

Les procédés algorithmiques au prisme des approches sémiotiques

Introduction au 7^e numéro du *Cygne noir*

Emmanuelle CACCAMO & Fabien RICHERT

Cygne noir, no 7, 2019 : « Algorithmes »

Résumé

Assistants virtuels, objets connectés, intelligence artificielle, données massives, apprentissage machine ou apprentissage profond ; nul jour sans que les journaux ne rapportent une nouvelle spectaculaire sur les technologies algorithmiques dites « intelligentes ». Aucune activité humaine ou presque ne semble échapper à la mainmise algorithmique et à la volonté de contrôle sans limite qu'elle traduit. Ce constat sur l'emprise des technologies algorithmiques sur nos vies n'est pas nouveau, mais nous avons souhaité le transmuter en enquête en consacrant un numéro du *Cygne noir* à l'exploration des liens entre sémiotique et procédés algorithmiques. Relevant tantôt de la sémiotique théorique, tantôt de la sémiotique appliquée, les textes ici rassemblés travaillent à leur façon ces rapports. Ce dossier témoigne de la manière dont la sémiotique continue d'étonner par sa capacité de renouvellement conceptuel, par sa force de saisie d'objets toujours nouveaux et par sa disposition épistémologique à se « suturer » à de nombreux cadres théoriques. Il réaffirme en dernier lieu à quel point la sémiotique peut tenir un rôle éminemment critique et démystifiant, à contre-pied de la technolâtrie dominant l'espace social, dans un contexte marqué par une numérisation toujours plus croissante des activités humaines et sociales.

Pour citer cet article

CACCAMO, Emmanuelle & Fabien RICHERT, « Les procédés algorithmiques au prisme des approches sémiotiques. Introduction au 7^e numéro du *Cygne noir* », *Cygne noir*, no 7, 2019. En ligne : <http://www.revuecygnoir.org/numero/article/cn7-algorithmes-introduction> (consulté le xx/xx/xxxx).



Cet article de *Revue Cygne noir* est mis à disposition selon les termes de la licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Pas de Modification 2.5 Canada.

LES PROCÉDÉS ALGORITHMIQUES AU PRISME DES APPROCHES SÉMIOTIQUES. INTRODUCTION AU 7^e NUMÉRO DU *CYGNE NOIR*

Assistants virtuels, objets connectés, intelligence artificielle, données massives, apprentissage machine ou apprentissage profond ; nul jour sans que les journaux ne rapportent une nouvelle spectaculaire sur les technologies algorithmiques dites « intelligentes ». Parmi les annonces fantasques, l'on apprenait par exemple récemment l'existence du projet Augmented Eternity. Relié au Massachusetts Institute of Technology (MIT), le projet en intelligence artificielle vise à transformer les individus en agents conversationnels virtuels de manière à ce que, après leur mort, leurs proches puissent continuer à converser avec eux – du moins avec une sorte de version numérique très partielle de ce qu'ils auront été jadis¹. Comme en témoigne cet exemple, l'ensemble des dimensions de l'existence humaine ou presque ne semble échapper à la mainmise algorithmique et à la volonté de contrôle sans limite qu'elle traduit.

L'emprise des technologies algorithmiques sur nos vies s'observe à différents degrés dans le monde social. Elle se manifeste par une transformation de nos activités, de nos échanges, de nos rapports avec les autres. Les procédés algorithmiques influencent également nos comportements, nos attitudes, nos habitudes de pensée, nos manières de communiquer. Partant de ce constat, nous avons souhaité consacrer un numéro du *Cygne noir* à l'exploration des liens entre sémiotique et procédés algorithmiques. À l'origine, nous nous demandions quel rôle la sémiotique peut jouer dans la compréhension des phénomènes algorithmiques. Qu'apporte en effet la sémiotique à la compréhension de l'intelligence artificielle, de l'Internet des objets, des données massives (*big data*), des médias socionumériques, notamment? Comment peut-elle éclairer ces objets si contemporains?

Par cette courte introduction, notre projet initial était de faire le point sur les rapports que l'étude des signes entretient avec les algorithmes. Il fallait toutefois se rendre à l'évidence : dresser un tel portrait est un projet hautement ambitieux, si ce n'est impossible, et ce pour plusieurs raisons. La notion d'algorithme elle-même est sujette à de nombreuses définitions, se confondant parfois avec la notion de « numérique ». D'abondantes recherches en sémiotique, tant théoriques qu'appliquées, menées dans différents domaines et disciplines (communication, design, éducation, ingénierie, informatique, arts et lettres, notamment) coexistent. Enfin, les objets et les corpus étudiés

sont d'une telle hétérogénéité qu'aucun travail de recension ne saurait être qualifié d'exhaustif.

Notre enquête nous a alors menés à formuler trois grands constats. Ce faisant nous suggérons au lectorat des pistes bibliographiques pour explorer plus avant certains sujets reliés à la thématique de ce numéro.

(1) La sémiotique s'est emparée très tôt de la question algorithmique.

Les premières études sémiotiques sur les algorithmes et les systèmes « intelligents » émergent en effet dès les années 1960-1970, en Russie et aux États-Unis². Précurseur de la sémiotique computationnelle, Dmitri Pospelov invente à Moscou le *Semiotic Situational Control* (SSC), qui s'appuie sur une série de règles syntactiques et sémantiques pour manipuler des systèmes (commerciaux, industriels, sociaux, politiques) dont la complexité échappe aux techniques de contrôle plus conventionnelles. À la même époque, aux États-Unis, Eugene Pendergraft crée l'Autognome. Inspiré notamment par la sémiotique de Charles S. Peirce, Pendergraft imagine une machine auto-apprenante capable de collecter des informations diverses à partir de l'environnement avec lequel elle interagit³. Longtemps ignorées, ces études qui allient théories sémiotiques et travaux en intelligence artificielle ont connu un nouvel élan dans les années 1990.

(2) La rencontre entre la sémiotique et l'objet « algorithme » (et les thèmes connexes) a donné lieu à un foisonnement d'enquêtes dans de nombreux domaines disciplinaires, et pas seulement en sciences humaines et sociales, comme on pourrait le penser à première vue. Elle a aussi donné lieu à des champs spécifiques plus ou moins institutionnalisés.

Les procédés algorithmiques font par exemple l'objet de recherches sémiotiques et narratologiques en arts et lettres. La sémiotique permet d'observer les productions narratives machiniques⁴ et semble parfois accompagner un curieux espoir d'automatisation de l'écriture de scénarios fictionnels⁵.

L'étude des représentations fictionnelles des modèles algorithmiques, de l'intelligence artificielle, des données massives ou plus largement des nouvelles technologies numériques dans les productions culturelles (cinéma, série télévisée, littérature, jeu vidéo, etc.) forme par ailleurs un champ d'investigation sémiotique à développer⁶.

Dans les études en communication, les données massives⁷, l'Internet des objets⁸ et les médias socionumériques constituent quelques-uns des objets de prédilection des chercheur·e·s développant des approches sémiotiques. Les médias socionumériques constituent à ce titre un terrain d'enquête fructueux pour les sémioticien·ne·s intéressé·e·s par les formations identitaires et la constitution du sujet⁹, par la médiation de rites

funéraires¹⁰, par les logiques de captation de l'attention et de manipulation¹¹ et par la circulation de fausses nouvelles¹².

Tous domaines confondus, plusieurs travaux en sémiotique portent plus largement sur les interfaces humain-machine¹³, sur l'analyse du langage assisté par ordinateur¹⁴, sur la visualisation de données numériques (ou datavisualisation¹⁵), sur l'énonciation, la syntaxe et la rhétorique des langages de programmation¹⁶. Parmi tous ces travaux, notons que certains s'inscrivent dans des champs dont la reconnaissance est parfois lacunaire, tels que la sémiotique computationnelle¹⁷, la robot-sémiotique¹⁸ et la cybersémiotique¹⁹.

Ricardo Gudwin et João Queiroz définissent la sémiotique computationnelle comme un champ qui a émergé de la recherche sur l'intelligence artificielle²⁰. Mobilisant théories sémiotiques et philosophie pragmatiste, ce courant de pensée constitue pour eux une « alternative » aux approches généralement valorisées dans la recherche traditionnelle sur les systèmes intelligents, à l'exemple du cognitivisme. Selon Jean-Guy Meunier, plusieurs formes de sémiotiques computationnelles développant des projets différents se seraient succédé²¹. Meunier distingue en effet au moins deux grands programmes scientifiques : d'un côté, les recherches ayant débuté dans les années 1970 portant sur les systèmes intelligents considérés comme des systèmes sémiotiques à analyser et à définir ; de l'autre, les recherches développées dans le cadre des « humanités numériques », qui utilisent des outils algorithmiques pour traiter certains objets culturels (à l'exemple des textes) et participent à l'élaboration critique des méthodes et des technologies d'analyse²².

La robot-sémiotique, quant à elle, peut être définie comme étant le fruit de la rencontre entre l'ingénierie robotique, l'intelligence artificielle et les théories sémiotiques. En son sein, plusieurs recherches en sémiotique s'intéressent à l'interaction entre les humains et les robots (*human-robot interaction*)²³. On peut citer à titre d'exemple les travaux de Stéphanie Walsh Matthews et Jamin R. Pelkey autour du projet en cours intitulé « Studying speech practices of children with ASD through Human Robot Interaction »²⁴. Les chercheur-e-s travaillent d'un point de vue sémiotique sur la manière dont l'interaction robotique permet de mieux comprendre les pratiques langagières des enfants souffrant d'un trouble du spectre de l'autisme.

Mobilisant la biosémiotique, Winfried Nöth s'est par ailleurs interrogé sur l'*Umwelt* des robots²⁵. Son travail pourrait se rattacher à la robot-sémiotique, mais aussi au champ de la cybersémiotique. D'après Søren Brier, qui en développe le concept, la cybersémiotique désigne un courant réunissant les théories cybernétiques de l'information (seconde cybernétique), la biosémiotique et la sémiotique peircienne²⁶.

Un autre ensemble de travaux allie sémiotique et algorithmes. Il s'agit de ceux portant sur le sémiocapitalisme, dont les plus importants ont été réalisés par Maurizio Lazzarato²⁷, Franco Berardi²⁸ et Gary Genosko²⁹. Ces travaux ont en commun de s'inspirer du philosophe et psychanalyste Félix Guattari qui, dès les années 1970, a insisté sur l'existence de sémiotiques a-signifiantes constituée de signaux artificiels, qualifiés de « points-signes », qui fonctionnent indépendamment des effets de sens qu'ils peuvent engendrer. Le champ du sémiocapitalisme formule une critique d'ordre sociopolitique qui cherche à établir l'influence de certaines procédures algorithmiques sur les modes de subjectivation, suivant le processus de devenir-machine de l'humain (automatisation des comportements et des attitudes) et du devenir-humain de la machine (développement de l'intelligence artificielle)³⁰.

(3) La réflexion sur les signes, le sens et les lois de l'interprétation forme un terrain hautement fertile pour la compréhension des procédés algorithmiques. Dans ce cadre, la sémiotique tient un rôle éminemment critique en regard des procédés algorithmiques et des nouvelles orientations de la science actuelle.

Indépendamment de la variété des objets sur lesquels elles se penchent, les études sémiotiques jouent un rôle fondamental dans la compréhension et la problématisation épistémologique des modèles algorithmiques, des systèmes numériques dits « intelligents » et des sémiotiques dites « artificielles »³¹.

Rappelons que les recherches en intelligence artificielle se concentrent par exemple à faire émerger des machines dites « sensibles » capables de saisir certains aspects des choses du monde avant de les transformer automatiquement en flux de données. Accompagnées d'un champ lexical propre (apprentissage, décision, intelligence, traduction, langage, cognition, sémantique, etc.), ces recherches investissent des champs d'application qui intègrent, dans leurs problématiques centrales, les questions du sens, de la signification et de l'interprétation. Qu'il s'agisse de représenter symboliquement la connaissance, de formaliser des modes de raisonnement, de simuler des aptitudes humaines, comme l'apprentissage, la reconnaissance de la parole ou la prise de décision, les recherches en intelligence artificielle commandent des domaines d'investigation finalement très proches de ceux de la sémiotique.

À l'égard de tels procédés algorithmiques, les travaux en sémiotique se penchent alors sur différents aspects. Si les algorithmes fonctionnent à partir d'un pseudolangage qui mobilise des opérateurs variés (logiques, arithmétiques, relationnels, notamment), ordonne des instructions, appelle des fonctions, affecte des valeurs à des variables, ils sont alors susceptibles d'être analysés en tant que composantes d'un *système sémiotique* à part entière, avec des caractéristiques qui lui sont propres³².

Les enquêtes menées visent de surcroît à saisir les *types de signes* impliqués dans les procédures algorithmiques. Maude Bonenfant et ses collègues mettent par exemple au jour l'« hypersymbole » ou l'« hypersymbolisation » à l'œuvre dans la production de sens reliée à l'exploitation des données massives³³. À partir des catégories peirciennes, les chercheur·e·s associent la numérisation du monde à un processus sémiotique hypersymbolique qui a pour effet de naturaliser les données qui tiennent pour la réalité du monde social qu'elles sont censées refléter.

Les approches sémiotiques évaluent les phénomènes *d'interprétation algorithmique* et les *formes de signification* produites par les machines³⁴. Elles cherchent à déterminer à quel point les données sont signifiantes³⁵.

En somme, elles s'appliquent à *définir les machines d'un point de vue sémiotique*. Dans ce contexte, les travaux de Charles S. Peirce sont devenus des références incontournables pour conceptualiser les ordinateurs comme des machines sémiotiques³⁶ – Peirce ayant proposé une formalisation rigoureuse de son modèle sémiotique et de sa classification des signes. Son concept de *sémiose* est particulièrement fécond pour réfléchir aux modalités suivant lesquelles les machines seraient en mesure de simuler une partie de l'activité psychique humaine (raisonnement, pensée logique, interprétation du langage, etc.) par des procédés algorithmiques plus ou moins complexes³⁷.

Si certains travaux en sémiotique, en particulier en ingénierie, accompagnent parfois le développement de machines impliquant des modèles algorithmiques, il faut souligner que la sémiotique permet d'en révéler les rouages et, parfois, d'en déconstruire certains mythes. Pour reprendre Jean-Guy Meunier,

bien qu'une technologie crée l'illusion, souvent spectaculaire d'ailleurs, de manipuler du sens et de la signification, elle ne peut, par définition que manipuler des signes sans signification. À l'entrée, un ordinateur ne reçoit que des symboles et il ne peut que les reconfigurer en d'autres symboles. [...] Une sémantique peut toujours être associée aux données, aux traitements et aux extrants, mais elle est toujours externe. Et c'est l'objectif même de la sémiotique de travailler au dévoilement de cette signification³⁸.

La sémiotique invite donc à bien repositionner la machine en tant qu'outil de manipulation et de production de symboles au service de nos propres capacités sémiotiques. Cependant, au vu de la sophistication actuelle des procédés algorithmiques (gestion du langage et de la parole, reconnaissance faciale, résolution de certains problèmes auparavant impossibles à traiter par certains spécialistes), la séparation nette entre ce qui revient à l'humain ou à la machine tend à devenir problématique.

C'est pourquoi, en dernier ressort, la sémiotique endosse un rôle nécessaire dans un projet d'*épistémologie critique* à l'égard des nouvelles orientations scientifiques en

intelligence artificielle – par ailleurs fortement ancrées dans un paradigme de productivité économique. Ce faisant, la sémiotique tient aussi un rôle critique d'un point de vue sociopolitique. En effet, dans le contexte des « humanités numériques » et du *big data*, les universitaires sont fortement encouragés à recourir à des méthodes issues de l'intelligence artificielle pour analyser des phénomènes sociaux. D'aucuns annoncent à ce propos un changement épistémologique, ou une « fin de la théorie » au profit d'une science menée par les données (*data-driven science*)³⁹. Certains travaux en sémiotique examinent et critiquent ce changement de paradigme et analysent la rhétorique des discours scientifiques qui en font la promotion. C'est par exemple le cas du travail de Lucile Crémier et ses collègues dont nous publions une traduction dans ce dossier⁴⁰.

*

Dans ce 7^e numéro du *Cygne noir*, nous avons tâché de poursuivre les recherches sémiotiques sur les algorithmes.

Le premier texte qui ouvre ce dossier nous offre les clés de lecture essentielles pour aborder la sémiotique computationnelle. Dans son texte « L'analyse de texte assistée par ordinateur : introduction à l'un des champs fondamentaux de la sémiotique computationnelle⁴¹ », Davide Pulizzotto commence par situer la sémiotique par rapport à l'objet d'étude qu'est l'ordinateur. Il propose de classer la sémiotique computationnelle en trois branches. L'auteur aborde ensuite l'un des champs d'application les plus marquants de ce courant, à savoir la pratique de l'analyse de texte assistée par ordinateur (ATO). Il propose notamment de discuter des limites liées à cette pratique en rappelant l'importance de certaines théories sémiotiques (celle de Saussure et de Hjelmslev notamment) pour développer des approches automatisées au traitement du langage naturel qu'il convient d'adapter au phénomène d'accumulation massive de données (*big data*).

Le thème de ce numéro est aussi l'occasion pour Jean-Guy Meunier de revenir sur les enjeux de la modélisation formelle en sémiotique computationnelle⁴². Pionnier de ce courant au Québec, Meunier rappelle le rôle déterminant d'un modèle conceptuel pour la modélisation formelle dans le domaine de la sémiotique mobilisant des modèles informatiques. Il montre également quelles contraintes une modélisation computationnelle impose à la modélisation formelle et en expose clairement les limites ainsi que les critiques dont elle peut faire l'objet. Comme l'annonce l'auteur, si la sémiotique computationnelle s'inscrit dans une épistémè positiviste, empirique et possiblement expérimentale, il faut s'attendre à ce qu'elle fasse l'objet des critiques que l'on adresse habituellement à ce type de théories de la connaissance.

Le troisième texte que nous proposons, intitulé « Données brutes ou hypersymboles? Signification et données numériques, entre processus discursif et procédure machinique⁴³ », démontre sans conteste le potentiel de renouvellement théorique de la sémiotique. Après avoir proposé une analyse critique de certains discours qui vantent de manière emphatique les mérites des *big data* dans la science des données, Lucile Crémier, Maude Bonenfant et Laura Iseut Lafrance St-Martin développent un modèle sémiotique qui relativise la prétendue neutralité des données. En s'inspirant de la philosophie pragmatiste de Peirce, les auteur-e-s développent la notion suggestive d'hypersymbole pour souligner l'écart entre les données collectées dans le cadre du *big data* et leurs référents. L'article met aussi l'accent sur les importantes différences qui existent entre *procédure sémiotique* et *processus sémiotique*. De cette manière, Crémier, Bonenfant et Lafrance St-Martin réaffirment avec force l'importance de la sémiotique pour mieux critiquer le mouvement croissant d'hypersymbolisation de la réalité auquel contribuent les divers usages liés à l'accumulation massive des données.

Dans son texte « Médias socionumériques alternatifs. Étude sémiotique et rhétorique de diaspora*⁴⁴ », Emmanuelle Caccamo propose quant à elle une critique de la rhétorique des médias socionumériques libres. L'originalité de la perspective sémiotique adoptée par l'auteure repose sur la prise en compte non seulement des discours et des argumentaires de certains médias socionumériques, mais aussi du design d'interaction des interfaces et de leurs multiples parcours de navigation en ligne. En s'appuyant notamment sur la notion de rhétorique procédurale, elle montre que les médias socionumériques dits « libres » ne parviennent pas toujours à se différencier clairement des plateformes propriétaires hégémoniques. L'auteure nous invite alors à réfléchir au statut « alternatif » revendiqué par certains de ces médias socionumériques.

Ancré dans la théorie critique, le texte de Fabien Richert, « Du fétichisme des sémiomarchandises "parlantes" à l'asservissement machinique : le cas des assistants personnels intelligents⁴⁵ », enquête sur une classe d'objets dont font partie Amazon Echo et Google Home. À rebours de l'idée doxique selon laquelle les effets de la technologie dépendent des usages ou de dispositions morales, l'auteur nous invite à déplacer notre regard et à réfléchir aux logiques normatives et performatives intrinsèques à ces objets parlants. Richert démontre notamment que ces technologies reposent sur un régime de vérité particulier et qu'ils renouvellent les formes d'interpellation althusseriennes associées à un pouvoir de former un type de sujet spécifique. Par le biais du concept de fétichisme de la marchandise de Marx, l'auteur s'attache à déconstruire le rapport que les individus entretiennent avec ces objets paradigmatiques du sémiocapitalisme ainsi que l'imaginaire fallacieux dont ces derniers sont porteurs. Finalement, en revenant sur

les conditions de fabrication de ces technologies, Richert met en lumière les logiques d'asservissement machinique qui leur sont propres.

Enfin, dans le texte « Les aléas de la subjectivité sous le sémiocapitalisme : Félix Guattari et le potentiel politique des hybrides humains-machines⁴⁶ », Gary Genosko réinvestit les réflexions de Félix Guattari et s'intéresse aux modes de production de la subjectivité en lien avec l'informatique généralisée. Mobilisant le concept de « sémiocapitalisme » forgé par Franco Berardi, Genosko se penche sur la manière dont la subjectivité des travailleurs est transformée en matière première et intégrée aux « sémiomarchandises » dans un contexte d'exploitation cognitive liées au travail dit immatériel. Dans cet article issu d'une conférence prononcée à Kyoto en 2018, Genosko définit la subjectivité d'un point de vue machinique et se questionne sur l'éventualité qu'une hybridation humain-machine puisse constituer une occasion pour le sujet de s'émanciper de la logique capitaliste.

Relevant tantôt de la sémiotique théorique, tantôt de la sémiotique appliquée, les textes ici rassemblés travaillent à leur façon les rapports entre sémiotique et procédés algorithmiques. Ce dossier témoigne de la manière dont la sémiotique continue d'étonner par sa capacité de renouvellement conceptuel, par sa force de saisie d'objets toujours nouveaux et par sa disposition épistémologique à se « suturer » à de nombreux cadres théoriques. Il réaffirme en dernier lieu à quel point la sémiotique peut tenir un rôle éminemment critique et démystifiant, à contre-pied de la technolâtrie dominant l'espace social, dans un contexte marqué par une numérisation toujours plus croissante des activités humaines et sociales.

Bibliographie

- ALAC, Morana, Javier MOVELLAN & Fumihide TANAKA, « When a Robot is Social: Spatial Arrangements and Multimodal Semiotic Engagement in The Practice of Social Robotics », *Social Studies of Science*, vol. 41, no 6, 2011, p. 893-926.
- ANDERSEN, Peter B. & Berit HOLMQVIST, « Interactive fiction: Artificial intelligence as a mode of sign production », *AI & Society*, vol. 4, no 4, 1989, p. 291-313.
- ANDERSEN, Peter B., *A Theory of Computer Semiotics: Semiotic Approaches to Construction and Assessment of Computer Systems*, Cambridge, Cambridge University Press, 1990.
- ANDRY, Tiffany, « Une approche sémiotique de la conception d'outils d'analyse de données massives », *Communication et organisation*, no 54, 2018, p. 133-148.
- BARRICELLI, Barbara Rita & Stefano VALTOLINA, « A Semiotic Approach to EUD for the Internet of Things », dans B. R. Barricelli et al. (dir.), *CoPDA 2015. Cultures of Participation in the Digital Age*, CEUR Workshop Proceedings, vol. 1641, 2015, p. 45-51. En ligne : <<http://ceur-ws.org/Vol-1641/paper8.pdf>>
- BENTO, Luis Felipe Hussin, Raquel O. PRATES & Luiz CHAIMOWICZ, « Using Semiotic Inspection Method to Evaluate a Human-Robot Interface », *2009 Latin American Web Congress*, Merida, Yucatan, 2009, p. 77-84.
- BERARDI, Franco, *Tueries : forcenés et suicidaires à l'ère du capitalisme absolu*, trad. de l'anglais par P. Dardel, Montréal, Lux, 2016.
- BONACCORSI, Julia, « Le monde de l'Open data : les jeux sémiotiques et esthétiques de la "visualisation" comme rhétorique de la transparence », dans S. Badir & F. Provenzano (dir.), *Pratiques émergentes et pensée du medium*, Louvain-la-Neuve, L'Harmattan, 2016, p. 131-150.
- BONENFANT, Maude, André MONDOUX, Marc MÉNARD & Maxime OUELLET, « Big Data, médiation symbolique et gouvernementalité », dans É. Broudoux & G. Charton (dir.), *Big Data - Open Data : Quelles valeurs? Quels enjeux?*, Rabat, 2015, Louvain-la-Neuve, De Boeck Supérieur, 2015, p. 31-41.
- BORGES, Priscila Monteiro & Renira Rampazzo GAMBARATO, « The Role of Beliefs and Behavior on Facebook: A Semiotic Approach to Algorithms, Fake News, and Transmedia Journalism », *International Journal of Communication*, vol. 13, 2019, p. 603-618.
- BRIER, Søren, *Cybersemiotics: Why Information is Not Enough!*, Toronto, University of Toronto Press, 2008.
- CACCAMO, Emmanuelle, *Imaginer les technologies de mémoire totale avec la science-fiction audiovisuelle occidentale (1990-2016). Étude sémiotique, intermédiaire et technocritique des représentations de la mémoire personnelle*, thèse de doctorat en sémiologie, Montréal, Université du Québec à Montréal, 2017.

- , « Médias socionumériques alternatifs. Étude sémiotique et rhétorique de diaspora* », *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnenoir.org/numero/article/caccamo-medias-socionumeriques-alternatifs>>.
- COMPAGNO, Dario & Matteo TRELEANI, « Introduction to Meaningful data/Données signifiantes », *Semiotica*, vol. 2019, no 230, 2019, p. 1-17.
- CRÉMIER, Lucile, Maude BONENFANT & Laura Iseut LAFRANCE ST-MARTIN, « Raw data or hypersymbols? Meaning-making with digital data, between discursive processes and machinic procedures », *Semiotica*, no 230, 2019, p. 189-212.
- , « Données brutes ou hypersymboles? Signification et données numériques, entre processus discursif et procédure machinique », *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnenoir.org/numero/article/cremier-et-al-hypersymboles>>.
- DE SOUZA, Clarisse Sieckenius, « The Semiotic Engineering of User Interface Languages », *International Journal of Man-Machine Studies*, vol. 39, no 5, 1993, p. 753-773.
- , *Semiotic Engineering of Human-Computer Interaction*, Cambridge, The MIT Press, 2005.
- FERRARI, Marcia Ikezaki, Silvia Amelia BIM & Plinio Thomaz AQUINO Jr., « The Signs of Semiotic Engineering in the IoT Interaction Design », dans C. Baldwin (dir.), *Advances in Intelligent Systems and Computing*, 2018, p. 421-433.
- GENOSKO, Gary, *Critical Semiotics: Theory, from Information to Affect*, New York, Bloomsbury Academic, 2016.
- , « Les aléas de la subjectivité sous le sémiocapitalisme : Félix Guattari et le potentiel politique des hybrides humains-machines », trad. de l'anglais (Canada) par S. Levesque & F. Richert, *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnenoir.org/numero/article/genosko-semiocapitalisme-guattari>>.
- GEORGES, Fanny, « Identité numérique et représentation de soi : analyse sémiotique et quantitative de l'emprise culturelle du web 2.0 », *Réseaux*, no 154, 2009, p. 165-193.
- GOMEZ-MEJIA, Gustavo, *Les fabriques de soi? Identité et industrie sur le web*, Paris, MKF éditions, 2016.
- GUARDA, Rebeka F., Marcia P. OHLSON & Anderson V. ROMANINI, « Disinformation, dystopia and post-reality in social media: A semiotic-cognitive perspective », *Education for Information*, vol. 34, no 3, 2018, p. 185-197.
- GUATTARI, Félix, *La révolution moléculaire*, Paris, Les prairies ordinaires, 2012 [1977].
- GUDWIN Ricardo & João QUEIROZ, *Semiotics and Intelligent Systems Development*, Hershey, Idea Group Pub., 2007.

- , « Towards an Introduction to Computational Semiotics », dans C. Thompson & H. Hexmoor (dir.), *International Conference on Integration of Knowledge Intensive Multi-Agent Systems*, 2005. En ligne : <<https://ieeexplore.ieee.org/document/1427113>>.
- HAYLES, N. Katherine, « Can Computers Create Meanings? A Cyber/Bio/Semiotic Perspective », *Critical Inquiry*, vol. 46, no 1, 2019, p. 32-55.
- HEYMAN, Rob, *Facebook & Users: Who is Using Who? A material semiotic approach to the irreversibilisation of Facebook as a case of lifeworld colonisation by social media*, thèse de doctorat en communication, Bruxelles, Vrije Universiteit Brussel, 2016.
- JULLIARD, Virginie & Fanny GEORGES, « Produire le mort. Pratiques d'écriture et travail émotionnel des deuillements et des deuillemes sur Facebook », *Réseaux*, vol. 4, no 210, 2018, p. 89-116.
- KETNER, Kenneth Laine, « Peirce and Turing: Comparisons and Conjectures », *Semiotica*, vol. 68, no 1-2, 1988, p. 33-62.
- KITCHIN, Rob, « Big Data, new epistemologies and paradigm shifts », *Big Data & Society*, vol. 1, no 1, 2014, p. 1-12.
- LAZZARATO, Maurizio, *Signs and Machines: Capitalism and the Production of Subjectivity*, Los Angeles, Semiotext, 2014.
- MANSARD, Nicolas, *Semiotics of Motion: Toward a Robotics Programming Language*, HDR, Toulouse, Université Paul Sabatier - Toulouse III, 2013. En ligne : <<https://tel.archives-ouvertes.fr/tel-01061112/document>>.
- MEUNIER, Jean-Guy & Dominic FOREST, « Lecture et analyse conceptuelle assistée par ordinateur : premières expériences », dans J.-P. Desclés & F. Le Priol (dir.), *Annotation automatique et recherche d'information*, Paris, Hermès, 2009, p. 211-230.
- MEUNIER, Jean-Guy, « Vers une sémiotique computationnelle? », *Applied Semiotics/ Sémiotique appliquée*, no 26, 2018, p. 98. En ligne : <<http://french.chass.utoronto.ca/as-sa/ASSA-No26/26-6.pdf>>.
- , « Enjeux de la modélisation formelle en sémiotique computationnelle », *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnenoir.org/numero/article/meunier-modelisation-semiotique-computationnelle>>.
- , « Sémiotique computationnelle : sémiotique de la computation ou la computation en sémiotique », conférence, Montréal, Université de Montréal, 2020.
- NADIN, Mihai, « Semiotic Machine », *Public Journal of Semiotics*, vol. 1, no 1, 2007, p. 57-75.
- NÖTH, Winfried, « Semiosis and the Umwelt of a Robot », *Semiotica*, no 134, 2001, p. 695-699.

- O'LEARY, Daniel E., « "Big Data", the "Internet of Things" and the "Internet of Signs" », *Intelligent Systems in Accounting, Finance and Management*, vol. 20, no 1, 2013, p. 53-65.
- PULIZZOTTO, Davide, « L'analyse de texte assistée par ordinateur : introduction à l'un des champs fondamentaux de la sémiotique computationnelle », *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnoir.org/numero/article/pulizzotto-ato>>.
- QUEIROZ, João & Floyd MERRELL, « On Peirce's Pragmatic Notion of Semiosis—A Contribution for the Design of Meaning Machines », *Minds & Machines*, vol. 19, no 1, 2009, p. 129-143.
- RANDELL, Joseph, « The relevance of Peircean semiotic to computational intelligence augmentation », *SEED – Semiotics, Evolution, Energy, and Development*, vol. 3, no 3, 2003, p. 5-36.
- RICHERT, Fabien, « La critique de la valeur à l'ère du sémiocapitalisme », *Cygne noir*, no 4, 2016. En ligne : <<http://revuecygnoir.org/numero/article/critique-valeur-semiocapitalisme>>.
- , « Du fétichisme des sémiomarchandises "parlantes" à l'asservissement machinique : le cas des assistants personnels intelligents », *Cygne noir*, no 7, 2019. En ligne : <<http://www.revuecygnoir.org/numero/article/richt-fetichisme-semiomarchandises>>.
- RIEGER, Burghard B., « Computing Fuzzy Semantic Granules from Natural Language Texts. A Computational Semiotic Approach to Understanding Word Meanings », dans M. H. Hamza (dir.), *Artificial Intelligence and Soft Computing, Proceedings of the IASTED International Conference*, Anaheim, IASTED/Acta Press, 1999, p. 475-479.
- ROUVROY, Antoinette & Thomas BERNIS, « Gouvernamentalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation? », *Réseaux*, vol. 1, no 177, 2013, p. 168-173.
- SALAMON, Daniel Kurniawan & Ryan Pratama SUTANTO, « Representation of Power in a Smartphone - Semiotic Analysis of Smartphone Culture in Watch Dog Game », *4th Bandung Creative Movement International Conference on Creative Industries*, 2017. En ligne : <<https://www.atlantis-press.com/proceedings/bcm-17/25892198>>.
- SALLABERRY, Arnaud, *Visualisation d'information : de la théorie sémiotique à des exemples pratiques basés sur la représentation de graphes et d'hypergraphes*, thèse de doctorat en informatique, Bordeaux, Bordeaux 1, 2011.
- SEQUEIRA, João Silva & Maria Isabel RIBEIRO, « Semiotics and Human-Robot Interaction », *Informatics in Control Automation and Robotics*, vol. 15, 2008, p. 277-291.

- STEINER, Pierre, C.S. *Peirce and Artificial Intelligence: Historical Heritage and (New) Theoretical Stakes*, dans V. C. Müller (dir.), *Philosophy and Theory of Artificial Intelligence*, Berlin/Heidelberg, Springer, 2013, p. 265-276.
- TANAKA-ISHII, Kumiko, *Semiotics of programming*, New York, Cambridge University Press, 2010.
- VALLE, Andrea & Alessandro MAZZEI, « Sapir–Whorf vs Boas–Jakobson Enunciation and the Semiotics of Programming Languages », *Lexia*, 27-28, 2017, p. 505-525.
- VICKERS, Paul, Joe FAITH & Nick ROSSITER, « Understanding Visualization: A Formal Approach Using Category Theory and Semiotics », *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no 6, 2013, p. 1048-1061.
- WALSH MATTHEWS, Stéphanie & Marcel DANESI, « AI: A Semiotic Perspective », *Chinese Semiotic Studies*, vol. 15, no 2, 2019, p. 199-216.
- WINFIELD, Alan F. T., « When Robots Tell Each Other Stories: The Emergence of Artificial Fiction », dans R. Walsh & S. Stepney (dir.), *Narrating Complexity*, Cham, Springer, 2018, p. 39-47.
- ZIEMKE, Tom, « Robosemiotics and embodied enactive cognition », *SEED – Semiotics, Evolution, Energy, and Development*, vol. 3, no 3, 2003, p. 112-124.

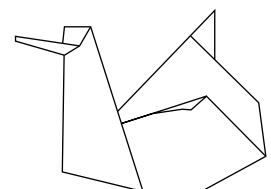
Notes

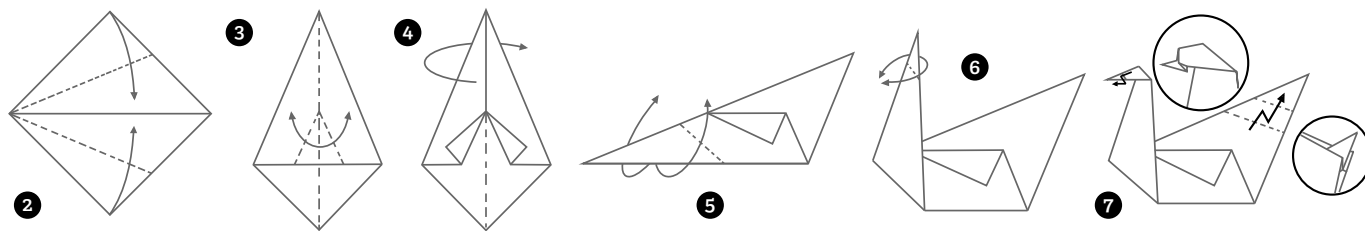
- 1 Augmented Eternity est développé par Hossein Rahnama depuis 2016. Voir en ligne : <<https://www.media.mit.edu/projects/augmented-eternity/overview/>>.
- 2 R. GUDWIN & J. QUEIROZ, « Towards an Introduction to Computational Semiotics », dans C. Thompson & H. Hexmoor (dir.), *International Conference on Integration of Knowledge Intensive Multi-Agent Systems*, 2005. En ligne : <<https://ieeexplore.ieee.org/document/1427113>>.
- 3 *Idem*.
- 4 A. F. T. WINFIELD, « When Robots Tell Each Other Stories: The Emergence of Artificial Fiction », dans R. Walsh & S. Stepney (dir.), *Narrating Complexity*, Cham, Springer, 2018, p. 39-47.
- 5 P. B. ANDERSEN & B. HOLMQVIST, « Interactive fiction: Artificial intelligence as a mode of sign production », *AI & Society*, vol. 4, no 4, 1989, p. 291-313.
- 6 D. K. SALAMOON & R. P. SUTANTO, « Representation of Power in a Smartphone - Semiotic Analysis of Smartphone Culture in Watch Dog Game », 4th Bandung Creative Movement International Conference on Creative Industries, 2017. En ligne : <<https://www.atlantis-presse.com/proceedings/bcm-17/25892198>> ; E. CACCAMO, *Imaginer les technologies de mémoire totale avec la science-fiction audiovisuelle occidentale (1990-2016). Étude sémiotique, inter-médiale et technocritique des représentations de la mémoire personnelle*, thèse de doctorat en sémiologie, Montréal, Université du Québec à Montréal, 2017.

