

POUR UNE EXPLORATION NUMÉRIQUE DES POLÉMIQUES SUR LE THÉÂTRE

FRANÇOIS LECERCLE, CHIARA MAINARDI ET CLOTILDE THOURET¹

Le projet « Haine du théâtre » du Labex Obvil est né du sentiment que l'heure était venue d'avoir une vue surplombante sur les polémiques théâtrales qui se sont développées dans une grande partie de l'Europe, entre le milieu du XVI^e et celui du XIX^e siècle. Elles avaient fait l'objet d'études ponctuelles², mais personne n'avait essayé d'embrasser, dans son ensemble, une production pléthorique et très répétitive. Ce ressassement n'exclut pas, du reste, de fortes différences d'un pays à l'autre et d'étranges inversions. Quand la bataille fait rage en Angleterre, au tournant des XVI^e et XVII^e siècles, elle semble presque inexistante en France ; quand la polémique se tarit en Angleterre, avec la prise de pouvoir des puritains puis la Restauration, elle prend une force et une évidence inédite en France dans les années 1660 et suivantes, où sont publiés les pamphlets les plus retentissants (Conti, Nicole, Voisin et Bossuet).

Pour dominer une telle masse de documents, sur une base européenne, il fallait réunir une équipe internationale. Celle-ci s'est fixé un double objectif. La première tâche était documentaire : il s'agissait de recenser un corpus proliférant et mal exploré, d'établir des bibliographies plus complètes et de rassembler, à usage interne, des fac-similés des textes. Le second objectif était d'analyser ces documents, dans une série de séminaires, journées d'étude, colloques et publications³.

LA CONVERSION NUMÉRIQUE DU PROJET « HAINE DU THÉÂTRE »

Les humanités numériques (HN) se sont greffées sur un projet qui était, au départ, très « classique ». Peu ou pas formés au numérique, nous avons tout d'abord considéré les HN comme un moyen de mettre à la portée des chercheurs des textes dont beaucoup sont rares et certains inconnus : il s'agissait pour nous, essentiellement, de mettre en ligne un corpus commodément accessible. L'intégration dans le Labex Obvil nous a poussés à privilégier le corpus français, parce qu'il s'intégrait mieux avec les projets des autres équipes qui, au départ, travaillaient uniquement sur la France⁴. En outre, le corpus français était moins

¹ Ce travail a bénéficié d'une aide d'État gérée par l'Agence Nationale de la Recherche dans le cadre des Investissements d'Avenir portant la référence ANR-11-IDEX-0004-02. Nous remercions Frédéric Glorieux et Vincent Jolivet, ingénieurs à l'Obvil, sans l'aide desquels il n'aurait pas été possible.

² Voir, notamment, Henry Phillips *The Theatre and its Critics in Seventeenth Century France*, Londres-New York, Oxford University Press, 1980 ; Jonas Barish, *The Antitheatrical Prejudice*, Berkeley, University of California Press, 1981 ; Laurent Thirouin, *L'Aveuglement salutaire. Le réquisitoire contre le théâtre dans la France classique*, Paris, Champion, 1997 ; Sylviane Léoni, *Le Poison et le remède. Théâtre, morale et rhétorique en France et en Italie (1694-1758)*, Oxford, Voltaire Foundation, 1998 ; Laura Levine, *Men in Women's Clothing : Anti-theatricality and Effeminization 1579-1642*, Cambridge, Cambridge University Press, 1994 ; Michael O'Connell, *The Idolatrous Eye, Iconoclasm and Theater in Early-Modern England*, New York, Oxford University Press, 2000 ; Ubaldo Floris, *Teorici, teologi e istrioni. Per e contro il teatro nella Francia del Cinque-Seicento*, Rome, Bulzoni, 2008 ; Donatella Pallotti et Paola Pugliati, dir., *La Guerra dei teatri. Le controversie sul teatro in Europa dal secolo XVI alla fine del Ancien Régime*, Pisa, ETS, 2008 ; Déborah Blocker, *Instituer un « art ». Politiques du théâtre dans la France du premier XVII^e siècle*, Paris, Champion, 2009.

³ Voir le site du projet, <http://obvil.paris-sorbonne.fr/projets/la-haine-du-theatre>.

⁴ Ce corpus polémique est susceptible de s'articuler à la Bibliothèque critique numérique de l'Obvil : si ces textes ne relèvent pas de la critique à proprement parler, ils s'appuient néanmoins sur un système de valeurs qui contribue à construire les normes et les canons de la production dramatique et plus largement littéraire, que ce

accessible que d'autres, comme le corpus anglais, largement disponible sur des sites spécialisés, comme Early English Books Online (EEBO, pour les ouvrages entre 1500 et 1700) et Eighteenth Century Collection Online (ECCO, 1701-1800).

La nécessité de répertorier des textes pour les mettre en ligne nous a amenés à des prospections qui ont sérieusement élargi la liste des traités connus. C'est ainsi que la date du premier traité théâtrophobe a été avancée d'une quarantaine d'années, puisqu'un traité théâtrophile peu connu nous a mis sur la piste d'un texte donné comme « allemand⁵ » mais qui s'est avéré être un traité publié en français par un prédicateur luthérien d'origine silésienne⁶.

