



HAL
open science

Sémantique de corpus numérique. Emmanuel Macron, président thaumaturge (2017-2023)

Damon Mayaffre, Laurent Vanni

► To cite this version:

Damon Mayaffre, Laurent Vanni. Sémantique de corpus numérique. Emmanuel Macron, président thaumaturge (2017-2023). *Espaces Linguistiques*, 2023, 6, 10.25965/espaces-linguistiques.562 . hal-04353800

HAL Id: hal-04353800

<https://hal.science/hal-04353800>

Submitted on 19 Dec 2023

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.



Sémantique de corpus numérique. Emmanuel Macron, président thaumaturge (2017-2023)

Digital corpus semantics. Emmanuel Macron, thaumaturge president (2017-2023)

Damon MAYAFFRE

CNRS - Université Côte d'Azur - BCL

damon.mayaffre@unice.fr

Laurent VANNI

CNRS - Université Côte d'Azur - BCL

laurent.vanni@unice.fr

URL : <https://www.unilim.fr/espaces-linguistiques/562>

DOI : 10.25965/espaces-linguistiques.562

Licence : CC BY-NC-SA 4.0 International

Résumé : L'Intelligence artificielle questionne la sémantique de corpus. En prenant en compte l'axe syntagmatique (CNN, *convolution*) et l'axe paradigmatique ou « rapport associatif » (RNN, *transformer*), l'architecture que nous présentons fournit une aide à l'interprétation, pour une sémantique de corpus augmentée. L'algorithme implémenté dans le logiciel Hyperbase est ici appliqué au corpus d'Emmanuel Macron (2017-2023), qui cultive, par son discours, la dimension thaumaturgique de son pouvoir : soigner, protéger, prendre soin.

Mots clés : texte, sémantique de corpus, intelligence artificielle, logométrie, Macron

Abstract: Artificial Intelligence raises questions about corpus semantics. By taking into account the syntagmatic axis (CNN, *convolution*) and the paradigmatic axis or "associative relationship" (RNN, *transformer*), the architecture we present provides an interpretation aid for enhanced corpus semantics. The algorithm implemented in Hyperbase software is applied here to the corpus of Emmanuel Macron (2017-2023), who cultivates, through his discourse, the thaumaturgical dimension of his power: to heal, to protect, to care.

Keywords: text, corpus semantics, artificial intelligence, logometry, Macron

Sémantique de corpus numérique

Damon Mayaffre et Laurent Vanni

Introduction

La sémantique de corpus – entendons, *de corpus numériques* – s'est construite en France, depuis la fin du XX^{ème} siècle et l'œuvre de François Rastier (notamment *La Mesure et le Grain. Sémantique de corpus*, 2011), un fondement épistémologique solide.

De fait, la corporalité numérique de nos données révolutionne notre rapport à l'empirie textuelle ; les outils numériques, particulièrement la statistique et l'Intelligence artificielle, font émerger de nouveaux observables linguistiques ; le Web donne accès aux intertextes nécessaires à la compréhension des textes étudiés ; ou encore, les liens hypertextuels balisent et architecturent nos parcours de lecture, pour une sémantique augmentée.

La présente contribution entend illustrer ces plus-values heuristiques et l'émergence de nouveaux observables en analyse du discours grâce à l'Intelligence artificielle et au traitement *deep learning*. Ces nouveaux observables permettront, en l'occurrence, d'objectiver des parcours interprétatifs dans le grand corpus numérique présidentiel français de la V^e République de de Gaulle à Macron, de 1958 à 2023, pour mettre au jour un discours thaumaturgique dans les termes de Marc Bloc ou un discours du *care* dans les termes de Carol Gilligan, chez le dernier président en date.

Au sein du corpus (considération macro ou globale) et au sein du paragraphe (considération meso ou régionale), les modèles *Transformers*, réputés depuis ChatGPT, repèrent l'activation ou le mécanisme d'attention (*self-attention*) existant entre les différentes grandeurs textuelles étudiées : les mots, la ponctuation, les étiquettes morpho-syntaxiques, les n-grams, les motifs (considération micro ou locale). Ce sont ainsi des unités simples ou des unités complexes mais toujours des unités *en contexte* – en relation global/local – qui se trouvent mises en relief, au service d'une sémantique non pas atomique, formelle et ontologisante mais d'une sémantique relationnelle, contextualisante et interprétative ; bref, au service d'une sémantique de corpus ou d'une *herméneutique matérielle numériques* (Mayaffre, 2010 ; Guaresi et Mayaffre, 2021).

1. Modèles

L'IA sait prédire mais peine encore aujourd'hui à décrire. Sans erreur, l'IA sait *classer* les textes et attribuer à leurs auteurs des productions anonymisées. De manière satisfaisante, l'IA sait également *traduire*, comme en atteste par exemple la plateforme DeepL (<https://www.deepl.com/translator>). Depuis l'automne 2022 et les améliorations de GPT-3 puis GPT-4, l'IA sait encore *générer* des textes dont on peut discuter la monstruosité – au sens littéral – et l'intérêt, mais constater la cohésion impeccable et la cohérence acceptable. Mais si l'IA sait classer, traduire ou générer des textes, elle peine encore à *décrire* la matière textuelle sur laquelle son attention se porte pour performer. Or c'est cette matière décrite qui intéresse le linguiste, le stylisticien ou l'herméneute.

De manière générale, l'*interprétabilité* des modèles IA est ainsi identifiée comme l'enjeu scientifique majeur pour les informaticiens du XXI^{ème} siècle¹. La boîte aujourd'hui n'est certes plus aussi noire qu'avant, mais la profondeur (le *deep*) de l'apprentissage, l'empilement itératif des couches cachées des réseaux de neurones artificiels, la quantité des (hyper)paramétrages possibles, les jeux algorithmiques démultipliés d'essais-erreurs empêchent souvent des retours-machine clairs et heuristiques (ie. interprétables). Opaque pour les informaticiens eux-mêmes, le deep learning devient ainsi problématique pour les SHS. Concrètement pour nous, la seule question qui vaille est aussi simple qu'irrésolue : que fait la machine ? C'est-à-dire, quelles sont les saillances ou réalités linguistiques du texte pertinemment prises en compte par les réseaux de neurones artificiels dans leurs tâches réussies de classification ou de génération ? Au-delà, et de manière générale pour la recherche scientifique comme pour la société, « l'intelligibilité du numérique », selon la récente revue éponyme (<https://intelligibilite-numerique.numerev.com/>), devrait être au cœur de toutes nos préoccupations puisque le numérique ne saurait plus être considéré aujourd'hui comme un nouveau support, comme le furent successivement par exemple le papyrus ou le papier, et comme une simple *techne* : le numérique est devenu depuis une vingtaine d'années une matrice des connaissances et la nouvelle *épistème* universelle, aux conséquences académiques, sociétales et anthropologiques majuscules ; *Homo numericus*. La « civilisation » qui vient titre récemment un auteur à succès (Cohen 2022).

1.1. Du syntagmatique

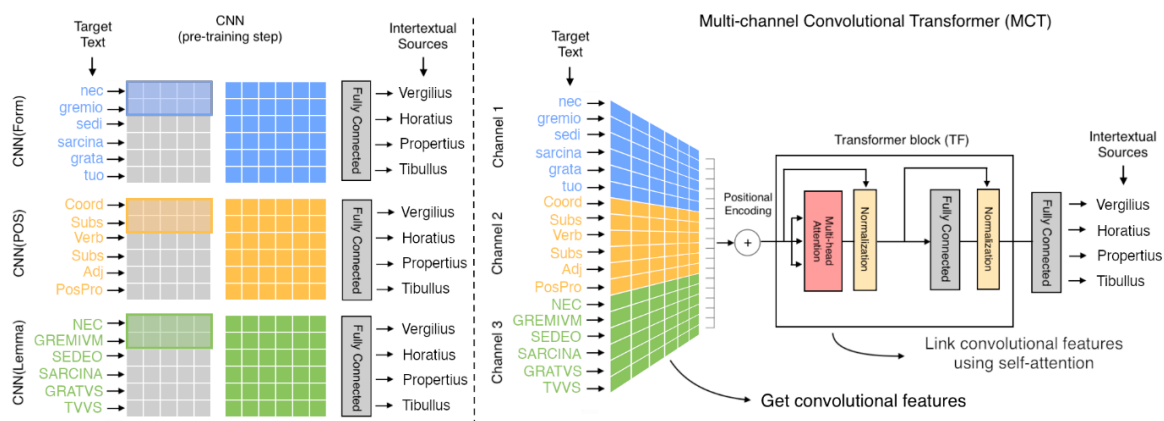
Les progrès de l'IA en termes d'interprétabilité des modèles et des descriptions textuelles sont désormais importants et ouvrent le champ d'une sémantique de corpus augmentée : de nouveaux observables apparaissent en effet que la philologie classique avait souvent décrits, parfois pressentis, quelques fois ignorés (Mayaffre et Vanni *éds.*, 2021). Quoique le terme ne soit pas théorisé, nous considérons ces nouveaux observables numérico-linguistiques comme des *herménèmes* : des endroits du texte hautement interprétés par la machine et donc – c'est une injonction – interprétables par l'humain, pour des parcours de lecture renouvelés.

