



HAL
open science

Décrire les textes politiques par le deep learning : à la recherche de nouveaux observables

Magali Guaresi

► **To cite this version:**

Magali Guaresi. Décrire les textes politiques par le deep learning : à la recherche de nouveaux observables. JADT 2020 : 15es Journées internationales d'Analyse statistique des Données Textuelles, Jun 2020, Toulouse, France. hal-03167188

HAL Id: hal-03167188

<https://hal.science/hal-03167188v1>

Submitted on 1 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Décrire les textes politiques par le deep learning : à la recherche de nouveaux observables

Magali Guaresi¹

¹Université Côte d'Azur (BCL), Université Libre de Bruxelles (ReSIC) –
magali.guaresi@gmail.com

Abstract

The methods of deep learning, particularly deconvolution, have recently made it possible to go beyond simple classification tasks in order to develop text description tasks. This article proposes to apply the methods of deep learning on a corpus of electoral proclamations by the left and the right between 1958 and 2017. We highlight some salient results to interpret electoral speeches under the French V^o Republic in order to emphasize, from a methodological point of view, the added value of deconvolution protocols. We thus point to well-known units of textual statistics (such as specificities, lexical and grammatical co-occurrences). But we also show how the model allows for the formal capture of complex syntagmatic units, which the analysis of textual data has often conceptualized without being able to identify them automatically, such as patterns (« motifs ») or passages.

Keywords: Deep learning, Textual data analysis, Description, Motifs, Passages, Cooccurrences, Political discourse

Résumé

Les méthodes du deep learning, en particulier la déconvolution, permettent depuis peu de dépasser les simples tâches de classification pour enrichir la description des textes. Cet article propose d'appliquer les méthodes d'apprentissage profond sur un corpus de proclamations électorales de gauche et de droite entre 1958 et 2017 en France. Au fil de l'exposition de quelques résultats saillants pour comprendre les discours électoraux sous la V^o République, on soulignera, d'un point de vue méthodologique, la plus-value de la déconvolution. Celle-ci met au jour des unités bien connues de la statistique textuelle (telles que les spécificités, les cooccurrences lexicales ou grammaticales). Mais elle permet également de saisir formellement des unités syntagmatiques plus complexes, que l'analyse des données textuelles a souvent conceptualisées sans parvenir à les repérer automatiquement, tels que les motifs ou les passages.

Mots clés : Deep learning, Statistique textuelle, Description, Motifs, Passages, Cooccurrences, Discours politiques

1. Introduction

Depuis peu, l'intelligence artificielle (IA) et en particulier les méthodes du deep learning dépassent les simples tâches de classification ou de prédiction pour enrichir les tentatives de description des textes (Ducoffé et al., 2016 ; Vanni, Mayaffre et Longrée, 2018). Cette contribution pose l'hypothèse d'une plus-value heuristique des observables linguistiques traités dans les protocoles de deep learning. Elle vise à en donner des exemples, tirés d'une

application sur un corpus de discours électoraux français et, affermis par un contrôle a posteriori des outils traditionnels de l'analyse de données textuelles (ADT).

1.1. Méthode : la déconvolution

Les réseaux de neurones du deep learning (ici un modèle convolutionnel – CNN – implémenté dans le logiciel Hyperbase) ne se contentent plus d'attribuer un texte à son auteur avec de bons scores de réussite mais permettent désormais d'observer les items linguistiques qui ont présidé à l'apprentissage profond puis à la bonne reconnaissance : c'est la déconvolution (Kim 2014, Vanni et al. 2020-*submitted*). L'algorithme de déconvolution permet de souligner les éléments linguistiques qui ont participé, dans les couches profondes du système, à la prise de décision. Ces unités sont dotées d'un indice quantifiant leur activation lors de l'apprentissage : le Text Deconvolution Saliency (TDS) (Vanni et al., 2018). Les données les plus fortement activées sont celles qui ont contribué à la bonne identification du texte. Précisions que chaque mot, ou *token*, est associé à un TDS unique puisque le modèle convolutionnel le considère toujours dans une fenêtre cotextuelle (de 3 à 10 mots)¹. Aussi, les différentes occurrences d'un mot dans le corpus ne sont jamais identiques, contrairement à la spécificité en ADT. Par ailleurs, les données considérées relèvent des trois niveaux linguistiques : la forme graphique, le lemme et l'étiquette morpho-syntaxique que l'algorithme considère simultanément. En somme, la déconvolution est potentiellement capable de restituer à l'analyste des unités textuelles complexes tels que des patterns grammatico-lexicaux ou des motifs au sens de Longrée et Mellet (2013) ou encore des passages, dans la perspective herméneutique rastirienne (Rastier 2007), susceptibles de nourrir nos tâches de description et d'interprétations des discours politiques contemporains.

1.2 Corpus et protocole d'apprentissage

Dans cette contribution, nous proposons d'appliquer les protocoles de déconvolution sur un jeu de professions de foi électorales, rédigées par les vainqueurs des législatives entre 1958 et 2017², dans la perspective de décrire les évolutions du discours électoral de gauche et de droite au fil de la V^e République. Pour ce faire, un entraînement a été mené sur le corpus de proclamations électorales, regroupées selon l'appartenance partisane (gauche/droite) et la périodisation chronologique³. Série textuelle chronologique (Salem, 1991), notre corpus a été soumis à une analyse factorielle des correspondances préalable dans le but d'établir une chronologie endogène (Guaresi, 2018). Ont ainsi été formés 6 textes différents : la gauche 1 et la droite 1 (de 1958 à 1973), la gauche 2 et la droite 2 (de 1978 à 1997) et la gauche 3 et la droite 3 (de 2002 à 2017). Sur la base de ces données, la machine apprend de manière

¹ Dans cet article [inscrit dans le projet UCAJEDI portant la référence n°ANR-15-IDEX-01], le modèle balaye le texte via une fenêtre coulissante de 6 mots.

