

Université de Montréal

Interprétation littéraire numérique

Modélisation algorithmique, cybersémiotique et herméneutique

Par

Yann Audin

Département de littérature et de langue du monde

Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade

de maître en littérature comparée

Août 2022

©Yann Audin 2022

Université de Montréal

Département de littératures et de langues du monde, Faculté des arts et des
sciences

Ce mémoire intitulé

Interprétation littéraire numérique

Modélisation algorithmique, cybersémiotique et herméneutique

Présenté par

Yann Audin

A été évalué par un jury composé des personnes suivantes

Joyce Boro

Présidente-rapporteuse

Michael E. Sinatra

Directeur de recherche

Marcello Vitali-Rosati

Codirecteur

Dominic Forest

Membre du jury

Résumé

Dans ce mémoire, nous explorons les possibilités d'une herméneutique algorithmique littéraire en faisant d'abord des parallèles entre la lecture humaine et la modélisation textuelle numérique. À partir des caractéristiques de ces deux formes de rapports au texte littéraire, nous différencions les modèles littéraires nés de la cognition et de la computation, en plus de rester critique des formes représentatives ainsi générées. Nous utilisons ensuite les cadres théoriques de la sémiotique et de la cybersémiotique pour placer l'interprétation humaine et computationnelle sur un même continuum d'évolution biotechnologique. Ainsi, nous envisageons la possibilité de sens littéraire numérique à partir de méthodes d'exploration de données appliquées aux modèles littéraires algorithmiques. Finalement, nous considérons les potentielles conséquences, limites et avantages d'une pratique de l'herméneutique algorithmique, et la forme qu'une telle pratique pourrait prendre.

Mots clés : Herméneutique littéraire ; herméneutique algorithmique ; modélisation littéraire ; cybersémiotique ; sémiotique littéraire ; exploration de données.

Abstract

In this master thesis, we explore the possibilities of digital criticism, first by creating parallels between human reading and computational text modelling. From the characteristics of both these approaches to literary work, we differentiate literary models born of cognition and computation, and stay critical of these representative forms. We then use literary semiotics and cybersemiotics as theoretical frameworks to set human and computer interpretations as two steps on the same biotechnological evolutionary ladder. Thus, we consider the possibility of digital literary meaning produced with advanced data mining tools applied to algorithmically generated literary models. Finally, we contemplate what consequences, limits and advantages such a digital hermeneutic practice would have and offer, and the shapes it may or may not take.

Keywords: Literary hermeneutics ; digital criticism ; literary modelling ; cybersemiotics ; literary semiotics ; data mining.

Remerciements

Bien qu'il ne porte que mon nom, ce mémoire n'existerait pas sans le support académique, émotionnel ou matériel de plusieurs personnes.

Je tiens donc à remercier Dr Marcello Vitali-Rosati et Dr Michael E. Sinatra pour leurs conseils et leurs contributions à ma pensée. Je souhaite aussi souligner l'enseignement de Katharina Clausius, Terry Cochran, Najat Rahman et Simon Harel qui, en dépit de la pandémie de Covid-19, ont réussi à m'enseigner des leçons que je porterai à jamais en moi. De surcroît, je me dois de mentionner le Département de Littératures et de langues du monde mené par Joyce Boro, ainsi que le travail exemplaire de Kathy Leduc.

Je remercie également mes parents dont l'aide me permet d'écrire ces lignes sous un toit, et qui ont nourri ma curiosité bien au-delà de ce que d'autres auraient toléré. Je souligne le support émotionnel de ma copine des neuf dernières années (j'écris ces lignes deux jours avant notre anniversaire), ainsi que celui de mes ami.e.s qui m'ont suivi au cours des deux dernières années durant nos rencontres hebdomadaires sur Zoom, à Argans, puis en personne.

Finalement, je remercie mes collègues étudiant.e.s du Centre de recherche inter-universitaire qui ont subi mon stress constant, et de mon programme qui ont alimenté discussions et débats, en classe et en dehors. D'entre ces derniers, il convient de nommer Sarah pour avoir réussi à briser mon isolement académique et pour ses commentaires éclairants, ainsi que Gabrielle qui, après m'avoir embarqué dans l'AÉLCUM, la SBL et VocUM, m'a également embarqué dans un avion duquel on nous a poussé en plein vol ; ce texte est un autre *leap of faith* dont la chute, bien plus longue que celle en parachute, m'a amené beaucoup de joie.

Table des matières

	Page
<i>introduc</i>	1
Mise en contexte	2
Les définitions littéraires	4
Des chiffres et des lettres	5
Sémiotique, modélisation et inférences	6
Thèse	9
Chapitre I - <i>lectio</i>	13
Lecture humaine et littérature	15
<i>Embodied, Enworlded</i>	15
Comment l'humain lit-il la littérature	16
Lecture algorithmique	19
Formalisme	22
Modélisation textuelle algorithmique	23
Caractéristiques de la lecture numérique	26
Chapitre II - <i>interpretatio</i>	31
Cadre théorique	32
Bio et cybersémiotique	34
Méta-modèles et inférences littéraires	38

Interprétations humaines	43
Au-delà du sens littéral	43
Abduction littéraire et autres inférences	45
Lecture humaine : conséquences interprétatives	48
Interprétation numérique	50
Induction algorithmique	53
Inférences bayésiennes	55
Statistiques bayésiennes et inférence	58
Il y aura des humains	61
Chapitre III - <i>experientia</i>	63
Modélisation textuelle algorithmique	64
Modélisation thématique	66
Vectorisation de mot	68
Découverte de connaissances dans les bases de données	70
<i>Digital Criticism</i>	74
Pratiques passées et actuelles	76
Vers une interprétation littéraire numérique	79
 <i>conclusio</i>	 83
 Bibliographie	 87

introduction

Si le développement des humanités numériques fait peu de vagues chez les sciences formelles¹, l'intrusion du nombre dans les lettres a provoqué une levée de boucliers chez certains humanistes, et une forte suspicion chez d'autres (Piper 2018). Dans le paradigme des deux cultures (Snow 1959), l'algorithme est une force a-littéraire, un outil de la méthode et absolutiste. Quelques siècles de séparation entre les questions qui touchent au monde et celles qui touchent à l'humain sont à l'origine de la distance qui a été artificiellement maintenue entre les sciences formelles et la critique littéraire (Binder 2020) malgré de nombreuses incursions multidisciplinaires de la part, entre autres, des formalistes. Les efforts de conservations de la ligne entre les deux cultures se séparent entre deux impératifs, le premier étant une prohibition romantique comparable à l'impératif du *Butlerian Jihad* de Frank Herbert : “Thou shalt not make a machine in the likeness of a human mind.” D'un autre côté, un rejet épistémologique du positivisme pour lequel l'algorithme et le nombre sont vus comme des hérauts ; Gadamer prévenait : il est nécessaire de choisir entre la méthode et la vérité (Gadamer 1989).

Le présent mémoire cherche à réconcilier la critique littéraire et la recherche algorithmique, d'un côté en reformulant cette dernière en termes compatibles avec la recherche dans le domaine littéraire, et de l'autre en limitant nos objectifs et définitions du sens littéraire. Soyons prudent de mentionner que nous

1. Sauf dans des domaines directement connexes comme l'informatique et les statistiques.

n'approchons pas le texte comme un artéfact culturel, ce genre d'études littéraires historiques s'inscrivent dans le domaine des *Cultural Analytics* alors que nous cherchons à contribuer au domaine naissant du *Digital Criticism* (Ramsay 2011 ; Rockwell et Sinclair 2016). Autrement dit, nous cherchons à mieux comprendre les textes et corpus, et bien qu'il puisse être souhaitable d'utiliser des informations sur le contexte sociohistorique de la production ou réception d'un texte, ces dernières prennent ici une place secondaire : notre objectif est l'interprétation algorithmique de textes littéraires. En 2011, Stephen Ramsay écrivait dans *Reading Machines* : "But"algorithmic criticism" – criticism derived from algorithmic manipulation of text – either does not exist or exists only in nascent form" (Ramsay 2011). En effet, certains des *digital humanists* les plus importants de la recherche littéraire tels que Franco Moretti, Matthew Jockers, Ted Underwood et Martin Paul Eve laissent de côté le sens et la critique littéraire, ou encore ne leur sont qu'adjacents dans leurs recherches. Leurs travaux sur la modélisation littéraire, la lecture distante et les expérimentations algorithmiques font de ceux-ci la base sur laquelle repose une partie importante des prochaines pages ; toutefois, celles-ci n'exploreront pas principalement les sociétés (So 2020), les auteurs (Jockers 2013), les genres (Underwood 2019) ou une combinaison de ces sujets (Moretti 2013). Plutôt, nous nous concentrons sur l'interprétation des textes littéraires, en nous questionnant d'abord sur la lecture (humaine ou algorithmique) avant d'en venir à notre sujet principal. Nous théorisons la lecture comme une forme de modélisation littéraire dans le premier chapitre, et l'interprétation comme des inférences à partir de modèles littéraires dans le second. Le dernier chapitre traite des possibilités actuelles d'interprétation littéraire numérique à partir des méthodes et outils d'analyse du discours et de découverte de connaissances dans les bases de données dans leur forme actuelle.

Mise en contexte

L'histoire de l'utilisation de méthodes quantitatives dans le domaine littéraire marque l'importance du dénombrement, de la position textuelle et des répétitions

de motifs dans la recherche textuelle. Par exemple, les index et les répertoires de concordances sont dès le Moyen ge des outils importants pour l'exégèse et herméneutique (Milic 1967). Dès la fin du XIXe siècle, des efforts furent déployés entre autres par les formalistes et les structuralistes pour répliquer certains succès du scientisme dans les domaines artistiques (voir (Hennequin 1888 ; Barthes 1981))². L'avènement de l'ordinateur permet d'accélérer les méthodes quantitatives déjà en place et de nouvelles formes d'analyse, l'exemple par excellence étant l'étude textuelle menée par Roberto Busa et son *Index Thomisticus* (Busa 1980), un outil numérique de découverte des concordances dans l'oeuvre de Saint-Thomas d'Aquin³. Suivent ensuite plusieurs succès numériques, par exemple en stylométrie avec l'attribution des auteurs des *Federalist Papers* (Mosteller et Wallace 1963), et l'expansion des éditions numériques (Earhart 2012). Ces premiers projets ont en commun qu'ils se basent sur des expériences réalisables par des humains... si un temps suffisant leur était accordé. Aujourd'hui, la numérisation des corpus, la puissance de calcul disponible, et des algorithmes sophistiqués rendent possibles de nouvelles formes de recherche algorithmique. En parallèle, le regain d'intérêt académique de la dernière décennie pour les humanités numériques, les avancées en linguistique computationnelle, et les études post-humanistes offrent de nouveaux outils numériques et cadres épistémologiques pour l'étude textuelle numérique. Par exemple, l'opportunité d'analyser plusieurs milliers de textes à la fois permet d'effectuer des expériences sur des genres littéraires, des époques où plusieurs corpus nationaux à la fois (Moretti 2013 ; Underwood 2019). *Distant reading* est l'expression consacrée pour ce type d'études à très grande échelle et trouve son origine dans un article écrit par Franco Moretti (Moretti 2000) ; *Conjectures on World Literature*.

2. Parmi ces tentatives de formalisation littéraire se trouvent les outils d'analyse proposés dans une édition spéciale de *Communications* dirigée par Roland Barthes : *L'analyse structurale du récit* (Barthes 1981). Neuf auteurs y proposent des représentations visuelles, des formes diagrammatiques et des formules quasi-mathématiques, autant de modèles qui attestent d'une volonté forte pour une science de la littérature. Les succès de ces méthodes furent mitigés, tant pour leurs difficultés à modéliser les textes que dans leur utilisation limitée dans les cercles littéraires.

3. Bien que ce soit le nom de Roberto Busa qui soit associé à ce projet, il est important de reconnaître le travail des dizaines de femmes qui rendirent possible l'*Index Thomisticus* bien qu'elles n'aient pas été créditées (Terras et Nyhan 2016).

Les définitions littéraires

Le champ littéraire, par sa résistance au paradigme néolibéral et à l'épistémè scientifique, est également résilient aux définitions formelles et contraignantes (Bourdieu 1991). Une polyphonie de justifications et d'explications de ce domaine de recherche coexistent, parfois, celles-ci se contredisent, parfois, elles se renforcent, mais reste toujours constant la réévaluation constante de ses fondements (Culler 2011 ; Hutchinson 2018). Ce dont nous pouvons nous permettre de ne pas douter, c'est de la centralité de la fiction, de l'écriture, du discours et de l'écriture dans nos sociétés, et qu'il convient alors d'étudier ces aspects de l'être humain (Culler 2011). Au-delà de ce fait, les méthodes, approches et épistémès divergent, quoique les questions d'analyse, d'herméneutique et de généralisation restent au moins communes, sinon centrales dans le champ littéraire. La centralité du texte est supplémentée par des modèles, connaissances ou intuitions pluridisciplinaires⁴ qui interagissent pour maintenir une discussion de plus en plus sophistiquée sur les objets littéraires et ce qu'ils représentent. C'est donc sans surprise que le concept d'interprétation littéraire soit difficile à formaliser, de même qu'il est virtuellement impossible de générer un consensus sur ce qui constitue une telle chose. Il est toutefois possible de dire que l'herméneutique littéraire existe quelque part sur un spectre où se trouve à un extrême une interprétation absolue (similaire à celle que se propose de faire l'interprétation juridique), et à l'autre l'interprétation subjective (pour laquelle le sens est uniquement une fonction de l'être pensant qui est en contact avec le texte) (Peter Szondi 1978). L'interprétation littéraire est plus ou moins codifiée ; plusieurs formes d'interprétation coexistent, et plusieurs interprétations peuvent se superposer pour un même système interprétatif. La quête pour une interprétation littéraire numérique ne peut être celle d'une lecture absolue, une idée contre laquelle les littéraires s'insurgent avec raison, bien que de manière quasi-unanime les humanistes ne la proposent pas. Les projets récents les plus ambitieux sont fermement campés

4. La littérature est un domaine toujours qui, même lorsque son étude est à son plus "pur", tire ses racines dans des questions d'ordre philosophique ou esthétique. Les outils d'analyse de la littérature offrent un panorama de techniques et de théories tirées de toutes les sciences sociales et humaines (Tracey et Morrow 2006).

dans un même camp : l'étude numérique des textes de fiction et de la poésie est, pour les auteurs qui ont inspiré le présent mémoire (Rockwell et Sinclair 2016 ; Underwood 2019 ; Piper 2018), une source de questions et de premiers pas vers des découvertes sociohistoriques, herméneutiques, biographiques ou poétiques.

Des chiffres et des lettres

Dans *The Two Cultures and the Scientific Revolution*, C. P. Snow décrivait le *gap* entre les communautés scientifiques et littéraires (Snow 1959) ; un paradigme de séparation artificielle de deux épistémès dont l'origine remonte aux romantiques qui séparèrent le domaine l'humain pour le protéger du scientisme (Binder 2020). Les humanités numériques s'inscrivent dans une volonté de réunification des républiques du physique et de l'esprit, portée par la compréhension de l'artificialité de cette fragmentation et les développements des dernières années en sciences humaines et sociales (Binder 2020). Le rapprochement entre les deux cultures prend plusieurs formes ; par exemple, les *Cultural Analytics* : l'utilisation de méthodes quantitatives et cadres épistémologiques tirées des sciences et de la statistique sur des objets culturels ou données sur les sociétés humaines. Cette récupération du mode scientifique dans le cadre de textes littéraires permet aux humanistes de poser des questions aux réponses claires ; soit en posant des hypothèses falsifiables, en implémentant un protocole de recherche rigoureux et en analysant les résultats de leurs expériences. Effectivement, la majorité des projets littéraires numériques ne sont pas *littéraires*, mais historiques, linguistiques, ou stylistiques (Moretti 2013 ; Underwood 2019 ; Jockers 2013). D'autres types de projet utilisent des outils computationnels pour générer de nouvelles formes de représentations des objets culturels ; par exemple, l'édition numérique permet d'augmenter le texte, de le réduire sélectivement, ou d'attirer l'attention du lecteur vers certains passages, motifs ou autre aspect selon des règles précises et paramétrables (Earhart 2012). L'édition, comme toute transformation téléologique, est productrice et porteuse de sens ; les transformations numériques automatiques (dont les paramètres sont changés par le texte étudié ou par

l'utilisateur) peuvent avoir des conséquences inattendues sur le sens de l'oeuvre (Souchier 1998 ; McLuhan 1964). Le médium numérique, par son agencement propre et son réagencement (éditorial, involontaire, ou automatique) des textes, est porteur de sens—les humanistes se doivent de rester suspicieux des *Black Box* (Underwood 2014). Le médium numérique est structurellement porteur de sens et les humanistes doivent rester suspicieux des mécanismes cachés qui offrent des résultats rapides (*Black Box* (Underwood 2014)) : un regard critique sur la technologie et sur les textes qui sont représentés de par celle-ci n'est possible qu'en connaissant les règles algorithmiques qui produisent les résultats analysés. L'arrivée du nombre dans la recherche en sciences humaines et dans les lettres est à la fois un champ de bataille épistémologique et une réconciliation (Piper 2018 ; Binder 2020), résultats de la lente convergence entre la phrase et l'équation. Si les sciences humaines ont embrassé les statistiques et les arts visuels la technologie, pourquoi la littérature offre-t-elle plus de résistance à l'intégration d'outils numériques dans son étude des textes littéraires (Jannidis 2020)? Le cadre théorique nécessaire à une herméneutique littéraire numérique et à la recherche algorithmique en littérature est lent à être produit et à être adopté, mais ce processus est bien enclenché (Piper 2018). Toutefois, la majorité de ces études ne sont pas littéraires au sens propre ; elle n'est pas du registre du *Digital Criticism*, mais bien des *Cultural Analytics* (Underwood 2019 ; Jockers 2013 ; Moretti 2013). Le *leap of faith* conceptuel que demande une interprétation textuelle littéraire est plus grand que celui des *Cultural Analytics* ; notre quête demande une approche mi-phénoménologique, mi-naïve de la lecture et une ouverture de la notion d'interprétation.

Sémiotique, modélisation et inférences

Dans le présent mémoire, nous considérons le texte littéraire comme un assemblage sémiotique. La sémiotique de Pierce nous fournit un cadre où l'interprétant est nécessaire pour générer du ou des sens à partir d'un réseau de signes. Le signifiant existe donc dans un état constant de superposition anticipative : en

attente d'un interprétant, son espace de sens possibles (un espace sémiotique) est immense. C'est l'interprétant qui peut le faire s'effondrer en un seul signifiant, même si généralement l'effet du signe sur l'interprétant laisse plusieurs sens en état de superposition.

We may now try out a definition: a literary work is a discourse in which an important part of the meaning is implicit. This is a Semantic Definition of 'literature', since it defines 'literature' in terms of meaning. (Olsen 1982)

Dans le cadre des études littéraires, le rejet de l'interprétation littérale du texte est tenue pour acquis, et c'est d'ailleurs ce vide herméneutique qui motive la création de nouvelles formes interprétatives (Culler 1976). Le présent mémoire utilise les définitions de la sémiotique puisque cette dernière permet de faire le pont entre la production de réseaux de significations par les êtres vivants et par les algorithmes (bio et cybersémiotique), tout en étant un outil herméneutique littéraire (Culler 2001). Le texte littéraire dans toute sa complexité est opaque pour un ordinateur, c'est pourquoi nous ferons appel à la notion de modélisation. Cela n'est pas trop loin du processus humain de lecture :

la lecture est un processus de *construction* [puisque le texte est] un pur *artefact* dénué de toutes significations. Mais d'autre part, la prégnance des stéréotypes est telle que, sitôt qu'il est situé dans un contexte socioculturel donné, le texte devient un *objet social* dont les signifiants peuvent être référés à des schémas sémantiques de ce contexte, et la lecture devient quant à elle un processus de reconnaissance et de combinaison d'une matière préexistante. (Dufays, Gemenne, et Ledur 2005)

Le modèle comme forme de représentation du texte n'est pas le texte lui-même, mais encore le texte n'est pas significatif sans l'interprétant et son travail de modélisation. Toutefois, un modèle du texte, malgré la perte importante d'information

intrinsèque à toute intermédiation⁵, est plus facile à traiter numériquement que du texte simple. À ce propos, il faut faire la différence entre un *sign-of* et un *sign-for* : selon Noel Burton-Roberts, le premier n’est pas représentationnel, et puisque la littérature est figurative, fictionnelle, non littérale, nous devons la voir comme représentationnelle (*sign-for*) (Burton-Roberts 2013). Donc, le sens n’est pas une propriété sémiotique du signe, ce dernier ne possédant pas de telles propriétés. Burton-Roberts dissocie la sémantique du signe pour en faire deux sujets à part : “meaning is not a (semantic) property but a semiotic *relation* (to semantic properties)” (Burton-Roberts 2013). L’étude du signe est ainsi l’étude de réseaux ; l’accent est mis sur le fait que les signes sont arbitraires, et sur la manière dont les conventions de signification sont dépendants d’un horizon, un *umwelt* de signes.

Meaning is by its nature *communicative*. I take ‘communicative’ when it involves meaning, to be equivalent to ‘leads [the subject] S to entertain a thought’. [...] I am concerned with communication, but only insofar as it can held to involve *meaning* (and thus thought and inference). (Burton-Roberts 2013)

Il précise : “meaning-X [the signifier]-to-S [the subject] depends on S making inferences based on her beliefs” (Burton-Roberts 2013). La croyance a un équivalent formel : les statistiques bayésiennes simule un degré de certitude de manière intelligible pour un ordinateur ; de plus, ces “croyances” sont modulées en suivant des règles statistiques au contact de nouveaux cas de figure. Une relation significative cognitive produite par une inférence (dans ce cas précis, une abduction⁶) a également un équivalent dans la computation : des algorithmes permettent de produire des inférences, et ces dernières peuvent être rassemblées et traduites en relations computationnelles (Fayyad, Piatetsky-Shapiro, et Smyth 1996). Un corpus, ou le modèle d’un corpus, constitue le *umwelt*, l’horizon de l’ordinateur

5. Ce terme exploré plus en détails dans le deuxième chapitre, voir (Hayles 2019).

6. L’inférence par abduction à l’avantage sur les autres formes d’inférence de ne pas mettre l’accent sur la vérité, mais sur les probabilités.

; ses réseaux de signes limitent les structures qu'il peut dégager et interpréter grâce à des inférences numériques.

Thèse

L'expression "interprétation littéraire numérique" représente plusieurs pratiques potentielles ou actuelles, et si nous explorerons plus d'une avenue de recherche dans notre troisième chapitre, il convient de restreindre et de qualifier le sujet de cette étude. "Littéraires" est des trois mots celui que nous rendons le plus contraignant en faisant de celui-ci un cadre rigide pour les objets que nous désirons étudier et le type d'analyse que nous désirons en faire. Si nous théorisons comment des données biographiques, sociohistoriques, psychologiques et linguistiques peuvent contribuer à l'interprétation littéraire numérique, les questions que nous sommes disposés à poser sont celles d'un littéraire : nous ne considérerons pas le texte comme un artéfact historique, pas plus que nous ne mettrons l'accent sur le contexte ou l'auteur (Ramsay 2011). Les textes de fiction n'ayant pas de sens fixe ou propre (le rejet du sens littéral étant une caractéristique quasi-universelle de la recherche littéraire (Culler 1976)), nous devons mettre en place des mécanismes par lesquels l'interprétation restera littéraire ; soit ouverte et multiple. Le mot "numérique" a également une double fonction ; il dénote que nous désirons interpréter à l'aide de la computation plutôt que la cognition, mais aussi que nous mettons l'accent sur les méthodes qui sont nécessairement numériques plutôt que possiblement numérique⁷. Ainsi, nous considérerons des algorithmes de modélisation thématique et de découverte de connaissances dans les bases de données qui outrepassent les limites de la cognition biologique, ou utilisent des présuppositions qui sont si Autres⁸ que l'être humain ne saurait les imiter. Notre dernier mot, "interprétation" est problématique, parce qu'il dépend de valeurs et de conventions qui changent dans le temps : certains objectifs

7. Par exemple, les définitions quantitatives des formalistes russes et de la stylométrie du milieu du dernier siècle peuvent être automatisées, mais ne dépendent pas de l'ordinateur.

