

Les dessous du numérique.

La transformation des approches critiques, Paradigme d'une nouvelle exploration littéraire ?

Comme nous l'avons vu durant cette semaine de réflexions et d'échanges, grâce à l'avènement du numérique, les techniques de recherche littéraire sont devenues plus précises et efficaces qu'auparavant. La digitalisation des données permet d'accéder en peu de clics à des pistes de recherche à explorer par la suite. La constitution d'outils numériques suscite l'espoir que des recherches automatiques puissent être accessibles et utilisables, en s'appuyant sur les habitudes perceptuelles des chercheurs et puissent faciliter la tâche de la recherche sur une grande masse de textes. Néanmoins, ces technologies numériques métamorphosent les approches critiques traditionnelles, en écartant les conséquences pratiques et intellectuelles qui en sont strictement liées.

Afin d'explorer cette conversion liée aux humanités numériques, nous proposons avec cette communication de représenter les différents enjeux du Projet Haine du Théâtre sur un ensemble de textes théâtrophiles et théâtrophobes entre le XVII^e et le XVIII^e siècle. Nous verrons, en particulier, la place qu'occupent les humanités numériques pour dévoiler les divers aspects du corpus en question.

1. Le Projet HdT

Conçu en 2012 par François Lecercle et Clotilde Thouret, le projet La Haine du Théâtre est l'un des nombreux projets du Labex Obvil. La HdT¹ se propose d'examiner les manifestations récurrentes d'hostilité au théâtre tout en explorant les caractéristiques du théâtre occidental et ses modalités d'expression, au cours des siècles, dans divers pays d'Europe, de l'antiquité à nos jours. Les réflexions de l'équipe scientifique de la HdT se basent principalement sur la découverte des circonstances et des arguments qui étaient utilisés dans les controverses théâtrales, entre la fin du XVI^e et le début du XIX^e siècles en particulier en France, Italie, Espagne et dans l'espace germanique.

1.1 Equipe

Ce projet ressemble une équipe internationale : plus de 70 chercheurs, en France et à l'étranger. En particulier, il combine des membres intérieurs appartenant au CRLC, CELLF, CLEA, ELCI, REIGENN et VALE, et une trentaine de membres extérieurs rattachés aux Universités françaises. À l'étranger, une quinzaine de chercheurs et de post-doctorants œuvrent à enrichir le projet la HdT. La nécessité de couvrir l'ensemble des

¹ Cette abréviation désigne dorénavant le projet *Haine du Théâtre*.

domaines européens sur une longue période, du XVI^e au XXI^e siècle, est facilitée par le grand nombre de participants ; cette diversité constitue une plus-value pour la création du corpus numérique, les textes étant parfois édités par les membres de l'équipe. De plus, des collaborations avec des universités étrangères ont été nouées dès le lancement du projet et se montrent très actives (Université de Chicago, Université de Bologne et le théâtre de Modène, le Festival d'Avignon et l'Université de Berkeley).

Dès la naissance de la HdT, les porteurs du projet, François Lecercle et Clotilde Thouret, se sont fixés un double objectif. Une première tâche, parfaitement documentaire, a pour but de recenser un corpus proliférant et mal exploré, ainsi que d'établir des bibliographies plus complètes et de rassembler, à usage interne, des fac-similés des textes. Une seconde tâche, abordant l'analyse documentaire, a été initiée dans le cadre d'une série de séminaires, journées d'étude, colloques et publications, dont les résultats sont consultables en ligne sur le site de l'Obvil².

1.2 Corpus

Le projet Haine du Théâtre réunit un corpus critique de textes polémiques. Ces textes aident à comprendre les normes et les canons de la production dramatique de la période analysée. Ils permettent d'interpréter le point de vue des débats sur les disputes de la vie culturelle, de la place accordée aux formes d'expression culturelle vis à vis de l'instance politique et des normes sociales.

Ce corpus de controverses est encore en cours de constitution, mais représente d'ores et déjà plusieurs dizaines de milliers de pages (téléchargées pour la plupart en pdf). Les porteurs du projet se sont principalement concentrés sur le corpus français en raison de sa faible accessibilité, *a contrario* du corpus anglais, largement disponible sur des sites spécialisés, tels que Early English Books Online³ (EEBO, pour les ouvrages entre 1500 et 1700) et Eighteenth Century Collection Online (ECCO, 1701-1800). Dès à présent, le corpus compte plus de 400 titres pour la bibliographie française, dont une petite centaine a désormais été éditée et se trouve en ligne sur le site du projet.

