



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 : 10 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 un mois afin d'identifier les thématiques communes mais aussi leurs spécificités.

Candidats : les 12 candidats validés par le conseil constitutionnel pour cette seconde analyse : Nathalie Artaud, Nicolas Dupont-Aignan, Anne Hidalgo, Éric Jadot, Jean Lassalle, Marine Le Pen, Emmanuel Macron, Jean-Luc Mélenchon, Valérie Pécresse, Philippe Poutou, Fabien Roussel, Éric 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, soit 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 :

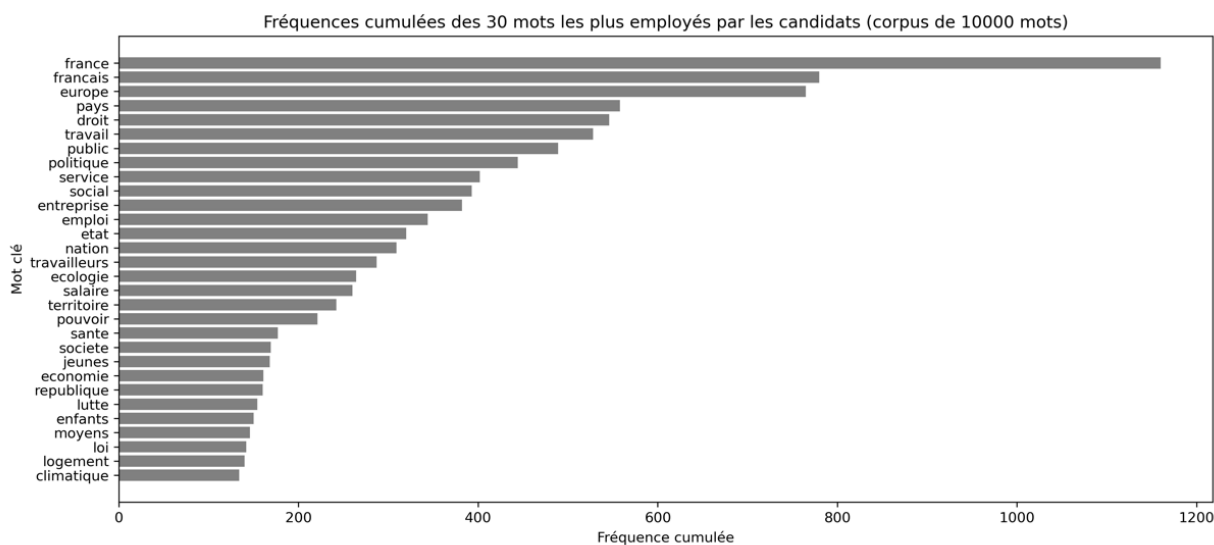
Analyses individuelles :

Par candidat, les termes **France** et **Europe** sont les termes de loin les plus cités et figurent systématiquement dans le Top 20 de chacun des candidats hormis pour Nathalie Artaud et Philippe Poutou où Europe n'apparaît pas (dans les 30 termes les plus fréquents). Le terme **français** apparaît chez tous les candidats sauf pour Roussel, Lassalle, Artaud et Poutou.

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 12 candidats (cumul des fréquences relativement à un corpus de 10 000 mots) sont représentés par le graphique ci-dessous :



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

Les autres termes les plus cités sont : **pays**, **droit**, **travail**, **public**, **politique**, **service**, **social**, **entreprise** puis viennent ensuite les termes, **emploi**, **état**, **nation**, **travailleurs**, **écologie**, **salaires**, **territoire**, **pouvoir**, **santé**.

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 :

Emmanuel Macron	Valérie Pécresse	Anne Hidalgo	Yannick Jadot	Marine Le Pen	Jean-Luc Mélenchon	Eric Zemmour	Fabien Roussel	Jean Lassalle	Nicolas Dupont-Aignan	Nathalie Artaud	Philippe Poutou
confiance	liberté	parlement	respect	tourisme	commun	histoire	égalité	développer	suppression	travailleurs	urgence
engagement	ordre	citoyens	accord	constitution	gouvernement	argent	salarie	aides	indépendance	patronat	finir
virus		collectivités	renforcer	immigration	prix	puissance		conseil	rendre	capitalistes	contrôle
chemin		biodiversité	ensemble	article	propositions	impôts		vote	soutien	classe	jeunesse
culture		salaries	produits	situation	eau	achat		élèves	valeurs	intérêts	richesses
épidémie				animaux	espace	vivre			mérite	parti	arrêt
choix				identité		reconquête				force	
continuer										conscience	
décisions										bourgeoisie	

