



## **Livrable en Humanités Numériques**

### **Compte rendu critique de l'utilisation d'un outil appliqué à notre corpus : Voyant Tools**

Premièrement, il est sûrement de bon ton de définir les Humanités Numériques, d'autant plus que cette définition fait l'objet de débats car elles désignent un domaine de recherche interdisciplinaire, donc une approche globale qui permet de dépasser la dichotomie entre sciences « dures » et « molles ». En 2015, Alexandre Gefen dit même qu'il n'y a pas de définition figée de ce champ disciplinaire qui est en train de s'inventer. Néanmoins, il semble y avoir un consensus sur le fait qu'elles peuvent nous aider à étudier de nouveaux objets, qu'elles nous apportent de nouvelles méthodes également et qu'elles permettent la diffusion des recherches et du patrimoine en invitant notamment au dialogue entre les disciplines, ainsi que de nouvelles méthodes d'apprentissage avec le *e-learning*. Le dialogue interdisciplinaire est notamment prépondérant dans mon sujet.

Je réalise un mémoire en histoire culturelle et sociale des XX<sup>e</sup> et XXI<sup>e</sup> siècles sous la direction de Christian Delporte. Le sujet s'intitule : « L'Affaire Violette Nozière<sup>1</sup> : du fait divers à la fiction (1975-aujourd'hui) ». La problématique principale que soulève ce sujet – et de laquelle découle toute une arborescence de questionnements – porte sur l'évolution de la figure mythique de Violette Nozière depuis les années 1970 dans les représentations. Cela à travers des productions culturelles, dont les auteurs sont eux-mêmes travaillés par les représentations de leur temps et qui se réapproprient l'image de Violette, en oscillant entre héritages et ruptures avec les fictions antérieures. Ma méthode consiste donc en l'étude des œuvres elles-mêmes, en l'immersion dans leur mise en contexte et en leur remplacement dans un contexte plus large. Ce sujet s'inscrit à la croisée de divers champs historiques, notamment l'histoire culturelle et son sous-champ l'histoire des représentations<sup>2</sup>.

L'histoire des représentations se définit en relation avec l'analyse des pratiques comme une histoire de la manière « dont les hommes voient, vivent voire produisent le monde dans lequel » ils évoluent. Après avoir été touchée par la « crise de l'histoire<sup>3</sup> » et la crainte liée au *linguistic turn*, aujourd'hui l'usage de la notion de représentations semble apaisé, éprouvé par plus de trente ans de travaux. Les

1 Paris, 1933. Violette Nozière a empoisonné ses parents, son père meurt et sa mère en réchappe. Une fois arrêtée, après s'être enfuie, ce qui passe comme un aveu de sa culpabilité, elle avoue le parricide mais donne comme mobile l'inceste que son père lui aurait fait subir depuis ses 12 ans. Fait indicible et inaudible dans l'entre-deux-guerres, elle est condamnée à mort, graciée puis réhabilitée en 1963. Tous les tenants et les aboutissants avérés de l'affaire sont étudiés par Demartini, Anne-Emmanuelle, *Violette Nozière, la fleur du mal. Une histoire des années trente*, Ceyzérieu, Champ Vallon, 2017.

2 La définition qui va suivre est inspirée de celle de VENAYRE, Sylvain, « Représentations », in DELPORTE, Christian, et al., *Dictionnaire d'histoire culturelle de la France contemporaine*, Paris, PUF, 2010, p. 697-701.

3 L'on doit cette impression de « crise de l'histoire », partiellement, à la fin des grands paradigmes, tel l'effondrement du marxisme à la suite de la chute des régimes communistes, résultant de la Guerre froide. Et cette impression s'amplifia avec la fin des ventes miraculeuses d'ouvrages historiques de la génération de Georges Duby. Plusieurs historiens comme Paul Veyne exprimèrent ces doutes : voir VEYNE, Paul, *Comment on écrit l'histoire. Essai d'épistémologie*, Paris, Éditions du Seuil, 1971, ouvrage dans lequel il réfute l'idée même de l'existence d'une vérité historique, car une vérité unique n'existerait pas. Les tenants des *Annales* préfèrent l'expression de « tournant critique ».

années 1980 marquent le tournant entre mentalités et représentations<sup>4</sup>. Cet infléchissement historiographique se caractérise notamment par la notion de « système de représentations », prouvant la primauté nouvelle dédiée au discours dans la démarche historique. Mon sujet s'y inscrit donc puisque ma réflexion se fonde sur la distinction et l'intrication des différents types de discours qui construisent mon objet et émanent des différentes sources, en tentant ainsi le repérage de systèmes de représentations et le re-traçage des moments significatifs. Ainsi, si je développe ce qu'est l'histoire des représentations, c'est parce que cela permet de mettre en lumière le terme fondamental « discours ». Cela justifie mon choix de réaliser un livrable sur l'outil Voyant Tools, qui pourrait me permettre d'observer justement ces moments significatifs.

