



HAL
open science

La logométrie en histoire: une herméneutique numérique. Exploration d'un corpus de professions de foi électorales de député-e-s (1958–2007)

Magali Guaresi

► **To cite this version:**

Magali Guaresi. La logométrie en histoire: une herméneutique numérique. Exploration d'un corpus de professions de foi électorales de député-e-s (1958–2007). *Digital Studies/Le champ numérique*, 2019, 9 (1), 10.16995/dscn.349 . hal-02292881

HAL Id: hal-02292881

<https://hal.science/hal-02292881>

Submitted on 20 Sep 2019

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.



Open Library of Humanities



Part of the Ubiquity
Partner Network

Digital Studies /
Le champ numérique

Research

How to Cite: Guaresi, Magali. 2019. "La logométrie en histoire: une herméneutique numérique. Exploration d'un corpus de professions de foi électorales de député-e-s (1958–2007)." *Digital Studies/Le champ numérique* 9(1): 14, pp. 1–15. DOI: <https://doi.org/10.16995/dscn.349>

Published: 12 August 2019

Peer Review:

This is a peer-reviewed article in *Digital Studies/Le champ numérique*, a journal published by the Open Library of Humanities.

Copyright:

© 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Open Access:

Digital Studies/Le champ numérique is a peer-reviewed open access journal.

Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

RESEARCH

La logométrie en histoire: une herméneutique numérique. Exploration d'un corpus de professions de foi électorales de député-e-s (1958–2007)

Magali Guaresi^{1,2}

¹ Université Libre de Bruxelles, BE

² Université Côte d'Azur, FR

magali.guaresi@gmail.com

Le passage du papier au numérique constitue, pour les historien-ne-s ayant affaire aux textes, une rupture majeure. En enrichissant les archives textuelles – devenues des hypertextes, délinéarisés et dénaturalisés –, le numérique augmente les capacités de navigation dans les corpus, révèle de nouveaux observables linguistiques et transforme nos pratiques de lecture. La logométrie, méthode de lecture assistée par ordinateur, participe depuis quatre décennies à ces évolutions en proposant des outils alliant lecture qualitative et approche quantitative de corpus numériques dans le but de formaliser des parcours interprétatifs heuristiques. Cette contribution en présentera les principales fonctionnalités, appliquées à un corpus de professions de foi électorales rédigées sous la Cinquième République (1958–2007).

Mots-clés: Logométrie; corpus numérique; herméneutique numérique; discours électoral; histoire du discours politique

For historians dealing with text, the changeover from paper to digital sources represents a major change. By expanding text archives – turned into hypertexts, un-linear and denaturalised –, the digital increases the navigation capabilities within corpora, reveals new linguistics aspects that can be observed, and transforms our reading practices. Logometry, the method using computer-assisted reading, has been contributing to these evolutions over the last four decades by proposing tools that combine qualitative reading and quantitative approaches to digital corpora, with the goal of formalising the hermeneutic interpretive process. This article will present its principle functions through its application to a corpus of electoral manifestos written under the Fifth Republic (1958–2007).

Keywords: Logometry; digital corpus; digital hermeneutics; electoral discourse; history of political discourse

Introduction

« L'introduction de l'informatique dans la pratique du travail historique est aujourd'hui chose faite » écrivait déjà Jean-Philippe Genet (1986, 7) dans une livraison de la revue *Histoire et Mesure* à la fin des années 1980. L'historien s'interrogeait à cette occasion sur les conséquences de cette mutation technique sur les pratiques historiennes. Trois décennies plus tard, à l'heure du triomphe du numérique dans les sciences humaines et sociales, nous nous proposons de prolonger modestement la réflexion. Plus précisément, cette contribution s'attachera à questionner les formes contemporaines de la lecture des archives textuelles, au cœur du travail des historien-ne-s, et à montrer la fertilité de l'assistance de l'ordinateur – et plus particulièrement de la méthode logométrique (Mayaffre 2010) – dans l'exploitation et l'interprétation d'importants corpus de textes.

Après avoir brièvement évoqué les conséquences épistémologiques majeures du passage au numérique sur les pratiques de lecture des archives (ou des corpus) textuelles par les historien-ne-s, nous présenterons quelques exemples de traitements logométriques, non exhaustifs (ni du point de vue méthodologique, ni du point de vue de l'interprétation historique) mais seulement suggestifs. L'étude matérielle, assistée par une méthode d'analyse informatisée, d'un corpus de professions de foi électorales écrites par des femmes et des hommes élu-e-s député-e-s en France entre 1958 et 2007¹ ambitionne, en effet, d'illustrer le propos épistémologique et méthodologique par des résultats concrets obtenus en histoire politique contemporaine. L'objectif de souligner la fertilité d'une hyperlecture des textes et de mettre en avant les avantages d'une lecture non pas seulement intuitive mais outillée et contrôlée dans l'objectivation de parcours interprétatifs susceptibles de nourrir l'enquête historique sera poursuivi.

