



HAL
open science

Explorations textométriques d'un corpus foucauldien. Le Désordre des familles : au plus près de la naissance du genre du rapport

Hugo Dumoulin

► **To cite this version:**

Hugo Dumoulin. Explorations textométriques d'un corpus foucauldien. Le Désordre des familles : au plus près de la naissance du genre du rapport. 11èmes Journées de Linguistique de Corpus, Jul 2023, Grenoble, France. hal-04627781

HAL Id: hal-04627781

<https://hal.science/hal-04627781v1>

Submitted on 27 Jun 2024

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.

Explorations textométriques d'un corpus foucauldien. Le Désordre *des familles* : au plus près de la naissance du genre du rapport

Hugo Dumoulin¹

¹ Laboratoires Modyco & Sophiapol, Université Paris Nanterre
hugo.dumoulin@parisnanterre.fr

Introduction

En 1968, lorsque le cercle des étudiants en épistémologie de l'ENS l'interroge sur sa méthode, Foucault répond : « À tant d'incertitudes, je voudrais substituer l'*analyse du discours* lui-même dans ses conditions de formation, dans la série de ses modifications et dans le jeu de ses dépendances et de ses corrélations » (1994, I : 682, nous soulignons). Si une telle déclaration ne manque pas de situer le philosophe parmi les penseurs fondateurs de l'analyse du discours, elle n'en est pas moins énigmatique : comment repérer séries, dépendances, corrélations, conditions de formation du discours ? Y'a-t-il une théorie qui formaliserait les structures de la discursivité ? Déclarant finalement que « le temps n'est pas encore venu de la théorie » (1969) Foucault ne nous a pas laissé de clef disciplinaire définitive pour contrôler ses procédures de découverte. Il nous laisse cependant des corpus, constitués par « la mise en travail de problématiques philosophiques au sein d'archives historiques » (Revel 2010).

Cette étude veut montrer à partir d'un exemple qu'une lecture de linguistique de corpus de ces matériaux est possible, et que les explorations textométriques permettent de répondre à certaines questions laissées ouvertes par Foucault. Il se trouve que Foucault a publié quelques-uns de ses corpus, nous rendant la tâche plus facile : c'est le cas du *Désordre des familles* (Foucault & Farge 1982), abordé sous l'angle d'une lecture outillée dans notre thèse, de laquelle cette étude est extraite.

Corpus et méthodologie

Corpus

Comme l'indiquent les éditeurs des *Dits et Écrits* (1994, III : 237), Foucault a longtemps poursuivi le projet d'une grande publication des archives de l'enfermement à l'hôpital général et à la Bastille, sur lesquelles il n'a cessé de travailler depuis l'*Histoire de la folie* (1961). De ce grand projet cependant, il ne reste que *Le Désordre des familles*, publié en collaboration avec Arlette Farge dans la collection corpus de Gallimard (1982), et présentant 287 textes tirés des dossiers de police versés aux archives de la Bastille pour les années 1728 et 1758 conservées à la bibliothèque de l'Arsenal à l'époque de leur consultation par Foucault et Farge.

Ces dossiers sont liés à l'institution des lettres de cachet, ordres émis par le lieutenant général de police par délégation du roi à dessein d'enfermer un individu à la Bastille sans jugement. En rupture avec l'historiographie révolutionnaire, Foucault et Farge montrent que ces lettres sont émises massivement suite à la demande des familles elles-mêmes (1982 : 9) pour régler tout un ensemble d'affaires d'ordre privé qui vont du conflit de voisinage à des questions plus graves qui relèveraient aujourd'hui du pénal. Dans les dossiers l'on trouve ainsi des lettres écrites (via l'écrivain public) par les familles requérant l'enfermement d'un tiers (les « placets »), des lettres écrites par un ensemble de personnages se présentant comme les garants des précédentes, enfin des textes écrits par les policiers faisant la synthèse de leurs démarches d'enquête suite aux demandes et rendant un avis sur l'ordre d'enfermement.