## **I/ Justification du choix de Voyant Tools**

### ***Processus de conception***

D'abord, quelle question voudrais-je poser à mon corpus ? Dans le cadre de mon sujet, je cherche à analyser quels sont les thèmes prépondérants dans les différents discours qui peuvent-être tenus sur l'affaire Nozière ; mais aussi de voir s'il y a une évolution temporelle de ces thèmes et si, selon l'émission, l'article, la fiction ou la station de radio, ce sont les mêmes. Aussi, puisque le mystère de cette affaire repose sur l'accusation d'inceste, et l'inceste étant tabou – bien que les verrous du secret pesant sur sa dicibilité semblent sauter timidement peu à peu –, j'aimerais analyser la façon dont on en parle ; si c'est par euphémisme (paraphrase voire adjectif), selon les médias et dans le temps. J'ai ensuite interrogé le type de données que j'avais : pour cet exercice, ce sont des retranscriptions d'émissions radiophoniques écoutées à l'Inathèque. Puis j'ai cherché quel outil me permettrait avec ces données de répondre à la question que je souhaitais poser à mon corpus. Puisque c'est une analyse textuelle que je voulais mener, j'ai choisi Voyant Tools. Enfin je me suis demandée sous quelle représentation visuelle j'aimerais montrer les résultats. J'ai conclu que je souhaitais utiliser des visualisations montrant la prépondérance de certains mots (nuages de tags), des courbes de fréquence ou des histogrammes pour traduire visuellement le corpus et plus précisément l'utilisation d'un mot dans le temps. Et si possible peut-être voir les co-occurrences.

### ***Place de Voyant Tools dans les logiciels de visualisation***

Il existe différents types de visualisations selon l'usage que l'on veut en faire : si elle doit servir une démonstration ou à la navigation dans son corpus, ou encore à la formulation ou à l'analyse. L'on peut citer Tableau Public, sauf que ce logiciel ne permet pas une analyse textuelle ; Xmind peut permettre quant à lui de représenter le modèle conceptuel d'une base de données, je l'ai d'ailleurs utilisé pour ma « culture générale ».

4 C'est pour cela que je préfère parler d'histoire des « représentations » plutôt que de « mentalités » car la première se substitua à l'autre, jugée obsolète – y compris par ses tenants comme Georges Duby en raison de sa faible valeur heuristique. Ce tournant s'opère sous l'impulsion puis la diffusion de la pensée de Michel Foucault – malgré les critiques à son encontre résumées par Jacques Léotard –, ou sous celle de Michel de Certeau pour d'autres. Voir *Ibid.*, p. 699.

### ***Place de Voyant Tools parmi les outils d'analyse textuelle***

Sur Dataviz catalogue, si l'on cherche un outil par fonction et que l'on sélectionne « *Analysing Text* » donc une méthode de visualisation qui révèle les motifs d'un corpus de texte, la plateforme propose le nuage de mots et plusieurs outils permettant d'en réaliser. Parmi ceux-ci, beaucoup nécessitent de savoir coder, ce dont je n'ai, actuellement, pas la compétence (AnyChart, D3, R Graph Gallery & Python Graph Gallery, etc.). Les autres ne génèrent – presque – que des nuages de mots (Wordclouds.com, Wordcloud.pro, Wordle, etc.). La page ne cite pas Voyant Tools, cependant c'est celui que je choisis car je trouve que le seul nuage de mots est insuffisant car davantage esthétique qu'exploitable. En effet, il peut ne pas être aisé de distinguer la taille de deux mots, sûrement en partie à cause d'un effet d'optique, les divers mots étant disposés couleurs variées, horizontalement ou verticalement. C'est là l'intérêt de Voyant Tools : il propose différentes représentations visuelles, ce qui permet de choisir la plus adaptée à son corpus. Autrement, il existe Tropes mais il semble moins facile d'utilisation. Il existe d'autres logiciels de textométrie<sup>5</sup>, mais je ne les ai découverts qu'après avoir réalisé ce livrable, tel que HyperBase dont certaines fonctionnalités semblent se rapprocher de celles de Voyant Tools : le calcul d'évolution du vocabulaire pour des corpus diachroniques ressemblant aux « Tendances » de Voyant Tools.

### ***II/ Présentation de l'outil Voyant Tools***

#### ***Place de Voyant Tools dans l'histoire des Humanités numériques<sup>6</sup>***

Selon Lou Burnard dans *Du literary and linguistic computing aux digital humanities : retour sur 40 ans de relations entre sciences humaines et informatique* paru en 2012, il existerait trois âges des Humanités numériques dont le premier correspond aux années 1960-1980 et qu'elle nomme le *Literary and linguistic computing*<sup>7</sup>. Je ne reviendrais pas sur les travaux précurseurs du Père Busa. Les années 1960-1970 sont donc marquées par les débuts de la linguistique computationnelle avec la naissance d'outils pour le traitement automatique en linguistique de corpus, notamment. La lexicométrie, rebaptisée dans les années 1990 textométrie, permet de faire des statistiques grâce aux machines afin de repérer des irrégularités dans les textes, donc des schémas ou des tendances, d'établir des concordances également ; *grosso modo* d'opérer une lecture distante des textes. Procédés philologiques centenaires, les machines permettent un gain de temps incroyable. Les années 1990 sont aussi celles où les Humanités Numériques se développent et se structurent fortement au Canada, pays d'origine de Voyant Tools<sup>8</sup>. Dès les années 1960, la lexicométrie est appliquée aussi en histoire, surtout dans le cadre de l'analyse des discours politiques. Mais du fait du côté très conservateur de la discipline, cette dernière

5 Voir : Pincemin, Bénédicte, « Sept logiciels de textométrie », HAL-SHS, 2018.

6 Mounier, Pierre, « Du *Literary and linguistic computing* aux *Digital Humanities* : retour sur 40 ans de relations entre sciences humaines et informatique », Hypothèses, Open Edition, Digital Humanities > EHESS > Séminaire du CLÉO, publié le 30/01/2010, mis à jour le 16 mars 2010, consulté le 04/12/2020.