8. Jeffrey M. Binder utilise le terme *alien* pour signifier une différence de cette magnitude (Binder 2016).

et limites de l'herméneutique et de la recherche littéraire du début du siècle sont même incompatibles avec ceux d'aujourd'hui. Toutefois, à défaut d'une définition formelle stable ou d'une expression aussi évocatrice, nous devons resserrer le domaine de ce mot : pour le présent mémoire, une interprétation est une transformation de la forme littéraire en connaissances pertinentes et intelligibles sur le texte, et possiblement sur d'autres objets, contextes, ou sujet. Ainsi, une définition fonctionnelle, littéraire, et utilisable dans le contexte du *Digital Criticism* serait : l'interprétation littéraire est la traduction⁹ d'un assemblage sémiotique¹⁰ en signes intelligibles qui ne sont pas équivalents au sens littéral du premier assemblage. Nous justifierons et établirons des définitions complémentaires au cours de nos deux premiers chapitres, mais sommes désireux de rassurer que les enjeux épistémologiques seront couverts dans les sections sur les inférences du second chapitre. Il est important de noter que l'interprétation littéraire numérique ne sera pas un décodage d'un code à l'aide d'une clé (d'un système interprétatif fixe), elle se doit de résister à l'unicité de solution et à la méthode correcte ; qui plus est, elle devra éviter d'investir l'épistémè scientifique. L'algorithme se retrouve ici dans une étrange situation puisque ce dernier était par défaut destiné à effectuer une série de transformations mécaniques¹¹ prédéfinies et répliquables. Trois avenues nous permettront de circonvenir ce paradoxe ; d'abord la collaboration entre l'ordinateur et le chercheur ; puis l'utilisation de statistiques capable de considérer plusieurs chemins valables et de les présenter ; enfin, certains cadres numériques sont en mesure de concevoir et traiter des états simultanés différents et des superpositions de réponses contradictoires.

Le problème que nous souhaitons poser en est d'abord un de sémiotique, de modélisation, et d'émergence du sens. La tâche à accomplir est donc triple :

9. Le mot traduction est choisi avec soin, car il implique une remédiation, une transformation linguistique, et l'introduction d'un nouvel auteur.

10. Un assemblage sémiotique est une série de signes qui représente quelque chose pour lequel il n'y a pas de signe préétabli ou pour lequel aucun signe ne saurait être établi.

11. Searle les appelle des manipulations symboliques tout en mentionnant que ces symboles en particulier ne symbolisent rien pour la machine (Searle 1980) ouvrant un autre débat : la sémiotique nécessite-t-elle que le signifiant soit réel ou même conceptuel? Un signe vers un point d'un espace mathématique abstrait répond-il aux mêmes règles qu'un signe linguistique humain?

1. Nous devons situer la modélisation automatique de textes littéraires par rapport à la lecture humaine et justifier les modèles numériques comme des représentations textuelles.
2. En utilisant un cadre conceptuel capable de représenter le rapport humain et algorithmique aux modèles ; nous devons exprimer les processus d'analyse comme générateur de sens.
3. Il nous faut finalement démontrer que le sens produit par l'analyse numérique de modèles littéraires générés automatiquement peut être intelligible et pertinent, un point qui est malheureusement en dehors de la portée du présent mémoire.

Alors qu'une part importante des recherches littéraires quantitatives ont des visées du domaine des *Cultural Analytics*, nous cherchons plutôt à contribuer au champ naissant du *Digital Criticism*.

Chapitre I - *lectio*

Nous divisons l'interprétation littéraire en deux étapes, d'abord la lecture, activité humaine naturalisée à un point où elle en devient parfois invisible, et l'interprétation que nous traitons dans le prochain chapitre. Dans son *Essai sur la notion de lecture*, la philosophe française Simone Weil place l'origine de sa notion de lecture dans l'anticipation et l'association parfois arbitraire :

Le mystère est que des sensations en elles-mêmes presque indifférentes nous saisissent de la même manière par leur signification. Quelques traits noirs sur du papier blanc, cela est bien différent d'un coup de poing dans l'estomac. Mais parfois l'effet est le même. [...] Il ne s'agit pourtant pas de quelque chose d'analogue au réflexe conditionnel; il s'agit de quelque chose d'analogue à la lecture, où parfois une combinaison de signes toute nouvelle, et que je n'avais jamais vue, me saisit l'âme, où la signification qui blesse pénètre, avec le blanc et le noir, aussi irrésistiblement qu'eux. [...] Ainsi les significations, qui examinées abstraitement sembleraient de simples pensées, surgissent de toutes parts autour de moi, s'emparent de mon âme et la modifient [...] (Weil 1946)

Les processus derrière la lecture humaine sont nombreux et complexes, tant et si bien qu'il est difficile de théoriser cet acte vieux de 6 000 ans que nous accomplissons tous les jours de manière automatique (Glavanakova 2020). La

lecture résiste aux descriptions et à l'analyse de la même façon que l'esprit humain s'en échappe ; il n'en reste pas moins que plusieurs modèles offrent une meilleure compréhension de la lecture. Plusieurs modèles compétitionnent (Dufays, Gemenne, et Ledur 2005) ; la psychologie, les sciences cognitives, la narratologie, les approches esthétiques, et la philosophie offrent des théories parfois complémentaires, parfois incompatibles. Malgré l'absence d'une théorie définitive ou unifiée de la lecture¹², et l'incertitude quant à la place exacte de la lecture dans l'expérience humaine, il est possible de caractériser la lecture. En utilisant ces caractéristiques (développées dans la section **Comment l'humain lit-il la littérature**) comme des principes premiers, nous sommes à même développer une théorie limitée de la lecture, suffisante à nos buts, car c'est en explorant de manière similaire les caractéristiques de la lecture artificielle que nous pouvons produire un cadre de comparaison entre ces deux formes de lecture. Un tel cadre servira d'outil dans le prochain chapitre pour trouver les différences et parallèles entre l'interprétation humaine et l'interprétation numérique. En envisageant ces caractéristiques comme des principes premiers de la lecture, les différences entre la lecture humaine et la lecture algorithmique se traduisent en différents modes interprétatifs. Dans le cas de la lecture humaine, l'accent est mis sur la production de modèles à partir de textes littéraires en utilisant le concept de *Dasein* (Heidegger 1962) et la (bio)sémiotique dans un contexte herméneutique (Hayles 2019). La lecture numérique est approchée du point de vue du formalisme et du structuralisme, ainsi que du côté des différences de limites respectives de la cognition humaine et de la computation. Les discussions d'interprétation numérique textuelle se trouvent dans le chapitre II - *interpretatio* dans lequel des parallèles entre la biosémiotique et la cybersémiotique (Hayles 2019) servent de point de départ à la construction d'un cadre théorique de l'interprétation algorithmique.

12. Il est difficile de totalement développer une théorie de la lecture sans se reposer sur une théorie de l'être humain (Furlong 1995).

Lecture humaine et littérature

La lecture est une fonction cognitive de haut niveau qui se base sur des fonctions cognitives plus simples à propos desquelles nous n'avons que des connaissances partielles (Glavanakova 2020) et qui tombent en dehors de la portée du présent mémoire. Dans *Lenses on Reading. An Introduction to Theories and Models* Daine Tracey et Lesley Morrow identifient plusieurs approches à une théorie de la lecture : comportementale, constructiviste, développementale, physiologique, cognitif, et l'apprentissage social, affectif. Elles tirent des leçons des domaines de l'éducation, la psychologie, la sociologie, la linguistique et la neuroscience (Tracey et Morrow 2006). Notre approche à la lecture humaine est donc ici limitée, nous ne considérerons qu'indirectement les fonctionnements de la mémoire, de la psychologie, de la cognition ou de l'affect. La lecture peut toutefois être présentée comme étant composite, inter(et intra-)textuelle, diachronique et approximative. Différentes formes de lectures se superposent, différents textes se parlent chez le lecteur qui rencontre ceux-ci de manière séquentielle, interrompue, et sous des modes de lecture changeants. Le problème de la mémoire est également à considérer, les textes n'étant mémorisés que de très rares cas. Similairement à la lecture, la fiction est une forme discursive difficile à formaliser, et même à analyser ; comme dans le cas de la lecture, plusieurs cadres d'analyse (Barthes 1981) permettent d'étendre notre connaissance des formes narratives. Ces théories de la littérature divergent parfois dans leur explication de la création de sens littéraire (Furlong 1995), nous nous concentrons donc sur le modèle sémiotique du texte littéraire.

Embodied, Enworlded

On ne se baigne jamais deux fois dans le même fleuve.

- Héraclite

Even in Kyoto

Hearing the cuckoo's cry

I long for Kyoto

- Basho (traduction de Lucien Stryk)

L'être-là de l'être humain dans l'espace-temps est un facteur premier dans son rapport à la littérature : en tant qu'individu *embodied* et *enworlded*, la lecture humaine n'est pas neutre, elle est fortement teintée par les expériences du lecteur. Les modèles physiques, émotionnels, relationnels, etc. permettent aux humains une forme d'intertextualité intermédiaire : leurs références sont complexes, et leurs référents sont bâtis en dehors du monde littéraire. Ainsi, le lecteur humain reconnaît la fiction comme ayant des règles différentes de la réalité, et peut mettre en relation sa lecture de la fiction avec sa lecture du monde : ses modèles sont comparés et contrastés avec des modèles extra-littéraires. Une conséquence importante de *Dasein* est la prolifération de modes de lecture et de cadres interprétatifs, parfois chez le même individu. Les littéraires ne considèrent pas tous les référentiels comme équivalents, plusieurs référents peuvent être utilisés pour lire un texte ouvert avec des résultats différents sans qu'aucun ne soit préférentiel (Dufays, Gemenne, et Ledur 2005). Les modes de lecture et cadre interprétatifs des lecteurs dépendent de facteurs linguistiques et culturels, des séquences textuelles, et des expériences du lecteur. L'être-là des lecteurs génère pour chaque texte un espace interprétatif (voir la section sur le sujet) dont les possibilités s'effondrent (au moins partiellement) au contact d'un lecteur particulier. Ainsi, la sémiotique de Charles Sanders Peirce (Atkin 2013 ; Eco 1976) doit être privilégiée sur celle de Ferdinand de Saussure (Culler 1976); dans un signe repose un espace de signifiant qui, comme une particule subatomique, existe dans un nuage de probabilités qui n'est resserré qu'au contact d'un observateur ou interprétant (Cohen-Tannoudji, Diu, et Laloë 2005).

Comment l'humain lit-il la littérature

La lecture humaine est **diachronique** (Miller 2010), l'être humain ne pouvant concentrer son attention que sur un texte à la fois ; même dans les cas de lectures simultanées, par exemple dans le contexte de recherches comparatistes, le lecteur

doit passer d'un texte à l'autre faute de pouvoir lire les deux en même temps. Ses lectures sont de plus séparées, interrompues par les périodes passées loin du texte : la distraction passagère, le fait de poser le livre pour un temps et de sortir du monde littéraire sépare le temps du lecteur entre la lecture de la fiction et la lecture du monde (Weil 1946). En conséquence de son diachronisme, la lecture humaine est **séquentielle et ordonnée** : la lecture d'un texte par un humain est toujours une étape intermédiaire de son parcours littéraire (et de ses expériences de vie). Elle est de plus informée par ce qui a été lu (ou s'est passé) précédemment, et envisagée comme faisant partie d'un continuum. La série de rencontres textuelles d'un lecteur ne suit pas de règles précises, il est virtuellement impossible de trouver deux lecteurs ayant l'expérience d'une même séquence littéraire étendue. Ces variations entre lecteurs donnent lieu à déclinaisons infinies de leurs grilles de lecture, cadres interprétatifs et modèles littéraires. L'ordre des lectures est un sujet d'importance pour l'exégèse (Thiselton 1992) car il affecte grandement la production du sens pour trois raisons : 1. Les références (réelles ou accidentelles) aux textes précédemment lus informent la lecture ; 2. Les lectures informent les modèles littéraires et grilles d'interprétation du lecteur, et ces cadres de lecture transforment l'expérience de lecture et d'interprétation¹³ ; 3) La lecture est influencée par l'état de conscience du lecteur, et l'état de conscience du lecteur est influencé par ses lectures antérieures (Thiselton 1992). La lecture humaine est donc **intertextuelle**, chaque texte tisse des liens avec d'autres textes, mais aussi avec le monde et avec lui-même. Dans le premier cas, la récursion de thème, de motifs, et de passages joue sur plusieurs niveaux à la fois. Par exemple, le motif du père absent dans l'univers cinématique de Marvel déclenche une réaction émotionnelle amplifiée par sa récurrence (Padnick 2019 ; Hale-Stern 2018) et l'ubiquité du thème de la fin du monde et du sentiment pré-apocalyptique rend possible la production de *Don't Look Up* et sa déconstruction des scènes d'*Armageddon* et d'*Independence Day*. L'intertextualité avec le monde

13. Thiselton s'inquiète des modifications conscientes et inconscientes que les lecteurs appliquent aux évangiles à cause de leur parcours littéraire. Par exemple, la distance entre l'horizon littéraire auquel a été habitué le lecteur et celui des premiers chrétiens peut mener à des interprétations dangereuses (Thiselton 1992).

perçu par le lecteur lui permet d'envisager le texte à partir de sa relation à un contexte sociohistorique, et de transformer son rapport au monde. Toute lecture est également **intratextuelle**, c'est-à-dire que chaque partie du texte (phrase, paragraphe, chapitre, etc.) voit sa lecture affectée par les parties de texte précédentes¹⁴. La poétique et la stylistique sont des formes d'intratextualité puisque dans chacun de ces cas, l'oeuvre est génératrice de règles qui la structurent (même si ces dernières viennent d'une tradition extérieure). Par exemple, la structure supplante le sens comme point focal de l'analyse pour les *New Critics* dont l'approche est notamment immanente (Dufays, Gemenne, et Ledur 2005) : le système structural qui met en relation les signes du poème est l'objet d'étude (McGrath 1985).

La mémoire humaine étant imparfaite et cette dernière étant (comme la lecture) sujette aux changements d'état de conscience, la lecture est **approximative**. Le processus de référencement aux autres textes est inefficace puisque les textes ne sont pas mémorisés sauf en de très rares cas ; il est donc commun qu'un lecteur ne remarque pas les rappels thématiques, les récursions de motifs ou encore les concordances textuelles. Pire encore, le souvenir du texte présentement lu est sujet à des oublis partiels, ce qui mène parfois à des situations où le lecteur n'a pas toutes les informations pour comprendre l'oeuvre bien qu'elles aient été établies quelques chapitres ou pages plus tôt. Le lecteur retient de manière sélective des fragments d'oeuvre pour un temps limité, ce qui rend son expérience de lecture unique, subjective et arbitraire : la même personne ne saurait lire deux fois le même texte de la même façon. L'état de conscience et le cadre littéraire utilisé par le lecteur lors de sa lecture transforme également son expérience de l'oeuvre, cette dernière est le sujet d'une négociation. En nommant la lecture comme un acte **négocié**, nous attirons l'attention d'abord vers la différence entre les codes de production d'une oeuvre et les codes utilisés pour la décoder, et ensuite vers la remédiation nécessaire du texte lors de la lecture. L'asymétrie d'encodage est une conséquence de la nature ordonnée de la lecture et de *Dasein*, et elle est décrite,

14. L'intratextualité est au coeur du modèle sémiotique du sens littéraire et sera développée plus en détails dans le chapitre II.

dans un contexte médiatique, par Stuart Hall dans *Encoding and Decoding in the Television Discourse* (Hall 1999). Hall identifie trois rapports à l'information ; la position hégémonique, négociée, et d'opposition, mais certains de ses critiques préfèrent utiliser un spectre. C'est le cas de Jarrett Cole qui répond à Hall que toute lecture est au moins infinitésimalement négociée dans "Negotiating Code: The Decoding of Stuart Hall" (Cole 2020). La correspondance entre le texte et la pensée est imparfaite ; la lecture comme l'écriture ne permet pas le transfert direct de l'information, tant à cause des différences linguistiques inhérentes entre individus qu'à cause des limites et de la stabilité du texte par rapport à la pensée. Finalement, la lecture humaine est une activité **composite** qui repose sur plusieurs rapports au texte : se superposent les lectures émotionnelle, phonétique, poétique, métaphorique, etc. (Barthes 1981). Dans le cadre de la fiction, la lecture sémantique (soit l'interprétation littérale) est supplémentée et même supplantée par les autres formes de lecture telles que la connotation, les niveaux d'ironie, d'humour ou de fiabilité, la reconnaissance (consciente ou non) des structures des récits, l'identification à des personnages... La tension entre deux formes de lectures qui se contredisent et la superposition de formes de lectures qui se renforcent sont des points d'intérêt en littérature (Culler 2011). Les paradoxes de la littérature (qui offrent des interprétations multiples, superposées, incompatibles, concomitantes) imitent ceux de la réalité et leur étude à partir de théories tirées de recherche en sciences humaines enrichie de surcroît la lecture humaine.

Lecture algorithmique

L'impression d'incompatibilité entre la lecture et les méthodes algorithmiques a deux sources distinctes ; la première, que nous devons mettre de côté plutôt qu'adresser, est une crainte : celle que la machine puisse remplacer le chercheur dans le domaine qui est, dans le paradigme des deux cultures (Snow 1959), le plus humain d'entre tous. Toujours dans ce paradigme, l'absolutisme algorithmique, dans sa rigidité formelle, est à l'opposé du récit ; leur incompatibilité est d'abord

une perception, une impression d’incalculabilité du texte littéraire (Meunier 2017) qui ne tient pas compte du fait que nous calculons sans cesse l’incalculable dans plusieurs domaines (Cohen-Tannoudji, Diu, et Laloë 2005). Dans le cadre de la lecture algorithmique, comme dans des problèmes impossibles à résoudre en physique quantique, il est possible de modéliser des approximations qui tendent vers un résultat inatteignable, ou de faire preuve d’originalité.

Algorithms, when thought of as remote, inflexible mathematical structures underlying computer programming and the more deterministic branches of science and engineering, can seem irrelevant or even antithetical to the work of the humanities. By concentrating our efforts on the building of algorithmic “text analysis” tools, do we unthinkingly imply that the craft of scholarship can be mechanized? Are we tacitly putting algorithms forth as substitute for contemplation and insight? Or (a far more insidious assumption) are they, as the quiet servants delivering us the “content” of an archive, simply beneath our notice and contempt? (Nowvskie 2004)

The craft of scholarship et *l’insight* sont des concepts chargés de jugements, de valeurs et d’histoire : les dissocier du domaine de l’humain est une attaque contre l’orthodoxie et le romantisme latent du domaine littéraire (Binder 2020 ; Piper 2018). Dans sa thèse doctorale, Bethany Paige Nowvskie fait l’éloge d’un art de la transformation algorithmique pour provoquer l’intuition humaine et comme nouvelle avenue de recherche académique. Sa vision ludique de l’expérimentation herméneutique algorithmique rappelle l’attitude de Sinclair et Rockwell dans *Hermeneutica* ; la recherche assistée par ordinateur commence par le jeu et le jouet (“to play” et “toys”) (Rockwell et Sinclair 2016). La vision cartésienne de l’algorithme est toutefois dépassée, puisqu’elle repose sur une compréhension purement mécanique et non statistique de la computation. Aujourd’hui, des algorithmes peuvent générer des modèles (heuristiques) sémantiques (Mikolov et al. 2013), imiter (grossièrement) des aspects de la lecture humaine (Hutto et

Gilbert 2014), et même développer des stratégies qui rappellent (mais ne sont pas) l'intuition (Saldler et Regan 2019). Cela ne signifie pas que la machine sera en mesure de vivre l'expérience humaine, mais permet néanmoins d'espérer que la lecture d'un texte par un algorithme soit générateur de sens. La transcendance dans toutes ses formes, la métaphore et autres figures de style, la description des merveilles, l'incarnation et de l'omniprésence du trauma qui lui est lié : la littérature parle d'une réalité qui n'existe pas pour l'ordinateur (Searle 1980). Si la lecture algorithmique est porteuse de sens, ce dernier ne sera pas du même domaine que celui produit par un être humain, mais cela n'invalide ni cet hypothétique sens ni cette lecture, c'est plutôt là une avenue de recherche inexplorée (Binder 2016).

Tenter de répliquer numériquement la lecture humaine sous toutes ses facettes est une idée attirante, mais une telle démarche serait au mieux une batardisation partielle dans le contexte algorithmique actuel. À ce jour, la réplication des fonctions cognitives de haut niveau est au mieux une approximation à partir de présuppositions mathématiques. C'est aussi un enjeu d'intertextualité, les relations entre le texte et la réalité (modèles sociohistoriques ou *Dasein*) sont porteurs de sens, et notre rapport à la réalité est corporel, intentionnel, et *enworlded* (Searle 1980). Ainsi, la lecture numérique peut éviter d'être teintée par les traumas et le désir, l'impuissance ou le contrôle, les deuils et les joies ; l'algorithme peut être libre des affects de la nature humaine¹⁵. La lecture algorithmique n'est pas la lecture humaine et ne saurait l'être sans la simulation de l'expérience humaine, mais ce n'est pas là le but de l'interprétation littéraire numérique. Le processus intuitif dépend de la possibilité de découvrir structures non triviales entre différents niveaux de lecture. L'ordinateur peut également avoir plusieurs rapports différents au texte, mais son expérience des signifiants est dissociée du contexte (humain) qui les a produits. La solution à ce problème n'est pas l'apathie, car dans cette incapacité se trouve une opportunité pour la

15. Dans les mois qui précèdent le dépôt de ce mémoire, un ingénieur de Google fut licencié après avoir décrété que l'algorithme conversationnel PaLM soit doté d'une conscience (Wertheimer 2022). Nous ne nous prononcerons pas ici sur cette affaire.

création de nouvelles connaissances inusitées. Les réseaux de neurones sont à même de découvrir des liens inusités qui rappellent l'intuition humaine à certains égards (Binder 2016).