- 7 M. BONENFANT *et al.*, « Big Data, médiation symbolique et gouvernementalité », dans É. Broudoux & G. Charton (dir.), *Big Data - Open Data : Quelles valeurs? Quels enjeux?*, Rabat, 2015, Louvain-la-Neuve, De Boeck Supérieur, 2015, p. 31-41.
- 8 D. E. O'LEARY, « "Big Data", the "Internet of Things" and the "Internet of Signs" », *Intelligent Systems in Accounting, Finance and Management*, vol. 20, no 1, 2013, p. 53-65 ; B. R. BARRICELLI, & S. VALTOLINA, « A Semiotic Approach to EUD for the Internet of Things », dans B. R. Barricelli *et al.* (dir.), *CoPDA 2015. Cultures of Participation in the Digital Age*, CEUR Workshop Proceedings, vol. 1641, 2015, p. 45-51. En ligne : <<http://ceur-ws.org/Vol-1641/paper8.pdf>> ; M. I. FERRARI, S. A. BIM & P. T. AQUINO Jr., « The Signs of Semiotic Engineering in the IoT Interaction Design », dans C. Baldwin (dir.), *Advances in Intelligent Systems and Computing*, 2018, p. 421-433.
- 9 F. GEORGES, « Identité numérique et représentation de soi : analyse sémiotique et quantitative de l'emprise culturelle du web 2.0 », *Réseaux*, no 154, 2009, p. 165-193 ; G. GOMEZ-MEJIA, *Les fabriques de soi? Identité et industrie sur le web*, Paris, MKF éditions, 2016.
- 10 V. JULLIARD & F. GEORGES, « Produire le mort. Pratiques d'écriture et travail émotionnel des deuilés et des deuilées sur Facebook », *Réseaux*, vol. 4, no 210, 2018, p. 89-116.
- 11 P. M. BORGES & R. R. GAMBARATO, « The Role of Beliefs and Behavior on Facebook: A Semiotic Approach to Algorithms, Fake News, and Transmedia Journalism », *International Journal of Communication*, vol. 13, 2019, p. 603-618 ; R. HEYMAN, *Facebook & Users: Who is Using Who? A material semiotic approach to the irreversibilisation of Facebook as a case of lifeworld colonisation by social media*, thèse de doctorat en communication, Bruxelles, Vrije Universiteit Brussel, 2016.
- 12 R. F. GUARDA, M. P. OHLSON & A. V. ROMANINI, « Disinformation, dystopia and post-reality in social media: A semiotic-cognitive perspective », *Education for Information*, vol. 34, no 3, 2018, p. 185-197.
- 13 C. S. de SOUZA, *Semiotic Engineering of Human-Computer Interaction*, Cambridge, The MIT Press, 2005 ; C. S. de SOUZA, « The Semiotic Engineering of User Interface Languages », *International Journal of Man-Machine Studies*, vol. 39, no 5, 1993, p. 753-773.
- 14 B. B. RIEGER, « Computing Fuzzy Semantic Granules from Natural Language Texts. A Computational Semiotic Approach to Understanding Word Meanings », dans M. H. Hamza (dir.), *Artificial Intelligence and Soft Computing, Proceedings of the IASTED International Conference*, Anaheim, IASTED/Acta Press, 1999, p. 475-479 ; J.-G. MEUNIER & D. FOREST, « Lecture et analyse conceptuelle assistée par ordinateur : premières expériences », dans J.-P. Desclés & F. Le Priol (dir.), *Annotation automatique et recherche d'information*, Paris, Hermès, 2009, p. 211-230.
- 15 A. SALLABERRY, *Visualisation d'information : de la théorie sémiotique à des exemples pratiques basés sur la représentation de graphes et d'hypergraphes*, thèse de doctorat en informatique, Bordeaux, Bordeaux 1, 2011 ; P. VICKERS, J. FAITH & N. ROSSITER, « Understanding Visualization: A Formal Approach Using Category Theory and Semiotics », *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no 6, 2013, p. 1048-1061 ; J. BONACCORSI, « Le monde de l'Open data : les jeux sémiotiques et esthétiques de la "visualisation" comme rhétorique de la transparence », dans S. Badir & F. Provenzano (dir.), *Pratiques émergentes et pensée du médium*, Louvain-la-Neuve, L'Harmattan, 2016, p. 131-150 ; T. ANDRY, « Une approche sémiotique de la conception d'outils d'analyse de données massives », *Communication et organisation*, no 54, 2018, p. 133-148.
- 16 A. VALLE & A. MAZZEI, « Sapir-Whorf vs Boas-Jakobson Enunciation and the Semiotics of Programming Languages », *Lexia*, 27-28, 2017, p. 505-525.
- 17 R. GUDWIN & J. QUEIROZ, « Towards an Introduction to Computational Semiotics », *loc. cit.*
- 18 T. ZIEMKE, « Robosemiotics and embodied enactive cognition », *SEED – Semiotics, Evolution, Energy, and Development*, vol. 3, no 3, 2003, p. 112-124 ; N. MANSARD, *Semiotics of Motion*:

- Toward a Robotics Programing Language*, HDR, Toulouse, Université Paul Sabatier - Toulouse III, 2013. En ligne : <<https://tel.archives-ouvertes.fr/tel-01061112/document>>.
- 19 S. BRIER, *Cybersemiotics: Why Information is Not Enough!*, Toronto, University of Toronto Press, 2008. Voir également, en ligne : <<https://www.cybersemiotics.com/>>.
 - 20 R. GUDWIN & J. QUEIROZ, « Towards an Introduction to Computational Semiotics », *loc. cit.*, p. 393.
 - 21 J.-G. MEUNIER, « Sémiotique computationnelle : sémiotique de la computation ou la computation en sémiotique », conférence, Montréal, Université de Montréal, 2020.
 - 22 J.-G. MEUNIER, « Vers une sémiotique computationnelle? », *Applied Semiotics / Sémiotique appliquée*, no 26, 2018, p. 77. En ligne : <<http://french.chass.utoronto.ca/as-sa/ASSA-No26/26-6.pdf>>.
 - 23 J. S. SEQUEIRA & M. I. RIBEIRO, « Semiotics and Human-Robot Interaction », *Informatics in Control Automation and Robotics*, vol. 15, 2008, p. 277-291 ; L. F. H. BENTO, R. O. PRATES & L. CHAIMOWICZ, « Using Semiotic Inspection Method to Evaluate a Human-Robot Interface », *2009 Latin American Web Congress*, Merida, Yucatan, 2009, p. 77-84 ; M. ALAC, J. MOVELLAN & F. TANAKA, « When a Robot is Social: Spatial Arrangements and Multimodal Semiotic Engagement in The Practice of Social Robotics », *Social Studies of Science*, vol. 41, no 6, 2011, p. 893-926.
 - 24 Voir le carnet du projet sur ResearchGate : <<https://www.researchgate.net/project/Studying-speech-practices-of-children-with-ASD-through-Human-Robot-Interaction>>.
 - 25 W. NÖTH, « Semiosis and the Umwelt of a Robot », *Semiotica*, no 134, 2001, p. 695-699.
 - 26 Voir note 19 *supra*.
 - 27 M. LAZZARATO, *Signs and Machines: Capitalism and the Production of Subjectivity*, Los Angeles, Semiotext, 2014.
 - 28 F. BERARDI, *Tueries : forcenés et suicidaires à l'ère du capitalisme absolu*, trad. de l'anglais par P. Dardel, Montréal, Lux, 2016.
 - 29 G. GENOSKO, *Critical Semiotics: Theory, from Information to Affect*, New York, Bloomsbury Academic, 2016.
 - 30 Cf. F. RICHERT, « La critique de la valeur à l'ère du sémiocapitalisme », *Cygne noir*, no 4, 2016. En ligne : <<http://revuecygnenoir.org/numero/article/critique-valeur-semiocapitalisme>>.
 - 31 Cf. P. B. ANDERSEN, *A Theory of Computer Semiotics: Semiotic Approaches to Construction and Assessment of Computer Systems*, Cambridge, Cambridge University Press, 1990 ; M. NADIN, « Semiotic Machine », *Public Journal of Semiotics*, vol. 1, no 1, 2007, p. 57-75 ; R. GUDWIN & J. QUEIROZ, *Semiotics and Intelligent Systems Development*, Hershey, Idea Group Pub., 2007 ; K. TANAKA-ISHII, *Semiotics of programming*, New York, Cambridge University Press, 2010 ; S. WALSH MATTHEWS & M. DANESI, « AI: A Semiotic Perspective », *Chinese Semiotic Studies*, vol. 15, no 2, 2019, p. 199-216.
 - 32 K. TANAKA-ISHII, *Semiotics of programming*, *op. cit.*
 - 33 M. BONENFANT *et al.*, « Big Data, médiation symbolique et gouvernementalité », *loc. cit.*, p. 37.
 - 34 N. K. HAYLES, « Can Computers Create Meanings? A Cyber/Bio/Semiotic Perspective », *Critical Inquiry*, vol. 46, no 1, 2019, p. 32-55.
 - 35 Cf. D. COMPAGNO & M. TRELEANI, « Introduction to Meaningful data/Données signifiantes », *Semiotica*, no 230, 2019, p. 1-17.
 - 36 Cf. K. L. KETNER, « Peirce and Turing: Comparisons and Conjectures », *Semiotica*, vol. 68, no 1-2, 1988, p. 33-62 ; J. RANSDALL, « The relevance of Peircean Semiotic to Computational Intelligence Augmentation », *SEED – Semiotics, Evolution, Energy, and Development*, vol. 3, no 3, 2003, p. 5-36 ; P. STEINER, *C.S. Peirce and Artificial Intelligence: Historical Heritage and (New)*

- Theoretical Stakes*, dans V. C. Müller (dir.), *Philosophy and Theory of Artificial Intelligence*, Berlin/Heidelberg, Springer, 2013, p. 265-276.
- 37 J. QUEIROZ & F. MERRELL, « On Peirce's Pragmatic Notion of Semiosis—A Contribution for the Design of Meaning Machines », *Minds & Machines*, vol. 19, no 1, 2009, p. 129-143.
 - 38 J.-G. MEUNIER, « Vers une sémiotique computationnelle? », *loc. cit.*, p. 98.
 - 39 Voir à ce sujet la critique formulée par Rob Kitchin : R. KITCHIN, « Big Data, new epistemologies and paradigm shifts », *Big Data & Society*, vol. 1, no 1, 2014, p. 1-12.
 - 40 L. CRÉMIER, M. BONENFANT & L. I. LAFRANCE ST-MARTIN, « Raw data or hypersymbols? Meaning-making with digital data, between discursive processes and machinic procedures », *Semiotica*, no 230, 2019, p. 189-212.
 - 41 D. PULIZZOTTO, « L'analyse de texte assistée par ordinateur : introduction à l'un des champs fondamentaux de la sémiotique computationnelle », *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnoir.org/numero/article/pulizzotto-ato>>.
 - 42 J.-G. MEUNIER, « Enjeux de la modélisation formelle en sémiotique computationnelle », *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnoir.org/numero/article/meunier-modelisation-semiotique-computationnelle>>.
 - 43 L. CRÉMIER, M. BONENFANT & L. I. LAFRANCE ST-MARTIN, « Données brutes ou hypersymboles? Signification et données numériques, entre processus discursif et procédure machinique », *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnoir.org/numero/article/cremier-et-al-hypersymboles>>.
 - 44 E. CACCAMO, « Médias socionumériques alternatifs. Étude sémiotique et rhétorique de diaspora* », *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnoir.org/numero/article/caccamo-medias-socionumeriques-alternatifs>>.
 - 45 F. RICHERT, « Du fétichisme des sémiomarchandises "parlantes" à l'asservissement machinique : le cas des assistants personnels intelligents », *Cygne noir*, no 7, 2019. En ligne : <<http://www.revuecygnoir.org/numero/article/richtert-fetichisme-semiomarchandises>>.
 - 46 G. GENOSKO, « Les aléas de la subjectivité sous le sémiocapitalisme : Félix Guattari et le potentiel politique des hybrides humains-machines », trad de l'anglais (Canada) par S. Levesque & F. Richert, *Cygne noir*, no 7, 2019. En ligne : <<http://revuecygnoir.org/numero/article/genosko-semiocapitalisme-guattari>>.





L'analyse de texte assistée par ordinateur : introduction à l'un des champs fondamentaux de la sémiotique computationnelle

Davide PULIZZOTTO

Cygne noir, no 7, 2019 : « Algorithmes »

Résumé

La sémiotique computationnelle étudie l'interaction entre les processus d'émergence du sens et les systèmes formels, computables et numériques. En effet, l'une de ses hypothèses est la possibilité de décrire la sémiose à travers des métalangages formels et de la simuler par des procédés algorithmiques. Dans ce contexte, plusieurs pratiques d'analyse sémiotique se sont développées, à l'exemple de l'analyse de texte assistée par ordinateur (ATO). Avec cette dernière, en adoptant des techniques et des méthodes issues de l'informatique et de l'intelligence artificielle, les formes plus classiques de l'analyse de texte se joignent aux champs de recherche des humanités numériques. La sémiotique est ainsi appelée, entre autres, à discuter les enjeux de l'usage de ces techniques dans la recherche en sciences humaines et sociales. L'objectif de cet article est de présenter un survol de la sémiotique computationnelle et d'introduire le lectorat à certains aspects théoriques et méthodologiques de l'assistance informatique à l'analyse de texte. Plus particulièrement, le texte expose les étapes et les hypothèses de la transformation vectorielle du texte que présuppose l'ATO et discute des enjeux sémiotiques de deux procédures : la lemmatisation et la fonction de pondération.

Pour citer cet article

PULIZZOTTO, Davide, « L'analyse de texte assistée par ordinateur : introduction à l'un des champs fondamentaux de la sémiotique computationnelle », *Cygne noir*, no 7, 2019. En ligne : <http://www.revuecygnoir.org/numero/article/pulizzotto-ato> (consulté le xx/xx/xxxx).



Cet article de *Revue Cygne noir* est mis à disposition selon les termes de la licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Pas de Modification 2.5 Canada.

L'ANALYSE DE TEXTE ASSISTÉE PAR ORDINATEUR : INTRODUCTION À L'UN DES CHAMPS FONDAMENTAUX DE LA SÉMIOTIQUE COMPUTATIONNELLE

La « révolution numérique » en train de modifier profondément les sciences humaines et sociales (SHS) est loin d'être ignorée par la sémiotique. L'interaction, de plus en plus grandissante, entre les SHS et les sciences informatiques démontre la pertinence de définir un nouveau domaine d'études, à savoir celui des *humanités numériques*. La sémiotique est partie intégrante de ce nouveau domaine de recherche et y contribue par le développement d'un nouveau champ d'études appelé *sémiotique computationnelle*. Celle-ci regroupe des travaux de nature différente, mais qui partagent une relation, plus ou moins profonde, avec l'informatique. Ainsi, les travaux afférents permettent d'étudier, sous différentes perspectives, la nature computationnelle du *sens*.

Constitué de deux parties, cet article a une visée introductive et didactique. Fournissant des clés de compréhension pour appréhender un champ relativement récent de la sémiotique, ce texte s'adresse aux chercheurs en sciences humaines, plus précisément aux sémioticiens, intéressés à explorer plus avant les liens entre sémiotique, analyse de texte et computation. Dans la première partie, je présenterai un survol de la sémiotique computationnelle, en proposant une classification en trois branches. Il s'agira de dresser un portrait disciplinaire et méthodologique des champs d'études de la sémiotique computationnelle. Par la suite, j'approfondirai la nature de l'un de ses champs d'études les plus importants : l'analyse de texte assistée par ordinateur (ATO). J'ai en effet choisi d'isoler et d'explorer ce champ en vue de répondre à la question suivante : quels sont les enjeux sémiotiques de l'usage des outils numériques et quels éléments permettent d'en juger ? L'ATO se prête bien à l'exploration des relations entre sémiotique et usage des outils numériques puisque son cadre épistémologique est compatible avec l'une des pratiques les plus communes de la sémiotique, soit l'analyse des artéfacts sémiotiques. Ainsi, dans la deuxième partie, j'accompagnerai le lectorat à réfléchir sur certains aspects ayant trait aux dynamiques sémiotiques de l'usage des outils de traitement automatique du langage naturel (TALN) et de l'algèbre linéaire. Pour ce faire, je présenterai l'une des étapes qui permet le passage de l'analyse de texte traditionnelle à l'assistance par ordinateur, soit la conversion du texte en un modèle computable, réalisée grâce à l'adoption du *modèle sémantique vectoriel*. Plus particulièrement, je décrirai deux opérations spécifiques de la transformation vectorielle du texte, la *lemmatisation*

et la *fonction de pondération*, puisqu'elles sont fonctionnelles à la compréhension de la transposition mathématique du texte dans le cadre de l'ATO.

Bien que l'article ait une portée pédagogique, les opérations décrites dans la deuxième partie de ce texte me donneront la possibilité de m'arrêter plus longuement sur des questions supplémentaires et de problématiser certains éléments de l'approche concernée qui me semblent être souvent négligés par la communauté de chercheurs. L'opération de lemmatisation, notamment, permet de réfléchir à la nature sémiotique des caractéristiques qui composent les vecteurs du modèle sémantique. Le plus souvent, les auteurs se limitent à présenter ces caractéristiques comme un ensemble de mots. Peter D. Turney et Patrick Pantel, pour citer l'un des textes les plus connus, affirment que le sens d'un texte est décrit par les éléments qui le composent et leurs fréquences¹. Chaque texte est ainsi conçu comme un *sac-de-mots*, c'est-à-dire comme l'ensemble des signes linguistiques qui apparaissent dans le texte. Cet ensemble constitue le vecteur qui représente un texte dans un espace géométrique. Or, trop peu est dit à propos de la nature linguistique, sémantique ou lexicale de ces unités. Pour cette raison, je m'attarderai sur la lemmatisation. Je mettrai en évidence quelques limites des outils qui sont généralement utilisés en ATO – limites qui, à dire vrai, affectent peu la valeur heuristique du modèle vectoriel, mais qu'il est important de garder à l'esprit.

Dans cette deuxième partie de l'article, j'ai aussi la possibilité de rebondir sur la complémentarité entre l'hypothèse distributionnelle, qui est à la base du modèle sémantique vectoriel, et la linguistique de Saussure. Teun A. van Dijk est, probablement, le premier à signaler cette complémentarité². Il considère identiques le problème que se pose Zellig Harris, un des pères putatif de l'hypothèse distributionnelle, et celui de la sémiotique textuelle des années 1960 et 1970, laquelle se préoccupe de la construction d'outils pour l'analyse des textes. De plus, selon Dijk, l'analyse du discours de Harris propose des solutions qui coïncident avec la vision structuraliste du langage. Ces liens sont décrits de manière explicite par Magnus Sahlgren, qui souligne à plusieurs reprises la nature structuraliste de la sémantique vectorielle³. En particulier, l'auteur a mis en évidence le rôle de la « vision différentielle du sens »⁴, tirant ses racines de la linguistique structuraliste, pour la définition de l'hypothèse distributionnelle, énoncée la première fois par Harris. Ainsi, il montre que les motivations théoriques de deux types de modèles de la sémantique vectorielle, le syntagmatique et le paradigmatique, se basent sur la notion de *valeur* et sur la définition de l'axe syntagmatique et de l'axe paradigmatique en linguistique. De mon point de vue, la notion saussurienne de valeur peut également être utilisée pour motiver la fonction de pondération, ce qui constitue un prolongement logique des premières réflexions de Sahlgren.

Enfin, l'article s'adresse principalement à ceux qui abordent l'ATO depuis peu. Ces derniers trouveront les notions de base pour comprendre son cadre théorique. Mais il offre également quelques pistes de réflexion pour ceux qui cherchent à clarifier la valeur sémiotique de la transformation vectorielle du texte. Cette dernière discussion est basée sur plusieurs travaux empiriques que j'ai personnellement conduits⁵.

L'article est divisé en deux parties : la première présente les trois branches de la sémiotique computationnelle et détaille davantage l'ATO ; la deuxième décrit les procédures et les hypothèses de la transformation vectorielle du texte que l'ATO met en place et discute davantage les fondements sémiotiques de la procédure de lemmatisation et de la fonction de pondération.

1. Un survol de la sémiotique computationnelle

La sémiotique est une discipline intrinsèquement interdisciplinaire et son histoire en témoigne. En effet, en Europe, elle s'est développée avec le concours d'une autre discipline, la linguistique. Pendant plusieurs décennies, les histoires de ces deux disciplines se sont entrecroisées alors que la linguistique englobait la sémiotique en tant que champ d'études. Parallèlement, en Amérique du Nord, la sémiotique se définissait à l'intérieur du pragmatisme américain, par le biais des travaux de Charles S. Peirce. Avec le temps, la sémiotique a redéfini ses approches, ses méthodologies et ses objets d'études spécifiques, mais elle n'a pas cessé de produire des « mélanges disciplinaires » de toutes sortes. La sémiotique computationnelle est le produit de l'un de ces « mélanges ». Celle-ci se caractérise par la rencontre de la sémiotique avec les domaines de l'informatique. La sémiotique computationnelle constitue aujourd'hui un champ des humanités numériques, ces dernières recouvrant « un ensemble de pratiques de recherche à l'intersection des technologies numériques et des différentes disciplines des sciences humaines⁶ ».

Les volets de la sémiotique computationnelle sont nombreux et interagissent à différents niveaux avec l'informatique, mais aussi avec les différents champs de l'intelligence artificielle. Pour définir la sémiotique computationnelle, il est nécessaire d'aborder la relation entre la sémiotique et les *propriétés computables* d'un mécanisme signifiant. L'une des perspectives les plus intéressantes pour appréhender cette relation vient de l'intelligence artificielle. Ainsi, je propose de résumer cette relation avec le concept de « simulation » d'opérations cognitives par une machine, ce qui constitue, historiquement, le point de départ du programme de recherche de l'intelligence artificielle. La naissance de cette dernière, comme celle du cognitivisme, est enracinée dans le mouvement cybernétique des années 1940-1950. C'est dans ce contexte que l'intérêt pour les

mécanismes de la pensée humaine et la possibilité de les simuler à l'aide d'un ordinateur a commencé à émerger. Elaine Rich nous fournit probablement une définition de l'intelligence artificielle des plus précises et des plus actuelles⁷. Elle souligne comment le transfert des fonctions cognitives vers l'ordinateur est effectivement le principal but de cette discipline : « L'intelligence artificielle étudie la manière de faire faire aux ordinateurs des choses pour lesquelles, à l'heure actuelle, l'être humain est plus habile⁸. »

Cette définition de l'intelligence artificielle offre une perspective idéale pour appréhender les dynamiques sémiotiques par ses aspects computables. Dans ce cadre, si l'objectif de la sémiotique est de comprendre les mécanismes qui *mettent les « choses » en condition de signifier*⁹, la sémiotique computationnelle aurait le but de « simuler » les « conditions de signification ». Le verbe « simuler » devrait être appréhendé dans une acception large, car il doit représenter la sémiotique computationnelle et ses diverses expériences de recherche, dont seule une portion utilise effectivement l'informatique et l'intelligence artificielle pour implémenter des dynamiques sémiotiques.

Compte tenu de leur grand nombre et de leur diversité, il est très difficile de résumer et de classer les études pouvant faire partie de la sémiotique computationnelle. Pour des raisons de simplicité, je propose de diviser la sémiotique computationnelle en trois grandes branches : (1) l'étude de l'*ordinateur comme machine sémiotique*¹⁰ ; (2) l'étude de l'*ordinateur comme artéfact*¹¹ ; et (3) l'étude de l'*ordinateur comme outil au service de la recherche sémiotique*¹². Cette classification, inspirée des travaux de Jean-Guy Meunier¹³ et de Kumiko Tanaka-Ishii¹⁴, se caractérise par trois approches différentes du même objet : l'*ordinateur*. Le terme « ordinateur » est ici utilisé de manière métonymique et, selon ces approches, réfère à des éléments de nature différente. Toutefois, ces approches concernent toutes, d'une manière ou d'une autre, l'aspect computable sous-entendu dans la notion générale d'ordinateur. Cette classification ne peut toutefois être considérée comme étant définitive puisque ce champ est en évolution constante. De plus, plusieurs auteurs qui sont inclus dans la classification que je propose ne reconnaissent pas explicitement la nature sémiotique de leurs recherches, ce qui rend la détermination de cette classification encore plus difficile.

1.1 L'ordinateur comme machine sémiotique

La première de ces trois branches de la sémiotique computationnelle considère l'ordinateur comme une machine sémiotique, c'est-à-dire une machine qui *manipule des signes*. Dans cet axe, la sémiose est considérée comme *un fait ou un événement computable* et l'objectif principal des travaux qui y sont menés consiste à simuler des processus de si-

gnification. Ainsi, l'une des questions fondamentales posées dans ce contexte concerne la manière de « modéliser » les processus de construction du sens dans des systèmes artificiels¹⁵. La sémiotique joue un rôle très important dans ce genre de recherches et plusieurs modèles sémiotiques ont été proposés pour la simulation des systèmes et des processus du sens. Dans les travaux de João Queiroz et Floyd Merrell¹⁶, par exemple, la sémiotique de Peirce est employée comme cadre conceptuel à partir duquel il est possible de comprendre la sémiose et de construire un modèle formel pour sa simulation par l'ordinateur. La sémiose est entendue comme un processus qui s'auto-organise comme le font les organismes biologiques¹⁷.

Toutefois, les études en sémiotique qui se sont concentrées sur le lien entre les dynamiques sémiotiques et la computation sont peu nombreuses. Ces travaux ont surtout essayé de mettre en relation les caractéristiques formelles des modèles sémiotiques avec celles des modèles computationnels¹⁸. Ainsi, dans ce genre de travaux, le raisonnement logique est constitué par l'identification de fonctions cognitives qui peuvent, en premier lieu, être transformées en un *modèle formel et logique* et, successivement, en un *modèle computable*, pour enfin passer à la véritable phase d'*implémentation informatique*. Ce raisonnement constitue une approche générale pour la sémiotique computationnelle et, de manière plus spécifique, pour la troisième branche de la sémiotique computationnelle. En effet, identifier les fonctions cognitives que, involontairement ou volontairement, certains outils numériques peuvent simuler permet d'entrevoir avec plus d'attention leur potentiel heuristique.

1.2 L'ordinateur comme artéfact

La deuxième branche considère l'ordinateur en tant qu'*artéfact*. Un artéfact doit être appréhendé comme un objet construit par l'humain, qui lui sert pour exécuter des tâches. Dans ce contexte, la sémiose est regardée comme *un événement qui se génère à partir de l'interaction entre l'humain et la machine*. Dès lors, les concepts de machine et d'artéfact se chevauchent. L'artéfact prend une grande importance en sémiotique pour la définition du concept de culture. Pour François Rastier, par exemple, les artéfacts constituent la composante principale du niveau sémiotique de la culture, et ils regroupent les *objets culturels*¹⁹. L'ordinateur peut être étudié comme un *instrument* (une des catégories des objets culturels), c'est-à-dire un objet construit par l'humain et qui interagit avec lui dans différentes situations. Ici, la sémiotique est souvent utilisée pour supporter l'étude du design et de la conception des systèmes informatiques afin d'optimiser l'interaction humain-machine. Cette deuxième branche est probablement

la plus vaste. En simplifiant, on peut la subdiviser en trois axes, soit les études sur l'interaction humain-machine (IHM), comme les travaux en ingénierie sémiotique²⁰, les études sur les systèmes d'information comme systèmes sémiotiques qui sont liés à la sémiotique organisationnelle²¹, et enfin les études sémiotiques de la programmation et de ses langages, projet de recherche qui remonte aux années 1960²².

1.3 L'ordinateur comme outil

La troisième branche appréhende l'ordinateur comme un *outil pour l'analyse sémiotique*. Dans ce contexte, la sémiose qui est prise en considération *émerge d'un artéfact sémiotique qui est indépendant de l'outil informatique*. Les textes écrits ou autres artéfacts sémiotiques sont pris dans leur contexte et convertis en un objet formel sur lequel il est possible d'appliquer des outils pour l'analyse assistée par ordinateur. Le mot « ordinateur » n'est alors qu'un mot-contenant pour indiquer plusieurs éléments de nature mathématique, statistique, formelle, computationnelle et informatique. Pour le dire simplement, les travaux de cette branche continuent à étudier des phénomènes sémiotiques et ses artéfacts, mais au moyen de méthodes hybrides d'analyse sémiotique qui se servent de l'assistance informatique. La troisième branche se distingue des deux précédentes par son objet de recherche, lequel n'est pas la détermination de modèles computationnels pour la simulation sémiotique ni les aspects computables de la signification ou l'ordinateur comme artéfact. Son objectif est de répondre à des questions traditionnelles en sémiotique ou en SHS, à l'exemple de problématiques liées à l'analyse d'un corpus de textes, et de le faire à l'aide de méthodes assistées par l'informatique.

L'analyse du texte écrit domine les travaux de cette troisième branche, et la majorité des travaux qui y sont développés le sont en ATO. Le traitement automatique du langage naturel est en effet l'une des branches de l'intelligence artificielle qui s'est grandement développée dans les dernières décennies. À l'inverse, en ce domaine, les outils d'analyse de l'image ou du son sont encore en phase de développement. Il est probable que la recherche en apprentissage profond (*deep learning*) pourra faire émerger des méthodes pour l'analyse d'artéfact sémiotique dans lesquels plusieurs systèmes sémiotiques interagissent entre eux, comme le langage verbal écrit, visuel, audio, etc. Pour le moment, il existe très peu de travaux qui se sont penchés sur l'analyse d'artéfacts sémiotiques complexes. Pour toutes ces raisons, je présenterai dans les prochains paragraphes le cadre théorique de l'ATO, lequel se prête bien à une première exploration de l'usage des outils numériques pour l'analyse d'artéfact sémiotique.

1.3.1 L'analyse de texte assistée par ordinateur (ATO)

La troisième branche de la sémiotique computationnelle regroupe plusieurs travaux, parmi lesquels ceux déjà accomplis en ATO, dont l'objet de recherche est à la fois le développement de méthodes d'analyse de textes à l'aide d'outils statistiques, mathématiques ou informatiques ainsi que l'analyse de corpus textuels. En résumé, l'ATO s'applique à des corpus qui sont construits à des fins d'analyse et afin de répondre à des problématiques typiques en SHS. Par exemple, il est possible d'assister l'analyse d'un corpus composé d'articles de journaux, d'œuvres d'un philosophe, de romans d'un courant littéraire particulier, etc. Ce domaine est vaste et regroupe des travaux qui naissent dans différents contextes, comme l'analyse de contenu²³, l'analyse du discours²⁴, la statistique textuelle et la linguistique de corpus²⁵, pour ne nommer que cela. Il n'est donc pas simple de définir ce champ de recherche de manière univoque. L'expression « analyse de texte assistée par ordinateur » ne fait pas l'unanimité d'autant plus qu'elle n'est pas la seule utilisée²⁶. Il demeure toutefois possible d'esquisser un cadre général des disciplines et des méthodes qui contribuent aux avancées de ce champ de recherche. Sauf erreur de ma part, il n'existe aucune tentative de décrire l'interdisciplinarité de l'ATO. J'en propose ainsi une description schématique (fig. 1) afin de clarifier le plus possible la nature disciplinaire et méthodologique de l'ATO. Pour ce faire, je présente un survol qui va des SHS jusqu'à la science de données. La description de l'ATO que je propose est constituée de six axes de contribution.

D'abord, chaque recherche en ATO se situe à l'intérieur d'un domaine d'application dans différentes disciplines des SHS, comme la sociologie, l'histoire, la communication, les sciences politiques, la philosophie, etc. (premier axe). Par exemple, en histoire de la philosophie les chercheurs étudient, entre autres, l'évolution d'un concept en analysant un recueil de texte d'un philosophe particulier. En science politique, les chercheurs analysent les caractéristiques du discours électoral d'un candidat aux élections à partir des verbatim de ses discours télévisés. En sociologie, l'intérêt de recherche se concentre davantage dans la description de communautés socio-sémantiques à partir de corpus de textes journalistiques, et ainsi de suite. Ce sont donc les intérêts de recherche de ces disciplines qui déterminent le point de départ et les objectifs de recherche de l'ATO. Ainsi, chacune de ces disciplines caractérise de manière spécifique l'approche d'analyse utilisée. Il est donc important de ne pas négliger la contribution qu'elles apportent au développement des applications en ATO.

Ensuite, formant un deuxième axe, des disciplines comme la linguistique et la sémiotique contribuent en fournissant un cadre théorique pour l'analyse du texte. Des concepts relatifs aux paliers textuels (sème, mot, syntagme, phrase, paragraphes,

etc.), les axes syntagmatique et paradigmatique ou la dénotation et la connotation favorisent la compréhension de l'artéfact sémiotique qui est utilisé pour l'extraction des connaissances.

Un troisième axe de contribution à l'ATO est constitué par l'analyse de contenu et l'analyse du discours. Ces pratiques ont certainement fourni à l'ATO des méthodes et des cadres théoriques à la fois qualitatifs et quantitatifs²⁷ qui apportent des perspectives différentes à l'analyse du texte. Un de ses mérites est, par exemple, d'avoir mis en évidence l'importance d'une approche herméneutique qui se concentre davantage sur l'extraction de connaissances extratextuelles. En analyse du discours, les chercheurs ajoutent le plus souvent à l'analyse du texte une description des relations de son contenu avec son contexte historique, social, politique, etc.

Un quatrième axe regroupe les disciplines qui mettent en lumière plusieurs méthodes quantitatives d'analyse du texte, soit la textométrie, la statistique textuelle et la linguistique de corpus. Depuis le début des années 1960, ces disciplines contribuent à la définition de l'approche statistique à l'étude du langage. Chacune avec ses propres spécificités, elles ont su montrer comment utiliser différents outils de la statistique dans un contexte d'analyse de texte, comme la notion de corrélation²⁸ ou l'analyse factorielle de correspondances²⁹.

Un cinquième axe de contribution est celui du traitement automatique du langage naturel (TALN)³⁰ et de la linguistique computationnelle³¹. À la différence de l'axe précédent, celui-ci met en évidence un autre aspect de l'ATO, soit la description et l'annotation automatique des caractéristiques linguistiques d'un texte. En effet, c'est à l'intérieur de cette discipline que des outils comme les lexicographes automatiques ou les analyseurs morphosyntaxiques³² se développent et se mettent constamment au point. L'analyse linguistique du texte est, sans doute, une étape fondamentale de l'ATO, et le TALN contribue à améliorer son assistance par ordinateur, mettant ainsi en évidence l'importance de l'informatique et de l'intelligence artificielle. Plusieurs autres applications et outils sont développés dans ce domaine (à l'exemple du résumé automatique³³), et peuvent, à l'occasion, être intégrés dans une chaîne de traitement en ATO.

Enfin, la science des données met à disposition de l'ATO toute une série de méthodes développées en apprentissage automatique³⁴ et en apprentissage statistique³⁵, comme le regroupement automatique de documents³⁶, les techniques d'analyse thématique³⁷, etc. Ces outils peuvent être présentés sous des formes et des perspectives différentes, comme celles de la sémantique distributionnelle³⁸, de la recherche d'information³⁹ et, plus particulièrement, de la fouille de texte⁴⁰.

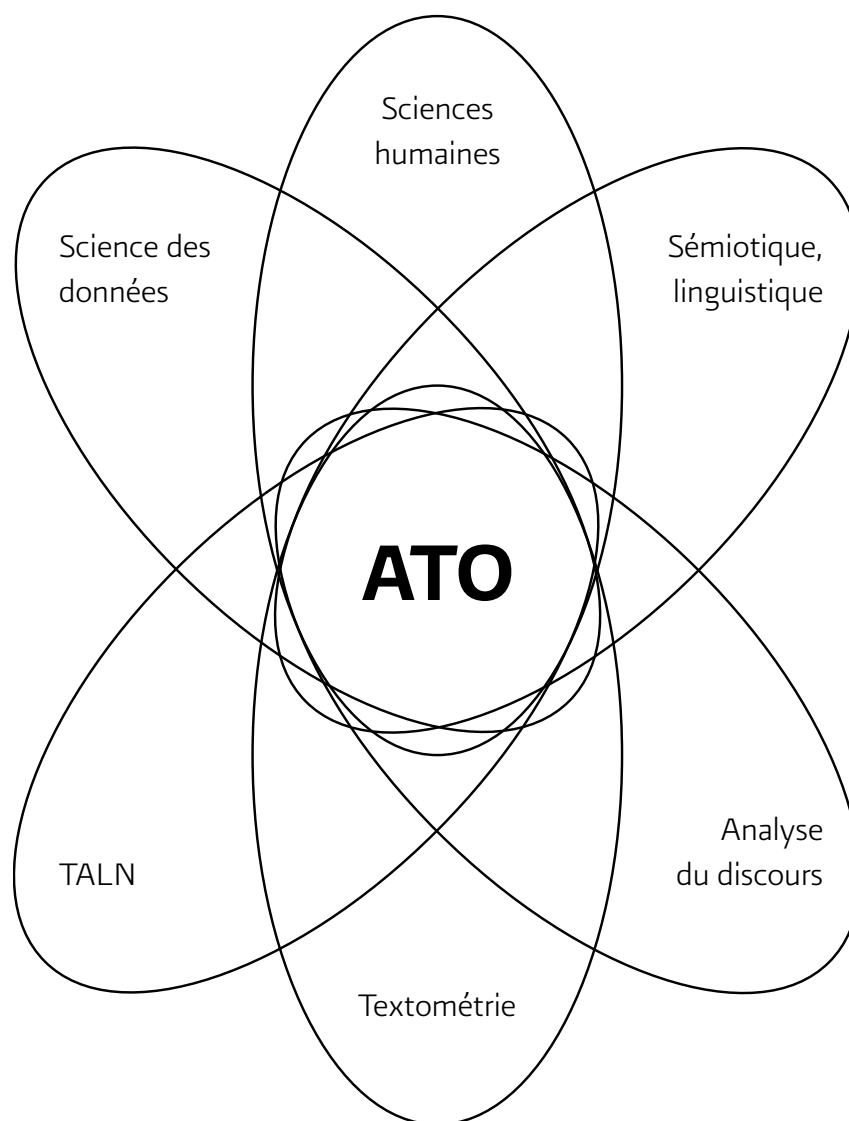


Figure 1. Schéma définissant l'interdisciplinarité de l'analyse de texte assistée par ordinateur.

L'ATO est alors une pratique d'analyse qui, à travers un bricolage de cadres théoriques et de méthodes qualitatives et quantitatives, répond à des questions de recherche typiques en SHS. Une étape particulière de ce bricolage a le mérite d'élargir les possibilités de l'assistance par ordinateur : il s'agit de l'étape de transformation du texte en un modèle formel et computable. C'est seulement lorsque cette transformation est accomplie que la plupart des outils de la statistique, du TALN ou de la science de données peuvent

être impliqués. Cette étape est aussi celle dans laquelle la sémiotique et l'informatique entrent plus facilement en contact, puisque c'est à ce moment que l'artéfact textuel subit des modifications pour s'adapter à un paradigme différent, celui numérique. En d'autres termes, les éléments linguistiques normalement utilisés pour la compréhension d'un texte, comme les mots, les lemmes, les phrases, etc., sont transférés dans un objet mathématique particulier appelé *matrice*. Il est alors important de comprendre, lors de ce passage du *texte* à la *matrice*, ce qui est véritablement conservé et ce qui est perdu.

Avant de passer à la description de la transformation vectorielle du texte en ATO, je tiens à faire une précision. En fait, il me semble que l'enthousiasme qui accompagne l'apparition des outils de l'apprentissage profond (*deep learning*) porte parfois la communauté qui s'intéresse à l'ATO à confondre différentes approches et applications. L'apprentissage profond est un sous-champ de recherche de l'apprentissage automatique. Les outils que s'y développent sont basés sur un modèle mathématique existant depuis les années 1960⁴¹, à savoir les réseaux de neurones. Sans rentrer dans les détails techniques, je veux ici souligner la différence entre les applications typiques de l'apprentissage profond et l'ATO. En fait, en raison des caractéristiques des réseaux de neurones, l'apprentissage profond est surtout développé pour des applications d'apprentissage de type supervisé. Or, en ATO, il est rare d'utiliser cette approche, car l'extraction de connaissance se fait à partir de matériel non structuré, sans annotation, de taille relativement petite et dans un cadre de type non supervisé. Les outils de l'apprentissage profond ne sont donc pas toujours adaptés à l'ATO. Toutefois, il est vrai qu'il est possible d'accéder à des modèles de langage entraînés par de tels outils pour accomplir des tâches supervisées qui sont fonctionnelles à l'analyse de texte. Par exemple, il est possible d'utiliser des modèles entraînés pour exécuter l'annotation sémantique, l'annotation morphosyntaxique ou pour obtenir des représentations vectorielles de mots ou de documents qui sont généralement appelées *word embedding*. Il est vrai également que des modèles de type non supervisés sont en train d'être développés. Mais, en l'état actuel, il me semble que le cœur des outils de l'ATO ne nécessite pas d'intégrer impérativement l'apprentissage profond.

2. Textes et vecteurs

Le plus grand défi pour l'ATO est la *conversion d'un objet sémiotique comme le texte en un objet de nature computable*. Cette étape est fondamentale, car les différents outils informatiques peuvent être seulement utilisés si l'objet d'étude assume une forme mathématique. La modélisation mathématique de phénomènes humains et sociaux n'est

pas un fait nouveau. Il s'agit d'une pratique nécessaire lorsque la statistique devient un outil d'analyse. Dans ce contexte, le phénomène humain ou social à étudier est transposé en un jeu de données qui est ensuite analysé. Par exemple, en fonction d'hypothèses particulières, il est possible d'étudier l'impact des collaborations entre chercheurs de différentes universités sur la productivité en analysant le nombre de collaborations et le nombre de publications. Dans ce cas, un phénomène social est converti en modèle statistique à travers des faits observables, comme le nombre de publications ou de collaborations. Ce type de pratique est requis également en ATO où, toutefois, les données sont composées à partir des textes. Dans le cadre de cet article, je décrirai le modèle utilisé en ATO pour la transformation du texte : la sémantique vectorielle.

2.1 Sémantique vectorielle

La sémantique vectorielle est un modèle pour la représentation formelle et computable du langage naturel. Son but est d'obtenir une *représentation vectorielle de la sémantique des textes*. Ce modèle partage les mêmes caractéristiques et propriétés que le modèle vectoriel, ce qui permet d'utiliser les outils mathématiques de l'algèbre linéaire. Le modèle sémantique vectoriel a été développé, en Amérique du Nord, par l'équipe de Gerard Salton dans les années 1970⁴², et, en Europe, par Jean-Paul Benzécri⁴³.