Assez rapidement, nous nous sommes rendu compte qu'il ne s'agissait pas simplement de favoriser l'accès à des textes dont certains étaient introuvables sur le web ; il fallait également en faciliter l'interrogation. Puisque la bibliographie était très vaste, il fallait trouver le moyen de récupérer des données à l'intérieur d'une masse énorme de texte. Pour cela, il était plus commode de renoncer aux transcriptions diplomatiques que nous avons d'abord envisagées. Certes, il était possible de conserver les graphies d'époque en élaborant un lexique de formes normalisées, pour permettre les recherches automatiques. Mais pour un grand nombre de textes très disparates, la procédure aurait été excessivement lourde. Nous avons donc décidé d'harmoniser les graphies et de moderniser l'orthographe. À partir de là, s'ouvrait une perspective nouvelle : exploiter pleinement les possibilités de l'informatique pour explorer un corpus qui dépasse de loin les possibilités d'analyse d'un individu.

S'est ainsi opéré un véritable retournement. Au départ, l'informatique était simplement un moyen de rendre le corpus accessible en ligne. Mais rapidement elle est devenue instrument de recherche à part entière. Cela s'est fait en plusieurs étapes.

En un premier temps, l'informatique nous a paru un instrument commode pour retrouver rapidement un mot ou un syntagme dans un corpus étendu. Elle offrait le moyen de savoir en quelques secondes qui cite S. Augustin ou fait référence au « plaisir ». Ce faisant, l'informatique permet deux choses : confirmer ce que l'on pressent (on a une intuition qu'on peut mettre à l'épreuve) ou chercher, sur un corpus plus large, ce qu'on a déjà observé dans un fragment. Dans les deux cas, il faut savoir ce qu'on veut trouver : il faut avoir les clefs de sa recherche, c'est-à-dire connaître les syntagmes que l'on a déjà repérés et que l'on voudrait retrouver dans un plus large ensemble de textes.

Ce n'est pas d'une mince utilité, parce qu'on peut interroger des corpus considérables, pour y retrouver une phrase, une citation ou un mot clef. On peut aussi repérer – à travers un mot significatif comme « plaisir » – la présence d'une idée ou d'un argument, voire évaluer son importance quantitative. Dès le début de la mise en ligne des textes, on a pu interroger le corpus pour connaître immédiatement le nombre d'occurrences d'un terme dans chaque texte et localiser les endroits où ce terme apparaissait le plus. Pour le mot « plaisir », cela donnait : Voisin (116), Conti (42), Nicole, éd. de 1675 (24), Guillot-Gorju (18)⁷.

soit en formulant des interdictions ou en stigmatisant des auteurs particuliers. Le meilleur exemple est Molière qui, à partir des années 1660 et jusqu'à la fin du XVIII^e siècle, devient dans les textes français (et même dans quelques textes étrangers) le point de mire des théâtrophobes. En désignant obsessionnellement la même cible, ceux-ci confirment paradoxalement son rôle de parangon, dans le moment même où ils s'emploient à le détruire.

⁵ Mlle de Beaulieu, *La Première Atteinte contre ceux qui accusent les Comédies, par une Demoiselle Française*, Paris, Richer, 1603, f. 2v.

⁶ Daniel Tilenus, *Traité des jeux comiques et tragiques, contenant instruction et résolution de la Question: assavoir si esbats et passe temps sont permis aux chrestiens*, s.l. [=Sedan], Salesse, 1600.

⁷ Voir 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 ; Prince de Conti, *Traité de la comédie et des spectacles*, Paris, Billaine, 1666 ; Pierre Nicole, « De la Comédie », in *Essais de morale*, Paris, Desprez, 1675 ; Guillot-Gorju, *Apologie de Guillot-Gorju. Adressée à tous les beaux Esprits*. Paris, Blageart, 1634.

On peut ainsi dominer des masses textuelles et obtenir, sur la présence des termes, des informations qu'aucune lecture patiente n'aurait pu obtenir. Il ne s'agit pas d'un simple gain de temps car on peut, grâce à l'informatique, faire des choses dont, sans elle, on ne ferait qu'une infime partie, avec en outre le risque de laisser échapper des occurrences. Le gain de temps permet de multiplier les sondages, c'est-à-dire de mener des recherches qu'autrement on ne ferait jamais.

Une nouvelle phase a commencé avec l'utilisation de logiciels qui permettent d'évaluer la fréquence relative des termes. Le numérique ne permet pas seulement de repérer et compter les occurrences mais d'apprécier les fréquences relatives. Il ne suffit pas en effet de dénombrer les occurrences du mot « plaisir », il faut pondérer les chiffres bruts en fonction de la longueur du texte pour évaluer une présence relative qui, en général, bouleverse totalement la hiérarchie. C'est éloquent pour le terme « plaisir » (**Figure 1**).

Bien sûr, chercher un mot ne suffit pas, il faut aussi chercher les synonymes possibles (délectation, délice, agrément, contentement, ravissement, récréation, etc.). Mais la recherche de ce seul terme est déjà pleine d'enseignements. Car les écarts peuvent être significatifs : il y a un abîme entre Guillot-Gorju, qui fait largement place au plaisir, et Saint-Evremond qui l'ignore. L'évaluation chiffrée attire ainsi l'attention sur le premier texte, qui est une apologie du théâtre par un farceur de l'Hôtel de Bourgogne : la surreprésentation du terme « plaisir » signale en réalité le caractère exceptionnel d'un texte aux accents assez hétérodoxes. Mais ces comptages peuvent aussi être trompeurs : le comédien La Porte, qui arrive au bas de l'échelle avec le plus faible nombre d'occurrences et le plus faible pourcentage, a une utilisation originale du plaisir théâtral, auquel il trouve deux vertus : assurer la cohésion d'une société et faire plus facilement passer des leçons rudes mais profitables⁸.