L'architecture IA raffinée que nous présentons en un schéma ci-dessous (figure 1) et que nous appelons Multi-channel Convolutional Transformer (MCT) (Vanni et al. 2023) combine deux modèles informatiques dont les performances sont bien établies par la littérature en IA, tout en faisant écho, sans nécessairement le revendiquer, au fondement de la théorie linguistique ; écho

¹ Qu'il nous soit permis de ne pas gloser ici, faute de place, sur la différence entre « interprétabilité » et « explicabilité » des modèles, pour nous en tenir à l'idée générale de l'opacité problématique des sorties-machine deep learning.

linguistique fondamental, donc, puisque le double impératif syntagmatique et paradigmatisé, décrit dès l'origine par Saussure, dans le chapitre V du Cours, semble pour partie satisfait.

Figure 1. Multi-channel Convolutional Transformer



Le premier modèle qui sert de pré-apprentissage du corpus (partie gauche du schéma) est dit convolutionnel et est internationalement connu sous l'acronyme CNN (Convolutional Neural Network) : l'unité est considérée dans une fenêtre ou une *convolution*. Chaque mot a ainsi une représentation au sein d'un empan co-textuel identifié. Ici, dans cet article, l'empan convolutionnel est de 3 mots : par fenêtre glissante 3 mots qui précèdent le mot étudié, 3 mots qui lui succèdent. Ainsi la représentation numérique, c'est dire la considération ou le traitement du mot « chat » sera différente dans

- (i) Le (1) tout (2) petit (3) CHAT mange (1') une (2') souris (3')

et

- (ii) J'(1)ai (2) un (3) CHAT dans (1') la (2') gorge (3')

puisque les deux occurrences de « chat » ne partagent ou ne combinent aucun mot commun sur la chaîne, ni à gauche, ni à droite.

En revanche, le mot « chat » dans

- (iii) Le (1) très (2) gros (3) CHAT mange (1') une (2') souris (3')

aura, pour le modèle, une représentation numérique ou une identité numérique proche du premier exemple puisque les tokens qui l'environnent – le mot *in praesentia* dirait Saussure – sont proches (et même strictement identiques pour la partie droite, en position 1', 2' et 3').

Ajoutons encore que dans

- (iv) Le (1) très (2) gros (3) MATOU mange (1') une (2') souris (3')

« Chat » et « matou » auront une représentation numérique équivalente au regard de leur convolution identique.

Si les modèles convolutionnels – que l'on utilise également pour l'image en regardant les pixels dans leurs contiguïtés horizontales et verticales – apparaissent performants pour traiter des textes, c'est, fondamentalement, parce qu'ils restituent l'axe syntagmatique de la langue. La linéarité, l'ordre et les combinaisons des mots sur la chaîne du texte ne sont plus négligés, à rebours de la plupart des approches informatico-statistiques classiques qui fonctionnent le plus souvent selon un schéma d'urne (le texte comme un sac de mot ou comme une urne, dans lesquels les mots sont isolés, décontextualisés, et tirés indépendamment au hasard lors des tests statistiques). Les effets destructeurs du texte et destructeurs du sens que porte en elle la tokenisation sont pour partie corrigés. Le token ou l'item n'est plus seul : il est considéré dans son co-texte syntagmatique strict et immédiat. La tokenisation – étape nécessaire aux traitements informatiques – décontextualise, et atomise le sens. Elle prend le risque d'une posture sémantique ontologisante : à une forme (« chat ») correspondrait une représentation unique et définitive c'est-à-dire *in fine* un sens invariable ; et dans le corpus tous les « chats » se vaudraient jusqu'à compter ensemble et de manière indifférenciée, dans l'urne ou dans le sac, toutes leurs occurrences. Quoique cette position soit indéfendable linguistiquement dès le premier abord, lorsqu'on pense à l'homographie, à la polysémie, au sens figuré, à la connotation, etc. (sans même parler d'une sémantique interprétative aboutie), la tokenisation préside à la plupart des traitements informatiques.

La convolution, elle, reco(n)textualise – on pourrait dire « syntagmatise » – à l'image des n-grams dont on sait les performances récemment rappelées par Brunet et Vanni (2019) ou à l'image des segments répétés théorisés dès la thèse de Salem (1986) : le sens naît du co(n)texte, et ici, de manière minimale mais très forte, car de manière mathématisée, du co(n)texte syntagmatique immédiat (la convolution).

En résumé, une approche informatique initiale a pu considérer naïvement le texte comme un « sac de mots ». En complément, l'IA le considère aujourd'hui comme un « sac de mots *continus* » (CBOW : *continuous bag of words*). Le premier modèle utilisé, convolutionnel, de notre architecture MCT n'est pas le seul modèle IA relevant du CBOW, mais il en fait partie intégrante.

1.2. Du paradigmatique

Le deuxième modèle de notre architecture, combiné au premier, est connu internationalement sous le nom de *RNN Transformer* (Recurrent Neural Network Transformer) et préside, par exemple, à Chat-GPT dont le succès d'estime est important depuis l'automne 2022. Ce sont les mécanismes

d'attention entre les mots, en corpus, au sein du paragraphe, qui sont calculés ; techniquement, dans la littérature anglo-saxonne, nous parlons de *self attention*.

Si les *Transformers* apparaissent performants pour traiter des textes, c'est qu'ils semblent mettre au jour des relations paradigmatiques entre les mots du discours ; toutefois dans la définition ambitieuse et élargie que Rabatel donne du paradigme et que nous discuterons ci-après (Rabatel 2017). Le modèle sélectionne en effet les mots qui, de son point de vue, vont ensemble, sont peut-être substituables, ou entretiennent on ne sait encore quelle relation textuelle ; relation non pas en langue, mais, de manière effective, dans le corpus et dans le texte. Ce sont ainsi les affinités sélectives ou les « rapports associatifs » entre mots pour reprendre la formulation forte et initiale de Saussure dans le chapitre V du *Cours* que l'IA repère automatiquement dans le corpus d'apprentissage, puis éclaire dans les textes étudiés².

Sans *a priori* linguistique, ni syntaxique, ni phonologique, ni sémantique, les mots liés sont empiriquement corrélés statistiquement – ils ont quelque chose à voir entre eux numériquement au regard de la norme ou de l'usage dans le corpus d'apprentissage. Dans les termes logométriques traditionnels, issus de la statistique textuelle historique, les modèles *Transformer* repèrent donc des « cooccurrences » pertinentes entre mots, pour construire des isotopies parlantes, dans une tradition quantitative largement éprouvée en France (Lafon 1981, Heiden et Lafon 1998, Viprey 1997 et 2006, Brunet 2012, Martinez 2012, Mayaffre 2008a, 2008b et 2014, Guaresi 2016, Vanni et Mittemann 2016 etc.). Seulement, comme le terme de « cooccurrence » est réservé depuis 30 ans au strict et simple calcul de probabilité de la *co-présence* de deux mots (ou deux lemmes ou deux codes) dans un paragraphe, nous préférons dire qu'il s'agit de repérer ici des *corrélations* entre mots ou rapports associatifs au regard de leur usage linguistique en corpus. Pour reprendre strictement la terminologie actuelle des *Transformers*, il s'agit donc de repérer l'*attention* d'un mot pour un autre au sein d'une séquence de texte donnée ; pour reprendre la philosophie connexionniste même de l'IA, il s'agit de repérer la *connexion* entre les mots au sein du corpus et plus régionalement au sein du paragraphe.