² Plus précisément, le corpus est composé de la quasi-exhaustivité des professions de foi électorales des femmes députées et d'un échantillon comparable de textes d'hommes. Comptant 864 392 occurrences, il incarne un solide observatoire des discours électoraux partisans depuis 1958. En raison des difficultés posées par le classement des discours issus du mouvement En Marche en 2017, qui se revendique souvent « ni de gauche, ni de droite », nous avons exclu les textes de la future majorité de la phase d'entraînement du réseau de neurones.

³ Si le classement des professions de foi à gauche ou à droite a été effectué par nos soins sur la base de l'étiquette affichée par les élu.es, la périodisation en trois temps a été retenue sur la base d'une AFC sur l'ensemble du vocabulaire du corpus.

autonome à reconnaître le discours de la droite et de la gauche à chacune des périodes déterminées. Dans un second temps, elle nous restitue les zones d'activation des textes susceptibles de guider l'interprétation historico-linguistique des discours des partis de l'ordre et du mouvement, dont nous exposons quelques-uns des résultats les plus saillants dans le développement de l'article. Ce faisant, nous nous interrogerons sur les vertus descriptives et interprétatives des unités mises au jour par les protocoles de déconvolution, à la lumière des résultats issus des méthodes éprouvées de l'analyse de données textuelles (i.e. spécificités et cooccurrences principalement).

2. 1958-1973: l'opposition des modèles

2.1. La droite gaulliste ou le chaos

Les passages les mieux reconnus pour la droite du début de la V^e République, et donnés à lire par le mécanisme de la déconvolution, décrivent le discours électoral gaulliste. Ils évoquent le « changement » et le « renouvellement » de la classe politique, accusée d'avoir mené le peuple de France au bord de l'abîme. Typiquement, l'algorithme de déconvolution souligne parmi les passages les plus fortement activés pour le classement à droite entre 1958 et 1973 ces phrases de Léon Delbecque (nous soulignons les zones de forts TDS) :

« Oui » le peuple de France a pris un bon départ. Derrière le général de Gaulle il s'est engagé résolument sur le chemin du renouveau national. Il l'a prouvé le 28 septembre en votant massivement pour un changement complet des méthodes périmées de l'ancien système et le renouvellement total des hommes qui l'avaient mené au bord de l'abîme⁴.

On pourrait s'étonner que le mot « changement » que l'histoire politique a bien identifié à gauche apparaisse dans les zones d'activation de cet extrait classé à droite, d'autant plus que ce terme n'est une spécificité lexicale des professions de foi de droite (Figure 1).

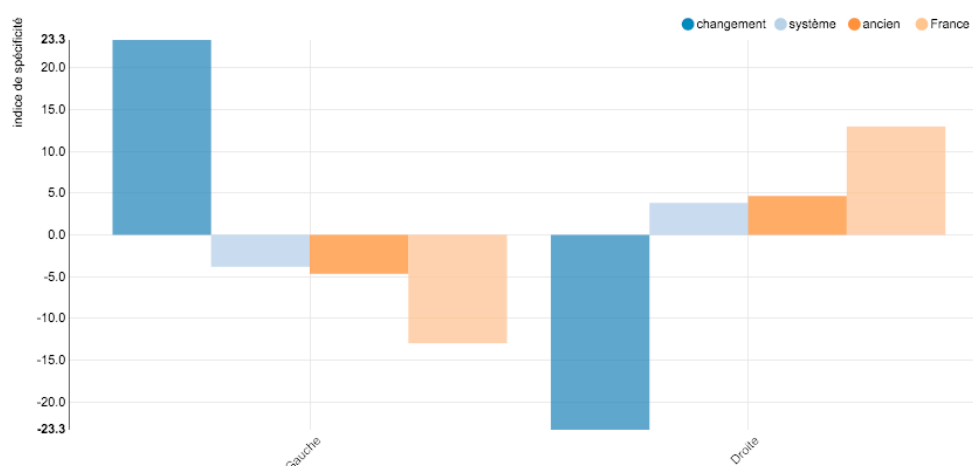


Figure 1 : Sous-utilisation du mot « changement » à droite et sur-utilisation des mots cooccurrents dans le passage : « système », « ancien » et « France »⁵.

Mais, le calcul des spécificités de l'ADT nous montre que les mots « système », « ancien », ou « France », présents dans le co-texte de « changement » dans l'exemple, sont eux sur-

⁴ Léon Delbecque, 1958, Nord.

⁵ Les graphiques sont tirés du logiciel Hyperbase Web Edition.

utilisés par la droite. Le deep learning ne s'y trompe donc pas et s'attache au terme « changement » en considérant son contexte d'emploi, c'est à dire les autres mots figurant dans le passage. L'algorithme active la reconnaissance à droite en raison de la cooccurrence du mot avec d'autres, plus typiques de la droite de cette époque comme « système », « ancien », ou « France » utilisés par les gaullistes pour pourfendre la IV^o République.