Un décompte particulièrement utile, dans le corpus des polémiques théâtrales, est celui des autorités invoquées, et en particulier des autorités patristiques, parce qu'elles constituent un enjeu essentiel, du moins aux XVI^e et XVII^e siècles, car leur importance recule nettement au XVIII^e siècle, où le débat se laïcise. Et il est important d'observer les écarts dans les références. D'après les premiers résultats (**Figure 2**), Augustin et Tertullien reculent, de la première à la deuxième moitié du XVII^e siècle, tandis que Thomas d'Aquin augmente fortement. Il serait prématuré d'en tirer des conclusions, car il faut attendre de pouvoir interroger un bien plus grand nombre de textes, mais on peut d'ores et déjà percevoir l'intérêt d'un tel graphique. D'abord, il fait apparaître la promotion des autorités païennes au XVIII^e siècle : la proportion des références à Aristote, Cicéron et Platon croît. Ensuite, il permet de mesurer la disparité entre l'importance de la condamnation platonicienne du théâtre, qui nourrit bien des démonstrations, et son peu de présence explicite dans les textes, qui tient au fait que les Pères de l'Église ne le citent pas ou presque. Enfin, il donne des bases objectives pour choisir les objets de nouvelles recherches qui porteraient sur l'usage polémique des autorités.

Une troisième phase s'est ouverte quand nous avons testé les logiciels de comparaison de textes. On dépasse alors le stade de la recherche fondée sur des intuitions à vérifier et on peut découvrir des phénomènes qui n'étaient pas frappants ou pas visibles, même pour un lecteur expert. Certes, une comparaison patiente des deux états du *Traité de la comédie* de Pierre Nicole (1667 et 1675)⁹ peut donner l'intuition que, de l'un à l'autre, il y a un affaiblissement du vocabulaire religieux, mais une comparaison des deux versions à l'aide du logiciel Médite¹⁰ le montre nettement, sans passer par une lente et scrupuleuse confrontation

⁸ Mathieu Lefebvre de La Porte, *Prologue de La Porte, comédien, Prononcé à Bourges, le 9 septembre 1607*, voir le texte de ce ms, http://obvil-dev.paris-sorbonne.fr/corpus/haine-theatre/la-porte_prologue_1607/.

⁹ Pierre Nicole, « L'Hérésie imaginaire », in *Les Visionnaires*, Liège, Beyers, 1667 ; Pierre Nicole, « De la comédie », in *Essais de Morale*, t. III, Paris, Desprez, 1675.

¹⁰ Issu d'une collaboration entre des spécialistes de génétique textuelle de l'ITEM (Institut des Textes et Manuscrits Modernes) et des spécialistes d'intelligence artificielle de l'équipe ACASA du LIP6 (Laboratoire

(Figure 3). Un graphique des mots supprimés dans la deuxième version est un bon indice de cette transformation (Figure 4).

Reste bien entendu à s'interroger sur les raisons d'un tel changement. Une hypothèse possible est que Nicole ait approfondi l'idée que le théâtre suscite des passions néfastes chez le spectateur, en la développant dans le sens d'une psycho-physiologie en partie héritée de la pensée cartésienne¹¹. Il y avait là un gain polémique essentiel, car l'effet du spectacle, obéissant d'abord à des lois physiques, devient un processus naturel et nécessaire sur lequel le spectateur n'a aucune prise puisqu'il n'en est pas conscient.

Si l'on n'avait affaire qu'à un corpus limité, ne dépassant pas quelques dizaines de textes, il n'y aurait assurément pas besoin de faire appel à des logiciels. Mais avec un corpus immense et qui augmente au fil des découvertes, il est nécessaire de recourir à l'informatique. Si celle-ci est particulièrement adaptée à l'analyse des polémiques théâtrales, ce n'est pas seulement à cause de l'ampleur du corpus, mais à cause de son caractère très répétitif : les polémistes invoquent sans cesse les mêmes autorités, citent les mêmes anecdotes et reprennent les mêmes arguments. Ils reformulent le propos de l'adversaire pour le retourner et allèguent les mêmes autorités pour les tirer à eux. Il faut donc franchir un pas de plus avec des logiciels qui dépassent le simple repérage des mots, syntagmes ou fragments qui sont repris quasiment à l'identique d'un texte à l'autre. On entre alors dans une troisième phase, bien plus exploratoire, en utilisant des instruments capables de saisir des idées et non pas simplement des mots. Un saut qualitatif s'opère quand on passe de la reprise textuelle à la reformulation, avec des logiciels qui permettent de repérer la reprise modulée et l'allusion.

L'instrument le plus utile est le logiciel DeSeRT¹², qui interroge un corpus pour repérer des blocs sémantiquement proches. Il ne saisit pas des mots (des suites de caractères) mais leurs synonymes. Alors que Médite identifie des reprises plus ou moins identiques, DeSeRT fait apparaître des proximités sémantiques. Il ne repère pas une citation ou un paragraphe largement repris mais des passages qui traitent du même sujet. Il permet ainsi d'observer bien plus largement les interférences entre un texte et un autre et d'identifier les phénomènes de plagiat.