7 Le deuxième âge correspond aux années 1980 à 1994 des *Humanities computing* puis de 1994 à aujourd'hui par les *Digital humanities* (ou Humanités numériques).

8 Auréline Berra, repository « Voyant Tools », GitHub.  
[https://github.com/aurelberra/voyant\\_tools/blob/master/tutorial/voyant\\_tools\\_intro\\_fr.md](https://github.com/aurelberra/voyant_tools/blob/master/tutorial/voyant_tools_intro_fr.md)

reçoit difficilement ces méthodes conduisant à leur éclipse. Mais aujourd'hui, elle est de retour en grâce avec un retour de l'analyse du discours ; notamment grâce à la mise à disposition de corpus textuels numérisés par Ortolang. Ainsi, utiliser Voyant Tools est en quelque sorte renouer avec les premières amours des Humanités Numériques.

### ***Voyant Server***

Étant donné que nos retranscriptions sont issues de la consultation d'émissions de radio conservées par l'Inathèque et que des droits les encadrent, j'ai fait le choix d'installer la version serveur du logiciel sur ma machine, grâce au *repository* GitHub édité par le co-créateur de l'outil, « VoyantServer Desktop<sup>9</sup> ». Autrement, l'on peut accéder librement à Voyant Tools via son site canadien<sup>10</sup>, ou sur son site miroir français hébergé par Huma-Num, attestant alors du sérieux du projet<sup>11</sup>. C'est à partir de la page d'accueil que l'on peut ajouter des textes ou son corpus : soit en ajoutant un ou des URL de textes en ligne ou en « copiant-collant » un texte, soit en chargeant un corpus d'entraînement proposé par Voyant Tools, soit en téléchargeant ses propres documents stockés localement.



### ***Définition de Voyant Tools et historique***

Voyant Tools est un outil d'analyse, de lecture distante et de visualisation de textes. C'est le fruit d'un projet *open source* (licence GPL) canadien, dont le code est disponible sur GitHub et qui a été dirigé par Stéfán Sinclair et Geoffrey Rockwell – deux universitaires canadiens à la tête d'une petite équipe de spécialistes des sciences humaines numériques. Aurélien Berra<sup>12</sup> indique son histoire : pour ses recherches doctorales sur l'Oulipo, le co-concepteur du présent Voyant Tools, développe dans les années 2000 le logiciel HyperPro, dont la plateforme s'ouvre à partir de 2008<sup>13</sup>, d'abord sous le nom de Voyeur, puis de Voyant. La plateforme recourt aux technologies Flash et Java<sup>14</sup>. C'est un logiciel qui est mis à jour comme le montre la mention d'Aurélien Berra de la parution en 2015 de Voyant 2.0 comportant « une réécriture en HTML 5 » qui améliore « les filtres de requête et introduit des outils fondés sur le calcul de proximité et les séquences d'éléments, ou n-grammes », c'est-à-dire la possibilité de travailler sur les co-occurrences, ce que je rechercherais peut-être.

9 <https://github.com/sgsinclair/VoyantServer/wiki/VoyantServer-Desktop>

10 <https://voyant-tools.org>

11 <http://voyant.tools.huma-num.fr>

12 Aurélien Berra est un enseignant-chercheur dont les domaines de recherche sont les langues et littératures anciennes. Il a, en 2016, fait la traduction française de l'interface ; et les deux années suivantes, il a dressé des listes de mots-outils ou « stopwords » pour le Grec et le Latin qui ont été intégrées au logiciel.

13 Wikipédia indique que la date initiale de sortie du logiciel est 2003.

14 C'est pourquoi, pour faire fonctionner le logiciel dans sa version locale sur nos machines, une version de « Java (8+) » est requise.

### ***Voyant Tools ou l'incarnation de l'esprit des Humanités Numériques ? Les principes du logiciel.***

Le fait que le code soit disponible montre que le logiciel repose sur un principe d'ouverture, encourageant toute contribution possible et reposant alors probablement sur l'intelligence collective. L'incipit du fichier Readme de l'entrepôt GitHub contenant le code du logiciel indique que – ici la traduction automatique par le logiciel DeepL – « Voyant Tools est conçu pour un très large éventail d'applications et d'utilisateurs, des étudiants aux chercheurs et journalistes en passant par les analystes de marché. » Aurélien Berra rappelle, sur le *repository* GitHub qu'il édite sur Voyant Tools en Français, que : « L'élaboration de cette plateforme s'inscrit dans un projet plus vaste, présenté dans un livre : Geoffrey Rockwell et Stéfan Sinclair, *Hermeneutica. Computer-Assisted Interpretation in the Humanities*, Cambridge, Massachusetts, MIT Press, 2016, dont [Hermeneuti.ca](http://Hermeneuti.ca) est le site compagnon. » Au-delà d'être mis à jour, Voyant Tools est pensé par et pour l'évolution des Humanités numériques avec une volonté d'adapter « la tradition du *literate programming* aux sciences humaines et sociales », comme le souligne Aurélien Barré en parlant de « sa modularité et son caractère évolutif » – qui sont deux des principes de conception de l'outil. Cela a été amorcé par la possible intégration d'« outil ou de vues dynamiques dans une page Web » et selon lui « la prochaine étape importante » sera sûrement « celle des carnets Spyral ». Car en plus d'intégrer des outils au sein du texte, il est question de la publication de « notebooks » qui allie « code et commentaire, analyse et argumentation ». Enfin, pour ce livrable, j'ai suivi le principe – intuitif finalement – de Voyant Tools puisque j'ai fouillé, manipulé donc interagi avec les données aisément « pour apprendre du processus et produire des séries de résultats ». Ce choix de Voyant Tools est une réponse « à la conviction que la théorie et la pratique sont intimement mêlées ».