² On se rappelle que le *Cours* n'utilise pas le mot « paradigmatique ». Aux « rapports syntagmatiques » répondent explicitement les « rapports associatifs ». Nous trouvons qu'une seule fois le terme « paradigme ». Les transcritteurs du *Cours* répètent en revanche « rapports associatifs », « associations », « séries associatives » ou « groupes associatifs » (Cf. la version du *Cours*, éd. Payot 1972, Chapitre V, p. 170-175). L'exégèse du *Cours* nous permet de préciser : le terme de « paradigme » est utilisé (une seule fois) pour mentionner *un cas particulier* de rapport associatif (« les paradigmes de flexion », p. 175). Notre propos sera *infra* d'étendre la notion de « paradigmatique » à des rapports qui ne se limitent pas à ce cas particulier ni aux seuls paradigmes grammaticaux ou flexionnels évidents (les pronoms par exemple, le nominatif, etc.) : des paradigmes ou « rapports associatifs » d'ordre textuel et herméneutique autrement plus délicats à déterminer.

Ces *rappports associatifs*, cette *attention*, cette *connexion* entre mots est dite *paradigmatique* en élargissant le concept de Saussure (éd. 1972, p. 170-175) à Benveniste (1970, p. 83-85), et de Benveniste à Rabatel (2017) car nous faisons nôtre la théorisation du paradigme non au niveau formel de la langue mais au niveau herméneutique du texte. Pour Rabatel en effet, dans son article « La notion de paradigme au défi du texte », il s'agit désormais de :

Plaider en faveur d'une extension de la notion de paradigme [...] au niveau des textes, articulant aussi syntaxe fonctionnelle et énonciation, description et interprétation, afin de montrer que la notion de paradigme n'est pas superposable avec celle des catégories prédéterminées (et bien fragiles) sur lesquelles se fondent la plupart de nos traditions grammaticales. (Rabatel, 2017, parag. 38, souligné par nous)

Rabatel précise ainsi ce que notre modèle repère de manière empirique :

Il vaudrait donc mieux *penser [...] que les rapports associatifs ne naissent pas seulement en fonction du stock de mots qui figurent dans notre mémoire, mais se construisent dans le texte*, en fonction des possibilités qu'offre (et tout autant ferme) l'organisation syntagmatique. (Rabatel, 2017, parag. 36, italique de l'auteur, souligné par nous)

Et plus loin, parlant, dans une expression herméneutique d'une « sémiose du global » (parag. 50), et semblant pressentir l'apport des traitements connexionnistes des textes, de conclure :

Le paradigmatique est ainsi reformaté – à partir d'une syntagmatique concevant les énoncés comme les maillons d'une chaîne textuelle, dans un cadre transphrastique –, rassemblant des éléments syntaxiquement différents, mais que rapproche leur aptitude à provoquer des effets interprétatifs similaires, sur la base d'une homologie fonctionnelle, au plan du texte.

Dans ce cadre, les paradigmes sont ouverts sur des virtualités de la langue comme sur celles qu'offrent les mises en relations effectuées par la parole vive en discours, avec des rapports associatifs inédits, produits par le texte lui-même. (Rabatel, 2017, parag. 50 et 51, souligné par nous).

Précisons deux éléments de complexité pour aider à la compréhension du traitement concrètement engagé par notre modèle. Ces *relations* entre mots, *rappports associatifs inédits* ou *cooccurrences* ne sont pas ceux de deux unités (par exemple deux mots), mais la relation de deux convolutions (les deux mots pris dans leurs contextes syntagmatiques respectifs immédiats, en vertu de la convolution décrite *supra*). D'autre part, les tokens repérés, qui entretiendraient entre eux une relation paradigmatique textuelle, sont le plus souvent représentés, en amont, par un embedding, comme dans Word2Vec et la plupart des logiciels de traitement informatique des textes. Les mots sont en effet représentés par une suite numérique (embedding) et mathématiquement représentés par un vecteur, de façon à ce que la machine, en corpus, en contexte, attribue, par exemple, à « professeur » et à « enseigner »

des profils numériques voisins (*embedding* voisin), au motif qu'ils ont été utilisés dans des co-textes proches ou dans une même perspective linguistique (ici, sans doute, une relation isotopique autour de l'éducation) ; des *embeddings* semblables donc, puis des vecteurs ressemblants (tirant dans la même direction) pour les deux mots, c'est-à-dire au fond un même *sens* du point de vue mathématique sinon directement du point de vue linguistique³.

Au final c'est donc, hautement, au sein du corpus, une corrélation (ou *self attention*) entre les vecteurs (*embedding*) d'unités textuelles considérées en cotexte (ou en *convolution*) que notre modèle repère. Par exemple, nous développerons ci-dessous la relation calculée dans le discours présidentiel entre le pronom « nous », la forme verbale « protégez » et le substantif « soignant » dans des paragraphes d'essence thaumaturgique chez Emmanuel Macron (*cf infra*, section 2).

1.3. Multichannel Convolutional Transformer

Enfin, notre architecture MCT est dite multichannel. Le multiniveau consiste, pour les images, à traiter indépendamment, avant de les croiser, les pixels bleus, rouges et jaunes. Pour le texte, et avant même que ne se développe l'IA, la communauté de linguistique de corpus a pris l'habitude de traiter les données textuelles au moins à trois niveaux : la forme graphique qui renvoie à la matérialité linguistique (« liront »), l'étiquette morpho-syntaxique qui renvoie à la grammaire voire à la syntaxe (verbe à la troisième personne du pluriel au futur) et le lemme qui renvoie au lexique et à l'unité de sens (lemme = « Lire »).

Un des principaux programmes se trouve là : le modèle IA doit savoir croiser ces trois niveaux élémentaires pour faire ressortir des *motifs* linguistiques complexes, possiblement lexicaux-grammaticaux ; autrement dit le modèle doit mettre au jour des marqueurs ou des observables linguistiques, non obviaux comme pouvait l'être un mot isolé ou un n-grams simples niveaux, aussi statistiquement significatifs qu'aient pu être ce mot ou ce n-grams. À noter, également, que ces nouveaux observables, de taille variable, peuvent être discontinus sur la chaîne ; la proximité entre termes et/ou codes grammaticaux est repérée dans la fenêtre (disons

³ Nous avons volontairement illustré la proximité possible des *embeddings* et des vecteurs de « professeur » et de « enseigner », pour des raisons sans doute sémantiques. Nous aurions pu reprendre notre exemple précédent : « chat » et « matou » auraient des *embeddings* proches au titre de leur usage co-textuel avoisinant... c'est-à-dire de leur synonymie. Mais précisons que les *embeddings* de « chantera » et de « partira » pourront se ressembler, dans leurs contextes, pour des raisons morpho-syntaxiques qui n'échapperont pas à la machine (verbe futur à la troisième personne du singulier). La vocation de l'*embedding* est de représenter un mot selon une multitude d'informations sans *a priori* ni exclusive ; empiriquement, une partie de l'*embedding* pourra être d'ordre sémantique, là où une autre partie de l'*embedding* sera d'ordre syntaxique. Au total, l'*embedding* que nous pratiquons contient une série de 76 nombres pour représenter un mot : chaque nombre donne une certaine information (peut-être sémantique, morphologique, phonologique, syntaxiques, etc.) sur le mot en contexte c'est-à-dire dans le corpus.

provisoirement le paragraphe) mais sans exiger leur stricte contiguïté : c'est bien le *motif* au sens de (Longrée et Mellet 2013) qui est recherché et non seulement le syntagme figé, la colligation ou le n-grams.

En résumé, la figure 1 (supra) montre la chaîne de traitement MCT que nous venons de présenter en trois points : (i) un pré-apprentissage du corpus par convolution (CNN – une représentation du mot *ad hoc*, selon son co(n)texte ou enchaînement syntagmatique), (ii) un traitement Transformer comme dans GPT (les mots dans leurs relations, rapports associatifs ou *self-attention*) ; et ceci (iii) en essayant de croiser trois niveaux de granularité linguistique : le mot-forme, le mot-lemme et la *part-of-speech*.

2. Parcours et résultats : Macron le président thaumaturge

Il n'existe pas de modèle valide sans résultat convaincant. Les deux premières sorties-machine du traitement des vœux d'Emmanuel Macron pour l'année 2023 sont éloquentes. Elles nous paraissent fertiles d'interprétation pour l'analyse du discours politique et illustrent ainsi concrètement l'herméneutique matérielle numérique que nous programmons grâce à l'IA.

En l'occurrence, Macron fait résonner la dimension thaumaturgique de son pouvoir présidentiel comme les rois, touchant les écrouelles, cultivaient cette dimension symbolique pour raffermir leur légitimité (Marc Bloch 1924). Si un de Gaulle asseyait l'autorité de son discours avant tout sur le/la geste régalien/ne, Macron joue ainsi de la thaumaturgie ou du *care* (cf. ci-après) durant une période politique difficile pour les Français.