Le livre donne accès au corpus d'archives à travers plusieurs couches d'un traitement accompli par Foucault et Farge. Les deux historiens transcrivent et dactylographient les manuscrits d'archives, tout en indiquant leurs difficultés de lecture de certains mots. Ils normalisent aussi une mise en page d'origine où du texte se trouve parfois écrit dans la marge ou en surimpression, ce dont ils gardent systématiquement la trace en le présentant de manière homogène en crochets dans le fil du texte ou en annexe à la fin de celui-ci. Surtout, Foucault et Farge ajoutent à la classification archivistique certains éléments de classement qui sont le reflet de leur propre lecture et qui se concrétisent dans le découpage du livre en parties thématiques. Du reste, cette lecture historique se fait discrète dans le livre : la part du commentaire est grandement réduite par rapport à l'espace alloué à la retranscription de pièces⁴⁰. Pourtant, elle n'est pas absente, et c'est en gardant trace de ces multiples couches de lecture par l'annotation que l'on peut espérer capter quelque chose des procédures de découverte des deux historiens à travers par la relecture textométrique.

Méthodologie

Publié tardivement par rapport à la période du travail entre Foucault et Farge, *Le Désordre des familles* est marqué par la problématisation foucauldienne du milieu des années 1970. À ce moment, Foucault réalise une histoire de ce que l'on pourrait appeler les « technologies de pouvoir » (Foucault 1975), qui met au jour l'émergence à partir du XVIII^e siècle du régime des « disciplines » (Foucault 1973), clef de voute des deux infinitifs du titre foucauldien : « surveiller » et « punir » (Foucault 1975), lesquels articulent « savoir » et « pouvoir » (Foucault 1973). Foucault et Farge avancent l'hypothèse que l'institution des lettres de cachet est une des premières mises en œuvre de la technologie de pouvoir « disciplinaire » (Foucault & Farge 1982 : 346). Les lettres de cachet seraient donc l'un des premiers instruments de « pouvoir-savoir ». Notre hypothèse est qu'une enquête de linguistique de corpus sous la forme d'une analyse des données textuelles permet d'éprouver l'hypothèse de Foucault et Farge, en cernant le fonctionnement *discursif* de ce dispositif de normalisation disciplinaire.

On avance que l'insertion de ces textes dans un certain genre joue un rôle décisif. Foucault lui-même indique que les « disciplines » s'attachent à certains types d'écrits administratifs qui émergent à la fin du XVIII^e siècle, Foucault parlant du début de « l'ère du *rapport* comme forme des relations entre savoir et pouvoir » (1973 : 238 nous soulignons). Si Foucault ne cite pas Bakhtine, il est manifeste qu'il s'intéresse à ce que le dernier définit comme des *genres* de discours situés au carrefour du langagier et du social et référés à certaines sphères d'activité (Bakhtine 1984 : 269). Et en effet, bien qu'empruntant des formes diverses, les textes du corpus gardent pour fonction de rapporter et de rendre compte, ce qui autorise de les aborder à partir de catégorie du macro-genre du « rapport », aux contours bien définis : double modalité de la « description » des faits et de leur « évaluation » axiologique (en vue de la prescription d'une action), caractère « adressé » de textes marqués par le différentiel hiérarchique entre un scripteur et le commanditaire auquel il s'adresse (Née, Oger, Sitri, 2017).