1.3 Objectif

L'objectif de la HdT est d'explorer les zones d'ombre de la controverse pour esquisser une vue d'ensemble des points centraux qui amènent à ces polémiques, et découvrir où et comment elles commencèrent, leurs divergences chronologiques dans les différents Pays, les liens entre elles ainsi que leurs réapparitions contemporaines. Ainsi, le recours au numérique permet d'examiner de façon inédite le rapport entre les auteurs théâtrophiles et théâtrophobes au sein des différentes cultures : repérer les citations et

² http://obvil.paris-sorbonne.fr/corpus/haine-theatre/seminaire_haine-theatre_2014/

³ <https://eebo.chadwyck.com>

les reprises, les sources utilisées, pour en identifier les similitudes dans les modalités argumentatives.

2. La place du numérique

Selon leur manifeste, « les *digital humanities* ne font pas table rase du passé. Elles s'appuient, au contraire, sur l'ensemble des paradigmes, savoir-faire et connaissances propres à ces disciplines, tout en mobilisant les outils et les perspectives singulières du champ du numérique. Les *digital humanities* désignent une transdiscipline, porteuse des méthodes, des dispositifs et des perspectives heuristiques liés au numérique dans le domaine des Sciences humaines et sociales »⁴.

Dès la naissance du projet, les humanités numériques ont été envisagées pour la HdT comme un moyen pour rendre accessible aux chercheurs des textes souvent inconnus, rares, et introuvables auparavant sur le web. Pour les porteurs du projet, il s'agissait ainsi de faciliter la découverte, l'accessibilité et la lecture des textes traitant des polémiques théâtrales.

Le corpus de la Haine du Théâtre est assez vaste. La bibliographie du domaine français que les porteurs du projet ont constitué compte 400 œuvres. Le corpus est mis en ligne au format XML-TEI⁵, respectant la mise en forme des éditions originelles. Il est composé de 166 fichiers avec une taille totale de 21,7 Mo. De ces 166 titres, 93 ont été chargés sur le site de l'Obvil de manière provisoire à travers une édition diplomatique. Ce choix facilite la récupération des données, bien que la langue n'ait pas été modernisée et les textes n'aient pas été encore annotés. En revanche, sur les 166 textes, les 73 restants ont été édités par des membres de l'équipe, avec une modernisation de la langue et l'ajout de notes explicatives pour les passages dont la compréhension du texte s'avérait problématique.

Ce corpus contient des textes de longueur très disparate : du texte de Bardou⁶, avec ses trois pages (format A4) au *Traité de Voisin*⁷, qui compte plus de 500 pages (format A4). Ce corpus HdT est hébergé actuellement à deux endroits différents : d'une part, sur le site de l'Obvil, où l'on peut consulter les textes, les télécharger un par un, utiliser des fonctions de recherche de mots choisis, lire les comptes rendus des séminaires, les rapports d'activités ; d'autre part, sur la plateforme d'hébergement Github⁸, où les données concernant le projet se retrouvent stockées et en libre accès. Les chercheurs intéressés peuvent très facilement télécharger le corpus dans sa totalité.

⁴ Manifeste des Digital Humanities, <http://tcp.hypotheses.org/318>

⁵ <http://www.tei-c.org/>

⁶ Bardou Pierre, *Epistre sur la condamnation du theatre a Monsieur Racine*, Paris, Chez la veuve de Jean Baptiste Coignard, 1684.

⁷ Abbé Joseph Voisin, *Défense du traité de Mgr le Prince de Conti touchant la comédie et les spectacles ou la réfutation d'un livre intitulé Dissertation sur la condamnation des théâtres*, Paris, Coignard, 1671.

⁸ <https://github.com/OBVIL/haine-theatre>

L'hébergement sur le site du Labex Obvil et sur GitHub permet de nombreux avantages, parmi lesquels le téléchargement d'un texte édité en plusieurs formats (html, tei, epub, txt).

