



HAL
open science

Intelligence artificielle et discours politique. Quelles plus-values interprétatives? Application aux corpus parlementaire et présidentiel contemporains

Magali Guaresi, Damon Mayaffre

► To cite this version:

Magali Guaresi, Damon Mayaffre. Intelligence artificielle et discours politique. Quelles plus-values interprétatives? Application aux corpus parlementaire et présidentiel contemporains. L'intelligence artificielle des textes. Des algorithmes à l'interprétation, 17, Honoré Champion, pp.131-182, 2021, Lettres numériques, 9782815937467. hal-03347997

HAL Id: hal-03347997

<https://hal.science/hal-03347997v1>

Submitted on 17 Sep 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.

INTELLIGENCE ARTIFICIELLE ET DISCOURS POLITIQUE

QUELLES PLUS-VALUES INTERPRÉTATIVES ? APPLICATION AUX CORPUS PARLEMENTAIRE ET PRESIDENTIEL CONTEMPORAINS

Magali Guaresi et Damon Mayaffre

INTRODUCTION

En France, dès l'origine à la fin des années 1960 et au début des années 1970, l'analyse du discours politique insiste sur la nécessaire « compétence idéologique »¹ du locuteur. Par cette reprise explicite du concept linguistique de « compétence », elle affirme l'importance du langage dans notre rapport au monde et devient partie prenante du *linguistic turn* qui caractérise, dans les SHS, l'épistémologie de la période. Cependant, par le détournement du concept de Chomsky du côté de l'idéologie, elle signifie l'insuffisance d'une approche linguistique formelle des textes politiques : la dimension politique qui informe nos corpus se situe précisément là où le sens interprété divorce de la signification prétendument établie par la langue ; là où la phrase grammaticale et la proposition vériconditionnelle sont impuissantes pour dire la complexité des rapports sociaux et l'Histoire ; là où la grammaire universelle ou la « compétence linguistique » proprement dites ne peuvent épuiser le monde, le langage et les textes.

En France, l'analyse du discours politique s'épanouit ainsi en insistant en même temps sur notre rapport nécessairement langagier au monde et sur la nature idéologique des discours. Le cadre général est certes dominé par le matérialisme marxiste mais Alphonse Dupront

¹ Denis Slatka, « Esquisse d'une théorie lexico-sémantique pour une analyse d'un texte politique (les cahiers de doléances) », *Langages*, n° 23, 1972.

reconnaît rapidement « l'autonomie du niveau langagier »¹. L'analyse du discours politique prend alors avec les travaux de Jacques Guilhaumou une posture herméneutique fondamentale face à l'archive politique et aux corpus textuels. Les mots n'ont pas de sens en eux-mêmes, pour eux-mêmes, et encore moins de référents isomorphes dans le monde. Les mots sont eux-mêmes contextes, acteurs, enjeux, « pistolets chargés » comme disait Sartre ; idéologèmes. Et le sens global qui est en jeu dans les mots utilisés dans un texte politique peut sceller le sort d'une Révolution, d'une République, d'une Commune, de la guerre d'Algérie ou de Mai 68 pour faire allusion aux corpus brûlants que l'AD politique traite à ses débuts². Il n'y a ni transparence du langage, ni « déjà-là » sémantique dans un texte politique, mais toujours interprétation, configuration et reconfiguration du monde par le langage aussi bien de la part du locuteur politique qui s'exprime – c'est pour cela qu'il prend la parole : pour changer le monde ! –, que pour l'auditoire qui écoute, que pour l'analyste qui analyse. Jacques Guilhaumou peut ainsi poser définitivement que « toute description discursive est déjà interprétation du social » et que « le contexte d'un énoncé n'est pas étranger à sa description même ». Il situe l'analyse du discours politique « à l'intérieur du paradigme herméneutique » et définit l'analyse du discours « comme une discipline interprétative à part entière »³. Et dans l'interprétation de ce *texte-contexte* ou de ce *langage-monde*, les compétences linguistique et idéologique sont toutes deux nécessaires.

Si un détour introductif par les origines de l'analyse du discours politique nous a paru nécessaire, c'est que l'Intelligence artificielle (IA) des textes politiques pose aujourd'hui les questions que Jean Dubois,

¹ Alphonse Dupront cité par Régine Robin, « Les historiens devant la linguistique. Nouveau gadget ou constitution de l'objet discursif ? », *Dix-Huitième Siècle*, n° 5, 1973, p. 112.

² Voir les thèses de Jean Dubois autour de la Commune, *Le vocabulaire politique et social en France de 1869 à 1872. À travers les œuvres des écrivains, les revues et les journaux* (1962) ; de Denise Maldidier sur la guerre d'Algérie, *Analyse linguistique du vocabulaire politique de la guerre d'Algérie d'après six quotidiens parisiens* (1969) ; de Maurice Tournier sur 1848, *Un vocabulaire ouvrier en 1848, essai de lexicométrie* (1976) ; de Jacques Guilhaumou sur la Révolution, *Idéologies, discours et conjoncture. L'exemple des discours révolutionnaires* (1978). Sur Mai 1968, on se reportera au livre fondateur de la lexicométrie politique, M. Demonet *et al.*, *Des tracts en mai 68. Mesures de vocabulaire et de contenu*, Paris, Colin, 1978.

³ L'ensemble des citations est tiré de J. Guilhaumou, « À propos de l'analyse de discours. Les historiens et le 'tournant linguistique' », *Langage et Société*, 1993, p. 6-7.

Michel Pêcheux, Régine Robin, Antoine Prost ou Maurice Tournier ont égrené dès le début des années 1970 puis tout le long des années 1980 ou 1990 ; notamment *l'inquiétude méthodologique*, c'est-à-dire le protocole mis en place face au corpus. Nous voulons donc ici (ré)-interroger les *conditions méthodologiques de l'interprétation des textes politiques*.

Dès lors qu'il s'agit d'aller plus loin qu'une lecture intuitive, le retour d'expérience de nos devanciers en matière d'analyse assistée par l'ordinateur – de l'Analyse Automatique du Discours (AAD) de Michel Pêcheux à l'Analyse de contenu de Laurence Bardin en passant, centralement pour nous, par la lexicométrie politique de Saint Cloud – doit en effet permettre de répondre, plus vite et mieux, à l'exigence méthodologique qui domine toute pratique scientifique. Concrètement, dans ce chapitre nous posons la question *des plus-values attendues du traitement des corpus politiques par les réseaux de neurones artificiels*.

Allons donc à l'essentiel. L'Intelligence artificielle, en l'occurrence le *deep learning*, ne saurait constituer pour nous la dernière traduction d'un positivisme qui se réinvente à l'occasion de chaque avancée technologique.

Nous postulons que le sens – particulièrement le sens nécessairement polémique d'un discours politique – n'est pas une donnée positive du texte mais un artefact ou un construit (ou encore un co-construit par la triade auteur-lecteur-analyste). En conséquence, nos méthodes de traitement ne peuvent avoir une prétention probatoire (prouver le sens ?) mais ont une vocation heuristique (chercher un sens !). Il s'agit d'objectiver des parcours de lecture et non de démontrer le sens, prétendument littéral, que l'on aurait préalablement essentialisé ou naturalisé dans le texte. Si l'on ne renonce, ni à la méthode ni à l'objectivation, puisqu'elles sont synonymes de science, celles-ci concernent l'établissement de parcours interprétatifs contrôlables, reproductibles, semi-automatisables, crédibles ; et non la réification d'une vérité sémantique supposément inscrite dans le marbre, le papier ou le fichier numérique, que la machine pourrait prétendre formaliser¹.

Concrètement, dans le quotidien des chercheurs, et au cœur de cette contribution, on ne saurait demander à l'ordinateur – pas plus

¹ Entendons-nous néanmoins : il ne s'agit pas non plus de retrouver un sens crypté, comme dans les allégorèses antiques ou religieuses. Le numérique doit être une plus value interprétative, pour une herméneutique matérielle, mais pas parcours initiatique.

aujourd'hui à l'IA qu'hier à la Logométrie – de produire le résumé machinal d'un texte, d'extraire automatiquement le sens d'un discours, de générer mécaniquement des Serments de Strasbourg ou une Déclaration des droits de l'Homme. Il s'agit simplement, dans une modeste prétention, de mettre au jour des clefs ou chevilles interprétatives fécondes, si possible inédites, sans présumer de leurs formes linguistiques (lettres, morphèmes, mots, syntagmes, étiquettes morpho-syntaxiques, segments répétés, cooccurrences, motifs lexico-grammaticaux, etc.). À l'usage, ces chevilles interprétatives se révéleront être des « passages » au sens de [Rastier 2007] : des zones de textes, de taille variable, non nécessairement contiguës, qui articulent, à l'occasion d'un effort de contextualisation, expression et contenu, et sur lesquelles nous pourrions construire du sens ; des herménèmes¹. Techniquement, il s'agira possiblement de « motifs » au sens de [Longrée et Mellet 2013] : des observables linguistiques, pas simplement élémentaires comme des mots, mais complexes (possiblement lexico-grammaticaux) caractérisant et structurant un texte.

Après avoir rappelé le type de représentations néo-connexionnistes du texte que l'Intelligence artificielle autorise et, dès lors, les types de parcours de lecture qu'elle nous permet d'envisager (1. *Deep learning et textes politiques profonds. Cadrage*), nous nous arrêterons sur deux apports de l'IA des textes politiques promis au meilleur avenir. La description des textes d'abord : sur un jeu d'apprentissage important de discours parlementaires de 1958 à 2017, nous décrirons le discours de gauche et de droite en diachronie. Ce faisant, nous mettrons à l'épreuve les députés En Marche lors de la législature 2017 quant à l'identité politique de leurs discours (2. *Vers une description des discours gauche / droite. Les professions de foi des députés sous la V^e République*). L'explicitation de l'intertextualité ensuite : sur le jeu d'apprentissage des discours élyséens depuis le début de la V^e République jusqu'à aujourd'hui (1958-2020) – discours jugés homogènes quant aux

¹ La définition des « passages » fait l'objet d'un article [Rastier 2007], puis d'un ouvrage [Rastier 2011]. Rappelons le résumé simple que l'auteur donne en note infrapaginale : « Le passage est une grandeur [du texte] établie ou reconnue par l'analyse, non une unité au sens logico-grammatical, qui suppose une délimitation stricte et constante. » [Rastier, 2007, § 6 note 4). On trouvera également une définition résumée §24 : « le passage est une *zone de localité*, définie par une sémiosis propre (mode d'appariement entre contenu et expression) et, sur chacun de ses plans (fragment et extrait) par des relations contextuelles internes fortes ».

conditions d'énonciation et au genre – nous analyserons cette parole autre qui traverse la parole de Macron, l'actuel président. L'IA est-elle capable de repérer les empreintes cachées ou explicites de de Gaulle ou de Mitterrand dans les discours d'Emmanuel Macron ? Est-il possible de pressentir les emprunts – à défaut de plagiats – que Macron consent à Pompidou, Sarkozy ou Hollande ? L'intertexte – notion fuyante s'il en est – peut-il être approché, sinon objectivé, par l'Intelligence artificielle ? (3. *Vers une objectivation de l'intertexte. Les emprunts de Macron à ses prédécesseurs à l'Élysée*).

1. **DEEP LEARNING ET TEXTES POLITIQUES PROFONDS. CADRAGE**

La linguistique textuelle dans sa version anglaise avec Halliday ou Sinclair ou sa version française avec Adam ou Rastier procède par co(n)textualisation, et il se trouve que l'Intelligence artificielle des textes, particulièrement dans les modèles convolutionnels (CNN), implémente cette exigence, en explorant systématiquement les unités du corpus dans leurs relations avec les autres.

L'Intelligence artificielle relève en effet du connexionnisme en proposant une modélisation mathématique, obscure dans ses détails (et qu'il nous reste donc à éclaircir), de la réticularité textuelle, c'est-à-dire des mots en co(n)texte, des mots en réseau, des mots en bonne « intelligence »¹ avec eux-mêmes au sein du texte ou du corpus. De manière schématique, l'idée fondamentale de l'algorithmique de l'IA est qu'un mot est un neurone et que des relations synaptiques entre ces mots-neurones finissent par faire réseau, par faire texte, par faire sens.

Conformément à son étymologie, le texte n'est plus seulement une « suite » comme dans la linguistique traditionnelle, pas plus qu'il ne se réduit à un « sac de mots » ou à une « urne » aux boules indépendantes comme dans la statistique textuelle : le texte (re)devient un tissu, un tissage, une toile, une grappe (cluster) ou un réseau.

Soulignons immédiatement deux choses sous 1.1. et 1.2.

¹ On se rappelle l'étymologie du mot intelligence. *Intellegere* composé du préfixe inter- (« entre ») et du verbe legere (« lire, cueillir, choisir »). L'intelligence d'un texte consisterait donc à lire ses mots en s'efforçant de les relier entre eux ; voire, nous le verrons, de les relier à d'autres mots et d'autres textes extérieurs (intertextualité).