### ***Préparation des données***

Voyant Tools intervient au moment de l'analyse. J'ai d'abord collecté mes données (en réalisant les retranscriptions des archives audiovisuelles sur Violette Nozière à l'Inathèque). J'ai fait le choix de retranscrire intégralement les passages évoquant Violette Nozière, que ce soit pour l'affaire même ou via sa fictionnalisation, ceci afin de mieux ressaisir *a posteriori* les émotions, les hésitations, etc. Étant tributaire du temps, je retranscris les émissions avec de nombreuses abréviations qui sont toujours les mêmes car un travail de reprise est nécessaire ensuite. Mais avec la fonctionnalité « rechercher & remplacer » de mon logiciel de traitement de texte (Libre Office), c'est un gain de temps. Mais je ne suis pas à l'abri de biaiser mes données – avec des doubles espaces par exemple qui pourrait fausser les scores de co-occurrences potentiellement. Il faut donc passer obligatoirement par l'étape du nettoyage de la donnée<sup>15</sup>. Ayant fait toutes mes retranscriptions dans le même fichier avec des « métadonnées » (précisions sur la chaîne, la date, le nom de l'émission, des remarques etc.), j'ai dû remettre chaque retranscription dans un fichier unique auquel je donnai pour nom la date, la chaîne,

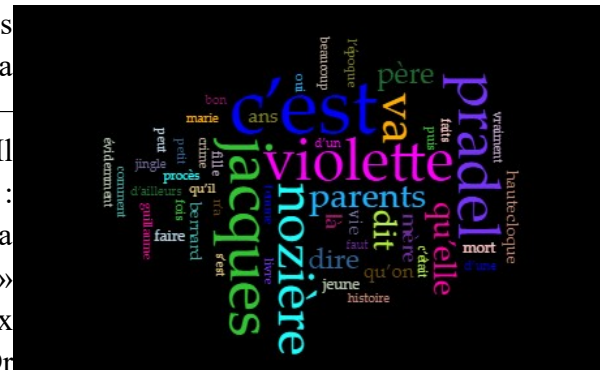
15 J'ai tenté d'utiliser Open Refine en regardant des tutoriels, en fouillant un peu partout mais je ne suis parvenue qu'à trouver comment, pour des données uniquement textuelles, enlever les doubles espaces, les majuscules, etc. L'expérience n'a donc apparemment pas été fructueuse.



puis le titre de l'émission, et enfin la durée. Puisque l'étape de préparation et nettoyage des données est indispensable mais chronophage, et que dans mon échéancier j'avais prévu cela pour les mois de janvier-février, j'ai choisi de ne mobiliser qu'un échantillon de mon corpus : les émissions de RTL. Cela pour plusieurs raisons : y était diffusée « L'Heure du crime » présentée par Jacques Pradel, émission pendant laquelle ce journaliste a reçu plusieurs des auteurs de mes sources principales – les fictions sur Violette Nozière. Aussi, car cela ouvre une fenêtre sur neuf ans (2009-2018), ce qui permet de voir une évolution, ou non, dans le temps –, mais également de comparer si selon l'invité le discours est différent – dans le fond et formellement – et à quel point. J'ai en effet pu constater « humainement » qu'il semblait y avoir une adaptation du discours du journaliste selon son invité, ce que j'aimerais vérifier statistiquement. Et choisir les émissions d'une seule station comme échantillon permettra de comparer les résultats avec ceux des autres émissions. Cet échantillon comporte également des biais : certaines des émissions ne sont pas dédiées exclusivement à Violette, donc je n'ai retranscrit intégralement que les passages sur elle car ce sont les représentations que l'on donne de cette jeune femme qui m'intéressent. Aussi, il ne couvre pas toute la période de mon sujet qui commence en 1975.

### *Analyse des résultats obtenus avec Voyant Tools*

Pour être honnête, je pense ne pas avoir réussi à nettoyer correctement mes données puisque j'ai laissé des sortes de tics de langage oral, le nom des personnes s'exprimant aussi. Cependant, Voyant Tools permet d'y remédier. Voici ce que Cirrus – un nuage de tags (donc une représentation « simplifiée » d'un corpus) pour lequel la couleur et l'orientation des mots sont choisies aléatoirement – donne après le chargement de mon corpus nettoyé au mieux. Il ressort parmi les mots avec le plus grand nombre d'occurrences : « jacques » et « pradel », le prénom et le nom du présentateur de la plupart des émissions qui composent cet échantillon, « Bernard » « Hauteclouque » est l'un des invités du journaliste pour deux émissions<sup>16</sup>. Et est mise en valeur surtout l'expression « c'est ». Or chacun de ces mots ne m'intéressent pas pour l'étude des discours et des thèmes prépondérants associés à Violette Nozière. J'ai donc tenté d'ajouter ces mots à la liste de *stop-words* (ou mots-outils) proposée pour le Français par Voyant Tools. Cependant, il semble impossible de se débarrasser du « c'est », donc j'ai procédé autrement en créant une liste des « mots inclus<sup>17</sup> ». J'ai une assez bonne connaissance de ces sources mais surtout de l'affaire Nozière elle-même. Je sais donc quels sont ses thèmes principaux : le parricide, le poison et l'inceste. Cependant, il se peut que ces mots ne soient pas employés comme