La représentation vectorielle du texte consiste dans la transformation de celui-ci en un point dans un espace géométrique, dont les coordonnées dépendent de ses caractéristiques lexicales et sémantiques. En d'autres termes, un document textuel est projeté dans un espace euclidien au moyen de sa conversion en *vecteur*. Ce dernier est constitué de valeurs numériques qui mesurent le poids de chaque caractéristique lexicale ou sémantique du document. Un ensemble de documents est ainsi transformé en une matrice sur laquelle, grâce aux outils classiques de l'algèbre linéaire, plusieurs opérations peuvent être effectuées. Une des premières opérations, qui a permis la naissance du champ d'études de la recherche d'information, est le calcul de la *similarité* ou de la *dissemblance* entre les vecteurs qui représentent les documents dans cet espace. Cette opération se concrétise notamment dans le calcul de la distance entre deux points de l'espace euclidien ou, plus fréquemment, dans le calcul de l'angle entre deux vecteurs. Elle est également l'opération la plus simple à exécuter et à interpréter. Tout simplement, plus les vecteurs sont proches, plus ils sont similaires. Par exemple, dans la figure 2, les points représentent des documents et la distance entre eux constitue leur similarité (ou dissemblance) sémantico-lexicale. Dans ce cas, trois groupes de documents similaires sont représentés.

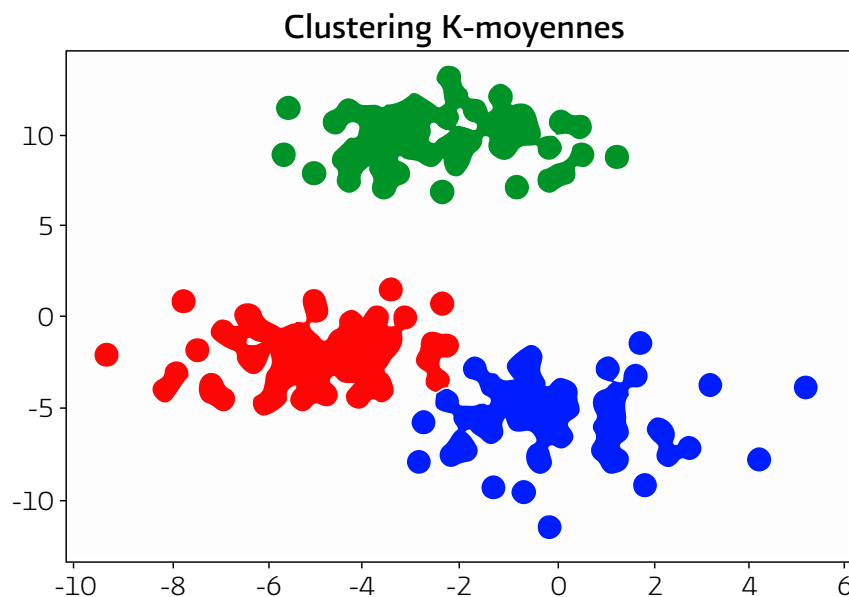


Figure 2. Représentation des similarités entre les documents d'un corpus.

Chaque point représente un document. La distance entre points représente la similarité lexico-sémantique. Les documents ont été organisés en trois ensembles distincts, identifiés par un algorithme de *clustering* (le *K-moyennes*). Chaque ensemble constitue un groupe de documents homogènes du point de vue sémantique. Il est possible d'interpréter ces groupes en termes de thèmes.

Du point de vue strictement mathématique, le modèle vectoriel correspond alors à la construction d'une matrice U . Cette modélisation est représentée par la formule 1 (fig. 3). Chaque ligne de cette matrice modélise un document sous la forme d'un vecteur $\vec{s}_i = (V_{i1}, V_{i2}, \dots, V_{ij})$ dans lequel V_{ij} correspond à la *valeur de pondération* du $j^{\text{ème}}$ mot dans le $i^{\text{ème}}$ segment.

$U =$

	mot ₁	mot ₂	...	mot _j
document ₁	V_{11}	V_{12}	...	V_{1j}
document ₂	V_{21}	V_{22}	...	V_{2j}
...	...	...	...	...
document _i	V_{i1}	V_{i2}	...	V_{ij}

Figure 3. Formule 1 : matrice Documents-Mots.

Dans ce cadre, chaque document est décrit par son lexique. Par exemple, la transformation vectorielle des phrases « Marc mange une pomme et une fraise par jour » (phrase 1) et « Marie et Julie mangent un ananas et une pomme par semaine » (phrase 2) s'accomplit avec la construction de la matrice suivante :

Phrase/ Lemmes	ananas	et	fraise	jour	Julie	manger	Marc	Marie	par	pomme	semaine	un	une
Phrase 1	0	1	1	1	0	1	1	0	1	1	0	0	2
Phrase 2	1	2	0	0	1	1	0	1	1	1	1	1	1
Phrase n	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 4. Exemple d'une matrice Documents-Mots.

Dans ces cas, la matrice représente les deux phrases par la fréquence d'apparition de chaque lemme. Pour la phrase 1, l'article « une » a une fréquence égale à 2, tandis que le nom « semaine » la fréquence est de 0.

Ce type de matrice constitue le modèle sémantique de type *sac-de-mots* (de l'anglais, *bag-of-words*). Ce modèle se base sur le postulat qui affirme que la sémantique d'un document peut être représentée par l'ensemble de ses mots et sans tenir compte de leurs relations syntaxiques. Dans ce contexte, alors, les textes qui partagent des entités lexicales similaires tendent à être similaires, ce qui implique que les vecteurs qui les représentent tendent à se rapprocher dans l'espace géométrique construit par le modèle sémantique vectoriel. Ainsi, chaque phrase qui est ajoutée à la matrice avec nos exemples (fig. 4) peut être comparée aux précédentes en calculant la distance sémantico-lexicale. Par exemple, le calcul de l'angle entre les vecteurs qui représentent les deux phrases est égal à 0,945, ce qui correspond à une similarité sémantique très grande, considérant que la valeur maximale est 1.

2.2 L'hypothèse distributionnelle

À ce point, le lectorat pourrait s'interroger sur la légitimation sémiotique du modèle vectoriel. Ou, en d'autres termes, quelles sont les raisons sémiotiques qui expliquent, au moins partiellement, pourquoi ce modèle fonctionne? La réponse classique est fournie

par l'*hypothèse distributionnelle*. Formulée par le linguiste John Rupert Firth, l'hypothèse distributionnelle est devenue un classique des travaux en fouille de texte⁴⁴. Cette hypothèse a été utilisée la première fois pour la construction de procédures de manipulation automatique du langage naturel par Zellig Harris, qui a assurément popularisé l'approche distributionnelle⁴⁵. Cette approche se base sur l'idée que *le sens des mots est distribué dans le contexte* dans lequel il apparaît. En d'autres termes, les contextes linguistiques ayant des éléments lexicaux distribués de manière similaire ont des significations similaires. Le paradigme distributionnel se base donc sur la notion selon laquelle la similarité entre entités sémantiques dépend de leurs propriétés distributionnelles. Cela implique que la similarité du contenu entre deux textes est corrélée à la similarité de leurs distributions lexicales. Enfin, deux mots tendent à être similaires s'ils côtoient souvent les mêmes mots et, par conséquent, les entités sémantiques qui partagent des contextes similaires tendent à être similaires.

Cette hypothèse fait entièrement partie du modèle sémantico-vectorel qui a été présenté dans la section précédente. En effet, chaque vecteur décrit un document sur la base de son lexique, lequel montre des relations avec les autres entités lexicales, c'est-à-dire des *relations de cooccurrence*, ce qui représente, selon l'hypothèse distributionnelle, un indice fiable et suffisant de la sémantique du mot et du texte. Or, l'hypothèse distributionnelle demeure la seule justification linguistique et sémiotique de toute recherche en ATO, mais aussi de toute application qui utilise le cadre théorique de la sémantique vectorielle. Ceci vaut pour la majorité des travaux en fouille de texte, qu'ils adoptent les approches de type supervisé, incluant l'apprentissage profond (par exemple, les représentations de *word embedding*), ou non supervisé. Sa solidité et sa pertinence ne sont pas remises en question dans cet article. Cependant, je veux aborder deux opérations spécifiques de la construction de la matrice, parce qu'elles sont fonctionnelles à la compréhension de la transposition mathématique du texte. La première renvoie à l'opération de *lemmatisation*, la seconde est la *fonction de pondération*. Ces opérations sont fondamentales pour bien construire la matrice et elles peuvent affecter négativement ou positivement le potentiel de l'hypothèse distributionnelle. Les prochaines sections permettent d'explorer si et comment *les caractéristiques textuelles transposées dans la matrice constituent de véritables simulacres des processus de signification qui sont en place dans le texte d'origine*.

2.3 La lemmatisation

La lemmatisation se rapporte à l'opération qui identifie le morphème porteur du signifié pour les unités lexicales des langues flexionnelles⁴⁶. Dans un cadre d'analyse d'ATO, cette opération est accomplie pour *normaliser les formes des mots*, c'est-à-dire pour regrouper les occurrences de chaque mot (*tokens*) en une seule forme normalisée (*type*). Ceci est justifié par le fait que, du point de vue sémantique, ce ne sont pas tous les caractères qui composent l'occurrence d'un mot qui sont pertinents. Par exemple, le mot « mangeait » est composé de deux morphèmes : (1) le radical, qui renvoie à l'unité lexicale abstraite d'un mot et qui est le signifiant qui véhicule le signifié du mot⁴⁷ ; et (2) l'affixe, qui est utilisé pour conférer un rôle syntaxique au mot. Par exemple, dans le mot « mangeait », le radical « mange » constitue le signifiant qui est en relation directe avec le signifié du verbe « manger », tandis que l'affixe « ait » spécifie le temps, le nombre et la personne de l'action. Le but principal de ce processus est de mettre en valeur les aspects sémantiques du texte, en simplifiant ainsi les formes et les flexions que chaque mot peut prendre. Ainsi, la matrice *U* sera constituée de lemmes (*type*) plutôt que d'occurrences des mots (*tokens*). Par exemple, le lemme « manger » substituera chaque occurrence du mot « mangeait ». Enfin, cette procédure *confère au modèle vectoriel une nature lexico-sémantique*.

Cette procédure de « normalisation » peut être accomplie avec d'autres méthodes, comme la *racinisation* et le *découpage en n-gram de caractères*⁴⁸. Cependant, surtout dans les applications d'ATO, il est préférable d'utiliser la procédure de lemmatisation puisqu'elle apporte un plus haut degré d'interprétabilité. En fait, les lemmes constituent des mots compréhensibles et font partie du vocabulaire de la langue cible, tandis que la racine ou le n-gram de caractères ne sont pas facilement reductibles à un mot spécifique. Les deux autres procédures peuvent toutefois être choisies dans d'autres contextes, comme dans des applications de recherche d'information⁴⁹. Cette procédure de sélection des lemmes, qui constitueront les caractéristiques de la représentation vectorielle du texte, peut également porter à la construction d'une matrice composée de n-gramme de lemmes, c'est-à-dire de segments de mots qui se répètent souvent dans le corpus à l'étude et qui peuvent constituer une unité de signification pertinente. Par exemple, de segments répétés⁵⁰ comme « agression sexuelle », « gaz à effet de serre », « accommodement raisonnable », etc.

2.3.1 Réflexions sur le processus d'identification des caractéristiques lexico-sémantiques d'un texte

Dans un cadre d'ATO, l'opération de lemmatisation est accomplie pour normaliser les formes des mots, c'est-à-dire pour regrouper les occurrences des mots (*tokens*) en types (forme lemmatisée du mot). La dichotomie *token* vs *type* est fondamentale en ATO. Umberto Eco a associé cette dichotomie au modèle du signe de Louis Hjelmslev⁵¹ :

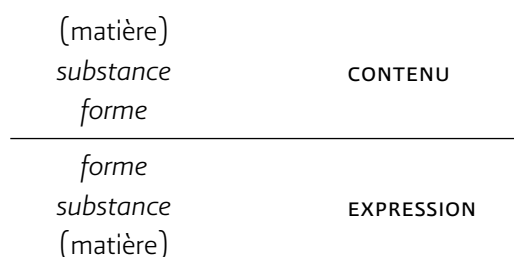


Figure 5. Expression et contenu selon Hjelmslev, reproduit d'après Eco⁵².

Ce modèle (fig. 5) implique une structure du signe à deux plans, celui du plan du contenu (signifié) et celui de l'expression (signifiant), lesquels se présupposent réciproquement et sont entièrement réversibles. Le modèle de Hjelmslev complète la théorie du signe proposée par Ferdinand de Saussure en ajoutant à chaque plan trois niveaux de transformation. Suivant Umberto Eco :

Selon ce modèle, on définit comme matière de l'expression tout continu amorphe auquel un système sémiotique déterminé donne forme en en découpant des éléments pertinents et structurés et en les produisant ensuite comme substance ; et l'on définit comme matière du contenu l'univers en tant que champ de l'expérience auquel une culture déterminée donne forme en en découpant des éléments pertinents et structurés et en les communiquant ensuite comme substance. La différence entre un élément de la forme et un élément de la substance est celle qui intervient entre un type et une occurrence concrète (*token*)⁵³.

En d'autres termes, la matière de l'expression et la matière du contenu sont un « tout continu amorphe » alors que la substance est la matière structurée par la forme. Ainsi, sur le plan du contenu, la matière amorphe est découpée en unités qui constituent la substance du contenu, ce qui est permis par la forme. Par exemple, le signifié général de « mourir » peut avoir différentes formes : « décéder » ou « partir », etc. Chacune de ces formes est porteuse d'un signifié (plan du contenu) qui se manifeste dans une forme particulière sur le plan de l'expression. Le rapport entre le plan du contenu et le plan

de l'expression est appelé fonction sémiotique. Chacune de ces formes peut avoir une nuance de signifié particulière, qui dépend du contexte d'utilisation de la forme. Ainsi, dans certains contextes, on dira « décédée » pour se référer à une personne qui est morte. Dans d'autres contextes, on utilisera le verbe « partir » pour se référer à la même personne décédée. Ceci implique que la substance du contenu demeure la même, soit la notion générale de « mourir ».

Quel est donc le lien entre forme et contenu, d'un côté, et le processus de lemmatisation et de conversion d'un *token* en *type*, de l'autre? Suivant les réflexions d'Umberto Eco, la dichotomie entre *type* et *token* peut être appréhendée à travers le modèle de Hjelmslev afin de considérer « la différence entre un élément de la forme et un élément de la substance [comme celle] qui intervient entre un type et une occurrence concrète (*token*)⁵⁴ ». Dans l'exemple précédent, le *type* « mourir » peut avoir différents *tokens*, comme « partir », « décéder », etc. Mais ces formes du contenu sont découpées à partir de la même substance du contenu, soit la notion générale de « mourir ».

Or, idéalement, la lemmatisation devrait adhérer à une *normalisation sémantique* du *token*, ce qui regrouperait les différentes formes de « mourir » sous le même *type*. Toutefois, la finesse d'une analyse de ce genre n'est pas encore praticable dans un cadre d'ATO avec les outils d'analyse sémantico-syntaxique disponibles aujourd'hui. Les procédures actuelles de lemmatisation permettent de normaliser les différentes occurrences et formes qu'un mot peut prendre, et donc de trouver les *types*, mais seulement dans un cadre linguistique du genre « partir » (*type*) *versus* « partais », « part » ou « partirais » (*token*). Elle ne permet pas de normaliser les formes du contenu et ses occurrences, comme pour le mot « mourir » (*type*) par rapport à « décéder », « partir », etc. Une piste d'amélioration de cette tâche consiste dans l'utilisation des outils d'annotation sémantique automatique, qui sont développés surtout dans le domaine de la recherche d'information et dans le web sémantique. Ces techniques sont utilisées pour détecter des entités sémantiques dans le texte à partir d'une base de connaissance préalablement construite, comme, par exemple, DBpedia⁵⁵. Bien que ces outils semblent être prometteurs, leurs performances ne sont pas encore assez bonnes pour leur faire confiance dans un cadre d'ATO. Selon la plus récente étude comparative, la précision de ces annotateurs ne dépasse pas les 40%⁵⁶.

Ceci constitue sûrement une limite des outils en TALN et, par conséquent, de l'ATO. La finesse de la relation sémantique entre *type* et *token* est très difficile à détecter par le biais d'un « lexicographe automatique », car il nécessite des algorithmes de lemmatisation qui, en analysant le contexte d'occurrence d'un mot (*token*) (ex. « partais »), peut trouver son *type lexical* (« partir »), mais aussi son *type sémantique* (« mourir »). Dans ce cadre, l'algorithme de lemmatisation devrait transformer la phrase « il est parti

serein » en la liste suivante de lemmes, « il », « être », « mourir », « serein », ce qu'il est actuellement très difficile d'exécuter avec les outils informatiques.

Malgré ces limites, l'opération de réduction du *token* à son lemme demeure nécessaire pour mettre en valeur les éléments sémantiques des textes. Ceci comporte, toutefois, la perte des informations morphosyntaxiques des mots. Dans un contexte d'ATO, il est préférable de retrancher ces informations, car lorsqu'elles sont conversées, elles réduisent la valeur sémantique du modèle vectoriel. Même si on perd ces informations, la lemmatisation demeure nécessaire dans un contexte computationnel, car chaque réduction lexicale permet d'augmenter la *significativité statistique* du *token*, ce qui est important dans un contexte d'analyse de données et de corpus de grande taille. Ainsi, par exemple, il n'est pas essentiel de garder toutes les différentes variantes du verbe « manger » (il mange, ils mangent, etc.), mais il suffit de conserver son lemme. Il demeure toutefois important de souligner les limites de l'opération de lemmatisation et d'en être conscient, afin d'interpréter correctement les résultats d'une chaîne de traitement en ATO basée sur la sémantique vectorielle et les outils classiques de lemmatisation.

2.4 La fonction de pondération

Le deuxième élément sur lequel je veux porter l'attention du lectorat est la fonction de pondération. Cette dernière est la fonction qui permet de calculer le *poids* que chaque mot possède pour chaque document. Retournons à la matrice U . Il s'agit de la représentation mathématique d'un certain nombre de documents. Les chiffres qu'elle contient constituent des valeurs qui résument l'information lexico-sémantique du document sous *forme numérique*. Cette représentation mathématique des textes est fondamentale pour le traitement informatique. Il est alors important de comprendre comment ces valeurs numériques sont calculées et, ensuite, d'analyser ce que ces valeurs « disent » du point de vue sémiotique.

Pour résumer autrement le modèle sémantique vectoriel, on peut dire que la représentation vectorielle d'un groupe de textes est constituée par la matrice U , qui est composée des vecteurs d_i et chacun d'entre eux représente un document faisant partie du corpus D . Ces documents sont décrits par un certain nombre de caractéristiques lexico-sémantiques v_j (les lemmes), qui font partie du vocabulaire V . Le poids de chaque caractéristique est pondéré par une fonction mathématique qui constitue le cœur du modèle sémantique vectoriel. C'est cette fonction qui concrétise la conversion d'un texte dans un modèle computable.

Différentes méthodes de pondération existent et chacune d'elles évalue différemment les poids des caractéristiques dans un document. Les principales fonctions sont : la fonction binaire, la fonction de la fréquence et la fonction TF-IDF (de l'anglais, *Term frequency-Inverse document frequency*). La fonction binaire, comme son nom l'indique, assigne seulement des valeurs binaires (0/1), signalant ainsi la présence ou l'absence d'une caractéristique. La fonction de la fréquence, quant à elle, assigne à chaque caractéristique sa fréquence pour chaque document. Par exemple, si un lemme est répété plus d'une fois, la valeur ne sera pas un, mais un chiffre correspondant au nombre d'apparitions du lemme dans le document.

La dernière fonction, le TF-IDF, est la plus utilisée et il en existe différentes versions. Sa puissance repose sur l'hypothèse que le poids d'une caractéristique (un lemme) est *proportionnel à sa fréquence dans chaque document et inversement proportionnel à sa fréquence documentaire*. C'est-à-dire que la valeur TF-IDF augmente si le mot est très fréquent dans un petit nombre de documents, puisque, si le mot est très répandu dans le corpus, la valeur tend à diminuer. Pour exécuter cette fonction, on calcule d'abord la fréquence documentaire, c'est-à-dire le nombre de documents dans lesquels les caractéristiques apparaissent. Par exemple, si le lemme « étudiant » apparaît dans 232 documents, alors 232 sera la fréquence documentaire du lemme « étudiant ». La fréquence documentaire est ensuite convertie en fréquence documentaire inverse comme suit :

$$Idf = \log \frac{n}{f(d_j)} + 1$$

Figure 6. Formule 2 : fonction *Idf*.

Dans cette formule (fig. 6), n est égal au nombre de documents de D et $f(d_j)$ est la fréquence documentaire du lemme j . Cette fonction est complétée par la normalisation logarithmique $\log(idf) + 1$. La composante inverse de la fréquence documentaire est donc représentée par la fraction entre le nombre total n de documents du corpus D et la fréquence documentaire $f(d_j)$ du lemme. La fréquence documentaire inverse est ensuite multipliée par la fréquence du lemme, ce qui donne la formule suivante (fig. 7) :

$$TfIdf = f(j) \cdot \log \frac{n}{f(d_j)} + 1$$

Figure 7. Formule 3 : fonction *Tf-Idf*.

Ainsi construite, la valeur de la fonction TF-IDF est plus élevée lors de l'augmentation de la fréquence du terme et elle est moins élevée lors de l'augmentation du nombre de documents qui contiennent le lemme. Donc, si un lemme est très fréquent et très répandu dans le corpus, il obtient une valeur TF-IDF plus basse, tandis qu'un lemme peu fréquent en général, mais très fréquent dans un sous-groupe de documents, obtient une valeur TF-IDF plus élevée. L'objectif de cette fonction est d'identifier *les caractéristiques les plus discriminantes dans un groupe de documents* ; en d'autres termes, les caractéristiques qui permettent de distinguer les documents en sous-groupes différents en raison de leur similarité lexico-sémantique.

2.4.1 La valeur saussurienne de la pondération

La pondération est la fonction qui assigne une valeur numérique à chaque caractéristique (par exemple, le lemme) de chaque document. Elle constitue la fonction mathématique la plus importante de l'étape de prétraitement du texte dans un contexte sémantique vectoriel. Or, il s'avère que les hypothèses et les postulats linguistiques en cause lors de la transformation d'un corpus en une matrice selon une fonction de pondération sont complémentaires à la conception structuraliste du sens⁵⁷. Tel que je l'ai déjà écrit en introduction, Sahlgren a déjà montré les liens entre Saussure et Harris, et il a souligné l'importance de la notion de valeur linguistique pour la compréhension des modèles sémantiques vectoriels de type syntagmatique et de type paradigmatique. Dans ce contexte, je veux souligner la complémentarité qui, de mon point de vue, existe entre la notion de *poids* présupposée dans la fonction de pondération TF-IDF et la notion de valeur linguistique.

Dans le *Cours de linguistique générale*, Saussure définit le concept de valeur en le distinguant de celui de *signification*. Il donne l'exemple suivant : en anglais le mot « *sheep* » possède la même signification que le mot français « mouton »⁵⁸. Dans la langue anglaise, le mot « *sheep* » ne peut pas être utilisé dans des contextes similaires à ceux dans lesquels le mot « mouton » est employé en français. En effet, pour indiquer une pièce de viande de mouton servie pendant un repas, on utilise en anglais le mot « *mutton* », alors qu'en français le même terme est gardé. Dans ce cas, « *sheep* » n'a donc pas la même valeur que « mouton », car il est remplacé par le mot « *mutton* » dans certains contextes. La valeur est alors une notion qui intègre plus directement le système-langue et les processus qui régissent l'agencement syntagmatique des mots. Comme on l'a vu, le mot « *mutton* » diffère du mot « *sheep* » et cette différence existe dans le système-langue, ce qui confère une valeur différente aux deux mots. Autrement

dit, la valeur indique *la position spécifique de chaque mot dans le système discret de la langue*. Chaque mot organise et structure la pensée d'une communauté linguistique qui, sinon, « n'est qu'une masse amorphe et indistincte⁵⁹ ».

Philosophes et linguistes se sont toujours accordés à reconnaître que, sans le secours des signes, nous serions incapables de distinguer deux idées d'une façon claire et constante. Prise en elle-même, la pensée est comme une nébuleuse où rien n'est nécessairement délimité. Il n'y a pas d'idées préétablies, et rien n'est distinct avant l'apparition de la langue⁶⁰.

Pour Saussure, la langue est un système discret de signes qui donnent forme à la pensée amorphe. Dans ce cadre, le signe linguistique se met « en condition de signifier » seulement à l'intérieur d'un *système discret et non analogique*, composé d'éléments qui *s'opposent entre eux*. Saussure propose donc ce type de schéma (fig. 8) :

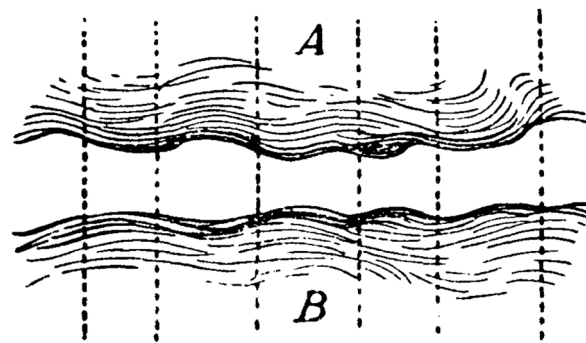


Figure 8. L'exemple de la feuille de papier⁶¹.

La lettre A illustre les « idées confuses » et la lettre B représente le flou continu de tous les sons possibles, alors que les lignes verticales constituent les mots qui donnent forme à la pensée. L'image qui permet de mieux comprendre ce schéma est une feuille de papier. La langue est comparée à une feuille où la pensée est le recto et le signe perceptible est le verso. Lorsqu'on coupe le papier, il est impossible de couper le verso sans couper le recto. De la même manière, la langue coupe la pensée amorphe et construit un système discret.

Par sa proposition selon laquelle « dans la langue, il n'y a que des différences⁶² », Saussure accorde de l'importance à *la position que le signe prend dans le système auquel il appartient*. Ce sont les différences entre les mots qui donnent la valeur à chaque forme de pensée constituée et perceptible par la langue. D'autre part, la signification se définit

verticalement à travers la relation arbitraire entre signifiant et signifié, alors que le plan horizontal est dominé par les relations discrètes et intrasystémiques entre les signes.

En outre, l'idée de valeur, ainsi déterminée, nous montre que c'est une grande illusion de considérer un terme simplement comme l'union d'un certain son avec un certain concept. Le définir ainsi, ce serait l'isoler du système dont il fait partie ; ce serait croire qu'on peut commencer par les termes et construire le système en en faisant la somme, alors qu'au contraire c'est du tout solidaire qu'il faut partir pour obtenir par analyse les éléments qu'il renferme [...]. [...] la langue est un système dont tous les termes sont solidaires et où la valeur de l'un ne résulte que de la présence simultanée des autres [...] ⁶³.

Pour revenir à la différence entre signification et valeur, la signification s'organise sur deux plans, soit celui de *l'interaction entre signifiant et signifié* et celui des *relations intrasystémiques*. La *valeur*, de son côté, est *l'élément qui donne du poids aux relations intrasystémiques* et permet l'intégration de ces dernières dans les processus de la signification. En d'autres termes, le processus de signification d'un mot n'est pas limité à la relation arbitraire entre signifiant et signifié, car sa valeur est établie en évaluant les relations que le mot entretient avec les autres. Enfin, pour Saussure, *la valeur d'un mot est le résultat d'une fonction qui permet de le distinguer des autres mots*.

À mon avis, la notion de valeur constitue une base théorique de nature sémiotique pour comprendre la notion de « poids » proposée par la fonction TF-IDF. En effet, de manière similaire à la notion de valeur chez Saussure, l'objectif de la fonction TF-IDF est d'intégrer la globalité du système dans lequel un mot (ou lemme) s'insère afin d'évaluer son « poids » dans un document. En assignant un poids à chaque occurrence d'un mot (lemme), la fonction TF-IDF évalue toujours sa valeur, c'est-à-dire la position qu'il occupe à l'intérieur de l'« univers clos » qui constitue le corpus. Autrement dit, on évalue la valeur sémiotique de la caractéristique qui fait partie du vocabulaire *V* à l'intérieur de l'univers-clos du corpus *D* et on lui assigne une valeur numérique. Si la valeur sémiotique d'un mot se définit à l'intérieur du système-langue, le poids TF-IDF d'un mot (lemme) est défini à l'intérieur du *système-corpus*. Le système dans lequel un mot se positionne, qu'il soit général comme la langue ou relatif comme un corpus, ne peut pas être construit en commençant « par les termes et [...] en en faisant la somme » ⁶⁴, mais en débutant par le « tout solidaire » ⁶⁵, c'est-à-dire, par les relations entre les éléments qui le constituent. François Rastier va dans le même sens quand il affirme que :

[...] sur l'axe syntagmatique, les classes sont définies par observation des cooccurrences : or, la création de listes de cooccurrences est une des opérations les plus simples et les plus éprouvées que permet la linguistique de corpus. Ainsi le concept de valeur étend-il également son efficacité aux groupements syntagmatiques que définissent les

contextes : il importe donc à présent de redéfinir la valeur comme un rapport du texte à ses unités constituantes d'une part, à son corpus d'autre part⁶⁶.

Ce dernier extrait met en évidence le rôle du concept de valeur dans un cadre d'analyse propre à la linguistique de corpus. La valeur a un rôle très important dans la détermination du poids que les unités du texte ont en fonction du corpus, ce qui implique d'identifier la valeur d'un mot à travers le *jeu de cooccurrences* dans les textes et dans le corpus, duquel le mot fait partie.

En résumé, je propose de considérer la fonction TF-IDF comme une tentative « non volontaire » pour l'approximation mathématique du concept de valeur chez Saussure. La fonction TF-IDF évite d'isoler le mot du corpus dont il fait partie et, pour ce faire, la fréquence des mots est combinée à la fréquence documentaire inverse. De cette façon, comme c'est le cas pour la valeur sémiotique, le TF-IDF met en évidence les oppositions et les différences entre les termes au moyen de l'identification de *seuils discriminants* entre les mots, seuils qui coupent le recto et le verso d'une feuille. Enfin, comme le dirait Saussure, *dans le corpus il n'y a que différences*.

Conclusion

Cet article à visée introductive a été réalisé en deux parties. Dans la première partie, un survol des champs d'étude de la sémiotique computationnelle a été présenté. J'ai proposé un portrait en trois branches, chacune d'elles caractérisant une approche particulière au regard de l'ordinateur : l'ordinateur comme machine sémiotique, où la sémiose est appréhendée comme un fait computable ; l'ordinateur comme artéfact, où la sémiose dérive de l'interaction entre l'humain et la machine ; et, enfin, l'ordinateur comme outil, où l'ordinateur est au service de l'analyse sémiotique classique. De ces branches, la troisième a fait l'objet d'un approfondissement. Ainsi, la pratique de l'analyse de texte assistée par ordinateur (ATO) a été présentée dans l'optique d'en interroger les enjeux sémiotiques. Poursuivant la démarche didactique adoptée dans cet article, j'ai examiné une étape spécifique de l'ATO, à savoir la transformation vectorielle du texte. Celle-ci est fondamentale pour la compréhension de l'usage de l'ATO et de l'usage d'outils numériques pour l'analyse sémiotique. J'ai ainsi présenté le modèle vectoriel, son cadre théorique et deux de ses opérations. J'ai souligné également l'importance de ces deux opérations pour le transfert des simulacres de la signification d'un modèle complexe et ambigu comme le texte à un modèle formel et computable comme la matrice.

L'une des contributions de cet article est d'avoir fourni une présentation originale de la sémiotique computationnelle, pour laquelle j'ai mis en évidence les différentes typologies de sémiose qu'elle étudie, et de l'ATO, pour laquelle j'ai fourni une description schématique de son interdisciplinarité. Ce texte contribue également à la discussion sur la procédure de lemmatisation et à prolonger les réflexions existantes sur la complémentarité entre la sémantique vectorielle et la linguistique structurale. À travers ces éléments, il met en lumière les fondements sémiotiques dans l'usage de l'ATO. En raison de sa nature introductive, l'article s'adresse principalement aux chercheurs des sciences humaines et sociales qui s'intéressent depuis peu à la sémiotique computationnelle.

En guise d'ouverture, ces réflexions sur les pratiques de l'ATO mènent à un élément important pour la sémiotique computationnelle, soit l'*approche empirique*, qui est favorisée par l'utilisation de corpus de très grande taille. Ceci correspond au contexte épistémologique qui caractérise l'analyse de mégadonnées (*big data*). La locution *big data* désigne le phénomène d'accumulation massive de données. Une grande masse de données sur le même phénomène constitue une valeur ajoutée pour l'analyse, mais elle apporte des défis, et plus particulièrement, l'emploi d'une approche méthodologique qui se caractérise par le concept de « recherche conduite par les données ». Cette approche s'inscrit dans l'approche empirique la plus classique et, en quelque sorte, la revitalise. En effet, cette approche, appelée « approche *data-driven* »⁶⁷, inaugure une nouvelle ère de l'empirisme, qui se caractérise par le volume et la diversité des données mises à disposition de l'analyse, ainsi que par une panoplie d'outils appartenant au domaine de la science de données. Dans ce contexte, l'approche empirique des données se radicalise, car l'analyse et l'extraction de connaissances sont effectuées sans les contraintes d'une théorie imposée d'avance⁶⁸. Ainsi, l'analyse de données devient un outil à disposition des SHS, mais au risque de laisser les données parler par elles-mêmes, sans (ou avec peu de) contraintes théoriques.

Avec les changements qu'apporte le numérique, la sémiotique doit affronter de nouveaux défis amenés par les nouvelles techniques offertes par la science des données, l'apprentissage automatique et l'accumulation massive de données. Elle doit également évaluer les implications sémiotiques des outils utilisés, afin d'assister l'interprétation des résultats et la construction d'algorithmes et de chaînes de traitement pour l'analyse d'artéfacts sémiotiques. Cette nouvelle aventure de la sémiotique n'est qu'au début de son parcours, mais le potentiel de recherche qui s'entrevoit est déjà très important.

Bibliographie

- ADAM, Jean-Michel, *La linguistique textuelle. Introduction à l'analyse textuelle des discours*, Paris, Armand Colin, 2011.
- AGGARWAL, Charu C. & ChengXiang ZHAI (dir.), *Mining Text Data*, New York, Springer, 2012.
- BEAUDOUIN, Valérie, « Statistical Analysis of Textual Data: Benzécri and the French School of Data Analysis », *Glottometrics*, no 33, 2016, p. 56-72.
- BENZÉCRI, Jean-Paul & Françoise BENZÉCRI, *Analyse des Correspondances : exposé élémentaire*, Paris, Dunod, 1980.
- BERNARD, Michel & Baptiste BOHET, *Littérométrie : outils numériques pour l'analyse des textes littéraires*, Paris, Presses Sorbonne nouvelle, 2017.
- BISHOP, Christopher M., *Pattern Recognition and Machine Learning*, Singapour, Springer, 2006.
- BOYD-GRABER, Jordan, Yuening HU & David MIMNO, « Applications of Topic Models », *Foundations and Trends® in Information Retrieval*, vol. 11, no 2-3, 2017, p. 143-296.
- CARLEY, Kathleen, « Content Analysis », dans R. E. Asher et al. (dir.), *The Encyclopedia of Language and Linguistics*, vol. 2, Édimbourg, Pergamon Press, 1990, p. 725-730.
- CHARTIER, Jean-François, Davide PULIZZOTTO, Louis CHARTRAND & Jean-Guy MEUNIER, « A Data-Driven Computational Semiotics: The Semantic Vector Space of Magritte's Artworks », *Semiotica*, vol. 2019, no 230, 2019, p. 19-69.
- CLARK, Alexander, Chris FOX & Shalom LAPPIN, *The Handbook Of Computational Linguistics and Natural Language Processing*, Malden, Wiley-Blackwell, 2010.
- COMPAGNO, Dario (dir.), *Quantitative Semiotic Analysis*, New York, Springer, 2018.
- DACOS, Marin & Pierre MOUNIER, *Humanités numériques. État des lieux et positionnement de la recherche française dans le contexte international*, rapport de recherche, Institut français, 2015.
- DE SOUZA, Clarisse Sieckenius, *The Semiotic Engineering of Human-Computer Interaction*, Cambridge, The MIT Press, 2005.
- DIJK, Teun A. van, « Grammaires textuelles et structures narratives », dans S. Alexandrescu et al. (dir.), *Sémiotique narrative et textuelle*, Paris, Larousse, 1973, p. 177-207.
- ECO, Umberto, « Pour une reformulation du concept de signe iconique », *Communications*, no 29, 1978, p. 141-191.
- ERTEL, Wolfgang, *Introduction to Artificial Intelligence*, Londres, Springer, 2011.

- ETXEBERRIA, Arantza & Jesus IBÁÑEZ, « Semiotics of the Artificial: The "Self" of Self-Reproducing Systems in Cellular Automata », *Semiotica*, vol. 127, no 1-4, 1999, p. 295-320.
- FABBRI, Paolo, *Le tournant sémiotique*, trad. de l'italien par Y. Jeanneret, Paris, Hermès Science publications-Lavoisier, 2008.
- FABRE, Cécile & Alessandro LENCI, « Sémantique distributionnelle », *Traitement automatique des langues*, vol. 56, no 2, 2015, 2015, p. 7-23.
- FETZER, James H., « Minds and Machines: Limits to Simulations of Thought and Action », *International Journal of Signs and Semiotic Systems*, vol. 1, no 1, 2011, p. 39-48.
- FIRTH, John Rupert, *Papers in Linguistics, 1934-1951*, Londres, Oxford University Press, 1957.
- GAGNON, Michel, Amal ZOUAQ, Francisco ARANHA, Faezeh ENSAN & Ludovic JEAN-LOUIS, « An Analysis of the Semantic Annotation Task on the Linked Data Cloud », *International Journal of Metadata, Semantics and Ontologies*, vol. 13, no 4, 2019, p. 317-329.
- GAMBHIR, Mahak & Vishal GUPTA, « Recent Automatic Text Summarization Techniques: A Survey », *Artificial Intelligence Review*, vol. 47, no 1, 2017, p. 1-66.
- GREIMAS, Algirdas J. & Joseph COURTÈS, *Sémiotique : dictionnaire raisonné de la théorie du langage*, Paris, Hachette, 1979.
- GUDWIN, Ricardo & Fernando A. C. GOMIDE, « Computational Semiotics: An Approach for the Study of Intelligent Systems-Part I: Foundations », Technical Report RT-DCA 09 - DCA-FEEC-UNICAMP, 1997.
- HARRIS, Zellig S., « Distributional Structure », *Word*, vol. 10, no 2-3, 1954, p. 146-162.
- HASTIE, Trevor, Robert TIBSHIRANI & Jerome FRIEDMAN, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York, Springer, 2013.
- HJELMSLEV, Louis, *Prolégomènes à une théorie du langage*, trad. du danois par U. Canger, avec la collab. d'A. Wewer, préface de V. Hjelm, Paris, Minuit, coll. « Arguments », 1968.
- JAMES, Gareth, Daniela WITTEN, Trevor HASTIE & Robert TIBSHIRANI, *An Introduction to Statistical Learning: with Applications in R*, New York, Springer, 2013.
- KETNER, Kenneth Laine, « Peirce and Turing: Comparisons and conjectures », *Semiotica*, vol. 68, no 1-2, 1988, p. 33-62.
- KITCHIN, Rob, « Big Data, New Epistemologies and Paradigm Shifts », *Big Data & Society*, vol. 1, no 1, 2014, p. 1-12.
- KRIPPENDORFF, Klaus, *Content Analysis: An Introduction to Its Methodology*, Thousand Oaks, SAGE Publications, 2004.