Du texte en histoire: la révolution numérique

Sans prétendre épuiser l'ensemble des aspects de la révolution numérique dans les sciences historiques, indiquons que le passage du support papier au support numérique ne saurait être considéré comme une simple transformation technique.

¹ Concrètement, le corpus est composé de la quasi-intégralité des professions de foi des femmes députées sous la Cinquième République. Il rassemble également, pour les besoins de la comparaison, un échantillon de professions de foi d'hommes rédigés dans des conditions de production comparables (même année, même département, même bord politique que les femmes). L'ensemble compte plus de 700 textes pour 500 000 occurrences. Pour consulter l'ensemble des analyses menées sur ce corpus, nous nous permettons de renvoyer à Guaresi 2018.

Posons plutôt qu'il s'agit là d'une révolution culturelle et épistémologique, inscrite dans une dimension pluriséculaire, qui modifie nos appréciations du texte et nos pratiques de lecture (Vandendorpe 1999; Goody 2007; Darnton 2011).

Outre la disponibilité immédiate des sources, l'abondance nouvelle des données ou encore la dynamique inédite des archives (Rygiel 2011), la généralisation du numérique entraîne surtout une évolution de notre rapport au texte et à la lecture, c'est-à-dire, plus concrètement, aux conditions de la connaissance historique, informée en grande partie par le langage (Guilhaumou 2006)². De façon évidente, le numérique a définitivement dénaturalisé et dématérialisé le texte. La pluralité des formats, des possibilités d'étiquetages et d'annotation révèle la dimension artefactuelle du texte. Ce dernier ne saurait plus être abordé comme un objet naturel, fixe et transparent qui donnerait un accès au sens et au passé de façon immédiate. Au contraire, c'est un objet construit qu'il convient d'analyser dans sa complexité linguistique.

Principalement, le support numérique et l'outillage informatique délinéarisent le texte. Longtemps considéré comme une suite linguistique continue et linéaire, le texte numérique appert également comme un réseau discontinu, fait de relations entre les mots, de renvois et de liens. Par là, c'est également la lecture, longtemps seulement linéaire, qui change de nature pour devenir également tabulaire, réticulaire ou paradigmatique. À la possibilité de lire les textes et les corpus dans leur linéarité s'ajoute celle de convoquer des tables de fréquences, de définir des listes de mots ou de concordances et de définir les réseaux lexicaux significatifs comme autant d'entrées dans la textualité et dans la compréhension des corpus. Le numérique offre une vision différente des textes en exhibant leur dimension paradigmatique, à côté de leur séquentialité (Adam 2008). Les textes soumis à l'observation de l'historien-ne bénéficient d'un enrichissement ou d'une augmentation (Sadin 2013). Leur support numérique et le traitement logiciel qu'on leur applique permettent de voir, de lire,

² Nous n'insisterons pas dans cette contribution sur le tournant herméneutique des sciences humaines et sociales. Posons néanmoins avec Guilhaumou (2006) ou Mayaffre (2010) que l'archive textuelle ou le discours ne racontent pas seulement l'histoire: ils y participent. Le discours, les textes, les mots, au cœur des analyses historiennes, ne figurent pas seulement le passé mais ils contribuent à le configurer, à la cofabriquer. À ce titre, le matériau langagier mérite d'être analysé non pas simplement intuitivement, comme si le sens était immédiat ou transparent, mais bien avec les égards méthodologiques qu'il mérite (et que la révolution numérique semble toujours mieux permettre).

de comprendre davantage: ils créent de nouveaux observables linguistiques (les spécificités statistiques d'un texte par rapport à d'autres, les cooccurrences, les réseaux lexicaux privilégiés, etc.) jusque là difficilement appréhendables à la seule lecture humaine et favorisent nos capacités de navigation et d'interprétation dans les corpus textuels (Rastier 2011).

Précisément, la logométrie du côté de l'histoire se fixe comme principal objectif, dès ses débuts (Prost 1974) d'observer différemment les corpus de façon à nourrir, baliser et encadrer l'interprétation du sens historique des archives textuelles. Pour cela, elle met à disposition de l'historien-ne du discours de nombreux outils de statistique textuelle destinés à susciter des interrogations heuristiques puis à objectiver des parcours de lecture au sein de vastes corpus numériques.