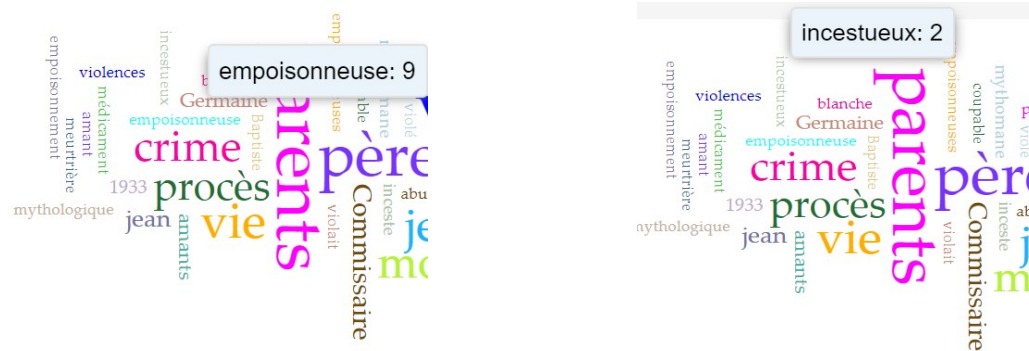


<sup>16</sup> HAUTECLOQUE, Bernard, *Violette Nozière. La Célèbre empoisonneuse des années trente*, Nantes, Normant, 2010.

<sup>17</sup> Parmi eux : violette, nozière, parents, père, mère, crime, meurtre, meurtrière, viol, violence, violences, violée, violé, violait, abus, abusé, abusait, inceste, incestueux, incestueuse, Guillaume, Commissaire, 1933, poison, poisons, empoisonne, empoisonnement, empoisonneuse, empoisonneuses, empoisonné, poudre, blanche, médicament, médicaments, jeune, histoire, époque, l'époque, procès, fille, livre, vie, mort, victime, coupable, Germaine, Baptiste, Jean, amant, amants, mensonge, mensonges, mythomane, menteuse, mythe, mythique, mythologique, etc.

tels, j'ai alors ajouté leurs déclinaisons : empoisonnement, empoisonneuse, crime, meurtre, meurtrière, inceste, incestueuses, incestueux, abus, abusé, etc. J'ai aussi utilisé les principaux mots qui étaient ressortis sans réglage et qui, eux, m'intéressaient, tels que : « vie », « procès », « parents », « famille », « mère », « père », etc. Rien d'étonnant, c'est un drame familial. Voici le résultat ci-contre :

Néanmoins, malgré l'esthétisme de Cirrus, j'aimerais aussi une visualisation plus « fine ». Je justifiais justement mon choix de Voyant Tools par le fait qu'il offre d'autres types de visualisations que le nuage de mots, qui paraît plus globalisant et esthétique qu'exploitable. On le voit avec l'exemple ci dessous au sujet des termes « empoisonneuse » et « incestueux » qui ont respectivement 9 et 2 occurrences dans le corpus :

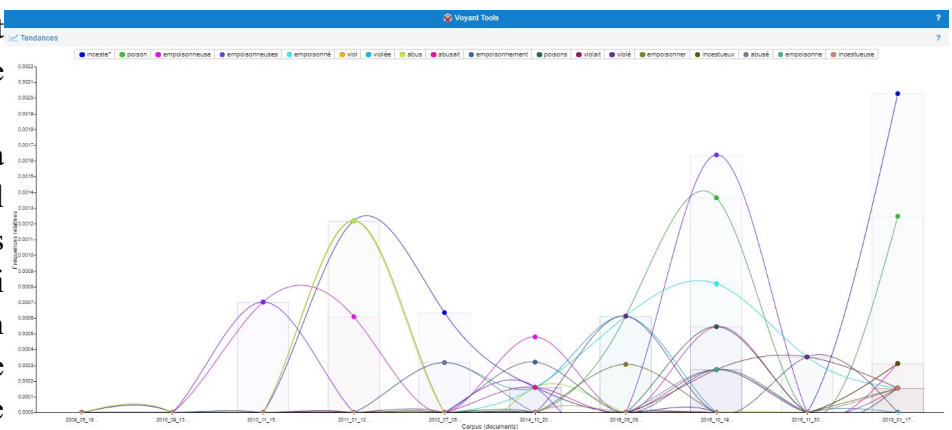


Disons donc que Cirrus offre un aperçu, une appréhension globale du corpus chargé. Mais puisque je souhaitais voir s'il y a une évolution des thèmes, je me suis ensuite tournée vers « Tendances » puisque l'axe des abscisses plaçait mes documents dans l'ordre chronologique des émissions (c'est pour cela que j'avais nommé les fichiers en commençant par la date de diffusion).