Formalisme

Le formalisme, comme le structuralisme, est antérieur aux ordinateurs ; similairement, l'algorithme n'est pas contingent à la computation. Plusieurs méthodes formelles d'analyse textuelle peuvent être implémentées par des humains, bien qu'elles soient grandement accélérées par l'utilisation d'ordinateurs. L'automatisation des travaux de Iarkho dans le cadre de la thèse doctorale de Inna Alekseyevna Wendell (Wendell 2021) et la découverte des concordances dans l'oeuvre de Saint Thomas d'Aquin (Busa 1980) étant de bons exemples. D'autres méthodes formelles se basent sur des modèles humains pour représenter mathématiquement ou visuellement des modèles, par exemple Tzvetan Todorov qui identifie ainsi les deux motifs principaux du *Decameron*, soit *avoided punishment* et *conversion* (Todorov 1969). On note ainsi deux avenues du structuralisme, soit la formalisation du texte avec des représentations visuelles ou mathématiques d'un côté, et de l'autre une approche quasi-scientifique du texte comme objet formel à caractériser et classifier. Dans un cas comme dans l'autre, le rapport entre le particulier et le général est maintenu comme étant direct, testable, et prouvable ; des concepts qui sont tirés des sciences plutôt que des lettres. Cette appropriation du langage scientifique court-circuite une possible réconciliation du paradigme des deux cultures en déplaçant la littérature dans le domaine de la science (Piper 2018). Ironiquement, le formalisme est une réaction au romantisme qui a produit la division entre les lettres et les sciences et qui est lui-même une réaction à la montée du scientisme (Binder 2020). Le formalisme agit donc comme une science, par hypothèses et expériences, et emprunte plusieurs aspects de la science tels que les représentations mathématiques et graphiques, la computation, la méthode, la description, et l'isolation du sujet d'étude. Si nous n'adhérons pas aux présuppositions philosophiques des formalistes, leur approche algorithmique

au texte est une source majeure d'inspiration en particulier le cas de Boris Iarkho que nous présenterons plus loin dans ce chapitre.

Modélisation textuelle algorithmique

Nous ne raisonnons que sur des modèles. - Paul Valéry

Le modèle du texte n'est pas le texte (Erb, Ganahl, et Kilian 2016), et si un texte littéraire n'est pas une base de données (Marche 2012), une base de données générées à partir d'un texte littéraire est fonctionnellement un modèle de ce texte. La modélisation algorithmique, comme la modélisation faite par un humain, est un processus politique et une intermédiation :

It would be naive to suggest that measurements are devoid of power relations, but neither are the acts through which proper names account for knowledge. Models make us aware of the externalities through which we arrive at truth claims. (Piper 2017)

specific levels formed by physiochemical media always contain more information within the level than they communicate upward or downward. (Hayles 2019)

models inscribe our beliefs within them. (Piper 2018)

En passant du texte au modèle, de l'information est perdue, mais de nouvelles formes d'information sont rendues visibles, et ces nouvelles informations portent la marque du modèle. Andrew Piper identifie cinq étapes de la modélisation littéraire dans *Think Small: On Literary Modeling* (Piper 2017): 1. *Theorization* : Selon Piper, la modélisation est téléologique ; elle sert à tester une hypothèse et le passage du texte au modèle encode une question. En passant d'un média à un autre, de l'information est perdue, mais des informations qui ne sont pas accessibles au niveau du texte sont rendues visibles, une idée aussi retrouvée chez N. Katherine Hayles (Hayles 2019). 2. *Conceptualization* : Cette étape demande

la particularisation (ou formalisation) de concept à des fins de falsification, le concept doit pouvoir être mis en relation avec le modèle théorique. Ce type de remédiation conceptuelle est imparfait, Piper donne l'exemple de Lancashire et Hirst qui ont utilisé la richesse du vocabulaire d'Agatha Christie au fil du temps comme indicateur de l'avancement de sa maladie mentale (Lancashire et Hirst 2009). La modélisation littéraire spécifique dépend, dans ce genre d'exemple, "on the specifications of other models in a larger representational web." (Piper 2017)

3. *Implementation* : L'implémentation de méthodes algorithmiques voit le chercheur générer les règles formelles nécessaires à la transformation du texte ou corpus en modèle. C'est une étape délicate d'un point de vue épistémique : c'est à ce point que l'hypothèse et les concepts qui lui sont associés entrent définitivement dans le domaine du nombre.

4. *Selection* : Le choix des textes à partir desquels le modèle sera produit est, selon Piper, "another approximation of the world through data" (Piper 2017). Cette étape est aussi critique puisqu'il faut choisir les textes en fonction de la question de recherche et du modèle généré pour y répondre, un processus similaire à "laying two maps on top of each other." (Piper 2017) Il est possible qu'une sélection mène à une perte d'information, ou une meilleure représentation, mais Piper note qu'il est préférable d'avoir plus de données que moins.

5. *Validation* : Pour valider un modèle, Piper propose un test statistique (Lancashire et Hirst utilise la régression) ou une validation par des lecteurs experts. La validation des deux premières étapes est toutefois une autre paire de manches : la formalisation est sujette à des choix arbitraires, et les concordances entre des caractéristiques mesurables et un concept peuvent être fortuites¹⁶.

Andrew Piper reconnaît que ces quatre premières étapes sont toutes d'une façon ou d'une autre des formes de réduction. La théorisation limite l'approche au texte en fonction de la question de recherche, la conceptualisation dénature les concepts pour les rendre mesurables, l'implémentation est une réduction littérale du texte à un modèle qui ne fait que représenter le texte (Erb, Ganahl, et Kilian 2016),

16. *Cum hoc ergo propter hoc*, corrélation n'est pas causalité.

et la sélection limite la recherche à un corpus et à ses règles propres. L’auteur note cependant que la réduction n’est pas unique aux humanités numériques : la généralisation en littérature est le résultat de considérer le particulier comme représentant le général¹⁷. Nous trouvons dans la littérature plusieurs exemples de modélisation automatique de la fiction à partir de minage de données, tels que la modélisation de réseaux de personnages dans l’oeuvre de Shakespeare (Iyyer et al. 2016) et les différences discursives formelles entre la fiction et la biographie (Underwood 2019).

L’exemple de Boris Iarkho Pour le formaliste russe Boris Iarkho, la biologie est la meilleure analogie scientifique pour une approche structurale de la littérature (Lvoff 2021 ; Gasparov 2016). La forme du cadre théorique adopté par ce formaliste permet d’effectuer des études comparatistes, de caractériser et classer des oeuvres, et de faire des découvertes sur les genres littéraires et la littérature. Ses travaux dénotent de l’importance des réseaux, relations, et structures dans la modélisation de la littérature dont l’étude en tant qu’objet vivant demande un cadre théorique capable d’envisager l’évolution, les variations infinies et les continuums. Iarkho utilise des définitions formelles pour caractériser les oeuvres, par exemple, il nomme *frequentative weight* (FW) le résultat du nombre de lignes d’un personnage divisé par le nombre de fois qu’il parle et *speech volume weight* le nombre de lignes d’un personnage divisé par le nombre de lignes de l’oeuvre (Wendell 2021).

$$FW = \frac{\textit{character_lines}}{\textit{character_utterances}}$$

$$SWW = \frac{\textit{character_lines}}{\textit{total_lines}}$$

17. “Measurement replaces charisma as the guiding vehicle of generalization. It difuses power, away from the persona (the proper name) and into a more dispersed array of technologies, techniques, and practices among which the individual is enmeshed (Latour).” (Piper 2017).

Ainsi, il peut définir des le *general dialogical weight* (GDW) d'un rôle comme étant :

$$GDW = \frac{FW + SWW}{2}$$

et fait la distinction entre les personnages ayant des scores supérieurs à 15.25 (les personnages principaux), entre 7.34 et 15.25 (secondaires), entre 7.33 et 2.1 (troisième rang), entre 0.1 et 2.09 (quatrième rang), et inférieur à 0.1 (sans dialogue). “Iarkho arrives at these ranges by looking at the role weight distributions and identifying quartiles” (Wendell 2021). L’analyse de Iarkho de la tragédie en cinq actes offre d’impressionnants résultats : à partir des données collectées et de définitions formelles, il réussit à différencier les pièces classiques des pièces romantiques, et à caractériser ces deux genres en fonction de ces mêmes caractéristiques. Qui plus est, son analyse permet de détecter les motifs d’évolution du théâtre au fil du temps : les caractéristiques bougent par vagues, d’auteurs en auteurs, avec des périodes de résurgence et de stabilisation (Lvoff 2021). Les progrès de Iarkho sont freinés d’abord par ses moyens techniques¹⁸ et ensuite son exil à Omsk à la suite de sa participation à l’écriture d’un dictionnaire russe-allemand jugé comme fasciste par la Cour suprême de l’U.R.S.S. (Gasparov 2016).

Caractéristiques de la lecture numérique

À l’inverse de la lecture humaine qui est diachronique, la lecture algorithmique peut être **synchronique** ; l’ensemble des textes d’un corpus peuvent être considéré à la fois et de la même façon. Ce synchronisme est parfois virtuel : certains algorithmes considèrent les textes les uns après les autres, mais leur ordre n’a pas d’importance (la même opération sera faite sur chaque texte), ou encore tous les ordres sont équivalents (les différences seront marginales si l’ordre des textes

18. Iarkho travaillait seul et à la main, en utilisant des méthodes statistiques ne nécessitant que du papier quadrillé et une règle (Lvoff 2021).

est changé). Par exemple, les calculs de Iarkho sur la tragédie en cinq actes ne changeront pas si l'ordre de deux pièces est inversé (Wendell 2021), alors que l'inversion de deux textes aura un effet sur la vectorisation de mot, mais les modèles générés par deux séquences différentes des mêmes textes seront aussi précis (Mikolov et al. 2013). La lecture algorithmique est génératrice de modèles, elle agit donc comme intermédiaire entre deux médias. Cette **intermédialité** signifie que la lecture algorithmique est réductrice, mais la nature mathématique de cette traduction la rend **modulable** : des nouveaux outils peuvent être ajoutés, les anciens peuvent être modifiés, de sorte que les modèles produits soient plus sophistiqués, spécialisés, efficaces, etc. La modulation des algorithmes de lecture artificielle peut mitiger la perte inhérente à l'intermédialité pour assurer une meilleure adéquation entre le texte et le modèle. Un autre aspect intrinsèque au numérique est sa capacité à considérer l'ensemble des textes de manière **exacte**, ce qui est à l'opposé du rapport humain au texte. Ainsi, la lecture algorithmique ne manque aucune concordance, du moins aucune concordance qui réponde à la définition que l'on en fait. L'**intertextualité** est primordiale dans le cadre de la lecture artificielle : l'ordinateur n'ayant pas accès à une représentation organique de la réalité, toute l'information à laquelle ce dernier a accès est encodée dans les textes que nous lui fournissons. Même lorsqu'un algorithme n'est mis au contact que d'un seul texte, le seul sens qu'il soit capable de produire vient de l'**intratextualité** que l'on peut résumer en la segmentation du texte en sous-sections. La lecture algorithmique produit donc des résultats à la manière d'un **comparatiste désincarné** : ses modèles ne sont pas le résultat d'une intention interne ou d'une force autonome (Searle 1980).

La connotation associée à l'expression lecture algorithmique laisse à croire qu'une telle entreprise peut être objective, mais c'est en fait loin d'être le cas, plusieurs niveaux de subjectivités s'y superposent. Certaines sources de subjectivités sont liées à la programmation du programme, d'autres à la nature même des algorithmes et du *hardware*, et finalement d'autres sont le résultat de choix éditoriaux dans la sélection du corpus (par le chercheur, ou à cause des facteurs

de disponibilité et d'indisponibilité).

It would be naive to suggest that measurements are devoid of power relations, but neither are the acts through which proper names account for knowledge. Models make us aware of the externalities through which we arrive at truth claims. (Piper 2017)

Une lecture artificielle n'est pas mue par sa propre volonté, mais elle n'en reste pas moins **téléologique** et **politique**. La programmation est informée par les buts, biais et présuppositions des chercheurs ; en récupérer les étapes de la modélisation d'Andrew Piper, nous voyons qu'elles sont toutes sujettes aux influences de ces trois éléments à divers degrés. De la théorisation à la validation, l'aspect téléologique de la programmation guide et resserre les choix, il contient et limite les possibilités. Les présuppositions du chercheur forment le cadre épistémique utilisé dans la programmation, ce qui influence l'angle d'approche de la question, mais aussi informe consciemment les choix de corpus et d'algorithmes¹⁹. Les biais sont plus facilement détectables dans les questions posées, les choix de corpus, et la validation des modèles. Les préjugés négatifs et partis pris crée un horizon pour les hypothèses valables et dignes d'intérêt, et pointent les chercheurs vers certains corpus alors que d'autres sont ignorés. Similairement, la lecture rapprochée et l'évaluation d'un corpus sont toutes deux affectées par les biais implicites du chercheur. C'est ainsi que la lecture algorithmique est politiquement chargée, que les chercheurs en soient conscients (l'analyse des discours publics d'Erdogan par Senem Aydın-Düzgit en est un bon exemple (Aydın-Düzgit 2016)) ou non (les choix de corpus des premiers projets de stylométrie révèlent un important biais favorable aux Pères fondateurs des États-Unis).

Pour plusieurs algorithmes, leur structure même génère une certaine chance dans la création de leur modèle. Les réseaux de neurones sont un bon exemple

¹⁹. Par exemple, la vectorisation de mots est une conséquence du structuralisme en linguistique (Wiedemann et Fedtke 2021).

d'introduction d'un état de base semi-aléatoire ; cet algorithme commence avec un modèle créé au hasard, et chaque itération de l'algorithme change les poids du modèle à la recherche d'une configuration donnant le meilleur résultat possible. Toutefois, les problèmes complexes possèdent plusieurs états pour lesquels l'algorithme est satisfait, des minimums locaux dans lesquels le réseau de neurones s'arrête sans savoir si d'autres points sont plus efficaces à accomplir la tâche en cours (plusieurs techniques permettent d'éviter les minimums locaux, voir (Gori et Tesi 1992)). Ce n'est pas là la seule forme de "subjectivité algorithmique", les méthodes basées sur les statistiques bayésiennes débutent également avec un état pseudo-aléatoire qui est soit une estimation humaine ou le résultat d'un algorithme simple d'approximation (Brownlee 2018). Finalement, plusieurs méthodes de classification ne fournissent pas des résultats définitifs durant leur tâche, mais offrent plutôt des probabilités d'appartenance à chaque catégorie. C'est là un exemple de *fuzzy mathematics*, une branche des mathématiques qui utilise des spectres non binaires, et où les objets peuvent appartenir à plusieurs catégories à la fois selon certaines proportions (Bonissone 1980). La linguistique quantitative utilise les *fuzzy sets* pour représenter les incertitudes et imprécisions associées à l'étude de la langue, une méthode plus efficace que la théorie probabiliste puisqu'elle permet à des éléments linguistiques d'avoir plusieurs identités sémantiques simultanément plutôt qu'une probabilité d'appartenir à l'une plutôt qu'à une autre (Bonissone 1980).

Chapitre II - *interpretatio*

Then began the epoch of interpretation, an epoch in which we are doubtless still living and from which one should rightly be allowed to expect (more so than from positivism and Geistesgeschichte) a new revival of hermeneutics, of the *ars interpretandi*. But what the epoch of interpretation took over from hermeneutics was little more than the name: the art of interpretation, and the concept of the hermeneutic circle. (Peter Szondi 1978)

Nous avons précédemment défini la lecture comme une modélisation textuelle, un cadre au sein duquel l'interprétation peut être approchée comme la création d'un méta-modèle : un travail d'inférence à partir de modèles littéraires ou extratextuels (sociohistorique, linguistique, biographique, etc.). Dans ce chapitre, nous montrons d'abord que l'abduction et, dans une moindre mesure, l'induction permettent d'imiter plusieurs tâches herméneutiques et interprétatives des littéraires. Ces formes d'inférences ont des équivalents algorithmiques parfois directs, l'induction par exemple est mathématisable selon certaines conditions, parfois indirects (nous discuterons des liens entre l'abduction et l'inférence bayésienne dans le dernier tiers du présent chapitre). Nous ne désirons pas réduire l'interprétation littéraire à la modélisation et des opérations logiques, notre argument est plutôt que des parallèles importants existent entre des formes limitées d'interprétations littéraires et des méthodes algorithmiques. En outre, pour légitimer notre

démarche, nous créons un parallèle similaire entre la biosémiotique, la sémiotique littéraire et la cybersémiotique, le but étant de réduire la distance entre la création de signes et de sens par un être vivant et une machine à une différence de degrés sur un même continuum d'évolution biotechnologique (Hayles 2019). Pratiquement, nous utilisons donc l'idée d'un méta-modèle comme le résultat d'inférences à partir des informations structurées générées par la modélisation de textes littéraires, ce méta-modèle peut prendre la forme de généralisations, d'une classification, de définitions, etc. (Fayyad, Piatetsky-Shapiro, et Smyth 1996). Les différences et particularités des modélisations humaines ou algorithmiques favorisent certaines structures et formes de modèle, dans un cas comme dans l'autre. Par exemple, la lecture émotionnelle humaine est encore difficile à répliquer mécaniquement pour des textes complexes (Nalisnick et Baird 2013a) alors que la production d'un réseau de concordance est beaucoup plus rapide avec un algorithme (Busa 1980). Ces différences humain-machine auront par ailleurs un impact sur le type d'inférences que les deux formes de cognitions (humaine ou computationnelle) peuvent produire. Notamment, les humains sont généralement incapables de trouver des tendances dans une base de données à plus de quatre dimensions alors que les méthodes algorithmiques excellent à ce genre de tâche (Fayyad, Piatetsky-Shapiro, et Smyth 1996). Les découvertes de la cognition ne sont pas toujours facilement formalisables, et nous ne parvenons pas toujours à rendre les découvertes de la computation intelligibles (Rhody 2012), mais notre utilisation du mot "sens" étant dissociée de la compréhension, nous approchons ces inférences et méta-modèles comme du possible sens herméneutique.

Cadre théorique

Dans son article *Pathologies of Epistemology in Literary Studies*, Rory Ryan décrit avec prudence le sillage du paradigme d'une autorité interprétative, qu'elle soit autoriale ou transcendantale. Selon Ryan, "It is difficult, having challenged objectivism, not to go 'the whole way' and allow pure subjectivism, pure chaos" (Ryan 1985). Nous avons précédemment déclaré que nous n'étions

pas en quête d'un algorithme permettant une interprétation absolue ou idéale du texte littéraire : nous ne cherchons pas à substituer la machine à l'auteur ou à Dieu, et rejetons la thèse d'une méthode capable de découvrir la vérité d'un texte littéraire. Parallèlement, nous ne pouvons pas nous résoudre à *allow pure subjectivism, pure chaos* ; l'interprétation doit être limitée selon des critères d'intelligibilité et de pertinence. Nous cherchons plutôt à théoriser le lecteur comme étant producteur de modèles, le lecteur algorithmique est alors un programme capable de générer des modèles littéraires complémentaires à partir de données intra-, inter-, et même extratextuelles. Nous ne pouvons nous assurer complètement du respect de ces critères en amont, considérant d'un côté que nos démarches sont expérimentales et (j'emprunte l'expression à (Rockwell et Sinclair 2016)) joueuses et de l'autre que ces critères sont difficiles à formaliser. Nos visées numériques sont limitées d'un côté par notre sujet d'étude, la littérature, et nos outils d'analyse : ces derniers peuvent produire des catégories de manière non supervisée, découvrir des règles internes pour un jeu de données, nous aider à formaliser des définitions, ou effectuer des prédictions pour des données incomplètes (Fayyad, Piatetsky-Shapiro, et Smyth 1996). Dans notre dernier chapitre, nous donnerons des exemples précis de projets de *Digital Criticism*, mais annonçons d'emblée que ceux-ci reposeront sur les avantages précis de la lecture (ou modélisation) et de l'interprétation numérique.

Des cinq étapes de la modélisation textuelle d'Andrew Piper présentées dans le précédent chapitre, quatre contribuent directement à rendre intelligible et pertinente une expérience herméneutique numérique²⁰. Les étapes de **conceptualisation** et de **théorisation** permettent de limiter et de diriger la modélisation littéraire en fonction d'un but, l'interprétation et la modélisation numérique sont téléologiques par nature. Si cela peut être suffisant pour assurer la pertinence des démarches, un processus cyclique de calibration et de reconsidération des étapes précédentes par la **validation** est nécessaire pour assurer l'intelligibilité de la modélisation. Ces allers-retours entre le chercheur et l'algorithme augmente

20. L'implémentation étant une étape technique, nous ne la considérons pas puisqu'elle dépend directement des autres étapes.

la marque de l'humain sur la modélisation, une relation qui peut multiplier les instances où les biais conscients et inconscients du chercheur dirigent l'algorithme vers les réponses ou paradigmes de l'humain. Une autre étape de la modélisation a un impact considérable sur la pertinence et, dans une mesure à peine moindre, sur l'intelligibilité d'une interprétation littéraire numérique : la **sélection** (Piper 2017 , 2018; Barthes 1966). Les textes passés aux méthodes de modélisation et dont les modèles sont étudiés informent plus que le domaine de découverte : ceux-ci forment l'horizon sémiotique de l'algorithme, son *umwelt* (Hayles 2019 ; Uexküll 1982). Le *umwelt* peut être supplémenté par des modèles sur le monde, tels que des informations biographiques des auteurs, sociohistoriques, linguistiques, culturelles, etc. La portée et la forme de cet horizon module les inférences et types d'inférences possibles ; l'interprétation littéraire numérique est confinée par les limites du réseau de signe qui est présenté à l'algorithme. Notre définition de sens et la sémiotique nous servent d'un côté de justification : puisque nous approchons le texte comme un signe complexe, nous sommes en mesure d'évacuer le sens littéral et plusieurs autres considérations linguistiques ; mais aussi à accentuer la continuité entre l'humain et la machine. En effet, pour démontrer certaines équivalences entre la computation et la cognition, il est plus simple de se baser sur un cadre théorique pouvant traiter les deux similairement. Pour mettre en relation les trois médias que sont la vie cellulaire²¹, la cognition et la computation, nous devons les placer sur le même continuum d'évolution biotechnologique ; cela signifie isoler le signe comme un phénomène émergent de la matière. L'inscription de la sémiotique littéraire dans le cadre de la bio et de la cybersémiotique nous permet finalement d'appréhender le sens de textes littéraires à partir de méthodes formelles et algorithmiques.

Bio et cybersémiotique

Le signe, plus petite unité sémiotique, ne possède pas de propriété sémantique ; il est, au sens informatique, un pointeur qu'un interprétant aura le loisir

21. Nous incluons la biosémiotique puisqu'elle nous permet de commencer notre description à son niveau le plus simple, ainsi que pour démontrer la continuité entre les trois médias.

de décrypter comme pointant vers différentes propriétés sémantiques (Burton-Roberts 2013). Nos intérêts de recherche nous amènent à nous concentrer sur le signe de type symbolique et à un interprétant qui raisonne par inférence, que ce dernier soit humain ou machine. Nous nous intéressons également au signifiant ; toutefois, la définition de Ferdinand de Saussure est limitante : il fait du signifiant un objet, une action, un processus ; dans le cadre linguistique, nous parlerions de d'une propriété sémantique du sens, le sens littéral (Searle et Latraverse 1979). Notre approche au signifiant est plus complexe, d'abord à cause de l'importance que nous donnons à l'interprétant ; nous avons théorisé le texte comme générateur d'un espace sémiotique qui est resserré au contact d'un lecteur puisque celui-ci est détenteur d'un réseau de signes qui entre en dialogue avec celui du texte (Culler 1976). Ainsi, les sens d'un texte ne sont pas limités aux exemples saussuriens, le texte peut pointer vers des expériences et concepts qui ne sont pas encore linguistiquement établis et stables, et même des idées qui n'entrent pas dans les paradigmes expérientiels et conceptuels. Appréhender le sens comme étant le résultat du contact entre deux réseaux de signes nous amène aux cadres théoriques de Terrence Deacon et de N. Katherine Hayles, le premier décrivant la production de sens comme une dynamique (Deacon 2011), et la seconde comme une intermédiation (Hayles 2019). Dans *Incomplete Nature*, Deacon classe différents types de système en fonction de leurs rapports aux contraintes et à l'anticipation. Selon lui, un système producteur de sens renforce les contraintes qui rendent possibles son existence et développe des mécanismes de réaction à son environnement qui arrivent à faire un lien entre un phénomène absent et un phénomène présent (Deacon 2011). Ces liens sont, en termes biosémiotiques, des signes, et le sens est un réseau complexe de ceux-ci. L'approche de Hayles est plus libre puisqu'elle ne se limite pas à la biologie ; elle utilise plutôt le principe de médias : entre différents niveaux d'organisation de la matière, les interactions mènent à la création de signes par anticipation. Similairement à Deacon, elle place le sens dans le système complexe de signes qui permettent la représentation et la découverte de connaissance dans d'autres niveaux d'organisation (d'autres médias) (Hayles 2019). Dans un cas comme

dans l'autre, la relation entre le signe et le sens se trouve dans deux aspects de l'interprétant : ses caractéristiques (nous les avons décrites dans le premier chapitre pour un lecteur humain ou algorithmique) et le réseau de signes dont il est porteur. Ce type de relations entre (et d'inférences à partir de) deux réseaux ou systèmes est théorisé dans la biosémiotique, la sémiotique littéraire, et la cybersémiotique : le même processus est présent dans ces trois médias (Hayles 2019). Nous utilisons ici les définitions de Hayles : soit qu'un média est un niveau d'organisation de la matière et le sens est une relation comportementale (ou décisionnelle) entre un interprétant et un réseau de signes (Hayles 2019) ; et de Deacon : soit qu'un signe est une relation entre ce qui est présent et absent, produite par un système téléodynamique de contraintes orthogrades (Deacon 2011). En d'autres mots, pour qu'un signe existe, il doit être généré et entretenu par un système récursif qui reproduit ses propres contraintes et y génère des corrélations (Deacon 2011).

