	1933	100949	3995	3949	97



# Résultats (Évaluation)

## **Classement des corrélations :**

0 - non valide à la lecture du contexte des communiqués

1 - valide mais relation évidente (grand nombre de cooccurrences directes dans le texte)

2 - valide mais indirecte (peu de cooccurrences directes) - ne peut être mise en évidence sans modèle

➔ Explorer le contexte pour chacune de ces associations pour valider ou non une relation de causalité dans l'usage de ces termes.

# Résultats (Analyse)

## Exemple 1 : civilisation - islam (corpus RN)



Mise en contexte :

- classement = 2 - mise en évidence par le modèle
- corrélation positive pour chacun des modèles testés
- contexte partagé : vocabulaire autour de l'islam, dénonciation de l'islamisme et des islamistes (attentats, immigration massive...), du port du voile (détournement de la défense des droits des femmes)

# Résultats (Contexte)

## Exemple 1 : civilisation - islam (corpus RN)

Détournement de la thématique “défense des droits des femmes” = dénonciation de l’islam

le niqab est à ses yeux une forme d'oppression contre les femmes

société elles ne feront pas le poids face au grand aveuglement des collectifs « néoféministes » à l'égard de l'immigration massive et de son corollaire l'islam radical toutes les données statistiques sont à ce sujet édifiantes selon une étude de l'observatoire national de la délinquance et des réponses pénales on drp 52% des viols commis à paris l'ont été par des étrangers ce qui rappelle les tristes viols et agressions sexuelles commis à cologne le 31 décembre 2015 après lesquels près de 600 femmes avaient déposé plainte \n plus inquiétant encore un sondage réalisé par l'ifop que le magazine elle après l'avoir pourtant commandé a finalement renoncé à publier démontre que le multiculturalisme et notamment la progression du rigorisme islamique dans notre société nuit gravement aux droits des femmes par exemple 27% des hommes de confession musulmane déclarent avoir effectué une pratique sexuelle sans le consentement du partenaire féminin contre 13% pour l'ensemble des français 14 points de pourcentage de plus !\n « couvrez cette pancarte que je ne saurai voir »

famille à diriger un parti politique marine le pen sera toujours aux côtés des femmes qui aspirent à reconquérir leur liberté en brisant les chaînes de ces ghettos culturels l'élection présidentielle de constitue au delà d'un choix politique un enjeu de civilisation

dans son journal télévisé du décembre france télévisions a diffusé un reportage alarmant sur ces quartiers de banlieues où les droits des femmes s'effondrent sous le poids de cultures étrangères rétrogrades adaptation de leur tenue ou de leur maquillage violence verbale asservissement et contrôle social dans l'indifférence générale et en premier lieu de celui des pouvoirs publics les femmes subissent un recul des libertés fondamentales se voyant contraintes de déserrer l'espace public chaque jour un peu plus l'islam est le code pénal

# Résultats (Contexte)

## Exemple 2 : souveraineté - france (corpus PS)



Mise en contexte :

- classement = 2 - mise en évidence par le modèle
- le PS retourne l'argument de la perte de souveraineté de la France, émis par le FN/RN notamment
- accuse également LREM / Macron d'être en contradiction avec ses annonces dans les politiques menées

# Conclusions et Travaux Futurs

Méthodologie pour révéler des associations indirectes entre termes employés dans une communication politique :

- recherche de corrélations
- analyse comparative entre sous-corpus (partis politiques / personnalités politiques)

→ Prolongement des travaux :

- objectiver la proximité des thématiques sur des corpus de textes différents
- circulation au sein de l'espace public

★ Wikipédia



Jeanne Vermeirsche  
jeanne.vermeirsche@univ-avignon.fr

Eric Sanjuan  
eric.sanjuan@univ-avignon.fr

Tania Jiménez  
tania.jimenez@univ-avignon.fr