1.1. CONVOLUTION, CO-TEXTE, COOCCURENCES

La nature des relations synaptiques entre les mots-neurones, qui forment ainsi réseau ou texte, reste en partie aujourd'hui à définir linguistiquement¹. Les relations les plus évidentes que la machine met au jour sont les collocations ou les cooccurrences : deux mots-neurones entretiennent des liens-synapses pour le modèle à partir du moment où ils figurent dans une même fenêtre syntagmatique du texte. Un mot-pôle peut être ainsi facilement défini par l'univers co-textuel dans lequel il gravite : l'ensemble de ses cooccurrents avec qui il entretient un lien géographique dans la phrase ou le paragraphe [historiquement Firth 1957 ; puis de nombreux auteurs dont Brunet 2012]. Si, à la suite de [Guiraud 1954] nous définissons le sens d'un mot comme la somme de ses co(n)textes d'utilisation, et que nous précisons que la forme minimale mais calculable du co(n)texte est la cooccurrence, alors la somme des cooccurrents d'un mot en donne une bonne approximation sémantique [Guaresi 2015 ; Mayaffre 2008]. La généralisation du traitement des cooccurrences (le croisement de chaque mot du texte avec tous les mots du texte, pour mesurer systématiquement les liens plus ou moins forts qu'ils entretiennent) permet de représenter le réseau textuel ou la texture dans son ensemble. Ainsi en France, à la suite de [Viprey 1997], de nombreuses illustrations factorielles ont été données de la matrice cooccurrentielle Mots x Mots. À partir d'autres modèles mathématiques que ceux de l'AFC, des logiciels de clustering comme Gephi permettent également de calculer et visualiser l'entrelacs textuel [Ben Hamed et Mayaffre 2014]. Dans le domaine de l'IA, enfin, l'*embedding* proposé par *Word2Vec*, internationalement utilisé, implémente efficacement ce principe de fond. Précisément, dans cette contribution, les tokens du texte ont une valeur (*embedding*) ou une représentation, données selon leurs affinités cooccurrentielles : c'est la matrice essentielle de départ.

Cependant, au-delà des relations statistiques entre les mots dans un empan ou une fenêtre (cooccurrences donc) ci-dessus évoquées, bien connues et fondamentales, les réseaux de neurones artificiels semblent

¹ En effet aucune règle linguistique n'est implémentée au départ dans le programme. Par exemple, nous ne programmons pas de règle syntaxique avant de demander au *deep learning* de découvrir les relations qui existent entre les mots ; pas plus nous ne définissons d'ontologies qui pourraient classer les mots. En conséquence, l'heuristique de la démarche consiste à comprendre la nature des relations mises au jour par la machine.

remarquer d'autres types de relations qui tissent les mots entre eux pour former le texte : relation de nature distributionnelle notamment, au sens harrisien [Harris 1952/1969]. La machine pourra ainsi comprendre la relation d'équivalence qui existe entre « peuple » et « parti » dans le discours communiste, au motif que leurs occurrences respectives apparaissent dans un environnement distributionnel identique : *le peuple doit se mobiliser/réunir/rassembler* ou *le parti doit se mobiliser/réunir/rassembler* ; *il faut l'unité/union du peuple / il faut l'unité/union du parti*. Le distributionnalisme n'est pas étranger à l'approche cooccurentielle (ici « peuple » et « parti » ne sont certes pas directement cooccurents entre-eux, mais ils partagent strictement les mêmes cooccurrences), mais le distributionnalisme introduit d'autres éléments comme l'enchaînement syntagmatique strict des mots ou leur nature et dépendance morpho-syntaxiques après étiquetage. Prometteur voire fondateur au début des années 1970 pour l'AD [Harris traduit dans *Langages* 1969], puis quasi abandonné dans les années 1990-2000, le distributionnalisme trouve dans l'IA, aujourd'hui, une forme de renaissance à laquelle les praticiens pourraient réfléchir.

1.2. RÉSEAUX, COUCHES CACHÉES, COMPLEXITÉ

Le modèle connexionniste de l'IA est complexe, au point de devenir humainement inintelligible : c'est ici que se situe la marge de progrès et le programme de recherche. Le réseau est composé de plusieurs couches, qualifiées de couches cachées ou de couches profondes (*deep*), qui s'empilent jusqu'à rendre difficile la description, et fragile l'édifice de l'interprétation. Le réseau de neurones artificiels met en effet non seulement en relation synaptique, à un premier niveau évident, les mots-neurones entre-eux (lien synaptique de premier ordre notamment sur une base cooccurentielle). Mais, considérant ces liens, leur poids et leur valeur comme une nouvelle représentation du texte, le réseau va mettre en relation, à un deuxième niveau, ces liens premiers entre eux : un réseau du réseau ou une représentation de la représentation ; des liens synaptiques entre synapses (et non plus seulement entre neurones). Et les couches vont ainsi s'enchaîner, dix fois, cent fois s'il le faut, pour définir un connexionnisme dont l'épaisseur garantit l'opacité pour le linguiste : c'est ce phénomène de boîte noire qu'il convient d'essayer d'éclaircir. À ce stade, convenons donc que le rapport entre le « réseau textuel » (composé de mots-neurones reliés, et

point de départ de l'analyse) et le « réseau neuronal » de l'IA (composé de liens synaptiques, de poids et de couches cachées) est un rapport évident au départ (nous partons bien du texte, de ses mots-neurons et du réseau élémentaire que ces mots forment entre eux) mais qui devient rapidement analogique (nous avons vite affaire à des réseaux de réseaux, ou réseaux de synapses dans lesquels l'information linguistique initiale, le mot ou la cooccurrence par exemple, n'est plus évidente).

Quoi qu'il en soit, notwithstanding les limites théoriques et empiriques de l'IA, notwithstanding l'opacité aujourd'hui des algorithmes¹ que cet ouvrage essaye d'éclairer (cf. particulièrement chapitre 1), l'idée fondamentale reste de modéliser le texte comme un réseau *c'est-à-dire comme un parcours de lecture* qu'il s'agit de décrire ou de parcourir à partir (i) de ses unités matérielles discrètes (le *token*, le lemme ou l'étiquette morphosyntaxique attachée à ce *token*), mais (ii) sous conditions que ces unités soient « reliées » ou co(n)textualisées (le token en relation cooccurrentielle ou distributionnelle avec son environnement textuel proche ou lointain ; en fonction de sa représentation textuelle dans le corpus). Au niveau supérieur, l'objectif est de dégager des structures profondes (*deep*) ou sous-jacentes du texte, et de faire émerger de nouveaux observables linguistiques (motifs complexes, zones ou grandeurs textuelles à définir, « formes sémiotiquement saillantes » [Rastier 2007 : §82], passages non obviés) qui auraient pu échapper à la lecture naturelle.

1.3. DU SENS EN CONTEXTE

L'effort de co(n)textualisation, mise en réseau ou mise en résonance des unités du texte, qui, nous l'avons dit, renvoie à l'étymologie d'intelligence, ne doit pas être minimisé, parce qu'il s'opère à trois niveaux imbriqués dans le traitement : micro, méso, macro.

— Au niveau micro ou local, c'est le co-texte *stricto sensu* de l'unité qui est considéré dans les modèles convolutionnels (CNN) : grâce à une fenêtre coulissante de n mots – nous avons choisi pour cette contribution une fenêtre de convolution de 6 mots – le texte est balayé,

¹ Outre la multiplication des couches cachées, l'opacité du traitement est garantie par la *backpropagation*. Celle-ci consiste à modifier la représentation initiale des mots dans l'*embedding* (pour nous, les valeurs cooccurentielles du mot qui servent de matrice de départ) pour permettre à la machine d'optimiser ses décisions.

afin que chaque mot soit pris dans son environnement ou plutôt ses environnements syntagmatiques essentiels et immédiats. Ici on remarquera que le mot est successivement considéré en fonction des 5 mots qui le précèdent, puis considéré en fonction des 4 mots qui le précèdent et d'1 mot qui lui succède, puis des 3 mots qui le précèdent et des 2 mots qui lui succèdent ; *et caetera* pour tous les mots du texte jusqu'à son épuisement. On le comprend, l'approche cesse alors d'être seulement paradigmatique et discrète comme souvent dans l'approche informatique (tokenisation puis *sélection* de tel ou tel mot en fonction de sa fréquence, de son étiquette morphosyntaxique, de sa lettre à l'initiale, etc). L'approche devient également syntagmatique et continue (cotextualisation glissante et *combinaison* des mots entre eux par collocation ou convolution). L'idée fondamentale est que nous ne considérons pas le mot isolé, pour lui-même, mais que nous en donnons des représentations différentes selon ses co-textes immédiats¹.

— Au niveau méso ou régional, le corpus est divisé en sections ou fenêtres ou empanns – ici de 100 mots. C'est à ce niveau de granularité ou de co(n)textualisation que l'algorithme repère les régularités, les combinaisons, les saillances, l'identité linguistique d'un auteur ou d'une œuvre. Ce repérage se fait en fonction des composants que nous venons d'identifier au niveau micro, et de leur combinaison, dans ces sections : les tokens dans leur co-texte local immédiat, combinés entre-eux, dans la section régionale. Donc, nous analysons, dans les sections ou empanns de 100 mots, la présence, absence, fréquence et la combinaison (ou cooccurrences) de tokens eux-mêmes co-textualisés dans une fenêtre de convolution glissante de 6 mots. Dans une tâche de classification, chaque section de 100 mots pourra ainsi être attribuée à un auteur, un genre, une époque selon sa composition linguistique et selon la tâche de l'analyse. Et la fiabilité de cette attribution (classification ou *prediction*) sera dûment calculée (taux *d'accuracy*). Dans une tâche de description et d'interprétation, les composants linguistiques bien appris par le réseau, c'est-à-dire responsables de la bonne attribution du paragraphe à son auteur, seront exhibés : l'indice

¹ Concrètement, dans l'infinité des exemples possibles, le mot « homme » n'a plus la même « valeur » pour le système, c'est-à-dire la même représentation numérique, dans les énoncés « la déclaration des droits de l'homme est fondatrice » et « l'homme et la femme sont égaux par nature ». À strictement parler, le système ne considère plus que « homme » et « homme » soient le même mot. Il n'y a plus dès lors d'identité du signe et d'isonomie, mais seulement une valeur différentielle, en contexte.

intitulé *Text Deconvolution Saliency* (TDS) [Vanni *et al.* 2018] montre en effet les « zones d'activation » du réseau, c'est-à-dire les observables linguistiques (simples tokens, collocations, cooccurrences, motifs complexes) significatifs. Ces observables représentent, à nos yeux, des « grandeurs textuelles » pertinentes, des « formes sémiotiquement saillantes » ou des « passages » définis par [Rastier 2007] : ils sont décisifs pour l'interprétation contrôlée des textes ou pour l'intelligence (artificielle) des textes. Dans le détail, « passages » en eux-mêmes, ces « zones d'activation » ou observables linguistiques sont susceptibles, lorsqu'ils sont concentrés dans une même fenêtre (empan de 100 mots), d'ériger celle-ci en parangon, en passage-clef, en passage-idéal ou passage-exemplaire [Rastier 2007 et à sa suite Vanni, Mayaffre et Longrée 2018].

— Enfin au niveau macro ou global, c'est tout le texte ou plutôt tout le corpus qui est traité minutieusement malgré sa taille. C'est ici le postulat de la sémantique de corpus et un avantage repéré de longue date dans tout traitement informatique des textes. Le corpus que la machine est capable d'embrasser est la forme objectivable maximale du contexte ; au delà de lui, nous entrons dans le hors-corpus et tombons dans un magma textuel certes passionnant mais qui externalise l'analyse. Le corpus numérique est un hypertexte ou un architexte dans lequel les textes du corpus prennent sens. Il est le global (global statistique notamment) qui détermine le local. Il constitue la norme ou le fond par rapport auxquels le texte et ses unités se détachent, c'est-à-dire prennent sens. Les régularités par récurrence, les contrastes, les singularités et ici « l'apprentissage » des motifs linguistiques d'un auteur ou d'un genre se font toujours par rapport à un corpus donné. Les vieilles notions de corpus contrastifs et de norme endogène, et plus encore la notion récente de *corpus réflexif* prennent ainsi une dimension décisive dans le parcours interprétatif (Rastier 2001 ; Mayaffre 2002). Le corpus est la matérialisation de l'horizon d'attente de Gadamer ou [Jauss 1990] et, nous le verrons dans la troisième partie, il peut être considéré comme l'intertexte des textes constituants – un certain intertexte, parmi d'autres évidemment ; et charge à la machine de retrouver les emprunts qu'un auteur fait à cet intertexte, les marques qu'il puise dans cet intertexte, c'est-à-dire l'empreinte globale du corpus sur le texte analysé : la détermination du local par le global ; la compréhension du micro et du méso dans ou par le macro.