Dès que l'on fait face à l'apport computationnel sur ce corpus, on s'aperçoit que sa contribution ne se limite pas à faciliter la quête des textes : il bouleverse radicalement la tradition de notre rapport à la littérature. Loin de permettre seulement un accès plus favorable aux textes, d'être purement un accessoire, l'informatique provoque une révolution des objets littéraires dans leur constitution, leur qualité et, en conséquence, leur critique. La création des données de la HdT ne s'est réduite en aucun cas à une opération subite et universelle. Sur le plan intellectuel, les principes éditoriaux institués par les porteurs du projet ont privilégié la modernisation des textes afin qu'un lecteur dépourvu de connaissances préalables puisse utiliser le moteur de recherche sans difficultés. Ce choix a également été entrepris pour que les œuvres des différentes époques se retrouvent éditées dans un même type de langue.

Les textes enregistrés sous un format numérique se prêtent bien à des questionnements ludiques, métaphasiques ou expérimentaux. Leur transplanation⁹ d'un support, d'un logiciel ou d'un hypertexte à l'autre, ouvre un nouvel éventail d'applications. La littérature qui se déploie à l'usage computationnel est par conséquent livrée à des analyses en profondeur. L'informatique pour la HdT est donc devenue, de simple moyen pour rendre le corpus accessible en ligne, à instrument privilégié pour expérimenter et sonder les corpus.

La mise à disposition des données, des résultats des analyses, ou tout simplement la visualisation des textes de controverse, sont souvent pris pour acquis par les utilisateurs. L'aisance avec laquelle l'on visualise et l'on récupère sur le web des données en *open access* occulte promptement l'activité préalable qui avait été indispensable pour les rendre intelligibles et maîtrisables sur une plateforme. Les conditions de production des données, leurs règlements, le lien avec le contexte et les difficultés rencontrées sont très rarement mises en évidence. De la sorte, elles sont immanquablement oubliées.

La constitution d'un corpus ne se résume jamais en une mise à disponibilité immédiate, simple et permanente. A la base du processus, trois étapes névralgiques peuvent être constatées : l'identification, l'extraction et le traitement des données. Après avoir constitué le corpus et en avoir extrait les éléments, traiter¹⁰ la donnée devient une exigence pour qu'elle soit facilement réutilisable en différents contextes.

Les humanités numériques transforment des apparences sous-jacentes en objets concrets, quantifiables et évaluables. Des éléments non visibles à la lecture humaine peuvent ressortir à travers l'analyse des données. Mais les résultats et leurs examens

⁹ Référence au « transplanage » d'Harry Potter. Nous nous prenons la liberté d'associer la téléportation magique et l'informatique. La première, permettant de quitter un lieu et de réapparaître dans un autre, se comporte tout comme la deuxième. A l'aide d'interfaces, qui cachent le code et le jeu des données, l'informatique consent la jonction de différents éléments, sans que l'utilisateur final s'aperçoive de ses secrets, qui restent invisibles.

¹⁰ Le parallèle avec la « brutification » des données apparaît clairement. Cfr. Camille Paloque-Berges. *Les sources nativement numériques pour les sciences humaines et sociales*. 2015. <halshs-01239053>

sont assurément dépendant des outils informatiques qui exécutent ces analyses, selon différents types de calculs.

Considérant l'encodage en XML-TEI du corpus, l'équipe de la HdT a choisi de ne pas effectuer un balisage détaillé de tous les éléments sensibles d'un texte¹¹. Il a été décidé d'explorer cet aspect automatiquement, par le biais des programmes. Les logiciels libres créés par le Labex et par ses partenaires sont mis à disposition de la communauté sur une page décidée¹². En particulier, les logiciels d'alignement de textes, *Médite*¹³ et *Phoebus*¹⁴, créés par Jean-Gabriel Ganascia et issus d'une collaboration entre des spécialistes d'intelligence artificielle de l'équipe ACASA du LIP6 et des chercheurs en humanités numériques du Labex OBVIL et de l'ITEM, permettent de confronter les différentes éditions du même texte et de découvrir assez précisément les ajouts, les modifications et les suppressions apportées, ou bien faire ressortir les citations présentes dans deux textes différents. Le logiciel DeSeRT¹⁵, élaboré par Jean-Gabriel Ganascia, est également très utile pour analyser la controverse théâtrale, en permettant de trouver non seulement une citation ou un paragraphe identique dans des textes différents, mais des passages relevant du même champ conceptuel. Pour cette intervention aujourd'hui, nous nous limitons à renvoyer aux articles qui comprennent des résultats de nos recherches¹⁶.