Le projet a avancé encore avec la constitution d'une ontologie, c'est-à-dire d'un lexique structuré qui permet de 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 identiques¹³. Il s'agit d'établir les liens logiques entre les concepts (et les mots qui les expriment) pour construire des univers sémantiques. En liant les signes linguistiques à des champs sémantiques, on élabore un descripteur sémantique du contenu des textes qui doit permettre, à terme, d'analyser sémantiquement les textes sans que les a priori d'un lecteur

d'informatique de Paris 6 - UPMC, <http://www-poleia.lip6.fr/~ganascia/Médite/>), ce logiciel permet de repérer les différences entre des blocs de deux textes très semblables : on découvre ainsi assez précisément les ajouts, les modifications et les suppressions apportées. Un autre logiciel, Phœbus, issu d'une collaboration entre des spécialistes d'intelligence artificielle de l'équipe ACASA et des chercheurs en humanités numériques du Labex OBVIL (l'Observatoire de la vie littéraire, voir <http://arxiv.org/abs/1404.2997v1>), permet de repérer les récurrences entre deux textes, de les détecter automatiquement et d'explorer des emprunts, des réseaux de réutilisation textuelle.

¹¹ Voir notamment Pierre Nicole, « De la comédie », *op. cit.*, chap. V ; et l'analyse d'Erec Koch dans *The Aesthetic Body. Passion, Sensibility and Corporeality in Seventeenth-Century France*, Newark, University of Delaware Press, 2008, pp. 123-125, et dans « La contagion des passions, de Descartes à Malebranche », *Littératures classiques*, n° 68, *Les Émotions publiques et leurs langages à l'âge classique*, dir. Hélène Merlin-Kajman, pp. 177-188.

¹² DeSeRT a été créé par J.-G. Ganascia, professeur d'informatique à l'UPMC (Paris). Ce logiciel est en phase de développement mais une version en ligne se trouve à l'adresse <http://obvil-dev.paris-sorbonne.fr/desert/>.

¹³ 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.

biaisent la perception. Pour créer cette ontologie, nous avons tout d'abord entrepris une analyse lexicométrique des co-occurrences autour du fil rouge du corpus que sont les autorités citées – qu'il faut prendre au sens large d'*auctoritates* : on prend en compte les pouvoirs institués aussi bien que les doctes, Auguste aussi bien qu'Augustin. Les autorités fonctionnent en effet comme des points d'ancrage des démonstrations, autour desquels se distribue le discours polémique¹⁴. On repère ainsi des termes significatifs et non équivoques, qui composent le lexique de l'ontologie. Celui-ci est alors structuré dans différents champs sémantiques (ou classes), dans lesquelles les lemmes sont répartis selon des relations hiérarchiques ou transversales. Ainsi « pouvoirs institués » se décline en « pouvoir religieux » et « pouvoir profane », et « pouvoir religieux » en « cardinal », « évêque », « pape », etc. Grâce à l'élaboration de ces différentes classes (« autorités », « pouvoirs institués », « spectacle », « droit », « économie », « émotion », « femme », etc.), l'ontologie répertorie l'ensemble des dimensions des controverses sur le théâtre ; elle constitue un outil d'exploration sémantique adapté au corpus, qui peut ainsi être interrogé selon les différentes perspectives de recherche choisies par l'utilisateur.

C'est en utilisant l'ontologie que l'on peut constater, par exemple, le changement du rôle dévolu aux femmes dans l'argumentation. L'association entre la femme et le théâtre ne devient flagrante qu'à partir du milieu du XVII^e siècle. Avant, les femmes ne sont guère mentionnées qu'en rapport avec le déguisement des hommes¹⁵. C'est à partir du *Traité* de Philippe Vincent¹⁶ que le rôle des femmes au théâtre prend de l'importance dans l'argumentation, ce que confirment les traités postérieurs, de D'Aubignac¹⁷, Conti¹⁸, Lelevel¹⁹, Lemarcant²⁰, Nicole²¹, Pégurier²² et Voisin²³, parus entre 1666 et 1694. La thématique féminine est ainsi examinée selon différentes facettes : elle n'est plus seulement liée au recours au travesti pour les rôles féminins, mais touche à la participation des femmes au théâtre, en tant qu'actrices ou que public, aux dommages que la représentation causerait inévitablement à leur innocence et à leur pudeur. Au XVIII^e siècle, la question s'élargit encore, avec un nouvel éventail de lemmes : non plus seulement « fille », « femme » ou « comédienne » mais aussi « dame », « demoiselle », « maîtresse » et « actrice », notamment dans les textes de D'Alembert²⁴, Dancourt²⁵ ou Rousseau²⁶.

¹⁴ Nous avons calculé pour chaque texte les noms propres et éliminé ceux qui n'étaient pas pertinents (c'est-à-dire qui ne relevaient pas des autorités). Nous avons ensuite étudié les collocations de chaque occurrence en contexte, avec un seuil de cinq unités (à gauche et à droite du mot pivot), en éliminant les mots équivoques ou non pertinents. Nous avons ensuite étudié les collocations de chaque occurrence en contexte, avec un seuil de cinq unités (à gauche et à droite du mot pivot), en éliminant les mots équivoques ou non pertinents. Pour cette partie, la collaboration avec Frédéric Glorieux (Labex Obvil) a été essentielle. La mise en forme de l'ontologie a été faite grâce à l'aide de Vincent Jolivet (Labex Obvil) et de Zied Sellami (équipe ACASA). L'ontologie HdT [Haine du théâtre] est couplée avec un annotateur qui enrichit automatiquement les fichiers (encodés en XML-TEI) avec des balises pour marquer toutes les formes, leur lemme et également le champ sémantique relatif. Voir Chiara Mainardi, Zied Sellami et Vincent Jolivet, « A Semantic Exploration Method Based on an Ontology of 17th Century Texts on Theatre: la Haine du Théâtre », *First International Workshop on the Semantic Web for Cultural Heritage* (SW4CH 2015), Sept. 2015, Poitiers, France. *New Trends in Databases and Information Systems*, 539, pp. 468-476, Communications in Computer and Information Science.