Pour conclure ce cadrage, ajoutons que le dialogue que nous avons entrepris entre, d'un côté le traitement statistique des éléments du discours que propose depuis une cinquantaine d'années l'ADT et, de l'autre côté l'Intelligence artificielle, nous permet de montrer que si le *fréquentiel* reste important dans le *deep learning* (comme il est consubstantiel à l'ADT), l'IA enrichit l'approche statistique par une démarche *séquentielle*. L'unité n'est plus seulement comptable/comptée mais contextualisable/contextualisée; elle n'est plus discrète et indépendante, mais continue et dépendante de son environnement syntagmatique court (fenêtre convolutionnelle de 6 mots) et allongé (la section de 100); au plus haut niveau, elle est dépendante de l'ensemble du corpus. La co(n)textualisation fait ainsi le pendant à la tokenisation. Nous utiliserons dans cette contribution le modèle convolutionnel (CNN) qui explore par définition le corpus par fenêtres coulissantes, mais les modèles RNN (*Recurrent neural network*) considèrent également, d'une autre manière, le texte séquentiellement, et le co(nt)exte dans sa dimension linéaire. Dans les modèles RNN de type LSTM (*Long short-term memory*) chaque mot porte en mémoire le mot ou les mots qui le précède(nt). Et qu'il nous soit permis d'entrer superficiellement dans la technique RNN pour souligner que l'opération se fait possiblement de gauche à droite comme de droite à gauche, puisque la construction du sens n'est pas en sens unique: chaque mot porte la mémoire non seulement des mots pré-posés mais des mots post-posés. Dans la phrase « l'esprit des lois c'est la propriété privée », la mémoire, c'est-à-dire la représentation numérique, de « lois » contient non seulement « l'esprit des... » situé en amont mais aussi « ... c'est la propriété privée » situé en aval.

2. VERS UNE DESCRIPTION DES DISCOURS GAUCHE / DROITE. LES PROFESSIONS DE FOI DES DÉPUTÉS SOUS LA V^e RÉPUBLIQUE

L'Intelligence artificielle et en particulier les méthodes du *deep learning* dépassent donc les simples tâches de classification ou de prédiction pour enrichir la description des textes. Les réseaux de neurones convolutionnels, ici implémentés dans le logiciel Hyperbase, sur lesquels s'appuie cette contribution, ne se contentent pas d'attribuer un texte à son auteur avec de bons scores de réussite mais permettent d'observer les items linguistiques qui ont présidé à l'apprentissage

profond puis à la bonne prise de décision : c'est la déconvolution [Kim 2014 ; Vanni, Mayaffre et Longrée 2018].

Dans cette section, nous proposons d'appliquer le protocole de déconvolution sur un jeu de professions de foi électorales, rédigées par les élus des législatives entre 1958 et 2017, dans la perspective de décrire les évolutions du discours électoral de gauche et de droite au fil de la V^e République¹. Pour ce faire, un entraînement a été mené sur le corpus de proclamations électorales, regroupées selon l'appartenance partisane (gauche/droite)² et la périodisation chronologique. Une chronologie endogène a été préalablement établie grâce à une analyse factorielle des correspondances (AFC) menée sur notre corpus. Celle-ci a permis de faire émerger de cette *série textuelle chronologique* [Salem 1991] six grandes scansion temporelles : 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 structurées, la machine apprend de manière autonome à reconnaître le discours de la droite et de la gauche à chacune des périodes déterminées. Dans un deuxième temps, elle nous restitue *les zones d'activation des textes*, c'est-à-dire les passages remarquables qui ont présidé à la classification, susceptibles de guider l'interprétation historico-linguistique des discours des partis de l'ordre et du mouvement. Enfin, nous lui présentons les professions de foi des élus de la République En Marche, qui n'ont pas participé à la phase d'entraînement, de façon à ce que l'algorithme les classe sur l'échiquier politique et nous donne à lire les éléments linguistiques apparentés à la gauche et à la droite dans les discours de la majorité.

Chemin faisant, nous interrogerons les vertus descriptives et interprétatives des unités du texte mises au jour par l'IA et le protocole

¹ 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. Cette analyse s'inscrit ainsi dans la lignée des travaux d'Antoine Prost sur les professions de foi de la fin du XIX^e siècle [Prost 1974].

² 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 élus. 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 : précisément, ils seront soumis à l'analyse.

de déconvolution, en les croisant avec des résultats issus des méthodes éprouvées de l'analyse de données textuelles.

2.1. LA GAUCHE ET LA DROITE DÉCRITES PAR LE *DEEP LEARNING* (1958-2017)

2.1.1. 1958-1973: L'OPPOSITION DES MODÈLES

La droite gaulliste : le général ou le chaos

Les passages exemplaires les mieux reconnus par le *deep learning* pour la droite du début de la V^e République, et donnés à l'analyse par le mécanisme de la déconvolution, décrivent le discours électoral gaulliste, devenue la force politique dominante. 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 signale, parmi les passages les plus fortement sélectionnés pour le classement à droite entre 1958 et 1973, la déclaration de Léon Delbecque. Nous soulignons, en gras, les zones de texte activées (forts indices TDS) par le réseau de neurones :

« 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'abîme¹.

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. Le calcul des spécificités, appliqué à toutes les professions de foi du corpus, est formel : statistiquement, le lemme « changement » est sous-utilisé dans les candidatures de droite entre 1958 et nos jours (figure 1) alors qu'il est sur-utilisé à gauche :

¹ Léon Delbecque, 1958, Nord.

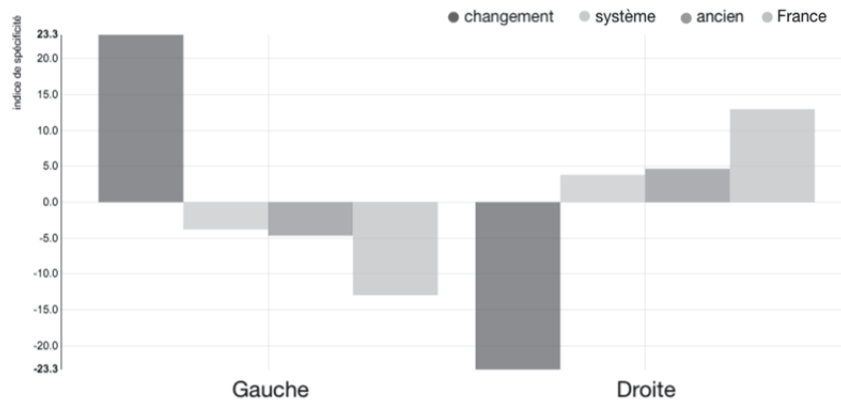


Figure 1. Sous-utilisation du mot « changement » à droite et sur-utilisation des mots cooccurrents dans le passage : « système », « ancien » et « France » dans le corpus Professions de foi (1958-2017)

Mais le calcul des spécificités de l'ADT nous montre également que les mots « système », « ancien » ou « France », présents dans le contexte de « changement » dans l'extrait de Delbecque, sont sur-employés 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^e République.

Le « changement » ici ne correspond en rien à un bouleversement de l'ordre social ou économique, comme à gauche, 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 profond, qui s'active sur les structures linguistiques faisant le parallèle entre l'adhésion à la France (ou à la V^e République), l'adhésion à de Gaulle et l'expression du peuple, détecte bien. Madeleine Martinache aux élections de novembre 1958 écrit :

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.¹

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^e République. À 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 indépendants.²

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 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** demeure.³

Ici aussi l'approche statistique des simples mots « parti » et « communiste » montrerait qu'ils appartiennent à la gauche de l'échiquier et, évidemment, aux locuteurs communistes (cf. ci-après), mais le syntagme nominal entier « la dictature du parti communiste » appartient bien pour le *deep learning* aux candidats de droite à la députation. À cette date, en tout cas, la tendance à la démocratie plébiscitaire et la *priorité* du maintien de l'ordre et de la stabilité du

¹ Madeleine Martinache, 1958, Nord.

² Léon Delbecque, 1958, Nord.

³ Nicole Hautecloque, 1968, Paris.

régime face aux mouvements estudiantins et sociaux de mai 1968 font du gaullisme un bonapartisme moderne [Vasseur-Desperriers 2006].

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 candidats 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 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 + VERBE » (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 candidats appellent de leur vœu un front « ouvrier, démocratique et progressiste » qui vise à se démarquer tant du socialisme soviétique que du pouvoir personnel gaulliste. Marie-Claude Vaillant-Couturier écrit en 1968 :

[...] exiger que soient durablement garantis les résultats des grandes **luttres 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 paix**¹.

¹ Marie-Claude Vaillant-Couturier, 1968, Val-de-Marne.

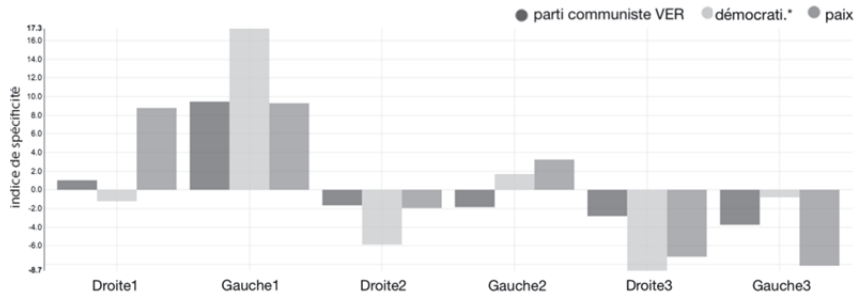


Figure 2. Sur-utilisation du motif « parti communiste VER », de la racine « démocrati.* » et du lemme « paix » dans les professions de foi de la gauche 1 (1958-1973)

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).



Figure 3. Les co-occurrences statistiques de « paix » dans le discours de la gauche 1 (en haut) et de la droite 1 (en bas) entre 1958 et 1973

En prenant en compte le co-texte, le réseau identifie à raison le programme pacifiste et anti-colonial de la gauche communiste du début de la V^e République.

2.1.2. ALTERNANCES ET COHABITATIONS (1978-1997)

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)¹. 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 « droite 2 », la critique ne se fait pas centralement sur la politique économique ou sociale mais plutôt sur la morale et sur le terrain régalien. À 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 candidats 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 [...]².

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, Édouard Balladur écrit :

¹ Notons toutefois que les corpus d'apprentissage, puis de test, ne contiennent pas les professions de foi de 1986 qui voient la victoire de candidats Front national. Ces textes ont été écartés en raison de l'hétérogénéité induite par la tenue exceptionnelle des élections législatives au scrutin de liste.

² Martine Aurillac, 1997, Paris.

Nous devons poursuivre notre lutte commune contre l'insécurité, la délinquance et le crime.¹

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 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.** ». Méthodologiquement, insistons seulement sur le fait que ce sont donc des *patterns séquentiels* structurants, des *motifs* ou des *collocations* qui activent fortement la reconnaissance des discours.

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 70². 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). 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.³

¹ Édouard Balladur, 1988, Paris.

² 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.

³ Edwige Avice, 1981, Paris.

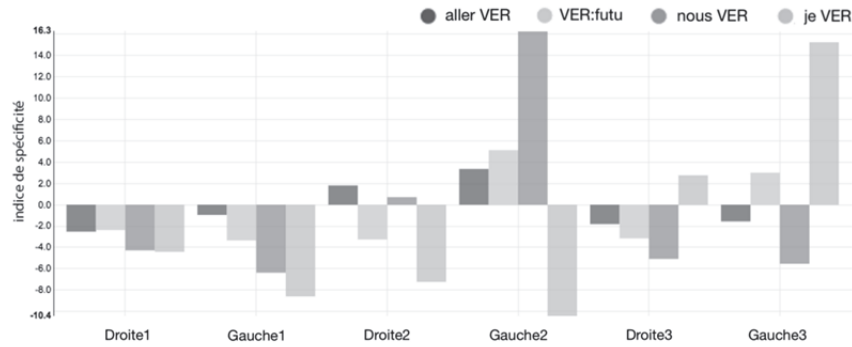


Figure 4. Sur-utilisation des structures grammaticales « aller + Verbe à l’infinitif » et « nous + Verbe futur », et sous-utilisation de « Je + Verbe » dans la gauche 2 (1978-1997)

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 futurs députés 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 valeur¹.

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. 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

¹ Jean-Michel Testu, 1981, Indre et Loire.

logement social. **Nous assurerons** la sécurité de proximité avec davantage d'agents [...]¹.

Les autres passages activés par l'algorithme 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 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 élus et les citoyens et non plus changer substantiellement les structures politiques, sociales ou économiques.

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** [...]².

Malgré les glissements lexicaux (non plus « changer la vie » mais désormais « changer la politique »), l'IA reconnaît les structures phrastiques typiques des promesses de mouvement de la gauche de gouvernement et attribue convenablement le texte à la bonne classe (ici la gauche 2-1978-1997).

2.1.3. L'APRÈS 2002

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^e tour de l'élection présidentielle, le discours de droite se décomplexé [Fougier 2012 ; Mayaffre 2012] et dépasse même les repères de

¹ Eric Besson, 1997, Drôme.

² Marie-Françoise Clergeau, 1997, Loire-Atlantique.

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 Nicolas Sarkozy sur les députés de son parti, bien après la fin de son mandat. Par exemple en 2017, Marine Brenier propose de :

[...] **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 **communautarisme**¹.

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, devenue 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 :

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 **communautarisme**².

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 candidats de la droite entre 2002 et 2017 adoptent massivement ce mode qui confère à l'exposition de leur programme un aspect impersonnel et intemporel.

Ceci semble témoigner à la fois de l'évolution historique du genre même de la profession de foi [Bonhomme 2016] qui s'apparente de plus en plus au fil de la V^e République à un catalogue de propositions (liste à puces, dispositions sous forme de tirets) et de la réassurance idéologique d'une droite décomplexée qui exprime de manière injonctive les mesures fortes à prendre (« restaurer », « éradiquer », « défendre », « renforcer »).