- LEBART, Ludovic, Bénédicte PINCEMIN & Céline POUDAT, *Analyse des données textuelles*, Québec, Presses de l'Université du Québec, 2019.
- LEBART, Ludovic & André SALEM, *Statistique textuelle*, Paris, Dunod, 1994.
- MAINGUENEAU, Dominique, *L'analyse du discours*, Paris, Hachette Supérieur, 1997.
- MALATERRE, Christophe, Jean-François CHARTIER & Davide PULIZZOTTO, « What Is This Thing Called Philosophy of Science? A Computational Topic-Modeling Perspective, 1934-2015 », *HOPOS. The Journal of the International Society for the History of Philosophy of Science*, vol. 9, no 2, 2019, p. 215-249.
- MALATERRE, Christophe, Davide PULIZZOTTO & Francis LAREAU, « Revisiting Three Decades of *Biology and Philosophy*: A Computational Topic-Modeling Perspective », *Biology & Philosophy*, vol. 35, no 1, 2020, p. 1-25.
- MANI, Inderjeet & Mark T. MAYBURY, *Advances in Automatic Text Summarization*, Cambridge, The MIT Press, 1999.
- MANNING, Christopher D., Prabhakar RAGHAVAN & Hinrich SCHÜTZE, *Introduction to Information Retrieval*, Cambridge, Cambridge University Press, 2008.
- MANNING, Christopher D. & Hinrich SCHÜTZE, *Foundations of Statistical Natural Language Processing*, Cambridge, The MIT Press, 1999.
- MARRONE, Gianfranco, *Corpi sociali: processi comunicativi e semiotica del testo*, Turin, Einaudi, 2001.
- MEUNIER, Jean-Guy, « Artificial intelligence and sign theory », *Semiotica*, vol. 77, no 1-3, 1989, p. 43-64.
- , « Humanités numériques et modélisation scientifique », *Questions de communication*, vol. 1, no 31, 2017, p. 19-48.
- MORENO, Juan Manuel Torres, *Automatic Text Summarization*, Somerset, Wiley, 2014.
- MORETTI, Franco, *Distant reading*, Londres, Verso Books, 2013.
- NADIN, Mihai, « Information and Semiotic Processes: The Semiotics of Computation », *Cybernetics & Human Knowing*, vol. 18, no 1-2, 2011, p. 153-175.
- NÉE, Émilie (dir.), *Méthodes et outils informatiques pour l'analyse des discours*, Rennes, Presses universitaires de Rennes, coll. « Didact Méthode », 2017.
- POLGUÈRE, Alain, *Lexicologie et sémantique lexicale : notions fondamentales*, Montréal, Presses de l'Université de Montréal, 2016.
- POUDAT, Céline & Frédéric LANDRAGIN, *Explorer un corpus textuel : méthodes, pratiques, outils*, Paris, De Boeck supérieur, 2017.

- PULIZZOTTO, Davide, Jean-François CHARTIER, Francis LAREAU, Jean-Guy MEUNIER & Louis CHARTRAND, « Conceptual Analysis in a Computer-Assisted Framework: Mind in Peirce », *Umanistica Digitale*, vol. 2, no 2, 2018, p. 185-205.
- PULIZZOTTO, Davide, Jean-François CHARTIER, Jean-Guy MEUNIER, Louis CHARTAND, Francis LAREAU & Louis HÉBERT, « Vers une sémiotique computationnelle : étude de cas et premières explorations », *Applied Semiotics / Semiotique appliquée*, no 26, 2018, p. 192-208.
- QUEIROZ, João & Floyd MERRELL, « On Peirce's Pragmatic Notion of Semiosis—A Contribution for the Design of Meaning Machines », *Minds and Machines*, vol. 19, no 1, 2009, p. 129-143.
- RAPAPORT, William J., « Semiotic Systems, Computers, and the Mind: How Cognition Could Be Computing », *International Journal of Signs and Semiotic Systems*, vol. 2, no 1, 2012, p. 32-71.
- RASTIER, François, *Sémantique interprétative*, Paris, Presses universitaires de France, 2009.
- , « Objets culturels et performances sémiotiques. L'objectivation critique dans les sciences de la culture », dans L. Hébert & L. Guillemette (dir.), *Performances et objets culturels. Nouvelles perspectives*, Sainte-Foy, Presses de l'Université Laval, 2010, p. 15-58.
- , *La mesure et le grain : sémantique de corpus*, Paris, Honoré Champion, 2011.
- , « Computer-Assisted Interpretation of Semiotic Corpora », dans D. Compagno (dir.), *Quantitative Semiotic Analysis*, New York, Springer, 2018, p. 123-139.
- RICH, Elaine, Kevin KNIGHT & Shivashankar B. NAIR, *Artificial intelligence*, New Delhi, Tata McGraw-Hill, 2009.
- ROSENBLATT, Frank, « The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain », *Psychological review*, vol. 65, no 6, 1958, p. 386-408.
- SAHLGREN, Magnus, *The Word-Space Model*, thèse de doctorat, Stockholm, Université de Stockholm, 2006.
- , « The Distributional Hypothesis », *Italian Journal of Linguistics*, vol. 20, no 1, 2008, p. 33-53.
- SALTON, Gerard & Michael J. MCGILL, *Introduction to Modern Information Retrieval*, New York, McGraw-Hill, 1983.
- SALTON, Gerard, Amit SINGHAL, Mandar MITRA & Chris BUCKLEY, « Automatic Text Structuring and Summarization », *Information Processing & Management*, vol. 33, no 2, 1997, p. 193-207.

- SALTON, Gerard, Andrew WONG & Chung-Shu YANG, « A Vector Space Model for Automatic Indexing », *Communications of the ACM*, vol. 18, no 11, 1975, p. 613-620.
- SAUSSURE, Ferdinand de, *Cours de linguistique générale*, publié par C. Bailly & A. Sechehaye avec la collab. d'A. Riedlinger, éd. critique préparée par T. de Mauro, Paris, Payot, 1995 [1916].
- SHAH, Neepea & Sunita MAHAJAN, « Document Clustering: A Detailed Review », *International Journal of Applied Information Systems*, vol. 4, no 5, 2012, p. 30-38.
- STAMPER, Ronald, *Information in Business and Administrative Systems*, New York, John Wiley & Sons, 1973.
- TANAKA-ISHII, Kumiko, *Semiotics of programming*, New York, Cambridge University Press, 2010.
- , « Semiotics of Computing: Filling the Gap Between Humanity and Mechanical Inhumanity », dans P. P. Trifonas (dir.), *International Handbook of Semiotics*, New York, Springer, 2015, p. 981-1002.
- TURNEY, Peter D. & Patrick PANTEL, « From Frequency to Meaning: Vector Space Models of Semantics », *Journal of Artificial Intelligence Research*, vol. 37, 2010, p. 141-188.
- ZEMANEK, Heinz, « Semiotics and Programming Languages », *Communications of the ACM*, vol. 9, no 3, 1966, p. 139-143.

Notes

- 1 P. D. TURNEY & P. PANTEL, « From Frequency to Meaning: Vector Space Models of Semantics », *Journal of Artificial Intelligence Research*, vol. 37, 2010, p. 141-188.
- 2 T. A. van DIJK, « Grammaires textuelles et structures narratives », dans S. Alexandrescu et al. (dir.), *Sémiotique narrative et textuelle*, Paris, Larousse, 1973, p. 177-207.
- 3 M. SAHLGREN, *The Word-Space Model*, thèse de doctorat, Stockholm, Université de Stockholm, 2006.
- 4 M. SAHLGREN, « The Distributional Hypothesis », *Italian Journal of Linguistics*, vol. 20, no 1, 2008, p. 33-54.
- 5 Voici une sélection des plus récents travaux : J.-F. CHARTIER et al., « A Data-Driven Computational Semiotics: The Semantic Vector Space of Magritte's Artworks », *Semiotica*, no 230, 2019, p. 19-69 ; C. MALATERRE et al., « What Is This Thing Called Philosophy of Science? A Computational Topic-Modeling Perspective, 1934-2015 », *HOPOS. The Journal of the International Society for the History of Philosophy of Science*, vol. 9, no 2, 2019, p. 215-249 ; C. MALATERRE, D. PULIZZOTTO & F. LAREAU, « Revisiting Three Decades of Biology and

- Philosophy: A Computational Topic-Modeling Perspective* », *Biology & Philosophy*, vol. 35, no 1, 2020, p. 1-25 ; D. PULIZZOTTO et al., « Conceptual Analysis in a Computer-Assisted Framework: Mind in Peirce », *Umanistica Digitale*, vol. 2, no 2, 2018, p. 185-205 ; D. PULIZZOTTO et al., « Vers une sémiotique computationnelle : étude de cas et premières explorations », *Applied Semiotics / Sémiotique appliquée*, no 26, 2018, p. 192-208.
- 6 M. DACOS & P. MOUNIER, *Humanités numériques. État des lieux et positionnement de la recherche française dans le contexte international*, rapport de recherche, Institut français, 2015, p. 7.
 - 7 Cf. W. ERTEL, *Introduction to Artificial Intelligence*, Londres, Springer, 2011, p. 2.
 - 8 E. RICH, K. KNIGHT & S. B. NAIR, *Artificial intelligence*, New Delhi, Tata McGraw-Hill, 2009, p. 3. Je traduis : « Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better. »
 - 9 G. MARRONE, *Corpi sociali: processi comunicativi e semiotica del testo*, Turin, Einaudi, 2001.
 - 10 Cf. A. ETXEBERRIA & J. IBÁÑEZ, « Semiotics of the Artificial: The "Self" of Self-Reproducing Systems in Cellular Automata », *Semiotica*, vol. 127, no 1-4, 1999, p. 295-320 ; J. H. FETZER, « Minds and Machines: Limits to Simulations of Thought and Action », *International Journal of Signs and Semiotic Systems*, vol. 1, no 1, 2011, p. 39-48 ; K. L. KETNER, « Peirce and Turing: Comparisons and conjectures », *Semiotica*, vol. 68, no 1-2, 1988, p. 33-62 ; J.-G. MEUNIER, « Artificial Intelligence and Sign Theory », *Semiotica*, vol. 77, no 1-3, 1989, p. 43-64 ; W. J. RAPAPORT, « Semiotic Systems, Computers, and the Mind: How Cognition Could Be Computing », *International Journal of Signs and Semiotic Systems*, vol. 2, no 1, 2012, p. 32-71.
 - 11 Cf. C. S. DE SOUZA, *The Semiotic Engineering of Human-Computer Interaction*, Cambridge, The MIT Press, 2005 ; M. NADIN, « Information and Semiotic Processes: The Semiotics of Computation », *Cybernetics & Human Knowing*, vol. 18, no 1-2, 2011, p. 153-175 ; R. K. STAMPER, *Information in Business and Administrative Systems*, New York, John Wiley & Sons, 1973 ; K. TANAKA-ISHII, *Semiotics of programming*, New York, Cambridge University Press, 2010 ; H. ZEMANEK, « Semiotics and Programming Languages », *Communications of the ACM*, vol. 9, no 3, 1966, p. 139-143.
 - 12 Cf. L. LEBART & A. SALEM, *Statistique textuelle*, Paris, Dunod, 1994 ; F. MORETTI, *Distant reading*, Londres, Verso Books, 2013 ; M. BERNARD & B. BOHET, *Littérométrie : outils numériques pour l'analyse des textes littéraires*, Paris, Presses Sorbonne nouvelle, 2017 ; D. COMPAGNO (dir.), *Quantitative Semiotic Analysis*, New York, Springer, 2018 ; F. RASTIER, « Computer-Assisted Interpretation of Semiotic Corpora », dans D. Compagno (dir.), *Quantitative Semiotic Analysis*, New York, Springer, 2018, p. 123-139 ; J.-F. CHARTIER et al., « A Data-Driven Computational Semiotics: The Semantic Vector Space of Magritte's Artworks », *loc. cit.*
 - 13 J.-G. MEUNIER, « Humanités numériques et modélisation scientifique », *Questions de communication*, vol. 1, no 31, 2017, p. 19-48.
 - 14 K. TANAKA-ISHII, « Semiotics of Computing: Filling the Gap Between Humanity and Mechanical Inhumanity », dans P. P. Trifonas (dir.), *International Handbook of Semiotics*, New York, Springer, 2015, p. 981-1002.
 - 15 R. R. GUDWIN & F. A. C. GOMIDE, « Computational Semiotics: An Approach for the Study of Intelligent Systems-Part I: Foundations », Technical Report RT-DCA 09 - DCA-FEEC-UNICAMP, 1997.
 - 16 J. QUEIROZ & F. MERRELL, « On Peirce's Pragmatic Notion of Semiosis—A Contribution for the Design of Meaning Machines », *Minds and Machines*, vol. 19, no 1, 2009, p. 129-143.
 - 17 A. ETXEBERRIA & J. IBÁÑEZ, « Semiotics of the Artificial: The "Self" of Self-Reproducing Systems in Cellular Automata », *loc. cit.*
 - 18 Cf. J.-G. MEUNIER, « Artificial Intelligence and Sign Theory », *loc. cit.* ; J.-G. MEUNIER, « Humanités numériques et modélisation scientifique », *loc. cit.*

- 19 F. RASTIER, « Objets culturels et performances sémiotiques. L'objectivation critique dans les sciences de la culture », dans L. Hébert & L. Guillemette (dir.), *Performances et objets culturels. Nouvelles perspectives*, Sainte-Foy, Presses de l'Université Laval, 2010, p. 17-19.
- 20 Cf. C. S. DE SOUZA, *The Semiotic Engineering of Human-Computer Interaction*, op. cit.
- 21 Cf. R. K. STAMPER, *Information in Business and Administrative Systems*, op. cit.
- 22 Cf. L. LEBART & A. SALEM, *Statistique textuelle*, op. cit.
- 23 Cf. K. CARLEY, « Content Analysis », dans R. E. Asher et al. (dir.), *The Encyclopedia of Language and Linguistics*, vol. 2, Édimbourg, Pergamon Press, 1990, p. 725-730 ; K. KRIPPENDORFF, *Content Analysis: An Introduction to Its Methodology*, Thousand Oaks, SAGE Publications, 2004.
- 24 Cf. J.-M. ADAM, *La linguistique textuelle. Introduction à l'analyse textuelle des discours*, Paris, Armand Colin, 2011 ; D. MAINGUENEAU, *L'analyse du discours*, Paris, Hachette Supérieur, 1997.
- 25 Cf. C. POUDAT & F. LANDRAGIN, *Explorer un corpus textuel : méthodes, pratiques, outils*, Paris, De Boeck supérieur, 2017 ; F. RASTIER, *Sémantique interprétative*, Paris, Presses universitaires de France, 2009.
- 26 On parle aussi de « statistique textuelle », de « textométrie », de « sémiométrie », de « distant reading », de « linguistique de corpus », etc. Les pratiques auxquelles ces expressions réfèrent, bien qu'avec des différences, sont très comparables à celles de l'ATO.
- 27 Cf. E. NÉE (dir.), *Méthodes et outils informatiques pour l'analyse des discours*, Rennes, Presses universitaires de Rennes, coll. « Didact Méthode », 2017.
- 28 Cf. C. POUDAT & F. LANDRAGIN, *Explorer un corpus textuel*, op. cit., p. 90.
- 29 Cf. J.-P. BENZÉCRI & F. BENZÉCRI, *Analyse des Correspondances : exposé élémentaire*, Paris, Dunod, 1980.
- 30 Cf. C. D. MANNING & H. SCHÜTZE, *Foundations of Statistical Natural Language Processing*, Cambridge, The MIT Press, 1999.
- 31 Cf. A. CLARK, C. FOX & S. LAPPIN, *The Handbook Of Computational Linguistics and Natural Language Processing*, Malden, Wiley-Blackwell, 2010.
- 32 Avec la locution de lexicographe automatique, je regroupe les techniques qui permettent d'associer une suite de caractères à une entrée de dictionnaire d'une langue, comme la lemmatisation ou, dans une moindre mesure, la racinisation. Les analyseurs morphosyntaxiques, quant à eux, sont des techniques qui permettent d'effectuer une analyse morphosyntaxique automatisée de chaque mot d'un texte.
- 33 Cf. G. SALTON et al., « Automatic Text Structuring and Summarization », *Information Processing & Management*, vol. 33, no 2, 1997, p. 193-207 ; I. MANI & M. T. MAYBURY, *Advances in Automatic Text Summarization*, Cambridge, The MIT Press, 1999 ; J. M. T. MORENO, *Automatic Text Summarization*, Somerset, Wiley, 2014 ; M. GAMBHIR & V. GUPTA, « Recent Automatic Text Summarization Techniques: A Survey », *Artificial Intelligence Review*, vol. 47, no 1, 2017, p. 1-66.
- 34 Cf. C. M. BISHOP, *Pattern Recognition and Machine Learning*, Singapour, Springer, 2006 ; T. HASTIE, R. TIBSHIRANI & J. FRIEDMAN, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York, Springer, 2013.
- 35 Cf. G. JAMES et al., *An Introduction to Statistical Learning: with Applications in R*, New York, Springer, 2013.
- 36 Cf. N. SHAH & S. MAHAJAN, « Document Clustering: A Detailed Review », *International Journal of Applied Information Systems*, vol. 4, no 5, 2012, p. 30-38.
- 37 Cf. J. BOYD-GRABER, Y. HU & D. MIMNO, « Applications of Topic Models », *Foundations and Trends® in Information Retrieval*, vol. 11, no 2-3, 2017, p. 143-296.

- 38 Cf. C. FABRE & A. LENCI, « Sémantique distributionnelle », *Traitement automatique des langues*, vol. 56, no 2, 2015, 2015, p. 7-23.
- 39 La recherche d'information est un domaine de l'informatique qui étudie les manières d'extraire de l'information d'une collection de documents à partir d'une requête précise fournie par l'utilisateur. Voir notamment G. SALTON & M. J. MCGILL, *Introduction to Modern Information Retrieval*, New York, McGraw-Hill, 1983.
- 40 Cf. C. C. AGGARWAL & CX. ZHAI (dir.), *Mining Text Data*, New York, Springer, 2012.
- 41 F. ROSENBLATT, « The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain », *Psychological review*, vol. 65, no 6, 1958, p. 386-408.
- 42 G. SALTON, A. WONG & C. S. YANG « A Vector Space Model for Automatic Indexing », *Communications of the ACM*, vol. 18, no 11, 1975, p. 613-620.
- 43 V. BEAUDOUIN, « Statistical Analysis of Textual Data: Benzécri and the French School of Data Analysis », *Glottometrics*, no 33, 2016, p. 56-72.
- 44 Firth a formulé une expression devenue phrase : « Il est possible de connaître la signification d'un mot en regardant avec qui il se tient. » J. R. FIRTH, *Papers in Linguistics, 1934-1951*, Londres, Oxford University Press, 1957, p. 11. Je traduis : « You shall know a word by the company it keeps. »
- 45 Z. HARRIS, « Distributional Structure », *Word*, vol. 10, no 2-3, 1954, p. 146-162.
- 46 Les langues flexionnelles, à la différence des langues isolantes, ont une morphologie complexe, ce qui justifie des opérations de lemmatisation.
- 47 Cf. A. POLGUÈRE, *Lexicologie et sémantique lexicale : notions fondamentales*, Montréal, Presses de l'Université de Montréal, 2016.
- 48 Cf. G. SALTON & M. J. MCGILL, *Introduction to Modern Information Retrieval*, op. cit.
- 49 Cf. C. D. MANNING, P. RAGHAVAN & H. SCHÜTZE, *Introduction to Information Retrieval*, Cambridge, Cambridge University Press, 2008, p. 39.
- 50 Cf. L. LEBART, B. PINCEMIN & C. POUDAT, *Analyse des données textuelles*, Québec, Presses de l'Université du Québec, 2019.
- 51 U. ECO, « Pour une reformulation du concept de signe iconique », *Communications*, vol. 29, 1978, p. 141-191.
- 52 *Ibid.*, p. 141.
- 53 *Idem.*
- 54 *Idem.*
- 55 Relié à Wikipédia, DBpedia forme un « projet universitaire et communautaire d'exploration et extraction automatiques de données dérivées de Wikipédia. Son principe est de proposer une version structurée et normalisée au format du web sémantique des contenus de Wikipedia ». En ligne : <<https://fr.wikipedia.org/wiki/DBpedia>>. Voir également le site web du projet : <<https://wiki.dbpedia.org/>>.
- 56 M. GAGNON et al., « An Analysis of the Semantic Annotation Task on the Linked Data Cloud », *International Journal of Metadata, Semantics and Ontologies*, vol. 13, no 4, 2019, p. 317-329.
- 57 M. SAHLGREN, « The Distributional Hypothesis », *Italian Journal of Linguistics*, loc. cit.
- 58 F. de SAUSSURE, *Cours de linguistique générale*, publié par C. Bailly & A. Sechehaye avec la collab. d'A. Riedlinger, éd. critique préparée par T. de Mauro, Paris, Payot, 1995 [1916], p. 160.
- 59 *Ibid.*, p. 155.
- 60 *Idem.*
- 61 *Ibid.*, p. 156.

62 *Ibid.*, p. 166.

63 *Ibid.*, p. 157-159.

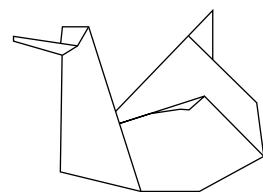
64 *Ibid.*, p. 157.

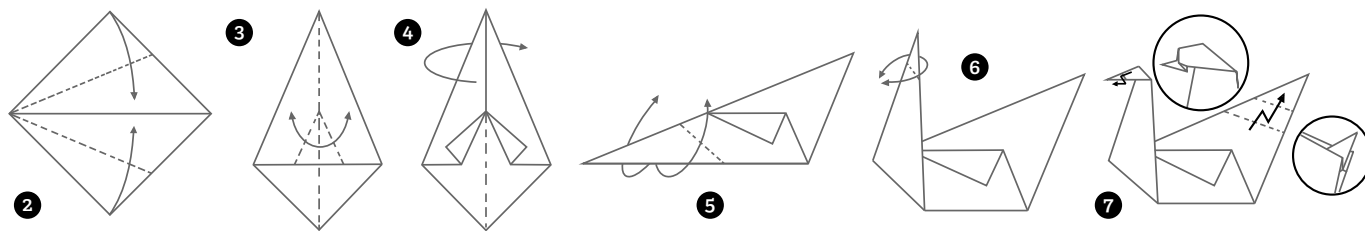
65 *Idem.*

66 F. RASTIER, *La mesure et le grain : sémantique de corpus*, Paris, Honoré Champion, 2011, p. 30.

67 R. KITCHIN, « Big Data, New Epistemologies and Paradigm Shifts », *Big Data & Society*, vol. 1, no 1, 2014, p. 1-12.

68 *Ibid.*





Enjeux de la modélisation formelle en sémiotique computationnelle

Jean-Guy MEUNIER

Cygne noir, no 7, 2019 : « Algorithmes »

Résumé

Une sémiotique computationnelle n'est pas une sémiotique de la computation ni des outils numériques ou de la relation personne-machine. Il s'agit d'une approche qui dans sa modélisation des artefacts sémiotiques construit des modèles computationnels qui en appellent à des modèles conceptuels et formels. Plus précisément, un modèle formel se construit en regard de ce qui a été conceptualisé à propos des artefacts sémiotiques à l'étude et est aussi contraint par les exigences de la modélisation computationnelle. Cet article traite des enjeux de la modélisation formelle pour une sémiotique computationnelle, à savoir : (a) la modélisation formelle dans les pratiques sémiotiques, (b) les définitions d'une modélisation formelle, (c) le rôle déterminant d'un modèle conceptuel sur la modélisation formelle, (d) la contrainte qu'exerce la modélisation computationnelle sur la modélisation formelle et, enfin, (e) les limites qu'impose la modélisation formelle à la recherche sémiotique.

Pour citer cet article

MEUNIER, Jean-Guy, « Enjeux de la modélisation formelle en sémiotique computationnelle », *Cygne noir*, no 7, 2019.

En ligne : <<http://www.revuecygnoir.org/numero/article/meunier-modelisation-semiotique-computationnelle>>
(consulté le xx/xx/xxxx).



Cet article de *Revue Cygne noir* est mis à disposition selon les termes de la licence Creative Commons :
Attribution - Pas d'Utilisation Commerciale - Pas de Modification 2.5 Canada.

ENJEUX DE LA MODÉLISATION FORMELLE EN SÉMIOTIQUE COMPUTATIONNELLE

[L]'analyse ne peut jamais se contenter de considérer les termes mais doit, par-delà les termes, saisir les relations qui les unissent. Ces relations seules constituent son véritable objet.

Claude Lévi-Strauss¹

1. Le défi de la sémiotique : la modélisation formelle

Vers la fin des années 2000, Peer Bundgaard et Frederik Stjernfelt ont entrepris une enquête au sujet des défis futurs de la sémiotique en interrogeant différents sémioticiens². La conclusion était sans appel : la sémiotique se renouvellerait par la puissance du structuralisme dominant. On voyait ce paradigme comme une approche plus rigoureuse que celle offerte par la philosophie classique. Juri Lotman, bien auparavant, avait lui aussi insisté sur l'importance des mathématiques comme voie naturelle pour la formalisation de la sémiotique³. En 2014, Kalevi Kull et Ekaterina Velmezova ont repris cette enquête, en interrogeant un autre groupe de sémioticiens⁴. Cette seconde investigation donna cependant des résultats bien différents. On se montrait très pessimiste sur l'avenir de la sémiotique, car, selon ces chercheurs, le structuralisme n'avait pas livré la rigueur attendue. L'application du structuralisme en sémiotique avait peu exploré la formalisation de la notion de structure⁵. La sémiotique était demeurée surtout analytique et conceptuelle. Comme l'affirmait Mihai Nadin, la sémiotique était devenue trop « spéculative »⁶. Aujourd'hui, une problématique similaire émerge au sein de la communauté sémiotique qui, en raison de l'introduction de l'informatique dans sa pratique, s'interroge sur les enjeux de la formalisation dans une sémiotique qui devient computationnelle.

Dans cet article, j'explorerai ainsi les enjeux de cette modélisation formelle pour une sémiotique computationnelle. Je préciserai (a) les tentatives de modélisation formelle dans la pratique sémiotique ; (b) les définitions techniques d'une modélisation formelle ; (c) le rôle déterminant d'un modèle conceptuel sur la modélisation formelle ; (d) les contraintes qu'exerce la modélisation computationnelle sur la modélisation formelle ; et enfin (e) les limites qu'impose la modélisation formelle à la recherche sémiotique.

2. La modélisation formelle dans les pratiques sémiotiques

Malgré les critiques élaborées par les chercheurs interrogés par Kull et Velmezova, la recherche d'une rigueur formelle a toujours été présente en sémiotique. En effet, on peut trouver de nombreux exemples qui illustrent ces recherches formalisantes. Par exemple, Charles S. Peirce, comme on le sait, a identifié la sémiotique avec la logique. Il a utilisé des graphes existentiels⁷. Robert Marty a traduit ses catégories peirciennes en algèbre de treillis⁸. Louis Hjelmslev représenta les relations entre la substance et la forme comme une relation fonctionnelle⁹. L'école sémiotique de Charles Morris utilisa la logique prédicative et modale de Frege-Carnap¹⁰ qui ultérieurement fut surtout appliquée à l'étude de la langue et dont les formes les plus complexes furent offertes par la grammaire applicative de Sebastian Shaumyan et de Jean-Pierre Desclés¹¹. Claude Lévi-Strauss utilisa des modèles algébriques¹² et Joachim et Michael Lambek, des grammaires catégorielles pour décrire des relations de parenté¹³. Avec des collègues, nous avons décrit des formes génériques iconiques par le biais de ces grammaires catégorielles¹⁴. A. J. Greimas proposa un carré logique pour sa structure actancielle¹⁵. Même Jacques Lacan et Roland Barthes proposèrent de « mathématiser » leurs objets d'étude¹⁶ – la psychanalyse et la mode –, mais sans trop de succès. Jaakko Hintikka travailla la sémiotique dans l'horizon de la logique modale¹⁷. Jean Petitot, Line Brandt et Wolfgang Wildgen s'inspirèrent de la topologie et des systèmes dynamiques (à la Thom) pour expliquer la morphogénèse du sens¹⁸. Dans cette même voie, David Piotrowski et Yves-Marie Visetti proposèrent une modélisation diagrammatique pour penser une sémiolinguistique dynamique¹⁹. Enfin, des efforts pour une vision formelle se sont aussi retrouvés dans les théories sémiotiques naturalistes comme la biosémiotique²⁰, la cybersémiotique²¹, la théorie de l'information²² et de la communication²³ et la sémiotique cognitive²⁴. Reste que la formalisation ne s'est pas constituée comme besoin de base de la pratique sémiotique. On la trouve réductionniste, positiviste, atomiste et rarement dynamique. Évidemment, on doute assurément qu'une sémiotique computationnelle arrangerait les choses. Se pose alors la question de la valeur épistémique d'une sémiotique computationnelle dont les fondements reposent sur la modélisation formelle que l'on voit décidément comme réductionniste.

Pourtant, malgré toutes ces critiques, et contrairement à toute attente, la communauté des informaticiens et des mathématiciens, en dehors même de la communauté sémiotique, a commencé à appliquer systématiquement de la modélisation formelle à des objets sémiotiques par le biais de traductions algorithmiques. En effet, les sciences informatiques, surtout celles d'inspiration cognitive, ont exploré, avec une vision sémiotique, des territoires qui traditionnellement n'appartenaient pas aux sémioticiens.

Par exemple, certains chercheurs en ont appelé à la sémiotique pour comprendre des dimensions de l'informatique elle-même²⁵.

On les retrouve en grand nombre dans les sous-disciplines de la fouille d'information, de la fouille textuelle, du traitement du langage ou du traitement d'images, etc. Des logiciels de divers types, souvent clé en main, sont utilisés pour effectuer certaines tâches d'analyse sémiotique. Cachés dans ces algorithmes et ces logiciels, se trouvent des grammaires formelles de tout type servant au traitement du langage, ou encore de l'algèbre linéaire et non linéaire pour le traitement d'images, de la parole ou de la musique, et enfin des modèles chaotiques, dynamiques, catastrophiques, évolutionnistes pour le traitement de la cognition. Quelques rares mathématiciens et informaticiens (à l'exemple de James Albus, Alexander Meystel, Jean Petitot, Ricardo Gudwin, Fernando Gomide, Burghart Rieger, Kumiko Tanaka-Ishii, Leonid Perlovsky, notamment)²⁶ verront explicitement un travail de nature sémiotique dans ces algorithmes et ces modèles formels.

Qui plus est, l'informatique contemporaine (l'intelligence artificielle, la fouille de données, l'apprentissage machine, l'apprentissage profond, etc.) a commencé à explorer de manière computationnelle des artefacts sémiotiques très classiques comme le langage naturel, les arts visuels, la musique, les jeux, les rituels, les médias, pour ne citer que cela. Là encore, on y traite de manière originale des données sémiotiques. Bien que ne quittant pas les approches classiques déductives descendantes (*top down*), elles introduisent maintenant des méthodes dites ascendantes ou inductives qui permettent la découverte de régularités structurales à propos des relations entre les données et donc ultimement du sens et de la signification de ces données.

Enfin un courant important issu de ces approches informatiques en est même émergé à la fin des années 1980 : les *humanités numériques* (HN)²⁷. Celles-ci ont de fait exploré des territoires appartenant classiquement à la sémiotique tels que le texte, le discours, l'argumentation, le genre, le style, les contenus iconiques, la musique, etc. Cependant, dans ces HN, l'informatique était souvent utilisée comme un outil qui, soutient-on, n'affectait pas la pratique classique des sciences humaines et sociales et des arts et des lettres : « Les Humanités numériques sont une extension des compétences et des méthodes traditionnelles du savoir, elles ne les remplacent pas²⁸. »

Pourtant, malgré la richesse des recherches menées dans le domaine, il semble bien que ces HN n'aient pas influencé de façon importante les sémioticiens – à l'exception de rares chercheurs en sciences humaines aux intérêts sémiotiques avoués qui ont intégré l'informatique dans leur démarche²⁹. La modélisation formelle de dimensions sémiotiques est maintenant entrée dans les projets des HN. Elle y demeure cependant peu

explicitée. Le formel reste caché sous les algorithmes ou sous la technologie informatique qui est souvent utilisée comme un outil dont on ignore les fondements formels³⁰.

Le courant semble aujourd'hui changer. Pour une nouvelle génération de jeunes sémioticiens et sémioticiennes, la sémiotique future doit non seulement utiliser ces outils informatiques, mais elle doit aussi comprendre et s'approprier la modélisation formelle et computationnelle qui lui est sous-jacente. Ce n'est qu'à cette condition qu'elle pourra l'appliquer avec rigueur et de manière heuristique à des artefacts sémiotiques. Ces modélisations formelles et computationnelles ouvriront de nouvelles voies de recherche pour la sémiotique. Si la communauté sémiotique ne s'approprie pas ces modélisations, les recherches sur les artefacts sémiotiques n'en continueront pas moins, mais il est fort possible qu'une bonne partie de celles-ci soient réalisées dans des sciences aux méthodes de type naturaliste, positiviste ou expérimental. N'est-ce pas ce que font déjà diverses disciplines des sciences cognitives, comme la psychologie, l'intelligence artificielle, la sociosémantique, les technologies de l'information et les études en communication? Les questions sémiotiques comme telles ne disparaîtront pas, mais les analyses sémiotiques demeureront conceptuelles et la philosophie y jouera un rôle d'épistémologie des théories sur le signe.

Par ailleurs il semble que la sémiotique est déjà confrontée aux problèmes que pose l'usage de l'informatique. De fait, elle en vit le défi sous divers couverts. Et, comme c'est le cas pour plusieurs autres disciplines scientifiques, l'informatique pourrait entraîner une modification profonde des objets d'étude de la sémiotique et de sa méthodologie³¹. On peut penser par exemple à l'impact des données massives (*Big data*) sur la sémiotique. Propp et Lévi-Strauss avaient réussi à étudier quelques dizaines de contes ou de légendes. Il leur serait maintenant possible d'en explorer des milliers. Et un tel nombre de données introduit nécessairement dans la recherche une note empirique plus forte. Ces données ne changent pas l'artefact sémiotique comme tel. Elles demeurent toujours des données « signifiantes ». Elles sont simplement en plus grand nombre. Cependant ce qui change profondément est notre vision de ces objets d'étude. Et une modélisation computationnelle invite nécessairement une modification de la méthode d'investigation et d'analyse.

Cependant, si une telle vision computationnelle de la sémiotique veut se développer en regard de l'informatique, elle doit mieux comprendre ses fondements théoriques et mieux les positionner en regard d'une formalisation; formalisation qui devient une condition *sine qua non* de son entrée dans l'univers computationnel et donc de l'utilisation de l'ordinateur.

3. La modélisation conceptuelle dans une sémiotique computationnelle

Souvent, on comprendra la formalisation comme une opération effectuée dans un « troisième monde » où elle est complètement détachée du monde réel ou actuel. Or comme le montre la philosophie récente des sciences³², la formalisation est une opération épistémique qui ne peut pas être détachée ou isolée de multiples autres opérations cognitives à l'œuvre dans une démarche scientifique et plus particulièrement de la modélisation conceptuelle.

Ceci est d'autant plus vrai dans le cas d'une modélisation formelle qui serait réalisée au sein d'une recherche sémiotique. En effet, il ne peut exister de modèles formels sémiotiques qui soient détachés de l'important travail conceptuel qu'une recherche sémiotique met en œuvre. Ce travail conceptuel est inhérent à la sémiotique en raison de la complexité des objets d'étude. Ceux-ci sont la complexité incarnée. Ils ne révéleront pas facilement leur signification à l'analyse, car celle-ci se tisse dans de multiples relations des plus intriquées. Pour paraphraser Peter Frederick Strawson, les signes ne portent pas leur signification sur leurs manches³³.

Cette modélisation conceptuelle est un travail omniprésent dans les recherches sémiotiques. Je dirais que le modèle conceptuel est celui qui consume le plus d'énergie épistémique. Et il ne doit pas être négligé. Tout comme les autres démarches scientifiques, la sémiotique doit conceptualiser ses propres objets d'étude pour révéler les constituants des artefacts sémiotiques et les relations structurantes qui les caractérisent. On ne peut pas penser que de tels objets complexes puissent être modélisés formellement sans, au préalable, avoir été conceptualisés d'une manière ou d'une autre. Comme le dit l'épistémologie classique, l'objet d'étude doit être *appréhendé*, c'est-à-dire non seulement senti, mais *conceptualisé*. Cette conceptualisation ne reste pas purement mentale pour un individu. Elle doit aussi en arriver à être exprimée dans un langage naturel pour en permettre la communication à une communauté épistémique. Ce travail de conceptualisation et d'expression linguistique est tantôt préalable tantôt concomitant à un éventuel modèle formel ou computationnel³⁴.

Comme le dit autrement Gilles-Gaston Granger, une modélisation formelle n'est qu'un horizon dans l'ensemble d'une démarche scientifique :

En tant que pensée en exercice, elle [la science] ne peut se présenter que comme une tentative de mise en forme, commentée par le truchement d'un langage non formel. La formalisation totale n'apparaît jamais que comme *un horizon de la pensée scientifique*, et l'on peut dire que la collaboration des deux langages est un caractère transcendantal, c'est-à-dire dépendant des conditions mêmes d'appréhension d'un objet³⁵.

Illustrons la nécessité d'un tel modèle conceptuel dans une analyse sémiotique. Commençons par un premier exemple simple : la sémiotique d'une fête d'anniversaire d'enfants dont on veut expliciter les constituants et les relations sémiotiques. Une analyse sémiotique de ce rituel ne peut pas débiter en projetant sur ce rituel un modèle formel, par ex. : un graphe des actions, une équation sur la dynamique d'échanges d'objets, etc. Il faut au préalable conceptualiser ce qui constitue ce rituel et exprimer ces concepts dans un langage naturel. Et selon les paradigmes théoriques dans lesquels le sémioticien se situera, il en appellera à diverses expressions conceptuelles pour désigner ces entités et ces relations. Certaines entités seront conceptualisées et catégorisées comme *acteur*, *actant*, *sujet*, *objets*, *instruments*, *lieu*, *temps*, etc. Elles se verront attribuer des propriétés, des traits, des caractéristiques, des valeurs, des désirs, des émotions, etc. On parlera de certaines relations en termes d'actions, de praxis, d'opérations, d'opposition, de dépendance, de proximité, d'association, de destination, d'orientation, d'intention, de comportements, de rituels, etc. On en appellera aussi à un métalangage théorique où des entités et des relations seront décrites comme des représentations, des désignations, des observations, de la signification, de l'interprétation, du sens, de la référence, des énonciations, des discours, des images, du culturel, du social, etc. Ces termes et les énoncés qui les contiennent varieront selon les paradigmes. Mais, quelles qu'en soient la liste et les formes, ces concepts et ces catégories relèvent tous, non pas du modèle formel et computationnel comme tel, mais bien d'un modèle conceptuel. Or c'est souvent ce modèle conceptuel qui est, comme je le disais précédemment, le travail central qui occupe une grande partie de l'énergie de la recherche sémiotique. Souvent transparent aux chercheurs eux-mêmes, il n'en demeure pas moins présent. C'est donc dans le modèle conceptuel qu'un modèle formel puise ultimement les référents et le sens de ses symboles et formules. En d'autres termes, c'est dans le modèle conceptuel que la sémantique des symboles et des formules est donnée. Encore une fois, sans lui, il n'y a pas de compréhension.

Voici un second exemple où la complexité de l'objet est plus évidente : la définition du signe. Encore une fois, l'objet d'étude doit être conceptualisé avant même d'être formalisé. Pour Saussure, le signe est conceptualisé *comme une relation entre le signifiant et le signifié*. Pour Peirce, il est *une structure relationnelle entre le representamen, l'objet, l'interprétant* et même *l'interprète*. Pour Frege, il est *une structure relationnelle entre un signe (zeichen) ou expression (ausdruck), un sens (sinn) et une référence (Bedeutung)*, notamment. Dans de multiples sciences cognitives, un signe (symbole) est conceptualisé comme un ensemble complexe de relations entre des facteurs *subjectifs, perceptuels, catégorisants, conscients, sociaux, culturels, contextuels*, etc.

Autrement dit, on conceptualise un artefact sémiotique comme un « objet » complexe structuré de plusieurs relations. Quel que soit le modèle formel que l'on pourrait projeter sur lui, sa sémantique référerà à ce qui est construit dans le modèle conceptuel. Ainsi avant même de penser à de la formalisation mathématique pour décrire ces relations structurantes, un grand travail de modélisation conceptuel est nécessaire.