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 :

Emmanuel Macron	Valérie Pécresse	Anne Hidalgo	Yannick Jadot	Marine Le Pen	Jean-Luc Mélenchon	Eric Zemmour	Fabien Roussel	Jean Lassalle	Nicolas Dupont-Aignan	Nathalie Artaud	Philippe Poutou
confiance	famille	climatique	climatique	tourisme	commun	histoire	développement	recherche	suppression	travailleurs	urgence
engagement	école	transition	transition	constitution	gouvernement	argent	égalité	développer	création	patronat	capitalisme
avenir	liberté	parlement	protection	immigration	prix	peuple	salarie	aides	recherche	capitalistes	finir
virus	ordre	citoyens	respect	article	avenir	école	création	conseil	indépendance	classe	profits
chemin		collectivités	accord	famille	propositions	puissance		vote	production	intérêts	contrôle
culture		biodiversité	états	protection	états	défense		peuple	rendre	parti	jeunesse
épidémie		salariés	renforcer	situation	développement	impôts		élèves	soutien	chômage	pandémie
choix			ensemble	défense	eau	achat			valeurs	capitalisme	richesses
chômage			produits	mesures	espace	vivre			mérite	profits	arrêt

Paris, le 10 mars 2022

Dans les mots simples (1-grams) utilisés par les candidats parmi les 30 mots les plus utilisés, certains sont communs aux **deux-tiers** des candidats :

Emmanuel Macron	Valérie Pécresse	Anne Hidalgo	Yannick Jadot	Marine Le Pen	Jean-Luc Mélenchon	Eric Zemmour	Fabien Roussel	Jean Lassalle	Nicolas Dupont-Aignan	Nathalie Artaud	Philippe Poutou
France	France	Europe	Europe	français	Europe	France	public	social	français	travail	travail
pays	français	France	droit	France	France	français	droit	France	France	entreprise	service
Europe	pays	public	service	droit	droit	pays	social	nation	pays	emploi	public
français	Europe	social	social	Europe	public	état	emploi	pays	travail	politique	droit
travail	nation	pays	France	territoire	social	Europe	service	territoire	politique		entreprise
emploi	travail	politique	travail	nation	service	politique	France	politique	droit		social
nation	entreprise	français	public	état	emploi	travail	travail	état	public		emploi
	politique	entreprise	politique	public	travail	entreprise	entreprise	service	service		politique
	état	service	pays	pays	français	droit	nation	Europe	nation		France

Version Alternative :

Dans les mots simples (1-grams) utilisés par les candidats parmi les 30 mots les plus utilisés, certains sont communs aux **trois-quarts** des candidats :

Emmanuel Macron	Valérie Pécresse	Anne Hidalgo	Yannick Jadot	Marine Le Pen	Jean-Luc Mélenchon	Eric Zemmour	Fabien Roussel	Jean Lassalle	Nicolas Dupont-Aignan	Nathalie Artaud	Philippe Poutou
France	France	Europe	Europe	France	Europe	France	public	France	France	travail	travail
pays	pays	France	droit	droit	France	pays	droit	pays	pays	entreprise	service
Europe	Europe	public	service	Europe	droit	état	service	politique	travail	politique	public
travail	travail	pays	France	état	public	Europe	France	état	politique		droit
	entreprise	politique	travail	public	service	politique	travail	service	droit		entreprise
	politique	entreprise	public	pays	travail	travail	entreprise	Europe	public		politique
	état	service	politique	politique	politique	entreprise	pays	entreprise	service		France
	droit	droit	pays	service	état	droit	politique	public	état		
	public	travail	état		pays	public	Europe	droit	Europe		

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

Paris, le 10 mars 2022

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 », « situation de handicap » en « situation handicap ».

