



La data intelligence au service de l'analyse des discours des principaux candidats à l'élection présidentielle

ou comment l'IA peut nous aider à mieux comprendre
les propositions et les valeurs des principaux candidats ?

selon **Acadys – Beyond Data Sciences**

Date : 2 Mars 2022

Objectif : utiliser la puissance de l'intelligence artificielle pour disséquer les discours et programmes politiques des principaux candidats à l'élection présidentielle française qui aura lieu dans 40 jours afin d'identifier les thématiques communes mais aussi leurs spécificités.

Candidats : les 6 principaux hors Emmanuel Macron pour cette première analyse. Anne Hidalgo, Eric Jadot, Marine Le Pen, Jean-Luc Mélenchon, Valérie Pécresse, Eric Zemmour

Méthode : nous avons identifié dans un premier temps l'ensemble des sources officielles des candidats déclarés qu'elles soient orales ou écrites pour les injecter dans un data lake en utilisant les techniques du big data. Les discours ont été « scriptés » au travers de méthodes de type NLP (Natural Language Processing) afin de pouvoir être comparés aux textes écrits. Cela inclut en particulier des techniques de reconnaissance vocale pour transcrire les discours enregistrés en texte exploitable. Des algorithmes d'analyse de textes de type Machine learning ont été entraînés afin d'extraire les termes et thématiques les plus fréquents (simples 1-gram et composés 2-gram) sur une base de 10000 mots. Un traitement préliminaire a été effectué pour « nettoyer » les corpus (pre-processing). Une analyse vectorielle utilisant la distance euclidienne permet de mieux comprendre le niveau de « proximité » entre les différents candidats.

Sources : seules les sources émanant directement des candidats : programmes officiels, sites web officiels, publications sur les réseaux sociaux par le candidat lui-même, discours politiques lors de meetings officiels de campagne ont été considérées. Les interviews pouvant être biaisées par les types de questions posées par les journalistes ou les informations reprises par les médias n'ont pas été considérées dans cette analyse.

Premiers résultats :

A. Termes les plus employés :

Analyses individuelles :

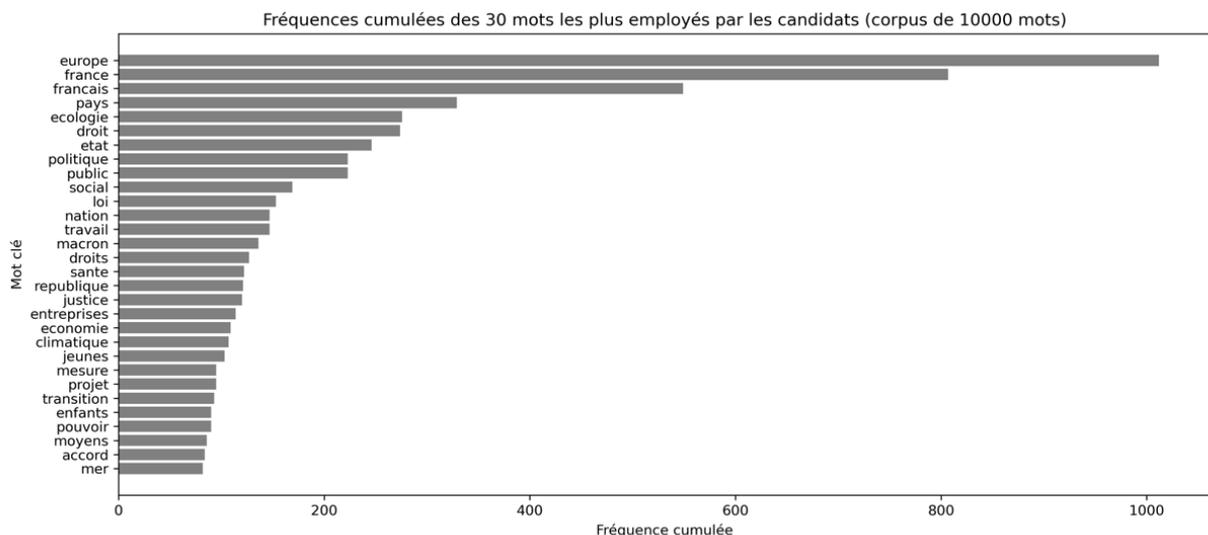
Par candidat, les termes **France** et **Europe** sont les termes de loin les plus cités et figurent systématiquement dans le Top 5 de chacun des candidats hormis pour Marine Le Pen où l'Europe n'arrive qu'en 9^{ème} position. Le terme **français** apparaît dans le top 10 de chaque candidat sauf pour Y. Jadot où il apparaît au-delà du 20^{ème} rang.