Exemples de traitements logométriques: applications à un corpus de professions de foi électorales (1958–2007)

La logométrie, anciennement appelée lexicométrie, est un « un traitement (semi) automatique global du texte dans toutes ses dimension: graphiques, lemmatisées, grammaticalisées. L'analyse ainsi portera sur toutes les unités linguistiques de la lettre aux isotopies, en passant par les n-grams, les mots, les lemmes, les cooccurrences, les codes grammaticaux, les bi-codes ou les enchaînements syntaxiques » (Mayaffre 2010, 22). La force de la méthode tient dans la complémentarité – visible jusque dans les interfaces des logiciels³ – entre un traitement d'essence qualitative outillée par l'ordinateur (recherche documentaire, extraction et retour systématique au texte, navigation hypertextuelle) et un traitement quantitatif (repérage des saillances – c'est-à-dire des surutilisations lexicales, grammaticales ou cooccurentielles statistiquement significatives – mise à jour des structures (sémantiques) sous-jacentes des corpus, méthodes de classifications des textes). Nous nous proposons, à la suite de cette contribution, d'en présenter les principales fonctionnalités en insistant sur les parcours interprétatifs qu'ils suggèrent.

³ Dans cette contribution, seuls Hyperbase et Iramuteq seront utilisés. Mais d'autres logiciels tels Lexico 3 ou TXM disposent de méthodes d'analyse des données textuelles comparables.

Les outils de classification

La vocation première de l'historien-ne est selon les mots de Le Goff (1996, 14) de réfléchir à la dialectique de la continuité et du changement. Sur un corpus chronologique, à l'image du nôtre, l'un des premiers questionnements portera sur les proximités ou les distances linguistiques et donc historico-politiques entre les différents ensembles textuels dans le temps. Ainsi, sur le corpus de professions de foi des députées élues entre 1958 et 2007, divisé en douze textes (soit un texte par années d'élection), il est possible d'appliquer un traitement non supervisé de l'ensemble des unités textuelles du corpus afin de faire émerger une chronologie endogène à la féminisation de l'Assemblée nationale (rappelons que la définition de chronologies endogènes est d'autant plus impérieuse en histoire des femmes et du genre en politique que celles-ci ont souvent été masquées derrière les chronologies traditionnelles [i.e. masculines]. Cf. Achin [2005] par exemple). Le décompte de l'exhaustivité des mots du corpus et le calcul de la distance (linguistique) entre les textes⁴ permettent-ils de rendre compte des continuums, des ruptures ou des accélérations de la parole féminine lors des campagnes aux élections législatives sous la Cinquième République ?

Ainsi, le logiciel Hyperbase propose d'observer la distance entre les textes par une représentation arborée (**Figure 1**). Sur l'arbre, il est possible d'appréhender les grands rythmes de la légitimation électorale des femmes à leur entrée dans l'Hémicycle. Une première époque se dessine au bout d'une ramification pour regrouper les professions de foi rédigées entre 1958 et 1973. Le scrutin de 1978 marque ensuite une mutation dans le parler électoral féminin et débute une longue évolution progressive. Un troisième temps se dessine, à partir de 1993, pour attester du développement de nouvelles rhétoriques qui tranchent avec les routines discursives précédentes. Cette chronologie endogène, issue du corpus dans un mouvement inductif, est heuristique à plus d'un titre. C'est d'abord la faiblesse de l'impact de la crise de mai 1968 et de l'éveil de la seconde vague féministe à partir de 1970 sur les stratégies de présentation des candidates à la députation qui est révélé en

⁴ Les statisticiens du texte ont proposé pour cela des calculs de distance intertextuelle: Par exemple, Muller [1977], Corpus [2003], Mayaffre et Luong [2003].

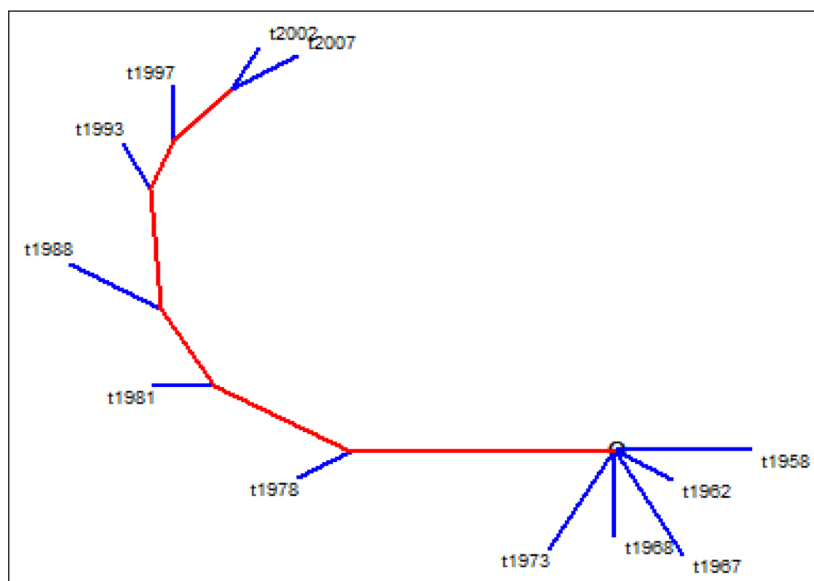


Figure 1: Analyse arborée – Corpus Professions de foi des Femmes (1958–2007).