¹ Marine Brenier, 2017, Alpes-Maritimes.

² Philippe Vigier, 2007, Eure et Loir.

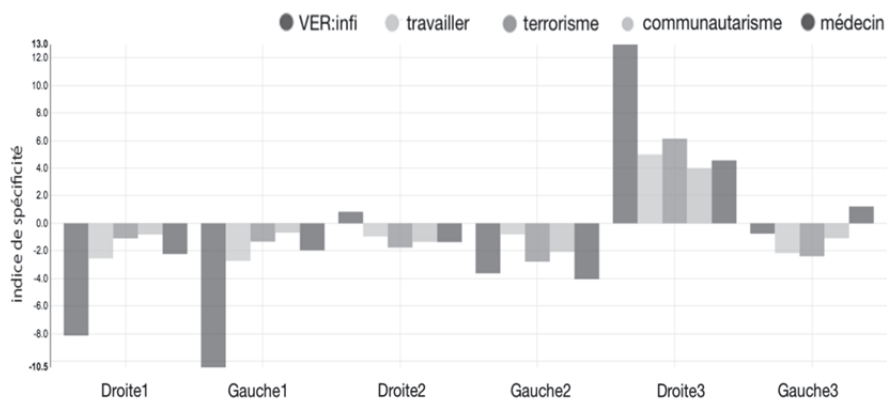


Figure 5. Sur-utilisation grammaticale et lexicale dans le corpus de la droite 3 (2002-2017) : l’infinitif, et les lemmes « travailler », « communautarisme », « médecin »

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 handicap¹.

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.

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 candidats de la gauche social-démocrate cèdent en effet

¹ Bérangère Poletti, 2017, Ardennes.

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 [Le Bart et Lefèbvre 2005]. La déconvolution pointe ainsi 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 renouvellement¹.

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 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 candidats pour devenir centrale dans les développements programmatiques de la gauche du XXI^e siècle [Mayaffre 2021]. À 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)².

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.

¹ Chaynesse Khirouni, 2012, Meurthe-et-Moselle.

² Pascale Crozon, 2012, Rhône.

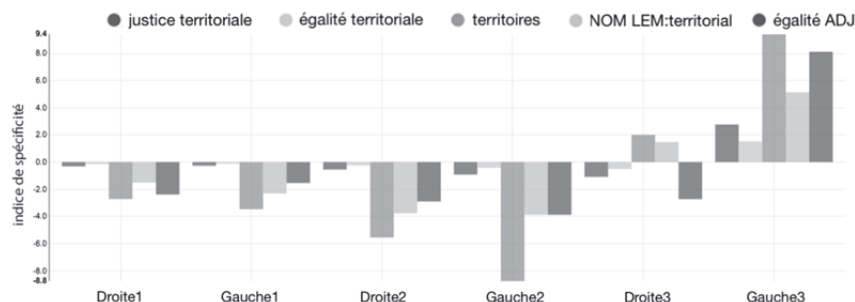


Figure 6. Sur-utilisation de « territoires », « NOM + LEM:territorial » dans le corpus de la gauche 3 (2002-2017)

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 et aux oppositions de groupes sociaux, et de plus en plus à une égalité et à une justice pensées dans l’espace géographique. Quelques futurs élus 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 d’Édith Gneugneau 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 vie¹.

Beaucoup d’autres, 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 ».

*

¹ Edith Gueugneau, 2012, Saône et Loire.

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 [Longrée & Mellet 2013] – ou les passages – c'est-à-dire des unités locales, continues ou discontinues, interprétables au sein du texte ou du corpus [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 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, soit dans son co-texte phrastique immédiat, qui active le réseau.

2.2. SITUER LES PROFESSIONS DE FOI MACRONISTES SUR L'ÉCHIQUIER POLITIQUE

La déconvolution fournit des observables linguistiques pour la description et l'interprétation des textes sur lesquels on a entraîné le réseau. Dans les sections précédentes, nous avons cherché à cerner les unités qui avaient présidé au bon apprentissage de la gauche et de la droite au fil du temps. Les textes appris et les textes à reconnaître étaient les mêmes. Mais il est également possible de soumettre à l'algorithme des textes qu'il n'a jamais vus, des données qui n'appartenaient pas au corpus d'apprentissage, afin de procéder tout de même à une classification. Concrètement, nous avons présenté les professions de foi de La République En Marche à notre réseau de neurones, entraîné sur les professions de foi de gauche et de droite,

entre 1958 et 1973, 1978 et 1997 et 2002 et 2017. L'enjeu est de faire attribuer les passages de ces discours électoraux à l'une de ces classes. Nous postulons qu'en étant contraint de situer les extraits des textes d'En Marche dans l'un des 6 groupes politico-chronologiques, le réseau nous suggérera les parentés idéologiques entre le récent mouvement d'Emmanuel Macron et les familles politiques historiques. Autrement dit, nous faisons l'hypothèse que la reconnaissance des passages macronistes à gauche ou à droite à telle ou telle période, procède de l'usage d'unités linguistiques partagées (repérées par le TDS) qui nous informera sur les influences voire la généalogie politique du mouvement.

La démarche est loin d'être anecdotique et doit servir la compréhension du macronisme puisque ses fondateurs n'ont eu de cesse de remettre en cause le clivage gauche/droite tout en s'y référant et en s'y rattachant¹. Aussi, pour comprendre la recomposition du paysage politique contemporain nous posons l'hypothèse que les proclamations électorales du dernier scrutin législatif ne dépassent pas le clivage gauche/droite traditionnel en adoptant des mots, thèmes ou arguments inédits mais qu'au contraire elles procèdent d'héritages et d'emprunts aux discours des deux grandes forces qui structurent la vie politique française depuis la Révolution, emprunts que la déconvolution permet de saisir formellement.

2.2.1. CLASSIFICATION

La classification des professions de foi des candidats de la République En Marche se fait selon le tableau suivant : plus de 56% des passages sont attribués à la droite entre 2002 et 2017, 30 % à la gauche à la même période, etc. Si ces résultats ne permettent pas encore la description d'unités linguistiques, ils sont intéressants à deux égards.

¹ Par exemple : Yolaine De Courson, candidate En Marche en Côte d'Or, écrit : « Les clivages gauche droite nous divisent, diminuent nos forces alors qu'il y a urgence à les unir pour faire de nos territoires que nous aimons tant, un avenir pour nos enfants, et retrouvant l'audace, la fierté et le plaisir d'entreprendre.» (2017, La République En Marche, Côte d'Or). Falliot, quant à lui, constate que : « L'enjeu de ces législatives est de taille. Il s'agit [...] de participer à la recomposition du paysage politique avec un grand rassemblement des modérés de droite, de gauche et du centre, partageant des valeurs sociales, humanistes, laïques et libérales, ouverts sur l'Europe et le monde.» (2017, La République En Marche, Tarn).

D'une part, ils révèlent le poids de la chronologie sur les discours politiques : les textes d'En Marche produits en 2017 s'apparentent d'abord aux textes du corpus d'apprentissage produits entre 2002 et 2017 puis à ceux rédigés entre 1978 et 1997 avant ceux écrits entre 1958 et 1973. Outre le fait que le genre de la profession de foi électorale ait évolué au fil de la V^e République (passant d'un argumentaire étayé à une forme de tract amélioré), l'actualité pèse fondamentalement sur les discours électoraux. On comprend alors que peu de phrases de 1958 trouvent un écho en 2017 ; les enjeux et les pratiques politiques étant sensiblement différents.

D'autre part, politiquement, l'algorithme note une plus grande proximité avec les professions de foi de la droite entre 2002 et 2017 qu'avec la gauche de la même époque. Pourtant les députés de la République En Marche se revendiquent plutôt, pour une large partie d'entre-eux, de la majorité socialiste sortante.

Classe	% de passages attribués
Droite 3	56,73
Gauche 3	30,74
Gauche 2	6,21
Droite 2	6,1
Droite 1	0,15
Gauche 1	0,06

Tableau 1. Classification des professions de foi En Marche de 2017

Une fois ce classement établi, l'intérêt de l'analyste du discours doit s'attacher à la description et à l'interprétation des unités linguistiques activées pour chacune des classes. Pour nourrir les recherches sur le mouvement macronien, on observera en particulier les unités de gauche et de droite des deux dernières périodes du corpus.

2.2.2. LE DISCOURS DE LA RÉPUBLIQUE EN MARCHE EN CAMPAGNE : LA 3^e VOIE

L'analyse du discours électoral de 2017 par les outils du *deep learning* donne peut-être des clefs pour définir le macronisme en campagne, que les observateurs de la vie politique ont peiné à saisir faute de substance ou de contenus marqués. Les professions de foi des marcheurs n'apparaissent pas tant « ni de gauche, ni de droite » que « et de gauche et de droite ».

Les phrases classées à droite relèvent du discours d'un libéralisme tempéré par une préoccupation sociale. Le passage le mieux reconnu pour cette classe, une citation de Macron reprise dans la quasi-totalité des candidatures, l'exprime bien :

Nous ne pouvons pas prendre le risque de retomber dans l'**immobilisme**. Notre pays a besoin d'une transformation profonde pour **retrouver** confiance, **libérer** les énergies **en protégeant** les **individus** au cœur d'une Europe qui protège.

Dans ce passage, le réseau s'active particulièrement sur le lemme « immobilisme », sur le verbe à l'infinitif « libérer » ou encore sur le lemme « protéger », qui sont des spécificités statistiques de la droite de la dernière période du corpus.

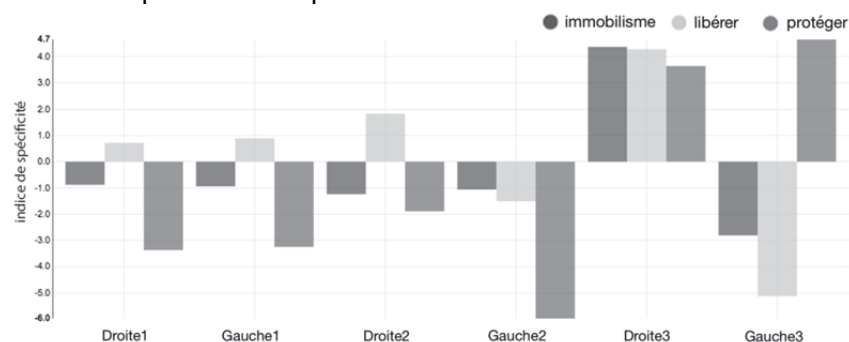


Figure 7. Les mots de la droite 3 (2002-2017) dans le corpus En Marche (2017)

Le discours En Marche s'inscrit, comme son nom même l'indique, dans la quête du mouvement permanent, face au risque de l'« immobilisme ». Ce positionnement aurait pu être un argument en faveur de sa catégorisation à gauche – la gauche étant historiquement désignée comme le parti du mouvement [Goguel 1946] – mais ici le terme contribue à son classement à droite. En effet, c'est la droite contemporaine, ressentant la nécessité de s'adapter aux évolutions de l'économie et aux soubresauts du marché, qui tient désormais le discours du changement ou de la réforme. Il s'agit par exemple de « libérer » les énergies ou « d'assouplir » le code du travail.

Dans la perspective macroniste, la libéralisation du marché et l'allègement nécessaire de l'État providence doivent se conjuguer avec des protections individuelles destinées à pallier les difficultés des plus démunis. Se dessinent dans les mots pointés par la déconvolution

(« libérer » mais aussi « protéger » que la gauche et la droite utilisent de concert au début du XXI^e siècle – figure 7), les marqueurs d'un discours de la troisième voie (qui se diffuse en Europe depuis les années 2000 et dont le Blairisme incarne un exemple éminent¹). Ainsi, un libéralisme égalitaire, au sens de Rawls [2008], que les spécialistes de la parole présidentielle ont identifié directement dans la parole du leader d'En Marche lorsqu'il prétend « en même temps » libérer et protéger [Mayaffre 2021], s'impose dans le discours : « Vous l'avez compris, je veux libérer l'énergie de ceux qui peuvent, et dans le même temps je veux protéger les plus faibles »².

A gauche, la filiation entre la social-démocratie qui domine l'Assemblée nationale entre 2012 et 2017 et La République En Marche s'opère sur le mot « territoires » et ses co-occurents privilégiés (« bassin », « acteurs ») dont on connaît l'usage important qu'en fait la gauche réformatrice après 2002 (figure 6). Nouvelle instance politique, les territoires sont également le lieu d'application du programme économique de La République En Marche, résumé en quelques mots : « transition énergétique (ou écologique) », « numérique » et « innovation ». Comme le montre la figure suivante, l'« innovation » et la « transition écologique » étaient déjà présentes dans la prose électorale de gauche du début du XXI^e siècle ; mais les marcheurs accentuent le sujet et l'orientent résolument vers la question de la technologie numérique (en particulier la couverture du territoire en « haut débit » internet).