En effet, la machine attire notre attention sur le passage-clef suivant :

[1] [...] avec vous, officiers et soldats dans notre armée, qui nous protégez, avec vous, policiers, gendarmes, pompiers, forces de sécurité, sauveteurs en mer, qui veillez sur cette soirée. Avec vous, médecins, soignants, patients. Avec vous, agents publics [...] (Macron, 31 décembre 2022, Vœux aux Français)

Ce passage extrait automatiquement, qu'il convient dans un premier temps de considérer dans sa globalité, est riche de contenu pour l'analyste politique puisqu'il se présente comme une longue énumération des professions, et finalement des Français, à qui Macron souhaite une joyeuse nouvelle année (*policiers, gendarmes, pompiers, sauveteurs, médecin, patients*, etc.). Mais notre modèle précise les mots, lemmes ou codes grammaticaux qui, au sein du passage, ont été (i) particulièrement activés par le réseau de neurones par convolution pour attribuer ce passage à Macron et qui (ii) sont reliés entre eux ou activés ensemble, selon la *self attention*, pour signer, selon Hyperbase *deep*, la prose macronienne. Se dessine alors aisément l'isotopie thaumaturgique du soin et de la guérison que Macron s'applique, de fait, à mobiliser dans son discours :

Figure 2. Passage-clé(I) des vœux de Macron, le 31 décembre 2022



Le lemme « soignant » (en vert, au singulier comme au pluriel) est ainsi activé par le réseau, avec un score d'activation important, et concentre, au sein du passage, le plus *d'attention*. Parlant en lui-même, le lemme se trouve en effet relié au verbe « protégez » et au pronom pluriel « nous » : Macron, dans une parole à la fois performative et curative, célèbre ainsi, dans cette période éprouvante post-Gilet Jaune, post-Covid et pré-inflation, la résilience, la guérison (« soignant »), la protection (« protégez ») de tous les Français, sinon déjà guéris et apaisés par les vertus du discours – au contraire le grave conflit autour de la réforme des retraites s'annonce déjà en ce début d'année 2023 – en tout cas, selon Macron, réunifiés et unanimes (« nous »)⁴.

A ce stade, il convient de rappeler que les sorties-machine IA articulent un traitement macro (global), meso (régional) et micro (local) du corpus. Traitement macro, évidemment, car l'identification du passage et de ses unités se fait par la considération de l'ensemble du corpus présidentiel (1958-2023) et, au sein de ce corpus de plus de 1000 discours et de 4 millions de mots, de l'ensemble du sous-corpus Macron (2017-2023) : c'est toujours par rapport à un corpus-norme – une norme interprétative que nous avons qualifiée ailleurs de *corpus réflexif* (Mayaffre, 2002) – que chaque événement local ou régional (ce mot-ci ou cet extrait-là) prend sens. Cette posture méthodologique voire épistémologique fondamentale (*le global détermine le local* selon les termes de l'herméneutique et de François Rastier, 2001) prend un tour concret en IA qui parle de « corpus d'apprentissage ». Ici cet extrait a donc été reconnu comme typiquement macronien au regard de la norme présidentielle depuis 1958, apprise par la machine. Le lecteur a pu peut-être penser un instant que la tonalité thaumaturgique était une caractéristique générique inévitable des discours de vœux ? Certes. Mais au-delà, la machine calcule qu'il s'agit d'une caractéristique de Macron *versus*

⁴ La performativité rassembleuse du « nous » macronien a été analysée ailleurs (Bouzereau et Mayaffre 2022). Notons ici, dans le passage, que le « avec vous » plusieurs fois répétés et le « nous » repéré sont tout simplement les slogans des affiches électorales de Macron, imprimées à des dizaines de milliers d'exemplaires, au printemps 2022 : « Avec vous ! » et « Nous tous ! ».

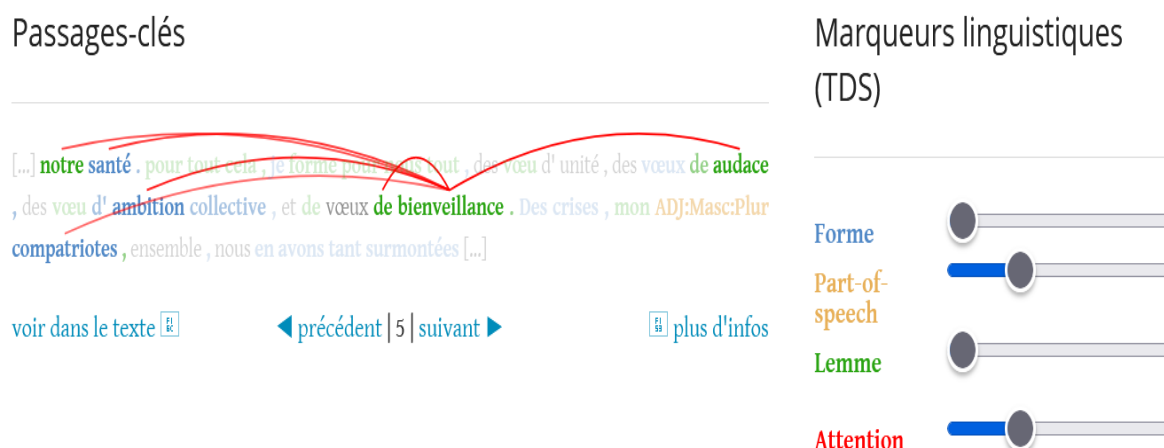
de Gaulle, Mitterrand ou Sarkozy, au sein du corpus présidentiel 1958-2023, tous genres de discours confondus.

Traitement meso ensuite, la fenêtre de traitement est une fenêtre paramétrable de 50 mots (voir l'extrait supra), dans laquelle les unités sont examinées. À ce palier intermédiaire, ce sont donc des extraits de texte relativement larges dans leurs cohésion/cohérence, dans leurs échos internes, dans leurs redondances ou particularités, qui sont considérés, et non pas les unités (les mots, ni même les phrases) prises isolément. Contrainte techniquement, la machine exige des fenêtres fixes pour qu'elles soient strictement comparables. Nous pouvons ainsi paramétrer le programme pour qu'il traite des passages de 25 mots, 50 mots ou 100 mots mais difficilement pour qu'il traite des fenêtres de tailles changeantes comme peuvent l'être les paragraphes naturels ou typographiques du texte. Quoiqu'il en soit, prolongeant la notion rastierienne de « passage » (Rastier, 2007), nous parlons ici de « passage-clefs » : des extraits sélectionnés pour leur pertinence interprétative (ie. interprétés par la machine / interprétables pour l'humain) (Vanni et al., 2022).

Traitement micro enfin, nous l'avons dit, car chaque token est considéré individuellement dans son co-texte particulier, enrichi de sa convolution propre, afin de lui attribuer une représentation numérique unique en fonction de sa place et de son importance locale ; et cette importance locale est dument chiffrée via un indice : le TDS (Vanni et al., 2018). De plus, la représentation locale unique se trouve enrichie dans le modèle par les relations individuelles avec les autres unités qui l'environnent (la self-attention). Ainsi dans cette phrase-ci « chat » aura une certaine représentation (un certain sens) et sera plus ou moins activé par le réseau de neurones (une certaine importance), enfin il entretiendra, à titre individuel, des liens avec tels ou tels mots du paragraphe (certaines affinités particulières). Dans cette phrase-là, en revanche, « chat » aura une autre représentation, pourra ne pas être activé par le réseau au regard de sa faible importance, et entretiendra, peut-être, des liens avec d'autres mots que précédemment.

Dans le discours de vœux pour l'année 2023 examiné pour l'occasion, le deuxième passage que l'IA repère comme typiquement macronien nous permet d'aller plus loin dans l'analyse du discours du président, et dans le parcours sémantique ou interprétatif suscité par l'IA.

Figure 3. Passage-clé (II) des vœux de Macron, le 31 décembre 2022



La tonalité de l'extrait est identique à celle du premier passage : c'est la célébration de l'unité d'un pays (« nous tous »⁵), à qui Macron présente ses meilleurs vœux, qui est à nouveau centrale. Seulement, ce n'est plus autour des « soignants » que le passage se structure mais autour du mot « bienveillance » à qui la machine accorde un score TDS maximal et un nombre de liens (self-attention) important avec les autres mots. Au-delà de la thaumaturgie blochienne, il convient ici de convoquer le *care* comme cadre rhétorico-politique fondamental du discours macroniste, comme il peut l'être dans certains courants de pensée américains contemporains (Gilligan, 1982).