creux; les textes de 1973 restent affiliés à ceux des premières législatures du régime quinto-républicain. C'est ensuite le constat traditionnellement affirmé d'un tournant mitterrandien (Sineau 2011) dans l'histoire politique qui est nuancé et précisé. La principale rupture dans la formulation des argumentaires électoraux des candidates à la députation s'opère moins en 1981, année de l'élection du premier président de la Cinquième République de gauche, qu'en 1978, date de l'élection de militantes du parti communiste français dans un contexte politisé. Enfin, l'émergence des discours de valorisation du féminin, souvent observés dans la période post-paritaire (2001–2002), trouvent d'après l'arborée leurs prémices dans les textes de 1993.

Caractériser les textes

Au-delà de ses capacités classificatoires, la logométrie permet, de façon plus incarnée, de caractériser les textes qui composent les corpus qu'elle traite. Cette caractérisation procède des différences constatées entre les discours. C'est là l'une des règles majeures des analyses logométriques: celles-ci s'effectuent dans le cadre d'une sémantique différentielle – posant que le sens n'est pas ontologique mais construit par la différence – et dans celui d'une statistique endogène. Par exemple,

les discours électoraux des femmes du début de la République ne prennent sens, ne se caractérisent ou ne s'individualisent que par rapport à la norme de la parole législative sous la Cinquième République, incarnée par le corpus qui rassemble l'ensemble des professions de foi réunies.

La fonction de caractérisation la plus éprouvée en logométrie réside dans le calcul des spécificités (Lebart et Salem 1994). Il s'agit par le décompte exhaustif de toutes les unités textuelles d'un corpus de repérer les formes remarquables (en termes quantitatifs) d'une partie par rapport à l'ensemble. Ainsi, sur les mêmes discours que précédemment – mais que l'on a divisé en deux périodes, le début et la fin du corpus, et en intégrant la comparaison avec les professions de foi des hommes élus aux mêmes dates – il est possible de connaître le vocabulaire spécifique des locutrices de chaque époque par rapport à leurs homologues masculins. Les principaux mots qui dominent la parole des femmes entre 1958 et 1973 d'une part et entre 1993 et 2007 d'autre part sont recensés dans **Figure 2**. Par ce tableau, ce sont les stratégies de légitimation et d'engagement des femmes sous la Cinquième République qui sont esquissées à grands traits, via le repérage de saillances lexicales (assorties d'un score de surutilisation autorisant l'interprétation politico-historique). Sans pouvoir entrer

1958 - 1973			
Spécificités des femmes	Score	Spécificités des hommes	Score
permettre	3,9	vous	3,7
lycée	3,5	ne	3,4
possibilité	3,3	campagne	3,4
milliard	3,3	tour	3,1
augmentation	3,2	république	3
économique (adj.)	2,8	liberté	2,8
construction	2,7	élection	2,6
1993 - 2007			
Spécificités des Femmes	Score	Spécificités des hommes	Score
je	12,3	nous	12,5
vous	5,9	changer	7,5
députée	5,8	Europe	5,1
moi	4,9	austérité	5,1
droit	4,4	diminution	4,5
engager	4,2	paix	4
écouter	4,1	agricole (adj.)	4

Figure 2: Les spécificités lexicales des députées élues entre 1958 et 1973 et entre 1993 et 2007 (Corpus femmes *versus* hommes, divisé en périodes).

dans les détails ou retourner au texte, on comprend toutefois que là où les candidates du début du régime se présentent essentiellement comme des promotrices de politiques publiques sur le terrain socio-économique qui délaissent le champ partisan et institutionnel, la parole des dernières élues du corpus se singularise davantage par la construction d'un ethos de femmes politiques susceptibles de répondre aux attentes citoyennes. Ce sont moins des idées ou des mesures programmatiques que les élues mettent en avant qu'une posture de représentation articulée sur la personne de la locutrice, son implication personnelle (« je ») et son lien avec l'électorat (« vous »), en partie lié à des stéréotypes de genre (« écouter »).

Pour l'historien-ne, une mention particulière doit être faite aux outils de description chronologique. En plus des outils de classification tenant compte des variables chronologiques, les logiciels de logométrie proposent également des graphiques de distribution de mots selon leur degré de surutilisation ou de sous-utilisation dans le temps. Par exemple, dans le corpus des professions de foi des femmes, nous pouvons observer l'évolution grammaticale du discours électoral au fil des trois grandes périodes définies plus haut (**Figure 3**). De façon éloquente, le graphique montre le changement de nature profonde du discours électoral au fil des cinquante années considérées. À un discours nominal (avec une surutilisation de noms, de déterminants et d'adjectifs) succède peu à peu un discours verbal (empli de