Dans la perspective d'une sémiotique computationnelle, une telle modélisation conceptuelle est essentielle. Il est illusoire de croire que des outils informatiques peuvent remplacer ce travail. Un modèle conceptuel est la base à partir de laquelle les concepts et les énoncés peuvent exprimer de manière compréhensible aux humains les relations structurantes à l'œuvre dans la sémiotité des artefacts étudiés. Ce seront justement ces conceptualisations des relations dans lesquelles les modèles formels et computationnels puiseront pour leur propre construction. Plus encore, ces modélisations conceptuelles seront ultimement l'horizon qui permettra l'interprétation des résultats de la formalisation et de la computation, même si celles-ci découvrent des relations nouvelles, passées inaperçues au chercheur au début de sa recherche.

En somme, en sémiotique, même si la modélisation conceptuelle est imprécise, générale, subjective, biaisée, influencée, etc., et même si le langage naturel dans lequel elle s'exprime est ambigu, métaphorique, fictionnel, narratif, inconsistent, etc., elle n'en demeure pas moins un moment épistémique incontournable dans l'explication et la compréhension du sens et de la signification des artefacts signifiants qui sont ses objets d'étude. Elle offre un premier dévoilement des relations structurantes qui sous-tendent ces types d'artefacts. Ces relations et leur relata seront souvent identifiées comme étant des propriétés, des objets, des événements, des traits, des caractéristiques, des états. Ce sont ces relations auxquelles référeront, certes de manière complexe, les symboles des modèles formels et qui deviendront des variables, des constantes, des opérations, etc. Et parce que ces relations structurantes sont souvent très nombreuses, mais surtout complexes, il devient nécessaire de construire une modélisation formelle réglée, contrôlée et qui s'exprime dans un langage non ambigu. Ce modèle formel n'aura donc ultimement de signification qu'en relation à ces relations structurantes conceptualisées dans le modèle conceptuel. Un modèle formel ne se comprend qu'en relation à ce modèle conceptuel. Comme le dit très justement Pierre Duhem, donner à un physicien une équation, sans lui donner d'explication, il n'y comprendra rien³⁶.

Ainsi, quelle que soit la démarche scientifique effectuée, un modèle formel est contraint par un modèle conceptuel. Ce dernier est l'horizon épistémique qui présente ce qui est à modéliser formellement. Il est aussi l'horizon à partir duquel les résultats de la manipulation formelle seront interprétés. Dans une telle perspective, un modèle conceptuel est une contrainte majeure auquel un modèle formel est soumis. Il ne s'agit

cependant pas de la seule contrainte. En effet, dans le cas d'une relation possible avec l'informatique, ce ne sont pas toutes les relations structurantes identifiées dans la modélisation conceptuelle et retenues par une modélisation formelle qui pourront être intégrées dans une modélisation computationnelle. Un modèle conceptuel exerce donc une contrainte importante sur un modèle formel, mais ce ne sera pas la seule contrainte que rencontre la modélisation formelle. Le modèle formel est aussi contraint en aval par le modèle computationnel auquel il sera associé. Dit autrement, ce n'est pas parce qu'un modèle formel est créé qu'il est immédiatement admissible à la computation! En cela la différence entre la perspective de la sémiotique structurale et la sémiotique computationnelle est immense. Cependant, avant d'expliquer cette contrainte de la computationnalité sur la formalisation, rappelons plus précisément, mais brièvement, les définitions techniques d'un modèle formel.

4. Définitions techniques d'une modélisation formelle

En philosophie des sciences et en épistémologie, le modèle formel a été considéré, de façon classique, comme le modèle idéal et prototypique de la science : pas de modèle formel, pas de science rigoureuse. Comme le résume bien Teodor Shanin :

Les disciplines savantes incorporent les éléments de formalisation. [...] D'une part, cela reflète le besoin de l'esprit humain de rationaliser et de s'orienter dans une réalité immensément complexe, d'y parvenir en construisant des images générales systématiques et cohérentes et des schèmes symboliques³⁷.

Et dans l'horizon d'une sémiotique qui se veut computationnelle, ce type de modèle est une condition *sine qua non* d'une modélisation computationnelle et ultimement de l'utilisation de la technologie informatique. Il est donc important de comprendre ce qu'il en est de ces modèles formels dans le cadre d'une sémiotique computationnelle.

Dans cette perspective, un modèle formel est un artefact sémiotique qui se définit classiquement sur trois plans interreliés : (1) un plan syntaxique, (2) un plan sémantique et (3) un plan pragmatique.

4.1 Définition syntaxique d'un modèle formel

Au plan syntaxique, un modèle contient nécessairement une structure sémiotique ou des expressions formées à partir d'un ensemble de symboles dont la manipulation est

régie par un ensemble de règles. Selon l'expression de Jean Ladrière : « Un formalisme se présente comme un système de symboles soumis à des règles précises de manipulation³⁸. » On retrouve une même définition simplifiée chez les scientifiques comme David Marr : « Dire que quelque chose est un modèle formel signifie seulement qu'il s'agit d'un ensemble de symboles avec des règles pour les assembler – ni plus ni moins³⁹. »

Certaines disciplines universitaires appelleront un tel artefact sémiotique un *langage formel* ou un *système symbolique formel*. Dans ce cas, on dira que ce langage est constitué d'un ensemble de symboles⁴⁰ appelés le *vocabulaire* ou le *lexique* et d'un ensemble de règles définissant la *syntaxe* ou la *grammaire* de ce langage. Selon le langage formel choisi, il existera diverses manières de représenter la diagrammatique, c'est-à-dire la manipulation formelle des symboles et de leur composition, mais surtout de leurs transformations.

Un même système formel peut cependant être présenté dans les formes symboliques différentes et qui sont régies par des grammaires qui leur sont propres. On parlera alors de *formalisme* ou de *notation*. Par exemple, le langage formel qui exprime une logique des propositions pourrait être un formalisme russellien ou un formalisme polonais⁴¹. Les deux pourront exprimer les mêmes structures logiques et ils pourraient construire des preuves équivalentes. Et selon les types de notations, la diagrammatique représentera des séquences de manipulations des symboles : *composition*, *dérivation*, *génération*, *tableau*, *arbre* et même de *calcul*, *raisonnement*, etc. Ce sont ces séquences de manipulations réglées qui entreront dans une *démonstration*, une *preuve*, etc.

Sur le plan épistémique, tous les formalismes ou notations ne sont pas équivalents. Tous n'expriment pas nécessairement les mêmes dimensions des structures logiques ou mathématiques. Par exemple, ils n'expriment pas tous les mêmes types ou catégories des symboles, formules et règles. Surtout, ils n'offrent pas toute une même facilité de manipulation et de compréhension pour les humains.

Enfin, dans certains cas, on assignera des statuts particuliers à certaines formules de ce langage (formules, équations, etc.). Par exemple, elles pourraient être prises comme des axiomes, des postulats, des théorèmes ou des corollaires. Et si, en plus, toutes les formules d'un langage peuvent être construites par divers types de règles strictes de composition, d'inférence, de transformation, etc., on parlera alors d'une formalisation de type axiomatique. Le prototype classique d'un système axiomatique est celui de la logique des propositions, mais plusieurs autres systèmes formels existent quoique leur structure axiomatique ne soit pas toujours transparente à l'utilisateur.

On ne doit pas confondre un *langage* axiomatique avec un *modèle* mathématique. L'un n'est pas l'autre. Un langage axiomatique est une forme particulière de système formel alors qu'un modèle mathématique est un artefact épistémique qui utilise un ou

plusieurs types de langages formels capables d'exprimer certaines structures mathématiques. Autrement dit, une même structure mathématique pourrait être exprimée dans des langages formels différents. Par exemple, certaines structures mathématiques pourraient être exprimées dans un formalisme algébrique, un formalisme géométrique ou un formalisme topologique, etc. Cependant tous les langages formels ne se prêtent pas à l'expression d'une même structure mathématique. Un langage formel est souvent formé en regard de la sémantique qui lui sera associée et du rôle pragmatique qui lui sera assigné. Par exemple, bien qu'il y ait une même structure mathématique pour calculer la surface d'un jardin de forme carrée, il y aura des différences importantes entre les langages pour exprimer cette même structure mathématique. Ainsi, les langages de l'arithmétique, de l'algèbre, de la géométrie, de la topologie et les treillis n'ont pas tous les mêmes facilités de manipulations des symboles pour les humains. Pourtant, ces langages arriveront à des résultants *équivalents*.

4.2 Définition sémantique d'un modèle formel

Techniquement parlant, le langage formel ne possède aucune sémantique. Il est un système de symboles sans signification. Or, un modèle ne peut pas n'être qu'un langage formel, si bien formé soit-il. Comme tous les autres modèles construits dans une recherche, un modèle formel doit être mis en correspondance avec quelque chose d'autre. En termes classiques, un modèle doit « représenter » quelque chose. Pour le dire autrement, il faut ajouter une sémantique à un langage formel. Comme le rappellent Mauricio Suárez et Albert Solé : « Dans une représentation scientifique, une source A – habituellement un modèle, un graphe, une équation – est employée pour représenter une cible B – habituellement un système, une entité ou un phénomène⁴². »

La nature spécifique de la sémantique des langages formels ne va évidemment pas sans générer des débats. Les Peirce, Frege, Russell, Husserl, Carnap, Tarski, Kripke et Bourbaki de la communauté des mathématiciens, par exemple, ne définissent pas la sémantique des langages formels de la même manière. Cependant, malgré des différences majeures entre ces sémanticiens du formel, il en ressort un point commun. Quel que soit le langage formel utilisé, une sémantique doit déterminer une référence : elle doit mettre le langage formel en correspondance avec autre chose. De fait, la thèse sémantique dominante dans ces débats sera de type référentiel. On ne nie pas le problème du « sens », mais on donne priorité à la référence. Dans cette perspective, les divers types d'expressions construites dans ce langage formel, comme les formules, les équations, les

graphes, etc., ne sont vrais ou faux que relativement à cette sémantique. On dit alors qu'un tel langage formel est *interprété*.

La question qui est alors soulevée par cette thèse est la suivante : à quoi alors réfère le langage formel le plus communément utilisé dans les modèles scientifiques, à savoir le langage mathématique ? Une des réponses classiques à cette question soutiendra que les mathématiques *réfèrent*⁴³ à des nombres. Ce que l'on interprétera souvent de manière rapide en identifiant une sémantique mathématique à une sémantique de la quantité. Or, la majorité des mathématiciens et mathématiciennes refuseront de réduire la sémantique des langages mathématiques à une référence à des nombres⁴⁴. Ce sera d'ailleurs l'une des grandes contributions de l'école mathématique de Bourbaki de montrer que la sémantique des langages mathématiques ne se réduit pas à l'univers des nombres. Au contraire, pour Bourbaki, l'essentiel de la mathématique est avant tout de porter sur des « structures », c'est-à-dire sur la diversité de relations qui peuvent exister entre des éléments, peu importe le monde dans lequel ils peuvent être instanciés. L'univers de nombre n'étant qu'un possible parmi d'autres.

Inspiré par les analyses de Michael S. Mahoney en histoire des mathématiques, je dirais aussi que les mathématiques ne portent pas sur le monde comme tel, mais sur la manière dont on se représente le monde et plus particulièrement sur la manière dont on se représente les relations que nous construisons à propos de ce monde, quel qu'il soit⁴⁵. Autrement dit, ce dont parlent les mathématiques (leur sémantique) est à propos de certains types particuliers de structures de relations entre des éléments, ce que Bourbaki appelle des *structures mathématiques*⁴⁶.

4.3 Définition pragmatique d'un modèle formel

La troisième dimension d'une modélisation formelle est de nature pragmatique. Celle-ci renvoie au rôle épistémique d'une modélisation formelle dans une démarche scientifique.

Une sémantique, si spécifique soit-elle, n'émerge pas du langage lui-même. Un langage formel ne génère pas sa sémantique de lui-même⁴⁷. Celle-ci vient plutôt des créateurs du modèle qui utilisent ce langage dans une modélisation. Ainsi, le choix d'une sémantique dépendra en partie des intentions épistémiques des créateurs du modèle⁴⁸ lorsqu'ils construisent un langage formel pour représenter leur objet d'étude.

À titre d'exemple, prenons un objet physique comme le pendule, que des chercheurs voudraient modéliser formellement. Pour ce faire, ils auront à choisir une modélisation spécifique parmi un grand nombre de modèles formels possibles. L'un d'eux pourrait servir à formaliser uniquement la *période de temps* qu'un pendule prend pour une os-

cillation. On pourrait ainsi vouloir modéliser *la friction* du mécanisme de suspension ou encore la *résistance du matériau* de la tige relativement à la masse. Et, pour chacun de ces « angles de vues » du pendule, une formalisation exprimant une structure mathématique spécifique pourrait être utilisée. Et pour chacune de ces structures mathématiques, plusieurs langages mathématiques sont disponibles. L'un d'eux pourrait être algébrique, l'autre topologique, etc.

Sur le plan pragmatique, ce modèle formel a un rôle de médiateur épistémique⁴⁹. Il sert à « représenter » *sous un angle ou un autre* son objet d'étude. Dans un contexte scientifique, il faut cependant bien comprendre que le terme « représentation » n'a pas la signification métaphorique que le discours ordinaire lui donne. En effet, un modèle formel ne « représente » pas le réel comme un miroir de la réalité. Il ne reflète pas la réalité. Il le rend simplement « présent » au chercheur sous une autre forme.

Un modèle formel est ainsi un outil épistémique, c'est-à-dire une construction cognitive qui permet à des chercheurs de mettre en correspondance des équations ou des formules d'un langage formel avec l'objet d'étude. De fait, le modèle *projette* des structures mathématiques sur ses objets d'étude. Pour parler encore métaphoriquement, un modèle est un filtre appliqué à l'objet d'étude et non un miroir réfléchissant. Il projette ainsi des schémas structurants sur l'objet d'étude. Comme l'affirme Giorgio Israel : « les schémas mathématiques unifient des réalités différentes, mais isomorphes⁵⁰. »

Cette projection elle-même peut être traduite formellement comme *une application (mapping)* fonctionnelle qui va d'une structure mathématique (SM) particulière à une structure spécifique de l'objet (SO) (fig. 1).

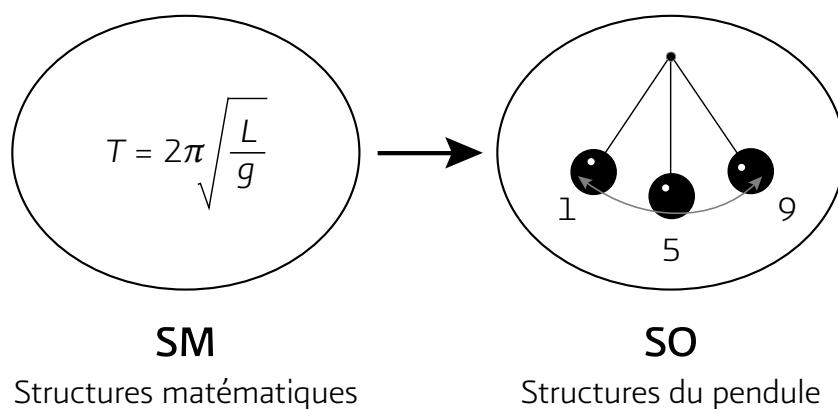


Figure 1. La modélisation formelle du pendule comme projection.

Cette formalisation tant des objets d'étude que de la relation de modélisation elle-même devient des plus intéressantes en science. Lorsque les objets d'étude sont complexes et dynamiques, elle permet d'unifier dans un langage formel commun une multitude de propriétés structurales de l'objet, mais aussi de comparer les modèles eux-mêmes. Ce type de formalisation ouvre alors à des analyses « différentielles » et « distributionnelles ». Elle fait ainsi apparaître de nouveaux « observables », c'est-à-dire qu'elle révèle de nouvelles propriétés de l'objet par la manipulation des symboles qui « représentent » ces propriétés dans le modèle formel lui-même.

La définition générale des modèles formels que j'ai présentée sous les angles syntaxique, sémantique et pragmatique est relativement classique. Je souhaite cependant souligner un point important de la sémantique : celle suggérée est d'inspiration bourbakienne. Elle lie directement le langage mathématique à des structures mathématiques. Or, cette sémantique est une des voies par lesquelles il devient possible de lier les langages formels mathématiques à la modélisation computationnelle pertinente à une sémiotique computationnelle. Dans cette définition une modélisation formelle ne peut être réduite à n'être qu'une modélisation quantitative.

Ainsi, si l'un des intérêts de la sémiotique computationnelle est d'étudier les propriétés signifiantes des artefacts porteurs de signification, il lui faut les approcher par une modélisation formelle dont la sémantique doit référer à des structures caractérisant ces artefacts, que celles-ci soient discontinues ou continues, stables ou dynamiques. Ce sont elles qu'un modèle formel projettera comme un filtre. Mais, comme nous le verrons dans ce qui suit, dans le cadre de l'intégration de l'informatique, ce modèle sera soumis à une autre contrainte : celle de la computation.

5. Modélisation formelle et computation

Il existe évidemment un grand nombre de structures de dépendances fonctionnelles qui peuvent être exprimées dans un langage formel. Cependant, dans un contexte informatique, il en est une sorte particulière qui est essentielle et nécessaire pour construire une modélisation computationnelle, à savoir celle qui, précisément, possède une structure de relations dites de type fonctionnel. En fait, ces types de relations forment un sous-ensemble de toutes les relations possibles entre des éléments. Elles sont définies techniquement de la manière suivante : une relation est dite fonctionnelle, si les éléments x_i d'un ensemble de départ X déterminent ou sont associés à un seul et unique élément y_i dans l'ensemble d'arrivée Y (fig. 2).

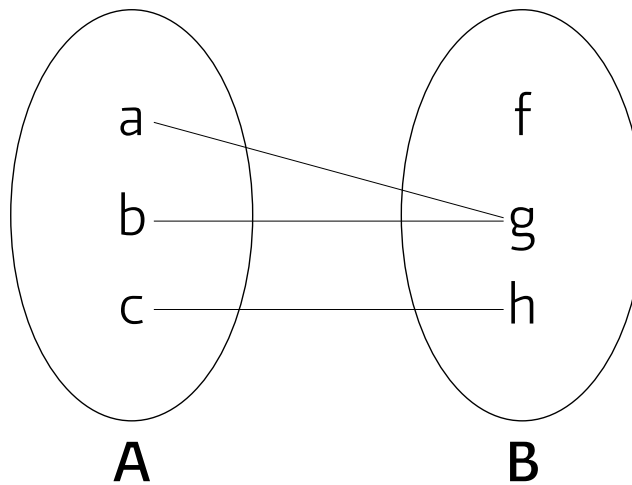


Figure 2. Graphe d'une relation fonctionnelle entre deux ensembles.

Autrement dit, une relation est dite fonctionnelle (ou fonction) à partir du moment où un élément de l'ensemble de départ ne peut pas avoir une relation avec plusieurs éléments de l'ensemble d'arrivée.

Pour illustrer ce type particulier de relations, nous prendrons deux exemples, l'un est arithmétique et l'autre non arithmétique.

L'exemple arithmétique est la relation fonctionnelle appelée le « carré d'un nombre ». Cette fonction détermine que pour tout nombre X d'un premier ensemble, il ne peut y avoir qu'un nombre Y dans l'ensemble d'arrivée. On peut représenter cette fonction de plusieurs manières. En algèbre, on écrira l'équation suivante :

$$y = x^2 \text{ ou } F(x) = x \cdot x$$

Cette manière de présenter une fonction est dite « intensionnelle ». Elle pourrait également être écrite dans le formalisme de la théorie des ensembles sous forme d'une liste de doublets :

$$F(x, y) : \{(1, 1), (2, 4), (3, 9) \dots (n, n)\}$$

Cette manière de présenter la même fonction sous la forme d'une liste est dite « extensionnelle ».

Notre deuxième exemple est non arithmétique. Il montre que les relations fonctionnelles n'existent pas uniquement dans l'univers des nombres. Ainsi, les sémanticiens considéreront que l'expression linguistique « Fils de » exprime une relation fonctionnelle entre deux personnes. En effet, cette relation détermine que, pour tout X qui est le fils de Y , il n'y a qu'un seul Y auquel X est lié. La langue française appellera par exemple ce Y « père ». Tout comme dans le cas précédent, il faut une procédure effective quelconque pour déterminer quelle est effectivement la personne Y qui est associée à la relation « X est le fils de ». Comme dans le premier exemple, cette relation fonctionnelle peut être exprimée « intensionnellement » ou « extensionnellement ».

Intensionnellement, elle sera formulée dans un langage logique prédicatif comme $\forall x \forall y (\text{FILS-de } [x_i, y_i])$. En langage naturel, cette expression logique sera traduite par la phrase : *une personne est le fils d'une autre personne*. Extensionnellement, elle sera formulée par une liste de doublets : $\{(\text{Jean Tremblay}, \text{Pierre Tremblay}), (\text{Louis Beaupré}, \text{Pierre Beaupré}) \dots (N_1, N_2)\}$.

Ces types de relations fonctionnelles se retrouvent régulièrement exprimées dans les langages mathématiques. En fait, elles forment un sous-ensemble particulier des structures mathématiques dont je parlais plus haut et que les langages mathématiques ont pour fonction d'exprimer. En contexte informatique, ces fonctions ont un rôle encore plus important : elles deviennent l'une des conditions *sine qua non* de leur computabilité. Voyons pourquoi.

Reprenons notre exemple arithmétique de la fonction « mise au carré ». Pour trouver la valeur de cette fonction « $F(y) = x \cdot x$ », il faut effectuer une opération que l'on appelle de manière générale un *calcul*. En l'occurrence, cette opération consiste à *multiplier une valeur de x_i par lui-même*. Or cette opération de « multiplication de x_i par lui-même » n'est pas simple : elle exige une manipulation sophistiquée de symboles⁵¹. On sait combien ces opérations exigent des efforts d'apprentissage chez les jeunes enfants.

Ces exemples montrent cependant quelque chose de très important : même si une expression formelle est bien formée, il ne s'ensuit pas que l'on connaît ou même qu'il existe une « procédure effective » pour trouver pour tous les éléments d'un ensemble quel est l'autre élément auquel il est associé. Dit autrement, il ne faut pas confondre l'expression d'une relation fonctionnelle avec les procédures de calcul effectif. Ce n'est pas parce qu'on a une équation formelle exprimant une relation qu'il existe une procédure effective de « calcul » qui donnera pour chaque valeur x_i une valeur y_i . Plus techniquement, on dit que si pour des relations fonctionnelles données il existe des « procédures effectives » pour trouver pour toute valeur de x une certaine valeur de y , alors cette fonction sera dite *calculable* (en anglais : *computable*).

Existe-t-il toujours de telles procédures de calcul? Est-ce que toutes les fonctions construites en mathématiques sont calculables? Ce sont d'importantes questions posées par David Hilbert au début du XIX^e siècle à propos de ce « calcul »⁵². Comment peut-on prouver qu'il existe ou n'existe pas une telle procédure pour toutes les relations fonctionnelles mathématiques? Autrement dit : qu'est-ce qu'une fonction calculable? Ou encore : qu'est-ce qui garantit qu'une fonction est calculable?

Deux voies de réponses ont balisé la compréhension de cette notion de calculabilité des fonctions et ont permis de la relier à la notion qui lui sera associée. Une première creusa la notion de calculabilité elle-même, l'autre explora la notion de structures mathématiques. Les deux démontrèrent des convergences théoriques importantes sur l'équivalence entre certaines formes de calculabilité et certaines structures mathématiques. Cette équivalence sera importante pour comprendre les liens de complémentarité sinon même d'équivalence entre certains modèles computationnels et certains modèles formels. J'y insisterai donc un peu.

La première voie qui explora ce lien de la calculabilité des fonctions avec la computationnalité commença avec Alan Turing en 1936⁵³. Une de ses grandes contributions fut, comme le souligna Gödel, de montrer que ces notions de décidabilité et de calculabilité pouvaient être comprises comme un ensemble de *procédures effectives* de manipulation de symboles représentant une fonction mathématique (appelée par Turing une *machine abstraite*) et qui seraient réalisables par un type particulier de *machine physique* (appelée plus tard « Machine de Turing »). On dira alors que si la fonction était « calculable », elle était aussi « computable » par cette machine.

Plus tard, une avancée théorique importante, appelée « thèse Church-Turing », affirma que les fonctions computables par une machine de Turing étaient elles aussi computables par une autre forme de calcul : le lambda-calcul de Church⁵⁴. Cette thèse fut ultérieurement généralisée à plusieurs autres formes de computation, par exemple : la logique combinatoire⁵⁵, les algorithmes⁵⁶, les règles de production⁵⁷ et bien d'autres. Qui plus est Robin Gandy démontra que la computation pouvait aussi être parallèle et dynamique⁵⁸.

Il suivra de toutes ces recherches la généralisation suivante : il y a des équivalences entre des relations fonctionnelles calculables et diverses formes de computation comme les algorithmes, les règles de production, les automates, certaines grammaires de Chomsky, etc.

La deuxième voie de recherche pour lier calculabilité et computationnalité sera parallèle à cette dernière et en reprendra certains jalons. Un peu plus complexe dans son argumentation, elle fera un lien entre les structures mathématiques et certaines formes de computation. Plusieurs mathématiciens, par des arguments formels différents et

sophistiqués, montreront que certaines de ces structures mathématiques, plus particulièrement les structures algébriques, lorsque comprises dans leur formulation d'algèbre abstraite, peuvent être mises en lien avec des fonctions computables. Par exemple, des formules algébriques construites par diverses sortes de *combinaisons ou transformations ou compositions* réglées sont vues comme des groupes, des sous-groupes, des monoïdes, des anneaux, etc. Von Newman montra que ces structures abstraites pouvaient être mises en lien avec les théories combinatoires et le lambda-calcul de Church. Enfin, il relia ces structures aux *automates* qu'il avait lui-même développés : « La théorie des automates [...] devra être, d'un point de vue mathématique, combinatoire plutôt qu'analytique⁵⁹. »

Chomsky relia à son tour ces automates aux structures d'un langage⁶⁰. Sa théorie montra qu'il était possible de faire une équivalence entre la théorie des « automates » finis de Von Neuman et des types de grammaires génératives⁶¹. De cette manière, certaines syntaxes formelles devenaient équivalentes à des automates finis déterministes : « La grammaire d'une langue peut être considérée comme une théorie de la structure de cette langue⁶². » Cette thèse chomskienne sera reprise par Marcel-Paul Schützenberger qui, lui, la remettra en lien avec la théorie des algèbres abstraites⁶³. Il montrera que certaines de ces structures pouvaient exprimer abstraitement la nature mathématique de la formation de séquences de symboles dans des langages. Il y avait donc un lien profond entre des automates, certains types de grammaires structurales et des structures mathématiques abstraites⁶⁴. Daniel Lacombe pour sa part montrera que des fonctions sur du continu étaient aussi computables⁶⁵.

Il manquait cependant un pont avec l'informatique classique issue de Turing et Church. Ce lien fut développé par John McCarthy⁶⁶. Vers les années 1960, celui-ci créa le célèbre langage de programmation Lisp qui lui-même avait un lien profond avec le lambda-calcul de Church. McCarthy montra alors qu'on pouvait comprendre des automates et des algorithmes comme des opérateurs de transformation d'expressions symboliques tels qu'ils sont construits par des grammaires formelles de Chomsky. Finalement, Christopher Strachey et d'autres proposèrent de se servir de ces types de structures comme fondement mathématique des langages de programmation⁶⁷.

Un dernier sillon de cette recherche rajouta un point intéressant. Il toucha une dimension importante de la nature des fonctions, à savoir leurs arguments, c'est-à-dire les symboles qui représentent ce sur quoi elles sont appliquées. Dans le langage informatique classique, ces arguments des fonctions sont nommés des intrants ou des données (*data*). Or, des fonctions peuvent être définies de façon extensionnelle, c'est-à-dire, soit par la présentation descriptive des éléments qui constituent une relation fonctionnelle particulière ou soit plus simplement sous la forme de liste. Dans un cadre informatique,

si la présentation extensionnelle est complexe, on imposera une structure aux données (*data structure*). C'est souvent sous cette forme qu'elle sera concrètement implémentée dans une base de données : « La programmation porte essentiellement sur certaines "structures de données" et les fonctions entre elles⁶⁸. »

Ainsi, cette dernière voie de recherche démontra, par un cheminement inverse, des équivalences entre des fonctions calculables du lambda-calcul, des algorithmes, des automates, des grammaires catégorielles, des langages de programmation, des structures de données et certains de type de structures mathématiques (algèbre, topologie, géométrie, treillis, etc.⁶⁹). Autrement dit, ce que l'un peut calculer, l'autre peut aussi le faire. On trouve cette idée chez Rodney Burstall et Peter J. Landin :

Une programmation porte essentiellement sur certaines « structures de données » [*data structures*] et sur des fonctions entre ces structures. [...] Ayant commencé avec des structures familières telles les groupes et les anneaux, les algébristes ont développé une notion plus approfondie des structures algébriques (ou algèbre) qui inclut celles-ci comme exemples, mais qui inclut aussi plusieurs des entités qui dans un ordinateur sont pensées comme des structures de données⁷⁰.

Richard Büchi écrit également :

Si la définition d'un « automate fini » est correctement choisie, il s'avère que tous les concepts de base et les résultats concernant la structure et le comportement des automates finis sont en fait seulement des cas spéciaux des concepts fondamentaux (homomorphisme, relation de congruence, algèbre libre) et des faits d'algèbre abstraite. La théorie des automates est simplement la théorie de l'algèbre universelle [...] avec des opérations unaires, et une mise en évidence sur les algèbres de type/groupe fini⁷¹.

Il faudra cependant attendre Dana A. Scott pour expliciter clairement les fondements mathématiques de la nature des fonctions computables. Dans un article de 1970, il soulignera la faiblesse des théories générales des fonctions : « La principale nouveauté mathématique de la présente étude est la création d'une théorie mathématique des fonctions adéquates qui accomplit ces objectifs (systématiquement!)⁷² [...] »

Malgré la richesse de ces liens, de nombreux problèmes persistent. Autant la théorie des algèbres abstraites que la théorie classique de la computation ne semblent pas les aborder de manière satisfaisante. On pense ici à des problèmes d'approximation, d'optimisation, de données massives, de la complexité, de la dynamique et surtout de non-compositionnalité de plusieurs fonctions. Il faudra attendre les développements récents des mathématiques pour que ces dimensions particulières soient reliées à des structures mathématiques.

Ainsi, pour qu'un modèle formel soit possiblement mis en relation avec un modèle computationnel, il faut que le langage formel utilisé exprime des structures mathématiques fonctionnelles. Mais encore faut-il que ces structures permettent la calculabilité. Cependant, si elles possèdent cette propriété, alors comme l'a montré le précédent résumé de la recherche en informatique théorique, il existe une équivalence entre certaines structures mathématiques calculables et de multiples langages exprimant des fonctions. Sachant cette équivalence, il sera alors possible pour la recherche scientifique de traduire leurs modélisations formelles directement en une modélisation computationnelle. Aussi verra-t-on souvent des modèles algébriques de grandes complexités exprimer directement les fonctions qu'ils contiennent en algorithmes ou en un langage de programmation adéquat. Et la raison en serait que l'on connaît maintenant mieux les fondements mathématiques de ces approches computationnelles. Comme l'écrit Mahoney :

En l'absence de structures mathématiques qui autorisent l'abstraction et la généralisation, les modèles computationnels ne disent pas grand-chose. Pas plus qu'ils ne fonctionnent comme les modèles traditionnels qui ont fourni une compréhension de la nature sur la base de ce que nous pouvons tester nos connaissances en faisant avancer les choses dans le monde⁷³.

En somme, on voit mieux comment un modèle formel est contraint en aval par la computation. Pour que les équations ou formules qu'utilise un modèle formel soient pertinentes pour une modélisation computationnelle, il faut que les relations structurantes possèdent des relations de dépendances fonctionnelles. Cette dépendance fonctionnelle est une condition *sine qua non*, de la traduction computationnelle, c'est-à-dire en algorithme ou autres « procédures effectives ».

Comme nous le verrons dans la section suivante, ces recherches auront une portée capitale pour une sémiotique computationnelle.

6. Critiques et limites de la modélisation formelle dans une sémiotique computationnelle

La sémiotique computationnelle fera nécessairement l'objet de différentes critiques. Par la mobilisation de l'informatique dans sa démarche, elle se positionne dans une épistémè positiviste, empirique et possiblement expérimentale. Cela n'ira pas sans soulever les critiques classiques élaborées par rapport à ce type d'épistémè. Entre autres critiques, on pourrait voir la sémiotique computationnelle comme (a) substituant le quantitatif au qualitatif, (b) éliminant la recherche analytique et conceptuelle et (c) réduisant ou

simplifiant trop la complexité intrinsèque des artefacts sémiotiques. Je présenterai ici ces possibles critiques en mes termes, tout en visant les points spécifiquement reliés aux modélisations formelle et computationnelle.

La première critique, générale, pouvant être formulée soutiendrait qu'une sémiotique computationnelle est simplificatrice et réductrice parce qu'elle substitue une approche quantitative à une approche qualitative et conceptuelle. Or la réalité de la pratique sémiotique n'est pas si simple. Une telle critique, en effet, relève d'une compréhension discutable de la nature et de la place d'un modèle formel dans une démarche de connaissance scientifique. Telle que formulée, cette critique reprend de manière implicite l'opposition classique de la philosophie des sciences de tradition allemande où l'on oppose les sciences de la nature et les sciences de l'esprit (*Geistwissenschaften*) ou de la tradition anglo-saxonne où on oppose l'empirisme/le positiviste et l'herméneutique. Ces deux oppositions sont vulgarisées en termes de sciences dites quantitatives *versus* qualitatives ou plus facilement encore de sciences dures et de sciences molles.

Mais comme le montre la philosophie contemporaine, cette vision ne correspond pas à la pratique concrète des sciences de quelque nature qu'elles soient. En effet les sciences sont des pratiques épistémiques qui ne peuvent être comprises comme opérant dans une seule et unique de ces positions épistémologiques. Au contraire une science ne peut approcher son objet d'étude que par le biais d'une multitude de points de vue ou, dit plus techniquement, de modèles interreliés. Aucune science dans sa démarche ne repose sur un modèle unique. Comme le formule Sara Green : « un seul modèle est insuffisant⁷⁴. »

Il en va de même en sémiotique : elle ne peut pas reposer que sur un seul et unique modèle. Une étude un peu attentive de ses pratiques montre rapidement que la sémiotique met en place de multiples modèles dont certains peuvent être vus comme modélisant du « qualitatif » et d'autres du « quantitatif », et où chacune des approches ne s'y investit pas avec la même énergie épistémique.

De fait, en raison de la complexité de ses objets d'étude, la sémiotique se concentrera classiquement sur un type de modèle plutôt qu'en autre. On conviendra facilement que celui qu'elle privilégie est le modèle « qualitatif » de type conceptuel. Celui-ci met en action des procédures épistémiques de type analytique et synthétique, des descriptions et des explications qui en appelleront à de nombreux paradigmes théoriques ; le but de ce modèle étant de construire une première conceptualisation des multiples relations qui structurent un artefact sémiotique et des interprétations qui peuvent lui être associées pour finalement les exprimer dans un langage naturel. Mais, on ne s'arrêtera pas là. Très souvent, pour confirmer la validité des énoncés produits dans ce modèle, on en appellera à diverses méthodes de validation. Par exemple, par déduction on validera les

hypothèses analytiques et interprétatives en mettant en correspondance des cas, des situations, concrètes, etc. Ceux-ci ne sont pas que des « illustrations », et ils serviront d'échantillons de validation empirique. Ou encore, l'hypothèse sera considérée vraie jusqu'à la rencontre d'un contre-exemple ou jusqu'à preuve du contraire. Dans certaines autres recherches, on procédera par induction et abduction. On partira des exemples ou des cas et on tentera par généralisation de trouver des « régularités » tant dans les structures des artefacts que dans leurs interprétations.

Or, il faut bien comprendre que ces méthodes de validation, en apparence simples et uniquement qualitatives, cachent habilement du quantitatif. En effet, les syntagmes comme « des cas », « un contre-exemple » ou « quelques situations » contiennent des marqueurs linguistiques (des articles déterminatifs) qui expriment linguistiquement de la quantification. Ces marqueurs signifient qu'il y a un certain *nombre* (bien que souvent limité) de situations ou de cas qui valident l'hypothèse. Bien que la méthode de validation ne soit pas exprimée formellement et ne mette pas en jeu des « expérimentations », il a y techniquement appel à du quantitatif, si minimal soit-il.

Par conséquent, il est trop général de soutenir qu'une modélisation computationnelle est réductrice parce qu'elle substitue une approche quantitative à une approche qualitative. Aucune personne en sémiotique ne dira que ses énoncés à propos des artefacts sémiotiques sont vrais uniquement parce que c'est elle qui les a déduits « analytiquement » de prémisses, de postulats ou d'axiomes ou encore que ces énoncés sont vrais par ce que c'est elle qui en a fait la synthèse et les généralisations. Aussi en appellera-t-on plutôt à des validations externes qui souvent seront « des » exemples ou « des » contre-exemples. Malheureusement ce type de formulation masque l'utilisation du quantitatif à travers le flou du langage naturel.

Par ailleurs, et ceci est plus important, il faut se rappeler que dans une perspective boubakienne, la sémantique d'un modèle formel de type mathématique renvoie avant tout à des structures dans l'univers de référence. Or ces structures dites mathématiques sont avant tout des relations de divers types entre des éléments qui dans certains cas peuvent être certes, des nombres, mais qui dans plusieurs autres cas peuvent être tout autre chose. Une belle illustration de cette formalisation autre que « quantitative » est la modélisation formelle de type grammaire catégorielle que Lambek et Lambek ont appliqué à des relations de parenté⁷⁵. Une grammaire générative (à la Chomsky) ou une grammaire fonctionnelle (à la Halliday) est une hypothèse de type mathématique non quantitative sur des relations structurelles dans la langue. Et on pourrait montrer que de nombreuses analyses de types géométriques utilisent des chiffres comme des artifices intermédiaires pour assister des calculs sur des vecteurs qui représentent des relations dans des phénomènes sémiotiques complexes. L'exemple donné plus haut sur la relation

Fils de montre qu'une structure mathématique de relation de dépendance fonctionnelle peut être projetée sur des personnes existant dans le monde réel.

La critique qui soutient qu'une sémiotique computationnelle qui en appelle à des modèles formels de type mathématique privilégie une approche quantitative aux dépens d'une approche qualitative ne correspond pas parfaitement aux pratiques effectives en sémiotique ou encore à la nature exacte d'un modèle formel. D'une part, la sémiotique utilise souvent de manière discrète sinon même implicite de la validation à teneur quantitative même si cela est minimal. D'autre part, une telle critique ne semble pas avoir compris que la modélisation formelle de type mathématique peut porter sur autre chose que des nombres tel que par exemple des structures de relation entre des éléments de toute sorte.

Une deuxième critique pourrait être élaborée. Plus complexe et problématique, elle soutiendrait que la sémiotique computationnelle élimine la modélisation analytique et conceptuelle. Cette critique me semble problématique, car elle ne correspond pas à une contrainte inhérente à la construction même de modèle formel. En effet, comme je l'ai montré plus haut, dans une démarche scientifique un modèle formel est contraint par ses relations aux autres modèles. Il ne se construit pas de manière indépendante et isolée des autres modèles qui participent à la connaissance. Par exemple, un modèle formel dépend en amont et en aval d'un modèle conceptuel sur lequel il est appliqué : dit autrement, un modèle formel ne peut construire des énoncés formels ni les interpréter qu'en relation à ce qui est conceptualisé à propos de l'objet d'étude. Un modèle formel est donc intimement contraint par un modèle conceptuel. Sans cette sémantique conceptuelle, un modèle formel est sans signification et, de ce fait, incompréhensible.

Il s'agit d'une illusion créée par un habile marketing de la science ou de vulgarisations simplificatrices que de penser qu'on puisse prendre des formules mathématiques clé en main et les appliquer à n'importe quel objet. La loi de l'attraction des masses, quelle qu'en soit la formalisation, ne peut expliquer l'attraction d'une personne pour une autre ! L'équation $F = Ma$ a des référents très spécifiques qui, dans le modèle conceptuel, sont nommés *masse* et *accélération* et *force*. Ce sont là des concepts qui reposent sur une conceptualisation technique et complexe et qui n'est pas simple et facile à comprendre. Ces concepts n'ont rien à faire avec les concepts qui décrivent une attraction amoureuse entre des personnes.