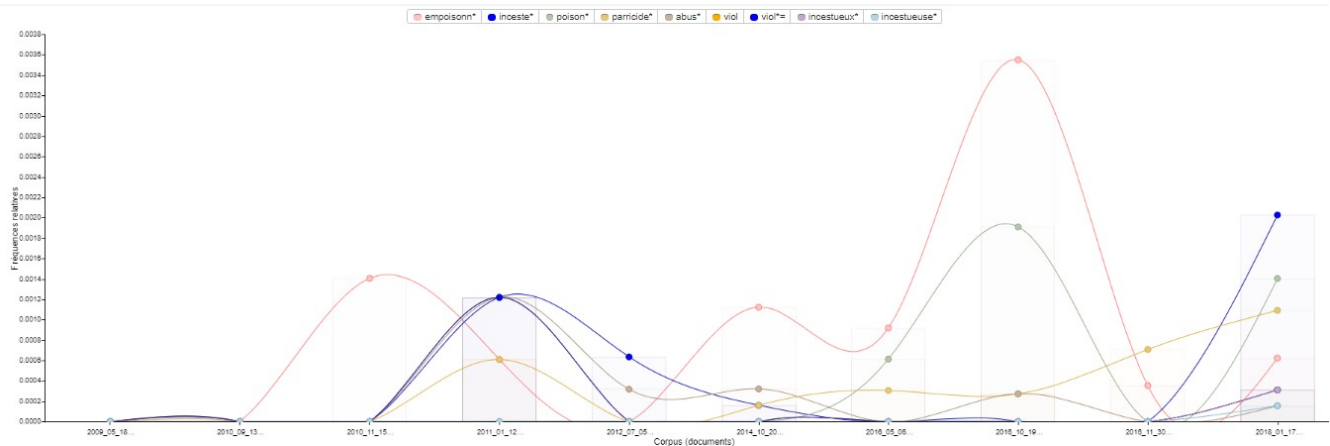
Étant donné que je souhaite voir lequel des thèmes est le plus exploité dans les discours entre le parricide, le poison et l'inceste, j'ai entré ces mots et leurs déclinaisons. Mais entrer toutes les déclinaisons donne un résultat

brouillon, presque illisible comme le montre la capture d'écran ci-contre :

C'est pourquoi j'ai décidé, à l'exception de l'inceste pour lequel j'ai gardé les déclinaisons (incestueuse et incestueux mais aussi abus et viol) – car la façon dont on parle de l'inceste m'importe également pour étudier la levée timide des verrous du tabou –, d'entrer pour les autres thèmes la forme racinedumot\* pour trouver les

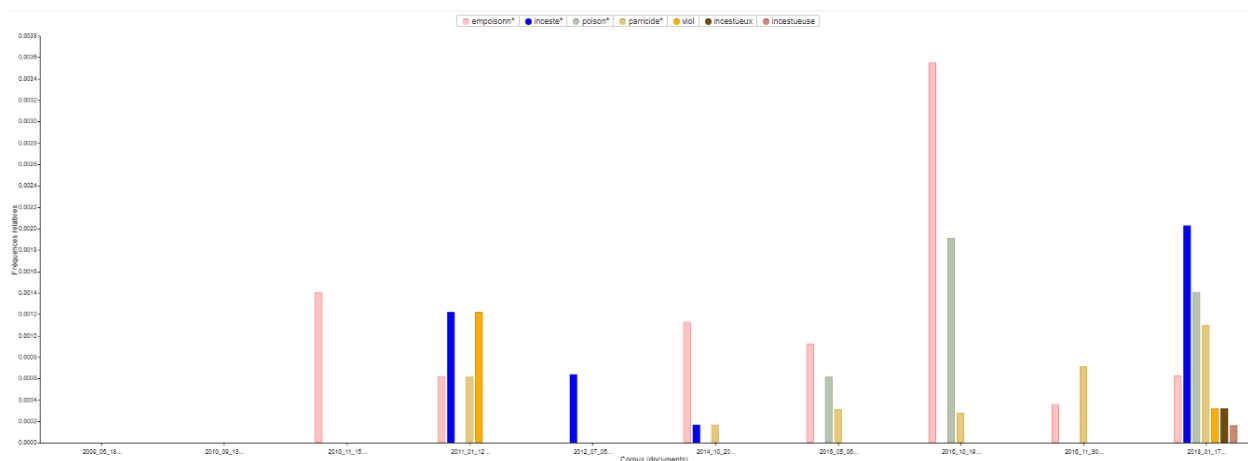


termes qui commencent par le terme exact *racinedumot* (exemple : empoisonn\* qui prend en compte empoisonnement, empoisonneuse, etc.) ; cf. le graphe ci-dessous :



J'ai essayé aussi avec seulement les termes d'inceste, de poison et de parricide mais le résultat était biaisé car l'on a l'impression que la question des abus sexuels subis – ou non – par Violette n'est pas évoquée dans les émissions de 2016 et que la question du poison, n'émerge qu'à partir de 2016. Dans les deux cas cités précédemment, je n'étais encore une fois pas totalement satisfaite de cette visualisation puisque les courbes se chevauchent quand il y a le même nombre d'occurrences dans un même document.