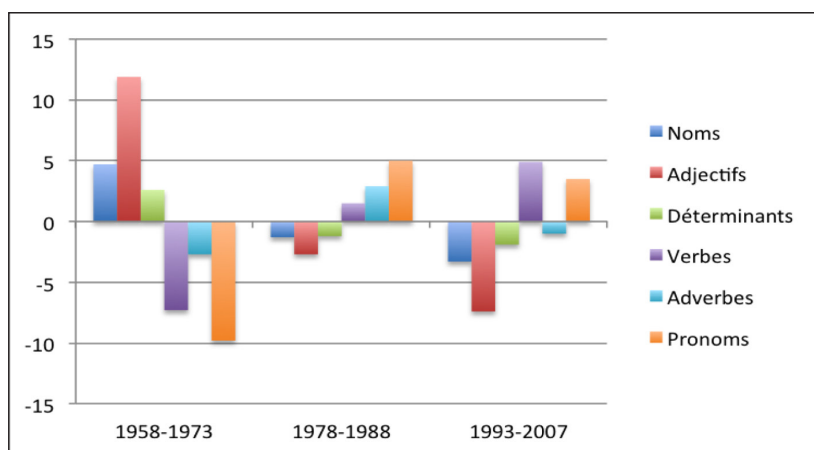


Figure 3: Évolution des codes grammaticaux dans le corpus de professions de foi des femmes divisé par périodes (1958–2007).

pronoms et verbes). En somme, le discours de la description de l'action publique ou de la proposition politique disparaît peu à peu au profit d'un discours articulé autour de la promesse de représentation des candidates, dont la production d'une image valorisée est le but principal.

La contextualisation

Souvent informatifs, les traitements des occurrences seules (comme ceux menés ci-dessus) réclament néanmoins à être prolongés pour mener à l'interprétation. En effet, répétons avec Rastier (2001) que le sens ne naît qu'en contexte. Aussi la logométrie autorise-t-elle la lecture traditionnelle des textes, qu'elle contribue à rendre systématique et exhaustive.

L'une des fonctions les plus répandues est celle du concordancier (Pincemin, Fabrice, Chanove et Mathieu-Colas 2006). Il permet de lister et de trier tous les passages contenant un mot choisi pour pôle pour en produire une interprétation contrôlée, dans le cadre de ses contextes d'emplois attestés en corpus. Par exemple ici, on donne un extrait du concordancier du lemme « femme », qui est l'une des grandes spécificités lexicales du corpus féminin en synchronie et qui évidemment recouvre une certaine polysémie et évolution sémantique dans le temps (**Figure 4**). Plus loin que le simple, mais informatif, retour au texte, la logométrie développe ces dernières années des méthodes relevant d'une véritable statistique contextualisante (Brunet 2012; Mayaffre 2014). Particulièrement, l'analyse des cooccurrences apparaît

01	8a	Les années ont passé . Les femmes sont devenues des citoyennes .
01	21a	désigner les hommes ou les femmes qui seront chargés de mettre en
01	28a	résumer a été rédigé par une femme qui a exercé depuis 19466 les ma
01	29a	ssemblée a besoin d' hommes et de femmes au patriotisme éclairé , aux
01	30a	Général de Gaulle eut donné aux femmes le droit de vote et d' éligibil
01	33a	dans l' avenir . Pour vous , femmes rurales , femmes artisanes , mè
01	33a	Pour vous , femmes rurales , femmes artisanes , mères de famille ,
01	34a	y ait au parlement que des femmes marxistes pour vous représenter
01	34a	, a besoin d' hommes et de femmes capables de courage , de sagess
02	42a	. Soutien et promotion de la femme sur le plan social , professionn
02	46a	personnel . Les ouvriers , les femme ; les vieux , les artisans et pe
02	54a	Femmes de cette circonscription , vous
02	54a	d' envoyer au parlement une femme comme vous , qui connaît vos pro
02	54a	et qui saura comme d autres femmes parlementaires remplir la tâche
02	54a	léguee nos ancêtres , hommes et femmes de bonne volonté nous voulons l
02	58a	mon mandat , j' ai pensé aux femmes , rurales et artisanes , aux fa
02	64a	angoissant , et à celle des femmes dont je connais , mieux que d_a
03	75a	Député _femme , la seule du Pas_de_Calais , d

Figure 4: Concordances autour du mot « femme » dans le corpus féminin (1958–2007).

comme l'une des voies de la formalisation de co(n)textes porteurs du sens des textes. En effet, le repérage de la coprésence matérielle et statistiquement significative de deux unités textuelles (typiquement deux mots) au sein des paragraphes constitue déjà un pas vers la détection de noyaux sémantiques, là où le mot seul demeurerait presque toujours ambigu. Il est évident que le mot « classe » seul ne saurait faire l'objet d'interprétation définitive là où la relation matérielle au sein des mêmes paragraphes du même terme avec les mots « sociale », « lutte » ou « grève » nous informe davantage du thème en question.