Le « changement » ici ne correspond en rien à un bouleversement de l'ordre social ou économique mais bien à un changement de régime, à des modifications institutionnelles qu'il s'agit de stabiliser et de maintenir en se fondant sur la légitimité du père fondateur, le général de Gaulle. La prose électorale de droite s'emplit alors d'une dimension référendaire que le réseau de neurones profonds, qui s'active sur les structures linguistiques faisant le parallèle entre l'adhésion à la France (ou à la V^o République), l'adhésion à de Gaulle et l'expression du peuple, détecte bien. « En votant oui à la France, oui à de Gaulle, vous avez également ratifié ce que la colère d'un peuple bafoué et la lassitude d'une armée avaient fait pour la patrie le 13 mai »⁶, écrit Madeleine Martinache aux élections de novembre 1958.

Pendant les quinze premières années du régime, la rhétorique de droite procède de rouages stables que la machine a su apprendre et pointe en sortie de réseau : celle de l'alternative entre « de Gaulle ou le chaos ». De Gaulle est présenté comme le seul rempart contre le retour à l'ancien système de la IV^o République. A la question qu'il pose lui-même : « Qui peut soutenir le général de Gaulle au Parlement ? » dans sa profession de foi, Léon Delbecque répond en reprenant l'alternative entre le gaullisme ou les hommes des partis de l'ancien système : « Electrices, électeurs vous n'avez le choix qu'entre les candidats des partis qui nous ont fait tant de mal et celui de l' UNR présenté par les associations gaullistes et soutenu par le mouvement des indépendants⁷. »

En 1968, la rhétorique du « de Gaulle ou le chaos » prend une dimension nouvelle lors des élections qui suivent les événements de mai. Les zones d'activation des passages attribués à la droite de cette époque mettent explicitement en balance « la dictature communiste » et « de Gaulle », le sauveur, comme par exemple dans l'exemple suivant :

Un plan diabolique de subversion a tenté d'imposer la dictature du parti communiste aux Français. Trois semaines de violences organisées ont menacé la paix, l'indépendance, le progrès. Une fois de plus, le général de Gaulle a arrêté le péril. Mais la menace demeure⁸.

A cette date, la tendance à la démocratie plébiscitaire et la priorité du maintien de l'ordre et de la stabilité du régime face aux mouvements estudiantins et sociaux de mai 1968 font du gaullisme un bonapartisme moderne (Vavasseur-Desperriers, 2006).

2.2. La gauche, le parti, la démocratie et la paix

A gauche, les passages les mieux reconnus sont tirés des professions de foi communistes qui dominent dans le corpus d'apprentissage pour la période 1958-1973 et qui se singularisent nettement par une énonciation spécifique que l'IA repère. Les candidat.es communistes ne concèdent rien à la logique de personnalisation des batailles électorales induite par le scrutin uninominal majoritaire à deux tours, qui tend à organiser la compétition entre ténors locaux bien établis. Au contraire, le réseau de neurones distingue la gauche de la droite de cette

⁶ Madeleine Martinache, 1958, Nord.

⁷ Léon Delbecque, 1958, Nord.

⁸ Nicole Hauteclouque, 1968, Paris.

période par le recours à une énonciation collective organisée autour du « parti ». Le mot, en position de sujet, est par ailleurs toujours activé dans le motif phraséologique (Longrée & Mellet, 2013) suivant : « le parti communiste + VER » (Figure 2). C'est donc le parti communiste qui « soumet », qui « adresse », qui « propose », qui « préconise » et qui finalement porte la candidature. La forte activation de ce motif pour la classification des discours de la gauche entre 1958 et 1973 signale également le primat du pôle communiste dans l'opposition de gauche qui tente de s'organiser autour du Programme commun. Les candidat.es appellent de leur vœu un front « ouvrier, démocratique et progressiste » qui vise tant à se démarquer du socialisme soviétique que du pouvoir personnel gaulliste. Marie-Claude Vaillant-Couturier écrit ainsi en 1968 :

[...] exiger que soient durablement garantis les résultats des grandes luttés ouvrières et que soient satisfaites les autres aspirations des travailleurs manuels et intellectuels. Alors que s'exprime avec force la volonté populaire d'en finir avec le pouvoir gaulliste, le parti communiste soumet aux électeurs son programme démocratique et social avancé [...]. Le parti communiste ne revendique pas pour lui seul le pouvoir. Les communistes sont pour un gouvernement populaire et d'union démocratique s'appuyant sur la volonté du peuple et dans lequel les communistes doivent avoir la place qui leur revient. Pour faire échec à la menace de dictature gaulliste. Pour la défense des libertés. Pour une politique de progrès social. Pour la démocratie et pour le socialisme. Pour l'indépendance nationale et la paix⁹.

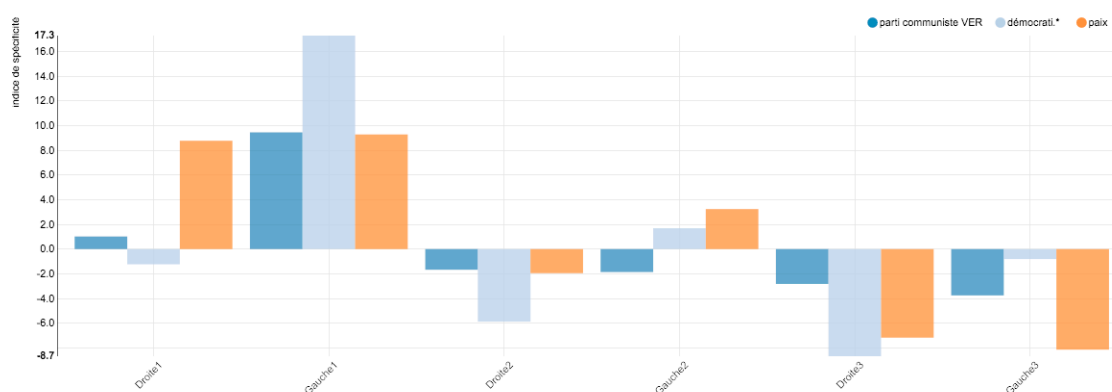


Figure 2 : Sur-utilisation du motif « parti communiste VER », de la racine « démocrat* » et du lemme « paix » dans les professions de foi de gauche en 1858-1973.