Notre méthode d'acquisition repose sur la conversion du corpus textuel d'archives transcrites et imprimées dans le livre du *Désordre des familles* en base de données textuelles (.xml) exploitable par des logiciels de textométrie. Partant d'une version .txt du livre, l'on réalise des choix de balisage visant à conserver certaines caractéristiques des textes en métadonnées ainsi que certains aspects de la composition textuelle à travers des balises <div> (Figure 1). Parmi les métadonnées de chaque texte l'on conserve des informations afférentes aux différentes couches d'exploitation du corpus : certaines représentent la classification adoptée à la bibliothèque de l'Arsenal, d'autres sont issues du traitement par Foucault et Farge du corpus. Parmi les premières, on trouve la date d'écriture de la pièce (« 1728 »), l'identifiant de celle-ci (« 11009f95 »), ainsi que le dossier dans lequel elle se trouve (représenté par nom d'affaire : « Barbe Blondel Duponchel »). Parmi les secondes, l'on trouvera le rang de la pièce dans l'ordre de présentation de chaque dossier dans le livre, ainsi que le regroupement des dossiers en thèmes selon les deux niveaux de classement du livre : une première division en deux grandes catégories « La discorde des ménages (DM) » et « Parents et Enfants (PE) », une deuxième division définissant dans

⁴⁰ Sur les 365 pages du livre, seules 55 sont consacrées à la présentation et aux analyses du corpus.

chaque catégorie des sous-thèmes comme « le déshonneur de l'errance » ; cette annotation donne en sortie une valeur d'attribut complexe du type « PE_le déshonneur de l'errance » (Figure 1).

Pour représenter certains aspects de la composition textuelle des pièces du corpus à travers le balisage, l'on retiendra des balises classiques représentant en-tête (<head>), division en paragraphes (<p>), et signatures (<signed>) de chaque pièce, ainsi qu'une balise <div> servant à représenter le fait que, sur une même pièce, plusieurs scripteurs peuvent parfois cohabiter : comme l'indiquent Foucault et Farge, la lettre du garant ou le rapport de police peuvent être écrits parfois à même le placet d'origine, dans la marge ou au verso de celui-ci. Ces segments de texte, bien que faisant partie de la même pièce d'archive, ont des statuts différents : ainsi, à travers les différentes valeurs d'un attribut « type » de la balise <div> l'on gardera trace non pas seulement des différents scripteurs, mais aussi des différentes démarches dans lesquelles ils insèrent explicitement leur écrit. Ce balisage est facilité par l'abondance des énoncés performatifs dans ces textes – qui représentent l'acte que le sujet dit accomplir par son dire (Authier-Revuz 2020 : 27) – au sein de formules dont on peut mesurer par ailleurs le caractère stéréotypé à l'aide d'un calcul des segments répétés⁴¹. L'on encode dix types de démarches par dix valeurs de l'attribut type⁴².

Après acquisition comme base de données textuelles, le corpus DÉSORDRE regroupe 77 049 occurrences pour 287 textes, soit une moyenne de 266 mots par texte.

Résultats

L'on aborde dans cette étude deux grandes analyses contrastives réalisées sur la base du balisage précédent.

L'ancrage lexical des contrastes dans la représentation de la norme

A l'aide du logiciel TXM (Heiden et al. 2010) on lemmatise le corpus et on réalise un balisage syntaxique de celui-ci selon l'étiquetage TreeTagger. L'on peut alors réaliser une analyse factorielle des correspondances (Benzécri 1973) entre les formes graphiques présentes dans le corpus et une partition de celui-ci selon les thèmes retenus par Foucault et Farge (Figure 2). Le résultat obtenu est particulièrement éclairant : le premier axe de l'AFC se laisse interpréter comme une opposition est-ouest entre parentalité d'une part et conjugalité, selon laquelle s'accordent exactement les parties du corpus inscrites dans les macro-thèmes de Foucault et Farge « Parents-enfants » (PE sur le schéma) – à l'est –, et « Discorde des ménages » (DM) – à l'ouest. En effet, les formes les plus contributives à l'axe 1 se trouvent partagées entre l'est – « fils » (contribution de 7,5%), « père » (5%) –, et l'ouest – « femme » (11,5%) et « mari » (6%). À l'ouest se trouvent les formes employées pour désigner les époux à l'intérieur du ménage, à l'est pour désigner les membres de la relation parentale.