Compte tenu de la complexité des objets d'étude de son domaine d'étude, une sémiotique dite computationnelle ne peut éliminer le travail analytique et interprétatif qui doit se construire dans un modèle conceptuel. Au contraire, il met une pression sur ce modèle dont la fonction est d'explorer et de préciser au meilleur de ses moyens les concepts et la conceptualisation de l'objet d'étude afin de le rendre accessible à un

éventuel modèle formel. C'est en effet sur le contenu du modèle conceptuel que le modèle formel projette des hypothèses structurantes. Même dans les cas où un modèle formel est offert *a priori*, ce sera le modèle conceptuel qui permettra d'interpréter les équations et les formules que propose une hypothèse formelle.

En somme, la modélisation formelle en sémiotique ne peut éliminer la modélisation conceptuelle. Aucune modélisation sémiotique computationnelle n'est possible si le modèle formel qui sous-tend les algorithmes n'est pas relié en amont, en parallèle ou en aval, à un travail épistémique conceptuel et interprétatif.

Une troisième critique plus profonde pourrait être adressée. Elle limiterait absolument une sémiotique computationnelle, mais du même coup en préciserait absolument les cibles. Elle soutiendrait que la modélisation formelle à horizon computationnel ne peut rendre compte de la complexité du sémiotique. De ce point de vue, une sémiotique computationnelle serait donc réductrice.

Un modèle formel qui, en aval, doit être éventuellement intégré dans une technologie informatique dépend effectivement des contraintes que lui impose un modèle computationnel. En effet, un modèle formel à horizon computationnel ne peut se concentrer que sur des relations structurantes de types dépendances fonctionnelles et qui sont calculables. Car c'est le seul type de structures qui peuvent être traduites dans des algorithmes. Autrement dit, une équation ou une formule d'un langage formel est traduisible en algorithmes si et seulement si elle est une fonction qui peut être calculée par des procédures effectives computables. Dit banalement, une fonction non calculable ne peut être computable! En conséquence, un modèle formel sémiotique si bien formé soit-il, si élégant soit-il, est inutile si la grammaire ou la manipulation des symboles exprimant des structures fonctionnelles ne permet pas des procédures effectives, c'est-à-dire du calcul ou de la computation. Ceci limite radicalement la construction de modèle formel pour une sémiotique computationnelle. Qui plus est, rien ne garantit que dans un artefact sémiotique seules les structures de relations de dépendances fonctionnelles soient les seules pertinentes ou intéressantes.

En d'autres termes, un modèle formel couplé à un modèle computationnel est limité. Il impose une vision spécifique de l'objet d'étude. Par sa nature même, un tel modèle est une sorte d'idéalisation⁷⁶. Il impose des choix. En conséquence, il est toujours réducteur. Il n'a jamais le type de richesse épistémique que possède un modèle conceptuel. Par exemple, un modèle formel mathématique de type *logique* qui contient des symboles atomiques et des règles spécifiques d'inférence exigera une atomisation et une discrétisation des éléments. Il ne se focalisera que sur des structures statiques consolidées. Il lui sera difficile, de par son formalisme même⁷⁷, de traiter le continu et la dynamique de certains phénomènes⁷⁸.

Un modèle formel mathématique sera encore des plus réducteurs. En effet comme celui-ci ne peut cibler que des structures relationnelles de dépendances fonctionnelles computables, il risque de rencontrer le mur de la non-computationnalité. Car, comme on le sait par Godel, tout système formel présente la non-complétude⁷⁹. Et on sait par Turing qu'il n'est pas possible de déterminer quand un algorithme s'arrêtera. Par Chaitin, on sait que plus un système est complexe, plus il tend vers le non-computable⁸⁰. En conséquence, un modèle formel si bien formé soit-il, peut s'avérer non traduisible en algorithmes parce que non computable. Voici un exemple presque contre-intuitif : une équation simple, élégante et ayant la forme suivante : $ax + bx = c$ (dite équation diophantienne) a été démontrée non computable⁸¹ ! Et il existe le nombre infini de de telles fonctions non calculables. Une sémiotique computationnelle est donc absolument limitée. Elle ne peut pas traiter cette infinité de fonctions ! Comme son nom l'indique : elle ne traite que des fonctions « computationnelles ».

Il s'avère cependant qu'il reste encore une infinité de fonctions calculables. Assez pour occuper aussi une infinité de recherches en sémiotique. La critique qui soutiendrait que la sémiotique computationnelle est limitée et réductrice est incontournable, mais on peut la voir positivement : elle définit précisément la cible. Une sémiotique computationnelle, au sens strict, ne peut avoir comme objets d'étude que des artefacts sémiotiques présentant des structures relationnelles formalisables et computables. Et comme il est pensable que plusieurs des relations qui tissent les artefacts sémiotiques ne puissent être formalisables et computables, il s'ensuit qu'une sémiotique computationnelle ne pourra jamais être totalement automatique et algorithmique.

En conséquence l'informatique devra accepter de n'être qu'un assistant dans la recherche sémiotique. La sémiotique computationnelle devra faire preuve d'humilité et accepter de ne travailler que sur ce qui est formalisable et computable. Il n'en reste pas moins que ce domaine de recherche demeure vaste. Les deux disciplines pourront naviguer dans deux infinis : l'infini du computable et l'infini du non-computable. Là est l'un des paradoxes de leur rencontre.

Bibliographie

- ALBUS, James S. & Alexander M. MEYSTEEL, « A Reference Model Architecture for Design and Implementation of Intelligent Control in Large and Complex Systems », *International Journal of Intelligent Control and Systems*, vol. 1, no 1, 1996, p. 15-30.
- ARMATTE, Michel & Amy DAHAN DALMEDICO, « Modèles et modélisations, 1950-2000 : nouvelles pratiques, nouveaux enjeux », *Revue d'histoire des sciences*, tome 57, no 2, 2004, p. 243-303.
- ASPERTI, Andrea & Giuseppe LONGO, *Categories, Types and Structures*, Cambridge, The MIT Press, 1991.
- AUDIN, Michèle, « Hommage à Claude Lévi-Strauss », *Images des mathématiques*, 2009. En ligne : <<https://images.math.cnrs.fr/Hommage-a-Claude-Levi-Strauss.html>>.
- BARBUT, Marc, « Sur le sens du mot "structure" en mathématiques », *Les Temps Modernes*, no 246, 1966, p. 791-814.
- BARTHES, Roland, *Système de la mode*, Paris, Seuil, 1967.
- BERNARD, Michel & Baptiste BOHET, *Littérométrie : outils numériques pour l'analyse des textes littéraires*, Paris, Presses Sorbonne nouvelle, 2017.
- BOUDON, Raymond, *À quoi sert la notion de « structure » ? Essai sur la signification de la notion de structure dans les sciences humaines*, Paris, Gallimard, 1968.
- BOURBAKI, Nicolas, *Théorie des ensembles*, Berlin, Springer, 2006 [1970].
- BRANDT, Line, *The Communicative Mind: A Linguistic Exploration of Conceptual Integration and Meaning Construction*, Newcastle, Cambridge Scholars Publishing, 2013.
- BRANDT, Per Åge, « Toward a Cognitive Semiotics », *Recherches en communication*, no 19, 2003, p. 21-34.
- BRIER, Søren, « Biosemiotics and the Foundation of Cybersemiotics: Reconceptualizing the Insights of Ethology, Second-Order Cybernetics, and Peirce's Semiotics in Biosemiotics to Create a Non-Cartesian Information Science », *Semiotica*, vol. 127, no 1-4, 1999, p. 169-198.
- BRUNET, Étienne, *Méthodes quantitatives et informatiques dans l'étude des textes*, Paris, Champion, 1986.
- BÜCHI, J. Richard, « Algebraic Theory of Feedback in Discrete Systems, Part I (1966) », dans *The Collected Works of J. Richard Büchi*, textes rassemblés par S. Mac Lane & D. Siefkes, New York, Springer, 1990, p. 338-369.
- BUNDGAARD, Peer & Frederik STJERNFELT, *Signs and Meaning: Five Questions*, New York, Automatic Press, 2009.

- BURDICK, Anne, Johanna DRUCKER, Peter LUNENFELD, Todd PRESNER & Jeffrey SCHNAPP, *Digital Humanities*, Cambridge, The MIT Press, 2012.
- BURSTALL, Rodney M. & Peter J. LANDIN, « Programs and Their Proofs: An Algebraic Approach », *Machine Intelligence*, vol. 4, 1969, p. 17-44.
- CHAITIN, Gregory, Francisco A. DORIA & Newton DA COSTA, *Gödel's Way: Exploits into an Undecidable World*, Boca Raton, CRC Press, 2012.
- CHARTIER, Jean-François, Davide PULIZZOTTO, Louis CHARTRAND & Jean-Guy MEUNIER, « A Data-Driven Computational Semiotics: The Semantic Vector Space of Magritte's Artworks », *Semiotica*, no 230, 2019, p. 19-69.
- CHOMSKY, Noam, *Syntactic Structures*, La Haye, Mouton, 1957.
- , « On certain formal properties of grammars », *Information and Control*, vol. 2, no 2, 1959, p. 137-167.
- , « Trois modèles de description du langage », *Langages*, no 9, 1968, p. 51-76.
- CHURCH, Alonzo, « An Unsolvability Problem of Elementary Number Theory », *American Journal of Mathematics*, vol. 58, no 2, 1936, p. 345-363.
- COMPAGNO, Dario (dir.), *Quantitative Semiotic Analysis*, New York, Springer, 2018.
- CURRY, Haskell B. & Robert FEYS, *Combinatory Logic. Volume 1*, Amsterdam, North-Holland Publishing Company, 1958.
- DAVIS, Martin, *Computability and Unsolvability*, New York, Dover Publications, 1982.
- DE FRANCESCHI, Élisabeth, « Le moment de conclure (1977-1978) », *Oxymoron, revue des médiations thérapeutiques par l'art*, no 0, 2010. En ligne : <<http://revel.unice.fr/oxymoron/index.html?id=3072>>.
- DE SOUZA, Clarisse Sieckenius, *The Semiotic Engineering of Human-Computer Interaction*, Cambridge, The MIT Press, 2005.
- DESCLÉS, Jean-Pierre, « Réseaux sémantiques : la nature logique et linguistique des relateurs », *Langages*, no 87, 1987, p. 55-78.
- , « La Grammaire Applicative et Cognitive construit-elle des représentations universelles? », *Linx*, no 48, 2003, p. 139-160.
- DIMINESCU, Dana & Michel WIEVIORKA, « Le défi numérique pour les sciences sociales », *Socio*, no 4, 2015, p. 9-17.
- DUHEM, Pierre, *La théorie physique. Son objet et sa structure*, Paris, Chevalier & Rivière, 1906.
- FETZER, James H., « Minds and Machines: Limits to Simulations of Thought and Action », *International Journal of Signs and Semiotic Systems*, vol. 1, no 1, 2011, p. 39-48.

- GANDY, Robin, « Church's Thesis and Principles for Mechanisms », *Studies in Logic and the Foundations of Mathematics*, vol. 101, « The Kleene Symposium », 1980, p. 123-148.
- GIERE, Ronald, « Using Models to Represent Reality », dans L. Magnani, N. J. Nersessian & P. Thagard (dir.), *Model-Based Reasoning in Scientific Discovery*, Boston, Springer, 1999, p. 41-57.
- GÖDEL, Kurt, *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*, trad. de l'allemand par B. Meltzer, New York, Dover Publications, 1992 [1931].
- GRANGER, Gilles-Gaston, *Pensée formelle et sciences de l'homme*, Paris, Aubier-Montaigne, 1967.
- GREEN, Sara, « When one Model is not Enough: Combining Epistemic Tools in Systems Biology », *Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 44, no 2, 2013, p. 170-180.
- GREIMAS, Algirdas J. & Joseph COURTÈS, *Sémiotique : dictionnaire raisonné de la théorie du langage*, Paris, Hachette, 1979.
- GROSS, Maurice & André LENTIN, *Notions sur les grammaires formelles*, Paris, Gauthier-Villars, 1970.
- GROUPE μ , *Principia Semiotica : aux sources du sens*, Bruxelles, Les Impressions nouvelles, 2015.
- GUDWIN, Ricardo & Fernando GOMIDE, « Computational Semiotics: An Approach for the Study of Intelligent Systems. Part I: Foundations », rapport, 1997. En ligne : <http://www.dca.fee.unicamp.br/~gudwin/ftp/publications/rep1_97.pdf>.
- HILBERT, David, *Les fondements de la géométrie*, trad. de l'allemand par P. Rossier, Paris, Dunod, 1971 [1899].
- HINTIKKA, Jaakko, « On the Development of the Model-Theoretic Viewpoint in Logical Theory », *Lingua Universalis vs. Calculus Ratiocinator An Ultimate Presupposition of Twentieth-Century Philosophy*, Dordrecht, Springer, 1997, p. 104-161.
- HJEMSLEV, Louis, *Prolégomènes à une théorie du langage, suivi de la structure fondamentale du langage*, trad. du danois par U. Canger, avec la collab. d'A. Wewer, Paris, Minuit, coll. « Arguments », 1968.
- HOFFMEYER, Jesper, « Surfaces Inside Surfaces. On the Origin of Agency and Life », *Cybernetics & Human Knowing*, vol. 5, no 1, 1998, p. 33-42.
- ISRAEL, Giorgio, *La Mathématisation du réel. Essai sur la modélisation mathématique*, Paris, Seuil, 1996.
- KETNER, Kenneth Laine, « Peirce and Turing: Comparisons and Conjectures », *Semiotica*, vol. 68, no 1-2, 1988, p. 33-62.

- KULL, Kalevi & Ekaterina VELMEZOVA, « What is the Main Challenge for Contemporary Semiotics? », *Sign System Studies*, vol. 42, no 4, 2014, p. 530-548.
- LACOMBE, Daniel, « La théorie des fonctions récursives et ses applications », *Bulletin de la Société Mathématique de France*, tome 88, 1960, p. 393-468.
- LADRIÈRE, Jean, *Les limitations internes des formalismes*, Paris, Gauthier-Villars, 1957.
- LAMBEK, Joachim & Michael LAMBEK, « The Kinship Terminology of Malagasy Speakers in Mayotte », *Anthropological Linguistics*, vol. 23, no 4, 1981, p. 154-182.
- LANDIN, Peter J., « The Next 700 Programming Languages », *Communications of the ACM*, vol. 9, no 3, 1966, p. 157-166.
- LÉVI-STRAUSS, Claude, « Réflexions sur l'atome de parenté », *L'Homme. Revue française d'anthropologie*, tome 13, no 3, 1973, p. 5-30.
- LONGHI, Julien, « Contours, perspectives et tensions des "humanités numériques" », *Sens-Dessous*, no 24, 2019, p. 123-135.
- LUHMANN, Niklas, *Social Systems*, trad. de l'allemand par J. Bednarz Jr. & D. Barcker, préface d'E. M. Knodt, Stanford, Stanford University Press, 1995.
- LUKASIEWICZ, Jan, *Elements of Mathematical Logic*, New York, MacMillan, 1963 [1929].
- MAHONEY, Michael S., « The Structures of Computation and the Mathematical Structure of Nature », *The Rutherford Journal*, vol. 3, 2010. En ligne : <<http://www.rutherfordjournal.org/article030107.html>>.
- MANNING, Peter K. « Semiotics and Data Analysis », dans M. Hardy & A. Bryman (dir.), *Handbook of Data Analysis*, Londres, SAGE Publications, 2004, p. 566-587.
- MARKOV, Andreï, « The Theory of Algorithms », *American Mathematical Society Translations: Series 2*, vol. 15, 1960, p. 1-14.
- MARR, David, « Selections from Vision », dans A. Noë & E. Thompson (dir.), *Vision and Mind. Selected Readings in the Philosophy of Perception*, Cambridge, The MIT Press, 2002, p. 229-265.
- MARTY, Robert, *L'Algèbre des signes. Essai de sémiotique scientifique d'après C.S. Peirce*, Amsterdam/Philadelphie, John Benjamins Publishing Company, 1990.
- MAYAFFRE, Damon, *Le discours présidentiel sous la V^e République. Chirac, Mitterrand, Giscard, Pompidou, de Gaulle*, Paris, Presses de Sciences Po, 2012.
- MCCARTHY, John, « Recursive Functions of Symbolic Expressions and their Computation by Machine – Part 1 », *Communications of the ACM*, vol. 3, no 4, 1960, p. 184-195.
- , « A Basis for a Mathematical Theory of Computation », *Proceedings of the Western Joint Computer Conference*, 1961, p. 225-238.

- MEUNIER, Jean-Guy, « Artificial Intelligence and Sign Theory », *Semiotica*, vol. 77, no 1-3, 1989, p. 43-64.
- , « Categorical Structure of Iconic Languages », *Theory and Psychology*, vol. 8, no 6, 1998, p. 805-827.
- , « Vers une sémiotique computationnelle? », *Applied Semiotics/Sémiotique appliquée*, no 26, 2018, p. 75-107. En ligne : <<http://french.chass.utoronto.ca/as-sa/ASSA-No26/26-6.pdf>>.
- MEUNIER, Jean-Guy, Ismail BISKRI & Dominic FOREST, « Classification and Categorization in Computer Assisted Reading and Analysis of Texts », dans C. Lefebvre & H. Cohen (dir.), *Handbook of Categorization in Cognitive Science*, New York, Elsevier, 2005, p. 955-978.
- MONTAGUE, Richard, *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, Yale University Press, 1974.
- MORGAN, Mary S. & Margaret MORRISON (dir.), *Models as Mediators. Perspectives on Natural and Social Science*, Cambridge, Cambridge University Press, 1999.
- MORRIS, Charles W., *Foundations of the Theory of Signs*, Chicago, The University of Chicago Press, 1938.
- NADIN, Mihai, « Information and Semiotic Processes. The Semiotics of Computation », *Cybernetics and Human Knowing*, vol. 18, no 1-2, 2011, p. 153-175.
- , « Reassessing the Foundations of Semiotics: Preliminaries », *International Journal of Signs and Semiotic Systems*, vol. 2, no 1, 2012, p. 1-31.
- PERLOVSKY, Leonid I., *Neural Networks and Intellect: Using Model-Based Concepts*, New York, Oxford University Press, 2001.
- PETITOT, Jean, *Morphogénèse du sens*, Paris, Presses universitaires de France, 1985.
- PIOTROWSKI, David & Yves-Marie VISETTI, « Connaissance sémiotique et mathématisation. Sémiogenèse et explicitation », *Versus*, no 118, 2014, p. 141-170.
- POST, Emil L., « Recursively Enumerable Sets of Positive Integers and their Decision Problems », *Bulletin of the American Mathematical Society*, vol. 50, no 5, 1944, p. 284-316. En ligne : <https://projecteuclid.org/download/pdf_1/euclid.bams/1183505800>.
- PRIETO, Luis, *Messages et signaux*, Paris, Presses universitaires de France, 1966.
- PULIZZOTTO, Davide, *La sémiotique computationnelle : entre narrativité et apprentissage automatique. Une démonstration de faisabilité avec un corpus automatique journalistique à propos du Printemps érable*, thèse de doctorat, Montréal, Université du Québec à Montréal, 2020.

- RAPAPORT, William J., « Semiotic Systems, Computers, and the Mind: How Cognition Could Be Computing », *International Journal of Signs and Semiotic Systems*, vol. 2, no 1, 2012, p. 32-71.
- RASTIER, François, *La mesure et le grain : sémantique de corpus*, Paris, Honoré Champion, 2011.
- RASTIER, François, Marc CAVAZZA & Anne ABEILLÉ, *Sémantique pour l'analyse de la linguistique à l'informatique*, Paris, Masson, 1994.
- RIEGER, Burghard B., « Computing Fuzzy Semantic Granules from Natural Language Texts. A computational semiotics approach to understanding word meanings », *Proceedings of the IASTED International Conference Artificial Intelligence and Soft Computing*, 1999, p. 475-479. En ligne : <<https://www.uni-trier.de/fileadmin/fb2/LDV/Rieger/Publikationen/Aufsaeetze/99/iasted99.pdf>>.
- ROBERTS, Don D., *The Existential Graphs of Charles S. Peirce*, La Haye, Mouton, 1973.
- SCHÜTZENBERGER, Marcel-Paul, « Some Remarks on Chomsky's Context-Free Languages », *Quarterly Progress Report of the MIT Research Laboratory of Electronics*, no 68, 1961, p. 155-170.
- SCOTT, Dana, « Outline of a Mathematical Theory of Computation », *Technical Monograph PRG-2*, Oxford University Computing Laboratory, 1970, p. 169-176.
- SEBEOK, Thomas A., *Biosemiotics*, Berlin, Mouton de Gruyter, 1992.
- SHANIN, Teodor, « Models and Thought », dans T. Shanin (dir.), *The Rules of the Game*, Hoboken, Taylor and Francis, 2013 [1972], p. 1-22.
- SHAUMYAN, Sebastian K., *Applicational Grammar as a Semantic Theory of Natural Languages*, Chicago, Chicago University Press, 1977.
- STRACHEY, Christopher, « Towards a Formal Semantics », dans T. B. Steel (dir.), *Formal Language Description Languages for Computer Programming*, Amsterdam, North-Holland Publishing Company, 1966, p. 198-220.
- STRAWSON, Peter Frederick, « On referring », *Mind*, vol. 59, no 235, 1950, p. 320-344.
- SUÁREZ, Mauricio & Albert SOLÉ, « On the Analogy between Cognitive Representation and Truth », *Theoria*, vol. 21, no 1, 2006, p. 39-48. En ligne : <<https://www.ehu.eus/ojs/index.php/THEORIA/article/view/552/401>>.
- TANAKA-ISHII, Kumiko, *Semiotics of programming*, New York, Cambridge University Press, 2010.
- TIJUS, Charles, Javier BARCENILLA, Javier CAMBON DE LAVALETTE & Jean-Guy MEUNIER, « The Design, Understanding and Usability of Pictograms », dans D. Alamargot, P. Terrier & J.-M. Cellier (dir.), *Improving the Production and*

Understanding of Written Documents in the Workplace, Amsterdam, Elsevier, 2007, p. 17-32.

TURING, Alan M., « On Computable Numbers, with an Application to the Entscheidungsproblem », *Proceedings of the London Mathematical Society*, vol. 42, no 1, 1937, p. 230-265.

VON BERTALANFFY, Ludwig, « Conclusion », *Human Biology*, vol. 23, no 4, 1951, p. 336-345.

—, « Problems of General System Theory », *Human Biology*, vol. 23, no 4, 1951, p. 302-312.

—, « Towards a Physical Theory of Organic Teleology: Feedback and Dynamics », *Human Biology*, vol. 23, no 4, 1951, p. 346-361.

VON NEUMANN, John, « The general and logical theory of automata (1951) », dans *John von Neumann - Collected Works*, vol. 5, textes rassemblés par A. H. Taub, Oxford, Pergamon Press, 1961, p. 288-326.

WILDGEN, Wolfgang, *Catastrophe Theoretic Semantics. An Elaboration and Application of René Thom's Theory*, Amsterdam, John Benjamins Publishing Company, 1982.

ZEMANEK, Heinz, « Semiotics and Programming Languages », *Communications of the ACM*, vol. 9, no 3, 1966, p. 139-143.

Notes

- 1 C. LÉVI-STRAUSS, « Réflexions sur l'atome de parenté », *L'Homme. Revue française d'anthropologie*, tome 13, no 3, 1973, p. 5.
- 2 P. BUNDGAARD & F. STJERNFELT, *Signs and Meaning: Five Questions*, New York, Automatic Press, 2009.
- 3 En 1972, Lotman écrivait, je traduis : « La capacité des différentes disciplines mathématiques de servir de métalangage aussi dans la description de phénomènes artistiques est évidente » (« The ability of various mathematical disciplines to serve as a metalanguage also in the description of the phenomena of art is evident. »), cité dans K. KULL & E. VELMEZOVA, « What is the Main Challenge for Contemporary Semiotics? », *Sign System Studies*, vol. 42, no 4, 2014, p. 531.
- 4 *Ibid.*
- 5 Il y avait bien dans le courant structuraliste des réflexions théoriques sur cette notion de structure (voir R. BOUDON, *À quoi sert la notion de « structure » ? Essai sur la signification de la notion de structure dans les sciences humaines*, Paris, Gallimard, 1968). Elle n'en demeurerait pas moins analytique et conceptuelle. Quelques mathématiciens cependant l'avaient explorée sur le plan formel des sciences appliquées en sciences sociales et humaines : voir M. BARBUT, « Sur le sens du mot "structure" en mathématiques », *Les Temps Modernes*, no 246, 1966, p. 791-814.

- 6 M. NADIN, « Reassessing the Foundations of Semiotics: Preliminaries », *International Journal of Signs and Semiotic Systems*, vol. 2, no 1, 2012, p. 1-31, cité dans K. KULL & E. VELMEZOVA, « What is the Main Challenge for Contemporary Semiotics? », *loc. cit.*, p. 532.
- 7 Voir à ce sujet D. D. ROBERTS, *The Existential Graphs of Charles S. Peirce*, La Haye, Mouton, 1973.
- 8 R. MARTY, *L'Algèbre des signes. Essai de sémiotique scientifique d'après C.S. Peirce*, Amsterdam/Philadelphie, John Benjamins Publishing Company, 1990.
- 9 L. HJEMSLEV, *Prolégomènes à une théorie du langage, suivi de la structure fondamentale du langage*, trad. du danois par U. Canger, avec la collab. d'A. Wewer, Paris, Minuit, coll. « Arguments », 1968.
- 10 C. W. MORRIS, *Foundations of the Theory of Signs*, Chicago, The University of Chicago Press, 1938.
- 11 Voir notamment S. K. SHAUMYAN, *Applicational Grammar as a Semantic Theory of Natural Languages*, Chicago, Chicago University Press, 1977 ; J.-P. DESCLÈS, « La Grammaire Applicative et Cognitive construit-elle des représentations universelles? », *Linx*, no 48, 2003, p. 139-160.
- 12 M. AUDIN, « Hommage à Claude Lévi-Strauss », *Images des mathématiques*, 2009. En ligne : <<https://images.math.cnrs.fr/Hommage-a-Claude-Levi-Strauss.html>>.
- 13 J. LAMBEK & M. LAMBEK, « The Kinship Terminology of Malagasy Speakers in Mayotte », *Anthropological Linguistics*, vol. 23, no 4, 1981, p. 154-182.
- 14 J.-G. MEUNIER, « Categorial structure of Iconic Languages », *Theory and Psychology*, vol. 8, no 6, 1998, p. 805-827 ; C. TIJUS *et al.*, « The Design, Understanding and Usability of Pictograms », dans D. Alamargot, P. Terrier & J.-M. Cellier (dir.), *Improving the Production and Understanding of Written Documents in the Workplace*, Amsterdam, Elsevier, 2007, p. 17-32.
- 15 A. J. GREIMAS & J. COURTÉS, *Sémiotique : dictionnaire raisonné de la théorie du langage*, Paris, Hachette, 1979.
- 16 É. DE FRANCESCHI, « Le moment de conclure (1977-1978) », *Oxymoron, revue des médiations thérapeutiques par l'art*, no 0, 2010. En ligne : <<http://revel.unice.fr/oxymoron/index.html?id=3072>> ; R. BARTHES, *Système de la mode*, Paris, Seuil, 1967.
- 17 J. HINTIKKA, « On the Development of the Model-Theoretic Viewpoint in Logical Theory », *Lingua Universalis vs. Calculus Ratiocinator. An Ultimate Presupposition of Twentieth-Century Philosophy*, Dordrecht, Springer, 1997, p. 104-161.
- 18 W. WILDGEN, *Catastrophe Theoretic Semantics. An Elaboration and Application of René Thom's Theory*, Amsterdam, John Benjamins Publishing Company, 1982 ; J. PETITOT, *Morphogénèse du sens*, Paris, Presses universitaires de France, 1985 ; L. BRANDT, *The Communicative Mind: A Linguistic Exploration of Conceptual Integration and Meaning Construction*, Newcastle, Cambridge Scholars Publishing, 2013.
- 19 D. PIOTROWSKI & Y.-M. VISETTI, « Connaissance sémiotique et mathématisation. Sémiogenèse et explicitation », *Versus*, no 118, 2014, p. 141-170.
- 20 T. A. SEBEOK, *Biosemiotics*, Berlin, Mouton de Gruyter, 1992 ; J. HOFFMEYER, « Surfaces Inside Surfaces. On the Origin of Agency and Life », *Cybernetics & Human Knowing*, vol. 5, no 1, 1998, p. 33-42.
- 21 S. BRIER, « Biosemiotics and the Foundation of Cybersemiotics: Reconceptualizing the Insights of Ethology, Second-Order Cybernetics, and Peirce's Semiotics in Biosemiotics to Create a Non-Cartesian Information Science », *Semiotica*, vol. 127, no 1-4, 1999, p. 169-198 ; N. LUHMANN, *Social Systems*, trad. de l'allemand par J. Bednarz Jr. & D. Barcker, préface d'E. M. Knodt, Stanford, Stanford University Press, 1995.

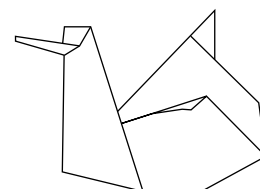
- 22 L. PRIETO, *Messages et signaux*, Paris, Presses universitaires de France, 1966 ; M. NADIN, « Information and Semiotic Processes. The Semiotics of Computation », *Cybernetics and Human Knowing*, vol. 18, no 1-2, 2011, p. 153-175.
- 23 L. BRANDT, *The Communicative Mind: A Linguistic Exploration of Conceptual Integration and Meaning Construction*, op. cit.
- 24 P. A. BRANDT, « Toward a Cognitive Semiotics », *Recherches en communication*, no 19, 2003, p. 21-34.
- 25 Cf. notamment H. ZEMANEK, « Semiotics and Programming Languages », *Communications of the ACM*, vol. 9, no 3, 1966, p. 139-143 ; K. L. KETNER, « Peirce and Turing: Comparisons and Conjectures », *Semiotica*, vol. 68, no 1-2, 1988, p. 33-62 ; J.-G. MEUNIER, « Artificial Intelligence and Sign Theory », *Semiotica*, vol. 77, no 1-3, 1989, p. 43-64 ; J. S. ALBUS & A. M. MEYSTELE, « A Reference Model Architecture for Design and Implementation of Intelligent Control in Large and Complex Systems », *International Journal of Intelligent Control and Systems*, vol. 1, no 1, 1996, p. 15-30 ; C. S. DE SOUZA, *The Semiotic Engineering of Human-Computer Interaction*, Cambridge, The MIT Press, 2005 ; K. TANAKA-ISHII, *Semiotics of programming*, New York, Cambridge University Press, 2010 ; J. H. FETZER, « Minds and Machines: Limits to Simulations of Thought and Action », *International Journal of Signs and Semiotic Systems*, vol. 1, no 1, 2011, p. 39-48 ; W. J. RAPAPORT, « Semiotic Systems, Computers, and the Mind: How Cognition Could Be Computing », *International Journal of Signs and Semiotic Systems*, vol. 2, no 1, 2012, p. 32-71.
- 26 Voir par ex. J. S. ALBUS & A. M. MEYSTELE, « A Reference Model Architecture for Design and Implementation of Intelligent Control in Large and Complex Systems », loc. cit. ; R. GUDWIN & F. GOMIDE, « Computational Semiotics: An Approach for the Study of Intelligent Systems. Part I: Foundations », rapport, 1997. En ligne : <http://www.dca.fee.unicamp.br/~gudwin/ftp/publications/rep1_97.pdf> ; B. B. RIEGER, « Computing Fuzzy Semantic Granules from Natural Language Texts. A computational semiotics approach to understanding word meanings », *Proceedings of the IASTED International Conference Artificial Intelligence and Soft Computing*, 1999, p. 475-479. En ligne : <<https://www.uni-trier.de/fileadmin/fb2/LDV/Rieger/Publikationen/Aufsaezte/99/iasted99.pdf>> ; F. RASTIER, M. CAVAZZA & A. ABEILLÉ, *Sémantique pour l'analyse de la linguistique à l'informatique*, Paris, Masson, 1994 ; L. I. PERLOVSKY, *Neural Networks and Intellect: Using Model-Based Concepts*, New York, Oxford University Press, 2001 ; D. COMPAGNO (dir.), *Quantitative Semiotic Analysis* New York, Springer, 2018.
- 27 A. BURDICK et al., *Digital Humanities*, Cambridge, The MIT Press, 2012, p. 8.
- 28 *Ibid.*, p. 16. Je traduis : « Digital Humanities is an extension of traditional knowledge skills and methods, not a replacement for them. »
- 29 On peut citer par exemple : J. PETITOT, *Morphogénèse du sens*, op. cit. ; É. BRUNET, *Méthodes quantitatives et informatiques dans l'étude des textes*, Paris, Champion, 1986 ; J.-P. DESCLÉS, « Réseaux sémantiques : la nature logique et linguistique des relateurs », *Langages*, no 87, 1987, p. 55-78 ; F. RASTIER, M. CAVAZZA & A. ABEILLÉ, *Sémantique pour l'analyse de la linguistique à l'informatique*, op. cit. ; B. B. RIEGER, « Computing Fuzzy Semantic Granules from Natural Language Texts », loc. cit. ; P. K. MANNING, « Semiotics and Data Analysis », dans M. Hardy & A. Bryman (dir.), *Handbook of Data Analysis*, Londres, SAGE Publications, 2004, p. 566-587 ; J.-G. MEUNIER, I. BISKRI & D. FOREST, « Classification and Categorization in Computer Assisted Reading and Analysis of Texts », dans C. Lefebvre & H. Cohen (dir.), *Handbook of Categorization in Cognitive Science*, New York, Elsevier, 2005, p. 955-978 ; F. RASTIER, *La mesure et le grain : sémantique de corpus*, Paris, Honoré Champion, 2011 ; D. COMPAGNO (dir.), *Quantitative Semiotic Analysis*, op. cit. ; J.-G. MEUNIER, « Vers une sémiotique computationnelle? », *Applied Semiotics/Sémiotique appliquée*, no 26, 2018, p. 77. En ligne : <<http://french.chass.utoronto.ca/as-sa/ASSA-No26/26-6.pdf>> ; J.-F. CHARTIER et al., « A Data-Driven Computational Semiotics: The Semantic Vector Space of Magritte's Artworks », *Semiotica*, no 230, 2019, p. 19-69.

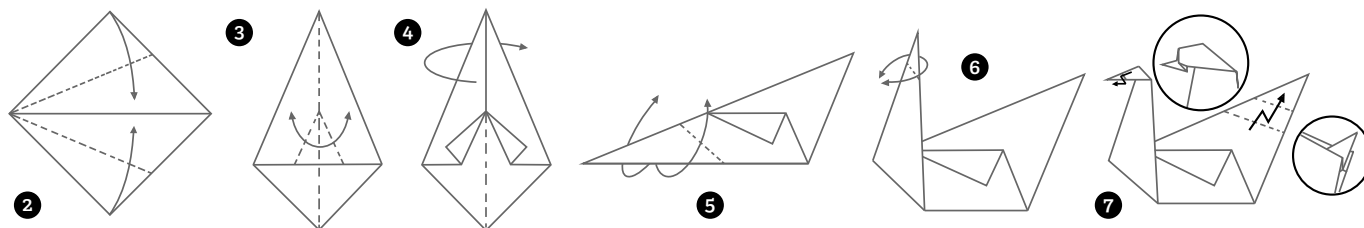
- 30 J. LONGHI, « Contours, perspectives et tensions des “humanités numériques” », *Sens-Dessous*, no 24, 2019, p. 123-135.
- 31 M. ARMATTE & A. DAHAN DALMEDICO, « Modèles et modélisations, 1950-2000 : nouvelles pratiques, nouveaux enjeux », *Revue d'histoire des sciences*, tome 57, no 2, 2004, p. 243-303 ; D. DIMINESCU & M. WIEVIORKA, « Le défi numérique pour les sciences sociales », *Socio*, no 4, 2015, p. 9-17.
- 32 Voir par exemple, pour ne citer que cela, M. S. MORGAN & M. MORRISON (dir.), *Models as Mediators. Perspectives on Natural and Social Science*, Cambridge, Cambridge University Press, 1999 ; R. GIERE, « Using Models to Represent Reality », dans L. Magnani, N. J. Nersessian & P. Thagard (dir.), *Model-Based Reasoning in Scientific Discovery*, Boston, Springer, 1999, p. 41-57.
- 33 P. F. STRAWSON, « On referring », *Mind*, vol. 59, no 235, 1950, p. 320-344.
- 34 Von Bertalanffy a si bien formulé la nécessité d'un modèle conceptuel en science : avant d'en arriver à une formalisation mathématique, il faut avoir conceptualisé l'objet même à mathématiser : « Les formulations mathématiques sont des expressions et des conséquences de certains modèles conceptuels que nous appliquons. » Cf. L. VON BERTALANFFY, « Problems of General System Theory », *Human Biology*, vol. 23, no 4, 1951, p. 341 ; L. VON BERTALANFFY, « Conclusion », *Human Biology*, vol. 23, no 4, 1951, p. 336-345 ; L. VON BERTALANFFY, « Towards a Physical Theory of Organic Teleology: Feedback and Dynamics », *Human Biology*, vol. 23, no 4, 1951, p. 346-361.
- 35 G.-G. GRANGER, *Pensée formelle et sciences de l'homme*, Paris, Aubier-Montaigne, 1967, p. 44 (je souligne).
- 36 P. DUHEM, *La théorie physique. Son objet et sa structure*, Paris, Chevalier & Rivière, 1906.
- 37 T. SHANIN, « Models and Thought », dans T. Shanin (dir.), *The Rules of the Game*, Hoboken, Taylor and Francis, 2013 [1972], p. 1. Je traduis : « Scholarly disciplines embody the formalization aspects. [...] It reflects, on the one hand, the need of the human mind for rationalizing and for orientation in an immensely complex reality, to be achieved by building up systematic and coherent general images and symbolic schemes. »
- 38 J. LADRIÈRE, *Les limitations internes des formalismes*, Paris, Gauthier-Villars, 1957, p. 80. Ladrière synthétise de manière juste, mais pédagogique, les définitions formelles de Peirce, Frege, Russell, Lukscievicks, Husserl, Hilbert, Tarski, Gentzen, Bourbaki, Church, notamment.
- 39 D. MARR, « Selections from Vision », dans A. Noë & E. Thompson (dir.), *Vision and Mind. Selected Readings in the Philosophy of Perception*, Cambridge, The MIT Press, 2002, p. 244. Je traduis : « To say that something is a formal scheme means only that it is a set of symbols with rules for putting them together—no more no less. »
- 40 Bien qu'un langage formel exige des symboles discrets et atomiques et des règles strictes de leur manipulation, il ne s'ensuit pas que ces langages formels sont incapables de traiter du continu, du parallèle et de l'infini. Ceci relèvera de leur sémantique.
- 41 À titre d'exemple simple, une certaine relation d'implication entre des propositions p et q sera dans le formalisme russellien écrit $p \rightarrow (p \rightarrow q \vee z)$ alors que, dans la notation polonaise, elle devient $CpAqz$ (cf. J. LUKASIEWICZ, *Elements of Mathematical Logic*, New York, MacMillan, 1963 [1929]). Mais deux formalismes ou notations exprimeront la même structure d'implication logique.
- 42 Pour Suárez et Solé, le rôle de la représentation comme mise en relation entre un objet d'étude et un modèle ne peut être le seul rôle d'un modèle. Il doit permettre des inférences. M. SUÁREZ & A. SOLÉ, « On the Analogy between Cognitive Representation and Truth », *Theoria*, vol. 21, no 1, 2006, p. 39. En ligne : <<https://www.ehu.es/ojs/index.php/THEORIA/article/view/552/401>>. Je traduis : « In a scientific representation some source A – typically a model, a graph, an equation – is used to represent some target B – typically a system, entity or phenomenon. »

- 43 La sémantique de ces symboles ne se limite évidemment pas à la question de la référence et la signification n'épuise pas le « sens ».
- 44 Dans un langage logique, une constante peut se référer à une personne et un prédicat à une propriété, par exemple la phrase formelle logique suivante « $S(j)$ et $B(j)$ » peut être interprétée de la manière suivante : « j » se réfère à « John », « S » à la propriété « Student » et « B » à la propriété « Brilliant ». Ce n'est donc pas parce qu'un modèle est formel qu'il fait nécessairement référence à des nombres ou, comme on le dit, qu'il est « quantitatif ». Dire que les symboles formels sont relatifs à des quantités est une proposition ambiguë : tous les symboles formels ne sont pas des chiffres. L'École de Bourbaki refusera précisément de voir la logique ou l'arithmétique comme le domaine premier de la mathématique.
- 45 Mahoney écrit par exemple : « Science is not about nature, but about how we represent nature to ourselves. » M. S. MAHONEY, « The Structures of Computation and the Mathematical Structure of Nature », *The Rutherford Journal*, vol. 3, 2010. En ligne : <<http://www.rutherfordjournal.org/article030107.html>>.
- 46 N. BOURBAKI, *Théorie des ensembles*, Berlin, Springer, 2006 [1970].
- 47 Ceci est un point important de la différence entre un langage ou langue formelle et un langage naturel. Certaines théories sémantiques soutiennent que la signification et le sens émergent des intentions de communication des locuteurs, plusieurs théories sémantiques soutiennent par ailleurs que la langue participe profondément à l'organisation et à la structuration de la sémantique d'un système formel.
- 48 Ici tant l'épistémologie sociale que la sociologie de la science peuvent jouer un rôle important dans l'explicitation du rôle épistémique du « créateur » de ces modèles. Ce « créateur » n'est souvent pas nécessairement un individu. Il peut être une communauté épistémique qui n'est pas sans être influencée par d'autres théories épistémiques antérieures ou périphériques, mais aussi des conditions sociopolitiques de production.
- 49 M. S. MORGAN & M. MORRISON (dir.), *Models as Mediators*, op. cit.
- 50 G. ISRAEL, *La Mathématisation du réel. Essai sur la modélisation mathématique*, Paris, Seuil, 1996, p. 75.
- 51 Il faut se rappeler que le symbole $\times$ de la multiplication est une abréviation qui réfère à une sommation complexe de la valeur d'un nombre. Ainsi 3×3 est une abréviation $(3 + 3) + (3 + 3) + (3 + 3) = 9$.
- 52 D. HILBERT, *Les fondements de la géométrie*, trad. de l'allemand par P. Rossier, Paris, Dunod, 1971 [1899].
- 53 A. M. TURING, « On Computable Numbers, with an Application to the Entscheidungsproblem », *Proceedings of the London Mathematical Society*, vol. 42, no 1, 1937, p. 230-265.
- 54 A. CHURCH, « An Unsolvable Problem of Elementary Number Theory », *American Journal of Mathematics*, vol. 58, no 2, 1936, p. 345-363.
- 55 H. B. CURRY & R. FEYS, *Combinatory Logic. Volume 1*, Amsterdam, North-Holland Publishing Company, 1958.
- 56 A. MARKOV, « The Theory of Algorithms », *American Mathematical Society Translations: Series 2*, vol. 15, 1960, p. 1-14.
- 57 E. L. POST, « Recursively Enumerable Sets of Positive Integers and their Decision Problems », *Bulletin of the American Mathematical Society*, vol. 50, no 5, 1944, p. 284-316. En ligne : <https://projecteuclid.org/download/pdf_1/euclid.bams/1183505800>.
- 58 R. GANDY, « Church's Thesis and Principles for Mechanisms », *Studies in Logic and the Foundations of Mathematics*, vol. 101, « The Kleene Symposium », 1980, p. 123-148.