De façon appliquée à notre corpus, il est possible de mesurer la plus-value de l'approche co-occurrence sur un mot très répandu dans le discours politique contemporain et repéré comme spécifique de la parole des dernières élues du corpus: le pronom personnel « je ». Face aux milliers d'occurrences du vocable dans le corpus, la définition des contextes d'emploi du mot trouve dans la statistique contextualisante un soutien précieux. Les principaux associés du pronom sont ainsi listés dans **Figure 5**.

À titre de comparaison, observons les mots articulés au pronom « nous » préféré par les hommes pour énoncer leur profession de foi entre 1993 et 2007 (**Figure 6**). À quelques exceptions près, le tableau indique que la pronominalisation différenciée des candidatures des hommes et des femmes implique la production de deux types de discours nettement distincts. Le discours électoral articulé au pronom « nous »

Co-occurents de « je » Corpus Femmes							
Spécificités	Score	Spécificités	Score	Spécificités	Score	Spécificités	Score
me	28,91	conviction	6,29	terrain	4,50	coeur	3,39
votre	20,04	présenter	6,23	préoccupation	4,46	entièrement	3,28
vous	19,15	attacher	6,09	attente	4,30	activement	3,24
mon	16,65	travailler	6,06	fier	4,21	attentif	3,01
engager	14,35	circonscription	6,04	disponible	4,17	humaniste	2,95
compter	8,45	femme	5,89	régulièrement	4,07	écouter	2,93
côté	7,96	entendre	5,87	remercier	4,05	porte-parole	2,91
battre	7,80	connaître	5,64	expérience	4,04	quotidien	2,82
solliciter	7,32	naître	5,50	rencontre	3,97	inquiétude	2,68
souhaiter	7,29	écouter	5,49	actif	3,64	connaissance	2,67
défendre	7,20	rencontrer	5,27	profondément	3,56	proche	2,60
savoir	6,82	confiance	5,10	permanence	3,52	engagement	2,55
toujours	6,61	convaincre	5,07	détermination	3,49	dialoguer	2,46
croire	6,57	efforcer	4,68	quotidiennement	3,44	pleinement	2,34
continuer	6,51	aimer	4,60	présent	3,43	attention	2,29
vouloir	6,42	dossier	4,57	représenter	3,41	disponibilité	2,08

Figure 5: Les co-occurents de « je » dans le corpus Femmes (1993–2007).

Co-occurents de « nous » Corpus Hommes			
Spécificités	Score	Spécificités	Score
ensemble	8,99	suppléant	3,07
vouloir	6,91	réussir	3,01
proposer	6,60	affronter	2,99
devoir	6,46	réduire	2,99
mobiliser	4,79	espoir	2,94
unir	4,41	faire	2,88
notre	4,25	défaire	2,86
Europe	4,07	capable	2,81
société	3,51	résigner	2,81
confiance	3,51	atteindre	2,79
pouvoir	3,31	montrer	2,76
changer	3,31	former	2,76
relever	3,31	connaître	2,68
partager	3,26	avenir	2,65
contraire	3,20	préparer	2,61
ferme	3,09	ultra-libéral	2,58

Figure 6: Les co-occurents du pronom « nous » dans le Corpus Hommes (1993–2007).

prend une dimension résolument politique et collective visant à créer une adhésion majoritaire autour d'un projet ou d'une vision. A contrario, le « je », statistiquement préféré dans le discours des candidates, apparaît davantage comme un mot apolitique axant l'essentiel de la légitimation discursive autour de la mise en scène d'une identité et de qualités personnelles et d'un lien avec les électeurs/trices.

Les logiciels de logométrie permettent également de puissants traitements globaux et généralisés de la cooccurrence dans le but de faire émerger les structures sémantiques, si ce n'est thématiques, sous-jacentes des corpus (Viprey 2006). Les progrès techniques permettent en effet d'établir les profils associatifs de plusieurs centaines de mots de manière à rendre compte des relations entre les mots au sein d'un ensemble textuel. Par là, c'est la dimension réticulaire des textes – anciennement pressentie mais difficilement appréhendable – qui est rendue tangible par le

traitement logométrique des corpus numériques. Plus que jamais, le texte se laisse saisir dans sa dimension réticulaire, dans son tissage, dans son organisation en réseau.