L'idée du soin reste omniprésente puisque la « santé », ou plutôt « notre santé », est repérée par le réseau. Mais du sanitaire *stricto sensu*, nous passons à un soin plus global, à une attention individuelle plus générale, au bien-être et à l'épanouissement des personnes, et finalement, de manière plus générale encore à la *virtus* ou à l'éthique : « la bienveillance », « l'audace », « l'ambition »,⁶ ou ailleurs « l'amour »⁷. La sémantique formelle et une statistique décontextualisée n'auraient peut-être pas remarqué ce passage, ce tissage textuel et ces liens, qui mettent en évidence la dimension thaumaturge-care comme identité récurrente du discours présidentiel de Macron.

⁵ Ici le slogan de la campagne 2022 est strictement reproduit « Nous tous » (masqué par les liens sur la sortie-machine de la figure 2).

⁶ On notera l'étymologie « d'ambition » non éloignée de « bienveillance ». Selon le dictionnaire de l'Académie française, il s'agirait « d'entourer quelqu'un pour le solliciter »

⁷ On sait quelques phrases remarquables que nous analyserons ailleurs comme le 15 octobre 2017 : « Ce que je peux vous dire ici avec beaucoup de force c'est que ce qui a été au cœur de ma conviction ce qui m'a poussé à m'engager dans la vie politique, à m'engager dans un combat inédit celui d'aller sans parti à l'élection présidentielle et aujourd'hui chaque jour à me battre, c'est l'amour de mon pays et de mes concitoyens. C'est l'amour de mon pays [...], je ne veux pas que le moindre de nos concitoyens puisse un instant penser que je n'ai pas pour lui une pleine estime et un plein respect. J'aime et estime l'ensemble de mes concitoyens parce que je suis le président de toutes et tous... » (Cf. Mayaffre, 2023, *Cahiers de praxématique*, à paraître)

3. Le macronisme et le care

Il est vrai ici que les sorties IA peuvent être nourries et étayées par divers indices. Macron est un président qui se veut thaumaturge et *bene volens* (*benivolentia*, « bienveillance », veut du bien). Et cette posture ne date pas du discours du 31 décembre 2022 ci-devant examiné, ni de la pandémie COVID en 2020 ou 2021 : il signe constamment le discours - sans quoi la statistique et l'Intelligence artificielle n'auraient pas pu la repérer dans le corpus d'apprentissage – depuis 2017 voire depuis 2016 ou 2015.

Sans fausse modestie, en effet, Macron milite depuis le début de sa carrière politique pour une re-sacralisation du pouvoir présidentiel qu'il estime historiquement dégradé sous les quinquennats Chirac, Sarkozy ou Hollande. Et dans cette re-sacralisation, la dimension thaumaturgique, miraculeuse, mystérieuse, mystique doit retrouver son rang.

Durant la campagne de 2017, interrogé sur quelques élans rhétoriques particulièrement habités, sur la *charisma* du leader ou sur des postures surprenantes de l'orateur en meeting les bras en croix, Macron avait par exemple déclaré :

[2] La dimension christique, je ne la renie pas ; je ne la revendique pas. [...] La politique, c'est mystique. [...] C'est tout mon combat. C'est une erreur de penser que le programme est le cœur d'une campagne [...]. J'ai toujours assumé la dimension de verticalité, de transcendance, mais en même temps elle doit s'ancrer dans de l'immanence complète, de la matérialité. (Macron, 11 février 2017, interview au JDD)

Précèdent ou s'ensuivent alors quelques moments symboliques, que la presse qualifiera de jupitériens, comme les célébrations appuyées de Jeanne d'Arc, le discours pharaonique de la victoire devant la pyramide du Louvre, la convocation extraordinaire et inédite du Congrès au Château de Versailles pour ouvrir le quinquennat, le discours du 7 septembre 2017 à Athènes face à l'Acropole⁸. Déjà en juillet 2015, le jeune futur président avait intrigué avec des propos jugés alors polémiques là où ils étaient seulement prophétiques des deux quinquennats à venir :

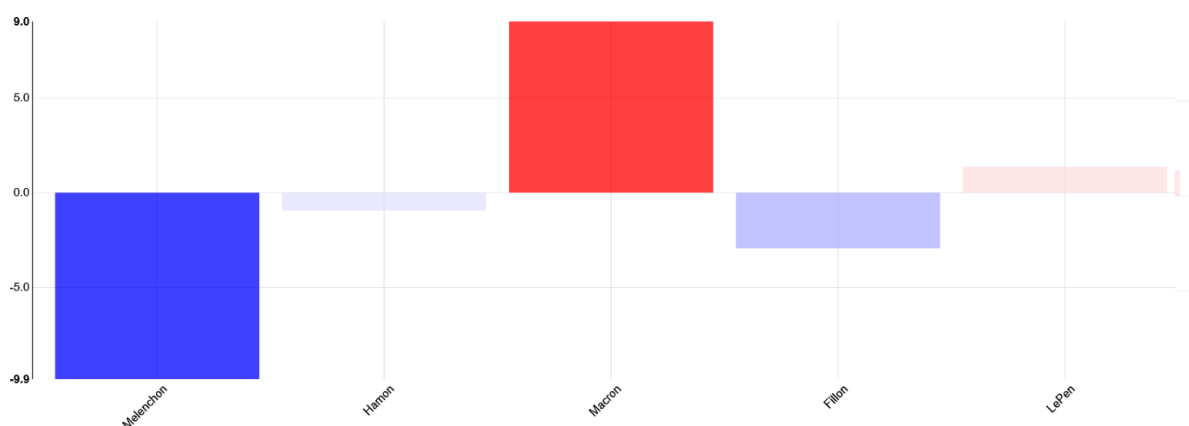
[3] La démocratie comporte toujours une forme d'incomplétude, car elle ne se suffit pas à elle-même. Il y a dans le processus démocratique et dans son fonctionnement un absent. Dans la politique française, cet absent est la figure du Roi, dont je pense fondamentalement que le peuple français n'a pas voulu la mort. La Terreur a creusé un vide émotionnel, imaginaire, collectif : le Roi n'est plus là ! On a essayé ensuite de

⁸ D'autres indices, plus ou moins anecdotiques, pourraient être versés comme le nom de l'agence de communication qu'Emmanuel Macron choisit pour mener sa campagne 2017 : « Gabriel et Jésus ». Ou, plus polémique, le « pèlerinage » d'Emmanuel Macron au Vatican, le 26 juin 2018, pour recevoir du Pape le titre honorifique de Chanoine de Latran, là où son prédécesseur avait refusé de faire le voyage, comme avaient refusé de le faire Georges Pompidou ou François Mitterrand.

réinvestir ce vide, d'y placer d'autres figures : ce sont les moments napoléonien et gaulliste, notamment. Le reste du temps, la démocratie française ne remplit pas l'espace. On le voit bien avec l'interrogation permanente sur la figure présidentielle, qui vaut depuis le départ du général de Gaulle. Après lui, la normalisation de la figure présidentielle a réinstallé un siège vide au cœur de la vie politique. Pourtant, ce qu'on attend du président de la République, c'est qu'il occupe cette fonction. (Macron, 8 juillet 2015, interview au journal Le 1)

Pour s'en tenir au corpus, à son traitement statistique, et reprendre seulement les mots activés par notre modèle dans les extraits précédents, le verbe *care* par excellence « protéger » est, de fait, quantitativement attesté dès la campagne 2017 : prononcé plusieurs centaines de fois, c'est une signature majeure du macronisme à sa naissance par rapport aux concurrents, Mélenchon, Hamon, Fillon ou Le Pen⁹ (figure 4).