3. La problématique du corpus et la relation avec le numérique

L'évolution digitale déconcerte le rapport entre le contenu et sa conversion, le passage du langage naturel au langage de programmation, les *datas* et leur support. L'expert est en conséquence amené à ouvrir constamment une réflexion épistémologique sur son travail. A titre d'exemple, les exigences de l'édition numérique ont conduit les porteurs du projet à relancer la réflexion sur l'intertextualité, à réorienter l'analyse de certains objets. La simple nécessité de mettre des textes en ligne en harmonisant la présentation, pour permettre l'interrogation automatique, a fait réfléchir aux potentialités exploitables par le biais du numérique. L'équipe de la HdT a ainsi examiné en détail les

¹¹ Dans certains projets du Labex, comme par exemple le projet *Mythographie*, les traductions et les citations indirectes ont été balisées.

¹² <http://obvil.paris-sorbonne.fr/developpements/>

¹³ <http://obvil.lip6.fr/medite/>

¹⁴ <http://obvil-dev.paris-sorbonne.fr/phoebus/>

¹⁵ <http://obvil.lip6.fr/desert/>

¹⁶ Mainardi C., « La querelle selon l'abbé d'Aubignac : les enjeux de la *Dissertation sur la condamnation des théâtres* », Colloque « La Haine du théâtre: débats et polémiques », Colloque international, Université Paris-Sorbonne, Maison de la recherche, Paris, 23-25 octobre 2014, à paraître, <http://obvil.paris-sorbonne.fr/projets/la-haine-du-theatre>; Ganascia J.-G., Mainardi C., « Crossed Semantic Analysis of Literary Texts with DeSeRT ». *Digital Humanities*, July 2016, Krakow, Poland. 2016, <<http://dh2016.adho.org/>> <hal-01483367>

formes et les enjeux de la mise en page, tout en se questionnant sur les marques typographiques, les allusions, les références et leur positionnement¹⁷.

Le corpus de la HdT étant extrêmement citationnel, les textes ne cessent jamais de se recopier les uns les autres. Leurs auteurs utilisent les mêmes autorités, allèguent les mêmes anecdotes et reprennent les mêmes arguments ; ils reformulent le propos de l'adversaire pour le retourner. En outre, la citation n'ayant pas encore été normalisée aux XVI^e et XVII^e siècles, elle se présente sous des aspects très variés et parfois assez imprécis.

Le caractère complexe du corpus ne se limite pas à l'aspect citationnel. Les textes de la HdT peuvent être divisés en trois grandes catégories : ceux à argumentation théâtrophile, ceux à caractère théâtrophobe et les textes adoptant une stratégie de défense double. Cette dernière consiste à dénoncer des abus pour mieux préserver l'utilité intrinsèque du théâtre. Le substrat universel pour ces trois types de textes est de posséder une logique déconcertante pour le lecteur : sous le déroulement des chapitres, on découvre parfois des connexions implicites, un usage de l'ironie très répandu et des nombreuses phases à la forme négative, qui infléchissent notablement la détection des contenus.

Avec ses reprises pour réitérer ou au contraire pour retourner l'argument contre l'adversaire, le corpus de controverses théâtrales se prête bien à une étude numérique. Pour explorer un corpus qui dépasse entièrement les possibilités d'analyse d'un individu, il était ainsi essentiel de trouver le moyen de récupérer des données à l'intérieur de cette masse de textes. Afin d'obtenir des résultats exploitables et utiles, l'équipe a dû devancer le simple repérage des mots et syntagmes, en utilisant des instruments capables de saisir des concepts, des idées, des champs sémantiques.

Plusieurs outils ont été testés pour parvenir à une analyse des thématiques du corpus, et produire une comparaison automatique entre auteurs et extraits. Nous nous permettons de citer, parmi d'autres IRaMuTeQ¹⁸ avec ses AFC¹⁹, Phoebus et DeSeRT. Les résultats²⁰ obtenus sont évidemment parlants, mais n'ont pas permis de développer les recherches autour de la HdT comme initialement souhaité. De plus, l'équipe nécessitait une comparaison thématique, assez fine, de différents passages choisis.