L'explication de Deacon est destinée à expliquer la formation de signes chez des êtres vivants, mais Hayles (dont le cadre théorique substitue les systèmes pour les médias) utilise ces définitions (et le cadre de l'intermédiation) pour démontrer qu'un programme informatique est également à même de produire des triades sémiotiques et des réseaux entre ces dernières. Hayles écrit :

As a relational operator, a sign creates possibilities for meaning making because it brings a behavior (the interpretant) into relation with the object, the absential phenomenon for which the sign stands.
(Hayles 2019)

Dans le cadre de la biosémiotique, Wendy Wheeler suggère que le sens est une fonction ; ce type de relation au monde est donc lié au potentiel de transformation de l'environnement par l'organisme :

The *meaning* of something is to be discovered in what it *does* in the world, in how it allows things to be and also to change. Thus we say

that a biological meaning is a function. (Wheeler 2016)

Cette approche au sens n'est toutefois pas conclusive dans le cadre de la cybersémiotique, d'abord à cause des particularités du numérique, mais aussi à cause du biocentrisme (ou de l'anthropocentrisme) sous-entendu par Wheeler (Hayles 2019). Pour répondre à Wheeler, Hayles se tourne vers une analogie à la compréhension :

With design and purpose displacing biological imperatives in the First Great Inversion²², a computer achieves meaning in this behavioral sense when it processes an algorithm, reads a data set, performs the calculations indicated, and produces results that it “understands” through its anticipations, interpretations, and information flows. (Hayles 2019)

Elle reste prudente, en mettant d'abord le terme *understand* entre guillemets, mais aussi en spécifiant que le *computational media* n'a pas de compréhension du même registre que la compréhension humaine ;

they are capable of meaning-making practices [...], *performing* actions in response to internal and external cues, making interpretations, and constructing relations between what is present and what is absent through their anticipations and operating constraints (Hayles 2019)

Pour Burton-Roberts le sens tel que produit par un interprétant “depends on [the subject] making inferences based on her beliefs” (Burton-Roberts 2013). Certains algorithmes peuvent produire des inférences, mais qu'en est-il de la croyance? Pour l'être humain, l'abduction est une inférence basée sur les croyances d'un individu, mais si l'abduction est formalisable, elle est difficile à mathématiser. En conceptualisant la croyance comme un degré de certitude à l'aide de l'inférence

22. La première grande inversion est définie ainsi par Hayles : “As humans continue evolving by implementing sign relations in computational media, design and purpose substitute for the biological imperative of survival and reproduction” (Hayles 2019)

bayésienne, nous obtenons un équivalent numérique à l'abduction. Les statistiques bayésiennes se basent sur des probabilités de vérité, et dont les présuppositions sont mises à jour au contact de nouvelles observations, ce qui est différent de l'abduction²³, mais valable dans la construction de connaissance. L'humain comme l'algorithme rencontre ses modèles tirés de la lecture à l'aide de la logique, mais cette dernière ne saurait être déductive, plutôt, des conclusions probabilistes et des inférences "to the best explanation" (Douven 2021) servent à générer du sens des modèles. Il suivra que l'algorithme est producteur de sens puisqu'il peut être mené à simuler la production humaine de sens sous la forme d'inférences guidée par des "croyances" dans le contexte de réseaux complexes de signes. Plus précisément, des méta-signes pertinents et intelligibles (au moins dans le cadre de la cybersémiotiques) peuvent être générés à partir du média numérique en se basant sur des signes dont la production rappellent le rapport du vivant à un environnement. Une partie de cette équivalence vient de la possibilité de simuler la croyance comme étant un degré de certitude, une méthode statistique basée sur le théorème de Baye (Weisberg 2009). Le sujet de la compatibilité entre l'inférence bayésienne et l'abduction est un sujet de débat dans lequel nous ne nous positionnerons pas ; cette compatibilité n'est pas nécessaire à ce que nous parlions d'équivalence et, de plus, nous ne nous basons pas sur les mêmes critères épistémologiques que la science (Douven (2021) ; (Weisberg 2009)). Notre agnosticisme dans cette discussion est également motivé par notre décision de ne pas formaliser l'inférence humaine (abduction ou induction)²⁴, la prochaine sous-section met donc l'accent sur le passage, grâce des inférences (rhétoriques ou logiques), de modèles de signes à un méta-modèle.

Méta-modèles et inférences littéraires

La modélisation de textes littéraires est un sujet établi des humanités numériques (Moretti 2013 ; Piper 2018 , 2017 ; Underwood 2019 ; Nalisnick et Baird 2013a), et

23. Ce sujet sera développé dans la section sur l'inférence bayésienne dans le dernier tiers du présent chapitre.

24. Nous ne donnerons pas plus que de définitions mathématiques très simples dans la prochaine section de ce chapitre sur l'interprétation humaine.

nous y retournerons dans notre dernier chapitre. L'analyse de modèles littéraires créés par des humains ont tendance à être aussi faite par des humains, bien que l'on trouve dans la littérature des exemples de représentations humaines analysées à l'aide d'algorithmes (Moretti 2013) et de modèles numériques étudiés par des humains (Nalisnick et Baird 2013a). Nous décrirons dans notre dernier chapitre ce genre de recherche collaborative humain-machine comme du *computer-assisted criticism* (Rockwell et Sinclair 2016). Plus généralement, une herméneutique algorithmique peut être conceptualisée comme une modélisation littéraire suivie de l'analyse du modèle pour générer un méta-modèle ; c'est là le paradigme précédemment présenté de création de signe, puis de sens. Qu'est-ce qui nous permet de parler de sens littéraire en référant à l'inférence algorithmique basée sur des modèles générés automatiquement? D'abord, le fait que ce travail herméneutique n'est pas fait en vase clos, mais plutôt le résultat d'une collaboration dialogique entre l'humain et l'ordinateur : les découvertes sont le résultat de cette rencontre et non de la mécanique seule (Rockwell et Sinclair 2016 ; Piper 2017 ; Moretti 2013 ; Ramsay 2011). La supervision humaine offre une certaine validité à notre démarche ; après tout l'humaniste connaît les objectifs, les limites, et la portée des études littéraires et reste central à la recherche ; le littéraire doit être en mesure de discerner les résultats pertinents et les rendre intelligibles. Ensuite, et de manière peut-être plus intéressante, parce que nous avons la possibilité de rejeter une conception anthropocentrique de l'herméneutique et d'appréhender de nouvelles avenues de recherche (Hayles 2019 ; Ramsay 2011 ; Binder 2016). Ce point est paradoxal considérant que nous venons de mettre l'accent sur la relation humain-machine comme étant une simple extension de la cognition, mais les structures de recherche proposées dans le dernier chapitre offriront des moyens pour pousser les limites de la cognition en terrains inconnus.

Certains intérêts de recherche des littéraires peuvent être réduits à des inférences, soit à l'abduction, à l'induction²⁵, et à l'inférence bayésienne. Ces intérêts sont la généralisation à partir du particulier (Piper 2020 ; Moretti 2013 ; Underwood

25. Nous ferons le cas que la nature de la littérature et de la recherche littéraire rendent difficiles l'induction, et que cette dernière est souvent une forme d'abduction déguisée.

2019), le positionnement d'une oeuvre dans un corpus (incluant la classification) (Wendell 2021 ; Nalisnick et Baird 2013b), et recherche de sens d'un texte à partir de modèles sociohistoriques, psychologiques ou philosophiques (Tracey et Morrow 2006). Dans ces cas (et d'autres que nous ne considérons pas directement), des observations sont combinées pour découvrir des classes de connaissances d'une autre nature, c'est un cas de double intermédiation selon la définition qu'en fait N. Katherine Hayles (Hayles 2019). En effet, les observations qui sont faites du texte sont le résultat d'une médiation, nous avons d'ailleurs mis l'accent sur cet aspect de la lecture en la qualifiant de modélisation. L'intermédiation est synonyme de perte d'information, ainsi, il est nécessaire de rappeler que l'étude d'un modèle n'est pas l'étude du texte (Erb, Ganahl, et Kilian 2016), à ce sujet dans le cadre de la biosémiotique, Hayles écrit :

In informational terms, the specific levels formed by physicochemical media always contain more information within the level than they communicate upward or downward. Indeed, it is precisely this reserve of dynamic information that characterizes the levels as such. (Hayles 2019)

Prenant l'exemple des particules subatomiques, elle ajoute que certaines caractéristiques d'un média (que nous pourrions qualifier de niveau d'émergence de la matière) ne sont visibles que dans un média adjacent ; l'émergence de nouvelles qualités est couplée à la découverte de qualité invisible aux niveaux précédents (Hayles 2019). Ainsi, les modèles littéraires produits ne sont pas les textes dont ils sont issus, mais lorsque nous considérons ce nouveau paradigme, des informations d'un nouvel ordre qui réfèrent au texte deviennent visibles (Erb, Ganahl, et Kilian 2016 ; Piper 2018 ; Jockers 2013). C'est le cas des genres littéraires qui ne peuvent être générés au niveau des textes eux-mêmes, et demandent une comparaison de modèles textuels, tels que la longueur, les émotions produites, les archétypes utilisés, les structures narratives, etc. C'est à partir de ces informations qu'un nouveau modèle (un méta-modèle) peut générer

des catégories à l'aide de l'abduction : le genre littéraire est une explication des différences et du groupage entre les textes qui n'est pas fondamentale, mais qui explique au mieux le phénomène observé (Douven 2021). Dans cette optique, notre choix de considérer le texte comme un assemblage de signes est à nouveau révélé comme n'étant pas innocent ; d'abord puisque la sémiotique nous permet de mettre l'accent sur l'inter- et l'intratextualité, et puis parce que considérer le signe comme un niveau d'émergence de la matière nous permet de théoriser le sens à son tour comme un niveau d'émergence (Hayles 2019 ; Deacon 2011). La sémiotique nous permet ainsi de ne pas nous frotter aux problèmes de la sémantique ; un dossier que nous préférons laisser aux linguistes et aux philosophes. L'herméneutique littéraire numérique que nous avons mise de côté dans le dernier chapitre refait ainsi surface dans le présent chapitre ; notre but final est, après tout, une contribution au champ du *Digital Criticism*. N. Katherine Hayles, à qui nous empruntons plusieurs définitions en ce qui a trait au sens, écrivait en 2019 : "Accepting that meaning making is not an exclusively human prerogative is a crucial step in the right direction" (Hayles 2019). Il s'agit à la fois d'un déplacement des questions de Searle et de Turing (Searle 1980 ; Turing 1950) sur la cognition et la compréhension, que nous mettons complètement de côté au profit d'un autre type de réconciliation humain-machine : en plaçant l'un comme l'autre sur le même continuum de l'évolution biotechnologique, nous avons un cadre théorique pour l'interprétation de texte littéraire qui couvre les deux à la fois.

In broad stroke, the first evolutionary leap bootstrapped complexity from simpler mechanistic reactions; the second emerged from the first, bootstrapping meaning-making practices from complexity; the third emerged from the second, bootstrapping artificial cognition from biologically derived signs and meanings. (Hayles 2019)

Cette continuité permet à Hayles de théoriser le signe et le sens en termes d'émergence au passage de différents niveaux d'*intermediation* qu'elle définit

ainsi :

intermediation can be understood as an alternation between elements considered as individuals, which interact among themselves to create the quasi-stable patterns that are incorporated into the emergence of individuals at a higher level of complexity, creating a process that results in increasingly complex patterns. (Hayles 2019)

Hayles, s'appuyant sur les catégories de Terrence Deacon, détermine que l'algorithme est une *meaning-making practice*, mais ajoute également que les niveaux d'émergence se succèdent en ordre de complexité (soit en ordre d'évolution). Grâce aux règles statistiques et à la généralisation, les médias supérieurs en ordre d'évolution ne sont pas nécessairement plus complexes, mais simplement différents, similairement, deux médias peuvent occuper la même strate de l'évolution biotechnologique tout en étant deux formes d'organisation de la matière entièrement différentes. Nous avons relevé une erreur similaire chez Wendy Wheeler qui définissait le sens comme une fonction ; dans un cas comme dans l'autre, le biocentrisme est en cause, ce que nous ne pouvons pas lui reprocher considérant les limites de son sujet d'étude (Wheeler 2016). Le concept de média est tout de même plus adapté à notre contexte puisque nous ne sommes pas toujours en mesure de quantifier le différentiel de complexité entre un lecteur humain et un algorithme, ou encore entre deux algorithmes²⁶. L'avantage de ce cadre théorique est toutefois teinté d'un compromis majeur : bien que plusieurs de leurs démarches puissent être réduites en termes de modélisation et d'inférence, ce n'est pas ainsi que les chercheurs pensent à la recherche. La prochaine section fait donc le cas qu'une part de la recherche herméneutique et littéraire utilise la modélisation et l'inférence, une recherche que nous assimilons au sens dans sa définition limitée précédemment présentée. En mettant l'accent sur les tâches que l'humain et l'ordinateur peuvent entreprendre, les possibilités du numérique

26. Autrement dit, le passage entre deux formes de représentations symboliques n'augmente pas forcément la complexité des motifs et de l'information, tout en répondant aux règles de l'intermédiation.

et ses différences avec le biologique n'en deviennent que plus apparentes.

Interprétations humaines

Hermeneutics is both science and art. (Robinson 1995)

Dans sa définition normative, le sens textuel est une propriété sémantique et communicative (Burton-Roberts 2013) : un sujet émetteur a encodé une information qu'un décodeur saurait plus tard décoder (Hall 1999) ; l'intention de l'auteur (rendre une information accessible) se reflète dans l'intention de l'interprétant (accéder à l'information), et le message n'est pas changé par le processus. Cela présuppose une symétrie d'encodage (linguistique, culturelle, etc.), des propriétés sémantiques linguistiques, et une intention de communication d'une information qui vaut la peine d'être encodée, et qui, par le fait qu'elle a été encodée, signale qu'elle vaut la peine d'être décodée. Ces trois aspects sont la cible de critique dans les prochains paragraphes, les deux premiers à cause de principes premiers ou universels, et le troisième dans le cadre de la littérature.

Au-delà du sens littéral

Selon Barthes, le sens littéral est codifié en fonction de la phonétique, de la phonologie, de la grammaire et du contexte (Barthes 1981) ; une vision du sens qui présuppose que le texte possède des propriétés sémantiques (Olsen 1982), ce qui est lourd de conséquences. En effet, en attribuant de la sémantique au texte, ce dernier devient chargé d'un "bon" sens, d'un sens "correct", et le pouvoir d'interprétation est enlevé au lecteur au profit d'un système interprétatif²⁷. Si certaines formes textuelles présupposent l'importance d'une communication la plus transparente possible (les textes légaux, traités mathématiques et manuels d'assemblage IKEA nous viennent en tête), ce n'est pas le cas de la littérature

27. Judith Butler décrit une langue ne permettant qu'une interprétation comme une langue morte dans *Excitable Speech*. Elle cite abondamment le discours d'acceptation du prix Nobel de littérature de Tony Morrison pour caractériser la langue vivante comme supportant une infinité d'interprétations (Butler 2017).

qui est dénaturée par l'utilisation d'un système rigide d'interprétation (Butler 2017). Une communication parfaite et directe par le texte présuppose également la transparence de l'émetteur ; ce dernier est pourtant en mesure d'encoder des informations sans en être conscient (Butler 2005), ou encore d'encoder des informations sans désirer les voir décoder par tous (Hall 1999). L'opacité de l'émetteur est supplémentée par l'opacité de l'interprétant ; de nombreux facteurs inconscients informent l'interprétation d'un texte de manière imprévisible pour l'émetteur. À ces biais inconscients s'ajoutent les différences linguistiques et culturelles entre les sujets qui rendent impossibles une symétrie entre les processus d'encodage et de décodage de l'information : un interprétant peut décoder un message de manière subversive ou contraires aux intentions de l'émetteur. Pour reprendre les catégories de décodage de Stuart Hall ; toute interprétation est au moins partiellement négociée (Hall 1999 ; Aligwe, Nwafor, et Alegu 2018).

L'asymétrie d'encodage est une caractéristique obligatoire de la communication, et dans le cas de la littérature, un trait assumé et même central. Nous reprenons ici la définition d'Olsen que nous avons précédemment utilisée : “a literary work is a discourse in which an important part of the meaning is implicit” (Olsen 1982). Des signes sociaux et linguistiques contribuent à différencier un texte dont le but est une communication claire et les textes littéraires ; le contexte du texte, son médium, et sa présentation contribue à indiquer le type d'exercice herméneutique qui est attendu au contact d'un texte, mais aussi des différences stylistiques qui sont facilement repérables par un algorithme (Underwood 2019). Si l'interprétation textuelle usuelle demande que le lecteur présuppose que le texte est porteur de sens et que l'information qui y a été encodée vaut la peine d'être décodée, l'interprétation littéraire est quelque peu différente. Nonobstant les lectures non-littéraires et les démarches absolutistes²⁸, le lecteur présume que le texte est porteur de plusieurs sens et que l'acte de décodage, plutôt que l'information qui en est retirée, vaut la peine. Le plaisir de l'interprétation littéraire est dérivé de deux aspects de la littérature, soit

28. Par exemple, l'originalisme légal aux États-Unis, l'exégèse historique biblique et l'étude du sens autorial.

l'esthétisme et la satisfaction du vide sémiotique (Culler 1976). Nous ne traitons pas de l'esthétisme dans le présent mémoire, mais le besoin de combler un vide sémiotique donne à l'interprétation un aspect téléologique que nous retrouvons chez le lecteur humain et l'algorithme. Ce vide combiné à la multiplicité de sens force l'utilisation de modèles précis pour permettre de faire s'effondrer les possibilités sémiotiques au profit d'un nombre limité d'interprétations. Le dernier aspect du sens littéral que nous nous devons d'évacuer est celui de l'adéquation entre la linguistique et la sémantique ; un sujet que nous avons déjà touché dans nos présuppositions. Comme le note Noel Burton-Roberts, la langue n'est pas la seule dépositaire du sens (Burton-Roberts 2013), ce qui indique que le sens n'est pas trouvé au même niveau que le texte, il s'agit d'un média différent. Selon Barthes, les unités sémantiques du texte servent une fonction supérieure, qu'elle soit narrative ou discursive (Barthes 1981) ; mais en s'arrêtant à la phrase comme unité sémantique, il commet une erreur : il confond les propriétés des conventions linguistiques avec le sens (Burton-Roberts 2013). Nous dissociions complètement les signes écrits de propriétés sémantiques, pour les traiter comme des signes qui ne pointent vers des propriétés sémantiques que dans le contexte de réseaux de signes. Le sens est produit par conventions linguistiques : le texte n'a pas de sens, mais plutôt un potentiel de sens actualisé au contact de l'interprétant (Atkin 2013 ; Eco 1976). Le texte, le modèle et l'interprétation, sont trois médias différents, liés ensemble par des contraintes qui permettent l'émergence de l'un par l'autre, une émergence qui peut être dissociée de la sémantique au profit de la sémiotique (Hayles 2019).

Abduction littéraire et autres inférences

Trois tâches communes du littéraire²⁹ reposent sur une combinaison de transformations (la modélisation), associations (intra- et inter- et extratextuelles) et raisonnements (logiques, rhétoriques et intuitifs). Ces raisonnements prennent

29. La classification, la généralisation et l'analyse (ou caractérisation) d'une oeuvre dans un cadre précis. D'autres formes d'analyse répondent aussi à ce modèle, mais nous n'y toucherons pas directement dans le présent mémoire.

la forme d'inférences qui, dans le cadre de la recherche littéraire humaine, sont surtout des abductions et des inductions. Plusieurs structures rhétoriques communes révèlent des traces de la logique formelle ; par exemple, la thèse supportée par des arguments est un calque direct de l'induction. Cette dernière cherche à généraliser à partir de cas particulier des règles générales et, au contraire de la déduction qui mène aux conséquences obligatoires de ses arguments, l'induction offre un degré de support pour ses conclusions (Hawthorne 2021). L'induction est mathématisable et formalisable, mais il n'est pas nécessaire d'employer la logique formelle pour l'utiliser ; l'induction peut supporter des conclusions de manière qualitative à l'aide d'arguments clairement binaires ou probabilistes³⁰. Toutefois, les arguments de la littérature ne se prêtent pas tous bien au paradigme binaire vrai-faux et probabiliste de la logique formelle, certains sont conditionnels et la majorité sont dépendants des modes de modélisations textuels utilisés. C'est ainsi que, bien que la structure logique de l'induction soit souvent répliquée dans la recherche littéraire, cette dernière utilise plus souvent l'abduction ; une forme d'inférence que nous utilisons dans la vie de tous les jours de manière instinctive, mais qui a aussi une place centrale dans les démarches scientifiques. En 1998, Niiniluoto formulait ainsi l'abduction :

Given evidence E and candidate explanations H_1, \dots, H_n of E , if H_i explains E better than any of the other hypotheses, infer that H_i is closer to the truth than any of the other hypotheses. (Douven 2021)

L'abduction est une recherche de la meilleure explication pour une série d'observations incomplètes, ce qui est d'une grande importance dans la recherche littéraire où les textes disponibles ne sont jamais l'ensemble des textes possibles, et où un texte est considéré "terminé" quand il est publié plutôt que lorsqu'il peut prétendre être achevé³¹. Comme nous l'avons précédemment mentionné, l'abduction est utilisée dans la vie de tous les jours pour produire des conclusions

³⁰. C'est d'ailleurs là une structure rhétorique des plus connues et enseignées dans les cours de littérature : celle d'une thèse globale supportée par une série d'arguments particuliers.

³¹. Pour citer à nouveau Paul Valéry : "un ouvrage n'est jamais achevé, – mot qui pour eux n'a aucun sens, – mais abandonné".