Le second axe se laisse interpréter comme une opposition nord-sud entre plaintes visant des hommes et plaintes visant des femmes : les formes les plus contributives à l'axe 2 se trouvent partagées entre formes désignant des hommes au nord – « il » (9,5%), « fils » (3,5%), « ledit » (1%) –, et des femmes au sud – « fille » (6%), « elle » (4%), « ladite » (3,5%). Cet axe permet de contraster selon une opposition homme/femme le découpage thématique proposé par Foucault et Farge, ce que confirme d'ailleurs la

⁴¹ Ainsi par exemple un segment relativement long comme « X représente très humblement à votre Grandeur le / que » affiche une fréquence de 69 occurrences dans le corpus selon un calcul réalisé par Le Trameur (Fleury & Zimina 2014).

⁴² La liste des valeurs est : “enfermement” = scripteur demandant l'enfermement d'un tiers ; “libération” = scripteur demandant la libération d'un tiers ; “libération_soi” = scripteur demandant sa propre libération ; “libération_repentir” = scripteur demandant la libération d'un tiers dont il a précédemment demandé l'enfermement ; “enfermement_garant” = scripteur se présentant comme un garant pour enfermer un tiers ; “libération_garant” = scripteur se présentant comme un garant pour libérer un tiers ; “police_rapport” = scripteur policier rédigeant un rapport d'enquête ; “police_DR_enfermement” = scripteur policier rapportant le texte d'un placet ; “police_DR_libération” = scripteur policier rapportant le texte d'une demande de libération ; “correspondance” = lettre de correspondance versée au dossier.

position du thème de « la débauche des *maris* » au nord et de celui de « l'enfermement des *épouses* » au sud.

Une fois ce cadre interprétatif posé, l'on peut se pencher sur la position de différents lexèmes contribuant à la représentation de l'anormalité elle-même, et observer ainsi leur contribution aux contrastes du corpus. Ainsi, l'on observe sur l'axe 1 que les formes « argent » et « livres » sont significativement attirées à l'est, c'est-à-dire vers le rapport parent/enfant, tandis que les formes « conduite » et « vie » sont significativement attirés à l'ouest, du côté du rapport entre époux. Les parents reprochent à leurs enfants leur prodigalité matérielle, tandis que les époux se reprochent mutuellement leur mauvaise conduite morale. Le long de l'axe 2, l'on est attentif à la présence des formes « jours » et « enfants » au nord – du côté des plaintes des femmes visant les hommes –, par opposition à la présence de « débauche » et « libertinage » au sud – du côté des plaintes des hommes visant des femmes. Des retours au texte permettent d'éclairer cette répartition : alors que les femmes reprochent aux hommes leur dissipation (« il la consomme tous les *jours* en dépenses ») et leur absence dans la gestion des affaires du foyer (concernant les *enfants* en particulier), les hommes se plaignent de l'inconduite supposée de leurs concubines. La norme n'est donc pas la même pour tout le monde : figure bariolée, le scandale de l'anormal change de forme en fonction des situations où il est invoqué.

L'analyse factorielle confirme l'ancrage *lexical* du découpage par thèmes de la « mauvaise conduite » selon Foucault et Farge. Cette détermination lexicale de l'expression de l'anormalité met en avant deux choses : d'une part il y a un ancrage discursif du dispositif de normalisation disciplinaire : ce sont certains lexèmes (« argent », « conduite », « enfants », « libertinage ») qui contribuent à fixer la représentation de l'anormal dans les différentes situations où il surgit ; d'autre part et conséquemment, l'on peut arguer que la procédure de classification par Foucault et Farge des thématiques du corpus – et donc des aspects de la norme – repose plus ou moins inconsciemment sur le repérage de ces précédents lexèmes dans les textes, dans la mesure où notre analyse statistique met en valeur leur caractère significatif pour la classification de Foucault et Farge.