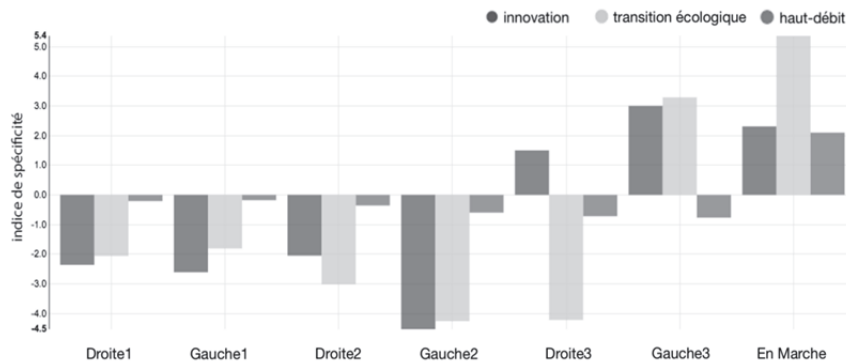


Figure 8. Les mots de la gauche 3 (2002-2017) dans le corpus En Marche (2017)

¹ Voir sur ce point Fairclough (2000).

² Macron, meeting à Bobigny, 16 novembre 2016.

Par exemple, Annaïg Le Meur écrit :

Je serai la députée de la cohésion et de la solidarité sur **notre territoire** et j'agirai pour :

1. Mobiliser tous les **acteurs** de **notre bassin** de vie pour faciliter l'accès à l'emploi.
2. Insérer **notre territoire** dans la **transition énergétique** et **numérique**, favoriser l'**innovation** dans nos filières clés (agriculture, agronomie, mer, tourisme)¹.

Ou encore Christophe Lejeune, en reprenant les priorités fixés nationalement, promet de :

Garantir l'**égalité des territoires** et engager la **transition énergétique** en couvrant la France en très **haut débit** ou en fibre [...]².

Dans ces passages, le progrès social (la « solidarité », l'« égalité » territoriales), sujet de prédilection de la gauche, devient surtout dans la perspective macroniste un progressisme technologique : l'innovation numérique et technique résoudra le problème de l'emploi, de l'égalité ou encore de la transition énergétique.

*

D'un point de vue méthodologique, ces traitements *deep learning* du corpus des candidats En Marche laissent entrevoir une plus-value pour l'analyse des textes politiques, par rapport à nos pratiques habituelles en logométrie. Par la statistique textuelle traditionnelle, on sait caractériser les discours d'un locuteur par rapport à d'autres, que l'on compare dans des corpus contrastifs qui incarnent une norme endogène. Par exemple, on cherchera les singularités du discours de La République En Marche en le confrontant, pour le distinguer, aux discours des communistes, des socialistes, des écologistes, des républicains, des nationalistes, etc. Dans cette logique, ce sont les caractéristiques fortes de chacune des sous-parties du corpus par rapport aux autres qui sont mises en relief : c'est l'identité ou

¹ Annaïg Le Meur, 2017, Finistère.

² Christophe Lejeune, 2017, Haute-Saône.

l'homogénéité du discours de chaque entité considérée qui est soulignée¹.

Ici, c'est l'hétérogénéité des discours qui a été mise en avant, ouvrant la voie à l'étude de l'intertexte que nous nous proposons de faire dans la section suivante. Les emprunts, les influences, les accents d'un autre locuteur ou d'un discours antérieur sont reconnus par l'algorithme à l'intérieur des professions de foi des députés En Marche. La modalité des emprunts controversés à la gauche et à la droite de gouvernement dans le discours de la majorité 2017 a été mise au jour.

Utilisés dans cette perspective, les outils du *deep learning* permettent une formalisation des interdiscours et des intertextes, encore difficilement saisissables dans nos pratiques quotidiennes en analyse de données textuelles et pourtant indispensables à tout parcours interprétatif des textes en corpus.

3. VERS UNE OBJECTIVATION DE L'INTERTEXTE. LES EMPRUNTS DE MACRON À SES PRÉDÉCESSEURS À L'ÉLYSÉE

Partageons l'affirmation de I. Kanellos sur l'importance de l'intertexte et sur la nécessité de l'explicitier pour permettre une interprétation critique² ; et transformons l'affirmation en programme de recherche pour la linguistique des textes et l'analyse du discours.

L'explicitation de l'intertexte est garante d'une certaine scientificité interprétative, dans la mesure où sa mise en place et sa communicabilité rendent une interprétation partageable. Il clarifie le terrain du face-à-face interprétatif en le rendant accessible aux regards extérieurs. En effet, une lecture qui expliciterait son parcours non seulement à travers les structures textuelles mais aussi les structures intertextuelles, offrirait une image claire, du moins une image moins fuyante de la compétence interprétative du lecteur et proposerait, en même temps, une grille d'appréciation, d'évaluation aussi de sa performance.³

Les notions d'intertexte et d'intertextualité sont en effet centrales dans les arts et sciences des textes depuis 50 ans, mais elles restent

¹ D'un point de vue mathématique, on lira ici le traitement factoriel du vocabulaire banal et spécifique proposé par [Bourgeois N. and *al.* 2012].

² Cette section reprend et développe la communication que nous devons faire au CMLF 2020, annulé pour des raisons sanitaires (COVID-19).

³ Ioannis Kanellos, « De la vie sociale du texte. L'intertexte comme facteur de la coopération interprétative », *Cahiers de praxématique* [En ligne], 33 | 1999, § 115.

encore aujourd'hui nébuleuses. Nul en effet ne conteste ni leur pertinence ni leur fécondité, mais personne ne réussit à les formaliser.

La définition aujourd'hui consacrée d'intertexte, proposée en 1973 dans l'*Encyclopaedia universalis* par Roland Barthes, sous l'entrée « Texte (théorie du) », dit tout de la force du concept en même temps que de sa fragilité scientifique :

Tout texte est un intertexte ; d'autres textes sont présents en lui, à des niveaux variables, sous des formes plus ou moins reconnaissables [...] ; tout texte est un tissu nouveau de citations révolues [...]. L'intertexte est un champ général de formules anonymes, dont l'origine est rarement repérable, de citations inconscientes ou automatiques, données sans guillemets. [Encyclopaedia universalis, ed. 1995, Tome 22, p. 372]

Admettons donc grossièrement depuis Lotman, Kristeva, Barthes, Genette, Rifatterre, etc. sinon depuis Bakhtine ou Voloshinov, que l'intertextualité est la « condition de tout texte » (*ibid.*). En utilisant les mots des autres et de tous, le locuteur articule un discours qui n'est jamais totalement un inédit. Quand on parle, l'héritage en langue et en discours ne laisse qu'exceptionnellement la place à l'originalité ou à la néologie, et pétri par 2000 ou 3000 ans de culture textuelle, l'auteur est condamné à être un contrefacteur. Sans évoquer le cas caricatural et réducteur du plagiat, les écrits riches et nombreux sur l'intertexte, sur le dialogisme, la polyphonie ou l'hétérogénéité¹ ont définitivement démontré qu'un texte n'est pas une monade ou un isolat mais un dialogue ou un carrefour, fait d'emprunts ou de reprises, de réponses ou de questions à l'endroit d'autres textes passés ou à venir, avec qui il communi(que) ; et sans lesquels il devient infra-interprétable.

Ajoutons encore, à la suite d'I. Kannelos et dans la posture herméneutique rastérienne qui est la nôtre, que lire c'est interpréter, et qu'interpréter c'est d'abord contextualiser, relier, mettre en correspondance, observer ce qu'il y a « entre ». Et cette contextualisation fondamentale se fait au plus haut niveau en versant le texte dans un *déjà-dit*, un *autrement-dit* ou un *bientôt-dit*, c'est-à-dire

¹ Depuis la thèse de J. Kristeva en 1967, la littérature sur l'intertexte et les notions associées est vaste et passionnante ; parfois des ouvrages entiers. Nous donnons ici arbitrairement seulement deux références : le numéro spécial des *Cahiers de praxématique*, « Sémantique de l'intertexte », n° 33, 1999 ; et l'ouvrage collectif de J. Bres *et al. (dir.)*, *Dialogisme et polyphonie – Approches linguistiques*, Éditions De Boeck-Duculot, 2005.

dans un certain intertexte. Dans cette perspective, nous considérons bien *l'intertexte comme la condition de l'interprétation de tout texte*.

Concrètement, la contextualisation pratique d'un texte à interpréter, dans une analyse, se fait à l'intérieur d'un certain corpus, défini de manière réflexive comme la matérialisation du contexte nécessaire à l'interprétation des textes qui le composent [cf. notamment Mayaffre 2002 et Rastier 2001 et 2004]. Dit autrement, les corpus réflexifs que nous avons proposés ailleurs [Mayaffre 2002 et 2007] deviennent pour cette contribution l'intertexte objectivable des textes qui le composent. Précisons bien, puisque l'intertexte est potentiellement infini : le corpus réflexif constitue une forme explicitée ou objectivée d'un *certain* intertexte choisi pour l'analyse.

Cela étant dit, même réduit et matérialisé dans des corpus réflexifs dûment constitués, l'intertexte reste un concept mou sinon impuissant. Sa valeur élastique semble décourager la formalisation et interdire l'implémentation informatique : « à des niveaux variables » nous dit Barthes dans *l'Encyclopaedia* ; « formes plus ou moins reconnaissables » imprécise-t-il ; « rarement repérable » sous-renchérit-il, « citations révolues », « inconscientes » et « sans guillemet ».

Nous souhaitons pourtant proposer ici pour conclure ce chapitre un usage particulier de l'Intelligence artificielle des textes susceptible d'envisager l'impossible, c'est-à-dire d'objectiver cet intertexte fuyant, mou ou liquide ; l'espoir est que le réseau de neurones artificiels relie, au sein du corpus, le texte-cible analysé (ici le texte de Macron dans son exercice de président entre 2017 et 2020) à son intertexte c'est-à-dire les textes qui le traversent (ici les discours présidentiels depuis 1958, soit ceux des prédécesseurs de Macron à l'Élysée).

Nous postulons en effet que l'intertexte – un des intertextes possibles – du discours de Macron est composé des discours que de Gaulle, Pompidou, Giscard, Mitterrand, Chirac, Sarkozy et Hollande ont pu prononcer avant lui, et dont le nouveau président s'inspire selon la contrainte générique (discours officiel présidentiel), la contrainte chronologique (le temps de la V^e République) et la contrainte politique (discours de synthèse, « en même temps » de gauche et droite, en même temps gaulliste, giscardien, sarkozyste ou hollandais).

3.1. PROPOSITION MÉTHODOLOGIQUE : UN USAGE ORIGINAL DU *DEEP LEARNING*

L'efficacité pratique des tâches d'apprentissage, puis de classification des textes par le *machine learning* ou le *deep learning* ne sont plus contestables. Depuis plusieurs années désormais, elle se chiffre par des taux d'exactitude (*accuracy rate*) spectaculaires. Ainsi, il n'y a pas plus de doute que la machine apprenne et reconnaisse un discours (une série de mots) de Mitterrand *versus* un discours de de Gaulle, qu'il n'y a de doute que la machine apprennent et reconnaisse une image de chat (une série de pixels) *versus* une image de chien [par ex. Ducoffe *et al.* 2016].

3.1.1. APPRENTISSAGE

L'algorithme implémenté dans le logiciel Hyperbase a ainsi appris (*learning*) à reconnaître les discours de de Gaulle (phrases longues, plutôt nominales ou adjectivales, avec la « France » ou « l'État » comme premiers noms) comme ceux de Mitterrand (phrases courtes, plutôt verbales, avec le « je » et le « moi » comme centre d'intérêt personnel et « l'Europe » comme horizon). Il a appris à reconnaître les discours de Sarkozy (un lexique fort et une syntaxe faible) et ceux de Hollande (une syntaxe compliquée et un lexique affadi). Il a appris à reconnaître les textes de Giscard (phraséologie technocratique ou didactique), ceux de Chirac (pleins de vide) ou ceux de Pompidou (style littéraire, riche voire ampoulé) ; et le jeu de validation (corpus de test composé de discours présidentiels anonymisés) permet d'affirmer que la machine retrouve automatiquement au-dessus de 92 % des fois le bon auteur des discours. Sur un jeu de textes littéraires composé de l'essentiel de la littérature française aux XVIII^e, XIX^e et XX^e siècles, des résultats supérieurs, proches de 100% sont même établis dans cet ouvrage par Étienne Brunet, Ludovic Lebart et Laurent Vanni, dans une étude qui brasse 50 auteurs de Racine à Giono, de Hugo à Yourcenar, de Proust à Le Clézio [Voir *supra* chapitre 2, cf. aussi Brunet et Vanni 2019].

Lorsque nous avons soumis à la machine les discours élyséens depuis 1958, nous avons construit un certain horizon d'attente ou, d'un autre point de vue, un certain point de départ pour les discours de Macron : selon nous, le module *deep learning* du logiciel a appris un certain intertexte, dans lequel il est plausible que Macron ait pu

emprunter ou dont il a pu s'inspirer pour construire ses propres discours ; ou encore, un certain intertexte dans lequel l'auditeur et l'analyste pourront faire résonner les discours de Macron pour les comprendre et les interpréter.