; Igor Douven exemplifie l'ubiquité de l'abduction dans son entrée sur ce sujet dans *The Stanford Encyclopedia of Philosophy* :

You happen to know that Tim and Harry have recently had a terrible row that ended their friendship. Now someone tells you that she just saw Tim and Harry jogging together. The best explanation for this that you can think of is that they made up. You conclude that they are friends again. (Douven 2021)

Similairement, il montre comment l'abduction est utilisée dans le domaine scientifique en prenant pour exemple les écrits de Sir Joseph John Thomson :

As the cathode rays carry a charge of negative electricity, are deflected by an electrostatic force as if they were negatively electrified, and are acted on by a magnetic force in just the way in which this force would act on a negatively electrified body moving along the path of these rays, I can see no escape from the conclusion that they are charges of negative electricity carried by particles of matter. (Thomson 1897)

Cette réduction à la meilleure explication se trouve aussi dans les études littéraires ; par exemple, la division entre les comédies et les tragédies de Shakespeare reconnaît certaines caractéristiques de ses pièces pour en tirer des conclusions (des catégories). C'est ainsi que des observations sur les pièces se transforment en règles et en paradigmes : 1. Nous observons que plusieurs des pièces de Shakespeare se terminent par un mariage, nous observons que la concentration de jeux de mots à caractère sexuel est haute dans certaines pièces de Shakespeare ; 2. Nous trouvons une corrélation entre les pièces de la première et de la seconde observation ; 3. Nous offrons comme explication que ces pièces font partie de la même catégorie que nous nommons "les comédies" (Barnet 1998). Nous sommes dans l'obligation de faire des inférences à partir d'observations incomplètes, d'un côté à cause des limites de nos modélisations, et d'un autre parce que l'ensemble des textes disponibles pour analyse est toujours incomplet.

Certains textes ne sont pas disponibles, certains textes ont été détruits, d'autres encore n'ont jamais été écrits, une problématique à laquelle il est possible de répondre en limitant l'étude à un corpus bien construit et, surtout, définit en fonction des barrières à la production, survie et disponibilité des textes. L'induction et l'abduction fournissent des outils logiques et rhétoriques malgré les limites des arguments disponibles pour ces inférences. Ces arguments sont contingents aux modèles littéraires produits, les conclusions le sont également ; les caractéristiques de la lecture humaine reviennent ainsi former les limites et possibilités de l'interprétation littéraire humaine.

Lecture humaine : conséquences interprétatives

Cette présente section a cherché à montrer les répercussions de l'asymétrie d'encodage, des limites des corpus, et de la sémiotique littéraire sur l'interprétation textuelle ; nous présentons maintenant comment les caractéristiques de la lecture humaine présentée dans le dernier chapitre influencent l'interprétation littéraire. Nous rappelons que ces caractéristiques sont des points saillants de la lecture qui servent de points de comparaison et de démarcation avec l'interprétation numérique. C'est parce qu'elles limitent, formes, permettent et informent les capacités herméneutiques de l'humain sur plusieurs plans différents que les frontières de l'expérience humaine prennent autant de place dans notre analyse (et parce que les frontières de la computation sont radicalement différentes). D'abord, l'interprétation, comme la lecture, est située dans le temps, elle fait donc partie d'une série d'interprétations qui s'entre-affectent : les leçons tirées de l'interprétation d'un texte suivent l'interprétant (à divers degrés) dans ses interprétations subséquentes. Cela est dû au fait que l'interprétant a accès aux modèles précédemment produits dans ses analyses, ces derniers l'informent méthodologiquement (différentes approches herméneutiques auront pu être testées et raffinées) et ont également permis de raffiner ses modèles non littéraires. Qui plus est, l'interprétant peut utiliser les catégories, inférences et connaissances dérivées de son *umwelt* littéraire (qui ne se limite pas aux modèles littéraires qu'il

porte, mais aussi aux inférences préalablement produites) pour appréhender le texte comme un réseau de signes vers des textes et concepts littéraires. Puisque les lectures sont ordonnées, les interprétations littéraires dépendent de facteurs semi-aléatoires ; la toile ainsi formée est unique à un interprétant dans un lieu et temps précis. Ces deux derniers éléments sont le résultat de la diachronie de la lecture, un concept qui ne se traduit pas directement dans le cadre de l'interprétation ; il est possible d'interpréter plusieurs modèles de textes littéraires à la fois, mais il n'est pas possible de produire plusieurs interprétations simultanément. Aux limites causées par la séquentialité de l'expérience humaine s'ajoute une autre caractéristique : l'introduction d'une distance temporelle entre les différents actes herméneutiques produit des déformations et des oublis dans les outils interprétatifs et méta-modèles générés par le passé par un interprétant. Les connaissances et inférences sont donc non seulement déformées par les interprétations subséquentes, mais également par les limites biologiques de la mémoire. D'autres facteurs contribuent à modifier l'interprétation, cette fois alors que l'interprétant accède les modèles littéraires ; ces derniers sont composites et certains des éléments de la lecture ont des effets sur l'état de conscience du lecteur qui peuvent avoir des effets permanents sur l'interprétant ou qui se répètent lorsque celui-ci réactualise son modèle d'une oeuvre qui l'a touché. Par exemple, l'affect émotionnel peut mener à oublier des aspects d'un texte (ou même d'oublier ce dernier entièrement), à mettre l'accent de manière quasi-arbitraire sur des éléments, et modéliser des parties du texte à l'aide d'outils herméneutiques inusités. En plus de l'affect, la relation sociohistorique de l'interprétant au texte change son rapport à ce dernier ; les lectures qui tendent vers l'hégémonique mènent à des interprétations radicalement différentes des lectures oppositionnelles (Hall 1999). Tous ces facteurs moulent les formes d'intertextualité et d'intratextualité que l'interprétant sera à même de convoquer et d'utiliser dans le cadre d'inférences. Le potentiel sémiotique des signes dépend des éléments listés dans le présent paragraphe, et il s'ensuit que les modèles littéraires en sont affectés, tout comme les méta-modèles et inférences qui en découlent. Pour paraphraser dans le jargon informatique ce que Judith Butler

répondrait aux obligatoires et importantes approximations qui sont ici décrites “It’s not a bug, it’s a feature!” L’herméneutique est après tout à la rencontre de la vérité et de la méthode, de l’art et de la science ; l’unicité des interprétations littéraires est centrale à la recherche littéraire.

Interprétation numérique

Qu’elle soit le fruit d’un humain ou d’un ordinateur, l’interprétation est un acte téléologique ; mais l’origine relatif des objectifs herméneutiques humains est différent de celle de l’algorithme (Searle 1980). L’interprétation humaine est motivée intrinsèquement, et elle dépend autant de processus conscients qu’inconscients pour les outils qu’elle emploie et ses motivations. Même si certaines traditions herméneutiques déconstruisent activement l’acte interprétatif en fonction de l’intra-, inter- et extratextualité, des facteurs arbitraires, informels, ou inconscients restent fermement enracinés dans les activités humaines (Butler 2005). Dans le cas de l’interprétation numérique, les motivations et biais ne sont pas internes, mais plutôt assujettis aux objectifs du chercheur, à ses biais inconscients, et aux conséquences inattendues de l’erreur humaine, multipliées à chaque itération des calculs touchés (Shah 2018). Cela signifie que l’aspect téléologique d’une interprétation algorithme doit être rendu la plus transparente possible, le chercheur doit rendre visible dans la pleine mesure de ses moyens ses décisions (Ramsay 2011). Plus important encore, les objectifs du chercheur, et par procuration de l’algorithme, doivent être formels, ce qui limite la portée téléologique de l’interprétation textuelle artificielle ; certaines formes d’objectifs littéraires n’ont pas encore su être formalisés. Cela est en partie dû au fait que l’approche du littéraire au texte n’est pas d’abord téléologique, elle ne le devient que lorsque ce dernier décide de porter son attention vers un texte en particulier dans un cadre académique ; les textes sont d’abord rencontrés pour le plaisir de la lecture et l’extension de son réseau sémiotique littéraire. L’interprétation numérique est donc contingente à la décision du chercheur d’étendre son contact au texte ou au corpus, c’est-à-dire que ces efforts herméneutiques

sous-entendent que quelqu'un croit que les textes traités en conjonction avec les méthodes utilisées sont à même de produire du sens pertinent et intelligible. La computation, à l'inverse de la cognition, peut rester stable dans le temps, par exemple en conservant parfaitement ses modèles littéraires, sociohistoriques ou linguistiques durant l'ensemble de son processus, ce qui ouvre la porte à certains enjeux éthiques et épistémiques : le moindre biais dans un de ces modèles voit son influence décuplée par l'application rigide et à grande échelle de méthodes d'inférence³². Ces différences permettent de s'assurer que tant qu'une forme d'information est portée par les modèles littéraires de la lecture numérique, et que cette dernière est reconnaissable par l'algorithme d'inférence, chaque instance sera de ce genre d'information sera intégré à l'analyse. C'est là un avantage important de l'interprétation numérique ; les capacités de reconnaissance et de mémoire de l'ordinateur permettent une intratextualité et une intertextualité qui ne sont pas approximatives (Ramsay 2011). Si les modèles qui lui sont fournis permettent à l'interprétant numérique de simuler certains niveaux de lecture humaine, le rapport entre un humain et ses modèles est différent que celui qu'à un algorithme avec les siens : l'extratextualité de l'algorithme est modulable, pouvant être changée en quelques lignes de code et dépendant totalement des décisions du chercheur. L'utilisation de ces modèles n'est pas, comme dans le cas de l'interprétant humain, approximative : la mémoire d'un ordinateur ne se dégrade pas³³ et ce dernier peut accéder pleinement et parfaitement à toutes les informations qui lui sont disponibles. C'est à la fois un problème à considérer (la propagation des biais pouvant grandement nuire aux résultats finaux de manière parfois difficile à détecter) et un avantage : les êtres humains ont beaucoup de difficulté à conserver les paramètres d'un modèle sur une longue période de temps. Sur ce point, l'interprétation littéraire numérique est favorisée par rapport à l'interprétation humaine, puisqu'elle a la possibilité d'être synchronique ou diachronique, et même de se mettre à jour pour simuler le synchronisme dans

32. L'exemple des algorithmes de recommandation des médias sociaux dénote du danger de l'application d'une règle ou d'un modèle à grande échelle (Shah 2018).

33. Nous parlons ici de périodes limitées, la pérennité des informations sur supports informatiques n'est pas assurée sur des périodes plus longues que quelques années.

le cas d'une interprétation diachronique. Les biais conscients et inconscients du chercheur affecte ainsi l'interprétation algorithmique, un piège qui ne saurait être complètement réglé par la création de réseaux sémantiques à partir de méthodes numériques et de données sociales ou linguistiques brutes puisque ces mêmes biais s'appliquent à la conceptualisation et à la sélection des données qui mènent à ces modèles (Shah 2018). Le rapport entre l'algorithme et le modèle sociohistorique ou linguistique est problématisé par un aspect supplémentaire : alors que l'humain est *enworlded* et *embodied*, le programme informatique dépend d'un support sans que ce dernier soit intrinsèquement similaire à son existence³⁴. Ce genre de modèle n'est toutefois pas nécessaire à notre démarche, l'analyse de textes pouvant se faire à partir d'un *umwelt* constitué uniquement de textes et de modèles littéraires. L'interprétation algorithmique de ce genre est comparatiste par nature ; c'est d'ailleurs là une forme d'étude littéraire qui se prête spécifiquement bien aux recherches interprétatives numériques. Comme nous l'avons précédemment mentionné, la rigidité de la computation lui permet d'éviter l'approximation une fois que les paramètres de sa recherche sont implémentés. La traduction des textes littéraires en modèles en est encore à ses débuts, mais la découverte de connaissances dans les bases de données est une branche de l'informatique qui est très mature en comparaison. Ainsi, un algorithme est à même de découvrir l'ensemble des concordances, répétitions, motifs, variations, etc., dont il connaît les définitions formelles. C'est sur ce point, le formalisme, que l'interprétation numérique doit se dissocier des mécanismes présentés dans la dernière section : nous y avons présenté l'abduction comme l'inférence de choix pour l'interprétation humaine. Cette forme d'inférence *to the best explanation* permet de comprendre ou d'expliquer des passages, oeuvres, corpus et courants littéraires à partir des informations limitées, celles-là mêmes qui sont soit disponibles, soit correspondent aux limites de la cognition. L'interprétation littéraire numérique n'est pas assujettie aux mêmes limites, bien que les informations qui sont disponibles à un algorithme soient également limitées, un autre type

34. L'algorithme est un média qui peut être dit désincarné, soit transférable de "corps", pouvant exister sous forme d'idée ou de signal avant de reprendre forme dans un support informatique.

d'inférence est donc nécessaire. De plus, l'abduction n'est pas formelle, mais il existe une forme d'inférence formelle qui partage plusieurs caractéristiques avec l'abduction ; il s'agit de l'inférence bayésienne, une forme spécifique d'induction.

Induction algorithmique

In a good inductive argument, the truth of the premises provides some *degree of support* for the truth of the conclusion, where this *degree-of-support* might be measured via some numerical scale. (Hawthorne 2021)

L'induction est d'abord un processus de généralisation qui peut mener à la caractérisation, la classification, ou la partition de données ; différentes méthodes de minage de données utilisent l'induction, par exemple les arbres de décisions (Quinlan 1986), la programmation logique inductive (Muggleton et Raedt 1994), et l'inférence bayésienne (Hawthorne 2021). Au contraire de la déduction, l'induction ne génère pas de conclusion qui soit nécessairement vraie (Douven 2021), mais plutôt pour laquelle il existe un haut degré de certitude. L'induction demande trois représentations (ou langage) ; celle des observations (\mathcal{L}_O), celle des connaissances externes (\mathcal{L}_B , background knowledge), et celle des hypothèses (\mathcal{L}_H) (Muggleton 1991).

The general inductive problem is as follows: given a consistent set of examples or observations $O \subseteq \mathcal{L}_O$ and consistent background knowledge $B \subseteq \mathcal{L}_B$ find an hypothesis $H \in \mathcal{L}_H$ such that $B \wedge H \vdash O$ (Muggleton 1991)

Nous cherchons donc une hypothèse formulée de sorte qu'elle puisse être utilisée en conjonction aux connaissances³⁵ pour expliquer les observations. Muggleton prévient :

35. Ces deux aspects nous demandent encore de mathématiser des concepts, voir (Piper 2017).

such a form of inference is not sound in the logical sense since H does not necessarily follow from B and O . In the worst case there could be an infinite number of contending hypotheses all of which fit the relationship shown above. (Muggleton 1991)

Notre réponse de la section précédente tient toujours, nos démarches ne sont pas assujetties aux standards de l'épistémè scientifique³⁶. Douven écrit de l'induction et de l'abduction que

both are *ampliative*, meaning that the conclusion goes beyond what is (logically) contained in the premises (which is why they are non-necessary inferences), but in abduction there is an implicit or explicit appeal to explanatory considerations, whereas in induction there is not; in induction, there is *only* an appeal to observed frequencies or statistics. (Douven 2021)

Cette différence nous rappelle que les raisonnements logiques humains et formels sont comme des comédiens qui ne jouent pas dans la même pièce de théâtre (j'emprunte cette image à Douven) : malgré les similarités, l'abduction et l'induction existent dans des paradigmes parallèles (Douven 2021). Alors que l'abduction repose sur des valeurs qui permettent de discriminer entre plusieurs hypothèses (par exemple, la simplicité de l'explication), l'induction cherche à ce que ses conclusions satisfassent le critère d'adéquation :

The logic should make it likely (as a matter of logic) that as evidence accumulates, the total body of true evidence claims will eventually come to indicate, via the logic's *measure of support*, that false hypotheses are probably false and that true hypotheses are probably true. (Hawthorne 2021)

36. Muggleton propose entre autres d'utiliser les statistiques bayésiennes pour discriminer entre plusieurs hypothèses ; quoiqu'une part importante de son approche soit dédiée à la compression de l'information, un enjeu des plus pressants en 1991.

Plusieurs formes d'induction algorithmique reposent sur des méthodes heuristiques pour que leurs conclusions répondent au critère d'adéquation, un tel exemple étant l'arbre de décision. Ce type de minage de données permet de découvrir des règles dans une base de données, de classifier des cas inconnus à partir de ces règles, et même de partitionner des jeux de données à partir d'un algorithme simple (Quinlan 1986). L'arbre de décision sélectionne les caractéristiques d'une base de données qui permet au mieux de séparer les données en parts égales, puis reprend ce principe jusqu'à ce que tous les cas puissent être décrits sur les branches de l'arbre métaphorique. Malgré sa simplicité, quelques tâches sont mieux accomplies à l'aide de cette méthode que des statistiques bayésiennes, par exemple l'attribution de données manquantes pour les cas incomplets d'une base de données (Quinlan 1986). De son côté, l'inférence bayésienne offre un cadre épistémologique, formel et probabiliste qui ressemble à l'abduction, ce sera donc le sujet du reste de cette section.

Inférences bayésiennes

L'abduction demande à l'interprétant de générer au moins une hypothèse répondant plus ou moins à "various 'theoretical virtues' such as simplicity, consilience and precision" (McGrew 2003). C'est *theoretical virtues* servent ensuite à évaluer les hypothèses, soit pour signifier que celle produite est satisfaisante, soit pour comparer les diverses hypothèses. Dans *Confirmation, Heuristics, and Explanatory Reasoning*, McGrew fait état de l'incompatibilité entre les critères de sélection de l'abduction et les mathématiques :

But attempts to bring such theoretical virtues into clear focus without explicitly invoking the probability calculus have not advanced the discussion beyond a clash of intuitions regarding their epistemic relevance, leaving critics with the impression that 'inference to the best explanation' remains a slogan rather than an accurate characterization of any form of non-demonstrative inference (Salmon [2001a], p. 60). (McGrew 2003)

La critique épistémologique de l'abduction de McGrew est virulente, mais bien que nous devons garder en tête qu'elle nous informe sur les limites mathématiques de ce genre d'inférence, nous devons nous rappeler des différences entre l'épistémè scientifique et nos propres démarches littéraires. Pour les sciences formelles, la production de connaissance nécessite l'accumulation d'une quantité phénoménale de données ("frequently the evidence is not strong enough to warrant an inference to any of the available explanations" (McGrew 2003)), McGrew ajoute :

IBE only sanctions inference when the 'best' is good enough; sometimes, the correct response to the data is agnosticism (Lipton [2001], p. 104). [W]hen the data are sparse, the virtues are worthless: they can do no epistemic work until there are enough of them in place to justify a definite conclusion. (McGrew 2003)

Nous mettons ainsi de côté l'abduction au profit d'une autre forme d'inférence, tant pour des raisons épistémiques que parce que nous avons besoin d'outils formels (lire ici algorithmiques) dans le cadre d'une interprétation numérique. Comme l'abduction, l'inférence bayésienne utilise une série d'observations comme point de départ, mais alors que l'abduction demande la création d'une ou plusieurs hypothèses répondant à des vertus qui serviront à juger et discriminer entre les explications, l'inférence bayésienne produit un degré de certitude pour ces hypothèses à partir de statistiques bayésiennes. Classiquement, un chercheur déciderait que son degré de certitude envers une hypothèse correspond à un nombre sur la plage $]0, 1[$, où 1 (ici exclu de la plage comme 0) représenterait une confiance absolue et 0 une impossibilité totale. L'inférence bayésienne est en mesure de confronter ce degré de certitude aux données pour donner une nouvelle mesure :

Van Fraassen (1989) interprets [inference to the best explanation] as a policy of favoring explanatorily superior hypotheses by giving their subjective probabilities a post-conditionalization boost. (Weisberg 2009)

C'est toutefois là une forme précise d'inférence bayésienne définie par Weisberg dans son article *Locating IBE in the Bayesian Framework* :

Subjectivist Conditionalization When you gain new evidence E , your new degree of belief in a hypothesis H , call it $q(H)$, should be your old degree of belief in H conditional on E : $q(H) = p(H | E)$. (Weisberg 2009)

En offrant à chaque hypothèse une chance égale, nous faisons plutôt appel à une inférence bayésienne basée sur une conditionnalisation objective :

Objectivist Conditionalization At any given time, your credence in an arbitrary proposition H ought to be $p(H | E)$, where p is the correct a priori probability distribution, and E is your total evidence at that time. (Weisberg 2009)

La conditionnalisation subjective offre plusieurs défis et défauts, quelques-uns étant épistémologiques, d'autres étant pragmatiques : il est difficile de formuler automatiquement des présuppositions probabilistes pour une hypothèse sans faire intervenir un humain, et ce type d'intervention par un humain n'est pas très solide (Weisberg 2009). Comme suggéré par McGrew, nous explorons l'avenue de l'inférence bayésienne en conjonction à la conditionnalisation objective pour offrir un équivalent algorithmique à l'abduction. Dans son argument pour mettre de côté la conditionnalisation subjective, il ajoute aux points déjà apportés précédemment que cette dernière n'est qu'une heuristique destinée à répliquer arbitrairement l'instinct et de l'expérience humaine avant de la combiner aux statistiques bayésiennes (McGrew 2003). Nous sommes toutefois confrontés à deux problèmes ; d'abord, l'inférence bayésienne ne permet pas de trouver de conclusion définitive au contact d'un jeu de données limitées. Ce premier enjeu est, dans notre cas très spécifique, un anti-problème, puisque nous ne cherchons pas à trouver de conclusions définitives, et bien que nous essayons de nous éloigner de l'idée que nos modèles sont des jouets (Rockwell et

Sinclair 2016), notre démarche se veut ouverte et incomplète, ce qui nous rend immunisés aux standards épistémiques du scientisme. Notre second problème est celui de la création d'hypothèses : ces dernières sont généralement produites par des chercheurs qui utilisent l'inférence bayésienne pour les tester. Notre projet d'interprétation littéraire algorithmique nous demande d'aller au-delà des hypothèses humaines, une tâche qui peut être accomplie à partir d'algorithmes génétiques. Noda et Freitas présentent les *Genetic Algorithms* comme appartenant au champ de la computation évolutive, et parlent d'individus qui dans notre cas correspondent à des hypothèses :

```

create (usually at random) an initial population of individuals; com-
pute the fitness (a quality measure) of each individual;
REPEAT
select individuals based on fitness;
apply genetic operators to selected individuals, creating offspring;
compute fitness of each offspring individual;
update the current population;
UNTIL (stopping criterion) (Noda et Freitas 2006)

```

Si générer une hypothèse demande des capacités importantes de conceptualisation, décoder une hypothèse produite par un algorithme génétique pour la rendre intelligible sera du domaine de l'exploit, mais c'est là un sujet qui dépasse la portée du présent mémoire.