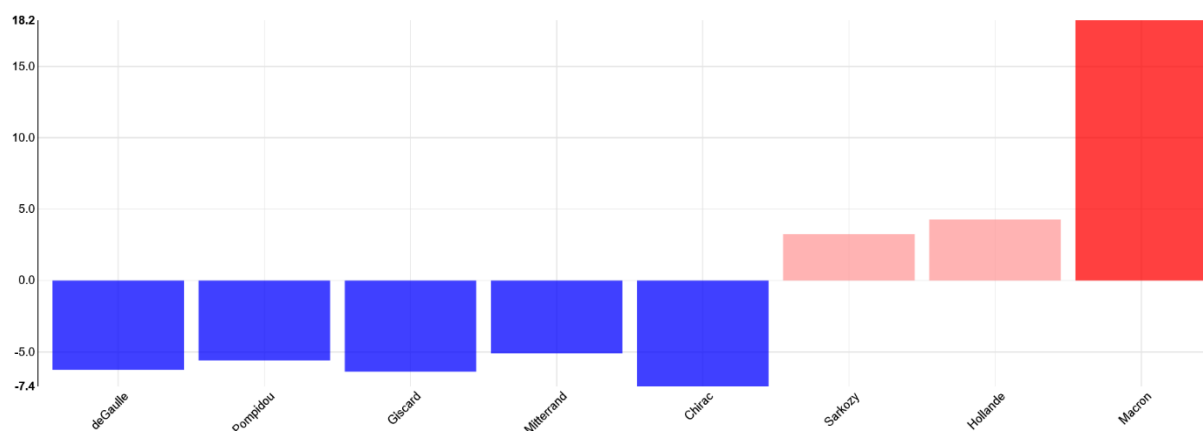
Figure 4. Surutilisation du verbe « protéger » chez Macron durant la campagne 2017 (Calcul des spécificités – sortie HYPERBASE 2023)



Tout le long de son premier quinquennat, Macron, en tant que président, martèle ainsi « protéger » ou « protection », avant, pendant, après le Covid-19 : on ne saurait l'imputer à la seule pandémie (figure 5).

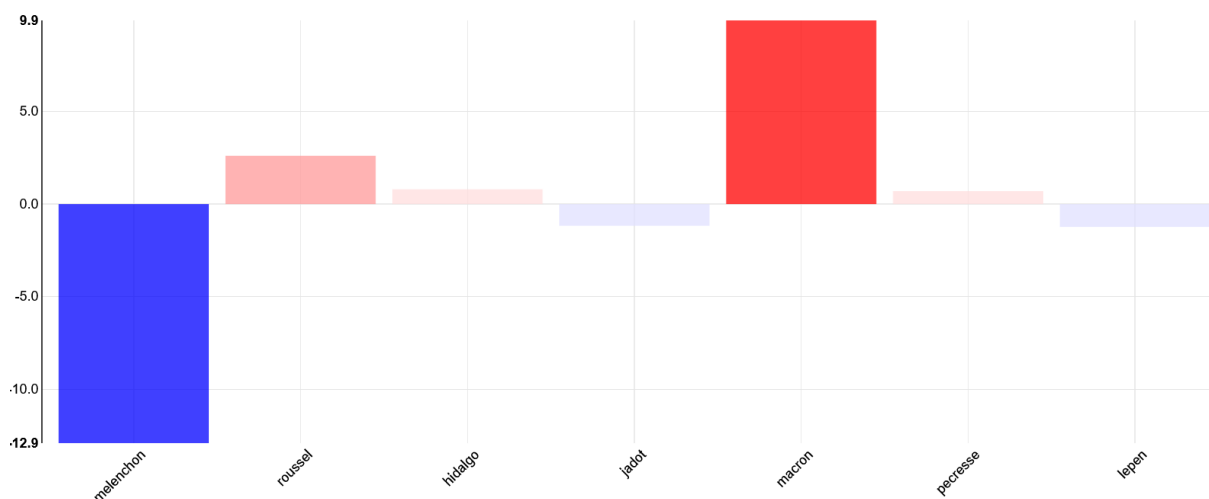
⁹ La récurrence du verbe chez Macron s'explique notamment par la sloganisation des formules : « libérer et en même temps protéger », ou simplement « libérer et protéger ».

Figure 5. Surutilisation du verbe « protéger » chez Macron dans le corpus présidentiel (1958-2023)
(Calcul des spécificités – sortie HYPERBASE 2023)



Et sans surprise, en 2022, pour sa réélection, Macron fait de « protéger » un des seuls mots fort d'une campagne muette par ailleurs pour cause de guerre en Ukraine ; ici la comparaison avec Mélenchon, Roussel, Hidalgo, Jadot, Pécresse, Le Pen est parlante (figure 6) :

Figure 6. Surutilisation du verbe « protéger » chez Macron durant la campagne 2022 (Calcul des spécificités – sortie HYPERBASE 2023)



Avec le verbe « protéger », nous posons que le terme « bienveillance », repéré par l'Intelligence artificielle comme central dans le deuxième extrait analysé, porte en lui beaucoup de la rhétorique macroniste.

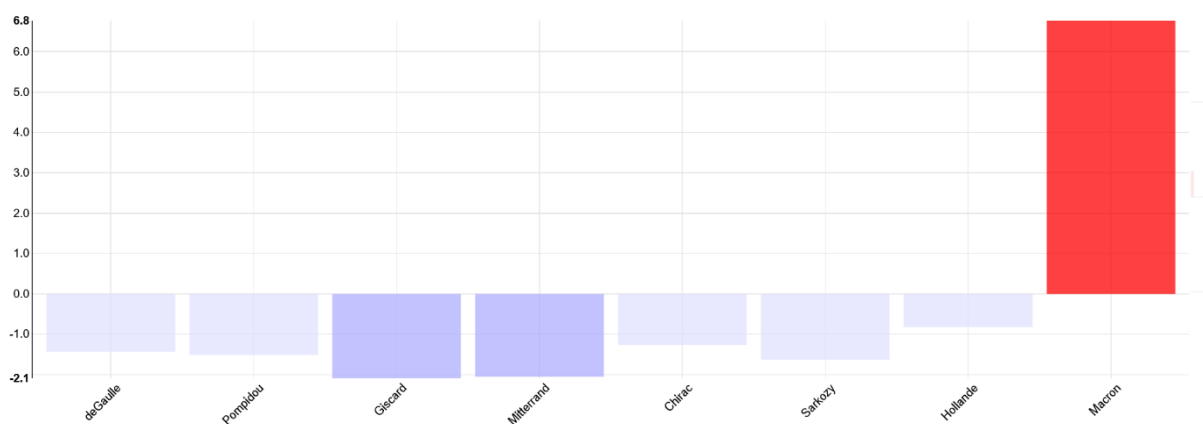
Nous trouvons des attestations significatives chez Macron dès la campagne 2017 lorsqu'il s'en prend, par exemple, à la « mondialisation non bienveillante » (28 février 2017, meeting à Angers) ou qu'il théorise, autour du mot, une autre façon de penser la politique et la démocratie :

[4] C'est pourquoi il nous faut toujours une exigence de bienveillance. Je vous la demande, non pas comme une idée simple, non pas comme une forme de naïveté de la vie politique, comme disent certains. La bienveillance, c'est d'abord et avant tout une hygiène démocratique. (Macron 4 avril 2017, meeting à Lyon)

En quoi la « bienveillance » serait une « hygiène démocratique » et peut-être même la condition de la démocratie ? Le discours n'ira pas plus loin, mais nous touchons ici à l'affirmation axiomatique d'une posture *care* fondamentale, plutôt que d'un programme politique, qui semble structurer le discours et l'idéologie d'Emmanuel Macron.

Statistiquement en tout cas, Macron est le seul président depuis 1958 qui sur-utilise massivement le mot dans ses discours :

Figure 7. Surutilisation de « bienveillance » chez Macron dans le corpus présidentiel (1958-2023)
(Calcul des spécificités – sortie HYPERBASE 2023)



Faute de « programme » politique établi en 2017 comme en 2022 – « C'est une erreur de penser que le programme est le cœur d'une campagne » avons-nous vu et souligné dans une précédente citation – le macronisme est difficilement identifiable. Néanmoins, nous pourrions le qualifier, d'un point de vue sociétal, de *carisme* au sens de Gilligan (1982), et plus encore au sens de Amartya Sen, l'économiste du bien-être selon le Prix Nobel, dont on perçoit l'influence chez Emmanuel Macron¹⁰. Grossièrement, la noblesse de la politique et l'idée de liberté et de justice, centrales dans

¹⁰ Nous pourrions peut-être aussi convoquer Ricoeur (1990) et ses réflexions sur l'éthique et le soi, quoique de manière plus éloignée.

tout projet politique, passent davantage par la prise en compte (bienveillante) des situations personnelles, concrètes et effectives, et par le développement des *capabilités individuelles* (*capability approach*), dont la bonne santé physique et psychologique, que par les lois aveugles, le Droit abstrait, l'Etat ou la Liberté désincarnée, les programmes généraux ou les planifications à grands traits (Sen, 2012).