Les termes **état**, **pays** et **politique** sont cités par tous les candidats parmi leurs 30 mots les plus utilisés.

Le terme **pouvoir d'achat** est cité fréquemment par Valérie Pécresse, mais aussi par Eric Zemmour et Marine Le Pen...

Analyse en fréquences cumulées :

En cumulant les fréquences, les 30 mots les plus employés par les 6 candidats (cumul des fréquences relativement à un corpus de 10 000 mots) sont représentés par le graphique ci-dessous :



En fréquence cumulée, les mots **Europe**, **France** et **français** sont de loin les plus employés.

Les autres termes les plus cités sont : **pays**, **écologie**, **droit**, **état**, **politique**, **public** puis viennent ensuite les termes **social**, **loi**, **nation**, **travail**, le président E. **Macron**, **droits**, **santé**, **république**, **justice**, **entreprises**, **économie**, **climatique**.

B. Termes spécifiques aux candidats

Dans les termes (mots simples 1-grams) utilisés par les candidats parmi les 30 mots les plus utilisés, certains sont spécifiques à chaque candidat :

Anne Hidalgo	Eric Jadot	Marine Le Pen	Jean-Luc Mélenchon	Valérie Pécresse	Eric Zemmour
Lutte	Accord	Constitution	Eau	Liberté	Argent
Formation	Respect	Territoire	Parlementaire	Ordre	Peuple
Parlement	Énergies	Article	Garantir	Besoin	Achat
Collectivités	Échange	Tourisme	Assemblée	Familles	Reconquête
Citoyens	Armes	Situation	Marche	Territoires	Vivre
Biodiversité	Règles	Mesures	Entreprise	Parents	Gauche
Ensemble	Produits	Identité	Proposition	Force	Campagne
Acteurs	Application	Conditions	Députés	Retraites	Élèves
Service	Véhicules	Dispositions	Concrètement	Famille	Terre
Éducation	Humaniste	Étranger	Instaurer		Défendre

Paris, le 02 mars 2022

Dans les mots simples (1-grams) utilisés par les candidats parmi les 30 mots les plus utilisés, certains sont spécifiques à un ou deux candidats :

Anne Hidalgo	Eric Jadot	Marine Le Pen	Jean-Luc Mélenchon	Valérie Pécresse	Eric Zemmour
Climatique	Accord	Étrangers	Mesure	Enfants	Histoire
Lutte	Transition	Constitution	Eau	École	Enfants
Transition	Respect	Immigration	Parlementaire	Liberté	Argent
Formation	Climatique	Mer	Mer	Ordre	Peuple
Parlement	Gouvernement	Territoire	Commission	Besoin	Puissance
Collectivités	Énergies	Article	Gouvernement	Femmes	Défense
Citoyens	Échange	Tourisme	Garantir	Président	Impôts
Biodiversité	Armes	Sécurité	Salariés	Protéger	Président
Ensemble	Règles	Défense	Marche	Familles	Protéger
Acteurs	Produits	Situation	Assemblée	Impôts	Reconquête

Concernant les groupements de deux mots (2-grams), voici ceux qui sont les plus fréquents.