L'ancrage énonciatif des contrastes entre positions socio-discursives

L'on réalise alors une deuxième analyse factorielle des correspondances entre les lemmes du corpus et une partition selon les types de scripteurs repérés par leurs démarches (Figure 3). La représentation obtenue est marquée la domination écrasante de l'axe principal (45% de l'inertie totale). Cet axe se laisse interpréter selon une opposition entre à l'est des pronoms personnels de 1^{ère} et 2^{ème} personne ainsi que certains termes d'adresse (« monsieur »), et à l'ouest des termes d'adresse comme « Monseigneur », ou des déterminants comme « Votre ». L'opposition radicale entre est et ouest est celle de la distance entre l'énonciateur et son interlocuteur : à l'est l'on a un rapport de relative égalité, à l'ouest il y a une distance hiérarchique écrasante. Ainsi, la répartition des types de termes d'adresse entre les scripteurs s'avère structurante des contrastes entre démarches au sein du corpus, ce qui confirme bien la prégnance sur ces textes de leur identité générique de rapports, c'est-à-dire de textes *adressés*.

L'on observe alors un phénomène spectaculaire : les textes des garants, qu'ils soient destinés à soutenir une demande d'enfermement ou de libération, se trouvent déportés à l'est ; à l'inverse, les textes des familles populaires, qu'ils visent l'enfermement ou la libération, sont déportés à l'ouest. Cela est décisif : les deux démarches (opposées d'un point de vue pragmatique) de la demande d'enfermement et de libération apparaissent tout de même moins différentes *énonciativement* que ce qu'on peut désigner comme les « positions socio-discursives » des scripteurs (petit peuple des « misérables », garants issus du clergé, police)⁴³.

Ces positions, hiérarchiquement articulées entre elles, soulignent le différentiel social qui fonctionne à l'arrière-plan du dispositif de normalisation disciplinaire. Cette répartition en trois « positions discursives » est confirmée par une classification ascendante hiérarchique des types de démarches selon

⁴³ L'on peut capturer les origines sociales des types de scripteurs en s'intéressant aux signatures qui figurent au bas des pièces, lesquelles comportent presque systématiquement la profession des signataires.

les lemmes employés (Figure 4). La généralité de ces textes participe donc à imprimer discursivement un ensemble de positions sociales différenciées.

Conclusion

Ainsi, l'on conclut de ces explorations que les discours et leurs caractéristiques génériques participent à l'impression sur les sujets de normes de conduite et d'une hiérarchie sociale, précisant par-là le fonctionnement du « pouvoir-savoir » décrit par Foucault. Cette étude a une vocation méthodologique : en gardant trace de la lecture opérée par Foucault et Farge, l'on peut indirectement éclairer leurs procédures de découverte. Ainsi l'on montre que la classification thématique des dossiers du corpus – et, partant, des différents aspects que peut prendre la norme – s'attache à la présence statistiquement significative de certains lexèmes dans les textes. De même, le repérage d'un dispositif de hiérarchisation sociale dans l'institution des lettres de cachet est attaché à la répartition statistiquement significative de différentes formes d'adresse, qui contribue à contraster le corpus selon les démarches des scripteurs, qui forment autant de sous-genres (placet, lettre de garant, rapport de police).

Cette étude s'ouvre sur les autres analyses menées dans la thèse, qui nous conduisent, au-delà des réflexions foucauldienne, à mettre en évidence l'influence de l'idéologie du petit artisanat parisien du XVIII^e siècle (Soboul 1981) sur l'état du genre du rapport à la même période. En menant sur Iramuteq (Marchand & Ratinaud 2015) une classification descendante hiérarchique « méthode Reinert » (1983) (Figure 5) on voit apparaître d'autres thématiques structurantes du corpus, comme par exemple la « matérialité économique » (classe 2). En menant des analyses déductives (calcul des spécificités, analyses des cooccurrences), on peut observer le comportement de formes de « préconstruit » caractéristique de l'emprise d'une idéologie sur le discours (Pêcheux 1975 : 152).