Cette posture *care*, en guise de programme, en guise « d'exigence démocratique » fondamentale et, ici, dans le discours analysé par la machine, en guise de vœux à la nation, sera illustrée de mille manières dans une contribution à venir. Faute de place et parce que cette présente contribution est avant tout méthodologique nous l'évoquerons seulement de manière cursive renvoyant aux Cahiers de Praxématique 2023 pour un développement abouti (Mayaffre et Vanni, 2023 à paraître).

Le rôle dévolu à Brigitte Macron attire d'abord l'attention. Infirmière ligue du médecin général en chef Macron, que l'on voit en blouse lors de ses déplacements au moment du Covid, Brigitte Macron collecte des pièces jaunes pour l'hôpital. Si elle refuse de faire de la politique, elle milite avec une rare ferveur en faveur des enfants hospitalisés, des autistes, de la lutte contre les handicaps ou contre les violences conjugales¹¹.

Du point de vue féministe, précisément, nous avons montré ailleurs combien le couple Macron, Brigitte mais ici plus encore Emmanuel, concentrent ses efforts pour un féminisme *care*, plus que politique, adressé à la personne plus qu'à la citoyenne, adressé à la victime de viol et de violence plus qu'à l'ouvrière (Mayaffre, 2021, p. 261-275]. Loin du féminisme historique du MLF, souvent d'obédience marxiste, la concentration du discours sur les soins médicaux et psychologiques à apporter aux femmes battues est remarquable dans le corpus Macron. Une sortie médiatique, via les réseaux sociaux, illustre cette posture *care*, individuelle, personnalisée, incarnée jusqu'au prénom des victimes de violence – la violence contre les personnes comme antonyme de la bienveillance ou de la sollicitude inter-individuelle.

[5] Monica, Pascale, Taïna, Séverine, Nadine, Guo, Michèle, Béatrice, Isabelle, Patricia, Gulçin, Sylvie, Céline, Caroline, Maureen, Josette, Gaëlle, Ginette, Nelly, Nicole, Hilal, Julie, Georgette, Dolorès, Babeth, Fabienne, Caroline, Stéphanie, Chantal, Céline, Dalida, Nathalie, Sandra, Marie-Alice, Chloé, Yaroslava, Sandra, Martine, Martine, Laura, Pierrette, Moumna, Marilyne, Gwenaëlle, Mambu, Nathalie, Mariette, Priscilla, Maïté, Audrey, Mayie, Michèle, Chantal, Coralie, Leïla, Isabelle, Mesdames, la République n'a pas su vous protéger. Vos prénoms nous sont familiers, vous êtes nos amies, nos mères,

¹¹ L'engagement de Brigitte Macron en faveur des Pièces jaunes prolonge celui de Bernadette Chirac. A noter néanmoins que Bernadette Chirac fut politiquement engagée toute sa vie (conseillère municipale, adjointe au maire, conseillère générale) là où Brigitte Macron se réserve à ce positionnement *care*.

nos sœurs, nos filles. La violence qui vous a coûté la vie nous écœure, nous révolte. La République française se doit de protéger toutes celles dont la vie est aujourd'hui en danger. (Macron, FaceBook, 6 juillet 2019)

Le *carisme* de Macron est surtout étayé par la logométrie lorsqu'elle permet de retourner systématiquement au texte et de convoquer certains extraits autour des mots presentis par l'algorithme comme « bienveillance », « soin », « solidarités » (au pluriel car il s'agit moins du concept idéalisé que de ses déclinaisons pratiques, individuelles, quotidiennes).

La convocation des passages contenant le motif « LEM =Prendre + soin » (*prendra soin, a pris soin, devoir prendre soin, prendre le plus grand soin, etc.*) est particulièrement éclairante, surtout si l'on considère qu'il constitue une traduction fidèle du verbe « to care ».

Devant les évêques de France, dans des mots très choisis, Emmanuel Macron explique :

[6] Vous avez ainsi établi un lien intime entre des sujets que la politique et la morale ordinaires auraient volontiers traités à part. Vous considérez que notre devoir est de protéger la vie, en particulier lorsque cette vie est sans défense. Entre la vie de l'enfant à naître, celle de l'être parvenu au seuil de la mort, ou celle du réfugié qui a tout perdu, vous voyez ce trait commun du dénuement, de la nudité et de la vulnérabilité absolue. Ces êtres sont exposés. Ils attendent tout de l'autre, de la main qui se tend, de la bienveillance qui prendra soin d'eux. (Macron, 9 avril 2018, discours aux évêques de France)

Le *carisme* macronien se confond, ici, avec la charité chrétienne. Et Macron pointe ici l'articulation entre ce qu'il appelle la « politique » et ce qu'il nomme la « morale », et que la littérature *care* nomme plutôt l'éthique. La formulation « la bienveillance qui prendra soin d'eux » confine au tautologique, mais aucune redondance n'est inutile à Macron pour dire la compassion, la sollicitude, le bien-être : le *care*.

C'est naturellement durant le COVID, c'est-à-dire lors d'un moment suspendu durant lequel les hommes et les sociétés semblent redécouvrir leur vocation à mourir, que Macron joue de la thaumaturgie et du *care* avec quelques élans remarquables à la télévision lors d'adresses aux Français, suivies par des millions de téléspectateurs.

[7] Chacun d'entre nous détient une part de la protection des autres, à commencer par ses proches. Je compte sur vous aussi pour prendre soin des plus vulnérables de nos compatriotes, ne pas rendre visite à nos aînés. [...]. Écrivez, téléphonez, prenez des nouvelles, protégez en limitant les visites. Je compte sur vous, oui, pour aussi aider le voisin qui, lorsqu'il est personnel soignant, a besoin d'une solution de garde pour ses enfants pour aller travailler et s'occuper des autres. (Macron, 3 décembre 2020, Adresse aux Français).

L'entraide, « aider le voisin », « s'occuper des autres » : durant le confinement, le politique semble s'arrêter, le Parlement est mis entre parenthèse, les conflits sociaux sont suspendus : le *care* devient l'alpha et l'omega du discours présidentiel.

Pour en revenir aux mesures féministes qui se concentrent sur la lutte contre les violences faites aux femmes plutôt que sur la lutte traditionnelle contre les inégalités politiques, économiques ou salariales, les extraits vont dans le même sens :

[8] Si elle [une femme victime] est prise en charge aux urgences, je souhaite qu'elle ne soit pas obligée ensuite d'aller porter plainte le lendemain au commissariat. Facilitons son parcours, mieux accompagner c'est prendre soin de ces femmes et les aider psychologiquement. Les violences subies ont un effet dévastateur et là aussi, les chiffres ne disent pas tout de l'épaisseur des vies, mais ils racontent beaucoup de l'horreur. Il y a 5 fois plus de tentatives de suicide chez les femmes victimes de violence, elles perdent une à quatre années d'espérance de vie. Il s'agit donc d'organiser la prise en charge psycho-traumatique des victimes. Nous souhaitons ainsi créer dès 2018 dans les centres hospitaliers des unités spécialisées dans la prise en charge globale du psycho-trauma. La résilience des femmes agressées ne passe pas seulement par une solution judiciaire mais par la consultation, par la reconstruction. (Macron, 25 novembre 2017, Discours sur la Grande cause nationale)

Ici c'est non seulement la solution politique mais la solution judiciaire qui sont secondarisées pour mettre en avant, en revanche, la bienveillance *care* : « prendre soin de ces femmes », « les aider psychologiquement », « organiser la prise en charge psycho-traumatique ».

Les deux derniers extraits (toujours sélectionnés autour du motif « prendre + soin ») confirment l'analyse, et mériteront commentaires plus approfondis ailleurs. Mais apprécions tout de suite la dernière phrase du premier, issu d'une très solennelle prise de parole à la télévision :

[9] Et les vertus qui, aujourd'hui, nous permettent de tenir, seront celles qui nous aideront à bâtir l'avenir, notre solidarité, notre confiance, notre volonté. Alors prenez soin de vous, prenons soin les uns des autres. (Macron, 13 avril 2020, Adresse aux Français)

Dans un écho intertextuel palpable le « aimez-vous les uns les autres » christique est mobilisé par l'Élu de la nation, à la première personne du pluriel (prenons) pour plus d'effet encore.

Quant au deuxième extrait qui date du 31 décembre 2021, son aspect remarquable tient dans la redite – parfois mot à mot –, de l'extrait que l'IA avait initialement remarqué dans les vœux en 2022 : le *care* chez Macron n'est pas une saute d'humeur ni un hasard, mais une ligne conductrice d'un discours général qui n'hésite pas à se répéter.