L'intelligence artificielle reconnaît ici la combinaison entre l'énonciation collective du parti et le programme idéologique très explicite de l'opposition de gauche, en particulier sur les questions démocratiques et sociales mais également sur la « paix », terme doté d'un fort TDS dans l'extrait. Le mot est pourtant très utilisé par la droite gaulliste également (Figure 2). Là encore l'activation du réseau semble prendre en compte le mot dans son co-texte. La présence, au sein de l'extrait, des lemmes « démocratie », « indépendance » et « peuple » contribue à la forte activation du terme « paix » et à la classification de l'ensemble à gauche, comme le confirme la comparaison statistique des co-occurrences du lemme « paix » dans les professions de foi de gauche (en haut) et de droite (en bas) entre 1958 et 1973 (Figure 3). En prenant en compte le co-texte, le réseau identifie à raison le programme pacifiste anti-militariste et anti-colonial de la gauche communiste du début de la V^e République.

⁹ Marie-Claude Vaillant-Couturier, 1968, Val-de-Marne.

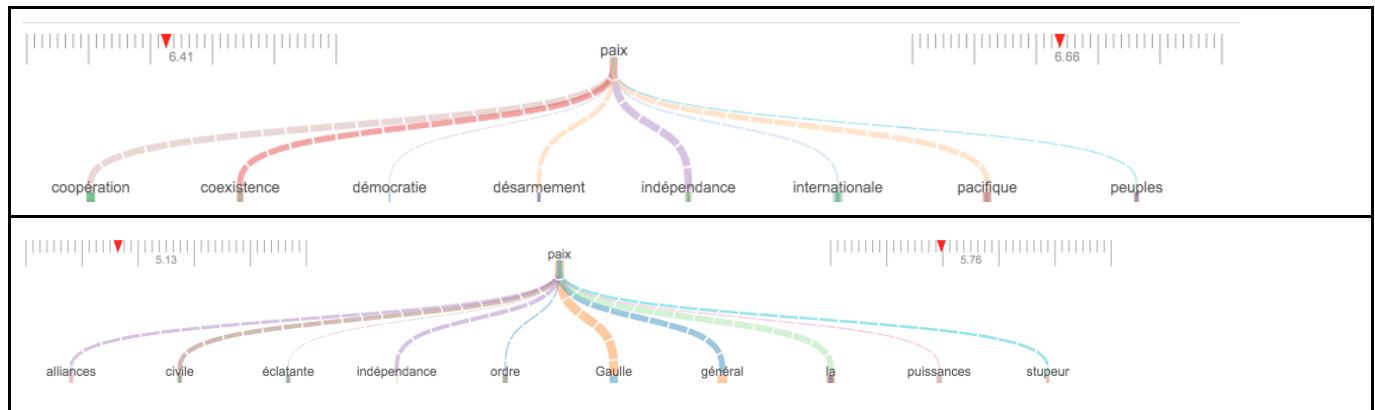


Figure 3 : Les co-occurrences statistiques de « paix » dans le discours de gauche (en haut) et de droite (en bas) entre 1958 et 1973.

3. Alternances et cohabitations (1978-1997)

3.1. La droite : descente dans l'arène et glissements thématiques

Entre 1978 et 1997, la droite ne peut plus s'exprimer depuis sa position surplombante de l'époque gaulliste. Régulièrement dans l'opposition (après les victoires des socialistes en 1981, 1988 et 1997), elle est aussi concurrencée sur sa droite par la montée du Front national (FN)¹⁰. Dans ce contexte, la droite descend dans l'arène pour mener une vive critique du bilan de la gauche et absorber les revendications de l'extrême droite. Comme le suggèrent les extraits des professions de foi les plus fortement reconnus pour la classe « droite2 », la critique ne se fait pas centralement sur la politique économique ou sociale mais plutôt sur la morale et sur le terrain régalien. A la fin des années 1980, et de façon encore plus marquée en 1993, date à laquelle l'union RPR et UDF gagne largement les élections, les candidat.es adoptent un discours décliniste et axiologique et proposent de redresser le pays, à renfort de verbes à l'infinitif préfixés en « re » que l'IA repère bien :

Rénover le modèle social et culturel français, en réaffirmant avec force, dans une morale politique retrouvée, nos valeurs de justice [...]¹¹.

Dans ce programme de restauration nationale, les passages qui activent avec les plus forts indices de TDS les réseaux du deep learning sont les énumérations contenant les lemmes « insécurité » ou « immigration ». Directement puisés dans les tracts électoraux du FN, ces extraits témoignent de la dérive thématique opérée dès la fin des années 80 vers les sujets de prédilection de l'extrême droite liant dans des raccourcis co-textuels insécurité et immigration (Barats-Malbreil, 1999). En 1988, Edouard Balladur écrit :

Nous devons poursuivre notre lutte commune contre l'insécurité, la délinquance et le crime.¹²

En 1993 et en 1997, la structure phraséologique (la lutte contre/ le combat contre + l'insécurité/l'immigration + x) se retrouve sous la plume de Marsaudon lorsqu'il se déclare en

¹⁰ Notons que les corpus d'apprentissage, puis de test, ne contiennent pas les professions de foi de 1986 qui voient la victoire de candidat.es Front national. Ces textes ont été écartés en raison de l'hétérogénéité induite par la tenue exceptionnelle des élections législatives au scrutin de liste.