Statistiques bayésiennes et inférence

L'inférence bayésienne repose sur le théorème de Baye pour lequel nous définissons A et B comme des événements qui ont une probabilité non nulle ($P(A, B) \neq 0$) :

$$P(A | B) = P(B | A) \frac{P(A)}{P(B)}$$

Ce qui se traduit en langage naturel par : la probabilité que l'évènement A aie lieu si B est vrai ($P(A | B)$) est égale au ratio entre les probabilités que A et B soient vrais ($P(A)$ et $P(B)$ respectivement), multiplié par la probabilité que l'évènement B aie lieu si A est vrai ($P(B | A)$). Ce théorème est dérivé d'une définition, celle des probabilités conditionnelles :

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

où $P(A \cap B)$ est la probabilité que A et B soient vrais. Puisque nous nous intéressons à l'inférence, nous utilisons E pour désigner les données (*the evidence*) et H pour désigner une hypothèse (Weisberg 2009). Chaque hypothèse peut être jugée indépendamment pour calculer notre degré de certitude en celle-ci en fonction des données disponibles (et pour mettre à jour notre confiance relative envers les hypothèses au contact de nouveaux cas de figure). C'est en comparant ces degrés de certitudes que nous pouvons utiliser les statistiques bayésiennes pour inférer quelle hypothèse est supérieure à une autre. Ainsi, pour une hypothèse H , nous donnons une probabilité initiale égale³⁷ à ce que l'hypothèse et sa non-hypothèse \bar{H} soient vraies ($P(H) = P(\bar{H}) = 0.5$).

$$P(H | E) = \frac{P(E | H)P(H)}{P(E | H)P(H) + P(E | \bar{H})P(\bar{H})}$$

L'utilisation de l'inférence bayésienne demande de pouvoir conceptualiser en termes mathématiques les données, les hypothèses et le rapport entre les deux. En d'autres mots, les hypothèses testées doivent être applicables algorithmiquement au modèle, de sorte que l'expression mathématique de l'hypothèse lorsque mise en relation au texte offre un résultat binaire vrai-faux. Pour reprendre notre exemple précédent de la séparation entre les tragédies et les comédies de Shakespeare : un modèle simple pourrait être la réponse binaire à la question "la pièce a-t-elle une

³⁷. Comme mentionné dans la section précédente, nous utilisons la conditionnalisation objective.

scène de mariage?” L’hypothèse “les comédies de Shakespeare ont un mariage” pourrait alors être testée contre l’anti-hypothèse “les comédies de Shakespeare n’ont pas de mariage”. Dans ce cas, nous pouvons donner un degré de certitude de 0.95³⁸ à notre hypothèse en tenant compte du fait que 15 des 16 comédies de Shakespeare présentent un mariage, l’exception étant *The Comedy of Errors*³⁹. Cela ne nous renseigne que sur les comédies, mais pas sur les autres pièces de Shakespeare, après tout, six des tragédies et quatre des pièces historiques de Shakespeare ont également des mariages, soit sur scène, soit sous-entendus, quoique ceux-ci aient tendance à avoir lieu au début de la pièce plutôt qu’à la fin. Si nous utilisons plutôt l’hypothèse : “les pièces de Shakespeare ayant un mariage sont des comédies”, nous découvrons que le degré de certitude pour cette hypothèse est de 0.55, un pourcentage qui monte à 0.68 si nous changeons notre hypothèse pour exclure les mariages ayant lieu dans un premier acte⁴⁰. Une hypothèse plus complexe permettrait d’identifier plus clairement les comédies des tragédies et pièces historiques, mais demanderait une modélisation plus complexe des pièces de Shakespeare. Pour en venir à ces résultats, nous avons assumé une probabilité égale pour les hypothèses H et les anti-hypothèses \bar{H} , ce qui diffère de l’abduction et de l’inférence bayésienne par conditionnalisation subjective. En effet, l’abduction donnerait aux hypothèses et anti-hypothèses des degrés de certitudes qui dépendent de sa *simplicity, consilience and precision* (McGrew 2003), le degré de certitude de .50 serait arbitraire ou, dans le cas d’une inférence bayésienne subjective, les degrés de certitudes initiaux seraient arbitraires. Dans le cas de l’abduction, notre seconde hypothèse a l’avantage d’être simple et de faire appel à une convention théâtrale établie⁴¹, bien qu’elle ne soit pas très précise. En assumant qu’un chercheur place son degré de certitude initial entre 60% et 90%, le degré de certitude après l’utilisation des statistiques

38. Ce qui correspond à un très haut niveau de certitude.

39. *The Winter’s Tale* et *Love’s Labour’s Lost* se terminent sur l’annonce d’un mariage imminent, nous les avons inclus puisque la fonction dramaturgique du mariage est remplie, ce qui dénote d’un modèle plus complexe que la simple présence d’une scène de mariage.

40. Quatre des tragédies et une pièce historique commencent avec un mariage ou juste après un mariage.

41. Conclure avec un mariage étant un motif récurrent dans les comédies d’autres dramaturges (Barnet 1998), nous pourrions faire le cas que notre hypothèse est conciliante.

bayésiennes place notre hypothèse n'importe où dans la plage $[0.76, 0.95]$. Ces variations importantes, même pour un modèle jouet et une hypothèse simple, sont la raison pour laquelle nous faisons la promotion de l'inférence bayésienne à partir d'une conditionnalisation objective comme substitut de l'abduction. Nous devons tout de même rappeler que si nous sommes en mesure de trouver de nombreux parallèles entre l'abduction et l'inférence bayésienne, ces deux formes d'inférences ne sont pas équivalentes : la première est basée sur une épistémologie des croyances (*belief*), et la seconde sur une épistémologie des degrés de croyances (Douven 2021).

Il y aura des humains

Nous avons précédemment énoncé notre objectif comme étant l'interprétation littéraire numérique, et les deux derniers chapitres nous ont mené à créer un cadre théorique où une herméneutique algorithmique est concevable. Bien que nos chapitres aient été divisés entre la lecture et l'interprétation, notre démarche double a cherché à réduire l'écart entre l'humain et la machine. Nous avons d'abord réduit certaines activités herméneutiques humaines à la modélisation et à l'inférence, des processus pour lesquels nous avons présenté des équivalents numériques. Ensuite, nous avons placé la cognition et la computation sur un même continuum d'évolution biotechnologique grâce à la biosémiotique, la sémiotique littéraire et la cybersémiotique. Loin de minimiser les différences entre l'interprétation humaine et algorithmique, nous avons mis de l'avant l'unicité de chaque *meaning-making practice* : ces deux rapports aux données textuelles et extratextuelles reposent sur des principes, présuppositions et épistémologies différentes, mais sont tous deux en mesure d'interpréter des réseaux complexes de signes. Nous n'avons toutefois pas explicitement ou formellement établi que l'interprétation littéraire numérique était elle-même critique ou littéraire, notre démarche s'étant arrêté à la notion de sens pour laquelle nous avons dû emprunter une définition tirée de la cybersémiotique. Le *Digital Criticism* doit répondre à des critères de pertinence et d'intelligibilité, un sujet auquel nous avons touché

dans le présent chapitre.

À ces deux critères s'ajoute un troisième, plus éluif et résistant aux définitions (formelles ou non) : celui de l'adéquation avec l'idée que se fait la communauté académique de ce qu'est la recherche en littérature. Bien que nous ayons à de multiples reprises souligné qu'une herméneutique numérique est indissociable du chercheur ; nous devons à nouveau nous replier sur ce fait : le *Digital Criticism* en étant encore à ses balbutiements, ce sera au chercheur de juger de la place des expériences d'interprétation littéraire, et de les assujettir au domaine littéraire. C'est donc en l'humain que repose encore le travail de discernement, ce qui remet l'accent sur l'aspect joueur des démarches d'interprétation numérique. Toutefois, cet aspect est encore présent non pas par nécessité ou à cause de caractéristiques intrinsèques à notre démarche, mais plutôt parce que nous n'avons pas encore de standards de modélisation littéraire, ou de pratiques établies de minage de données littéraires. En attendant le développement d'un certain accord (et parallèlement d'un débat sain) dans la communauté académique sur les outils d'analyse et leur place dans la critique littéraire, le *Digital Criticism* restera une activité en marge de la recherche normative. S'il est difficile de régulariser (et de faire adopter par un plus grand nombre) les idées et principes d'une interprétation littéraire numérique, il n'en est pas de même pour les conclusions que nous découvrons grâce aux méthodes quantitatives qui peuvent être traduites et présentées dans le langage académique des littéraires. Notre force à ce niveau doit venir des différences épistémologiques entre la cognition et la computation. C'est dans les différences entre l'interprétation littéraire humaine et numérique que se trouve le potentiel de découverte qui rend le *Digital Criticism* attirant au-delà du développement de ses outils numériques : l'interprétation en dehors de notre paradigme actuel est à même de mener à des connaissances qui sont également en dehors de ce paradigme.

Chapitre III - *experientia*

Nous avons consacré nos deux premiers chapitres à créer un cadre théorique dans lequel le sens littéraire n'est plus réservé aux humains. Pour ce faire, nous avons dû limiter et même déformer la définition du sens : les *meaning-making practices* d'un algorithme sont intrinsèquement différentes de celles d'un être vivant ou humain (Hayles 2019). Dans les prochaines pages, nous explorons en surface l'aspect pragmatique de l'interprétation littéraire numérique en décrivant quelques méthodes de modélisation textuelle qui sont nécessairement algorithmique⁴², et les fondements de la découverte de connaissances dans les bases de données. Cette étude, aucunement exhaustive, se contente de montrer que certains outils d'analyse du discours et de modélisation textuelle bien établis répondent à nos critères de lecture algorithmique. Nous décrivons également quelques pratiques du *Digital Criticism* qui existent déjà, sont en développement, ou que nous souhaiterions voir se développer à partir du cadre théorique des derniers chapitres. Pour qu'une communauté de pratique saine se mette en place, il est nécessaire que des chercheurs aux capacités littéraires et techniques soient à même de créer des méthodes de modélisation et d'interprétation fiable ; autrement la recherche littéraire algorithmique sera cantonnée à défendre ses méthodes avant ses résultats. Mais puisque nous en sommes encore aux débuts du *Digital Criticism*, les prochains paragraphes se devront de justifier

42. La modélisation thématique et la vectorisation de mot ne peuvent être implémentés qu'à l'aide d'outils computationnels, tant à cause de leurs demandes en puissance de calcul qu'à cause de leurs présuppositions linguistiques (Rhody 2012 ; Binder 2016 ; Piper 2018).

par l'exemple que des algorithmes peuvent contribuer aux études littéraires : nous n'en sommes pas encore au point où l'algorithme peut simplement être subalterne ou secondaire au texte (Ramsay 2011). Ainsi, ce chapitre culmine avec la description de différentes pratiques de *Digital Criticism*, des plus simples (et même parfois communes) aux plus spéculatives et complexes.

Modélisation textuelle algorithmique

While some of the earliest computational text-analysis projects, like Father Roberto Busa's famous collaboration with IBM on the Index Thomisticus, began by attempting to automate procedures that scholars had already been doing for centuries, topic modeling takes us well beyond the mechanical imitation of human action (Hockey). When we incorporate text-mining software into our scholarly work, machines are altering our interpretive acts in altogether unprecedented ways. (Binder 2016)

L'acte de transformation automatique du texte littéraire en objet mathématique structuré (le modèle) est, au sens utilisé par N. Katherine Hayles, une intermédiation qui permet d'encoder un aspect précis du texte dans une base de données (Hayles 2019). Ce que nous encodons et les processus par lesquels nous le faisons sont des choix éditoriaux dont la neutralité est loin d'être assurée ; quelques outils d'analyse modélisent le texte à partir de principes premiers, mais même dans ce cas, leurs modèles dépendent tout de même du corpus analysé (Dasu et Johnson 2003 ; Tan 2005). Cette intermédiation n'est donc pas neutre (Sculley et Pasanek 2008), mais en obligeant le chercheur à décrire formellement (sous la forme de lignes de code) ses actions et sa pensée, la modélisation algorithmique a l'avantage de mettre en lumière les biais du chercheur, ce qui permet de les considérer dans l'analyse ou de les adresser. Nous avons précédemment présenté une autre forme d'arbitraire : la traduction mathématique de définition en langage naturel des formalistes représente un autre enjeu épistémique de la modélisation

algorithmique ; il est difficile d'éviter les outils heuristiques (Rockwell et Sinclair 2016). Hors, un modèle généré à partir de présuppositions étrangères à celles d'un être humain ne produit pas d'objets assujettis à une définition simple en langage naturel : l'utilisation de concepts humains pour les caractériser est au mieux de l'abus de langage, au pire une erreur épistémologique majeure (Binder 2016 ; Ramsay 2011). Nous décrivons dans cette section deux formes de modélisation textuelle basées sur des principes premiers et des présuppositions de la linguistique computationnelle :

1. La modélisation thématique qui caractérise les documents d'un corpus à l'aide de thèmes (des listes de mots) qu'elle génère automatique.
2. La vectorisation de mots qui génère automatiquement un vecteur pour chaque mot du corpus en fonction de chacune de ses instances et de ses contextes.

Ces deux méthodes utilisent une série de présuppositions linguistiques—celles de la linguistique distributionnelle—qui repose sur les postulats suivants (Piper 2018) :

1. Le sens des mots est lié à leur fréquence d'apparition.
2. Le sens d'un mot est lié à ses cooccurrences avec d'autres mots dans un contexte donné.
3. Ces relations sont complètement contingentes aux échelles d'analyse utilisées.
4. Ces relations peuvent être considérées spatialement pour leurs associations sémantiques.

Selon Andrew Piper, l'accent mis sur la linguistique distributionnelle est peut-être lié à un biais d'outils : c'est une approche non triviale au texte et au langage qui utilise les notions très faciles à traiter pour un ordinateur, celles de positions et de distances entre des éléments séquentiels⁴³. En d'autres mots, ces postulats

43. La linguistique distributionnelle prend les mots en dehors de leur contexte syntagmatique

sont basés sur une présupposition cognitive selon laquelle du sens peut être généré de manière probabiliste, et une présupposition rhétorique selon laquelle les distances, la proximité, et les cooccurrences sont pertinentes à la découverte et à la construction du sens (Piper 2018).

Modélisation thématique

Topic modeling could stand in as a synecdoche of digital humanities. It is distant reading in the most pure sense: focused on corpora and not individual texts, treating the works themselves as unceremonious “buckets of words,” and providing seductive but obscure results in the forms of easily interpreted (and manipulated) “topics.” (Rhody 2012)

Dans *Alien Reading*, Jeffrey M. Binder suggère que certains types de modélisation textuelle sont à même d’approcher le texte de manières si Autres qu’un humain qu’il convient de les qualifier d’*alien* ; de radicalement étrangère. La modélisation thématique est un prime exemple de différence constructive entre la cognition et la computation ; il n’est pas rare que les sujets générés par cette méthode soient intelligibles pour un humain, bien que ce dernier ne soit pas en mesure de les reproduire (Binder 2016). Dans un tel cas, la modélisation littéraire est (nécessairement) numérique et son interprétation est humaine, c’est donc une forme *computer-assisted criticism* selon la définition offerte dans *Hermeneutica* par Rockwell et Sinclair. Toutefois, en nommant de manière intelligible les sujets produits par la modélisation thématiques à l’aide d’un langage naturel, le chercheur est à risque de faire une erreur d’ordre épistémologique. Binder décrit l’incompatibilité entre les résultats de la modélisation thématique et ce qu’un humain appelle un thème : superficiellement, un thème est une catégorie qui s’applique à certains textes alors que la modélisation thématique produit des listes de mots en fonction de leur utilisation dans un corpus. Plus profondément, le

pour les placer dans leur contexte paradigmatique, ce que la lecture rapprochée ne peut pas faire (Piper 2018).

thème est une forme d'archétype textuel produit par conventions qui sous-entend une intention et une continuité textuelle, alors que la modélisation thématique rassemble les mots dont la cooccurrence défie la chance en présupposant que les textes sont des mixtures normalement aléatoires de mots tirés de listes de mots.

The generative model underlying LDA is something like this: first, the writer picks a “mixture” of topics to write about; then the writer constructs the text word-by-word by first randomly choosing a topic from the mixture and then picking a word based on the probability table for that topic. This generative model allows a computer to perform two complementary operations: a topic-modeling program can “learn” what words are associated with what topics based on a corpus of text; then, it can use this model to infer the likely topics of other texts. (Binder 2016)

Nous sommes conscients que cette description de l'écriture ne décrit pas du tout le processus de production des textes par des auteurs humains, une réalité qui ne nous empêche pas d'explorer le texte avec cet outil⁴⁴. Toutefois, Lisa M. Rhody explique que dans certains cas, la linguistique distributionnelle est plus qu'un outil disponible et performant de modélisation ; parfois ses présuppositions sont similaires à celles du corpus étudié.

both LDA and our existing model of ekphrasis presuppose that latent patterns of language, when discovered, can be used to describe the corpus as a whole. (Rhody 2012)

À certains corpus correspondent mieux certaines présuppositions algorithmiques, le discernement de telles correspondances est du registre de la conceptualisation dans les étapes de la modélisation littéraire numérique (Piper 2017). Rhody iden-

44. Après-tout, il est commun d'appliquer des cadres interprétatifs modernes sur des textes anciens bien que des différences fondamentales existent entre les cultures ayant produit les premiers et les seconds.

tifie trois types de sujets pouvant être générés par un algorithme de modélisation thématique :

1. *“Chunk” topics* : Des sujets générés parce qu’un texte spécifique utilise les mots de manière si distincte qu’ils forment leur propre sujet compact.
2. *Semantically evident topics* : Des sujets pouvant être interprétés par un chercheur humain, composés de mots dont la juxtaposition produit une image, un thème ou même un sujet au sens humain.
3. *Semantically opaque topics* : Des sujets ne répondant pas aux deux premiers types, dont la juxtaposition des mots ne se traduit pas aisément (ou pas du tout) en termes humains.

La première et dernière catégorie nous rappellent que les résultats de la modélisation thématique sont intrinsèquement différents des thèmes générés par des humains, c’est toutefois une force de cet algorithme : il nous permet de contempler la possibilité d’un sens différent du sens humain. La modélisation thématique reste une forme utile et utilisée de modélisation textuelle dont les applications nombreuses sont garantes de sa maturité en tant qu’outil d’analyse numérique des textes.

Vectorisation de mot

Comme la modélisation thématique, la vectorisation de mots (parfois appelée le plongement lexical) est basée sur la linguistique distributionnelle, mais au lieu de générer des listes de mots, cet algorithme génère un vecteur par mot qui encode son contexte normal (Mikolov et al. 2013 ; Pennington, Socher, et Manning 2014). Pour ce faire, le chercheur définit un contexte caractérisé par un nombre entier m : chaque mot sera considéré en fonction des m mots le suivant et m mots le précédent, ce qui forme son contexte. D’un point de vue technique, chaque mot se voit attribué un vecteur ⁴⁵ \vec{w}_i aléatoire ; chaque vecteur de mot est ensuite mis à jour sur plusieurs itérations en fonction des autres vecteurs de son contexte,

45. Une série de nombres représentant une position dans un espace multidimensionnel.

de sorte qu'une fonction de perte soit minimisée. Cette dernière se base sur la probabilité que deux mots partagent un contexte, et les vecteurs sont modifiés en utilisant un algorithme de descente du gradient jusqu'à ce qu'un minimum local soit identifié (Chaubard, Mundra, et Socher 2016). La vectorisation de mot se décline en deux types ; d'abord CBOW (Continuous Bag of Words) pour lequel l'algorithme considère un contexte et cherche à prédire quel mot lui est associé ; puis Skip-gram qui prend l'approche inverse : à partir d'un mot, l'algorithme cherche à prédire un contexte normal (Mikolov et al. 2013). Les vecteurs ainsi générés ont plusieurs propriétés intéressantes qui reflètent les contextes normatifs dans lesquels les mots qui leur sont associés peuvent être trouvés, dans le cadre du corpus d'entraînement du modèle. Une telle propriété est que des mots au sens similaire tendent à se trouver dans des contextes similaires (Kusner et al. 2015), ainsi leurs vecteurs sont relativement proches. Une autre propriété est celle de relation sémantique entre deux concepts ; un exemple classique est celui de la relation entre les mots "roi", "reine", "homme" et "femme". Lorsqu'un modèle de vectorisation de mot est entraîné sur un corpus important dans lequel ces mots ont une présence normative et suffisante, la différence entre les vecteurs "femme" et "reine" sera équivalente à la différence entre les "homme" et "roi". Ainsi, le vecteur résultant de l'opération $\vec{w}_{roi} - \vec{w}_{homme} + \vec{w}_{femme}$ est très proche du vecteur \vec{w}_{reine} (Church 2016). Les modèles sont caractérisés par la taille de leurs vecteurs et certains modèles commerciaux entraînés sur d'immenses corpus sont disponibles aux chercheurs, le module Python Word2Vec par exemple a été entraîné sur quelque 100 millions de mots et ses vecteurs ont 300 dimensions. Ces vecteurs représentent donc les mots dans un très vaste corpus de nouvelles pour lesquelles la langue et les textes sont régularisés par défaut, ce qui offre aux chercheurs des vecteurs pour une utilisation très précise de l'utilisation des mots. Le processus peut toutefois être effectué sur n'importe quel corpus, quoique des résultats probants nécessitent une masse critique de documents. En tant que modèle, la vectorisation de mot est particulièrement difficile à analyser ; nous avons déjà mentionné que l'être humain est incapable de trouver des motifs dans les espaces à plus de trois dimensions sans aide algorithmique, et cet

algorithmes voit son efficacité augmenter avec la taille de ses vecteurs. Au niveau de l'interprétation des résultats de la vectorisation de mot, il est important de ne pas considérer que ces vecteurs ont des propriétés sémantiques, même si ces derniers peuvent des fois exemplifier des relations sémantiques entre mots. La vectorisation de mots, comme la modélisation thématique, nous offre un modèle qui n'est pas en accord avec la compréhension humaine des mots et documents, mais qui restent intéressants et pertinents à plusieurs activités d'analyse. Dans le cadre des études littéraires, la vectorisation de mots a servi à des études de genre (Grayson et al. 2016), l'étude de concepts et de mots dans une oeuvre donnée (Pollak, Martinc, et Poniz 2020 ; Dunder et Pavlovski 2019), et les variations culturelles dans les textes littéraires (Kozlowski, Taddy, et Evans 2019).