Figures

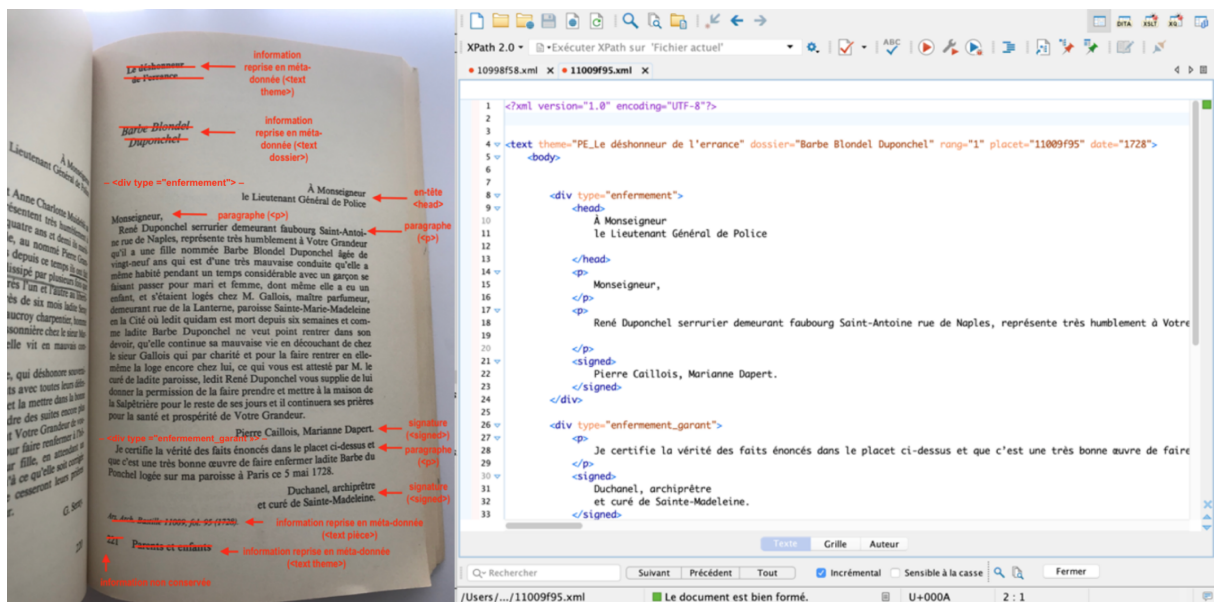


figure . 1 : Le balisage du corpus (à gauche une page de l'édition Gallimard, à droite la vue d'un fichier .xml dans OxygenXML)

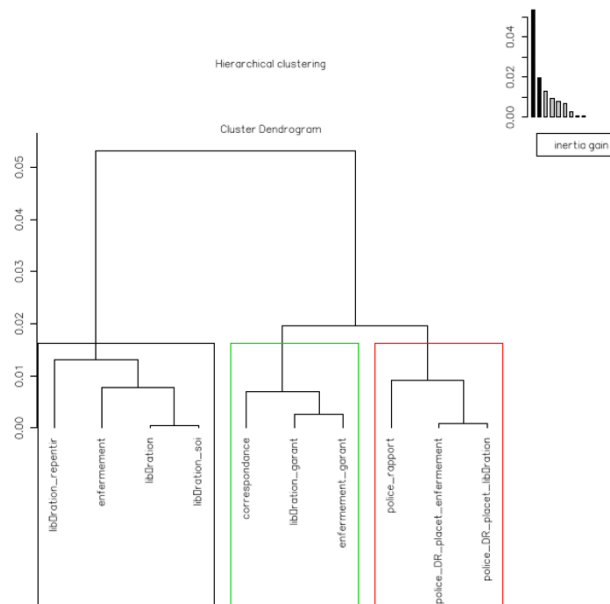


figure .4 : Classification ascendante hiérarchique du tableau de contingence types de <div> × lemmes