¹¹ Martine Aurillac, 1997, Paris.

¹² Edouard Balladur, 1988, Paris.

faveur d' « Une politique de lutte contre l'insécurité et l'immigration clandestine et pour restauration de la justice » ou de Tibéri qui dit se présenter « Pour continuer le combat contre la délinquance, la toxicomanie, l'immigration irrégulière. ». Ce sont donc des patterns séquentiels structurants, des motifs, que l'apprentissage profond active fortement pour la reconnaissance des discours de droite.

3.2. La gauche de gouvernement

En position d'accéder au pouvoir, la gauche de la deuxième période change radicalement son discours à la fin des années 70¹³. Elle passe de la contestation à l'exercice du pouvoir après 23 ans de règne de la droite et se projette avec enthousiasme, en 1981 au moins, dans l'avenir qu'elle veut changer. En effet, les extraits les mieux reconnus de cette époque sont composés de constructions pronominales et verbales exprimant le futur proche (« Nous + aller + verbe à l'infinitif » ou « Nous + verbes au futur » - Figure 4). Et au lendemain de la victoire de Mitterrand en 1981, Edwige Avice anticipe l'action parlementaire en ces termes :

Ensemble nous allons relancer l'économie. Nous allons donner la priorité à l'emploi. Nous allons construire une France où chacun aura sa place.¹⁴

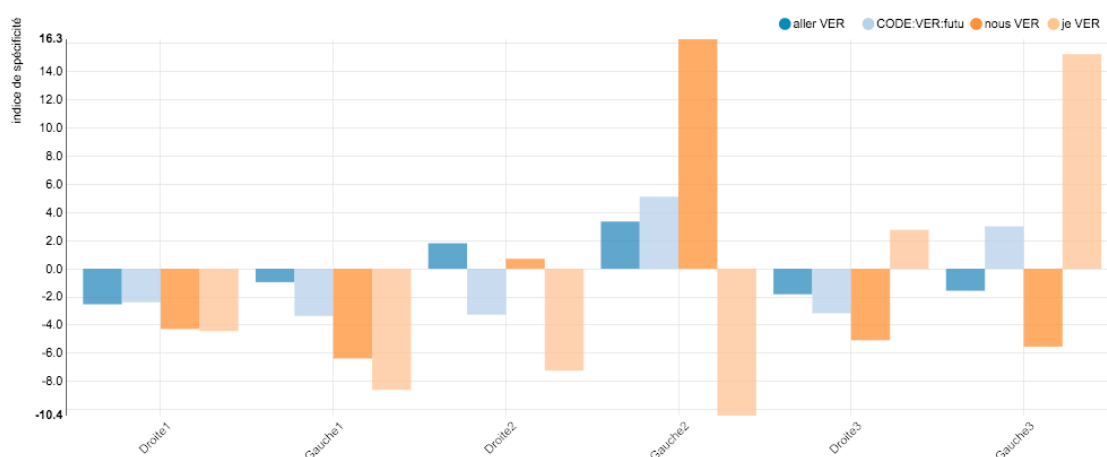


Figure 4 : Sur-utilisation des structures grammaticales « aller + Verbe à l'infinitif » et « nous + Verbe futur » dans la gauche 2 (1978-1997).

L'un des leitmotifs de la gauche de gouvernement, souvent conjugué au futur, réside dans la promesse d'un changement : il s'agit de « changer la vie » comme le promet le programme de conquête électorale du Parti socialiste (PS) en 1972 et comme le répètent les futur.es député.es dans les années 1980 :

C'est ensemble que nous changerons la vie, c'est ensemble que nous lutterons pour la paix, l'emploi, la liberté... alors, la devise républicaine, liberté, égalité, fraternité, retrouvera toute sa valeur¹⁵.

¹³ Ce constat est valable surtout pour l'après 1981. En 1978, alors que l'alternance est sans doute déjà possible, le parti communiste s'en tient à un discours idéologique fort, de taxation de grandes fortunes, de nationalisation, etc.

¹⁴ Edwige Avice, 1981, Paris.

¹⁵ Jean-Michel Testu, 1981, Indre et Loire.

Cette rhétorique persiste jusqu'en 1997 comme une signature du discours de gauche que l'IA reconnaît. Mais les passages alors activés informent également sur le glissement programmatique du parti socialiste au cours de la période (Marlière et Zappi, 2012). Certes, les candidats du parti du mouvement promettent toujours de changer « la vie quotidienne » mais la suite des unités activées (les verbes au futur) témoigne de l'attitude défensive de la gauche qui cherche avant tout à préserver ou maintenir des acquis sociaux ou à redonner ou relancer certains secteurs menacés et qui semble avoir renoncé à construire un nouveau système. Par exemple Eric Besson, candidat dans la Drôme, écrit en 1997 l'un des passages les mieux attribués à la gauche de cette période :

Nous changerons la vie quotidienne. Nous préserverons la sécurité sociale. Nous maintiendrons le régime de retraite par répartition. Nous redonnerons la priorité à l'éducation nationale. Nous relancerons le logement social. Nous assurerons la sécurité de proximité avec davantage d'agents [...] ¹⁶.