- 59 J. VON NEUMANN, « The general and logical theory of automata (1951) », dans *John von Neumann - Collected Works*, vol. 5, textes rassemblés par A. H. Taub, Oxford, Pergamon Press, 1961, p. 303. Je traduis : « The theory of automata [...] will have to be, from the mathematical point of view, combinatorial rather than analytical. »
- 60 N. CHOMSKY, « On certain formal properties of grammars », *Information and Control*, vol. 2, no 2, 1959, p. 137-167.
- 61 P. J. LANDIN, « The Next 700 Programming Languages », *Communications of the ACM*, vol. 9, no 3, 1966, p. 157-166.
- 62 N. CHOMSKY, « Trois modèles de description du langage », *Langages*, no 9, 1968, p. 52 (traduction de N. CHOMSKY, « Three Models for the Description of Language », *IRE Transactions on Information Theory*, vol. 2, no 3, 1956, p. 113-124).
- 63 M.-P. SCHÜTZENBERGER, « Some Remarks on Chomsky's Context-Free Languages », *Quarterly Progress Report of the MIT Research Laboratory of Electronics*, no 68, 1961, p. 155-170.
- 64 Ces travaux reçurent beaucoup d'échos en France par le biais des travaux de Maurice Gross et André Lentin (cf. M. GROSS & A. LENTIN, *Notions sur les grammaires formelles*, Paris, Gauthier-Villars, 1970) ainsi que ceux des mathématiciens linguistes du Centre des mathématiques appliquées (cf. J.-P. DESCLÈS, « La Grammaire Applicative et Cognitive construit-elle des représentations universelles? », *loc. cit.*). Ces travaux donnèrent une rigueur mathématique au structuralisme naissant.
- 65 D. LACOMBE, « La théorie des fonctions récursives et ses applications », *Bulletin de la Société Mathématique de France*, tome 88, 1960, p. 393-468.
- 66 J. MCCARTHY, « Recursive Functions of Symbolic Expressions and their Computation by Machine – Part 1 », *Communications of the ACM*, vol. 3, no 4, 1960, p. 184-195 ; J. MCCARTHY, « A Basis for a Mathematical Theory of Computation », *Proceedings of the Western Joint Computer Conference*, 1961, p. 225-238.
- 67 C. STRACHEY, « Towards a Formal Semantics », dans T. B. Steel (dir.), *Formal Language Description Languages for Computer Programming*, Amsterdam, North-Holland Publishing Company, 1966, p. 198-220.
- 68 R. M. BURSTALL & P. J. LANDIN, « Programs and Their Proofs: An Algebraic Approach », *Machine Intelligence*, vol. 4, 1969, p. 17. Je traduis : « Programming is essentially about certain “data structures” and functions between them. »
- 69 A. ASPERTI & G. LONGO, *Categories, Types and Structures*, Cambridge, The MIT Press, 1991.
- 70 R. M. BURSTALL & P. J. LANDIN, « Programs and Their Proofs: An Algebraic Approach », *loc. cit.*, p. 170. Je traduis : « Programming is essentially about certain “data structures” and functions between them. [...] Starting with such familiar algebraic structures as groups and rings algebraists have developed a wider notion of algebraic structure (or “algebra”) which includes these as examples and also includes many of the entities which in the computer context are thought of as data structures. »
- 71 J. R. BÜCHI, « Algebraic Theory of Feedback in Discrete Systems, Part I (1966) », dans *The Collected Works of J. Richard Büchi*, textes rassemblés par S. Mac Lane & D. Siefkes, New York, Springer, 1990, p. 339. Je traduis : « If the definition of “finite automaton” is appropriately chosen, it turns out that all the basic concepts and results concerning structure and behavior of finite automata are in fact just special cases of the fundamental concepts (homomorphism, congruence relation, free algebra) and facts of abstract algebra. Automata theory is simply the theory of universal algebra [...] with unary operations, and with emphasis on finite algebras. »
- 72 D. SCOTT, « Outline of a Mathematical Theory of Computation », *Technical Monograph PRG-2*, Oxford University Computing Laboratory, 1970, p. 170. Je traduis : « The main mathematical novelty of the present study is the creation of a proper mathematical theory of functions which accomplishes these aims (consistently!) [...] »

- 73 M. S. MAHONEY, « The Structures of Computation and the Mathematical Structure of Nature », *loc. cit.*, non paginé. Je traduis : « In the absence of mathematical structures that allow abstraction and generalization, computational models do not say much. Nor do they function as models traditionally have done in providing an understanding of nature on the basis of which we can test our knowledge by making things happen in the world. »
- 74 S. GREEN, « When one Model is not Enough: Combining Epistemic Tools in Systems Biology », *Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 44, no 2, 2013, p. 170-180.
- 75 J. LAMBEK & M. LAMBEK, « The Kinship Terminology of Malagasy Speakers in Mayotte », *loc. cit.*
- 76 Ce terme d'idéalisation est un des rôles épistémiques que plusieurs théoriciens donnent aux modèles scientifiques. Cf. M. S. MORGAN & M. MORRISON (dir.), *Models as Mediators*, *op. cit.*
- 77 Il est théoriquement possible de représenter des événements dynamiques par des propositions logiques. Mais cela exigera des symboles qui puissent représenter chacun des instants-moments de l'évolution d'un phénomène et donc ultimement de simuler le continu. Cela rendra lourd le système logique. Et on imagine comment il sera difficile de faire pour des systèmes de type dynamique et chaotique. Même les logiques du temps doivent se restreindre dans les objets dynamiques qu'ils peuvent représenter. Ce n'est pas sans raison que les pratiques scientifiques ne représentent pas les phénomènes dynamiques par le biais de la logique des prédicats, des propositions et même des logiques modales.
- 78 Voir la critique de Piotrowski et Visetti sur l'application de ces types de modèles à la sémiolinguistique. D. PIOTROWSKI & Y.-M. VISETTI, « Connaissance sémiotique et mathématisation. Sémiogenèse et explicitation », *loc. cit.*
- 79 K. GÖDEL, *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*, trad. de l'allemand par B. Meltzer, New York, Dover Publications, 1992 [1931].
- 80 G. CHAITIN, F. A. DORIA & N. DA COSTA, *Gödel's Way: Exploits into an Undecidable World*, Boca Raton, CRC Press, 2012.
- 81 M. DAVIS, *Computability and Unsolvability*, New York, Dover Publications, 1982.





Données brutes ou hypersymboles? Signification et données numériques, entre processus discursif et procédure machinique

Lucile CRÉMIER, Maude BONENFANT & Laura Iseut LAFRANCE ST-MARTIN

Cygne noir, no 7, 2019 : « Algorithmes »

Résumé

L'analyse de grands ensembles de données numériques (le plus souvent appelés « *Big Data* » ou « données massives ») constitue aujourd'hui une méthode de recherche de plus en plus populaire, consensuelle et couramment utilisée en sciences sociales. Avec l'automatisation de la collecte de données, la mathématisation de l'analyse et l'objectivation numérique, cette méthode gagne en efficacité et la valeur de vérité des résultats obtenus se renforce. Cet article établit, en premier lieu, une revue de littérature critique concernant la collecte et l'analyse des données massives et résume les débats éthiques actuels qu'occasionnent de tels outils informatiques de recherche. En second lieu, nous présentons un modèle sémiotique de la production et de la circulation des données numériques afin de problématiser l'idée selon laquelle les données donneraient à voir « le monde lui-même » (une présentation directe du monde surpassant tous les autres modes de représentation), plutôt que d'être un moyen de se figurer le monde (un mode de représentation parmi d'autres). Suivant l'approche sémiotique et la philosophie pragmatiste de Charles Sanders Peirce, nous définissons alors la numérisation comme un processus sémiotique d'hypersymbolisation. Nous mettons ainsi en lumière l'apparente naturalisation du sens, l'illusion d'iconicité et la rhétorique sur lesquelles repose la valeur de vérité des données dans le contexte économique et politique de leur usage à des fins d'application lucrative. Nous décrivons ainsi quelques implications épistémologiques et éthiques découlant de nos modes de représentation et d'usage des données, ainsi que de la valeur d'autorité qui leur est attribuée. Ceci nous permet d'entrevoir plusieurs pistes pour l'étude critique de l'analyse des données massives dans une perspective sémiotique.

Pour citer cet article

CRÉMIER, Lucile, Maude BONENFANT & Laura Iseut LAFRANCE ST-MARTIN, « Données brutes ou hypersymboles? Signification et données numériques, entre processus discursif et procédure machinique », *Cygne noir*, no 7, 2019.
En ligne : <<http://www.revuecygnoir.org/numero/article/cremier-et-al-hypersymboles>> (consulté le xx/xx/xxxx).



Cet article de *Revue Cygne noir* est mis à disposition selon les termes de la licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Pas de Modification 2.5 Canada.

DONNÉES BRUTES OU HYPERSYMBOLS? SIGNIFICATION ET DONNÉES NUMÉRIQUES, ENTRE PROCESSUS DISCURSIF ET PROCÉDURE MACHINIQUE *

1. Introduction

Cet article vise à examiner et à problématiser la production de la signification fondée sur l'analyse automatisée de données numériques afin de mieux comprendre les conditions de possibilité et les lacunes des méthodes de recherche basées sur les données massives dans le contexte de leur application en sciences humaines et sociales. En d'autres termes, nous cherchons à caractériser, dans une perspective sémiotique, ce qui explique l'« efficacité » et la valeur de vérité accordées aux données numériques. Nos principales questions de recherche sont les suivantes : à quel(s) type(s) de signes les données numériques s'apparentent-elles ? Comment la production de signification fonctionne-t-elle au cours de la production et du traitement des données ? Comment, pour qui et dans quel but les données deviennent-elles signifiantes ?

Afin de répondre à ces questions, nous résumons d'abord la littérature produite par les partisans de la science des données puis les critiques formulées concernant l'usage des données massives (Section 2). Ensuite, nous présentons un modèle sémiopragmatiste du fonctionnement des données comme signes (Section 3). Ainsi, nous questionnons leur mise en forme dans le contexte de leur mobilisation dans le discours. Sur la base de ce modèle, nous mettons en lumière un processus de naturalisation du sens, par lequel les données passent pour des images immédiates de phénomènes (illusion d'iconicité), ainsi que la construction d'effets symboliques (re/construction de symboles) grâce auxquels les données acquièrent leur valeur de vérité. C'est en cela précisément que la numérisation peut être considérée comme un processus hypersymbolique (Section 4). Nous décrivons ensuite quelques implications épistémologiques et éthiques découlant de nos modes de représentation et d'usage des données, ainsi que de la valeur d'autorité qui leur est attribuée. Ceci nous permet d'entrevoir plusieurs pistes pour l'étude critique de l'analyse des données massives dans une perspective sémiotique.

Tout au long de cet article, des questions sémiotiques plus générales orientent notre perspective : en quoi les mots « sens » et « signification » peuvent-ils s'appliquer aussi

* Publication originale anglaise : « Raw data or hypersymbols? Meaning-making with digital data, between discursive processes and machinic procedures », *Semiotica*, no 230, 2019, p. 189-212. Traduction et adaptation par Lucile Crémier, Maude Bonenfant & Laura Iseut Lafrance St-Martin.

bien en informatique qu'en sémiotique? Qu'implique le fait de « signifier » dans chaque discipline? Comment des outils informatiques de plus en plus complexes affectent-ils la manière dont nous comprenons les concepts de signification et de sémiologie eux-mêmes et que devrions-nous en tirer, en tant qu'individus et communautés politiques?

2. Revue de littérature

En science des données, de grands ensembles de données numériques sont constitués et analysés à l'aide d'une panoplie d'outils et de techniques. Il est de plus en plus répandu, légitime et à la mode d'employer de tels outils et bases de données massives pour la recherche en sciences sociales¹. Or, ces techniques mobilisent un mode spécifique de production de la connaissance. En effet, l'enregistrement et l'encodage des informations, d'après la collecte, la quantification et la compilation automatisées d'un ensemble de données qui seront communiquées en fonction d'une analyse mathématique et algorithmique, se démarquent d'autres méthodes de recherche : ce processus objective des informations qualitatives en fonction de critères essentiellement univoques. De ce fait, ces méthodes et les résultats qu'elles produisent se voient attribuer une forte valeur de vérité. En outre, la taille imposante des corpus de données numériques constitue un argument d'autorité en soi : la quantité « parle d'elle-même² » puisque la base de données massives donne à voir des corrélations permettant de tirer des conclusions. Cette logique quantitative³ tend à dominer les champs d'études où l'on vise avant tout à produire des conclusions solides et de l'information fiable.

Dans la section suivante, nous faisons état de quelques discours qui accompagnent les avancées récentes en science des données et des contributions scientifiques qui informent notre compréhension des données numériques aujourd'hui. Ainsi, on met en évidence les enjeux de l'analyse des données sous l'angle de la production de sens, laquelle est toujours déjà située à l'intersection entre la rhétorique, l'idéologie et le travail scientifique.

2.1 Les discours des données massives

Plusieurs chercheur·e·s et représentant·e·s de l'industrie en question affirment que le développement contemporain de la science des données annonce la « fin de la théorie »⁴. Dans un article provocateur publié il y a maintenant dix ans sur la plateforme *Wired*, Chris Anderson déclarait en effet que la recherche devrait délaisser dès maintenant les

disciplines du savoir modernes : « Oubliez la taxinomie, l'ontologie et la psychologie. Qui sait pourquoi les gens font ce qu'ils font? L'important, c'est qu'ils le font et nous pouvons mesurer cela avec une exactitude sans précédent. Avec suffisamment de données, les chiffres parlent d'eux-mêmes⁵. » En d'autres termes, les données massives libéreraient les chercheur·e·s des méthodes de recherche traditionnelles, c'est-à-dire de la nécessité de formuler des hypothèses avant de les confronter aux faits par l'analyse de données, puisqu'il serait possible de « sonder tout un ensemble de données à la recherche de régularités (*patterns*) révélatrices d'effets, pour produire des conclusions sans mener d'expériences scientifiques⁶ ». Si, selon Anderson, « [n]ous pouvons jeter des chiffres dans les plus grands ordinateurs que le monde ait jamais vus et laisser les algorithmes trouver des régularités statistiques là où la science ne le peut⁷ », alors nous serions témoins du processus de production le plus fiable et arrêté des connaissances les plus objectives et exactes sur les phénomènes naturels et sociaux.

À l'extérieur des communautés scientifique et informatique qui produisent et utilisent, entre autres technologies d'analyse de données, des algorithmes, les représentations courantes du potentiel de la science des données reposent donc sur des présupposés ontologiques et structurels particuliers, à commencer par la visée d'exhaustivité éventuelle des bases de données et la pertinence de la réduction quantitative, sur lesquelles nous nous penchons plus loin. À ce titre, Danah Boyd et Kate Crawford soulignent le fait que l'expression « *Big Data* » est le fruit d'un imaginaire scientifique particulier qui repose sur une croyance mythologique concernant la valeur de la quantification ainsi que sur une foi en sa capacité de modélisation du monde réel⁸. En effet, il semble que la recherche en science des données requiert une croyance absolue en l'objectivité, l'exactitude et la justesse des connaissances produites par la collecte et l'analyse automatisées de grands ensembles de données, surtout lorsqu'il s'agit de projets à but lucratif ou arrimés au marché. La mise en marché de l'information obtenue est habituellement un objectif explicite de ce type de recherche, ce qui normalise la rhétorique du marketing⁹. « Le marché perçoit les données massives de manière purement opportuniste : en vente, on les utilise pour produire des annonces ciblées, les compagnies d'assurance s'en servent pour optimiser leurs offres et les banquiers de Wall Street les mobilisent pour interpréter le marché¹⁰ » : les avancées récentes en science des données ont aussi révolutionné le mode de fonctionnement des entreprises¹¹. La valeur des données numériques est donc plus stratégique que jamais dans la logique de marché capitaliste, mais leur valeur est avant tout marchande, puisque la circulation des données est sujette à des droits d'auteur et à des frais d'utilisation¹².

Il est utile de souligner que les conséquences éthiques de tels développements ont été vivement critiquées, non seulement dans les cercles universitaires et scientifiques¹³,

mais aussi en dehors des institutions du savoir – de manière plus radicale, quoique plus candide. Par exemple, l'essor des productions littéraires et artistiques autour de dystopies numériques en science-fiction et dans la culture populaire se poursuit aujourd'hui. Par exemple, *Big Data Is Watching You!: A comic dystopia* de Bruce Hartman (2015), *Infomocracy* de Malka Older (2016) ainsi que plusieurs épisodes de la série britanico-étatsunienne *Black Mirror* de Charlie Brooker (2011-2017) explorent les retombées possibles d'une confiance totale en un mode automatisé de gouvernance et de contrôle de la vie privée et publique. Les discussions en ligne concernant la pertinence et les écueils de la numérisation massive participent aussi à ce débat¹⁴.

2.2 Pratiques de recherche

La science des données et ses techniques d'analyse actuelles produisent de nouvelles informations concernant les habitudes communicationnelles et comportementales en ligne. Affirmer qu'elles sont inutiles ou inadaptées à la recherche scientifique serait donc erroné. Néanmoins, mieux comprendre les limites et les biais de ces outils révélerait plus précisément le type de connaissance qu'ils peuvent produire. Leur application, dans le cadre de travaux scientifiques – ou, plus généralement, d'enquêtes concernant des phénomènes naturels et sociaux – qui ne prescrivent pas l'analyse d'unités échangeables ni n'astraignent les chercheur·e·s à une logique capitaliste, constitue plus qu'une simple transposition de méthodes d'un champ vers un autre. Généraliser l'utilisation des outils d'analyse de données dans les sciences humaines et sociales nécessite donc une analyse qui problématiser leurs principes opérationnels et leurs effets sur la connaissance produite.

Premièrement, les affirmations citées plus haut présupposent que nous croyions que nos bases de données peuvent être exhaustives. Dans cette perspective, produire des connaissances exactes à propos d'un phénomène donné signifie que toutes ses itérations ont pu être (ou seront) enregistrées et encodées. Ceci signifie que l'échantillonnage est non seulement exact et précis, mais aussi suffisamment objectif. En ce sens, l'analyste connaîtrait supposément la totalité *possible* des données. Cette remarque rappelle certaines objections formulées à l'encontre des approches réalistes de la recherche qualitative. En outre, selon les critiques de la science des données Craig Dalton et Jim Thatcher, qui ont analysé les processus de quantification du sujet, la complétude n'est pas seulement problématique en tant qu'objectif (*enregistrer* toutes les données existantes), mais aussi en tant que processus (*encoder* toutes les données existantes) et technique (*comment* les données sont encodées)¹⁵. En effet, « à mesure que la pleine expérience humaine est réduite en une série d'octets, nous ne devrions pas simplement

nous soucier de savoir si ces octets sont plus opérationnels que d'autres, mais plutôt nous demander ce qu'être quantifié-e signifie, quelles possibilités s'ouvrent ou se ferment ainsi pour l'expérience¹⁶ ? ».

Si ces critiques révèlent des problèmes spécifiques à la production des données numériques, d'autres chercheur·e·s ont souligné les enjeux propres à l'interprétation de ces données. L'argument principal et récurrent au sein des études critiques des données (*critical data studies*) est que « les données doivent être interprétées et l'interprétation est assujettie aux biais humains¹⁷ ». De plus, les résultats de l'analyse automatisée de données numériques sont sujets à interprétation au même titre que ces données. Lorsque l'analyste organise et présente les résultats d'une analyse de données à un public qui ne possède pas de compétences en informatique, une partie du processus interprétatif se perd puisque le public non initié n'a pas connaissance des potentielles simplifications du contenu et des choix de présentation effectués lors de la mise en forme des résultats pour qu'ils soient communicables. D'autres chercheur·e·s insistent aussi sur le fait que tous les phénomènes ne sont pas représentables, et encore moins visualisables¹⁸. Par conséquent, il semble utile de questionner l'idée selon laquelle mobiliser des outils de quantification et d'encodage automatisés dans le but de produire des connaissances scientifiques serait exempt des biais de la perception humaine et éviterait les lacunes de méthodes « moins objectives ».

Deuxièmement, l'analyse basée sur des corpus de données massives s'appuie sur un principe de corrélation quantitative. « [Si] sonder des données massives (*Big Data*) révèle des relations et des récurrences que nous n'avions même pas pensé à chercher¹⁹ », alors mobiliser des bases de données massives constitue une méthode précieuse pour identifier des rapports entre plusieurs phénomènes qui restaient invisibles sous d'autres angles d'approche. Or, la causalité perd alors en importance en tant que pilier dans l'évaluation de la logique, de l'exactitude et de la pertinence des connaissances scientifiques. Des chercheur·e·s en études critiques des données ont caractérisé plusieurs problèmes que pose cette tendance²⁰. Par exemple, transposer des phénomènes qualitatifs en données quantitatives et établir des corrélations qualitatives sur la base des données quantitatives requièrent à la fois que l'on croie en la continuité et l'équivalence entre qualité et quantité et que l'on accepte qu'une information quantitative puisse, en un sens, devenir qualitative²¹. Par ailleurs, la collecte de données n'est pas aussi fréquente ni assidue selon la zone géographique et elle est influencée par des critères sociaux, raciaux et de genre qui peuvent injustement mener à des situations discriminatoires à l'encontre de certains individus ou groupes sociaux²². Ceci constitue un « fossé numérique »²³ et produit des angles morts dans les bases de données (« *data shadows* »)²⁴ qui nuisent à la représentativité des bases de données²⁵.

Troisièmement, des spécialistes en science des données et des chercheur·e·s en informatique ont posé des questions d'ordre éthique et méthodologique à l'égard des protocoles de collecte et d'analyse de données. Par exemple, des études ont interrogé la représentativité statistique²⁶. Ces études sont néanmoins motivées par l'intention d'utiliser les outils et les technologies de la discipline, intention qui empêche une véritable critique, avec un regard plus extérieur. Utiliser de telles méthodologies basées sur la mobilisation de données massives en vue de produire un savoir sur le monde nécessite une certaine confiance en ses outils et, surtout, en leur capacité à s'aligner avec nos partis pris éthiques. Par exemple, la production d'une quantité massive de données exige que les universitaires, les chercheur·e·s et les professionnel·le·s étudient les actes et les habitudes des utilisateurs d'appareils numériques, ce que plusieurs ont rapproché ou caractérisé de phénomène de surveillance généralisée²⁷. Malgré la diversité des méthodes et des rapports que chaque chercheur·e entretient avec son corpus de données numériques, l'examen des « biais algorithmiques » et autres écueils de la recherche basée sur les données massives demeure insuffisant, et leurs conséquences scientifiques et sociales font l'objet de débats jusqu'à ce jour²⁸.

2.3 Synthèse

L'analyse critique et les expériences de pensée concernant la pratique scientifique basée sur des données numériques ne sont pas nouvelles. Cet article se situe dans la lignée de ces travaux critiques, mais vise aussi à contribuer à ce champ de connaissances sur les données dans une perspective sémiopragmaticiste, suivant le modèle du signe développé par Charles Sanders Peirce. En effet, les discours critiques restent souvent exclus du champ des arguments légitimes et objectifs, comme l'attestent les accusations de positions réactionnaires, de parti pris anti-technologiques ou encore d'ignorance quant au fonctionnement de la science des données²⁹. La division entre les disciplines dites « dures » ou « scientifiques », basées sur des données empiriques, et les disciplines dites « molles » ou « non scientifiques », basées sur une matière linguistique, contribuent sans doute au maintien de cette impasse.

Cet article affirme cependant que lorsqu'il est question de la production des connaissances scientifiques, les discours importent. Considérer le jeu de la production et de la circulation des savoirs universitaires et non universitaires est crucial, mais il est aussi important de laisser les sciences sociales et les approches linguistiques examiner le contexte des avancées scientifiques et la manière dont elles sont décrites, surtout lorsqu'elles sont vouées à influencer les méthodes de recherche en sciences humaines

et sociales. Par exemple, les discours qui vantent le potentiel lucratif lié à la marchandisation des données, mis de l'avant par les entreprises qui vendent des algorithmes, influencent non seulement les consommateur·trice·s et les représentations sociales par le biais des médias de masse et socionumériques, mais informent aussi la recherche, car les chercheur·e·s ne sont immunisé·e·s ni contre ces discours médiatiques ni contre les modes universitaires, la pression des pairs et le fétichisme technologique.

Cet article vise donc à faire dialoguer ces discours et à mesurer les enjeux relatifs à la pertinence scientifique, l'exactitude empirique et les effets de l'utilisation des données massives en recherche en regard de leurs fondements épistémologiques et éthiques. En utilisant la terminologie complexe de la sémiotique peircienne, nous voulons favoriser la compréhension d'un problème existant, évaluer les différentes critiques formulées ci-dessus et contribuer à la discussion. Enfin, nous devons souligner que l'argument développé par cet article s'applique principalement aux sciences informatiques modernes. Bien qu'ils soient intéressants pour de futures recherches, le paradigme et les concepts propres à la computation quantique ne font pas partie de notre objet d'étude.

3. Un modèle sémiopragmaticiste des données massives

3.1 Qu'est-ce qu'une donnée?

Les données numériques résultent de l'action d'outils tels que des capteurs (par exemple, en imagerie médicale) ou de l'interaction entre un sujet humain et une interface informatique. Cet article s'intéresse davantage au second cas de figure, mais considère les deux cas lorsqu'il s'agit de la visualisation, de l'analyse et de l'application des données à des usages sociaux. En termes simples, nous décrivons le circuit des données de la manière suivante : les clics et les actions des individus sur une interface sont encodés sous forme de données (essentiellement, des séries de 0 et de 1), qui sont transmises afin d'être entreposées dans une base de données. Une partie de ces données sera « forée » (procédure d'extraction ou *data mining* en anglais) et analysée de manière à ce que les résultats servent à informer les actions en ligne subséquentes. Le circuit de la production et de la circulation des données est diachronique, étendu dans le temps, mais sa vitesse d'exécution est si grande qu'à échelle humaine, il semble simultané.

Bien que ce soient encore des humains (plus précisément des spécialistes en programmation informatique) qui produisent de nouveaux algorithmes en amont et qui interprètent les résultats du traitement des données en aval, la procédure d'extraction est de plus en plus souvent automatique. En effet, les algorithmes sont autonomes et

peuvent, dans plusieurs cas, modifier le processus de traitement des données au fil du temps (apprentissage machine). Sur la base des modèles corrélatifs (*correlative patterns*) tirés de la base de données, des signaux précis et contextualisés sont envoyés aux individus via leurs interfaces. Ce cycle de production, de circulation et d'évolution des données numériques implique une influence réciproque entre, d'un côté, les utilisateurs ou les analystes et, de l'autre côté, les programmes ou les algorithmes. Ce phénomène pourra être mis en relation avec le concept d'affordance, sur lequel nous ne reviendrons pas davantage dans le cadre de cet article³⁰.

Il vaut la peine de mentionner que tous les individus qui interagissent avec des données numériques, comme les utilisateur·trice·s de technologies informatiques, les programmeur·euse·s, les analystes de données, les législateur·trice·s et les employé·e·s des compagnies privées n'entretiennent pas le même rapport à ces données. Leurs motivations, leurs intérêts et leurs niveaux de compréhension des langages informatiques diffèrent, ce qui influence l'application et l'évolution de ces outils de collecte et d'analyse de données.

3.2 Cadres théorique et méthodologique

Cet article aborde son objet d'étude et questionne les prémisses de l'usage de données massives par le biais d'un cadre méthodologique et théorique spécifiquement peircien. La pensée sémiotique et philosophique de Charles Sanders Peirce constitue une doctrine, c'est-à-dire un ensemble de principes, plutôt qu'une doctrine religieuse³¹. Elle fait figure de charnière dans le domaine de ce que Thomas Sebeok appelle la « science intégrée de la communication »³², qui s'intéresse à la constitution et à la circulation des codes. Les catégories conceptuelles du pragmatisme de Peirce dépendent directement de leur utilisation et de leur potentiel explicatif : il ne s'agit donc pas d'un cadre théorique structuraliste ou idéaliste. Le pragmatisme vise plutôt à étudier la mise en place et l'évolution des « habitudes » sémiotiques³³, lesquelles sont des modèles comportementaux, des pensées ou des actes basés sur des conventions sémiotiques.

En raison de l'importance des actions et des processus de production de sens dans le pragmatisme de Peirce, son modèle du signe est triadique. Un signe est, pour lui, une association d'un « representamen » avec un « objet » qu'il représente d'une certaine manière en fonction d'un « interprétant »³⁴. Cette relation triadique définit le processus de la sémiose, lequel représente, pour le dire simplement, le potentiel infini des associations et des dissociations entre les représentations, les références et les interprétations. Ce processus peut être résumé par la maxime *aliquid stat pro aliquo* (« quelque chose

tient lieu d'une autre chose ») de saint Augustin, ce qui veut dire que la sémiose est un processus à la fois actif et dynamique³⁵. En ce sens, cet article développe une approche sémiopragmaticiste des données massives puisque son objectif principal n'est pas seulement descriptif, mais aussi pragmatique, comme le montrent les questionnements éthiques et politiques mentionnés précédemment.

Peirce identifie trois niveaux de fonction de signes qui caractérisent le representamen – qualisigne, sinsigne, légisigne³⁶, l'interprétant – rhème, proposition, argument (EP 2, p. 172-173)³⁷ et la relation entre l'objet et le representamen – icône, index, symbole (MS [R] 491, p. 1-2)³⁸. Bien que nous ne définissions pas plus précisément chacune de ces catégories, il est tout de même important de mentionner qu'un signe ne se limite pas à un seul de ces niveaux de signification, celui-ci variant en fonction de l'interprétant. Ces trichotomies sont donc pragmatiques plutôt qu'essentialisantes puisqu'elles expriment les effets des signes sur des esprits particuliers dans des contextes particuliers. Donc, la sémiose est non seulement un phénomène potentiellement infini, mais aussi un processus non linéaire hétérogène, plutôt qu'un système organisé de manière auto-suffisante. Il est important de noter que ce modèle sémiotique et ce cadre épistémologique contrastent avec la sémiologie de Saussure, laquelle repose sur un schisme fondamental entre un « signifiant » et un « signifié », c'est-à-dire entre les mondes conceptuel et matériel – une fracture qui ne peut être palliée que par un certain parallélisme ou une certaine correspondance³⁹.

De plus, la sémiotique peircienne implique qu'un signe n'existe qu'*in actu* (CP 5.569)⁴⁰. Selon cette approche, un signe est donc la somme de ses effets. Une habitude sémiotique est une interruption de la sémiose et la stabilisation d'une signification à travers différents contextes. Dans le cadre de cet article, nous appréhendons la circulation des données et les processus de numérisation comme des moyens par lesquels des habitudes sémiotiques sont formées.

De plus, notre cadre épistémologique est complété par l'ajout de la distinction entre « processus » et « procédure ». Pour ce faire, nous nous appuyons sur la tradition de la philosophie processuelle⁴¹ et le concept de sémiotique procédurale. Bien que la philosophie processuelle soit largement comprise comme une approche « postmoderne », elle est cohérente avec la doctrine peircienne dans une certaine mesure. Le sémioticien John Deely la décrit d'ailleurs comme « précisément postmoderne » puisqu'elle a la capacité de problématiser les idées et les théories modernes⁴².

Ian Bogost distingue à ce titre l'analyse des processus (la manière dont les choses sont – *how things are*) de l'analyse des procédures (comment les choses fonctionnent – *how things work*)⁴³. Le niveau processuel se caractérise par un flux continu et dynamique d'événements dont les agents font l'expérience, une poïétique de l'évolution

de l'être à travers le changement. Près du concept de « devenir » développé par Gilles Deleuze et Félix Guattari⁴⁴, le concept de processus résiste à la réification et à l'essentialisation de l'être. Les significations évoluent à travers le temps, l'espace et en fonction des perspectives : les phénomènes se déroulent toujours *autrement*⁴⁵. Au contraire, le niveau procédural se définit par des opérations autoréférentielles utilisant des termes et une grammaire prédéfinie

Dans cet article, nous opérationnalisons la distinction entre la procédure et le processus afin de définir la numérisation et les procédures computationnelles et analyser les connaissances produites avec des données comme une sémiotique procédurale. Les processus sémiotiques ne peuvent pas être reproduits en termes machiniques, du moins pas dans un format d'encodage binaire. En d'autres mots, au niveau machinique, les résultats produits à partir des analyses de données sont issus uniquement de procédures sémiotiques. Cette prémisse signifie que nous ne pouvons analyser les procédures sémiotiques des ordinateurs que dans la mesure où nous, en tant qu'humains, comprenons les procédures des ordinateurs. Il serait d'autant plus osé de qualifier la sémiotique machinique de sémiose. Cependant, ce problème spécifique ne peut être traité dans le cadre de cet article.

Finalement, tout comme l'encodage de données empiriques qualitatives sous la forme de données numériques implique des critères de sélections spécifiques, le choix de la sémiotique de Charles Sanders Peirce comme méthodologie et cadre épistémologique pour modéliser les processus de signification implique certaines limites. Plus précisément, nous n'adopterons pas une perspective sociologique concernant l'utilisation de larges bases de données, nous n'expliquerons pas davantage les enjeux cognitifs en lien avec le traitement des données et nous ne développerons pas une critique politique de la quantification massive, bien que ces enjeux se situent en périphérie de notre critique. Ce choix souligne d'ailleurs la nature logique et abstraite du raisonnement humain. En effet, en référence à la pensée de Gilles Deleuze, le travail d'analyse sémiotique consiste à produire des signes sur d'autres signes en sélectionnant, schématisant et, sous plusieurs aspects, simplifiant notre expérience de la réalité pour exprimer quelques-uns de ses aspects⁴⁶. En d'autres mots, nous reconnaissons que notre modèle est un effort de symbolisation qui oriente nécessairement la compréhension de notre sujet.

3.3 En termes sémiotiques, les données numériques sont des symboles

Les données numériques sont produites, elles circulent, sont vendues, achetées, interprétées, etc., par des sujets humains ou par des machines (lesquelles ont été conçues

par des humains). En tant que telles, les données peuvent être décrites comme des signes (en fonction de la philosophie pragmatiste déjà mentionnée). Elles impliquent des procédures (au niveau machinique) et des processus (au niveau humain) de signification. À partir de notre définition de la production de données numériques dans le cadre d'analyse de données massives (Section 3.1), le circuit de la production, de l'entreposage et de l'utilisation des données peut être décrit à l'aide de la terminologie sémiotique et il peut ainsi être analysé. Il est important de noter qu'il s'agit d'une interprétation processuelle (sémosique) d'un phénomène procédural, un problème qu'il nous est impossible de dépasser étant donné la nature humaine et sémiotique de notre approche. Nous ferons un examen critique de la production de signes et de connaissances basées sur des données dans la section qui suit.

Selon la description du circuit des données mentionnée précédemment, les données ont une fonction symbolique. Lorsque nous considérons des représentations des données, il est possible de soutenir que les données sont des termes (le premier degré du symbole), c'est-à-dire des signes symboliques qui ne réfèrent pas à leur objet selon une référence intrinsèque et explicite. Les entités sémiotiques comme « tué » ou « creusé », par exemple, sont des signes qui ne *déclarent* pas leur sens, mais qui renvoient conventionnellement à celui-ci (MS 491, p. 9). En effet, Peirce fait la distinction suivante : « Les symboles se divisent en trois classes : les termes, qui appellent l'attention à des choses ou des quasi-choses ; les propositions qui déclarent des faits ; et des arguments, qui permettent de nous éclairer sur les connexions rationnelles des faits ou des faits possibles. »

Dans un sens descriptif, les signes collectés, utilisés et produits par l'encodage, le traitement algorithmique et l'apprentissage machine aident simplement à « attirer l'attention » sur des éléments enregistrables ou déjà enregistrés. Ils n'établissent pas de faits ou ne font pas en eux-mêmes de connexions puisque ce processus vient après qu'ils aient été stockés, forés, visualisés et analysés d'une certaine manière. Cela nécessite donc l'intervention de signes et d'interprètes de natures différentes – nous reviendrons sur ce point plus loin. Bien qu'affirmer cela ne soit pas révolutionnaire pour le champ de la programmation et des sciences informatiques (les 1 et 0 sont assez clairement des signes conventionnels), il semble tout de même important de le réaffirmer. Dans la section suivante, nous examinons la production et la circulation des données en utilisant le modèle des procédures sémiotiques numériques et les prémisses théoriques et conceptuelles établies jusqu'ici.