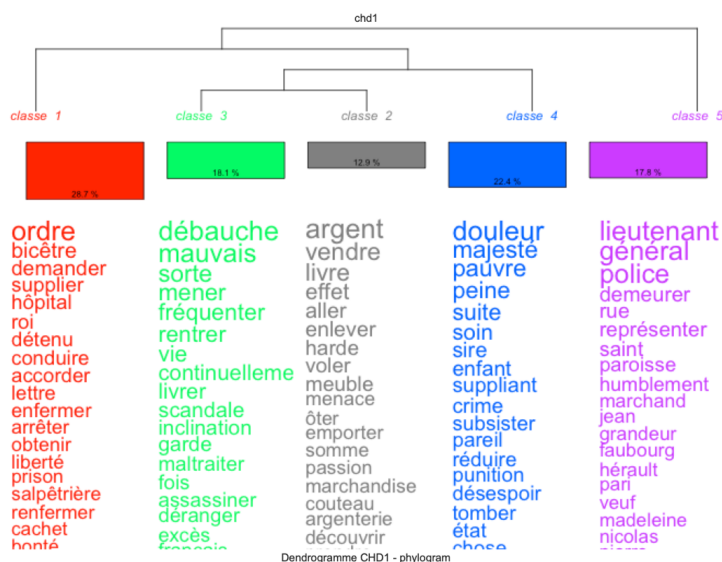


figure . 5 : Classification descendante hiérarchique méthode Reinert (Iramuteq)

Références bibliographiques

AUTHIER-REVUZ, J., 2020, *La Représentation du discours autre. Principes pour une description*, Paris, De Gruyter, 685 p.

BENZÉCRI, J.-P., 1973, *L'Analyse des données*, Paris-Bruxelles-Montréal, Dunod, 2 vol., 1234 p.

DUMOULIN, H., 2022, *Les théorisations du discours de Michel Pécheux et Michel Foucault à la lumière du concept d'énonciation*, thèse de doctorat soutenue le 9 décembre 2022, Université Paris Nanterre, 700 p.

FOUCAULT, M. & FARGE, A., 1982, *Le Désordre des familles. Lettres de cachet des Archives de la Bastille*, Paris, Gallimard-Julliard, 362 p.

FOUCAULT, M., 1961, *Histoire de la folie à l'âge classique*, Paris, Gallimard, 583 p.

FOUCAULT, M., 1969, *L'Archéologie du savoir*, Paris, Gallimard, 275 p.

FOUCAULT, M., 1973, *La société punitive*, Paris, Gallimard, Hautes Études, 349 p.

FOUCAULT, M., 1975, *Surveiller et punir*, Paris, Gallimard, 318 p.

FOUCAULT, M., 1994, *Dits et Écrits*, 4 volumes, Paris, Gallimard, 3427 p.

- FLEURY, S. & ZIMINA, M., "Trameur : A Framework for Annotated Text Corpora Exploration", *Proceedings of COLING 2014, the 25th international Conference on Computational Linguistics : System Demonstrations*, Dublin, pp. 57-61.
- HEIDEN, S., MAGUÉ, J.-P., PINCEMIN, B., 2010, « TXM : Une plateforme logicielle open-source pour la textométrie-conception et développement », *JADT 2010 : actes des 10^{èmes} Journées internationales de l'Analyse statistique des Données Textuelles*, Rome, vol. 2, n°3, pp. 1021-1031.
- MARCHAND, P., RATINAUD, P., 2015, « Des mondes lexicaux aux représentations sociales. Une première approche des thématiques dans les débats à l'Assemblée nationale (1998-2014) », *Mots. Les langages du politique*, 2015/2, n°108, pp. 57-77.
- NÉE, E., OGER, CL., SITRI, F., 2017, « Le rapport : opérativité d'un genre hétérogène », *Mots*, n°114.
- REINERT, M., 1983, « Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte », *Les Cahiers de l'Analyse des Données*, 1983, vol. VIII, n°2, pp. 187-198.
- REVEL, J., 2010, *Foucault une pensée du discontinu*, Paris, Mille et une nuits, 333 p.
- SOBOUL, A., 1981, *Comprendre la Révolution. Problèmes politiques de la Révolution française, 1789-1797*, Paris, Maspéro, 379 p.