Emmanuel Macron	Valérie Pécresse	Anne Hidalgo	Yannick Jadot	Marine Le Pen	Jean-Luc Mélenchon	Eric Zemmour	Fabien Roussel	Jean Lassalle	Nicolas Dupont-Aignan	Nathalie Artaud	Philippe Poutou
françaises français	pouvoir achat	service public	service public	union européenne	avenir commun	pouvoir achat	service public	proposition loi	peuple français	extrême droite	service public
chômage bas	forces ordre	transition écologique	transition écologique	projet loi	service public	France rurale	emploi formation	service national	service public	classe ouvrière	travailleurs travailleuses
président république	président république	collectivités territoriales	union européenne	article constitution	union européenne	service public	biens communs	développer recherche	écologie intelligente	femmes hommes	extrême droite
service public	politique familiale	union européenne	transition climatique	identité immigration	bifurcation écologique	peuple français	sécurité sociale	situation handicap	rendre pouvoir	propriété privée	réchauffement climatique
esprit résistance	Eric Zemmour	santé publique	femmes hommes	nationalité française	union populaire	argent public	fonction publique	taille humaine	regardons ailleurs	augmentations salaire	secteurs clés
campagnes rappel	États Unis	climat biodiversité	justice sociale	territoire national	France insoumise	république France	citoyennes citoyens	développer énergie	cohésion nationale	changer sort	industrie pharmaceutique
modèle social	île France	femmes hommes	accord paris	immigration projet	planification écologique	président république	jours heureux	statut salaire	récompenser travail	avenir société	exploitation oppressions
métropole mer	collectivités locales	énergies renouvelables	état droit	France puissance	seuil pauvreté	déserts médicaux	France jours	vote blanc	pouvoir peuple	camp travailleurs	programme rupture
avenir enfants	situation handicap	réchauffement climatique	droits humains	citoyenneté identité	situation handicap	États Unis	sociale solidaire	conseil économique	droit asile	conditions travail	gilets jaunes
préparer avenir	Nicolas Sarkozy	situation handicap	états membres	droit asile	commun programme	union européenne	prise charge	énergie solaire	Produire France	petits patrons	système capitaliste

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 $12 \times 30 = 360$ mots potentiels (30 mots retenus pour chacun des 12 candidats), il s'avère qu'il y a 140 mots uniques (soit un peu plus d'un tiers). Chaque candidat a donc été représenté par un vecteur de 140 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	Macron	Pécresse	Hidalgo	Jadot	Le Pen	Mélenchon	Zemmour	Roussel	Lassalle	Dupont-Aignan	Artaud	Poutou
Macron	0	241,9	317,3	351,7	316,2	302	258,7	361	328,7	282,4	521,1	387,4
Pécresse	241,9	0	277,4	334,7	248,9	269,8	166,1	310,4	283,1	200,2	509,2	355,8
Hidalgo	317,3	277,4	0	185,8	264,2	171,2	272,5	204,2	216,2	257,6	461	261,5
Jadot	351,7	334,7	185,8	0	309,8	210,2	328,9	270,6	277,6	321,8	490,7	306,8
Le Pen	316,2	248,9	264,2	309,8	0	238,1	237,4	281,6	266,8	226,9	498,8	325,7
Mélenchon	302	269,8	171,2	210,2	238,1	0	259,1	177,5	191,1	234,8	450,6	239,1
Zemmour	258,7	166,1	272,5	328,9	237,4	259,1	0	315,5	278,2	187,6	495,9	340,9
Roussel	361	310,4	204,2	270,6	281,6	177,5	315,5	0	223	280,9	450	229,8
Lassalle	328,7	283,1	216,2	277,6	266,8	191,1	278,2	223	0	255,6	444,7	245,8
Dupont-Aignan	282,4	200,2	257,6	321,8	226,9	234,8	187,6	280,9	255,6	0	468,2	302,9
Artaud	521,1	509,2	461	490,7	498,8	450,6	495,9	450	444,7	468,2	0	406,1
Poutou	387,4	355,8	261,5	306,8	325,7	239,1	340,9	229,8	245,8	302,9	406,1	0
		très proche										
		proche										

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 très proche d'Hidalgo.