Les autres passages activés soulignent la teneur des chantiers ouverts par le PS : la réforme des institutions. Mais loin de s'en prendre au présidentielisme de la V^e République, que la gauche a fini par adopter, la gauche réformatrice s'empare plutôt de la question de la crise de la représentation et entend rénover les liens entre les élites et le peuple par la fin du « cumul des mandats » et la féminisation de la vie politique (la « parité »). En témoignent ces phrases bien reconnues de Marie-Françoise Clergeau, candidate socialiste :

En donnant récemment la victoire aux socialistes, la Grande-Bretagne, un grand pays, vient de faire le choix de changer d'avenir. Pourquoi pas la France ? Mais changer l'avenir, c'est aussi changer la politique, changer la relation entre les élus et les citoyens. C'est pour cela que nous voulons limiter le cumul des mandats, assurer la parité hommes-femmes [...] ¹⁷.

En prenant exemple sur la Grande Bretagne, la candidate d'un parti que l'on rapprochera ensuite du « blairisme » (Marlière, 2007), Marie-Françoise Clergeau, énonce une nouvelle équivalence : changer l'avenir, c'est changer la politique, c'est à dire les modalités de la relation entre les élu.es et les citoyen.nes et non plus changer substantiellement les structures politiques, sociales ou économiques.

Malgré les glissements lexicaux (non plus « changer la vie » mais désormais « changer la politique»), l'IA reconnaît les structures phrastiques typiques des promesses de changement de la gauche de gouvernement et attribue convenablement le texte à la bonne classe.

4. L'après 2002

4.1. La droite décomplexée

Dans le contexte international de l'après 11 septembre et dans le contexte national de l'après 2002 et de la présence du Front national au 2^{ème} tour de l'élection présidentielle, le discours de droite se décomplexe (Fougier 2012 ; Mayaffre 2012) et dépasse même les bornes de l'identité historique de ce courant, par sa dérive sécuritaire. Les passages ayant le plus fortement activé la reconnaissance des professions de foi suggèrent l'influence de la rhétorique de Sarkozy sur les député.e.s de son parti, bien après la fin de son mandat. Par exemple en 2017, Marine Brenier propose de :

¹⁶ Eric Besson, 1997, Drôme.

¹⁷ Marie-Françoise Clergeau, 1997, Loire-Atlantique.

[...] promouvoir une France qui travaille : lutter contre l'assistanat et la fraude, baisser les charges, rétablir la défiscalisation des heures supplémentaires. Pouvoir d'achat : supprimer les droits de succession et plaider pour une baisse massive des impôts. Identité : défendre les valeurs de la république et nos traditions, et lutter contre le communautarisme¹⁸.

La candidate reprend ainsi les termes sarkoziens de promotion de la valeur travail (une « France qui travaille » fait sans doute écho au désormais célèbre « La France qui se lève tôt » de l'ancien président) ou encore de la préservation identitaire face au communautarisme.

Après 2001, la question sécuritaire, centrale à droite, trouve dans la lutte contre le « terrorisme islamique » un autre terrain d'expression, au côté de la lutte contre l'immigration illégale et le communautarisme, associées au sein d'énumérations de propositions :

Il faut enfin restaurer l'autorité de l'état, éradiquer le terrorisme islamique, l'insécurité, renforcer la sécurité et défendre notre identité en luttant fermement contre l'immigration illégale et le communautarisme¹⁹.

Notons que dans ces citations, outre le lexique spécifique de la doctrine sarkozyste, le réseau repère systématiquement l'usage de l'infinitif (Figure 5). Les candidat.es de la droite entre 2002 et 2017 adoptent massivement ce mode qui confère à l'exposition de leur programme un aspect impersonnel et intemporel et qui témoigne de l'évolution historique du genre même de la profession de foi, qui s'apparente de plus en plus au fil de la V^e République à un catalogue de mesures qu'à un engagement électoral substantiel (Bonhomme 2016).

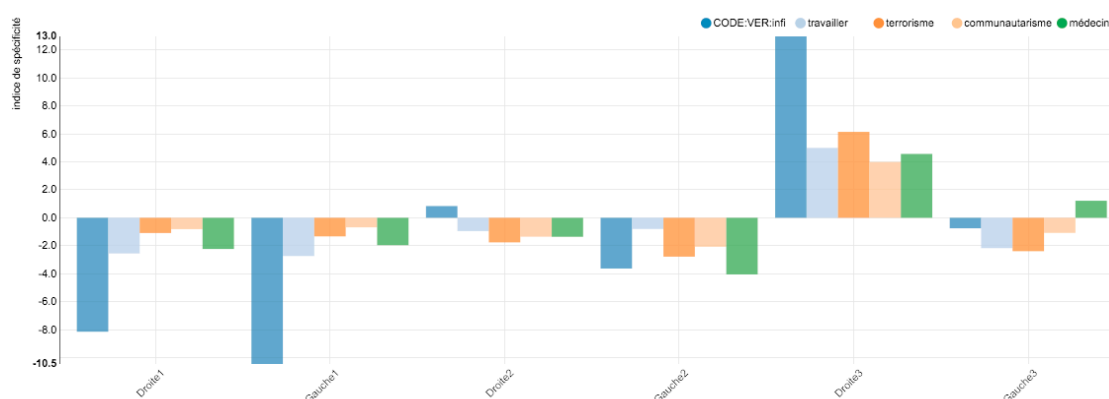


Figure 5 : Sur-utilisations grammaticales et lexicales à droite (2002-2017) : l'infinitif, les lemmes « travailler », « communautarisme », « médecin ».