Remarque : au cours du nettoyage du texte, certains mots sont supprimés pour l'analyse comme les mots de liaison. Ainsi, par exemple, « Président de la République » ressortira en « Président République », « Général de Gaulle » en « Général Gaulle ».

Anne Hidalgo	Éric Jadot	Marine Le Pen	Jean-Luc Mélenchon	Valérie Pécresse	Eric Zemmour
Transition écologique	Union européenne	Projet loi	France insoumise	Pouvoir achat	Pouvoir achat
Collectivités territoriales	État droit	Article constitution	Assemblée nationale	Forces ordre	France rurale
Service public	Droit européen	Union européenne	Bifurcation écologique	Président République	Peuple français
Union européenne	Présidence française	Identité immigration	Union européenne	Politique familiale	Argent public
Santé publique	Libre échange	Nationalité française	Crise sanitaire	États Unis	République France
Climat biodiversité	Transition écologique	Territoire national	Commission enquête	Éric Zemmour	Président république
Réchauffement climatique	Commission européenne	Immigration projet	Proposition loi	Collectivités locales	Général Gaulle
Énergies renouvelables	Énergies fossiles	Citoyenneté identité	Services publics	Ile France	Déserts médicaux
Services publics	Protection environnement	Droit asile	Députés insoumis	Situation handicap	États Unis
Femmes Hommes	Règles européennes	France puissance	Groupe parlementaire	Histoire France	Industrie défense

C. Niveau de proximité entre candidats

Une analyse de similitude a été réalisée pour identifier le niveau de proximité entre les candidats.

Dans la dernière sortie de l'algorithme, sur les $6 \times 30 = 180$ mots potentiels, il s'avère qu'il y a 92 mots uniques (soit à peu près la moitié donc). Chaque candidat a donc été représenté par un vecteur de 92 coordonnées dont les valeurs correspondent à la fréquence des mots.

En utilisant la distance euclidienne, voici les résultats sous forme d'une matrice croisée de distances :

candidat	pecresse	hidalgo	jadot	lepen	melenchon	zemmour
pecresse	0	268,6	686	255,8	255,5	155,4
hidalgo	268,6	0	640	271,1	176,1	264,5
jadot	686	640	0	704,2	635,5	703,7
lepen	255,8	271,1	704,2	0	252,2	248,8
melenchon	255,5	176,1	635,5	252,2	0	259
zemmour	155,4	264,5	703,7	248,8	259	0

On constate avec ce modèle, que le candidat le plus proche de Zemmour est Pécresse ...

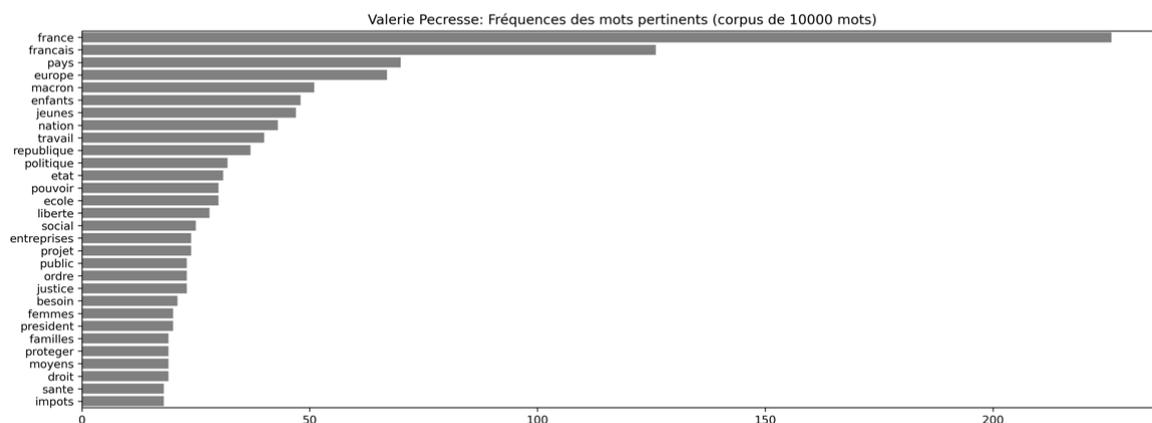
Le candidat le plus proche de Mélenchon est Hidalgo.