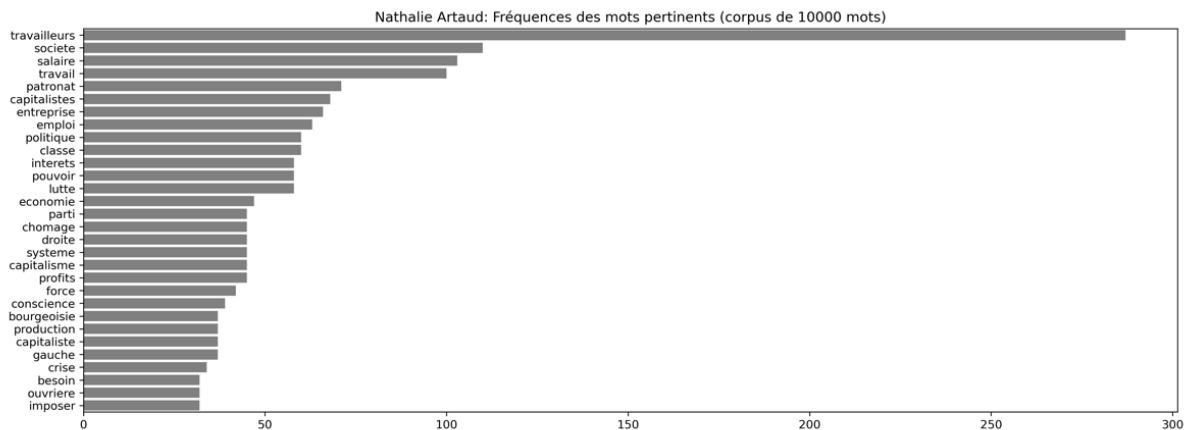
Artaud est très loin des autres candidats.

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

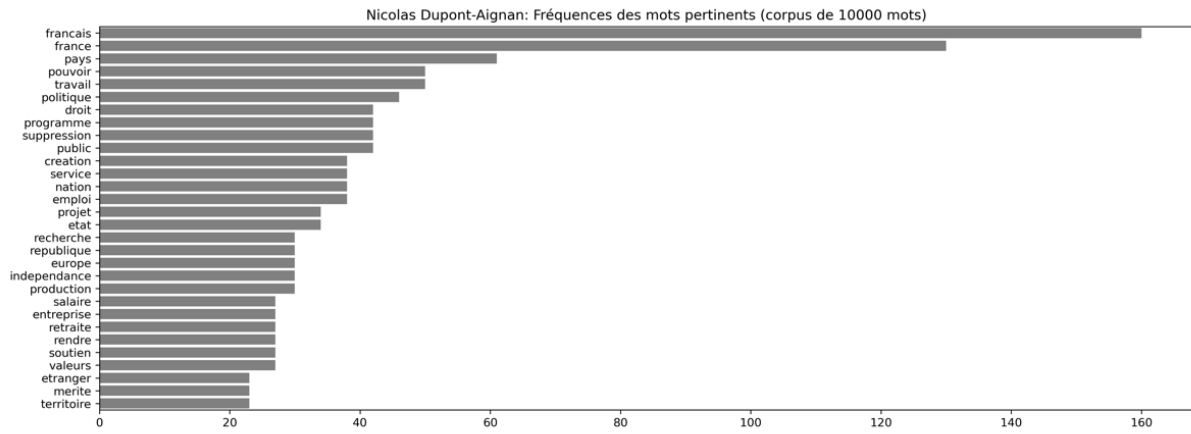
D. Les 30 termes sur 10000 les plus cités par les candidats

Les 30 termes les plus cités par ordre de fréquence :

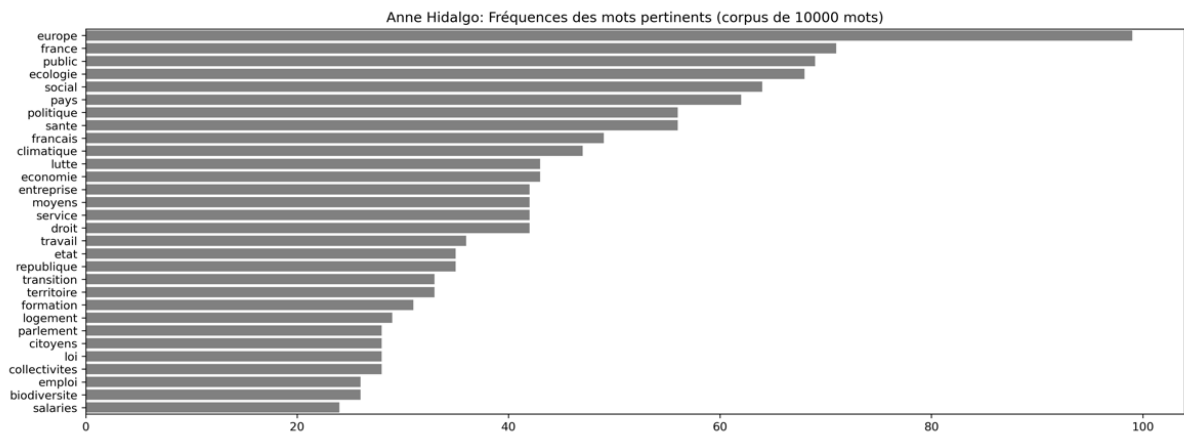
Pour **Nathalie Artaud** :