Ainsi, dans le corpus de professions de foi global, à partir de l'établissement et de la comparaison automatique des profils co-occurentiel des 300 substantifs les plus fréquents, nous avons pu définir l'architecture forte du discours de candidature sous la Cinquième République (Guaresi 2015). Nous reproduisons la projection factorielle (Figure 7) qu'en propose le logiciel Iramuteq (Ratinaud et Marchand 2012). La projection factorielle montre une opposition sur l'axe horizontal entre les thèmes politiques et institutionnels (« majorité », « république », « député », « élection ») et les questions économique et sociales (« impôts », « emploi », « éducation ») rejetées à droite

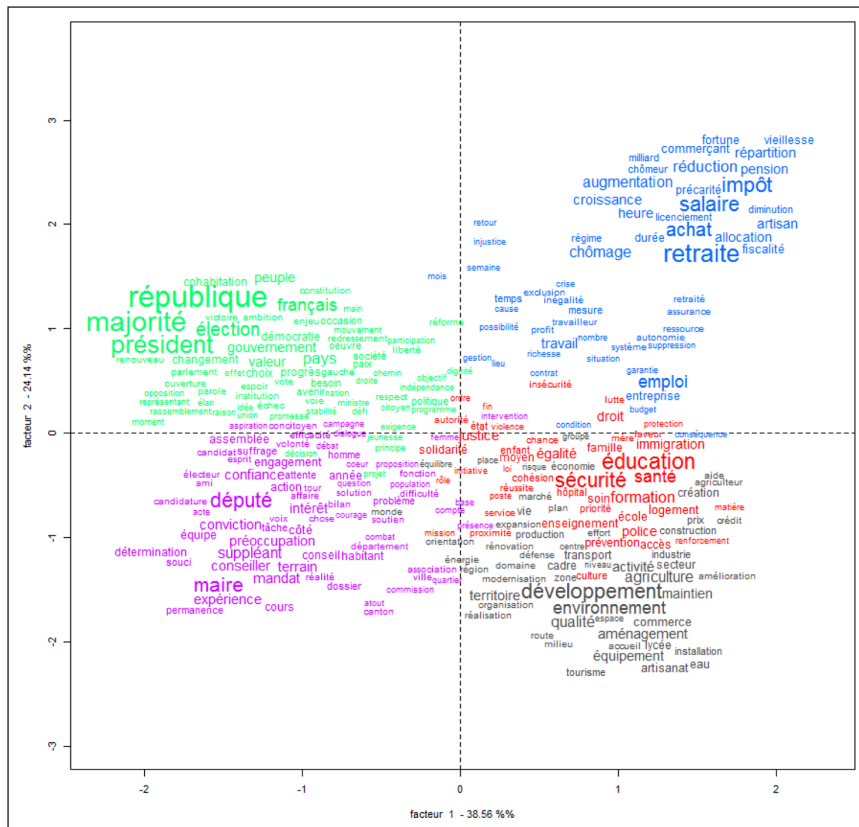


Figure 7: Analyse factorielle des cooccurrences (Corpus Professions de foi 1958–2007).

du graphique. Ces deux grands ensembles semblent ventilés dans les professions de foi dans une dimension nationale (« peuple », « français », « pays »), en haut de la figure, et un ancrage local (« maire », « canton », « territoire »), en bas. Par là, ce sont les représentations dominantes des rôles de parlementaires français et de leurs principaux engagements sous l'ensemble de la Cinquième République qui sont formalisés.

Conclusion

Si seuls quelques exemples de traitements logométriques ont été exposés, le potentiel heuristique d'une *hyperlecture* assistée par l'ordinateur apparaît déjà important. Qu'il nous soit permis, pour conclure, d'insister sur deux enrichissements de la lecture et de l'analyse historique de corpus d'archives textuelles par le numérique. D'abord, en permettant une meilleure formalisation de l'intertexte au sein de macro-corpus, la révolution numérique offre à l'historien-ne la possibilité de saisir des phénomènes discursifs locaux (i.e. la parole d'une locutrice en particulier, ou d'une élection) dans leur comparaison avec un horizon discursif global. La mise à disposition des sources discursives en abondance d'une part et leur traitement informatique possible facilitent toujours plus l'appréhension de l'intertexte, nécessaire à l'interprétation historique. Surtout, cette appréhension reste – par le traitement systématique et exhaustif de toutes les composantes linguistiques des corpus – très attachée à la matérialité textuelle, conformément aux attentes de l'herméneutique posant le discours comme acteur de l'histoire. Enfin, c'est le texte dans ses aspects paradigmatiques et réticulaires que le numérique contribue à dévoiler à l'œil des historien-nes pour enrichir de façon exogène leurs protocoles de lecture.

Déclaration de conflit d'intérêts

Les auteurs déclarent l'absence d'un conflit d'intérêts.

Références

- Achin, Catherine.** 2005. *Le Mystère de la Chambre basse. Comparaison des processus d'entrée des femmes au Parlement. France-Allemagne 1945–2000*. Paris: Dalloz.
- Adam, Jean-Michel.** 2008. *La linguistique textuelle. Introduction à l'analyse textuelle des discours*. Paris: Armand Colin.