¹⁷ Cfr. Lecercle F., Mainardi C., Thouret C., "Pour une exploration numérique des polémiques sur le théâtre", *RHLF*, n°116 / 4 dir. Didier Alexandre, *Littérature et humanités numériques*, PUF, pp. 773-790.

¹⁸ Logiciel développé par Pierre Ratinaud (LERASS – Toulouse 3), il s'agit d'une interface de R pour les analyses multidimensionnelles de textes et de questionnaires. <http://www.iramuteq.org/>

¹⁹ Analyse Factorielle des Correspondances.

²⁰ A titre illustratif, nous citons la comparaison des AFC des traités écrits dans les années 1660 et de la polémique autour de *La Lettre à d'Alembert*. Ce parallèle fait apparaître une inflexion générale du discours : si les préoccupations morales demeurent également importantes, le discours directement religieux s'estompe pour laisser une place plus grande au discours politique. Cfr. le propos de Clotilde Thouret dans la journée d'études sur la Haine de Molière (Paris, 8 décembre 2016) : *Molière, l'envers d'Augustin. Une exploration numérique des formes et des enjeux de la présence de Molière dans la controverse française sur le théâtre*, à paraître, <http://obvil.paris-sorbonne.fr/la-haine-de-moliere-actes-de-la-journee-detude>. Parmi d'autres résultats, l'utilisation de DeSeRT a mis en évidence dans le corpus des textes théâtrophobes l'idée de pureté dans la religion, tout comme la considération de l'idolâtrie comme « mère des spectacles ». Cfr. Ganascia J.-G., Mainardi C., « Crossed Semantic Analysis of Literary Texts with DeSeRT ». *Digital Humanities*, July 2016, Krakow, Poland. 2016, <<http://dh2016.adho.org/>> <hal-01483367>

3.1 Une ontologie

Afin de produire une analyse automatique entre thèmes, auteurs et extraits du corpus, les experts de la HdT se sont tournés vers une approche sémantique. Un lexique structuré a ainsi été élaboré pour saisir, dans les textes, un concept ou une association de concepts, de façon à repérer des passages exprimant des idées semblables ou traitant de questions identiques²¹.

Ce lexique est pour le moment encore en phase d'amélioration, mais dont une version « beta » est disponible sur le site de développement du projet. A terme, il sera possible d'obtenir une interprétation qui se dégage des textes eux-mêmes, tout en arrivant à découvrir des idées récurrentes, des partages de thèmes ou de mots-clés entre textes distincts, auteurs différents ou parties divergentes d'un même texte.

Afin de créer ce lexique, avec ses classes et ses sous-classes, l'équipe a réfléchi au fil d'Ariane qui associe les différents textes. Comme nous l'avons dit, le corpus HdT est extrêmement citationnel. En conséquence, dans le corpus des polémiques théâtrales les autorités invoquées sont très utiles car elles constituent un enjeu essentiel, du moins aux XVI^e et XVII^e siècles (leur importance recule nettement au XVIII^e siècle où le débat se laïcise)²². Les passages des autorités sont cités par les auteurs pour souligner leurs idées et leurs opinions sur un argument en particulier, pour mettre en évidence leur théorie ou point de vue et ils sont très souvent pourvus en commentaires de leur auteur. Grâce à ce leitmotiv des autorités, le lexique structuré de la HdT a été construit sous forme d'une ontologie. En général, cette dernière constitue un modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que des relations entre ces concepts, organisés à l'aide de relations hiérarchiques ou non hiérarchiques (transversales). L'analyse textométrique du thesaurus environnant les autorités convoquées a permis de créer un lexique structuré, hiérarchique, des termes discriminants de chaque thématique présente dans le corpus²³. L'ontologie de la HdT répertorie ainsi l'ensemble des dimensions des controverses sur le théâtre. Elle répartit les types de discours et les types d'arguments présents dans les textes.

L'ontologie en soi n'est utile à l'équipe de la HdT qu'à l'aide d'un dispositif d'annotation automatique²⁴. La combinaison des signes linguistiques et des champs

²¹ Elle a été élaborée à partir du logiciel TXM (<http://sourceforge.net/projects/txm/>) qui repère les co-occurrences de mots, permettant ainsi d'identifier les champs sémantiques dominants dans les textes et les listes de mots qui en relèvent.