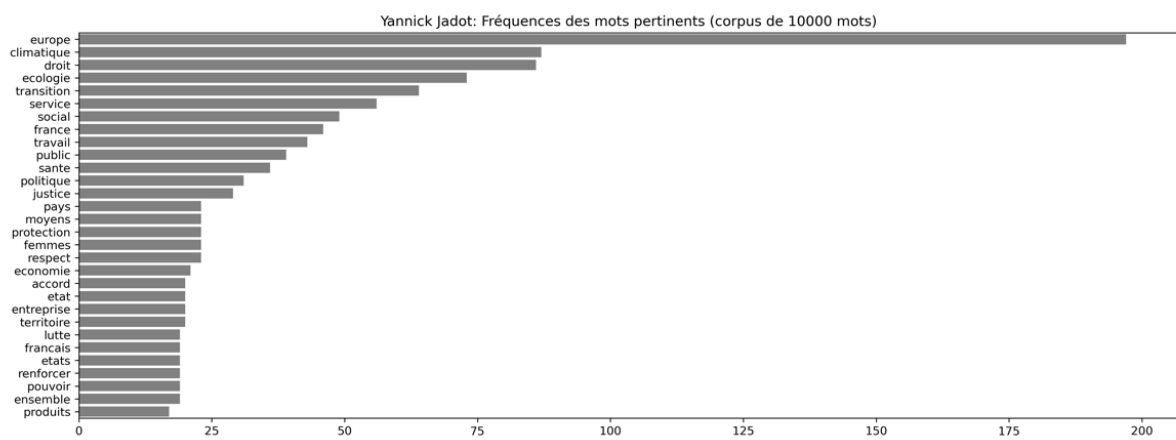
Pour Nicolas Dupont-Aignan :



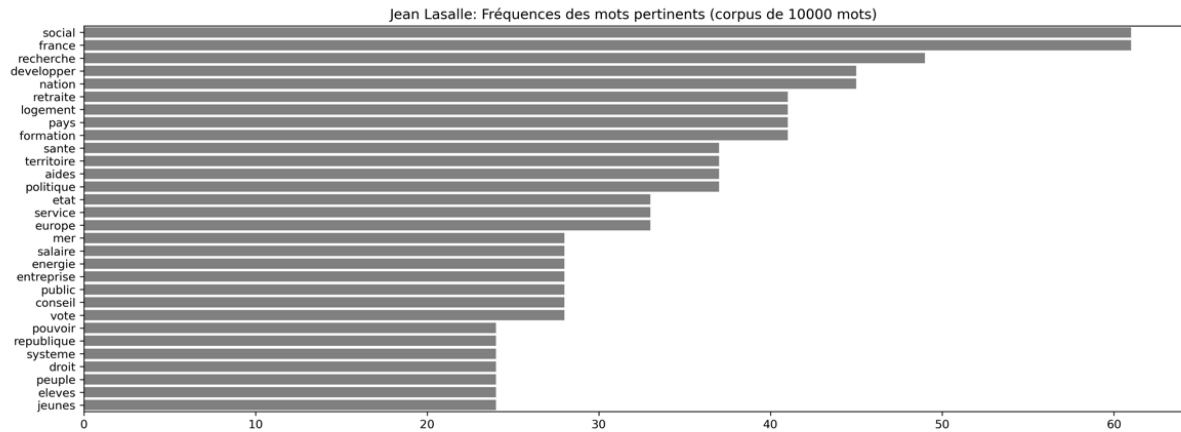
Pour Anne Hidalgo :