¹⁵ Voir, par exemple, Daniel Tilenus, *Op. cit.*, 1600.

¹⁶ Philippe Vincent, *Traité des Théâtres*, 1647.

¹⁷ Abbé d'Aubignac, *Dissertation sur la condamnation des théâtres*, Paris, Pépingué, 1666.

¹⁸ Prince de Conti, *Op. cit.*, 1666.

¹⁹ Henri Lelevel, *Réponse à la lettre du théologien, défenseur de la comédie*, Paris, Girard, 1694.

²⁰ Abbé Le Marcant, *La conduite du vrai chrétien*, Paris, Couterot, 1694.

²¹ Pierre Nicole, *Op. cit.*, 1667.

²² Laurent Pégurier, *Réfutation des Sentiments relâchés d'un nouveau théologien touchant la comédie*, Paris, Coignard, 1694.

²³ Joseph Voisin, *Op. cit.*, 1671.

²⁴ Jean Le Rond d'Alembert, *Lettre de M. D'alembert à M. J. J. Rousseau*, Amsterdam, Chatelain, 1759.

CE QUE LE NUMÉRIQUE APPORTE À L'ANALYSE DES POLÉMIQUES THÉÂTRALES

L'utilité du numérique ne tient pas seulement à l'ampleur immaîtrisable du corpus. La simple nécessité de mettre des textes en ligne en harmonisant la présentation pour permettre l'interrogation automatique s'est déjà avérée utile, puisqu'elle nous a conduits à réfléchir aux formes et aux enjeux de la mise en page, auxquels nous n'aurions peut-être pas prêté une attention suffisante. Ainsi des notes marginales : à quel endroit précis ou à quel fragment textuel renvoient-elles exactement ? Où les placer ? À quoi correspondent-elles ? Indexent-elles le schéma argumentatif ? Attirent-elles l'attention sur un sujet ? Donnent-elles la référence d'une allusion, d'une citation, d'une autorité alléguée ? Ainsi des citations, dont la présentation varie d'un texte à l'autre : beaucoup ne sont pas marquées typographiquement, ou bien ont des marques variables (l'italique) et des frontières incertaines.

Les exigences de l'édition numérique et les possibilités qu'elle ouvre conduisent donc à renouveler, ou du moins à réorienter, l'analyse de l'usage polémique de la citation, voire à relancer la réflexion sur l'intertextualité. En effet, il ne suffit pas de suppléer des guillemets absents. Même si on normalise les marques typographiques, les frontières d'une citation restent souvent floues et le texte instable, car la citation change de forme quand elle est empruntée à une édition ou une traduction différente et elle est souvent faite de mémoire ou sans aucun souci de littéralité. Et c'est là que les logiciels interviennent. Ils ne servent pas seulement à trouver les citations faites par un auteur et reprise par un autre mais ils identifient les reprises, avouées ou non, des propos d'un prédécesseur.

Ainsi, le Père Voisin reprend explicitement Pierre Nicole mais il se livre aussi à une reformulation sporadique de ses propos, ce que les logiciels Phœbus et Médite font apparaître très clairement. La *Défense du traité de Mgr le Prince de Conti* présente de nombreuses reprises ponctuelles du texte de Nicole qui ne sont accompagnées d'aucune référence. Par exemple, pour réfuter un argument de la *Dissertation* de d'Aubignac sur la présence des femmes au théâtre²⁷, Voisin reste allusif en évoquant des « gens d'esprits » alors qu'il décalque directement le traité du janséniste :

Mais en attendant qu'on s'acquitte envers lui, de ce devoir, voyons si les gens d'esprit, et de piété approuvent que les femmes se plaisent à la Comédie. Comme la passion de l'amour, disent-ils, est la plus forte impression que le péché ait fait dans nos âmes, ainsi qu'il paraît assez par les désordres horribles qu'elle produit dans le monde ; il n'y a rien de plus dangereux que de l'exciter, de la nourrir, et de détruire ce qui la retient.²⁸

Ces réutilisations subreptices, mises au jour par le logiciel de confrontation, précisent au moins deux aspects de la stratégie polémique de Voisin : d'abord, accumuler les arguments, en dépit de la répétition, pour faire nombre et donner une impression de masse à laquelle on ne peut pas répondre ; ensuite, même si la Paix de l'Eglise a été signée en 1669, se protéger d'éventuelles accusations d'accointances avec le jansénisme qui avait été condamné par le Pape Innocent X en 1667.

Des logiciels sémantiquement plus puissants permettent de faire apparaître les topiques. Il est essentiel, pour tirer des conclusions un peu générales sur les grandes tendances de la polémique, de repérer les arguments et les thématiques. Un logiciel de visualisation,

²⁵ Louis Hurtaut Dancourt, *Dancourt, arlequin de Berlin*, à M. J.-J. Rousseau, *citoyen de Genève*, Berlin et Amsterdam, Schneider, 1759.

²⁶ Jean-Jacques Rousseau, *Lettre de J. J. Rousseau à M. D'Alembert*, Leigh, 1758.

²⁷ Abbé d'Aubignac, *Op. cit.*, pp. 243-244.

²⁸ Joseph Voisin, *Op. cit.*, p. 346 ; le texte se trouve dans Nicole, *Op. cit.* éd. 1667, p. 456.