²² Cfr. Lecercle F., Mainardi C., Thouret C., "Pour une exploration numérique des polémiques sur le théâtre", *RHLF*, n°116 / 4 dir. Didier Alexandre, *Littérature et humanités numériques*, PUF, pp. 773-790.

²³ Pour produire cette ontologie, et repérer des termes significatifs et non équivoques qui composent le lexique de l'ontologie, nous avons tout d'abord entrepris une analyse lexicométrique des co-occurrences autour des autorités citées.

²⁴ Pour les détails sur l'annotateur et la méthode expérimentale utilisée pour la création du lexique cfr. Mainardi C., Sellami Z., Jolivet V., "A Semantic Exploration Method Based on an Ontology of 17th Century Texts on Theatre: la Haine du Théâtre", *First International Workshop on Semantic Web for Cultural Heritage (SW4CH 2015)*, *New Trends in Databases and Information Systems*, 539, 2015, pp. 468-476, *Communications in Computer and Information Science*.

sémantiques relatifs apparaît comme un descripteur sémantique du contenu du corpus. La densité des termes balisés est très élevée : environ 1 mot sur 10 (mots vides inclus).

La visualisation des données est un sujet d'importance indéniable. A présent, sur le site de développement du Labex, il est possible d'afficher en interne l'exploitation de l'ontologie couplée avec l'annotateur automatique. L'utilisateur peut accéder aux rubriques thématiques et voir comment se répartissent les termes à l'intérieur d'un texte choisi. Il s'agit d'un travail en cours : l'outil pour des interrogations fines et complexes n'est pas encore mis en place, mais il est possible de questionner l'ontologie avec un outil web de gestion de base de données²⁵.

Les requêtes à utiliser pour interroger l'ontologie sont suffisamment complexes, longues et absolument pas intuitives pour des littéraires. De ce fait, être en lien avec un informaticien est essentiel pour la création de ces questionnements. Très souvent, à la vue des premiers résultats, qu'ils soient prometteurs ou pas, les experts ont besoin de revenir sur leurs requêtes pour affiner les recherches. C'est ainsi qu'une nouvelle phase d'échange avec un ingénieur s'ouvre, et cela souligne l'importance, dans les humanités numériques littéraires, d'un dialogue étroit entre chercheurs et experts informatiques. Pour effectuer des analyses en toute autonomie, le double profil, à la fois littéraire et informatique, est un atout de plus en plus recherché parmi les chercheurs.

Ce lexique structuré sous forme d'ontologie amène tout d'abord à localiser les topiques, à voir ensuite quels auteurs couvrent l'ensemble du domaine, en faisant référence à tous les principaux arguments de la controverse et, ultérieurement, quels auteurs ne soulèvent que des questions particulières. En outre, il permet d'associer le questionnement à un auteur particulier. Pour une première exploration de l'ontologie, on s'est proposé d'essayer de saisir les enjeux de la mobilisation de Molière dans la polémique. Faute de temps, nous nous limitons ici à donner seulement deux exemples de résultats²⁶. En premier lieu, la thématique de la « société », interne à l'ontologie, se répartit également tout au long de la controverse. En revanche, une fois croisée avec Molière, elle est surreprésentée de la fin du 17^e et au 18^e siècle. Très brièvement, cela peut s'expliquer par une confirmation de l'enjeu du comique : ce qui se joue autour de la comédie, et donc des pièces de Molière, c'est la capacité structurante du théâtre qui fait du public un miroir du peuple. Il s'ensuit la nécessité d'attaquer ou de justifier le ridicule, susceptible de défaire le lien social, et le rire, susceptible d'introduire de la conflictualité morale entre les membres de la communauté politique. En deuxième lieu, après avoir distribué les textes selon deux positions, *pour* et *contre* le théâtre, l'analyse de l'ontologie a permis d'examiner les querelles théâtrales dans leur rapport avec Molière. Il a été possible de percevoir un changement de camp avec le temps : Molière il est mobilisé d'abord par les adversaires du théâtre, et ensuite par les défenseurs. En données relatives et sur l'ensemble du corpus, le dramaturge est d'autant plus mobilisé

²⁵ SQLiteManager : www.sqlitemanager.org

²⁶ D'autres résultats peuvent se lire dans tous nos articles précédemment cités.

par les théâtrophiles que par les théâtrophobes. Ceci est confirmé par des résultats qui tiennent compte du *théâtre* de Molière²⁷.