C'est pourquoi j'ai choisi d'afficher les « Tendances » par colonnes car c'est ce qui offre une meilleure lisibilité de la visualisation :

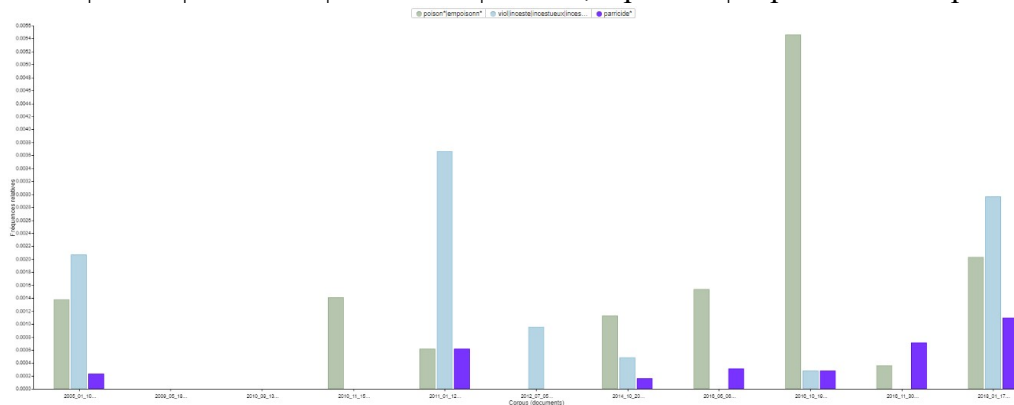


Cet histogramme montre que le thème du poison est prépondérant pendant la période – sauf pour la dernière émission de 2016 de ce corpus où c'est le parricide car l'on parle de la rareté de ce crime. Mais surtout sauf lorsqu'il s'agit d'émissions où l'on défend la validité de la thèse de l'inceste : la primauté va là au thème de l'inceste. Remarque : si l'on a l'impression qu'on ne parle ni de parricide ni d'empoisonnement dans l'émission de 2012, il n'en est rien, on en parle sous forme de paraphrases



avec la mention du « médicament » donné par Violette à ses parents. S'il y a une explosion de l'utilisation du terme inceste dans la dernière émission, c'est parce que l'invitée est la spécialiste de l'affaire, Anne-Emmanuelle Demartini, qui, pour des raisons scientifiques, défend la thèse de l'inceste. Aussi, les verrous du silence pesant sur la dicibilité de l'inceste sautent peu à peu, timidement. Si les deux premières émissions ne comportent aucun des mots recherchés, c'est parce que la première est une présentation courte du film de Claude Chabrol, *Violette Nozière*, diffusé ce soir-là sur Arte ; et la deuxième est un hommage à Chabrol qui vient de décéder et dont on rappelle la filmographie sans qualifier le crime de Violette. Cela montre que Voyant Tools est un outil et non pas une preuve et que pour utiliser les résultats qu'il génère, il faut avoir une relativement bonne connaissance de son corpus.

Par exemple, je sais que dans l'ensemble de mes retranscriptions j'ai une émission de France Inter de 2005 pendant laquelle Patrice Gélinet invita Véronique Chalmet pour parler de son livre *Violette Nozières*<sup>18</sup>, *la fille aux poisons*, sorti en 2004. Cette écrivaine-journaliste est la première pendant la période sur laquelle court mon sujet qui accrédite la thèse de l'inceste et en fait le cœur de son livre et ce, dès la deuxième ligne. Ainsi, il n'est pas étonnant que l'inceste soit le terme le plus courant du document correspondant à cette émission car c'est le thème central de son livre. Cf. l'histogramme ci-dessous où j'ai associé, pour une question de lisibilité, les termes correspondant au même thème « viol|inceste|incestueux|incestueuse|abus\* », « poison\*|empoisonn\* » et parricide :



Définitivement, l'imaginaire du poison prévaut en général dans les émissions sauf dans celles où l'on donne la parole à ceux qui croient en l'accusation faite par Violette à l'égard de son père. Cela atteste de la force de la figure de l'empoisonneuse. Évidemment, l'on peut recommencer avec d'autres thèmes comme la mythomanie pour comparer les résultats et monter en généralité. Si je ne présente que ces deux outils intégrés dans Voyant Tools, c'est parce qu'en testant les autres outils de visualisations, ils ne sont pas adaptés à ce que je recherchais : Grappe de termes est intéressant pour ses deux stratégies : termes les plus fréquents par défaut ou distincts. Mais, en capture d'écran c'est illisible, comme les Lignes de bulles, Mandalas, Nuages de points ou Arcs de texte. Et les outils de tableaux ne

<sup>18</sup> Le « s » final du patronyme n'est pas une erreur de frappe mais l'orthographe choisi par l'auteure et que l'on pouvait retrouver dans certains titres de journaux, indifféremment de l'orthographe sans « s » final.

sont pas très concluants (Contextes, Corrélations ou Collocations) pour ce que je voulais faire<sup>19</sup>. À la limite, par exemple, on peut noter dans « Collocation » que « inceste » (le « terme ») est d'abord corrélé avec « père » puis « accusations » et « n'a » (« Collocation » donc les mots trouvés à proximité du mot-clef « inceste »), ce que l'on retrouve aussi dans l'outil « Liens ». La négation n'est pas étonnante mais il faut chercher le contexte : si c'est pour dire que l'inceste était tabou donc inaudible ou pour nier la thèse. Ici, c'est le premier cas.