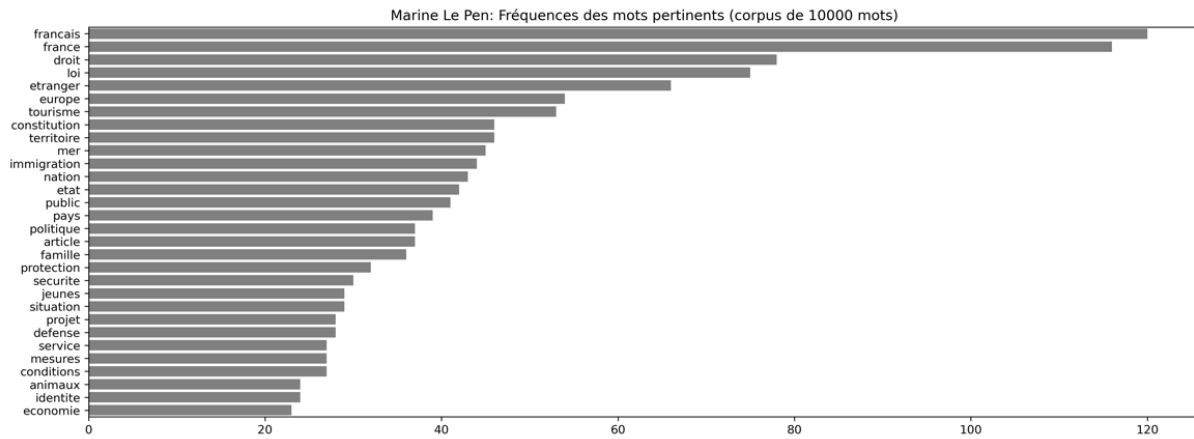
Pour Yannick Jadot :



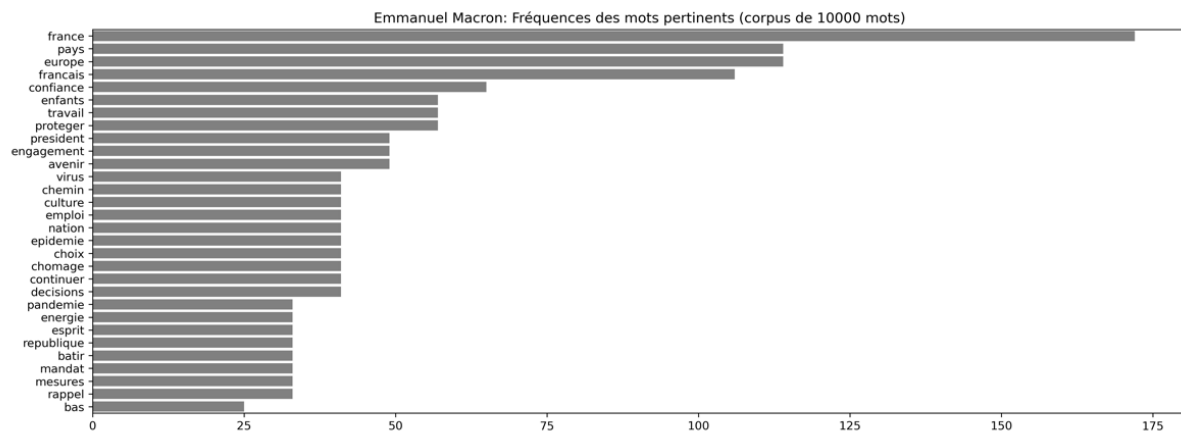
Pour Jean Lassalle :