3.1.2. PRÉDICTION

Puis, nous avons versé les discours de Macron dans ce corpus élyséen de référence (ou intertexte élyséen), en demandant à l'Intelligence artificielle de rapprocher chaque phrase ou chaque paragraphe de Macron d'un des présidents précédents ; non sans artificialité donc, nous forçons l'algorithme à attribuer (*prédiction*) les paragraphes de Macron à un de ses prédécesseurs au motif de ressemblances linguistiques détectées. Ainsi si Macron devait s'écrier à la tribune « Je vous ai entendu ! », la machine attribuera ce passage à de Gaulle en référence sans doute au « Je vous ai compris ! » d'Alger en 1958. Si Macron devait prononcer « vous n'avez pas le monopole des sentiments », ou peut-être « vous n'avez pas l'exclusivité des sentiments », la machine attribuera le passage à Giscard d'Estaing en référence au débat télévisé avec Mitterrand en 1974 durant lequel le candidat de droite avait répliqué au candidat de gauche « vous n'avez pas le monopole du cœur ». Pour donner un premier résultat réel du travail, lorsque Macron déclare à la télévision, lors de ses vœux aux Français, le 31 décembre 2018, « on ne peut pas travailler moins et gagner plus », Hyperbase attribue automatiquement la phrase à Sarkozy en référence au « il faut travailler plus pour gagner plus » que le président de droite avait souvent répété durant son mandat.

3.1.3. DESCRIPTION

Enfin, dernière étape décisive pour l'étude linguistique, l'algorithme de déconvolution mis au point par [Vanni *et al.* 2018] permet de décrire le corpus : il s'agit non seulement d'extraire les phrases de Macron attribuées à de Gaulle ou Pompidou, Sarkozy ou Hollande, mais de surligner les éléments linguistiques qui ont participé à la décision. Ici, la méthode d'Intelligence artificielle qui sert de guide (convolution puis déconvolution, et indice de reconnaissance des unités saillantes (*Text Deconvolution Saliency* - TDS)) est doublée par les procédés logométriques traditionnels (l'historique calcul des *spécificités*), afin d'affermir les résultats de l'IA par la statistique textuelle. Dans tous les

cas, que cela soit en matière de classification ou de description, rappelons que l'analyse se fait à trois niveaux linguistiques : les formes graphiques, les lemmes et les étiquettes morphosyntaxiques. Tant est si bien que tel extrait de Macron pourra être rapproché du discours de Giscard par exemple à cause de sa nature fortement nominale (combinaisons marquées de noms, d'adjectifs et de déterminants) ; tel extrait de Macron sera rapproché de Pompidou pour la variété et la richesse des lemmes utilisés (concentration de plusieurs lemmes rares et précieux) ; tel extrait sera rapproché de Hollande pour sa combinaison de mots graphiques significatifs comme « territoires » (au pluriel) ou « investissements » (également au pluriel).

3.2. RÉSULTATS : UN INTERTEXTE PLURIEL POUR UN DISCOURS PATCHWORK

L'intertexte macronien est varié : sur les 10.000 paragraphes du corpus Macron analysés¹, les taux d'emprunt, d'inspiration ou de rapprochement se hiérarchisent ainsi (tableau 2) :

TEXTE – Discours d'Emmanuel Macron	INTERTEXTE - Présidents inspireurs	Taux d'inspiration
	De Gaulle	2%
	Pompidou	7%
	Giscard	3%
	Mitterrand	12%
	Chirac	21%
	Sarkozy	29%
	Hollande	26%

Tableau 2. Intertexte de Macron ou sources d'inspiration de ses discours (en % de paragraphes)

L'intertexte de Macron répond donc grossièrement à la chronologie. Globalement, le discours de Macron emprunte davantage aux discours de ses prédécesseurs immédiats (Chirac, Sarkozy, Hollande) qu'à ceux de ses prédécesseurs plus lointains (de Gaulle, Pompidou, Giscard).

¹ Précisons le corpus : composé de 1000 discours équivalant 3 millions de mots, le corpus recueille les principaux discours de de Gaulle, Pompidou, Giscard, Mitterrand, Chirac, Sarkozy, Hollande et Macron entre 1958 et 2020. Afin de donner une bonne représentativité au corpus, nous avons saisi les allocutions solennelles, les grands discours d'estrade, les interviews télévisées, les vœux ou les tribunes dans la presse.

Aux deux extrémités chronologiques notamment, Macron s'inspire seulement à hauteur de 2% de de Gaulle alors qu'il s'inspire à hauteur de 26% de Hollande. Néanmoins, d'intéressantes exceptions contredisent cette logique temporelle : Pompidou inspire plus Macron que Giscard, et Sarkozy – en tête du classement avec 29 % – inspire plus le nouveau président que Hollande.

Si ce classement général est riche d'information, c'est l'examen linguistique systématique que nous avons fait des phrases attribuées à de Gaulle et Pompidou, Giscard et Mitterrand, Chirac, Sarkozy et Hollande qui est intéressant [Mayaffre 2021] : il en dit long sur le discours *patchwork* d'Emmanuel Macron qui diversifie son intertexte et multiplie les sources pour incarner une forme de syncrétisme discursif susceptible peut-être de réconcilier les mémoires et de résonner favorablement chez plusieurs types de citoyens.

Nous pourrions nous attarder sur les intertextes les plus évidents (ceux, par ordre hiérarchique, de Sarkozy et de Hollande) mais choisissons ici deux intertextes plus ténus – ténus mais effectifs – dans le discours, celui de Giscard et celui de Pompidou. Le premier est surtout grammatical, le second surtout lexical.

3.2.1. L'INTERTEXTE MORPHO-SYNTAXIQUE DE GISCARD CHEZ MACRON

Dans le corpus présidentiel, Macron n'emprunte guère à Giscard d'Estaing (3%, tableau 2), moitié moins qu'il n'emprunte à Pompidou. Et c'est un rapprochement peu flatteur que le *deep learning* finit par repérer à la marge du discours : un langage technocratique, notionnel, austère et austéritaire.

Tous les deux énarques, tous les deux inspecteurs des Finances, tous les deux ministres de l'Économie, Macron et Giscard utilisent en commun une langue technique et notionnelle qui fige et dépersonnalise le discours ; une forme de langue de bois a-politique qui pédagogise l'économie, la libre entreprise et l'équilibre des budgets ; une langue ou une « ENAlangue »¹ qui désidéologise la prose présidentielle pour en faire un discours professoral conceptuel et « techno », un cours magistral d'HEC ou de droit administratif, un mode d'emploi aseptisé sur l'économie de marché en temps de crise.

¹ Christian Delporte, *Une histoire de la langue de bois*, Paris, Flammarion, 2009.

Macron (parfois) comme Giscard (souvent) expliquent en effet plus qu'ils ne cherchent à convaincre, enseignent plus qu'ils n'arrivent à persuader, professent au lieu de débattre. Et ce discours explicatif du monde économique tel qu'il va résonne négativement au regard de la situation historiquement défavorable que les deux présidents sont condamnés à décrire (la crise pétrolière des années 1970 avec l'apparition du chômage pour l'un, la crise budgétaire des années 2010 avec la pérennité du chômage pour l'autre).

Par exemple, Macron explique la suppression des emplois-aidés en une phraséologie que le *deep learning* reconnaît comme giscardienne¹ :

C'est de la **subvention** déguisée vers les **collectivités locales** ou le **secteur associatif**. Ces **secteurs** ont toute leur utilité, mais, dans ce cas, cela doit entrer dans la **dotation des collectivités locales** ou dans les **subventions aux associations**, mais qu'on ne nous dise pas que c'est une **politique de l'emploi**. Le **taux de retour à l'emploi durable** des personnes concernées est en effet très faible.²

La formule nominale et absconse « le taux de retour à l'emploi durable », les simples noms « collectivités », « secteur », « dotation » ou « subvention » sont caractéristiques de cette novlangue politico-économique des services administratifs de l'État ou du monde des grandes entreprises.

Et plus loin dans cette interview, Macron précise la mise en place des Ordonnances travail :

Mais la **branche**, c'est-à-dire les **représentants des entreprises** et des salariés qui exercent des métiers similaires ou travaillent dans les mêmes **types d'activités**, c'est un progrès dans la meilleure prise en compte de la **réalité économique**. Ensuite, les **branches** sont indispensables parce que beaucoup **d'entreprises**, mal outillées ou trop petites, ne peuvent ou ne souhaitent pas négocier des **accords complexes**.³

¹ Répétons la démarche. L'algorithme examine tous les passages de Macron et mesure systématiquement leurs ressemblances avec les discours des présidents précédents. Dans cette section, les extraits de Macron qui seront cités ont donc été reconnus automatiquement comme très proches du parler de Giscard. Les passages en gras sont les « zones d'activation » (TDS) particulièrement pertinentes pour la reconnaissance.

² Macron, 31 août 2017, interview dans *Le Point*.

³ *Ibid.*

« Accord », « branche », « types d'activités », « prise en compte de la réalité économique », « représentants des entreprises » : le lexique et la phraséologie sont caractéristiques.

Linguistiquement, on remarquera également dans ce dernier extrait combien Macron abandonne une énonciation vivante et tendue, qu'il aime par ailleurs, et dans laquelle le locuteur fait habituellement face à l'interlocuteur, pour présenter, ici, les choses de manière impersonnelle. On objectera peut-être qu'il y a dans l'extrait des « représentants » et même des « salariés », mais l'impression générale est qu'il s'agit de la prose d'un système sans âme et sans acteurs, où des rouages « complexes » (pour reprendre le mot de Macron) fonctionnent de manière mécanique.

Cette technicité d'un langage désincarné atteint parfois des hauteurs chez Macron comme elle atteignait toujours des sommets chez Giscard. Par exemple, lorsque Macron parle de la santé, normalement vecteur d'émotion et de *pathos*, il s'enferme dans une langue dévitalisée :

Je souhaite que nous puissions créer, **sur cette base**, un véritable **hub des données de santé**, structure partenariale entre producteurs et utilisateurs des données, qui pilotera l'enrichissement continu mais aussi **la valorisation du système national de données de santé**, pour y inclure, à terme, l'ensemble des données remboursées par l'Assurance-maladie, en ajoutant **les données cliniques des hôpitaux**, les données de la médecine de ville, ainsi que les données de grande qualité, scientifiques et médicales, créées dans **le cadre de cohortes nationales**¹

Les mots utilisés sont ceux du XXI^e siècle bien sûr, et Giscard n'employait ni « *hub* de données », ni « structure partenariale entre producteurs et utilisateurs des données », mais cette succession de syntagmes nominaux et de notions techniques est, selon la machine, typiquement giscardienne. « La valorisation du système national de données de santé », en effet, qui se schématise grammaticalement en « déterminant/nom + déterminant/nom/adjectif + déterminant/nom + déterminant/nom » est caractéristique. La langue du septennat Giscard a marqué l'apogée de cette phraséologie nominale où les compléments du

¹ Macron, 29 mars 2018, discours *AI for Humanity* au Collège de France. À côté de l'analyse produite, on remarquera d'un point de vue rhétorique l'utilisation de verbes, de noms ou d'adjectifs valorisants pour décrire une réalité qui l'est moins : « véritable », « créer », « enrichissement », « valorisation », « grande qualité ».

nom se succèdent en cascade pour nommer des réalités ou des processus techniques de plus en plus compliqués : Giscard a, dans le corpus, le taux de noms (« système » ou « possibilité » pour reprendre nos exemples), ou de l'enchaînement « déterminant + nom » (« un système » ou « la possibilité »), ou de l'enchaînement « déterminant + nom + préposition + nom » (« un système de santé », « la possibilité de règlement ») le plus spectaculaire de la V^e République (figure 9).

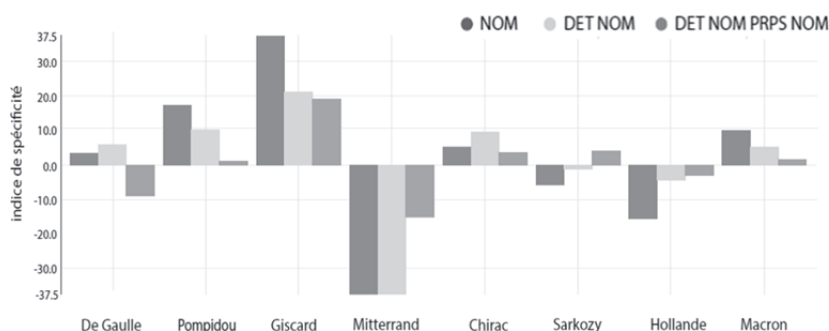


Figure 9. Utilisation des noms et du discours nominal dans le corpus présidentiel (1958-2020)

Cette rhétorique du nom, du complément du nom, du syntagme nominal est seulement légèrement fréquente chez Macron, mais lorsqu'il emploie « la réforme du marché du travail », « les acteurs du monde du logement », « les règles de recours au CDD », « la réalité des situations des collectivités territoriales », « la loi d'orientation des mobilités », etc., il semble vouloir resserrer la chose publique – c'est-à-dire la République – dans un jargon technico-technique et figer le débat dans des mécanismes naturalisés. Il euphémise, comme Giscard, la confrontation politique pour réduire le discours à la manipulation de notions et de sigles dont le but semble de réifier les rapports sociaux.