IRaMuTeQ²⁹, peut permettre un premier repérage grâce à une distribution graphique des thématiques qui tient compte de leur fréquence³⁰. En l'utilisant sur les traités écrits dans les années 1660 et sur la polémique autour de *La Lettre à d'Alembert*, on peut fait apparaître une inflexion générale du discours, où le religieux s'estompe pour laisser une place plus grande au politique. Mais pour le moment, les résultats ne sont pas totalement utilisables car le logiciel R³¹ qui contribue à générer les données visualisées ensuite grâce à IraMuTeQ, est un outil puissant mais très complexe et qui ne permet pas de contrôler les thésaurus utilisés ni le processus de génération des images.

En revanche, avec le logiciel DeSeRT, c'est l'utilisateur qui définit le thésaurus³², si bien que le logiciel sort des morceaux de textes « denses » en la thématique qu'on a définie, voire des morceaux qui croisent plusieurs thématiques – par exemple les passages qui associent « femme » et « adultère ». En permettant de sélectionner les textes qui font l'objet de la recherche, et donc de restreindre à une ou plusieurs querelles particulières, DeSeRT est un logiciel très prometteur pour le corpus des polémiques sur le théâtre. Car il effectue, en les accélérant et en en démultipliant les possibilités, les procédures de recherche appelées par les textes de controverses : entrée dans un corpus par une idée ou une intuition, passage rapide d'un texte à l'autre selon une logique thématique, etc.

Ces logiciels présenteront, à terme, un autre avantage. Si l'on pouvait interroger des corpus extérieurs, comme celui de Gallica – ou celui de Google Books, quand il sera suffisamment interrogeable – on pourrait élargir le corpus. Si l'on pouvait étendre les interrogations à un corpus non spécifiquement préparé comme celui de « Haine du théâtre », on pourrait identifier, à partir des caractéristiques observées dans les traités jusqu'ici recensés, des textes relevant partiellement de la polémique. Le théâtre a suscité des débats qui ont eu de larges échos dans le corps social et, du coup, la question s'est diffusée dans toutes sortes d'écrits où l'on ne s'attendrait pas, a priori, à la rencontrer. Actuellement, pour identifier les textes qui ne traitent que ponctuellement des querelles du théâtre, on ne peut compter que sur les rencontres de hasard. Ainsi, il faut avoir des raisons totalement étrangères au projet « Haine du théâtre » pour tomber, dans un récit de voyage, sur un chapitre comparant la France et l'Italie du point de vue de la tolérance que les autorités manifestent envers les spectacles³³. Si l'on parvenait à affiner suffisamment nos critères, la recherche automatique devrait permettre d'identifier les textes de tous ordres qui consacrent quelques pages aux polémiques théâtrales. Bien sûr, ces « coups de filet » doivent être suivis d'un tri expert : avant d'ajouter un texte au corpus des polémiques, il faut qu'un lecteur averti confirme que la machine a bien repéré un passage qui entre effectivement dans le débat.

La création d'une ontologie permet, elle, de localiser les topiques. C'est particulièrement utile pour apprécier le poids relatif d'arguments apparemment mineurs, comme les arguments économiques et politiques qui, en surface, sont beaucoup moins affichés que les arguments religieux ou moraux mais qui s'avèrent, à l'examen, souvent névralgiques. En utilisant l'ontologie, on voit immédiatement quels auteurs couvrent l'ensemble du domaine en faisant un sort à tous les principaux arguments et lesquels ne soulèvent que certaines questions particulières. Chez Voisin, Conti, Nicole et d'Aubignac, on rencontre une concentration élevée de termes relevant de chacune des principales

²⁹ 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/>.

³⁰ Les graphiques sont générés par l'AFC (Analyse Factorielle des Correspondances).

³¹ R est un logiciel libre de traitement des données et d'analyse statistiques.

³² Le thésaurus est créé selon les besoins de la recherche, notamment en précisant les topiques à rechercher.

³³ Voir Abbé Gabriel-François Coyer, *Voyages d'Italie et de Hollande*, Paris, Veuve Duchesne, 1775, t. II, chap. XI, « Des spectacles en Italie », p. 206-208.

thématiques³⁴. En revanche, certains auteurs comme Vincent, Lelevel, Guillot-Gorju ou Pégurier présentent des résultats élevés seulement pour la thématique économique (**Figure 7**).

Il est bon aussi de pouvoir préciser la répartition de certains arguments dans le temps ou en fonction des accointances idéologiques des polémistes. Qui évoque le travesti et à quelle période ? L'argument tient-il seulement au fait que la condamnation lancée par le Deutéronome (21, 11-12) est le seul passage biblique que l'on puisse invoquer contre le théâtre ? Qui invoque la présence des femmes sur scène, et dans quel sens ?

L'ontologie a, en outre, permis de faire une première évaluation chiffrée de l'intensité de la condamnation selon les différents textes, en mesurant la proportion de lemmes relevant de l'immoralité ou de la réprobation. Il faut assurément prendre quelques précautions pour interpréter les données puisque, parmi les textes en tête de classement, on trouve, outre les traités de Nicole et Conti, le traité de d'Aubignac, qui est une défense du théâtre. Mais, à la réflexion, il n'y a rien d'étonnant à cela, puisque l'abbé met en œuvre une stratégie de défense double, qui consiste à dénoncer des abus et à donner des gages aux adversaires pour mieux préserver l'utilité intrinsèque du théâtre et plaider pour une réforme des pratiques (dont il se présente comme le maître d'œuvre potentiel). On peut alors imaginer que la quantification de l'intensité de la condamnation, pour autant que les données soient soumises au contrôle d'un expert, permettra d'ordonner l'ensemble du corpus sur une ligne continue qui va de l'apologie sans nuance à la condamnation totale en passant par toute une série de positions réformatrices. En adoptant ainsi un point de vue surplombant, on pourra mieux saisir l'évolution du discours réformateur, pour élaborer de nouveaux questionnements et concevoir de nouvelles enquêtes.