Certains passages, à tonalité plus « sociale », sont également bien reconnus en raison de la co-présence de l'énonciation à l'infinitif et de lemmes renvoyant à l'électorat historique de la droite (les « médecins » par exemple); ce qui témoigne d'une argumentation électorale qui tend à s'axer vers les bases traditionnelles (médecins, ruraux), notamment en 2017 face à la concurrence de La République en Marche.

Soutenir les mesures d'équité entre tous les Français et tous les territoires. Favoriser l'installation des médecins dans nos territoires ruraux, encourager le regroupement en maisons médicales, améliorer la prise en charge des personnes âgées et en situation de handicap²⁰.

¹⁸ Marine Brenier, 2017, Alpes-Maritimes.

¹⁹ Philippe Vigier, 2007, Eure et Loir.

²⁰ Bérangère Poletti, 2017, Ardennes.

Le deep learning se révèle ici sensible à la combinaison syntagmatique de catégories grammaticales (les verbes à l'infinitif) et d'un lexique propre à la droite du début du XXI^e siècle.

4.2. La gauche des territoires

Les zones d'activation dans les discours de gauche à partir de 2002 suggèrent deux types d'évolution : d'une part, le passage d'une rhétorique collective (le « nous » encore en vigueur en 1997) à l'usage presque exclusif du « je » (Figure 6) et d'autre part, l'apparition notable de mots relatifs à la proximité et encore plus à l'ancrage territorial des politiques. Les candidat.es de la gauche social-démocrate cèdent en effet à la personnalisation et à l'exhibition de leur personne dans leur profession de foi. Ils espèrent renouer avec les habitants des circonscriptions par le développement d'une rhétorique de la « proximité », jamais vraiment définie et oscillant entre des dimensions affectives, sociales ou géographiques (Lebart et Lefèbre, 2005). La déconvolution pointe ainsi en sortie de réseau le vocabulaire de la proximité et le pronom personnel « je » dans la profession de foi de Chaynesse Khirouni :

Une députée proche de vous. Candidate du parti socialiste et soutenue par le parti radical de gauche, j'incarne le changement des pratiques politiques, la proximité et le renouvellement²¹.

On pourra dire que la tension entre la proximité et la distance, entre la circonscription et le pays, entre le territoire et la nation est constitutive du mandat de député.e et que les candidatures oscillent d'un pôle à l'autre au gré de l'histoire. Reste que, à partir de 2002, la question de l'ancrage territorial dépasse la présentation de soi des candidat.es pour devenir centrale dans les développements programmatiques de la gauche du XXI^e siècle (Mayaffre, 2020 sous presse). A la « justice sociale » et à la « justice fiscale » s'ajoute la « justice territoriale », comme dans l'engagement de la candidate socialiste Pascale Crozon en 2012 pour :

Rétablir la justice, et d'abord la justice sociale (retraite à 60 ans pour ceux qui disposent de leurs annuités, égalité salariale femmes-hommes, revalorisation de 25 % de l'allocation de rentrée scolaire), mais aussi la justice fiscale avec une tranche d'imposition de 75 % sur les revenus supérieurs à un million d'euros, et territoriale (nouvelle décentralisation, défense des services publics de proximité – école, santé, justice, sécurité)²².

L'adjectif « territorial », très fréquemment activé dans les réseaux de neurones, vient ainsi de plus en plus souvent qualifier les concepts politiques comme l'indique, d'un point de vue statistique la figure 6 qui confirme la sur-utilisation de la construction « NOM + LEM:territorial » à gauche entre 2002 et 2017.

²¹ Chaynesse Khirouni, 2012, Meurthe-et-Moselle.

²² Pascale Crozon, 2012, Rhône.

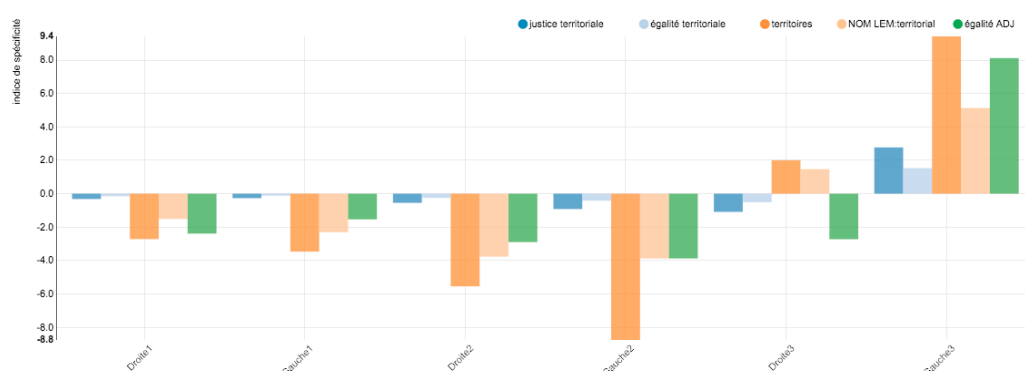


Figure 6 : Sur-utilisation de « territoires », « NOM + LEM:territorial » et de « égalité ADJ » dans le corpus de la gauche 3 (2002-2017).

Le substantif « territoires » - le pluriel soulignant la diversité (et non l'unité du territoire national) - est également souvent activé. Les « territoires » sont devenus le cadre des politiques d'une nouvelle gauche qui semble se référer de moins en moins aux luttes sociales, aux oppositions de groupes sociaux qu'à une égalité et une justice pensées dans l'espace géographique. Quelques futur.es élu.es défendent le retour de services publics sur l'ensemble des territoires conformément aux présupposés d'une République pensée comme une et indivisible, à l'image de celle Gneugneau Edith qui propose de :

Retrouver des services publics accessibles sur tous les territoires : L'égalité territoriale est une nécessité pour que chaque citoyen, où qu'il vive, ait le même accès aux services publics et la capacité de réussir sa vie²³.