Du reste, la contextualisation des structures nominales mentionnées ci-dessus est instructive sur cette ENAlangue giscardienne. La formule nominale issue du débat sur la *Loi travail* durant l'été 2017, « les règles de recours au CDD », mérite d'être précisée pour que le citoyen puisse la comprendre et en débattre. Mais Macron indique seulement « les règles de recours au CDD peuvent se faire au niveau de la branche »¹ ajoutant ainsi une notion technique supplémentaire (la « branche ») à un propos qui l'était déjà doublement (« les règles de recours » et le

¹ Macron, 31 août 2017, interview dans *Le Point*.

« CDD »). De la même manière, à propos de « la réforme du marché du travail », il croit bon d'ajouter : « La réforme du marché du travail est une réforme de transformation profonde »¹ ajoutant ici une notion abstraite (« transformation ») à des concepts fuyants ou complexes (« réforme » ou « marché du travail »).

En l'occurrence, l'opposition de gauche et les syndicats dénonceront immédiatement, derrière ces deux derniers exemples (« règles de recours au CDD », « réforme du marché du travail »), des Ordonnances travail qui, selon eux, réduisaient fondamentalement les droits des travailleurs et constituaient une des régressions sociales les plus importantes depuis l'après-guerre. Mais les phrases citées offrent objectivement peu de prise sémantique, politique ou idéologique à l'adversaire, au regard, soit de leur vacuité, soit de leur technicité, soit encore par la présentation valorisante des arguments ; et les Ordonnances de 2017 s'appliqueront sans beaucoup de contestation dans le pays.

Et les exemples pourraient être multipliés, comme cet extrait caricatural, issu d'un discours qui, ici, loin de réussir à neutraliser les oppositions politiques, participera au contraire au malaise des « gilets jaunes » à l'automne 2018 :

La loi d'orientation des mobilités, présentée hier en Conseil des ministres après une intense **phase de concertations**, propose **une partie de ces solutions** qui seront à construire sur le terrain.²

Ce type de discours fortement nominalisé autour de termes creux ou techniques de l'économie, souvent également autour de sigles plus ou moins opaques (TPE, APB, GPA, REP+, CSG, AMF, BTP, ALUR, APL, UNEDIC, IFI, etc.) a pu être décrit par Christian Delporte comme l'excellence de la langue de bois des décideurs, et comme une rhétorique de l'esquive d'hommes politiques soucieux de faire passer des réformes impopulaires par des discours aseptisés. Ce type de discours, dont l'effacement énonciatif est remarquable, renvoie en effet toujours aux politiques déjà en place (« loi d'orientation », « règles de recours », « réforme du marché du travail ») et à la *doxa* dominante, c'est-à-dire en l'occurrence au modèle économique en vigueur qui cherche à neutraliser les différends politiques, ou la lutte des classes,

¹ *Ibid.*

² Macron, 27 novembre 2018, discours sur la transition écologique.

derrière des notions administratives, institutionnelles ou naturelles qui semblent s'imposer d'elles-mêmes. Insister sur la nécessité d'une « réforme du marché du travail » par exemple, c'est imposer en la nommant une politique forte, tout en minimisant, en une formule nominale figée, ses conséquences sociales éventuelles (le gain de productivité réclamé aux agents ou la question des licenciements nécessaires pour rendre plus efficace le modèle). Ce discours permet en fait de techniciser un débat, audible seulement par une élite initiée, pour ne pas avoir à discuter avec les citoyens ou les travailleurs des principes et des fondements. Il permet de se soustraire aux jugements axiologiques ou idéologiques du mal et du bien, du juste et de l'injuste, du vrai et du faux, et d'éviter la contradiction politique derrière un allant de soi qui tire sa force de la langue. C'est un discours « notionnel » au sens étymologique c'est-à-dire un discours de la connaissance, du savoir ou plus précisément un discours de l'expert¹ : une rhétorique du « je sais pour vous » conclut Christian Delporte². Et, de fait, nous le verrons, Macron sur-utilise massivement le verbe « savoir » à la première personne (« je sais »), comme Giscard sur-utilise le verbe professoral « indiquer » : « je vous indique » (figure 10).

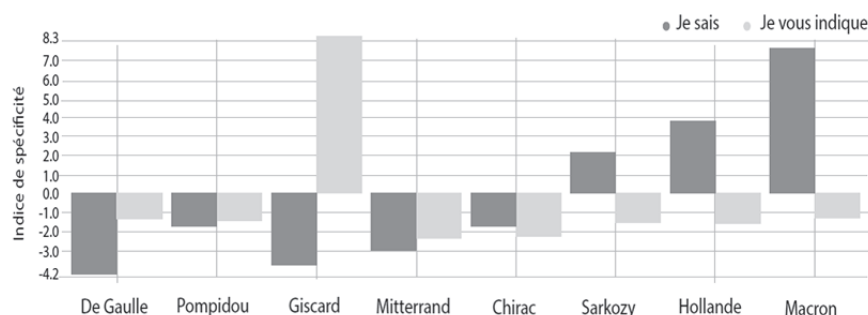


Figure 10. « Je vous indique » chez Giscard ou « je sais » chez Macron

¹ « Notion » est lointainement issu, *via* le grec et le latin, de l'indo-européen **gno* qui signifie « savoir », et que l'on retrouve par exemple en anglais dans le verbe *to know* ; une notion est un savoir élémentaire et les discours notionnels sont tenus par les gens qui savent (ici Macron et Giscard d'Estaing).

² Chr. Delporte, *Une histoire de la langue de bois*, *op. cit.* Sur le discours de l'expert, voir R. Cussó et C. Gobin dans un article au titre substantiel : « Du discours politique au discours expert. Le changement politique mis hors débat ? », *Mots*, 88, 2008, p. 5-11.

Quand Macron imite Giscard, nous versons ainsi dans un discours sans visage dont la conséquence est, au final, probablement contreproductive pour Macron comme elle le fut pour Giscard, en affaiblissant, par la forme, le message principal. En effet, le paradoxe de ce discours notionnel, descriptif, explicatif, stéréotypé est qu'il dégage une forme de fatalité – la fatalité de choses déjà-là, la fatalité d'une pensée unique et d'un discours pré-construit, la fatalité d'une mécanique économique-administrative ou gestionnaire que l'on ne saurait changer –, là où la modernité revendiquée par les deux présidents réclamerait au contraire action, débat, volontarisme, et pourquoi pas polémique ? Ce « je sais pour vous », ou pire encore « le système sait pour vous », « la langue parle à votre place » apparaissent comme un dessaisissement du politique au profit d'un ordre discursif (des formules figées) et d'un ordre économique (les politiques menées à l'échelle planétaire ou européenne depuis quelques décennies), c'est-à-dire comme un désinvestissement démocratique du président lui-même qui semble utiliser un prêt-à-parler technique et convenu ; et les citoyens, à leur tour, d'être plongés dans l'impuissance politique. Sans doute est-ce pour cette raison que Macron emploie cette éloquence giscardienne assez peu dans le corpus, pour rester fidèle au volontarisme horizontal et à l'élan démocratique vantés durant la campagne électorale.

3.2.2. L'INTERTEXTE LEXICAL POMPIDOLIEN

Emmanuel Macron a le goût pour les choses de l'esprit, particulièrement la littérature. Khâgneux, il aurait voulu être normalien comme avant lui Georges Pompidou, mais deux échecs successifs au concours de l'École normale l'empêchent de réaliser sa vocation. Comme Georges Pompidou, Macron est un lecteur passionné, voire lui-même prosateur, et il aurait sans doute aimé écrire une *Anthologie de la poésie française* et avoir une agrégation de grammaire si une vie trépidante dans le milieu des affaires ne l'en avait empêché.

Emmanuel Macron emprunte en tout cas à Georges Pompidou un niveau de discours, un registre de langage, une tenue littéraire sans équivalent sous la V^e République, que l'on ne retrouve pas depuis 20 ans dans les discours d'un Chirac, d'un Sarkozy ou d'un Hollande qui ont cherché, par souci de proximité, à imiter le langage populaire.

Dans le corpus présidentiel, Macron rivalise avec Pompidou par une richesse de vocabulaire hors normes qui constitue une des griffes de ses discours [Mayaffre 2020].

Mais au-delà de cette richesse lexicale générale qui rapproche Macron de Pompidou, et l'éloigne des autres présidents, il faut lire les phrases que le *deep learning* reconnaît comme pompidoliennes dans le discours du nouveau président. Elles représentent 7% du corpus Macron (*supra* tableau 2) et renvoient pour l'essentiel à des considérations littéraires et philosophiques. Elles déplacent, avec force références, le discours présidentiel du politique vers l'éthique ou vers la poétique, loin des facilités de la communication publicitaire qui a envahi le discours politique, et loin de la médiocrité – au sens étymologique – d'un Hollande ou la vulgarité – toujours au sens étymologique – de Sarkozy.

Je relisais il y a quelques jours **une lettre que Baudelaire** écrivait à sa mère, une nuit où il parlait de son découragement, du caractère insupportable de la vie. Et au milieu de la nuit, il décrivait ce moment, celui où souvent il se mettait à écrire de ce qu'il appelait « **le ressaisissement** ». C'est ce qui est en train de se passer dans la société française, c'est le moment du ressaisissement. Et il avait deux formules qui peuvent sembler paradoxales ou inconciliables pour certains, mais qui sont exactement ce que nous sommes en train de vivre et ce que nous devons continuer à faire. Il disait : c'est ce moment où j'arrive à **la netteté de la vérité** et où je retrouve la **puissance de l'espérance**.¹

« Le ressaisissement », « la netteté de la vérité », « la puissance de l'espérance » : ce niveau d'idéal et d'abstraction pour parler de la société française (dans cet extrait pour parler de l'égalité homme-femme) a été atteint par le seul Pompidou sous la V^e République. Et puisqu'il s'agit dans ce discours de la vision présidentielle de la société à travers les yeux de Charles Baudelaire, rappelons-nous ces mots, étrangement comparables, de 1970 :

Certes, nous avons l'habitude de la société dans laquelle nous vivons, nous en acceptons mieux les contradictions, nous oublions souvent de nous interroger sur les rivages vers lesquels nous entraîne ce que nous appelons le progrès. **Mais peut-être, comme le disait Baudelaire**, est-ce **l'épaississement de notre nature** qui seul nous empêche d'apprécier le milieu dans lequel nous respirons. L'avenir, c'est la réponse à trouver

¹ Macron, 25 novembre 2017, discours sur la Grande cause du quinquennat.

à la question que posait encore Baudelaire : « Qu'est-ce que le monde a désormais à faire sous le ciel ? », et de savoir si, comme il l'annonçait : « Nous périrons par où nous avons cru vivre. » C'est, en un mot, de donner un sens à la civilisation urbaine et technique dans laquelle nous sommes inexorablement engagés¹.

La convocation, commune aux deux hommes et que repère la machine, de Baudelaire, pour faire passer hautement un même message sociétal, est troublante. Plus profondément, si cela est possible, et toujours à propos de l'auteur des *Fleurs du mal*, Pompidou avait précisé à l'endroit de la jeunesse :

Vous êtes unis pour le progrès, pour tous les progrès, mais n'oubliez pas cette autre parole de Baudelaire : « Il ne peut y avoir de progrès vrai, c'est-à-dire moral, que dans l'individu et par l'individu lui-même. »²

Deux topiques fondamentaux de Pompidou, le progrès matériel en pleines années glorieuses et l'hymne à l'individu moral, figurent dans cet extrait. Or, c'est précisément ce que l'on entend chez Macron lorsqu'il s'inspire, selon la convolution/déconvolution, de Pompidou. De fait, la réflexion sur le progrès, revendiqué sous le terme de « progressisme », et la référence à l'individu comme horizon philosophique du politique, sont omniprésentes chez Macron comme dans cette solennelle déclaration, ramassée en quelques mots, prononcés à Versailles :

Le **progrès**, la dignité de **l'individu**, la force juste de la République sont nos boussoles et nous suffisent.³

« L'individu » est le terme-clef qui fait du pompidolisme un humanisme individualiste ou un humanisme libéral⁴ et qui fait du macronisme un pompidolisme. Sa distribution statistique est remarquable dans le corpus (figure 11) :

¹ Pompidou, 27 février 1970, discours prononcé au déjeuner du Commonwealth Club à San Francisco.

² Pompidou, 12 avril 1969, discours prononcé à Strasbourg.

³ Macron, 9 juillet 2018, discours devant le Congrès à Versailles.

⁴ Voir notre analyse du pompidolisme dans *Le discours présidentiel sous la V^e République*, chap. 3. « Pompidou : l'humaniste » (Presses des Sciences Po, 2012, p. 73-86). Voir également F. De Singly, *L'individualisme est un humanisme*, La Tour-d'Aigues, Les éditions de l'Aube, 2019.