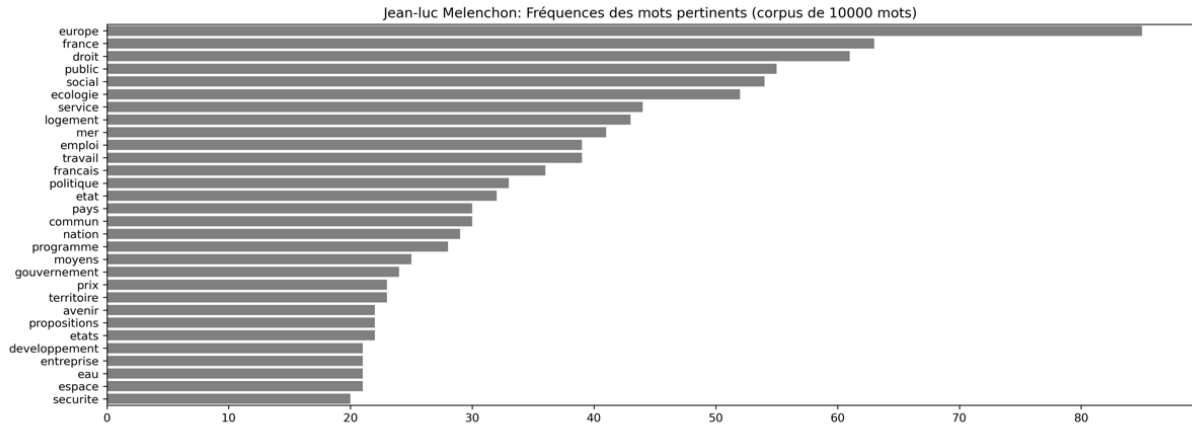
Pour Marine Le Pen :



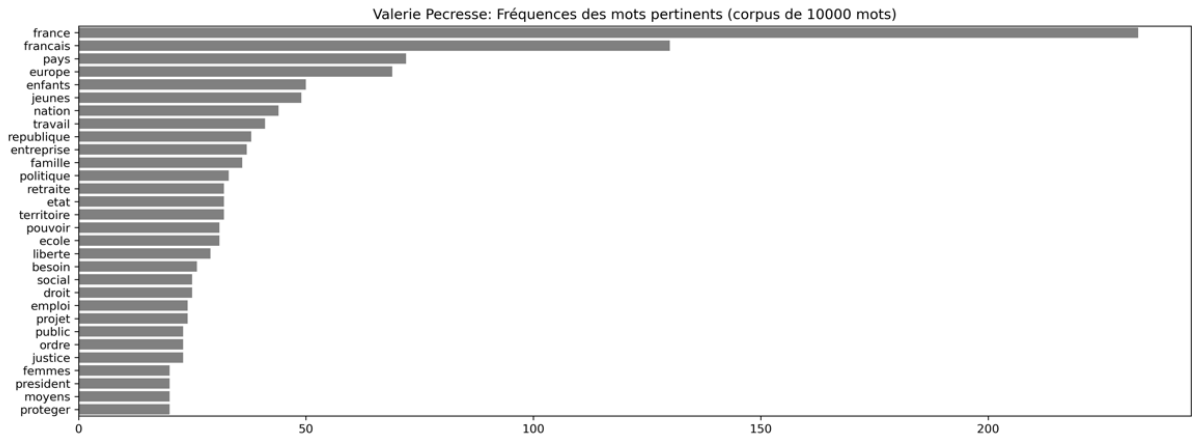
Pour Emmanuel Macron :



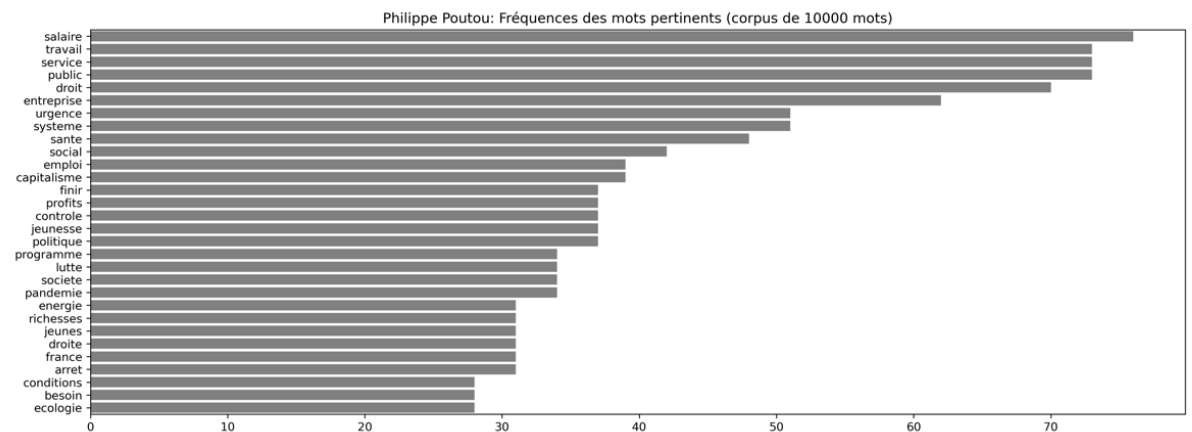
Pour Jean-Luc Mélenchon :