- Brunet, Etienne.** 2012. "Nouveau traitement des cooccurrences dans Hyperbase." *Corpus* 11: 219–48.
- Corpus.** 2003. "La distance intertextuelle." Consulté le 8 juin 2018. <https://corpus.revues.org/52>.
- Darnton, Robert.** 2011. *Apologie du livre*. Paris: Gallimard.
- Genet, Jean-Philippe.** 1986. "Histoire, Informatique, Mesure." *Histoire & Mesure* 1(1): 7–18. DOI: <https://doi.org/10.3406/hism.1986.904>
- Goody, Jack.** 2007. *Pouvoirs et savoirs de l'écrit*. Paris: La Dispute.
- Guaresi, Magali.** 2015. "Les thèmes dans le discours électoral de candidature à la députation sous la Cinquième République. Perspective de genre (1958–2007)." *Mots. Les langages du politique* 108: 15–37. DOI: <https://doi.org/10.4000/mots.21977>
- Guaresi, Magali.** 2018. *Parler au féminin. Les professions de foi des députés sous la Cinquième République (1958–2007)*. Paris: L'Harmattan.
- Guilhaumou, Jacques.** 2006. *Discours et événement. L'histoire langagière des concepts*. Besançon: Presses universitaires de Franche-Comté.
- Le Goff, Jacques.** 1996. *Saint Louis*. Paris: Gallimard. DOI: <https://doi.org/10.14375/NP.9782070733699>
- Lebart, Ludovic et André Salem.** 1994. *Statistique textuelle*. Paris: Dunod.
- Mayaffre, Damon.** 2010. *Vers une herméneutique matérielle numérique. Corpus textuels, Logométrie et Langage politique*. Nice: Université Nice Sophia Antipolis. Consulté le 8 juin 2018. <https://tel.archives-ouvertes.fr/tel-00655380>.
- Mayaffre, Damon.** 2014. "Plaidoyer en faveur de l'Analyse de Données co(n)textuelles. Parcours cooccurentiels dans le discours présidentiel français (1958–2014)." *JADT 2014*. Consulté le 8 juin 2018. <http://lexicometrica.univ-paris3.fr/jadt/jadt2014/>.
- Mayaffre, Damon et Xuan Luong.** 2003. "Arbres et généalogie politique. Représentation arborée du discours de Jacques Chirac (1995-2005)." *Histoire & Mesure* 18: 289–311. DOI: <https://doi.org/10.4000/histoiremesure.831>
- Muller, Pierre.** 1977. *Principes et méthodes de statistique lexicale*. Paris: Hachette.
- Pincemin, Bénédicte, Fabrice Issac, Marc Chanove et Michel Mathieu-Colas.** 2006. "Concordanciers: thème et variations." *JADT 2006*. Consulté le 8 juin 2018. <http://lexicometrica.univ-paris3.fr/jadt/jadt2006/tocJADT2006.htm>.

- Prost, Antoine.** 1974. *Analyse du vocabulaire des proclamations électorales de 1881, 1885 et 1889*. Paris: Presses universitaires de France.
- Rastier, François.** 2001. *Arts et sciences du texte*. Paris: Presses universitaires de France. DOI: <https://doi.org/10.3917/puf.rast.2001.01>
- Rastier, François.** 2011. *La mesure et le grain*. Paris: Presses universitaires de France.
- Ratinaud, Pierre et Pascal Marchand.** 2012. "Application de la méthode ALCESTE aux 'gros' corpus et stabilité des 'mondes lexicaux': analyse du 'CableGate' avec IRAMUTEQ." *JADT 2012*. Consulté le 8 juin 2018. <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/tocJADT2012.htm>.
- Rygiel, Philippe.** 2011. "L'enquête historique à l'ère numérique." *Revue d'histoire moderne et contemporaine* 58(4): 30–40. DOI: <https://doi.org/10.3917/rhmc.585.0030>
- Sadin, Éric.** 2013. *L'humanité augmentée. L'administration numérique du monde*. Montreuil: L'Echappée.
- Sineau, Mariette.** 2011. *Femmes et pouvoir sous la Cinquième République. De l'exclusion à l'entrée dans la course présidentielle*. Paris: Presses de science politique.
- Vandendorpe, Christian.** 1999. *Du papyrus à l'hypertexte. Essai sur les mutations du texte et de la lecture*. Paris: La Découverte.
- Viprey, Jean-Marie.** 2006. "Structure non-séquentielle des textes." *Langages* 163: 71–85. DOI: <https://doi.org/10.3917/lang.163.0071>

How to cite this article: Guaresi, Magali. 2019. "La logométrie en histoire: une herméneutique numérique. Exploration d'un corpus de professions de foi électorales de député-e-s (1958–2007)." *Digital Studies/Le champ numérique* 9(1): 14, pp. 1–15. DOI: <https://doi.org/10.16995/dscn.349>

Submitted: 04 April 2016 **Accepted:** 03 June 2019 **Published:** 12 August 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Digital Studies/Le champ numérique* is a peer-reviewed open access journal published by Open Library of Humanities. **OPEN ACCESS** 