L'objectif est, à terme, de pouvoir lancer des interrogations en variant le corpus, pour analyser la répartition des arguments dans le temps, en interrogeant les textes par tranches chronologiques. Quand on disposera d'un corpus conséquent dans diverses langues, on pourra spéculer sur les différences entre les aires culturelles. On pourra également affiner l'analyse « sociologique » en observant d'éventuels écarts en fonction de l'identité des intervenants. C'est pourquoi nous avons commencé à élaborer, pour chaque texte, une fiche précisant autant que possible l'identité du texte et de l'auteur au moment où il a écrit ou publié : le type de texte (sermon, traité de morale, traité théologique, pamphlet *ad hominem*, récit de voyage, etc.), l'affiliation religieuse de l'auteur, sa nationalité voire sa province d'origine, son appartenance professionnelle, son « camp » (est-il adversaire, défenseur ou réformateur du théâtre ?) et le lien du texte à une « querelle » particulière, puisque, si certains textes sont isolés, d'autres participent d'un tir groupé. L'affaire Caffaro, en 1694, a suscité près d'une dizaine de traités et pamphlets ; la *Lettre à d'Alembert* de Rousseau (1758) en a provoqué une demi-douzaine.

Filtrer le corpus permettra de lancer des interrogations ciblées sur tel ou tel sous-ensemble : interroger une tranche chronologique, les auteurs laïcs, les sermons, les textes publiés par des jésuites, etc. Dès lors, il deviendra possible de mener plusieurs types de recherche. Tout d'abord, explorer les productions d'un « groupe », comme les jésuites. On sait que ceux-ci ont donné, dans leur programme éducatif, une large place à l'image et au théâtre : la *ratio studiorum* inscrit les représentations à son programme, avec certaines réserves, comme l'exclusion des femmes et même des rôles féminins. On en a souvent conclu qu'ils étaient forcément parmi les défenseurs du théâtre. Si c'est vrai dans l'ensemble, il y a de nombreux contre-exemples : on compte nombre de jésuites parmi les adversaires du théâtre et certains sont virulents³⁵. Il serait intéressant d'explorer systématiquement les textes écrits

³⁴ Les différentes thématiques qui se sont dégagées lors de l'élaboration de l'ontologie sont les suivantes : économie, émotion, femme, musique, nationalité, religion, société et théâtre.

³⁵ La même année 1607, un jésuite défend à Lyon, contre un calviniste, une représentation de tragédie sainte (André de Gaule, *Conviction véritable du récit fabuleux divulgué touchant la Représentation exhibée en face de toute la ville de Lyon, au collège de la Compagnie de Jésus, le 7. d'Avoust, de la présente année 1607*, Lyon,

par les jésuites, non seulement pour repérer l'utilisation qu'ils font de leurs prédécesseurs, mais surtout pour voir s'il y a des spécificités dans les arguments développés, les anecdotes reprises et les autorités invoquées.

Autre enquête : étudier la distribution d'un argument, d'une anecdote ou d'une autorité non pas simplement dans la totalité du corpus mais selon des groupes d'auteurs ou des types de texte. Qui cite telle autorité : S. Augustin est-il plus invoqué par les jésuites, les calvinistes ou les jansénistes ? Qui reprend la diatribe contre les comédiennes dont l'exhibition sur scène est un poison qui, par les yeux, pénètre au fond de l'âme ? Qui raconte l'anecdote d'Alexandre de Phères, tyran sanguinaire que les tragédies d'Euripide forcent à verser des larmes sur les malheurs d'Hécube ou d'Andromaque³⁶ ? Ces arguments sont-ils particulièrement récurrents dans tel groupe, tel type de texte ou telle tranche chronologique ? En multipliant les interrogations de ce genre, on pourrait voir si certaines périodes ou certains groupes – confessionnels, professionnels, culturels – ont une propension significative à un type d'argument particulier.

Indépendamment de ces enquêtes, l'informatique ouvre des possibilités nouvelles pour l'édition. On peut imaginer un enrichissement des textes du corpus, pour constituer une édition hypertextuelle, en créant des liens avec les passages similaires dans d'autres textes du corpus. Au lieu de simplement signaler, comme une note d'édition critique, un texte comparable, il s'agirait de donner accès, par un simple clic, à la confrontation des textes similaires, dont les reprises textuelles ou les reprises synonymiques seraient surlignées en couleurs différentes³⁷. Dans des textes qui se citent et se reprennent à l'infini, c'est particulièrement utile. Le numérique n'offre donc pas seulement des possibilités de recherche nouvelles, elle offre une lisibilité informée que l'édition classique ne permet pas : le lecteur peut rebondir de texte en texte. A terme, cela devrait permettre de « naviguer » dans le corpus, comme on navigue sur le web. C'est d'autant plus indiqué pour le corpus polémique que les polémistes eux-mêmes ont en quelque sorte « navigué », en compilant les arguments de leur prédécesseurs, soit pour faire de leur traité la réédition déguisée d'un prédécesseur, soit pour constituer de véritables anthologies qui, comme le traité du Père Lebrun, cherchent à rassembler le plus possible d'arguments, d'autorités, de témoignages et d'exemples, dans des volumes qui enflent de réédition en réédition³⁸. En quelque sorte, la mobilité du web, qui permet de « sauter » de texte en texte et d'établir des connexions à l'infini à partir de liens entre des blocs apparentés, n'est pas sans homologie avec la façon dont beaucoup de polémistes – et pas seulement dans le domaine théâtral – procédaient, en reprenant, citant, déformant et invoquant leurs prédécesseurs.