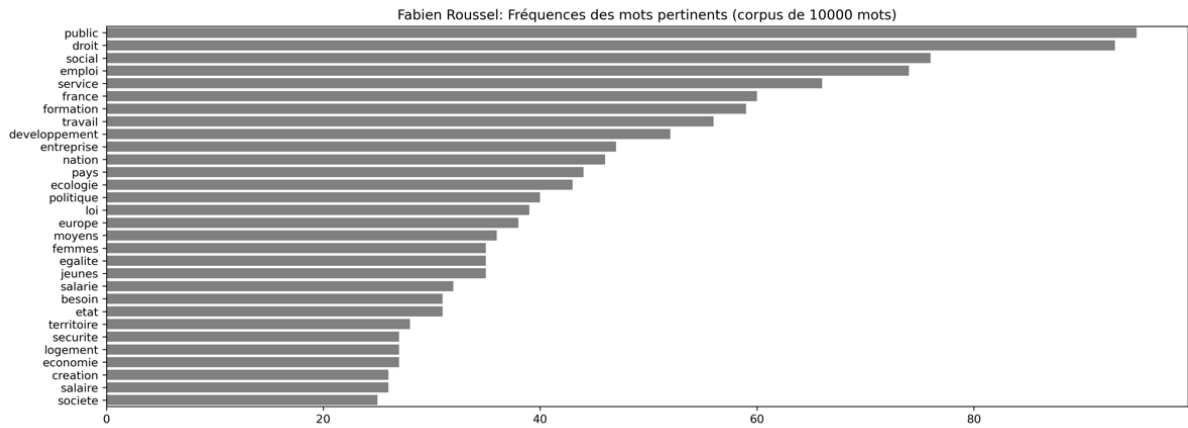
Pour Valérie Pécresse :



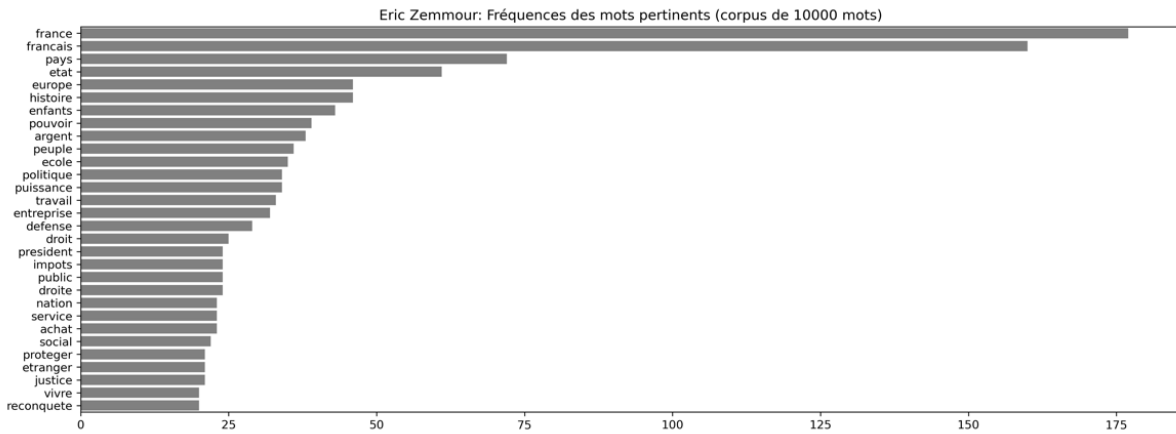
Pour Philippe Poutou :



Pour Fabien Roussel :



Pour Éric Zemmour :



Pour Fabien Roussel :



Pour Éric Zemmour :



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.

Pour toute demande d'informations complémentaires,
nous vous remercions de bien vouloir contacter :

Christophe LEGRENZI
christophe.legrenzi@acadys.com
+33 6 85 43 38 32

Myriam RAHMOUNI
myriam.rahmouni@acadys.com
+33 7 83 96 18 06

Camille BEN FERHAT
camille.ben-ferhat@acadys.com
+33 6 51 29 79 54