Jadot semble assez loin des autres candidats.

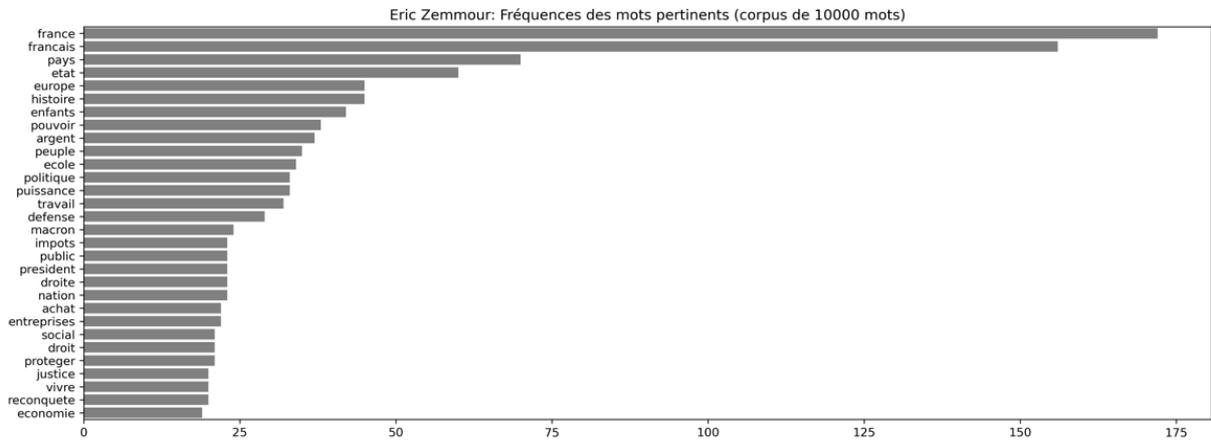
Le candidat le plus proche de Le Pen est Zemmour mais Zemmour est plus proche de Pécresse.

D. Les 30 termes sur 10000 les plus cités par les candidats

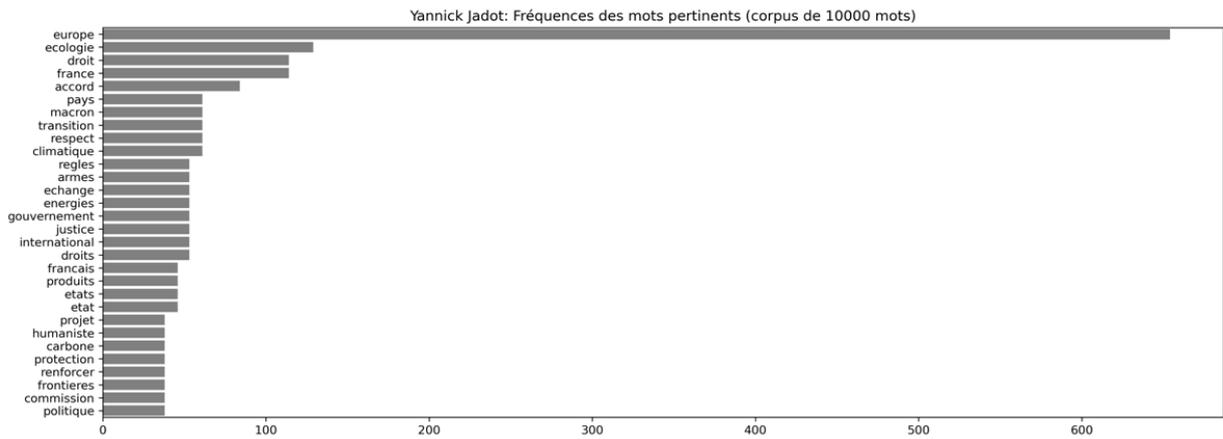
Pour Valérie Pécresse :