3.2 Les limites de l'ontologie

A présent, l'équipe de la HdT ne peut utiliser l'ontologie que pour des calculs de base. Premièrement, l'ontologie a été annotée selon la division en chapitres et tout en est lié. En particulier, pour connaître la valeur d'un certain thème dans une œuvre entière, il est nécessaire d'ajouter des calculs manuels pour tous les chapitres de l'œuvre en question. Deuxièmement, comme nous l'avons vu, les textes de la HdT ont de longueurs disparates. Connaître le nombre d'occurrences d'un signe linguistique n'est pas toujours pertinent, si l'on ne tient pas compte de la taille du texte, ou de la présence qualitative du mot en question. D'où l'importance du passage de la fréquence absolue à la fréquence relative, pour pondérer les chiffres bruts en fonction de la longueur du texte. La fréquence relative d'un thème choisi devra être calculée pour tous les chapitres liés à la recherche. Troisièmement, les typologies argumentatives des textes échappent à l'examen de l'ontologie. Pour aller au-delà de la présence de signes linguistiques semblables en différents passages, et étudier la ressemblance des propos, l'analyse fait face à un blocage persistant. A présent, l'outil ne peut pas prendre en compte et gérer l'allusion à un certain concept ou à une autorité particulière. En outre, des phénomènes aussi essentiels dans le discours polémique, que l'énonciation ironique ou la négativité d'une phrase, ne sont pas plus détectés. Enfin, dernier élément mais non des moindres, la visualisation des données et l'interface occupent une place très importante. L'équipe HdT ne peut pas encore disposer d'une interface pour requêter l'ontologie, ce qui oblige les chercheurs à revenir constamment aux « dessous » du numérique, et à travailler constamment avec les données, de manière directe. Bien que cela soit indéniablement une difficulté pour l'équipe, cette manière permet de prendre en compte la pertinence - ou pas - des résultats, et d'apprendre à gérer le corpus, ainsi qu'à améliorer les possibilités de questionnement.

4. Conclusions

Le numérique transforme-t-il donc les approches des littéraires ? Oui, en tant que complément aux recherches humaines. L'analyse digitale permet de dégager des connexions entre textes, des pistes à sonder qui étaient impossibles à voir à l'œil nu et, également, de les prouver à travers des approfondissements ultérieurs. La littérature qui a recours au numérique permet de dominer des masses textuelles en contournant la

²⁷ Ces résultats ont été présentés de manière plus détaillée par C. Thouret pendant la journée d'études sur la Haine de Molière (Paris, 8 décembre 2016) : *Molière, l'envers d'Augustin. Une exploration numérique des formes et des enjeux de la présence de Molière dans la controverse française sur le théâtre*, en cours de publication <http://obvil.paris-sorbonne.fr/la-haine-de-moliere-actes-de-la-journee-detude>.

contrainte temporelle. Aucune lecture humaine, autant persévérante qu'elle puisse l'être, n'aurait pu obtenir des informations si précises dans un délai si court. L'indéniable gain de temps va de pair avec la possibilité d'amplifier les questionnements avec un simple « click ».

L'apport du numérique dans la HdT est désormais indissociable du projet et en est de plus en plus partie prenante. Cependant, derrière les résultats, il y a toujours des calculs, des formules, des pensées structurées. Les outils informatiques ne peuvent pas tout : le chercheur occupe un rôle essentiel pour l'interprétation des données et dans la vérification des opérations. Un contrôle de qualité, un examen du corpus et un retour au contexte doit être constant.

Le traitement informatique exhibe nettement des idées ou des intuitions du chercheur, et ouvre des nouvelles perspectives de recherche. La littérature numérique transforme la question littéraire, met l'accent sur des problématiques qui lui étaient traditionnellement extérieures ou marginales. La digitalisation permet de faire le lien entre le concret et le concept, la pratique et la théorie. Plutôt qu'une innovation dans son essence, les humanités numériques offrent une autre dimension d'observation et d'analyse des sources littéraires. Tout résultat inédit qui peut ressortir des recherches numériques permet d'écrire un autre rapport au monde, une nouvelle sphère d'analyse.

Chiara Mainardi, Université Sorbonne Nouvelle Paris 3