Découverte de connaissances dans les bases de données

Les outils de modélisation littéraire sont plus lents à se développer en comparaison aux méthodes d'exploration de données, l'utilisation de ces derniers ayant été propulsée par la disponibilité du stockage numérique et la puissance de calcul des ordinateurs. Supportée par une industrie friande de mégadonnées et par les sciences observationnelles gérant des téraoctets à la fois, la découverte de connaissance dans les bases de données offre des techniques variées parmi lesquelles se côtoient des heuristiques et des cadres épistémologiques rigides. Entre autres avantages, l'exploration de données peut générer des prédictions, caractérisations et associations qui souvent échappent aux chercheurs humains en utilisant des algorithmes capables de fouiller systématiquement des jeux de données complexes et multidimensionnels (Han, Kamber, et Pei 2012 ; Fayyad, Piatetsky-Shapiro, et Smyth 1996). On distingue communément six tâches en exploration de données : une première tâche est la **régression**, une méthode statistique simple qui a plusieurs utilités, par exemple, elle permet de trouver la règle qui régit des données numériques (non-binaires ou discrètes) ou de

trouver le niveau de corrélation entre des données en fonction d'une telle ligne de régression. Ainsi, ces méthodes permettent de prédire des variables quantitatives et leur degré de précision sur un spectre donné, et s'il est possible pour un être humain d'effectuer des régressions linéaires⁴⁶ à deux ou même trois dimensions, un algorithme peut sans problème effectuer des régressions linéaires ou non linéaires à autant de dimensions que nécessaire. La **classification** est une forme de prédiction qualitative ; de multiples méthodes différentes en font un champ accompli et très divers ; c'est en partie parce que la classification a des applications importantes pour l'industrie et la médecine. Certaines approches sont simples et heuristiques, c'est le cas par exemple du K-Nearest-Neighbours qui compare un nouveau cas à des données d'entraînement et l'associe à la catégorie dominante de ses K voisins les plus proches (Fayyad, Piatetsky-Shapiro, et Smyth 1996). La classification peut être effectuée à partir de trois types de jeux de données différents, soit de manière supervisée, semi-supervisée, ou non-supervisée (Han, Kamber, et Pei 2012). Dans le cas des algorithmes proposant des prédictions, une méthode est dite supervisée si elle s'appuie sur des exemples de données auxquelles des résultats sont déjà associée pour classer les nouveaux cas qui lui sont présentés ou prendre des décisions en fonction de nouvelles circonstances. À l'inverse, une méthode non-supervisée mènent l'algorithme à générer ses propres catégories et classes à partir d'une base de données. Finalement, les méthodes semi-supervisées sont à même de compléter des cas supervisés avec des cas pour lesquels les catégories recherchées sont absentes. Ces méthodes plus sophistiquées sont capables d'augmenter leur précision considérablement en comparaison aux résultats qui seraient obtenus en utilisant seulement les jeux de données pré-catégorisés (Han, Kamber, et Pei 2012). La classification ne doit pas être confondue avec la **partition** qui divise des données en groupes similaires selon divers critères et règles algorithmiques. Une méthode heuristique est le partitionnement en k-moyennes pour lequel n points sont choisis dans l'espace des caractéristiques numériques. Ces points sont ensuite déplacés dans l'espace de

46. Une régression linéaire prend la forme $y = ax + b$ à deux dimensions et cherche simplement à trouver la ligne droite qui passe le plus près possible du plus grand nombre de points (de sorte que l'écart type soit minimisé).

sorte qu'ils minimisent leur distance moyenne avec les points de données qui leur sont les plus proches jusqu'à ce qu'il y ait convergence, c'est-à-dire que les points arrêtent de bouger (Han, Kamber, et Pei 2012). La partition automatique de données textuelles est un défi important dans le domaine de la bibliothéconomie et pour autres sciences de l'information. Dans le cadre de données relationnelles, la découverte de connaissances dans les bases de données peut prendre la forme d'un **apprentissage de règles d'association**. Un exemple classique est celui du panier d'épicerie, l'étude des données d'achats peut révéler des associations inattendues entre certains items, tel que la bière et les couches pour bébés qui, dans une des premières expériences de ce type, étaient généralement achetées de concert les samedis (Power 2002). Une série de méthodes de minage de données entrent sous l'ombrelle de la **réduction dimensionnelle** et du **résumé automatique**, celles-ci utilisent une variété d'outils statistiques et mathématiques pour transformer des données complexes en données plus simples. La modélisation thématique explorée dans la dernière section est parfois utilisée pour les tâches de résumé automatique puisqu'ils permettent d'isoler les thèmes des documents pour en offrir un aperçu. La réduction automatique est plutôt utilisée dans le cadre de données numériques puisqu'elle permet de mettre en évidence certaines caractéristiques, ressemblances ou différences fondamentales des données, tout en conservant les distances relatives des divers points de données (selon des règles ou décisions établies par le chercheur). Une dernière tâche est la **détection des anomalies**, une forme de découverte qui dépend habituellement des résultats d'autres méthodes d'exploration de données. Les anomalies peuvent être des données aberrantes dans une régression, un cas qu'aucune partition ne parvient à représenter, ou un jeu de données qui résiste aux règles de classification. La découverte d'anomalies permet de trouver les faiblesses d'un modèle ou encore nous renseigne sur les données disponibles, si un type de cas est rare ou un jeu de données diffère fortement des autres, il est possible que cela soit représentatif d'un problème dans la production des données ou d'une réalité incomprise par le chercheur.

Il existe pour chacune de ces tâches des exemples d'application dans l'histoire des études textuelles et littéraires. Un effort de réduction dimensionnelle des interactions entre les personnages des pièces de Shakespeare produite par Eric T. Nalisnick et Henry S. Baird mena à la création de réseaux de personnages dont les liens étaient gradés en fonction de la valence émotionnelle de leurs dialogues (Nalisnick et Baird 2013b). À partir de l'information relative au vocabulaire dans le cadre de ces échanges, Nalisnick et Baird furent en mesure de représenter visuellement ces données relationnelles qui dévoilent, entre autres, des groupes alliés et ennemis dans *Hamlet*. Dans ce même article, les auteurs découvrent une anomalie parmi les tragédies et comédies de Shakespeare quand ils comparent la valence moyenne des mots de chacune de ces pièces. En effet, une analyse rapide permet de voir que les comédies ont légèrement tendance à avoir une valence émotionnelle plus haute que celle des tragédies⁴⁷, avec une exception notable. *The Comedy of Errors* a la valence la plus basse de toutes les pièces considérées (Nalisnick et Baird 2013b), bien inférieure à la plus sombre des tragédies, en fait, c'est la seule pièce dont la valence moyenne soit négative. Cette donnée aberrante permet d'isoler *The Comedy of Error* comme étant une comédie atypique, en effet, "the plot features farcical but dark elements such as beatings, infidelity, theft, and madness, which probably have been taken at 'face value' by the shallow analysis" (Nalisnick et Baird 2013b). Le comique de cette pièce vient de la surabondance d'erreurs, tant et si bien que le titre de la pièce a passé dans la langue en devenant une expression qui signifie "an event or series of events made ridiculous by the number of errors that were made throughout" (Merriam-Webster s. d.). Des efforts de classification ont été faits par Ted Underwood dans *Distant Horizons* pour distinguer la fiction de la bibliographie, une tâche qui s'avère assez simples à accomplir à partir de modèles aussi simple que la durée de l'histoire (Underwood 2019). Andrew Piper réussit également à différencier la fiction de la non-fiction à partir de différences linguistiques formelles : son modèle se base sur des caractéristiques aussi simples

47. Un signe que, comme modélisation littéraire, la valence moyenne des mots dans une oeuvre de Shakespeare est n'a pas de pouvoir prédictif définitif sur le type de pièce dont il est question selon les définitions établies par la communauté de chercheurs.

que la proportion de point d'exclamation et d'interrogation et la fréquence relative de certains mots clés, d'où le titre de son ouvrage ; *Enumerations* (Piper 2018). Cherchant aussi (et bien plus tôt) à faire la différence entre deux genres littéraires, Boris Iarkho classifiait des pièces de théâtre en cinq actes à partir de caractéristiques formelles qu'il mesurait lui-même, un processus long et laborieux. Puisqu'il utilisait des variables numériques plutôt que quantitatives, plusieurs de ses découvertes étaient le fruit de régressions, qui lui permirent entre autres de montrer que les variations stylistiques littéraires n'étaient pas linéaires, mais venaient plutôt par vagues (Lvoff 2021). Finalement, des efforts de partitions sont trouvés chez plusieurs auteurs (Rockwell et Sinclair 2016 ; Moretti 2013), mais ceux-ci font face à des défis importants, le premier étant la définition de distances⁴⁸ entre les oeuvres en fonction de caractéristiques numériques ou qualitatives dont les échelle et schémas de gradation sont tous différents (Päpcke et al. 2022).

Digital Criticism

Parmi nos multiples exemples de recherche littéraire algorithmique ou assistée par ordinateur, Boris Iarkho revient encore et encore pour avoir réussi à combiner la modélisation littéraire des formalistes et la régression pour exposer la faiblesse d'un modèle littéraire historique. Son étude systématique des pièces en cinq actes lui a permis de différencier les pièces romantiques et classiques en fonction de leurs caractéristiques formelles, mais aussi de découvrir des motifs dans le passage d'une tradition à l'autre qui allaient à l'encontre de l'orthodoxie académique de son temps (Lvoff 2021). Les moyens limités de Iarkho ne l'ont pas empêché d'utiliser les outils mathématiques de la biologie évolutive (sa métaphore de choix pour décrire les variations stylistiques littéraires) pour découvrir que le passage d'un genre à l'autre était fait d'oscillations ou de vagues des caractéristiques qu'il avait modélisées plutôt que d'une transition linéaire. Il est toutefois nécessaire

48. Dans les espaces multidimensionnels, la notion de distance devient quelque peu floue, surtout quand les dimensions ne sont pas équivalentes dans leurs échelles (Mufikhah et Baharudin 2009).

de démontrer la relation entre les outils de Iarkho et ses découvertes : en utilisant les mathématiques de l'évolution des espèces (Gasparov 2016), ses découvertes ont pris cette forme exacte, celle d'une évolution⁴⁹. Pareillement, l'architecture algorithmique d'une interprétation littéraire numérique forme des contraintes et permet des possibilités herméneutiques qu'il nous faut explorer, un thème repris par Jeffrey M. Binder dans son article "Alien Reading". Il y distingue la lecture rapprochée, la lecture distante (voir *Distant Reading* (Moretti 2013), *Macroanalysis* (Jockers 2013) et *Distant Horizons* (Underwood 2019)) et, finalement, la lecture *alien*. Son approche, et le pilier sur lequel repose son analyse, est que nous devons interagir et comprendre les algorithmes de modélisation thématique comme une nouvelle forme de lecture : "an alien form of reading" (Binder 2016). L'article de Binder se concentre sur la modélisation textuelle et utilise l'exemple de la linguistique distributionnelle, mais d'autres formes de modélisation littéraire pourront être considérées comme *alien* si elles se basent sur des principes mathématiques et des présuppositions que nous ne pouvons répliquer à partir de la cognition. Binder laisse toutefois de côté la notion de sens, et bien que les formes de modélisation qu'il met de l'avant répondent à notre description d'une interprétation nécessairement numérique, nous devons nous en remettre aux travaux de Terrence Deacon, Noel Burton-Roberts et, surtout, N. Katherine Hayles pour parler d'interprétation, d'herméneutique ou de sens.

La modélisation littéraire est une forme de contrainte téléologique qui peut, sous certaines conditions, générer un système orthograde dans lequel des *meaning-making practices* peuvent avoir lieu, même si le modèle n'est pas en adéquation avec ce que l'être humain saurait générer. À propos de la modélisation thématique, Lisa M. Rhody écrit :

Understanding how topic modeling algorithms handle figurative language means allowing for a [...] beautiful failure — not a failure of

49. L'herméneutique à partir de théories sociales offre des résultats similaires : à partir d'une interprétation marxiste, féministe, queer ou psychologue d'un texte ne peuvent apparaître que des résultats qui correspondent à la lentille ainsi utilisée (Tracey et Morrow 2006).

language, but a necessary inclination toward form that involves a diminishing of language's possible meanings. (Rhody 2012)

Cet échec de la représentation éloigne les possibilités herméneutiques de celles d'un humaniste, mais le sens littéral du langage n'est pas le seul élément porteur de sens dans un texte. La modélisation peut encoder ces autres aspects, ces derniers peuvent être les motifs, les fréquences relatives, les répétitions, les associations, et d'autres caractéristiques facilement détectables à l'aide d'algorithme. Parallèlement, nous ne pouvons prétendre générer du sens à partir des mêmes codes utilisés par l'auteur lors de son écriture du texte : nous avons déjà établi une obligatoire asymétrie entre l'encodage et le décodage d'une oeuvre, une asymétrie que l'on trouve également entre chaque interprétant de l'oeuvre⁵⁰. Nous ne sommes pas en mesure de créer une équivalence entre un interprétant humain et un interprétant algorithmique (du moins pas avec les outils de modélisation et d'inférence numériques à notre portée), mais en nous reposant sur notre étude de la cybersémiotique, nous pouvons affirmer que l'algorithme peut être au mieux générateur de sens, et au pire influencer le sens.

Pratiques passées et actuelles

Dans notre introduction, nous avons cité Stephen Ramsay qui annonçait en 2011 qu'un "algorithmic criticism" n'existaient au mieux que dans une forme limitée et naissante (Ramsay 2011). Onze ans plus tard, le *Digital Criticism* en est encore à ses débuts, quoique nous puissions aujourd'hui trouver des exemples de plus en plus nombreux, concrets et probants d'analyses littéraires assistées par ordinateurs. Certaines transformations algorithmiques et formes de modélisation littéraire existaient toutefois bien avant la publication de *Reading Machines* (Ramsay 2011), et ces dernières étaient et sont encore utilisées à des fins interprétatives. Les exemples les plus simples sont les index et les fonctions de recherches ; et bien que ces outils semblent banals, triviaux, et

50. Chaque interprétation se base sur un réseau sémiotique différent, sur des modèles littéraires, psychologiques ou sociohistoriques, et sur des outils interprétatifs différents.

même complètement neutres, ils représentent bien comment toute intermédiation est, à un niveau ou un autre, éditoriale. Dans le cas de l'index, le choix des termes inclus informe la recherche qui peut être aidée par son utilisation, et Ted Underwood met en garde les chercheurs contre la sophistication cachée (*black box*) des moteurs de recherche. Dans son article "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago", Underwood explique que si la recherche booléenne de termes ou d'expressions dans un texte peut transformer la critique par ses possibilités, la recherche de sources secondaires à l'aide de moteurs de recherche sur des bases de données problématise la notion même de recherche académique en donnant un pouvoir disproportionné aux biais de confirmation.

[The search function] emphasizes a hermeneutic spiral that will be familiar to humanists, acknowledging that we approach every question with some previous assumptions (called "prior probabilities"), as well as particular kinds of uncertainty. [...] Researchers can never afford to treat algorithms as black boxes that generate mysterious authority. [...] To understand the interpretive limitations of an algorithm, one needs to understand its mathematical basis. (Underwood 2014)

Dans la recherche de sources secondaires dans une base de données, ce sont des algorithmes comme la modélisation thématique qui servent à choisir quels textes sont présentés au chercheur, et d'autres algorithmes encore qui décident de l'ordre dans lequel ces résultats sont montrés. La fonction de recherche est une transformation textuelle (en mettant de l'avant certains textes ou passages, elle en efface d'autres qui peuvent être ignorés) qui a des impacts sur l'interprétation (Underwood 2014), mais nous nous doutons que ce ne soit pas là ce que Stephen Ramsay avait en tête lorsqu'il parlait d'un "algorithmic criticism". D'autres formes de rapports entre le texte, le chercheur et l'algorithme, plus près du *Digital Criticism*, sont à l'oeuvre dans un renouveau du formalisme qui utilise des définitions tirées ou dérivées des méthodes des formalistes du dernier siècle.

Ces caractéristiques et analyses ont l’avantage d’être facilement répliquables à l’aide d’outils numériques, ce qui permet une continuation directe des démarches formalistes un siècle après leurs travaux, tout en les accélérant⁵¹ considérablement (Wendell 2021). Le formalisme inspire encore plusieurs pratiques de modélisation textuelle, par exemple dans le cadre de la stylométrie et l’étude des genres littéraires, dont les méthodes de base reposent sur des caractéristiques linguistiques superficielles et intelligibles, ou sur des méthodes quantitatives automatisées, mais qu’un humain pourrait accomplir avec un temps suffisant (Underwood 2019 ; Stamatatos 2009). Par contre, ces formes de modélisation et d’analyse littéraire sont limitées en comparaison aux techniques modernes de fouille de texte. L’analyse des sentiments est un exemple de modélisation qui vient de l’industrie (la plupart des modèles de ce type sont entraînés sur des revues de produits ou de films), mais qui trouve des applications dans le domaine littéraire. Cette méthode utilise un dictionnaire de termes auxquels sont associés des valeurs numériques : leur valence émotionnelle, en fonction de leur utilisation dans des exemples de textes positifs ou négatifs. À ce propos, nous avons déjà cité Eric T. Nalisnick et Henry S. Baird qui ont généré des réseaux à partir des dialogues d’*Hamlet* dans lesquels la valence émotionnelle des échanges caractérisait les relations entre personnages (Nalisnick et Baird 2013b). Pour cet exemple et d’autres, une modélisation textuelle est générée algorithmiquement, puis des chercheurs utilisent des outils de visualisation des données ou des méthodes simples d’exploration de données pour en analyser les résultats. Plusieurs critiques et auteurs rappellent que les résultats de telles démarches tendent à confirmer les connaissances des chercheurs plutôt qu’en créer de nouvelles, tout en mentionnant que cette approche se limite généralement à l’accumulation de preuves dans une imitation des sciences formelles plutôt que comme une recherche herméneutique et humaniste. Sur ce point, les exceptions, celles qui constitueraient un “algorithmic criticism” selon la définition de Ramsay, sont rares.

51. Rappelons encore l’exemple de Boris Iarkho qui comptait à la main les caractéristiques textuelles pour les analyser avec du papier quadrillé et une règle (Gasparov 2016).

Vers une interprétation littéraire numérique

Nous concluons ce chapitre en proposant une inférence qui résume l'ensemble du présent mémoire :

1. La lecture d'un texte littéraire est la création de modèles qui représentent ce texte à partir d'information intra-, inter-, et extratextuelle.
2. L'interprétation d'un texte littéraire est la création de sens à partir des modèles qui représentent ce texte.
3. Certaines interprétations littéraires sont le résultat d'inférences.
4. La modélisation textuelle algorithmique, bien que différente dans ses formes connues de la lecture humaine, constitue une représentation valable (bien que limitée) du texte.
5. Les inférences algorithmiques, bien que différentes dans leurs formes connues des inférences humaines, peuvent être des *meaning-making practices*.

Ergo, l'analyse par inférences algorithmiques de modèles textuelles algorithmiques peut générer du sens littéraire. Notre justification pour avoir une interprétation algorithmique est double, d'abord parce qu'en présence de multiples modèles complexes, l'analyse humaine est limitée et même inadéquate, puis parce que nous sommes motivés par la possibilité de possibilités herméneutiques en dehors du paradigme humain et de ses biais. Nous savons toutefois d'expérience que toutes les combinaisons de modèles et d'inférences ne génèrent pas du sens littéraire chez les humains ; similairement, l'interprétation littéraire numérique n'est pas forcément pertinente ou intelligible. Notre espoir est qu'en étendant la structure d'Andrew Piper pour modéliser algorithmiquement un texte⁵² à son interprétation, nous puissions également analyser de manière pertinente et intelligible les modèles textuels algorithmiques. Structurellement, une interprétation littéraire numérique commence ainsi par la modélisation d'un corpus à partir

52. Soit la théorisation, suivie de la conceptualisation, l'implémentation, la sélection et la validation (Piper 2017).

d'une ou plusieurs méthodes en représentations numériques ; des intermédiations dans lesquelles sont encodés des aspects précis du corpus. L'interprétation de ces modèles peut ensuite commencer à partir de méthodes d'exploration de données pour découvrir, à même les représentations du texte, des motifs, règles ou groupes dont la découverte n'est pas possible sans les intermédiations précédentes et leurs caractéristiques computationnelles. Selon les contraintes de la cybersémiotique, toutes les techniques d'exploration de données ne sont pas à même de générer du sens (d'être des *meaning-making practices*) ; Hayles exprime en ces mots le défi au coeur de l'interprétation numérique :

In computational media, interpretation is ultimately based on the logic gates, which generate the constraints that Deacon's scheme were accomplished through orthograde and contragrade dynamics. [W]ith computational media designed to implement symbolic processes through their gates, the crucial point is how these deterministic operations can create a basis for flexibility, adaptability, and evolvability (Hayles 2019)

Indirectement, Hayles décrit des méthodes comme les réseaux de neurones, les algorithmes génétiques et certaines techniques de réduction dimensionnelle.

these computational media do not understand in the same sense as do humans; nevertheless, they are capable of meaning-making practices within their *umwelten* [(their world horizon).] As von Uexküll emphasized, *umwelten* between different species may overlap, but they are never completely coincident with one another. (Hayles 2019)

L'interprétation littéraire numérique nous offre la possibilité de découvrir des orientations vers le texte qui échappent aux biais humains et à la "spirale herméneutique" (Underwood 2014), en plus de nous permettre de dépasser les limites biologiques de la cognition. Le *Digital Criticism* est ainsi à même de nous indiquer de nouvelles avenues de recherche littéraire en nous indiquant l'existence

de formes, de surfaces et de catégories littéraires ainsi que leurs définitions en fonction de modèles. La traduction de ces concepts algorithmiques en concepts naturels est donc une potentielle addition à nos outils herméneutiques, mais nous sommes également à même d'espérer une normalisation de certaines pratiques. Celle-ci devrait se développer naturellement une fois que le *Digital Criticism* aura des outils numériques fiables d'un point de vue pragmatique et autour desquels des conventions conceptuelles auront été adoptées par la communauté de chercheurs.

conclusio

“But what does this second turtle stand on?” persisted James patiently.

To this, the little old lady crowed triumphantly,

“It’s no use, Mr. James—it’s turtles all the way down.”

–(Ross 1967)

À quatre années d’intervalle, soit en 1927 et 1931 eu lieu une intéressante convergence entre les mathématiques et la philosophie ; d’abord avec le cercle herméneutique (*Being and Time* (Heidegger 1962)), puis avec les *incompleteness theorems* (Gödel 2007). Comme la relation entre l’auteur et le texte est constructrice et génératrice de l’un comme de l’autre (“Neither is without the other. Nevertheless, neither is the sole support of the other” (Heidegger 2008)), le sens du texte est construite à la lecture par le va-et-vient entre le lecteur et le texte, entre la partie et le tout. Le cercle herméneutique sous-entend aussi que l’intertextualité n’est pas seulement nécessaire à la compréhension d’une oeuvre, elle est fondamentale à son existence en tant qu’oeuvre d’art, un sujet mieux décrit par Roland Barthes dans sa *Théorie du texte* :

Le texte redistribue la langue (il est le champ de cette redistribution).

L’une des voies de cette déconstruction-reconstruction est de permuter des textes, des lambeaux de textes qui ont existé ou existent autour du texte considéré, et finalement en lui : tout texte est un intertexte

; d'autres textes sont présents en lui, à des niveaux variables, sous des formes plus ou moins reconnaissables : les textes de la culture antérieure et ceux de la culture environnante ; tout texte est un tissu nouveau de citations révolues. (Barthes 1974)

L'image évoquée par Barthes d'une surface qui ne tient que sur d'autres surfaces nous rappelle à la fois un cercle et une régression infinie qui ne saurait être arrêté qu'en posant des briques pour former, à une profondeur arbitraire, un sol qui lui-même repose dans le vide et sur lui-même. C'est là exactement l'état des mathématiques, Kurt Gödel l'a démontré en 1931 : aucun système mathématique formel n'est satisfaisant, puisque 1) ils ne sont pas en mesure de se valider ou s'invalider eux-mêmes, et parce que 2) chaque système mathématique contient des éléments axiomatiques qui ne peuvent pas être prouvés (Gödel 2007). D'un point de vue épistémologique, Hans Albert parle du trilemme de Münchhausen (basé sur une expérience par la pensée d'Agrippa) selon lequel il n'existe que trois formes possibles d'explication (Albert 1985) :

1. L'argument circulaire
2. Les régressions infinies
3. L'appel à l'autorité (ou aux axiomes)

La nature de l'art est ainsi similaire à la nature des mathématiques et de la connaissance, une idée qui représente bien l'*ethos* qui dominera la scène intellectuelle du XXe siècle, soit l'impression d'une perte d'équilibre alors que les fondements mêmes du monde semblent s'effriter. Elles représentent aussi une libération : l'art comme les mathématiques deviennent des jeux de langage (Wittgenstein 2004) et peuvent être considérés en fonction de leurs logiques internes, après tout, "it's turtles all the way down."