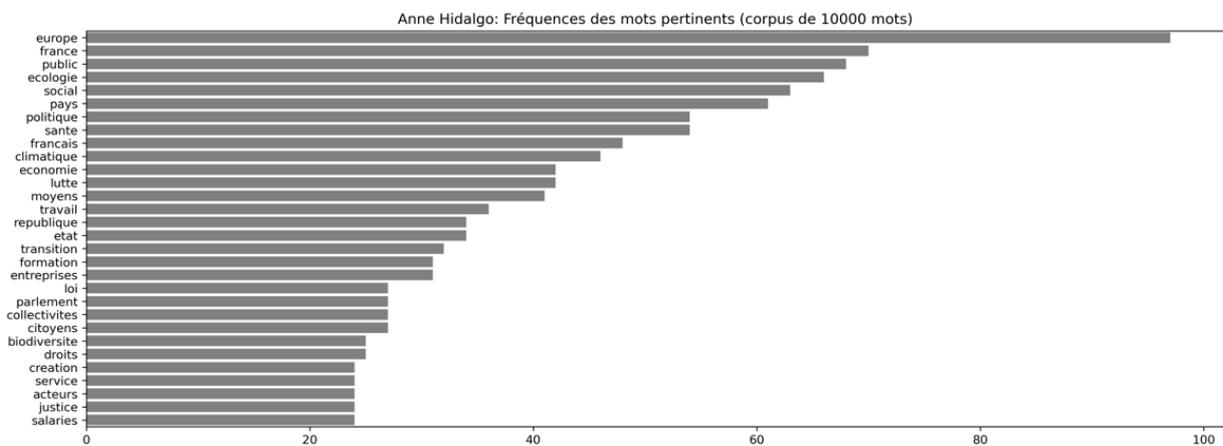
Pour Eric Zemmour :



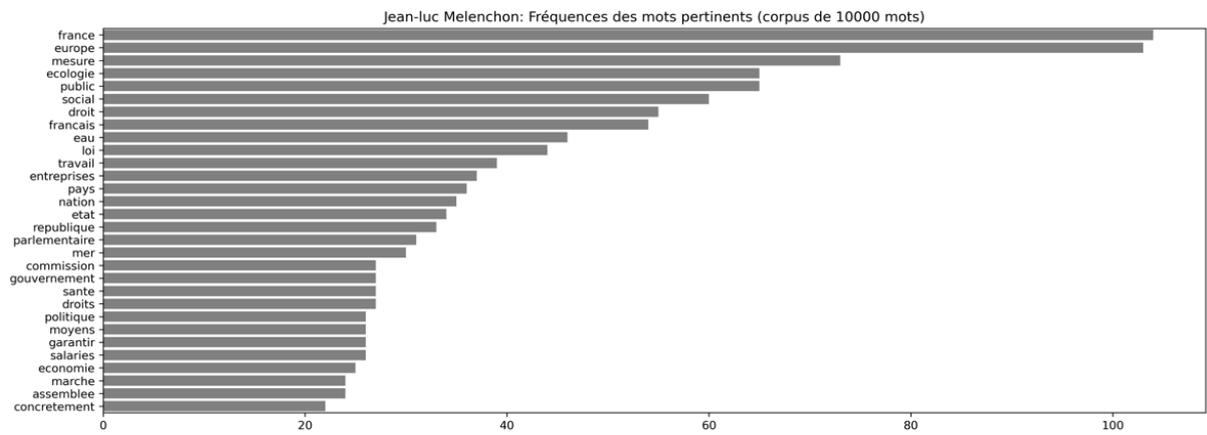
Pour Yannick Jadot :