Cloquemin, 1607) tandis que, à Bourges, un autre jésuite menace d'excommunication ses ouailles qui iraient à la comédie, s'attirant une violente réplique du comédien La Porte (*op. cit.*). Les jésuites français hostiles au théâtre sont, entre autres, François Guilleré (*Retraite pour les dames*, Paris, Michalet, 1684, 3e partie, Entretien X), Louis Bourdaloue (« Sermon pour le troisième dimanche après Pâques. Sur les Divertissemens du monde », in *Sermons du Pere Bourdaloue, de la Compagnie de Jesus*, t. 2, rééd. Lyon, Bruys et Ponthus, 1758. p. 48-88), Vincent Houdry (*Sermons sur tous les sujets de la morale chrétienne*, 5^e partie, t. II, Paris, Boudot, 1700, sermon 4), Jean Croiset (*Réflexions chrétiennes sur divers sujets de morale*, rééd., Paris, Sorin, 1777, vol. 1, « Des spectacles », p. 75-92).

³⁶ Il est assez délicat de repérer une anecdote. Elles sont en général liées à des noms propres ou à une conjonction de lemmes – mais il arrive assez souvent que le héros ne soit pas identifié (Alexandre de Phères devient un « tyran de l'antiquité »), voire que le fait soit attribué à un autre (Denys de Syracuse est substitué à Alexandre de Phères). Au stade où nous en sommes, la récolte des anecdotes, si on ne les a pas indexées en mettant les textes en ligne, est moins sûre que celle des autorités, pour lesquelles on peut associer plusieurs lemmes : à « S. Augustin » et « Augustin », associer « l'évêque d'Hippone » et la liste de ses principales œuvres.

³⁷ En utilisant le logiciel Médite. Voir *supra*, la figure 1.

³⁸ Pierre Le Brun oratorien, *Discours sur la Comédie, où l'on voit la réponse au Théologien qui la deffend, avec l'Histoire du Théâtre et les sentiments des Docteurs de l'Église depuis le premier siecle jusqu'à present*, Paris, Guérin et Boudot, 1694. Le traité connaît, au fil du XVIII^e siècle, plusieurs rééditions de plus en plus augmentées.

L'informatique a trois grandes vertus. Elle permet de mieux saisir de grands ensembles de texte, d'étayer des impressions sur des éléments de preuve factuels et de faire apparaître des éléments invisibles à la lecture cursive : par exemple, mettre en évidence des différences significatives dans la distribution des éléments dans le temps, selon les types de texte, les auteurs et groupes d'auteurs. Il n'y a là rien que ne puisse reconnaître un expert, à ceci près que la machine travaille à bien plus grande échelle et sans risque d'omission. Sur un corpus énorme, le lecteur le plus docte peut difficilement tout mémoriser et il laisse forcément échapper des reprises ou des échos.

Mais la machine ne peut pas tout. Elle ne saurait saisir nombre d'éléments essentiels pour l'analyse experte. Par exemple, à ce stade, les logiciels que nous utilisons ne font pas la différence entre affirmation et négation : ils ne permettent pas de distinguer adversaires et défenseurs du théâtre – impossible, pour eux, de reconnaître que l'abbé Voisin, quand il reprend l'abbé d'Aubignac, ne le fait pas pour s'emparer des plumes du paon mais pour retourner complètement son argumentation. Car les stratégies argumentatives échappent à l'appréhension de la machine : quand tout oppose les deux abbés, les logiciels ne les distinguent pas. Mais, ce faisant, ils posent une question intéressante : pourquoi les topiques sont-elles si proches ? Pourquoi le défenseur use-t-il d'un vocabulaire aussi négatif que l'adversaire ?

La machine ignore aussi des phénomènes aussi essentiels, dans le discours polémique, que l'énonciation ironique : elle ne peut – jusqu'à présent – repérer les réutilisations ironiques et les reprises en sens inverse. Que l'on reprenne pour réitérer ou au contraire pour retourner l'argument contre l'adversaire, la machine ne fait pas la différence.

D'où la nécessité d'un examen expert, au cas par cas, pour vérifier qu'il y a bien un rapport sémantique fort, et non pas un rapprochement fortuit de mots similaires. Il faut l'œil d'un spécialiste pour comprendre si on a affaire à une citation, une allusion, une reprise explicite ou dissimulée. Il faut un expert pour identifier les visées de la reprise : reprendre l'idée à son compte, la déformer ou la détourner – que ce soit pour l'utiliser à d'autres fins ou la ridiculiser ou la réfuter. Bref, il faut qu'un lecteur très informé « contrôle » les rapprochements effectués par le logiciel mais, par lui-même, il ne pourrait pas faire ces rapprochements sur une si grande échelle. Sur des corpus si vastes, l'expert est impuissant sans la machine, mais la machine est aveugle sans l'expert et elle produit des résultats que seul l'expert peut interpréter.