[10] Je veux ce soir, une fois encore, en votre nom à tous, témoigner notre reconnaissance pour nos personnels soignants, nos armées, nos forces de l'ordre, nos sapeurs-pompiers,

nos auxiliaires de vie nos aides à domicile et tant d'autres professions, tous engagés ce 31 décembre comme chaque jour pour nous protéger, pour prendre soin de nous. (Macron, 31 décembre 2021, Vœux aux Français. cf. nécessairement l'extrait du 31 décembre 2022 supra)

Conclusion

La sémantique de corpus numérique renonce prudemment à réduire la relation entre expression et contenu à une relation bijective. À une forme ne correspond évidemment pas un sens unique ; parfois des sens opposés. Et les contenus s'expriment, d'évidence, de diverses manières dans la matérialité linguistique des textes. Loin d'un rapport bijectif, ce qui relie un signe linguistique à son sens en discours est plutôt de l'ordre de l'analogie, du métaphorique, de l'interprétation contextuelle.

Conséquemment, la pertinence que le linguiste demande à un programme IA n'est pas d'extraire le sens ; puisque le sens est construit par la lecture et non déjà-là, supposément offert, dans le texte. Il s'agit donc non pas de prouver le sens, mais d'implémenter des parcours de lecture dont on jugera la pertinence à l'aune de leurs productivités sémantiques.

La lecture numérique que proposent les algorithmes d'intelligence artificielle mettent ainsi au jour des passages ou des zones du texte qui doivent être considérés comme des *textèmes*, des *herménèmes* ou des *interprétants* : la machine les a considérés comme pertinents pour identifier un texte, un auteur, un président, ou pour produire un résumé, une traduction, un texte pastiche.

Outre leur efficacité empirique – puisqu'ils président à des tâches de classification réussies ou des tâches de génération convaincantes ; comme ils ont réussi à nous mettre dans cette contribution sur la voie de l'idéologie macroniste – ces chevilles interprétatives, passages ou herménèmes numériques semblent performants car ils considèrent l'axe syntagmatique (la combinaison sur la chaîne, *in praesentia*) – et l'axe paradigmatique (la sélection des mots, non pas *in absentia* en mémoire, mais dans le corpus) pour rendre compte au final de l'entrelacs textuel et de motifs significatifs

Le modèle de deep learning présenté, et implémenté aujourd'hui dans le logiciel libre Hyperbase (<http://hyperbase.unice.fr>), a ainsi permis de montrer comment Macron développe une rhétorique thaumaturgique et multiplie dans son discours un vocabulaire *care*. Loin d'un programme revendiqué, loin d'un corpus idéologique avoué, Macron ramène l'idéal politique à la bienveillance individuelle, la justice sociale au soin personnel, et le politique à la sollicitude inter-individuelle. Au moment où ces lignes sont écrites, Macron prescrit aux Français une réforme des retraites contestée. Mais là où ses opposants dénoncent une régression sociale historique, un déni

démocratique sans précédent ou une injustice collective, Emmanuel Macron raisonne différemment et fait valoir que désormais chaque travailleur âgé bénéficiera d'un médecin attitré et d'une visite médicale personnalisée, susceptibles de l'exonérer, en cas de problème de santé personnel, des deux années de travail supplémentaires que la nouvelle loi promet.

Références

BENVENISTE Émile, 1970, « L'appareil formel de l'énonciation », *Langages*, 17, p. 12-18 ; nouv. éd., 1974, *Problèmes de linguistique générale*, n° 2, Paris, Gallimard, p. 79-88.

BLOCH Marc, [1924] 1983, *Les rois thaumaturges : études sur le caractère surnaturel attribué à la puissance royale, particulièrement en France et en Angleterre*, Paris, Gallimard.

BRUNET Étienne, 2012, « Nouveau traitement des cooccurrences dans Hyperbase », *Corpus*, n° 11 | 2012, URL : <http://journals.openedition.org/corpus/2275> ; DOI : <https://doi.org/10.4000/corpus.2275>

BRUNET Étienne et VANNI Laurent, 2019, « Deep learning et authentification des textes », *Textes et Cultures*, vol. XXIV, n° 1, p. 1-34. {hal-02561039}

COHEN Daniel, 2022, *Homo numericus. La « civilisation » qui vient*, Paris, Albin Michel.

GUARESI Magali, 2016, « Cooccurrences, contrastes et caractérisation textuels. Applications à un corpus de professions de foi électorales (1958-2007) », *13th International Conference on Statistical Analysis of Textual Data*, Université Nice Sophia Antipolis-CNRS, Nice, France. p. 439-451 (hal-01371551).

GILLIGAN Carol, 1982, *In a Different Voice*, Cambridge, Harvard University Press.

GUARESI Magali et MAYAFFRE Damon, 2021, « Intelligence artificielle et discours politique. Quelles plus-values interprétatives ? Application aux corpus parlementaire et présidentiel contemporains » in Damon Mayaffre et Laurent Vanni (éds), *L'intelligence artificielle des textes*, Paris, Champion, p. 131-182. [hal-03347997]

LAFON Pierre, 1981, « Analyse lexicométrique et recherche des cooccurrences », *Mots*, n° 3, p. 95-148.

LONGREE Dominique et MELLET Sylvie, 2013, « Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours », *Langages*, n° 189, p. 65-79.

HEIDEN Serge et LAFON Pierre, 1998, « Cooccurrences, La CFDT de 1973 à 1992 », *Des mots en liberté, Mélanges Maurice Tournier*, tome 1, Fontenay-aux-Roses, ENS Editions.

MAYAFFRE Damon, 2008-a, « De l'occurrence à l'isotopie. Les co-occurrences en lexicométrie », *Sémantique et Syntaxe*, n° 9, p. 53-72.

MAYAFFRE Damon, 2008b, « Quand travail, famille, patrie co-occurrent dans le discours de Nicolas Sarkozy. Étude de cas et réflexion théorique sur la co-occurrence », in Serge Heiden et Bénédicte Pincemin (ed.), *JADT 2008*, p. 811-822.

MAYAFFRE Damon, 2010, *Vers une herméneutique matérielle numérique. Corpus textuels, Logométrie et Langage politique*, Habilitation à Diriger des Recherches, soutenue à Nice le 30 avril 2010. [Vol. 1 sur Hal : <http://tel.archives-ouvertes.fr/tel-00655380>]

MAYAFFRE Damon, 2014, « Plaidoyer en faveur de l'analyse de donnée co(n)textuelles. Parcours cooccurrentiels dans le discours présidentiel français (1958-2014) », in Émilie Née, Mathieu Valette, Jean-Michel Daube, Serge Fleury (ed.), *JADT 2014*, p. 5-32.

MAYAFFRE Damon, 2021, *Macron ou le mystère du verbe. Ses discours décryptés par la machine*, La Tour-d'Aigues, L'Aube.

MAYAFFRE Damon et VANNI Laurent (éds), *L'intelligence artificielle des textes. Des algorithmes à l'interprétation*, Paris, Champion.

RABATEL Alain, 2017, « La notion de paradigme au défi du texte : l'exemple des paradigmes entrecroisés de l'énonciation et du point de vue », *Signata*, n° 8, p. 175-204.

RASTIER François, 2011, *La Mesure et le Grain*, Paris, Champion.

RICOEUR Paul, 1990, *Soi-même comme un autre*, Paris, Seuil.

SALEM André, 1986, « Segments répétés et analyse statistique des données textuelles », *Histoire et Mesure*, n° 1-2, p. 5-28.

MARTINEZ William, 2012, « Au-delà de la cooccurrence binaire... Poly-cooccurrences et trames de cooccurrence », *Corpus*, n° 11, p. 191-216.

SAUSSURE Ferdinand de, ([1916] 1972), *Cours de linguistique générale*, Lausanne-Paris, édition Payot.

SEN Amartya, 2012, *L'Idée de justice*, Paris, Flammarion.

VANNI Laurent et MITTMANN Adiel, 2016, « Cooccurrences spécifiques et représentations graphiques, le nouveau "Thème" d'Hyperbase ». *JADT 2016 - Statistical Analysis of Textual Data*, Nice, France. p. 295-305. (hal-01359413)

VIPREY Jean-Marie, 1997, *Dynamique du vocabulaire des Fleurs du mal*, Paris, Champion.

VIPREY Jean-Marie, 2006, « Structure non-séquentielle des textes », *Langages*, n° 163, p. 71-85.