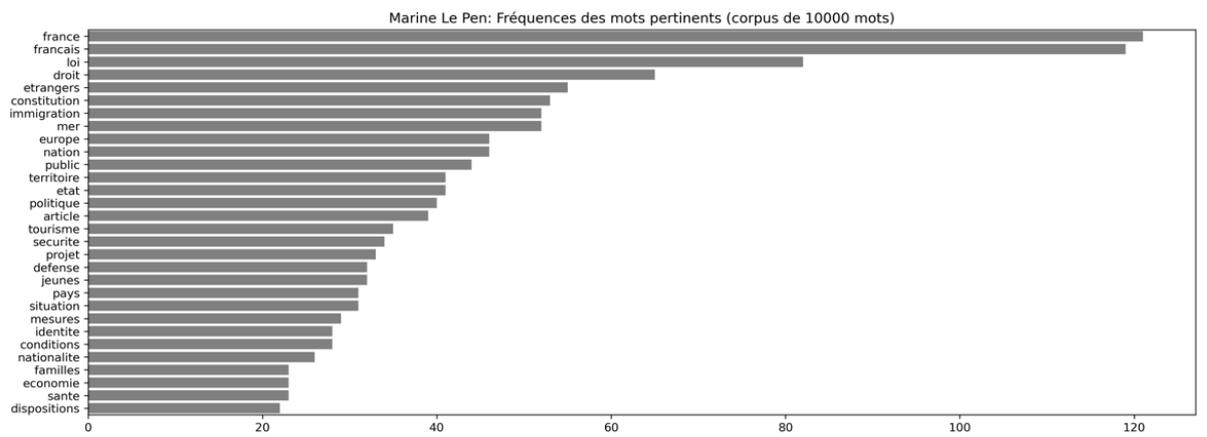
Pour Anne Hidalgo :



Pour Jean-Luc Mélenchon :



Pour Marine Le Pen :



F. Auteurs

Dr. Christophe LEGRENZI PDG ACADYS France et Président ACADYS International Prof Associé, PhD IAE/CNRS, PhD ISM, Ing. IT, Ing. SI Conseiller du Commerce Extérieur de la France	Dr. Georges BRESSANGE Directeur Associé ACADYS Expert data science PhD in Mathematics Responsable Business Unit BDS
---	--

À propos d'Acadys ?

Le cabinet ACADYS est **créé en 1996** par 3 associés fondateurs en France, Allemagne, Angleterre et Suisse, émanant à la fois du monde académique et de grands cabinets anglo-saxons.

Les dirigeants et consultants d'ACADYS ont toujours orienté leurs travaux et recherches autour d'un thème commun : **la performance et l'innovation d'entreprise**. Le **Système d'Information**, facteur incontournable de succès de la compétitivité d'entreprise, est devenu très vite le **domaine d'intervention** le plus connu d'ACADYS.

Selon nous, la transformation numérique et la modernisation des organisations n'a de sens que pour créer de la valeur humaine, durable et profitable.

La transformation numérique est un impératif et notre rôle est de faire de la transformation numérique une source certaine et durable de création de valeur.

La mission d'ACADYS est d'être le partenaire de confiance qui accompagne les décideurs dans leur transformation numérique pour créer de la valeur humaine, durable et profitable.

Présentation BDS – Beyond Data Sciences

Plus que technologique, la Data Intelligence est pour nous synonyme de démarche engagée pour faciliter le quotidien de vos équipes et valoriser vos actions.

BDS souhaite favoriser l'accès à toute organisation, publique comme privée, non initiée sur le sujet et éviter l'erreur d'appréhender l'intelligence de la donnée comme un enjeu purement technique en occultant les particularités sociales, managériales et culturelles des entreprises.

Constitué d'une équipe d'experts, à la fois sur la théorie et sur la pratique des enjeux liés à la donnée nous accompagnons les entreprises à comprendre, s'approprier et expérimenter la data intelligence, jusqu'à devenir encore plus performantes et inspirées.