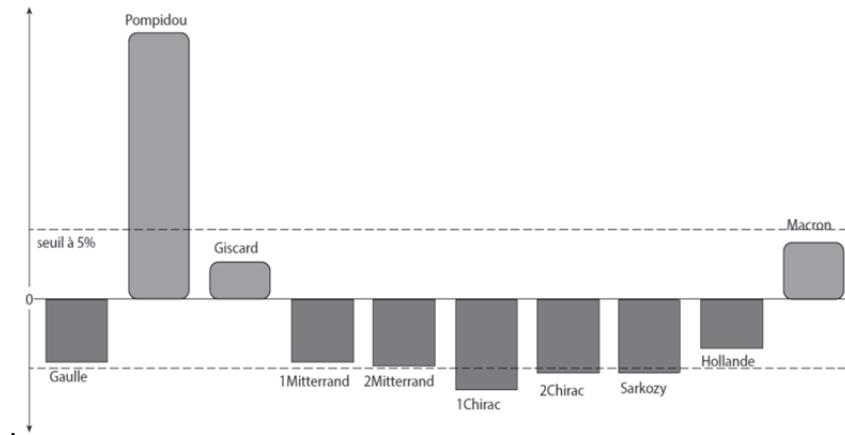


Figure 11. « Individu » dans le corpus (sur-utilisation de Pompidou et de Macron)

Macron l'utilise moins souvent que Pompidou certes, mais bien plus souvent que de Gaulle, Mitterrand, Chirac, Sarkozy ou Hollande ; et à ce constat on pourrait ajouter que Macron s'adresse non seulement très souvent aux « individus », comme Pompidou, mais également très souvent, plus simplement, aux « hommes » et aux « femmes » comme indiqué dans la figure 12.

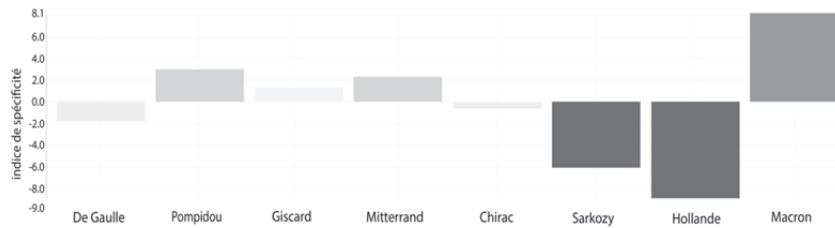


Figure 12. Les « homme(s) »+ les « femme(s) » dans le corpus sur-utilisés par Macron

Et comme chez Pompidou, les messages qu'il distille sont forts et explicites sur cet horizon libéral ou individualiste, aussi bien durant la campagne que durant la présidence.

Notre identité, c'est d'abord **cette démocratie respectueuse de l'individu**, des minorités, des droits fondamentaux, ce qu'on appelait du nom que je revendique encore « **la démocratie libérale** ».¹

En tout cas, l'appui intellectuel recherché auprès des grands auteurs pour exprimer une pensée philosophique, le plus souvent humaniste et individualiste, mais encore chrétienne, est une caractéristique incontournable et du corpus Macron et du corpus Pompidou. Ainsi, Hyperbase remarque cette déclaration de Macron à propos de la religion catholique pour la rapprocher de l'intertexte pompidolien :

[Ce] Questionnement [fondamental] qui lui donne la forme d'une architecture, d'une peinture, d'une philosophie, d'une littérature, qui toutes tentent, de mille manières, d'exprimer la nature humaine et le sens de la vie. « **Vénérable parce qu'elle a bien connu l'homme** », dit **Pascal** de la religion chrétienne. Et certes, d'autres religions, d'autres philosophies ont creusé le mystère de l'homme. Mais la sécularisation ne saurait éliminer la longue tradition chrétienne.²

De fait, en matière pascalienne, et en matière d'humanité, Pompidou avait lui aussi philosophé à haute voix dans ce qui peut être considéré dès lors comme une référence ou un prime-écho du corpus Macron, 50 ans auparavant :

C'est donc sur l'homme lui-même et sur la société que l'effort doit porter, pour les mettre en mesure de s'adapter aux données nouvelles de l'existence. **Le roseau pensant de Pascal est devenu maître de la nature**. Il n'en est, nous le voyons bien, que plus désemparé devant les problèmes que sa pensée lui pose à lui-même et qui, au bout du compte, se ramènent à définir le sens et le but de la vie. Dans cette recherche, qui est à la fois morale, sociale et métaphysique, un pays comme la France a vocation pour jouer un rôle important.³

Ainsi, de la même manière que Macron emprunte à de Gaulle le souffle historique et héroïque⁴, il partage avec Pompidou cette inquiétude « métaphysique » rare dans le discours d'un président.

¹ Macron, 17 avril 2018, discours devant le Parlement européen.

² Macron, 9 avril 2018, discours aux évêques.

³ Georges Pompidou, 12 février 1969, conférence au Cercle français à Genève

⁴ Sur l'intertexte gaullien cf. [Mayaffre 2021 ; *Un Jupiter ou un de Gaulle européen ?*, p. 127-144.]

Macron intrigue par sa complexité, son parcours, son discours, et Pompidou, avant lui, avait suscité l'interrogation. De manière volontairement énigmatique, *Le Monde* avait titré, en 1962, lors de la nomination de Pompidou Premier ministre de la France : « un banquier baudelairien à Matignon »¹. La triple dimension du personnage était alors pointée : l'« économiste » puisqu'il avait été banquier et homme d'affaires durant la période 1954-1962 ; le « poète » puisqu'il citait facilement Baudelaire ou Aragon ; le « politique » évidemment puisque l'homme accédait à une des plus hautes fonctions de la République

Plus jeune que Pompidou, Macron n'a pas toujours l'épaisseur de son devancier, mais il rivalise avec lui de richesse et de complexité. Comme Pompidou, il a été un cadre de la banque Rothschild où il devient banquier d'affaire en 2008 et décroche, selon la presse, le surnom de « Mozart de la finance »². Et à travers cet emploi, ce sont toutes les compétences économiques et la modernité d'un personnage lié aux milieux financiers qui peuvent être pointées. Comme Pompidou, Macron est un homme de livres et son discours est empli de références littéraires ou philosophiques ; le périmètre du discours de Macron dépasse ainsi souvent le champ étroit du technique ou de l'administratif pour embrasser l'orbe de la métaphysique et de la poésie. Comme Pompidou enfin, Macron est un politique, guidé par l'ambition de servir et de diriger, et un goût pour les lieux de pouvoir ; sa carrière ministérielle d'abord, présidentielle ensuite, ne peuvent faire oublier cet essentiel.

Et au soir du 14 mai 2017, n'aurait-on pas pu adapter la vieille manchette du *Monde* pour accueillir Macron à la présidence de la République : « Un banquier baudelairien à l'Élysée » ?³

*

En renonçant à l'idéal pour le réalisable, cette section a proposé des exemples de parcours méthodologique concrets pour baliser l'horizon interprétatif des textes : nous définissons les grands corpus numériques réflexifs comme une explicitation possible de l'intertexte – d'un certain intertexte comme il y a certains horizons d'attente – et convoquons le *deep learning* pour explorer – ou « apprendre » – cet intertexte.

¹ P. Viansson-Ponté, *Le Monde*, 14 avril 1962.

² *La Croix*, « Emmanuel Macron, ancien conseiller du prince aux manettes de Bercy » (26/08/2014).

³ Les observateurs ont très tôt remarqué sinon la triple dimension du personnage, au moins sa double personnalité [François-Xavier Bourmaud 2016].

Les développements récents de l'Intelligence artificielle des textes permettent ainsi de développer le programme des humanités numériques, dont Glenn Roe, par exemple, décline les ambitions dans les études littéraires. Pour l'auteur c'est bien une « phase nouvelle » qui s'ouvre devant et au-delà de l'académie :

...celle de la synthèse entre la textualité restreinte des collections numériques soignées et scientifiques et la contextualité grandissante des textes à l'ère de la numérisation massive. Nous entrons [...] dans le moment de l'intertexte; c'est-à-dire de l'interdépendance et de l'interconnexion des textes et des données... [C]ette intertextualité numérique requiert des approches synthétiques à l'étude littéraire, capables d'offrir en même temps, et la spécificité de l'œuvre d'art et la pluralité des systèmes d'information qui l'entoure; méthodes qui facilitent, d'une manière transparente et intuitive... le mouvement entre la macroanalyse ou la lecture à distance des collections massives et la lecture intensive ou la microanalyse des textes numériques; mobilité bilatérale et réciproque, enfin, entre le texte et le contexte par l'intermédiaire de l'intertexte. En prenant l'intertextualité comme base théorique de nos recherches informatiques, nous nous plaçons consciemment dans la riche tradition de l'étude intertextuelle...¹

Méthodologiquement, prolongeant les espoirs déjà anciens de Ludovic Tanguy et Théodore Thlivitis dans « Parcours interprétatifs (inter)textuels : vers une assistance informatique » [1999], nous demandons à l'algorithme de révéler les couches cachées ou profondes (*deep*) du texte, celles qui semblent receler ses empreintes inavouées et ses emprunts revendiqués : son intertexte. Si le texte a été comparé à un palimpseste – Macron écrit sur le parchemin de Pompidou ou de Giscard dans cette analyse –, gageons que son révélateur peut être numérique, sous condition que l'interprétation reste humaine.

CONCLUSION

Peut-être est-ce la nature politique, polémique, idéologique de nos corpus qui nous oblige, disions-nous en introduction ?

Quoi qu'il en soit, l'analyse du discours politique de ce chapitre épouse et rend impérieux les principes de la sémantique de corpus ou de la linguistique textuelle telles qu'elles se développent aujourd'hui.

¹ Roe G. (2014), « L'étude littéraire à l'ère du numérique: du texte à l'intertexte dans les 'digital humanities' », *Philologie im Netz Beiheft*, 7, p. 104.

Notre impératif est double et fonde une posture fondamentale : (i) reconnaître à l'interprétation sa force motrice dans l'établissement du sens – un sens politique toujours par définition en débat ; (ii) s'appuyer sur des *realia* textuelles, sinon objectives en tant que telles, toutefois objectivables et explicitées – pour cadrer notre subjectivité potentiellement partisane sur des discours idéologiques brûlants. En des mots plusieurs fois affirmés : il ne s'agit pas de prétendre prouver le sens – puisque le sens est pluriel, toujours interprété, toujours conte/stable/xtualisable – mais d'objectiver les parcours interprétatifs en les appuyant fermement sur des matérialités textuelles que le traitement repère et soumet au jugement.

Le traitement *deep learning* des textes, modèle convolutionnel puis déconvolutionnel, que propose l'Intelligence artificielle permet ainsi d'établir des « grandeurs textuelles » [Rastier 2007] inconnues, c'est-à-dire ces nouveaux observables linguistiques, qui ont été retenus par le réseau de neurones pour apprendre, reconnaître, classer un texte. La grammaire a ses mal nommées « parties du discours ». La linguistique textuelle n'aura jamais ses unités stables et univoques : elle doit composer avec des unités interprétées.

Nous avons vu que ces grandeurs – ici un mot, là un segment contigu, ailleurs une cooccurrence ou un motif lexico-grammatical discontinu – sont éligibles au rôle de « passages » en ce sens qu'elles apparaissent, *sur le plan de l'expression* comme des extraits du texte (à base de tokens ou d'étiquettes que l'algorithme identifie), en même temps qu'elles apparaissent *sur le plan du contenu*, pour l'Intelligence artificielle puis l'intelligence humaine, comme des carrefours ou rouages interprétatifs pertinents, grâce à une co(n)textualisation effective. En donnant une représentation mathématique complexe des mots (*embedding*) et en les embrassant en co(n)texte (*convolution*), le programme permet d'engrener, d'embrayer ou de convoluer l'interprétation ; en matière de traitement des textes, la *convolution* nous apparaît à ce stade comme la meilleure implémentation – certes réduite, locale mais effective – du cercle herméneutique ; la *déconvolution* – ici un indice appelé TDS (*Text Deconvolution Saliency*) – fait ressortir des zones de texte qui sonnent comme des éléments-carrefour du cercle herméneutique. Les relations entre le tout (le corpus d'apprentissage et de travail dans son ensemble) et les parties (les unités élémentaires, comme le mot) sont régulées par la puissance de machine et la complexité du réseau : particulièrement, la remise en

co(n)texte se fait de manière graduée et concentrique (la fenêtre de convolution, la section ou empan, le texte, le corpus). Au sommet, le traitement repère, au sein de corpus contrastifs ou plus précisément *réflexifs*, les éléments intertextuels qui traversent les discours. Ici, selon les principes mêmes de l'herméneutique, c'est le global du corpus qui informe localement les textes en suggérant les reprises, les échos, le dialogue ou le débat.

Car la politique est débat – c'est-à-dire fondamentalement intertexte et interdiscours, parcours et interprétation.

Les professions de foi sous la V^e République sont jeux et enjeux de langage dans une compétition électorale sans merci qui met en scène, depuis 60 ans, l'alternance gauche/droite jusqu'à l'élection, dans cette contribution, de la majorité macroniste. Le discours présidentiel de Macron, lui, ne dit pas seulement mais performe sa politique face à (ou avec) d'autres représentations langagières du monde et d'autres idéologies. Le président gouverne par les mots, sous condition d'arriver à leur donner un sens *ad hoc*.

En d'autres termes, notre démocratie contemporaine est logocratie. Dès lors, les méthodes d'analyse des discours et des textes – particulièrement la méthode que les pouvoirs privés ou publics promeuvent aujourd'hui à l'échelle planétaire : l'Intelligence artificielle – dépassent les simples querelles de laboratoire et la dispute scientifique, pour devenir enjeux globaux de société.

N.B. — Les auteurs ont plaisir à remercier ici Michèle Monte, Stéphane Lamassé et Benjamin Derouelle.