Beaucoup d'autres eux en revanche se déclarent en faveur d'un « nouvel acte de décentralisation » et d'un « renforcement des collectivités territoriales » qui laisserait la place à des « solidarités entre territoires » et à des « mesures d'équité entre territoires urbains et territoires ruraux ». Fort marqueur de distinction entre la gauche et la droite, l'« égalité » (l'équité ou la solidarité) se pense désormais sous l'angle territorial. C'est ce changement de paradigme politique profond que la déconvolution donne à voir à l'analyste en soulignant l'activation de collocations telles que « NOM + LEM:territorial » ou « égalité + territoriale ».

Conclusion

Pour conclure, on insistera sur l'une des plus-values des algorithmes d'apprentissage profond pour la description des textes politiques. L'ouverture, par la déconvolution, des couches cachées des réseaux de neurones met au jour des unités bien connues de la statistique textuelle (telles que les spécificités ou les co-occurrences lexicales ou grammaticales). Mais elle permet également de saisir formellement des unités plus complexes, que l'analyse des données textuelles a souvent conceptualisées sans parvenir à les repérer automatiquement, tels que les motifs - c'est à dire des patrons lexico-grammaticaux auxquels on peut attribuer une fonction de marqueurs discursifs structurants (Mellet & Longrée 2012, Longrée & Mellet 2013)- ou les passages - c'est à dire des unités continues ou discontinues interprétables au sein d'un empan textuel donné (Rastier 2007). Car la représentation du texte permise par la déconvolution dépasse l'approche fréquentielle historique de l'ADT pour une démarche plus séquentielle et contextualisée. Dans le modèle de la déconvolution, le taux d'activation d'un

²³ Edith Gueugneau, 2012, Saône et Loire.

mot (le TDS) varie selon sa position dans les différents passages ; ce qui confirme qu'il dépend des relations linguistiques du mot avec d'autres unités textuelles de l'empan considéré. Par exemple, ici, le deep learning a été capable de reconnaître les deux sens majeurs de la « paix » ou du « changement » dans le discours électoral sous la V^e République en fonction des autres mots qui composaient la fenêtre textuelle. C'est bien le mot replacé sur son axe syntagmatique, c'est à dire dans son co-texte phrastique immédiat, qui active le réseau. Par là, les nouvelles possibilités descriptives du deep learning, associées aux méthodes éprouvées de l'ADT, permettent d'alimenter la recherche de nouveaux observables linguistiques pour des protocoles de lecture renouvelés.

Références

- Barats-Malbré C. (1999). Immigration : carrefour de la suspicion (discours présidentiels et juridiques). *Mots. Les langages du politique*, 60(1), 43-58.
- Bonhomme, M. (2016). Les professions de foi à l'occasion des élections présidentielles françaises. *Mots. Les langages du politique*, 112(3), 21-27.
- Ducoffé M. et al. (2016). Machine learning under the light of phraseology expertise : use case of presidential speeches. De Gaulle-Hollande (1958-2016), In *Actes des JADTs*, 155-168.
- Fougier E. (2012). What is Sarkozysm ? The Sarkozy Revolution of the French Right-wing. *L'Europe en Formation*, 365(3), 217-228.
- Guaresi M. (2018). *Les professions de foi des députés.e.s sous la Cinquième République (1958-2007)*, Paris. L'Harmattan.
- Lebart C. et Lefebvre R. (2005). *La proximité en politique. Usages, rhétoriques, pratiques*. Rennes. Presses universitaires de Rennes.
- Longrée D. et Mellet S. (2013). Le motif : une unité phraséologique englobante ? Etendre le champ de la phraséologie de la langue au discours. *Langages*, 189, 2013, 65–79.
- Marlière, P. (2007). Retour sur une décennie de blairisme à la française. *Nouvelles Fondations*, 7-8(3), 14-19.
- Mayaffre D. (2012). *Nicolas Sarkozy: Mesure et démesure du discours (2007-2012)*. Paris. Presses de Sciences Po.
- Mayaffre D. (2020 sous presse). *Emmanuel Macron par l'Intelligence artificielle. Ses discours décryptés par la machine*. Paris. L'Aube.
- Kim Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 1746-1751.
- Rastier F. (2007). Passages. *Corpus*, 6 : 25-54.
- Salem A. (1991). Les séries textuelles chronologiques. *Histoire & Mesure*. 6(1/2), 149-175.
- Vanni L et al. (2018). Text Deconvolution Saliency (TDS) : a deep tool box for linguistic analysis. In *56th Annual Meeting of the Association for Computational Linguistics (ACL)*, Melbourne
- Vanni L, Mayaffre D. et Longrée D. (2018). ADT et deep learning, regards croisés. Phrases-clefs, motifs et nouveaux observables. In D. Iezzi et al. (dir.) *JADT' 2018*, UniverItalia, Rome, pp. 459-466.
- Vanni L. et al. (2020-submitted). From text saliency to linguistic objects: learning linguistic interpretable markers with multi-channels convolutional architecture. *57th Annual Meeting of The Association for Computational Linguistics*.
- Vavasseur-Desperriers J. (2006). *Les droites en France*. Paris. Presses Universitaires de France.