Au cours des dernières années, plusieurs initiatives (qui impliquent généralement de vastes réseaux de neurones) ont cherché à utiliser la masse grandissante et déjà incommensurable de données textuelles disponibles dans diverses bases de

données et sur les serveurs internet. Bien que les résultats dans le domaine des intelligences artificielles textuelles et linguistiques soient impressionnants⁵³ et même troublants⁵⁴, ces programmes révèlent sans cesse leur superficialité, et leur manque de compréhension des mots et concepts qu'ils utilisent. Dans un article publié en réponse à l'engouement autour du chatbot LaMDA, Jacob Browning et Yann Lecun résume en une phrase ce qu'ils considèrent comme noeud du problème et sa solution :

“Abandoning the view that all knowledge is linguistic permits us to realize how much of our knowledge is nonlinguistic.” (Browning et Lecun 2022)

La cognition (ou pensée) et le texte sont des médias différents : la connaissance ne saurait être supportée également par l'un comme par l'autre ; il convient de tenir compte de la perte importante d'information qui a lieu lors de toute intermédiation (Hayles 2019), l'écriture étant une telle intermédiation. Hayles spécifiait que du passage d'un média à l'autre, l'information était presque entièrement perdue, mais qu'ainsi des informations autrement invisibles (motifs, réseaux, etc.) devenaient visibles (Hayles 2019). Sur le sujet de la représentation visuelle des données, Alexander Galloway écrivait : “data have no necessary visual form” (Galloway 2011) ; une affirmation que nous pouvons généraliser à : l'information n'a pas de forme représentative nécessaire. Un exercice de différenciation équivalent à celui effectué sur la lecture puis l'interprétation (cognitive vs computationnelle) dans nos deux premiers chapitres révélerait des différences intrinsèques similaires entre les représentations cognitives et textuelles. La première est, nous l'avons déjà mentionné, *embodied* et *enworlded*, mais aussi multidimensionnelle, continue, modale, et en changeante. Le texte est généralement fini, séquentiel, linéaire, supporté⁵⁵, et statique : sa nature même le

53. Citons par exemple la traduction automatique, la recherche par mot clé dans les bases de données textuelles, ainsi que la lecture distante.

54. Rappelons que l'ingénieur Blake Lemoine fut renvoyé de son poste d'ingénieur chez Google après s'être fait l'avocat de l'intelligence artificielle LaMDA puisqu'il considérait qu'elle était consciente (Wertheimer 2022).

55. Le texte nécessite un support physique.

limite quant au type d'information qu'il peut représenter adéquatement ou même tout court. Les conséquences des limites du texte comme média et du langage comme support de la connaissance ont des répercussions importantes pour le développement d'une intelligence artificielle textuelle, mais nous nous sommes précédemment dissociés de la sémantique au profit de la sémiotique. Cette décision, loin d'être innocente, nous a permis de considérer l'oeuvre textuelle comme flottante dans un réseau (cercle) herméneutique d'autres oeuvres plutôt que comme un objet possédant des propriétés sémantiques. Le texte littéraire comme assemblage sémiotique faisant partie d'une toile intertextuel n'est pas (seulement) un artéfact culturel linguistique (comme l'utilisent les *Cultural Analytics*), ni une représentation directe de la pensée ou de la connaissance. Les présuppositions, outils et la philosophie de l'interprétation littéraire algorithmique sont à même d'envisager le texte littéraire pour ce qu'il est et de l'étudier en fonction de ses caractéristiques : en étudiant les motifs, réseaux et textures à sa surface, et en découvrant comment celle-ci est interconnectée avec d'autres surfaces. La modélisation littéraire algorithmique est une lecture, encore une fois, de surface (*shallow reading*) qui peut-être une représentation complexe, multidimensionnelle, inintelligible pour un être humain. Avec un peu de chance (et beaucoup de travail), nous saurons peut-être trouver des inférences algorithmiques pour rendre possible notre travail critique possible parmi une mer de nombres flottants sur une mer de lettres, flottant sur une mer de tortues.

Bibliographie

Albert, Hans. 1985. *Treatise on Critical Reason*. Princeton University Press.

Aligwe, Hygeinus Nwokwu, Kenneth Adibe Nwafor, et Johnson Chinasa Alegu. 2018. « Stuart Hall's Encoding-Decoding Model: A Critique ». *World Applied Science Journal* 36 (9). <https://doi.org/10.5829/idosi.wasj.2018.1019.1023>.

Atkin, Albert. 2013. « Peirce's Theory of Signs ». In *The Stanford Encyclopedia of Philosophy*, édité par Edward N. Zalta, Summer 2013. <https://plato.stanford.edu/archives/sum2013/entries/peirce-semiotics/>; Metaphysics Research Lab, Stanford University.

Aydın-Düzgüt, Senem. 2016. « De-Europeanisation through Discourse: A Critical Discourse Analysis of AKP's Election Speeches ». *South European Society and Politics* 21 (1): 45-58. <https://doi.org/10.1080/13608746.2016.1147717>.

Barnet, Sylvan. 1998. « Introduction ». In *Shakespeare : Four Great Comedies*. Signet Classics.

Barthes, Roland. 1966. *Critique et vérité*. Seuil.

———. 1974. « Théorie du texte ». Études à l'École pratique des hautes études.

———. 1981. « Introduction à l'analyse structurale des récits ». In *Communications, 8 : L'analyse structurale du récit*, 7-33.

- Binder, Jeffrey M. 2016. « Alien Reading: Text Mining, Language Standardization, and the Humanities ». In *Debates in the Digital Humanities*, édité par Matthew K. Gold et Lauren F. Klein. University of Minnesota Press.
- . 2020. « Romantic Disciplinarity and the Rise of the Algorithm ». *Critical Inquiry* 46 (4).
- Bonissone, Piero P. 1980. « A Fuzzy Set Based Approach: Theory and Applications ». *Proceedings of the 1980 Winter Simulation Conference*.
- Bourdieu, Pierre. 1991. « Le champ littéraire ». *Actes de la recherche en sciences sociales* 89: 3-46.
- Browning, Jacob, et Yann Lecun. 2022. « AI And The Limits Of Language ». *Noema Magazine*. <https://www.noemamag.com/ai-and-the-limits-of-language/>.
- Brownlee, Jason. 2018. « Why Initialize a Neural Network with Random Weights? » *Machine Learning Mastery*. 1 juillet 2018. <https://machinelearningmastery.com/why-initialize-a-neural-network-with-random-weights/>.
- Burton-Roberts, Noel. 2013. « Meaning, Semantics and Semiotics ». In *Perspectives on Linguistic Pragmatics*, édité par Alessandro Capone, Franco Lo Piparo, et Marco Carapezza, 1-22. Springer.
- Busa, Roberto. 1980. « The Annals of Humanities Computing: The Index Thomiscus ». *Computers and the Humanities*, 83-90.
- Butler, Judith. 2005. *Giving an Account of Oneself*. Fordham University Press.
- . 2017. *Le pouvoir des mots*. Traduit par Charlotte Nordmann et Jérôme Vidal. Éditions Amsterdam.
- Chaubard, Francois, Rohit Mundra, et Richard Socher. 2016. « Deep Learning for NLP : Lecture Notes ». Stanford University.

- Church, Kenneth Ward. 2016. « Emerging Trends: Word2Vec ». *Natural Language Engineering* 23 (1): 155-62.
- Cohen-Tannoudji, Claude, Bernard Diu, et Franck Laloë. 2005. *Quantum Mechanics*. Vol. 1. Wiley-VCH.
- Cole, Jarrett. 2020. « Negotiating Code: The Decoding of Stuart Hall ». *Winnsox*.
- Culler, Jonathan. 1976. *Ferdinand de Saussure*. Cornell University Press.
- . 2001. *The Pursuit of Signs : Semiotics, literature, deconstruction*. Routledge Classics.
- . 2011. *Literary theory: A very short introduction*. Oxford University Press.
- Dasu, T., et T. Johnson. 2003. *Exploratory Data Mining and Data Cleaning*. Wiley Series in Probability and Statistics. Wiley. <https://books.google.ca/books?id=HStUfJnCvT4C>.
- Deacon, Terrence. 2011. *Incomplete Nature: How Mind Emerged from Matter*. W. W. Norton & Company.
- Douven, Igor. 2021. « Abduction ». In *The Stanford Encyclopedia of Philosophy*, édité par Edward N. Zalta, Summer 2021. <https://plato.stanford.edu/archives/sum2021/entries/abduction/>; Metaphysics Research Lab, Stanford University.
- Dufays, Jean-Louis, Louis Gemenne, et Dominique Ledur. 2005. *Pour une lecture littéraire*. De Boeck Supérieur.
- Dunder, I., et M. Pavlovski. 2019. « Through the Limits of Newspeak: an Analysis of the Vector Representation of Words in George Orwell's 1984 ». In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 583-88. <https://doi.org/10.23919/MIPRO.2019.8756892>.

- Earhart, Amy. 2012. « The Digital Edition and Digital Humanities. » *Textual Cultures: Texts, Context, Interpretation* 7 (1): 18-28.
- Eco, Umberto. 1976. « Peirce's Notion of Interpretant ». *MLN - Comparative Literature* 91 (6).
- Erb, Maurice, Simon Ganahl, et Patrick Kilian. 2016. « Distant Reading and Discourse Analysis ». *Le foucauldien* 2 (juin): 8. <https://doi.org/10.16995/lefou.16>.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, et Padhraic Smyth. 1996. « From Data Mining to Knowledge Discovery in Databases ». *AI Magazine* 17 (3): 37. <https://doi.org/10.1609/aimag.v17i3.1230>.
- Furlong, Anne. 1995. « Relevance Theory and Literary Interpretation ». Thèse de doctorat, University College London.
- Gadamer, Hans-Georg. 1989. *Truth and Method*. Traduit par Joel Weinsheimer et Donald G. Marshall. continuum.
- Galloway, Alexander. 2011. « Are Some Things Unrepresentable? » *Theory, Culture & Society* 28 (7-8): 85-102.
- Gasparov, Mikhail. 2016. « Boris Yarkho's works on literary theory ». *Studia Metrica et Poetica* 3 (2). <https://doi.org/10.12697/smp.2016.3.2.05>.
- Glavanakova, Alexandra. 2020. « Reading Fiction in The Digital Age ». *The Online Journal for Arts and Culture - Reading Modes in the Digital Age* 20.
- Gori, Marco, et Alberto Tesi. 1992. « On the Problem of Local Minima in Backpropagation ». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (1).
- Gödel, Kurt. 2007. « On Formally Undecidable Propositions of Principia Mathematica and Related Systems ». In *God Created the Integers: The Mathematical*

- Breakthroughs that Changed History*, édité par Stephen Hawking. Running Press.
- Grayson, Siobhán, Maria Mulvany, Karen Wade, Gerardine Meaney, et Derek Greene. 2016. « Novel2Vec: Characterising 19th Century Fiction via Word Embeddings ». In.
- Hale-Stern, Kaila. 2018. « Marvel Movies Are a Universe of Terrible Fathers ». *The Mary Sue*. <https://www.themarysue.com/marvel-bad-dads/>.
- Hall, Stuart. 1999. « Encoding/Decoding ». In *The Cultural Studies Reader*, édité par Simon During, 507-17. Routledge.
- Han, Jiawei, Micheline Kamber, et Jian Pei. 2012. *Data Mining: Concepts and Techniques*. Elsevier.
- Hawthorne, James. 2021. « Inductive Logic ». In *The Stanford Encyclopedia of Philosophy*, édité par Edward N. Zalta, Spring 2021. <https://plato.stanford.edu/archives/spr2021/entries/logic-inductive/>; Metaphysics Research Lab, Stanford University.
- Hayles, N. Katherine. 2019. « Can Computers Create Meaning? A Cyber/Bio/Semiotic Perspective ». *Critical Inquiry* 46.
- Heidegger, Martin. 1962. *Being and Time*. Traduit par John Macquarrie et Edward Robinson. Blackwell.
- . 2008. « The Origin of the Work of Art ». In *Martin Heidegger: The Basic Writings*, traduit par David Farrell Krell. HarperCollins.
- Hennequin, Emile. 1888. *La critique scientifique*. Libr. Académique Didier.
- Hutchinson, Ben. 2018. *Comparative Literature: A Very Short Introduction*. Oxford University Press.
- Hutto, C. J., et Eric Gilbert. 2014. « VADER: A Parsimonious Rule-based

- Model for Sentiment Analysis of Social Media Text ». *Association for the Advancement of Artificial Intelligence*.
- Iyyer, Mohit, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, et Hal Daumé III. 2016. « Feuding Families and Former Friends: Unsupervised Learning for Dynamic Fictional Relationships ». *Proceedings of NAACL-HLT 2016*, 1534-44.
- Jannidis, Fotis. 2020. « On the perceived complexity of literature. A response to Nan Z. Da ». *Journal of Cultural Analytics*. <https://doi.org/10.22148/001c.11829>.
- Jockers, Matthew L. 2013. *Macroanalysis*. University of Illinois Press.
- Kozłowski, Austin C., Matt Taddy, et James A. Evans. 2019. « The Geometry of Culture: Analyzing the Meaning of Class through Word Embeddings ». *American Sociological Review* 84 (5): 905-49.
- Kusner, Matt, Yu Sun, Nicholas Kolkin, et Kilian Weinberger. 2015. « From Word Embeddings To Document Distances ». In *Proceedings of the 32nd International Conference on Machine Learning*, édité par Francis Bach et David Blei, 37:957-66. Proceedings of Machine Learning Research. Lille, France: PMLR. <https://proceedings.mlr.press/v37/kusnerb15.html>.
- Lancashire, Ian, et Graeme Hirst. 2009. « Vocabulary Changes in Agatha Christie's Mysteries as an Indication of Dementia: A Case Study ». *19th Annual Rotman Research Institute Conference, Cognitive Aging: Research and Practice*.
- Lvoff, Basil. 2021. « Distant Reading in Russian Formalism and Russian Formalism in Distant Reading ». *Russian Literature* 122-123: 29-65. <https://doi.org/https://doi.org/10.1016/j.ruslit.2021.07.003>.
- Marche, Stephen. 2012. « Literature Is not Data: Against Digital Humanities ». Los Angeles Review of Books. 28 septembre 2012. <https://lareviewofbooks>.

org/article/literature-is-not-data-against-digital-humanities/.

McGrath, Lynette. 1985. « Structural and Poetic Theory: Intention, Meaning and Privilege ». *College English* 47 (8): 809-23.

McGrew, Timothy. 2003. « Confirmation, Heuristics, and Explanatory Reasoning ». *The British Journal for the Philosophy of Science* 54 (4): 553-67.

McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man*. McGraw-Hill.

Merriam-Webster. s. d. « comedy of errors ». In *Merriam-Webster.com dictionary*. Consulté le 20 août 2022. <https://www.merriam-webster.com/dictionary/comedy%20of%20errors>.

Meunier, Jean-Guy. 2017. « Humanités numériques et modélisation scientifique ». *Digital Humanities and Scientific Modelling*, 19-48.

Mikolov, Tomas, Kai Chen, Greg Corrado, et Jeffrey Dean. 2013. « Efficient Estimation of Word Representations in Vector Space ». arXiv. <https://doi.org/10.48550/ARXIV.1301.3781>.

Milic, Louis T. 1967. « Winged words: Varieties of computer application to literature ». *Computers and the Humanities* 2: 24-31.

Miller, Geoffrey D. 2010. « Intertextuality in Old Testament Research ». *Currents in Biblical Research* 9 (3): 283-309. <https://doi.org/10.1177/1476993X09359455>.

Moretti, Franco. 2000. « Conjectures on World Literature ». *new left review* 1: 54-68.

———. 2013. *Distant Reading*. New York: Verso.

Mosteller, Frederick, et David L. Wallace. 1963. « Inference in an Authorship Problem ». *Journal of the American Statistical Association* 58 (302): 275-309.

- Mufflikhah, Lailil, et Baharum Baharudin. 2009. « Document Clustering using Concept Space and Cosine Similarity Measurement ». In *2009 International Conference on Computer Technology and Development*, 58-62. <https://doi.org/10.1109/ICCTD.2009.206>.
- Muggleton, Stephen. 1991. « Inductive Logic Programming ». *New Generation Computing* 8: 295-318.
- Muggleton, Stephen, et Lud de Raedt. 1994. « Inductive Logic Programming: Theory and Method ». *Logic Programming* 19 (20): 629-79.
- Nalisnick, Eric T., et Henry S. Baird. 2013a. « Character-to-Character Sentiment Analysis in Shakespeare's Plays ». In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 479-83. Sofia, Bulgaria: Association for Computational Linguistics. <https://aclanthology.org/P13-2085>.
- . 2013b. « Extracting Sentiment Networks from Shakespeare's Plays ». In *2013 12th International Conference on Document Analysis and Recognition*, 758-62. <https://doi.org/10.1109/ICDAR.2013.155>.
- Noda, Edgar, et Alex A. Freitas. 2006. « Discovering Knowledge Nuggets with a Genetic Algorithm ». In *Data Mining and Knowledge Discovery Approaches Based on Rule Induction Technique*, édité par Evangelos Triantaphyllou et Giovanni Felici. Springer.
- Nowviskie, Bethany Paige. 2004. « Speculative Computing: Instruments for Interpretive Scholarship ». Thèse de doctorat, University of Virginia.
- Olsen, Stein Haugom. 1982. « The Meaning of a Literary Work ». *New Literary History* 14 (1).
- Padnick, Steven. 2019. « How daddy issues drive the marvel cinematic universe ». *Tor.com*. <https://www.tor.com/2018/06/13/how-daddy-issues-drive-the-marvel-cinematic-universe/>.

- Päpcke, Simon, Thomas Weitin, Katharina Herget, Anastasia Glawion, et Ulrik Brandes. 2022. « Stylometric similarity in literary corpora: Non-authorship clustering and Deutscher Novellenschatz ». *Digital Scholarship in the Humanities*, août. <https://doi.org/10.1093/llc/fqac039>.
- Pennington, Jeffrey, Richard Socher, et Christopher Manning. 2014. « Glove: Global Vectors for Word Representation ». In *EMNLP*, 14:1532-43. <https://doi.org/10.3115/v1/D14-1162>.
- Peter Szondi, Timothy Bahti. 1978. « Introduction to Literary Hermeneutics ». *New Literary History* 10 (1): 17-29.
- Piper, Andrew. 2017. « Think Small: On Literary Modeling ». *PMLA* 132 (3): 651-58.
- . 2018. *Enumerations : Data and Literary Study*. The University of Chicago Press.
- . 2020. *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*. Cambridge University Press.
- Pollak, Senja, Matej Martinc, et Katja Mihurko Poniz. 2020. « Natural Language Processing for Literary Text Analysis: Word-Embeddings-Based Analysis of Zofka Kveder's Work ». In, 33-42.
- Power, Daniel J. 2002. « Ask Dan! » *DSS News*.
- TODO : What is the "true story" about using data mining to identify a relationship between sales of beer and diapers? + <https://www.kdnuggets.com/news/2000/n14/8i.html>
- Quinlan, J. R. 1986. « Induction of Decision Trees ». *Machine Learning* 1: 81-106.
- Ramsay, Stephen. 2011. *Reading Machines*. Champaign: University of Illinois Press.
- Rhody, Lisa M. 2012. « Topic Model Data for Topic Modeling and Figurative

- Language ». *Journal of Digital Humanities* 2 (1).
- Robinson, G. D. 1995. « Paul Ricoeur and the Hermeneutics of Suspicion: A Brief Overview and Critique ». *PREMISE* 2 (8).
- Rockwell, Geoffrey, et Stefan Sinclair. 2016. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. MIT Press.
- Ross, John Robert. 1967. « Constraints on variables in syntax ». Thèse de doctorat, Massachusetts Institute of Technology. Dept. of Modern Languages; Linguistics.
- Ryan, Rory. 1985. « Pathologies of Epistemology in Literary Studies ». *Journal of Literary Studies* 1 (1): 3-42.
- Saldler, Matthew, et Natasha Regan. 2019. *Game Changer : AlphaZero's Ground-breaking Chess Strategies and the Promise of AI*. New In Chess.
- Sculley, D., et Bradley M. Pasanek. 2008. « Meaning and mining: the impact of implicit assumptions in data mining for the humanities ». *Literary and Linguistic Computing* 23 (4): 409-24.
- Searle, John R. 1980. « Minds, brains, and programs ». *The Behavioral and Brain Science* 3: 417-57.
- Searle, John Rogers, et François Latraverse. 1979. « Le sens littéral ». *Langue française*, no 42: 34-47.
- Shah, Hetan. 2018. « Algorithmic Accountability ». *Philosophical Transactions A* 376 (2128).
- Snow, C. P. 1959. *The Two Cultures and the Scientific Revolution*. The Syndics of the Cambridge University Press.
- So, Richard Jean. 2020. *Redlining Culture: A Data History of Racial Inequality and Postwar Fiction*. Columbia University Press.

- Souchier, Emmanuël. 1998. « L'image du texte pour une théorie de l'énonciation ». *Les cahiers de médiologie*, no 6: 137-45.
- Stamatatos, Efstathios. 2009. « A survey of modern authorship attribution methods ». *Journal of the American Society for Information Science and Technology* 60 (3): 538-56. <https://doi.org/https://doi.org/10.1002/asi.21001>.
- Tan, Songbo. 2005. « Neighbor-weighted K-nearest neighbor for unbalanced text corpus ». *Expert Systems with Applications* 28 (4): 667-71. <https://doi.org/https://doi.org/10.1016/j.eswa.2004.12.023>.
- Terras, Melissa, et Julianne Nyhan. 2016. « Father Busa's Female Punch Card Operatives ». In *Debates in the Digital Humanities*, édité par Matthew K. Gold et Lauren F. Klein. University of Minnesota Press.
- Thiselton, Anthony C. 1992. *New Horizons in Hermeneutics: The Theory and Practice of Transforming Biblical Reading*. Zondervan Publishing House.
- Thomson, Joseph John. 1897. « Cathode Rays ». *Philosophical Magazine and Journal of Science* 44 (269).
- Todorov, Tzvetan. 1969. « Structural Analysis of Narrative ». Traduit par Arnold Weinstein. *NOVEL: A Forum on Fiction* 3 (1): 70-76.
- Tracey, Diane H., et Lesley Mandel Morrow. 2006. *Lenses in Reading : An Introduction to Theories and Models*. The Guilford Press.
- Turing, A. M. 1950. « Computing Machinery and Intelligence ». *Mind: A Quarterly Review of Psychology and Philosophy* 59 (236): 433-60.
- Uexküll, Jakob von. 1982. *The Theory of Meaning*. De Gruyter.
- Underwood, Ted. 2014. « Theorizing Research Practices We Forgot to Theorize Twenty Years Ago ». *Representations* 127 (1): 64-72.
- . 2019. *Distant Horizons - Digital Evidence and Literary Change*. Chicago:

- The University of Chicago Press.
- Weil, Simone. 1946. « Essai sur la notion de lecture ». *Les Études philosophiques, Nouvelle Série* 1 (1): 13-19.
- Weisberg, Jonathan. 2009. « Locating IBE in the Bayesian Framework ». *Synthese* 167 (1): 125-43.
- Wendell, Inna Alekseyevna. 2021. « A Statistical Analysis of Genre Dynamics: Evolution of the Russian Five-Act Comedy in Verse in the Eighteenth and Nineteenth Centuries ». Thèse de doctorat, University of California.
- Wertheimer, Tiffany. 2022. « Blake Lemoine: Google fires engineer who said AI tech has feelings ». *BBC News*. <https://www.bbc.com/news/technology-62275326>.
- Wheeler, Wendy. 2016. *Expecting the Earth: Life, Culture, Biosemiotics*. Lawrence & Wishart.
- Wiedemann, Gregor, et Cornelia Fedtke. 2021. « From Frequency Counts to Contextualized Word Embeddings: The Saussurean Turn in Automatic Content Analysis ». In *Handbook of Computational Social Science*, 2:366-85. Routledge. <https://doi.org/10.4324/9781003025245-25>.
- Wittgenstein, Ludwig. 2004. *Recherches philosophiques*. tel Gallimard.