### *Avantages et limites de Voyant Tools*

La première limite de la version française concerne les *stop-words*. Si le logiciel en propose une liste automatique, il laisse passer les expressions comme « c'est » et il est impossible de l'ajouter à cette liste pour qu'elle disparaisse des mots avec le plus d'occurrences. Aussi, le logiciel est ambigu avec les « l' » ou les « d' ». Car s'il prend en compte « l'histoire », il « oublie » « l'inceste » et « d'inceste ». Il a donc fallu que je recommence mes analyses de résultats après avoir modifié mes fichiers pour que toutes les mentions de la notion d'« inceste » apparaissent car cela faussait réellement les résultats : le terme inceste n'apparaissait qu'une fois dans tout le corpus chargé, dans le dernier document. Heureusement, ayant noté toute ma démarche, ce fut relativement rapide à refaire.

Voyant Tools présente néanmoins plusieurs avantages. Pour mes attentes, il semble être le meilleur du point de vue flexibilité-difficulté. L'on peut l'utiliser pour davantage de langues que Tropes, reste à voir à l'usage ce que cela donne. Voyant Tools est agréable et assez intuitif ; et même s'il est relativement facile à prendre en main, il existe un tutoriel en Anglais sur le site du projet Hermeuneti.ca, une documentation en ligne ou le tutoriel proposé par Aurélien Berra. Je l'ai déjà dit mais il propose plusieurs types de visualisations afin que l'utilisateur puisse choisir celle qui lui convient le mieux. Il offre de surcroît la possibilité de comparer des corpus entre eux ou de varier les échelles : prendre en compte plus ou moins de mots dans les visualisations. Aussi, les visualisations se recalculent sans cesse selon ses sélections. Il est également mis à jour. La version serveur a encore d'autres avantages : outre le fait crucial de respecter les droits sur les documents, une connexion Internet n'étant pas requise, il y a la possibilité de relancer le serveur s'il est trop lent, et il permet un maniement plus aisé de gros corpus. Sur une machine de plus de cinq ans dédiée au traitement de texte, malgré une installation de près d'une heure, il fonctionne très bien. Enfin, il semble exister une communauté autour de Voyant Tools, via GitHub ou par le biais du Twitter de la plateforme qu'Aurélien Berra dit être « un efficace canal d'information sur les évolutions de la plateforme et les usages que d'autres utilisateurs en font ». Cela montre l'inscription de Voyant Tools dans les Humanités Numériques étant donné que l'un des objectifs est de créer une communauté. Enfin, le site semble relativement *responsive* sur un smartphone en mode paysage.

<sup>19</sup> Je pense que les résultats qui ressortent de ces outils ne sont pas concluants parce que je n'ai pas dû réussir à nettoyer correctement mes fichiers. Mais cela ne m'handicape pas pour les analyser.

## **Bibliographie et sitographie**

### **Historiographie**

DELPORTE, Christian, et al., *Dictionnaire d'histoire culturelle de la France contemporaine*, Paris, PUF, 2010, p. 697-701.  
Demartini, Anne-Emmanuelle, *Violette Nozière, la fleur du mal. Une histoire des années trente*, Ceyzérieu, Champ Vallon, 2017.

### **Humanités Numériques et Voyant Tools : articles et ressources numériques**

Berra, Aurélien, « Voyant\_Tools », GitHub, 17/09/2018.  
Deschamps, Christophe, « Voyant Tools, un puissant service de *text mining* en open source », Wordpress, 05/02/2016.  
Mounier, Pierre, « Du *Literary and linguistic computing* aux *Digital Humanities* : retour sur 40 ans de relations entre sciences humaines et informatique », Hypothèses, Open Edition, Digital Humanities > EHESS > Séminaire du CLÉO, publié le 30/01/2010, mis à jour le 16 mars 2010, consulté le 04/12/2020.  
Pincemin, Bénédicte, « Sept logiciels de textométrie », HAL-SHS, 2018.  
Ribeca, Severino, « Home > Search by Function > Analysing Text > Word Cloud », The Data Visualisation Catalogue, datavizcatalogue.com  
« Voyant Tools Help », voyant-tools.org

## **Sources**

### **Fiction citées**

CHALMET, Véronique, *Violette Nozières. La Fille aux poisons*, Paris, Flammarion, 2004.  
HAUTECLOQUE, Bernard, *Violette Nozière. La Célèbre empoisonneuse des années trente*, Nantes, Normant, 2010.

### **Corpus analysé avec Voyant Tools**

INA RADIO, France Inter, 10/01/2005, « Violette Nozière », 27 minutes.  
INA RADIO, RTL 18/05/2009, « RTL Petit Matin. Tranche 06h00/06h30 », 18 minutes.  
INA RADIO, RTL, 13/09/2010, « Laissez vous tenter », 25 minutes 30.  
INA RADIO, RTL, 15/11/2010, « L'affaire Violette Nozière », 1 heure.  
INA RADIO, RTL, 12/01/2011, « le Commissaire Guillaume, source d'inspiration du Commissaire Maigret de Georges Simenon », 52 minutes 48.  
INA RADIO, RTL, 05/07/2012, « Émission spéciale à l'occasion de la sortie du hors-série 100 ans de la Crim' », 1 heure.  
INA RADIO, RTL, 20/10/2014, « Violette Nozière, l'ange noir de l'entre-deux guerres », 59 minutes 59.  
INA RADIO, RTL, 06/05/2016, « Sous le charme du fait divers », 59 minutes 59.  
INA RADIO, RTL, 30/11/2016, « La curiosité est un vilain défaut », 1 heure 59 minutes.  
INA RADIO, RTL, 17/01/2018, « L'Affaire Violette Nozière », 46 minutes 22.