4. Description des dynamiques sémiotiques dans la production et la circulation des données

4.1 L'illusion d'iconicité : la donnée, signal brut?

Au moment où une interaction entre un individu et une interface est enregistrée et encodée, un signal est saisi et sa trace est ensuite entreposée. Du point de vue des personnes non initiées au domaine de l'informatique, ce qui a été saisi peut ressembler à une image (premier degré de l'icône), c'est-à-dire que la donnée représente son objet par pure similarité. Cependant, la donnée reste un signe encodé dans une langue diagrammatique et les individus plus instruits en matière d'informatique, comme les spécialistes en programmation, la concevraient plutôt comme un diagramme (second degré de l'icône). Il est possible d'expliquer cette perception d'icônicité par l'entreposage et le traitement des données. L'entreposage des données nécessite en effet leur décontextualisation et leur désaffectation. En d'autres mots, la production de traces numériques nécessite l'évacuation de leur contexte affectif et symbolique d'actualisation.

Par exemple, lorsqu'un-e internaute interagit avec une interface (un clic sur un item affiché sur une page web, etc.), les actions sont enregistrées comme des traces « signifiant » l'évènement (le clic) de manière non ambiguë. Cette relation apparemment iconique entre le signe et l'objet influence la perception des données, lesquelles ne sont pas vues comme des informations textuelles sélectionnées et réifiées à propos d'un évènement, mais plutôt comme du contenu immédiat et nécessairement signifiant. Dans le contexte des algorithmes de traitement automatisé, le système de recommandation de YouTube par exemple, cette illusion se manifeste par la croyance en la pertinence véritable des vidéos dans la section « suivante ». Selon cette croyance, les clics précédents signifient à la fois un évènement, une intention et le plaisir de l'internaute. Bien que peu d'individus soient réellement capables de comprendre comment les algorithmes traitent les données et leur assignent une valeur, ceux-ci orientent souvent la consommation et les choix de navigation sur internet⁴⁷.

Dans le cas où le présupposé de similarité iconique est erroné, alors le processus peut être qualifié de dégénération du signe (CP 5.73)⁴⁸. En d'autres termes, on retranche le caractère conventionnel du symbole (passage de la tiercéité à la secondéité) ainsi que le caractère référentiel pour ne considérer que la ressemblance (passage de la secondéité à la priméité). Donc, les données sont principalement considérées comme des icônes. Suivant les travaux du sémioticien Jean Fissette, nous pouvons croire que le processus de dégénérescence *naturalise* les données numériques⁴⁹. Puisque les données apparaissent comme non médiées, représentant *directement* l'action, l'illusion d'iconicité peut être

définie comme un processus de naturalisation. En tant que présentation immédiate de la réalité, il ne semble pas être nécessaire de considérer la référence du signe à son objet afin d'en comprendre le contenu. Ceci fait en sorte que les données sont comprises non comme le résultat d'une méthode de formalisation par laquelle le monde est médié, mais comme une présentation directe du monde surpassant toutes les autres formes de représentation. Selon cette manière de concevoir les données, celles-ci ne sont pas des signes, c'est-à-dire quelque chose qui représente quelque chose d'autre (*aliquid stat pro aliquo*), mais plutôt un miroir univoque et naturel de la « réalité ».

La critique de l'iconisme formulée par le sémioticien Umberto Eco soutient notre argument en montrant que le mode iconique du signe repose lui aussi sur une construction culturelle⁵⁰. En effet, les signes iconiques deviennent signifiants dans le cadre de références culturelles partagées qui informent « les normes scripturales [et autres]⁵¹ ». Ceci implique un processus de sélection des traits qui sont considérés comme pertinents, importants ou nécessaires afin de reconnaître les objets. Selon Eco, les similarités de l'icône ne sont donc pas naturelles puisque nous devons apprendre ce qui est culturellement considéré comme similaires en fonction des objets. Par exemple, nous pouvons identifier facilement le bouton « maison » (*home*) sur nos appareils électroniques puisque nous avons appris à reconnaître les critères pertinents de la représentation schématique d'une maison.

Cependant, le contexte des données numériques ajoute une autre dimension au problème de l'iconisme. Il est en effet nécessaire de repenser le paradigme sous-jacent et la définition de la similarité. Premièrement, dans ce contexte, l'iconicité ne s'appuie plus seulement sur la reconnaissance de certains traits culturellement informés, mais aussi sur des instructions spécifiques (programmation) et sur des grammaires conventionnellement stables (langages informatiques). L'iconicité authentique est donc hautement improbable puisque le fait même de prendre les données pour des icônes présuppose un processus continu de symbolisation. Deuxièmement, la collecte de ces traces mises en forme dans un code binaire transforme ces *termes* (interprétation ouverte) en des *unités discrètes d'information* (système fermé). Cette transformation renforce la distinction entre les procédures et les processus déjà mentionnée. Les outils informatiques ne nous permettent pas de capter des sources d'événements d'interprétation, mais plutôt d'enregistrer des opérations logiques performées par des algorithmes dans le but d'obtenir des résultats particuliers.

En d'autres mots, lors de la phase initiale d'encodage du circuit des données, *les signes d'événements interprétatifs sont présentés en unités discrètes d'information traitées de manière procédurale*. Le fait que les données soient souvent réduites à leur fonction iconique – autant dans les discours encourageant leur utilisation massive que

dans la croyance scientifique en la représentativité des échantillons – peut être décrit comme une illusion précisément en raison de ce changement de niveau de signification (de processuels à procédural) et de fonctions des signes (par naturalisation du symbole vers l'icône).

4.2 (Re)construction des symboles : de la procédure d'encodage à la sémiose

Il est important pour nous, en tant que sémioticien·ne·s, et plus généralement en tant que chercheur·e·s, de comprendre ce qui se produit lorsque nous interprétons des données en nous basant sur des « répliques symboliques »⁵² de phénomènes réels produites à l'aide de procédures sémiotiques, ou plutôt sémantiques. Il est aussi nécessaire d'évaluer ce que les ordinateurs « apprennent » pour interpréter des données (apprentissage machine, apprentissage profond, etc.). Pour ce faire, nous devons les considérer comme des agents capables de traitements sémiotiques afin de mieux juger les effets de leurs procédures sur notre compréhension des processus signifiants et d'interprétation. Cela est particulièrement pertinent dans la mesure où les informations et les résultats que nous (en sciences humaines et sociales) manipulons visent souvent à affecter les groupes humains et/ou leurs manières de penser à propos des problèmes collectifs pragmatiques. Qu'est-ce que la conception commune des signes produits par les nouveaux outils de collecte et d'analyse des données dit de notre rapport aux données dans le contexte de la recherche et de notre compréhension des limites du concept de signification?

Afin de répondre à cette question, nous nous tournons vers la deuxième partie du circuit. Lorsque les données sont « forcées » pour être analysées, elles sont recontextualisées et réaffectées d'une valeur référentielle en fonction de critères humains d'analyse ou de fonctions algorithmiques prédéfinies. Cela signifie que lorsque la référentialité est réintroduite, les données sont (re)présentées et considérées en tant que signes indexicaux. Donc, les connaissances produites à l'aide de corrélations, que nous avons identifiées dans la revue de littérature, se basent plutôt sur des *inductions qualitatives* spécifiques aux interprétants indexicaux (ce qui est appelé, en termes peirciens, dici-signes propositionnels ou propositions). Autrement dit, le traitement automatisé des données permet l'induction de résultats exprimant des états de fait.

Cependant, la communication de résultats présuppose la réintroduction du sens symbolique. Bien que la recontextualisation des données puisse être supervisée par des agents humains, dans les recherches basées sur l'analyse de données massives, les machines effectuent de plus en plus souvent cette étape. Conséquemment, l'interprétation devient monosémique et prédéterminée, ou du moins pré-orientée, par le langage in-

formatique et les fonctions algorithmiques. Cette capacité à définir clairement les règles et les procédures servant au traitement de l'information est évidemment un avantage de la science des données et constitue l'un des principaux arguments en faveur de son efficacité à produire des résultats exploitables. Toutefois, il s'agit aussi de sa faiblesse : les humains doivent encore remplir des « vides interprétatifs »⁵³ lors de la lecture des résultats de l'analyse. Les analystes ou les scientifiques doivent donc (re)construire des symboles à partir des résultats produits procéduralement. En d'autres mots, *nous agissons en fonction de dynamiques sémiotiques procédurales comme si elles étaient équivalentes à des processus sémiotiques*.

De plus, l'usage fréquent des analyses de données dans le but d'établir non seulement des schèmes de probabilité, mais aussi d'affirmer des faits, nous permet de soutenir que les interprétants des données réaffectées et recontextualisées agissent comme des arguments plutôt que comme des propositions. Cependant, la production de lois nécessite des déductions qualitatives. Selon une perspective peircienne, cela produit une aberration sémiotique : l'objet d'un argument est « un signe selon une loi », plus précisément « la loi que le passage de ce genre de prémisses à ce genre de conclusions tend vers la vérité » (EP 2, p. 296). De plus, l'interprétant d'un signe ne peut pas être plus complexe que son representamen (EP 2, p. 296). En questionnant la fiabilité des résultats produits de cette manière, nous identifions ainsi qu'il y a un problème avec la volonté de créer des significations symboliques en fonction de signes inductifs et indiciaires, spécialement lorsque ce processus se déroule lors de la seconde phase de ce que nous appelons le cycle de la manipulation de la référence et du contenu contextuel.

Cela signifie aussi, sur le plan pragmatique, que les signaux, qui après avoir été traités de cette manière sont renvoyés aux internautes pour orienter leurs prochaines actions en ligne, sont intentionnels, c'est-à-dire motivés. De plus, puisque nous avons déjà montré que le traitement et l'analyse des données numériques entraînent un remaniement des valeurs contextuelles, affectives et sémiotiques, nous pouvons dire que ces signaux ne sont pas neutres : ils sont nécessairement influencés par les valeurs qui gouvernent la constitution des algorithmes, des critères et des techniques de collecte, etc.⁵⁴. En termes peirciens, cela signifie que les interprètes (par exemple, les publics sur YouTube) risquent d'identifier les résultats de l'algorithme (par exemple, les suggestions de visionnement) comme nécessairement et objectivement vrais, c'est-à-dire comme des « faits » signifiant leurs goûts.

Cependant, la valeur de vérité de ces résultats reste seulement *possible* (quasi-proposition). Par exemple, il est facile de confondre les propositions « vous pourriez aimer » et « vous aimerez ». La pertinence des vidéos suggérées (de la section « suivante ») n'est pas établie, mais plutôt inférée par l'algorithme en fonction de critères qui sont

inaccessibles pour la plupart des individus visés. Si le statut numérique des données est mis de côté, alors la chaîne interprétative qui place des vidéos les unes à la suite des autres peut sembler logique et factuelle, de même que fiable. Cela illustre le phénomène de la prophétie autoréalisatrice : les internautes peuvent en effet aimer les vidéos qui leur sont recommandées, mais cela risque d'être attribué à la qualité de la recommandation plutôt qu'au simple fait qu'elles ont été recommandées.

4.3 Les données numériques fonctionnent comme des hypersymboles

Les principales métaphores utilisées dans les discours scientifiques et non scientifiques au sujet des données confirment ce double mouvement entre l'illusion de la similarité iconique et le recadrement de l'efficacité symbolique. En effet, historiquement, les données ont souvent été utilisées comme si elles étaient des signes naturels. Cependant, les techniques de collecte et d'encodage produisent des signes symboliques qui n'aident qu'à « attirer l'attention » sur des éléments enregistrables ou déjà enregistrés. Par ailleurs, nous avons identifié des processus de dégénérescence et de naturalisation des signes à travers les procédures de collecte, d'encodage et d'entreposage. Il semble alors pertinent d'expliquer la transformation que les données subissent lors du processus de naturalisation à l'aide du concept d'hypoicône. Puisque les données sont des signes qui expriment leur relation avec l'objet selon une analogie (à travers un code binaire) et en fonction d'une loi (la rationalisation de leur collecte et de leur classification), elles peuvent être caractérisées comme des métaphores.

Dans le sens peircien du terme, les métaphores sont des signes iconiques qui n'expriment pas seulement une analogie entre des parties de différents objets, mais impliquent aussi le potentiel spontané de signification symbolique parce qu'elles possèdent « l'ascendance de la tiercéité, soit la généralité⁵⁵ ». Comme nous l'avons vu plus haut avec la critique de l'iconisme de Eco, cela signifie que les métaphores utilisées pour décrire les données ne sont pas des tropes neutres : elles orientent plutôt pré-symboliquement la manière dont nous conceptualisons, évaluons et interagissons avec les données. Bien que les liens que nous établissons entre le concept de métaphore de Peirce et la critique de l'iconisme de Eco soient importants pour notre discussion, il serait nécessaire d'en faire un commentaire beaucoup plus long afin d'en saisir toutes les ramifications, ce qui pourrait faire l'objet d'un autre article.

À travers les métaphores que nous utilisons pour parler des données, nous leur associons plusieurs qualités et habiletés dont la précision, l'objectivité et l'efficacité. Premièrement, les données sont régulièrement présentées comme une « force de la

nature » que nous devrions tenter de contrôler, ce qui contribue à naturaliser et objectiver davantage les données : « suggérer que le sens intrinsèque des données, comme les pépites d'or, est déjà là, dans l'attente d'être découvert, signifie qu'on l'on sépare l'interprétation de l'interprète et de sa subjectivité⁵⁶. » La métaphore géologique de la « mine de données » illustre comment les données, une fois dénuées de leur fonction symbolique de terme, sont interprétées comme de la matière brute extraite directement du monde matériel⁵⁷. Ces métaphores de matière brute encouragent à se représenter les données comme de la matière naturelle non raffinée⁵⁸.

Cependant, si les imposantes bases de données nous arrivent toujours déjà « cuites » (*baked*)⁵⁹, il semble aussi que la plupart de ces métaphores naturalisent un mode sémiotique spécifiquement capitaliste. Par exemple, lorsque les données sont comprises comme « une source de richesse », s'inscrivant ainsi dans la logique de la consommation et de la rareté⁶⁰, elles jouent un rôle dans le renforcement de la valeur positive associée à la vitesse d'analyse, à l'accumulation de ressources et à la course aux nouvelles informations.

De plus, la force évocatrice du concept de donnée lui-même peut être considérée comme une conséquence de sa nature métaphorique. À la fois le latin *data* (littéralement « chose donnée » ou « ce qui est donné ») et le grec *δεδομένα* (de *δίδωμι*, « donner, offrir ») impliquent une dynamique de communication directe et une stabilité des objets. Les deux termes sont des participes passés, connotant ainsi une action (de donner) achevée, même si le terme grec implique potentiellement une relation transcendantale (un cadeau des dieux aux humains). Le latin fournit au concept de donnée une base plus matérielle – comme l'illustre *Data*, ouvrage d'Euclide qui traite des prémisses de la géométrie. Ce dernier sens transparaît dans le terme anglais *data*, lequel a commencé à être utilisé pour parler des données numériques seulement au milieu du XX^e siècle⁶¹.

Plus significativement, les métaphores utilisées pour décrire les données ont une fonction importante. Pour les internautes, elles confèrent un sens aux agrégats d'informations anecdotiques et aux technologies plutôt complexes et opaques. Elles sont spécialement utiles pour les personnes non initiées à l'informatique (qui ne peuvent coder ou lire les codes). Les métaphores servent aussi à rendre signifiant ce qui resterait autrement a-signifiant et à lier des connaissances aux phénomènes afin de stabiliser leurs sens et ainsi rendre possible la formulation d'informations exploitables. En ce sens, les métaphores au sujet des données, en tant que signes pré-symboliques conventionnels, ne devraient pas être réduites à des productions discursives en marge des avancées scientifiques en informatique, mais devraient plutôt être considérées comme des déclencheurs de recherches scientifiques.

Cela est d'autant plus significatif que les jeux de données de plus en plus vastes doivent être formalisés, manipulés et visualisés. La quantité importante de données produites à travers les enregistrements et les encodages de masse peut être décrite comme hypersymbolique, c'est-à-dire comme une production « excessive » de symboles⁶². Autrement dit, l'ampleur des effets des procédures machiniques sur la perception et l'interprétation de la réalité est telle qu'elle crée un nouveau filtre de significations symboliques et le recourt à ce filtre pour interpréter et analyser les phénomènes du monde est de plus en plus fréquent.

En suivant la notion d'hyperréalité de Jean Baudrillard, nous qualifions ce phénomène d'hypersymbolisme. Un hypersymbole est plus qu'un signe conventionnel puisqu'il vise à *définir* et à *précéder* complètement la référence à son objet. Comme dans la théorie de Baudrillard où la carte devient signifiante indépendamment du territoire qu'elle représente, les bases de données constituent une couche hypersymbolique à partir de laquelle il est possible d'extraire du sens sans avoir besoin d'une référence externe de l'« origine » matérielle des données. Les données deviennent alors des objets en elles-mêmes, justifiant ainsi leur utilisation pour la production des discours et la construction de vérités au sujet de phénomènes réels – parfois produits par les données elles-mêmes.

5. Pistes de recherche

Cet article a présenté et opérationnalisé un modèle des procédures et des processus de production de sens dans le contexte des analyses de données numériques, afin de critiquer les effets de ce type de production sémiotique sur la production de la connaissance. La science des données, ses technologies et les discours qui les entourent convoquent souvent une définition de la donnée comme une présentation immédiate de la réalité. Certain·e·s affirment que les données numériques cessent d'agir comme une méthode de formalisation permettant la médiatisation du monde – un mode de représentation parmi d'autres – pour devenir une présentation directe du monde qui surpasserait tous les autres modes de représentation. Ce type de discours opère une naturalisation des signes symboliques qui attribue une importante valeur de vérité au reflet de la réalité produit à l'aide de signes binaires.

Or, notre modèle indique que loin de constituer une simple procédure, la numérisation et l'analyse automatisée de données numériques effectuent une hypersymbolisation de la réalité (les données étant des symboles), tout en naturalisant les signes produits par des humains. En termes peirciens, nous avons expliqué comment les données sont perçues et valorisées en tant que signes diagrammatiques. En manipulant

le contenu symbolique et référentiel des signes numériques, la sémiotique procédurale de la production des données a des conséquences processuelles : l'analyse de données finit par contribuer à l'élaboration du sens au niveau symbolique d'après des signes prétendument « naturels ». De plus, nous avons montré que les techniques discursives et les arguments d'autorité, tels que l'usage de tropes et de métaphores, renforcent ces tendances à la naturalisation et à la réorganisation du sens.

Afin de conclure notre discussion et pour problématiser les effets de ces dynamiques sémiotiques, nous pouvons poser trois questions pragmatiques. En effet, si le « *Big Data* » apparaît comme un « nouveau paradigme de vérité » ou un « régime de vérité » spécifique⁶³, alors modéliser les processus de production de sens à l'œuvre dans les procédures de numérisation et d'analyse des données numériques nous a montré en quoi un discours analytique et critique à leur sujet est non seulement possible, mais aussi nécessaire pour poser des questions épistémologiques, éthiques et politiques relatives à ces technologies. En formulant trois questions dans les sections qui suivent, nous indiquerons des pistes de recherche sémiotique au sujet de la science des données et des analyses de données numériques.

5.1 Quel type de connaissance ce régime sémiotique produit-il?

Notre revue de littérature a mis en évidence en quoi certaines applications de la science des données confondent la causalité et la corrélation, que la manière dont la quantité des données elle-même en vient à sembler suffisante pour justifier un argument et en quoi les signes symboliques deviennent obsolètes à travers un processus de naturalisation de la signification. Les progrès de l'élaboration et de la classification algorithmiques de l'information influencent la manière dont nous produisons des connaissances, menons des projets de recherche scientifique et appréhendons la notion de phénomène réel⁶⁴. Cet article a porté un regard critique sur la manière dont la « révolution des données »⁶⁵ peut sédimenter et orienter incorrectement les définitions du vrai, du factuel et du référentiel, et ce, aux niveaux autant pratique que théorique.

Malgré l'appel théâtral d'Anderson à la « fin de la théorie », les théories peuvent nous aider à comprendre les enjeux reliés au calcul informatique de pointe et à la prise de décision algorithmique. Cet article a montré que même un mode de production de connaissance inductif ne peut pas se passer de production de modèles, notamment à travers des visualisations iconiques, des tropes et des métaphores. La modélisation scientifique, en tant que stratégie organisationnelle, doit être considérée de façon explicite et critique au sein du processus d'évaluation des résultats de recherche. À ce

titre, la caractérisation des fonctions sémiotiques des données est bien une méthode de formalisation dont les fondements idéologiques et discursifs ne peuvent pas être minimisés.

Dans le contexte des nouvelles méthodes de recherche se basant sur des corrélations de données individuelles pour en induire un sens, peu importe l'hétérogénéité des ensembles de données, les analyses sémiotiques ont un rôle à jouer : elles permettent de cerner les modifications en ce qui a trait à la relation objet-représentation (le mode sémiotique des données), à la référence (ce de quoi la donnée tient lieu) et à l'agent-e de l'interprétation (pour qui) tout au long du circuit de la production et de la circulation des données.

5.2 Comment ce régime influence-t-il la définition du sujet et de la subjectivité humaine?

Les langages machiniques reposent principalement sur un encodage binaire. En effet, si les spécialistes en programmation utilisent des « langues » informatiques variées et des plus complexes, les commandes à l'intérieur de la machine fonctionnent toujours sur une base binaire. Comme nous l'avons montré précédemment, cette logique implique une réduction de la signification à un format (apparemment) diagrammatique, dans un paradigme procédural. En généralisant ce type de formalisation téléologique pour l'analyse de types de phénomènes variés, on risque de confondre la procédure d'encodage de l'information avec l'apparente homogénéité et commensurabilité des données (qui résultent de l'encodage). Affirmer qu'une subjectivité diagrammatique et procédurale est possible découle logiquement de la croyance selon laquelle l'expression diagrammatique sied à toutes les manifestations du sujet.

Si cette remarque peut s'appliquer à l'expression de la subjectivité (les agent-e-s s'expriment en produisant et en interprétant des données) et à la subjectivation (la production de données est constitutive du sujet, de sa vitalité et lui donnent accès à l'action collective), de récents travaux portant sur les dynamiques politiques de notre condition algorithmique proposent plutôt d'analyser notre devenir-machine⁶⁶. Ce devenir signifie non seulement une dépendance grandissante aux technologies numériques pour l'organisation de nos vies, mais aussi un certain « esclavage » ontologique⁶⁷. Certain-e-s chercheur-e-s défendent aussi l'idée selon laquelle la gouvernance algorithmique façonne un type de sujet qui lui est spécifique⁶⁸. Enfin, dans ce cadre d'analyse critique, des études récentes et des travaux de spéculation théorique examinent la fabrique biopolitique de la subjectivité posthumaine⁶⁹.

Ainsi, il paraît important de marquer la distinction entre, d'une part, une croyance en la possibilité de communiquer avec une interface et de façonner par ce biais la subjectivité humaine (sémiose machinique) et, d'autre part, le fait de problématiser cette croyance de communication avec une interface et notre capacité à façonner la subjectivité humaine (utilisation des outils d'analyse des données numériques informée idéologiquement). En ce sens, on pourrait qualifier l'horizon de l'identification ou de fusion avec la machine de leurre, de prophétie autoréalisante. Nous soutenons qu'il serait plus juste de dire que nous adaptons nos modes de subjectivation et notre subjectivité aux discours disséminés par le régime de vérité entourant les données massives. Ce débat nous conduit vers des considérations politiques.

5.3 Quel type de mise en commun et quel type de société la sémiotique procédurale prescrit-elle?

Cette question ne concerne pas l'ensemble des utilisations et justifications possibles de l'usage des données massives, mais plutôt les prémisses épistémologiques et certaines des orientations de la science des données contemporaine. Si un nombre suffisant d'agent·e·s (humain·e·s) en position de pouvoir soutenaient la quantification massive et la prédiction des comportements tels que la science des données contemporaine les permet, les processus décisionnels dans la sphère politique sembleraient devenir une simple question de calcul algorithmique. Ceci signifie que l'action politique risquerait d'être fondée sur une croyance en une vérité apolitique – croyance déjà profondément ancrée dans certains cercles transhumanistes et posthumanistes⁷⁰. Ce raisonnement laisse penser que l'idéologie elle-même pourrait être évacuée de nos sociétés ou même neutralisée puisqu'elle serait objectivable : les données donnent à voir ni plus ni moins que ce qu'il se passe et les résultats d'analyses automatisées n'ont aucun lien avec les croyances de qui que ce soit.

Cependant, si la politique se base sur des relations symboliques et affectives, suivant les prémisses de la philosophie de Peirce, alors l'occultation de ce phénomène social intrinsèquement processuel ne pourrait, en aucun cas, mener à une forme de communauté telle que nous l'entendons. L'ambiguïté, la contradiction et le changement à travers le temps sont essentiels aux relations collectives et au politique. Dans cette perspective, aspirer à un système social parfait, administré numériquement, semble dissimuler des processus discursifs et idéologiques de totalisation et d'universalisation au-delà de l'hégémonie, que nous pourrions apparenter à une manifestation autoritariste. Si cette affirmation peut renforcer les perspectives développées par la littérature

dystopique mentionnée plus haut (2.1), des recherches sémiotiques subséquentes pourraient examiner en quoi des conventions sémiotiques peuvent être « autoritaires ».

6. Conclusion

Cet article a montré en quoi les données sont des signes manipulés par une multiplicité d'agent·e·s au cours de leur production, de leur circulation et de leur interprétation. Nous avons soutenu que l'analyse des discours qui orientent notre perception et notre usage des données massives est tout aussi importante que la modélisation des processus et des procédures de production de sens, dans l'optique de comprendre quels types de connaissance, de subjectivité et de société la science des données promeut ou néglige. Nous avons suggéré que si certains *usages* et applications techniques des données constituent des formes de résistances et de subversions (par exemple, l'organisation et la visualisation de données médicales et l'utilisation de plateformes en ligne pour créer des communautés), ces stratégies n'affectent ni ne résolvent les problèmes intrinsèques à la *nature* des données numériques en tant que signes. Afin de problématiser la reproduction des illusions, des angles morts et des structures hégémoniques à travers l'usage des techniques propres à la science des données, il est nécessaire de dépasser l'usage subversif (pour la protection des données contre une utilisation marchande, par exemple). À celui-ci doit s'ajouter la production de contre-discours critiques par la publication et la pratique scientifiques qui visent un jugement plus approfondi, nuancé et argumenté des apports et limites de la science des données.

Bibliographie

- ANDERSON, Chris, « The end of theory: The data deluge makes the scientific method obsolete », *Wired*, 2008. En ligne : <<https://www.wired.com/2008/06/pb-theory/>> (consulté le 21 février 2020).
- ANDREJEVIC, Mark, « Big Data, Big Questions: The Big Data Divide », *International Journal of Communication*, vol. 8, 2014, p. 1673-1689. En ligne : <<https://ijoc.org/index.php/ijoc/article/view/2161/1163>> (consulté le 21 février 2020).
- BENNETT, Colin J., Kevin D. HAGGERTY, David LYON & Valerie STEEVES, *Transparent Lives: Surveillance in Canada*, Edmonton, Athabasca University Press, 2014.
- BERENDT, Bettina, Marco BÜCHLER & Geoffrey ROCKWELL, « Is it research or is it spying? Thinking-through ethics in Big Data AI and other knowledge sciences », *Künstliche Intelligenz*, vol. 29, no 2, 2015, p. 223-232.
- BOGOST, Ian, « Process versus procedure », communication présentée lors de la quatrième conférence internationale du Whitehead Research Project, Claremont, Californie, 2-4 décembre 2010. En ligne : <<http://bogost.com/downloads/Bogost%20-%20Process%20vs.%20Procedure.pdf>> (consulté le 21 février 2020).
- BOLLIER, David & Charles M. FIRESTONE, *The Promise and Peril of Big Data*, Washington, The Aspen Institute, 2010. En ligne : <https://assets.aspeninstitute.org/content/uploads/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf> (consulté le 21 février 2020).
- BONENFANT, Maude, André MONDOUX, Marc MÉNARD & Maxime OUELLET, « Big Data, médiation symbolique et gouvernementalité », dans E. Broudoux & G. Chartron (dir.), *Big Data-open data: Quelles valeurs, quels enjeux?*, Louvain-la-Neuve, De Boeck Supérieur, 2015, p. 31-41.
- BONENFANT, Maude, Fabien DUMAIS & Gabrielle TRÉPANIÉ-JOBIN (dir.), *Les pratiques transformatrices des espaces socionumériques*, Montréal, Presses de l'Université du Québec, 2017.
- BOYD, Danah & Kate CRAWFORD, « Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon », *Information, Communication & Society*, vol. 15, no 5, 2012, p. 662-679.
- BROOKER, Charlie, *Black Mirror*, Endemol, Zeppotron/House of Tomorrow, 2011-2017.
- BROWN, Ian & Christopher T. MARSDEN, *Regulating code: Good governance and better regulation in the information age*, Cambridge, The MIT Press, 2013.
- CACCAMO, Emmanuelle, *Imaginer les technologies de mémoire totale avec la science-fiction audiovisuelle occidentale (1990- 2016). Étude sémiotique, intermédiaire et technocritique des représentations de la mémoire personnelle*, thèse de doctorat, Montréal, Université du Québec à Montréal, 2017.

- CANADIAN INTERNET PUBLIC POLICY INTEREST CLINIC (CIPPIC), *On the Data Trail: How Detailed Information About You Gets Into the Hands of Organizations With Whom You Have no Relationship*, Rapport, Ottawa, 2006. En ligne : <<https://cippic.ca/sites/default/files/May1-06/DatabrokerReport.pdf>> (consulté le 21 février 2020).
- CASSIN, Barbara, « Intraduisible et mondialisation. Entretien réalisé par Michaël Oustinoff » *Hermès*, vol. 3, no 49, 2007, p. 197-204.
- CHASLOT, Guillaume et al., *Algotransparency*, 2018. En ligne : <<https://algotransparency.org/>> (consulté le 21 février 2020).
- CHRISTIAENS, Tim, « Digital subjectivation and financial markets: Criticizing social studies of finance with Lazzarato », *Big Data & Society*, vol. 3, no 2, 2016, p. 1-15.
- COVINGTON, Paul, Jay ADAMS & Emre SARGIN, « Deep Neural Networks for YouTube Recommendations », actes du 10th ACM Conference on Recommender Systems (RecSys'16), 2016, p. 191-198. DOI : 10.1145/2959100.2959190 (consulté le 21 février 2020).
- DALTON, Craig & Jim THATCHER, « What Does A Critical Data Studies Look Like, And Why Do We Care? », *Society and Space*, 2014. En ligne : <<https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care>> (consulté le 21 février 2020).
- DEELY, John, « The Green Book: The Impact of Semiotics on Philosophy », communication présentée lors du premier hommage annuel à Oscar Parland, Université d'Helsinki, décembre 2000, p. 12-13. En ligne : <http://www.commens.org/sites/default/files/news_attachments/greenbook.pdf> (consulté le 21 février 2020).
- DELEUZE, Gilles & Félix GUATTARI, *Mille plateaux : Capitalisme et schizophrénie II*, Paris, Minuit, 1980.
- DELEUZE, Gilles, *Francis Bacon. Logique de la sensation*, Paris, Seuil, 2002 [1981].
- DYCHE, Jill, « Big Data "Eurekas!" Don't Just Happen », *Harvard Business Review Blog*, 2012. En ligne : <<https://hbr.org/2012/11/eureka-doesnt-just-happen>> (consulté le 21 février 2020).
- ECO, Umberto, « Pour une reformulation du concept de signe iconique », *Communications*, no 29, 1978, p. 141-191.
- , *Lector in fabula. Le rôle du lecteur ou la Coopération interprétative dans les textes narratifs*, trad. de l'italien par M. Bouzaher, Paris, Le Livre de Poche, 1985 [1979].
- FISSETTE, Jean, « L'icône, l'hypoicône et la métaphore. L'avancée dans l'hypoicône jusqu'à la limite du non-conceptualisable », *Visual Culture*, no 14, 2009, p. 7-46.
- , « La photographie à l'infini. L'apport de la sémiotique de Peirce à l'étude de la photographie dans les médias », *Ocula*, no 15, 2014. En ligne : <<https://www.>

- ocula.it/files/OCULA-15-FISETTE-La-photographie-a-l-infini.pdf> (consulté le 21 février 2020).
- FLORIDI, Luciano, « Big Data and Their Epistemological Challenge », *Philosophy & Technology*, vol. 25, no 4, 2012, p. 435-437.
- FRIEDERICI, Nicolas, Sanna OJANPERÄ & Mark GRAHAM, « The impact of connectivity in Africa: Grand visions and the mirage of inclusive digital development », *Electronic Journal of Information Systems in Developing Countries*, vol. 79, no 2, 2017, p. 1-23.
- GALLOWAY, Alexander, « Are some things unrepresentable? », *Theory, Culture & Society*, vol. 28, no 7-8, 2011, p. 85-102.
- GARCIA, Megan, « Racist in the Machine: The Disturbing Implications of Algorithmic Bias », *World Policy Journal*, vol. 33, no 4, 2016, p. 111-117.
- GIBSON, James J., « The Theory of Affordances », dans R. E. Shaw & J. Bransford (dir.), *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, Hillsdale, Lawrence Erlbaum Ass., 1977, p. 67-82.
- GITELMAN, Lisa, « *Raw Data* » *Is an Oxymoron*, Cambridge, The MIT Press, 2013.
- GRAHAM, Mark, « Big Data and the end of theory? », *The Guardian*, 2012. En ligne : <<https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory>> (consulté le 21 février 2020).
- HALES, David, « Lies, Damned Lies and Big Data », *Aid on The Edge of Chaos. Rethinking International Cooperation in a Complex World*, 2013. En ligne : <<https://aidontheedge.wordpress.com/2013/02/01/lies-damned-lies-and-big-data/>> (consulté le 21 février 2020).
- HARTMAN, Bruce, *Big Data is watching you! A comic dystopia*, Philadelphie, Swallow Tail Press, 2015.
- HERN, Alex, « Google: 100,000 lives a year lost through fear of data-mining », *The Guardian*, 2014. En ligne : <<https://www.theguardian.com/technology/2014/jun/26/google-healthcare-data-mining-larry-page>> (consulté le 21 février 2020).
- ILIADIS, Andrew & Federica RUSSO, « Critical data studies: An introduction », *Big Data & Society*, vol. 3, no 2, 2016, p. 1-7.
- KITCHIN, Rob & Tracey P. LAURIAULT, « Small data in the era of Big Data », *Geo Journal*, vol. 80, no 4, 2015, p. 463-475.
- KITCHIN, Rob, « Big Data, new epistemologies and paradigm shifts », *Big Data & Society*, vol. 1, no 1, 2014, p. 1-12.
- KRAMPEN, Martin, Klaus OEHLER, Roland POSNER, Thomas A. SEBEOK & Thure von UEXKÜLL (dir.), *Classics of Semiotics*, New York, Springer, 1987.

- KURZWEIL, Ray, *Humanité 2.0: la bible du changement*, trad. de l'anglais par A. Mesmin, Paris, M21 éditions, 2007.
- LAZZARATO, Maurizio, « La machine », *Transversal*, 2006. En ligne : <<https://transversal.at/transversal/1106/lazzarato/fr>> (consulté le 21 février 2020).
- , *Signs and Machines. Capitalism and the Production of Subjectivity*, New York, The MIT Press, 2014.
- LEWIS, Paul, « Fiction is outperforming reality": How YouTube's algorithm distorts truth », *The Guardian*, 2018. En ligne : <<https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth>> (consulté le 21 février 2020).
- LORINO, Philippe, « Vers une théorie pragmatique et sémiotique des outils appliquée aux instruments de gestion », Document de recherche, ESSEC, 2002. En ligne : <<http://pfleurance.hautetfort.com/list/seminaire-7-l-intelligence-strategique-en-gestion/4140115640.pdf>> (consulté le 21 février 2020).
- MAYER-SCHÖNBERGER, Viktor & Kenneth CUKIER, *Big Data: A revolution that will change how we live, work, and think*, Londres, John Murray, 2013.
- MAZZOCCHI, Fulvio, « Could Big Data be the end of theory in science? A few remarks on the epistemology of data-driven science », *Science & Society*, vol. 16, no 10, 2015, p. 1250-1255.
- METCALF, Jacob & Kate CRAWFORD, « Where are Human Subjects in Big Data Research? The Emerging Ethics Divide », *Big Data & Society*, vol. 3, no 1, 2016, p. 1-14.
- MORE, Max & Natasha VITA-MORE, *The Transhumanist Reader: Classical and Contemporary Essays on The Science, Technology, and Philosophy of The Human Future*, New York, John Wiley, 2013.
- OLDER, Malka, *Infomocracy*, New York, A Tom Doherty Associates Book, 2016.
- ONLINE ETYMOLOGY DICTIONARY, « Data », En ligne : <<https://www.etymonline.com/word/data>> (consulté le 21 février 2020).
- PEIRCE, Charles S., *The Collected Papers of Charles Sanders Peirce*, vols 1-6 : C. Hartshorne & P. Weiss (dir.), Cambridge (MA), Harvard University Press, 1931-1935 ; A. W. Burks (dir.), vols 7-8, même éditeur, 1958.
- , *Manuscripts in the Houghton Library of Harvard University*, tel que rassemblé et annoté par Richard Robin (*Annotated catalogue of the Papers of Charles S. Peirce*), Amherst, University of Massachusetts Press, 1967.
- , *Écrits sur le signe*, trad. de l'anglais par G. Deledalle, Paris, Seuil, 1978.
- , *Essential Peirce: Selected Philosophical Writings. Vol. 2 (1893-1913)*, N. Houser & C. Kloesel (dir.), Bloomington, Indiana University Press, 1998.

- PIGLIUCCI, Massimo, « The end of theory in science? », *Science & Society*, vol. 10, no 6, 2009, p. 534.
- PRENSKY, Marc, « H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom », *Innovate*, vol. 5, no 3, 2009. En ligne : <<https://nsuworks.nova.edu/innovate/vol5/iss3/1/>> (consulté le 21 février 2020).
- PUSCHMANN, Cornelius & Jean BURGESS, « Big Data, Big Questions: Metaphors of Big Data », *International Journal of Communication*, vol. 8, 2014, p. 1690-1709. En ligne : <<https://ijoc.org/index.php/ijoc/article/view/2169/1162>> (consulté le 21 février 2020).
- RAHMAN, Rameez, « Big Data or Pig Data? », *Realm of the SCENSCI*, 2012. En ligne : <<https://thescensci.com/2012/12/14/big-data-or-pig-data/>> (consulté le 21 février 2020).
- RANCIÈRE, Jacques, *Le destin des images*, Paris, La Fabrique, 2003.
- RESNIK, David B, « Statistics, ethics, and research: An agenda for education and reform », *Accountability in Research*, vol. 8, 2000, p. 163-188.
- ROUVROY, Antoinette, « Big Data : de nouveaux outils à combiner aux savoirs établis et à encadrer par la délibération publique - Entretien avec Antoinette ROUVROY » *Statistique et société*, vol. 2, no 4, 2014, p. 33-41.
- ROUVROY, Antoinette & Thomas BERNIS « Gouvernamentalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation? », *Réseaux*, vol. 1, no 177, 2013, p. 168-173.
- SAUSSURE, Ferdinand de, *Cours de linguistique générale*, publié par C. Bailly & A. Sechehaye avec la collab. d'A. Riedlinger, éd. critique préparée par T. de Mauro, Paris, Payot, 1972 [1916].
- SCHRADIE, Jen, « Big Data not big enough? How the digital divide leaves people out », *Media Shift*, 2013. En ligne : <<http://mediashift.org/2013/07/big-data-not-big-enough-how-digital-divide-leaves-people-out/>> (consulté le 21 février 2020).
- SEBEEK, Thomas A., *Signs: An introduction to semiotics*, Toronto, University of Toronto Press, 2001 [1994].
- SHAMOO, Adil E. & David B. RESNIK, *Responsible Conduct of Research*, New York, Oxford University Press, 2003.
- SILVA, Selena & Martin KENNEY, « Algorithms, Platforms, and Ethnic Bias: An Integrative Essay », *Phylon*, vol. 55, no 1-2, 2018, p. 9-37.
- STEADMAN, Ian, « Big data and the death of the theorist », *Wired*, 2013. En ligne : <<https://www.wired.co.uk/article/big-data-end-of-theory>> (consulté le 21 février 2020).

- SWEENEY, Latanya, « Discrimination in Online Ad Delivery », 2013. En ligne : <<https://dataprivacylab.org/projects/onlineads/1071-1.pdf>> (consulté le 21 février 2020).
- SYMONS, John & Ramón ALVARADO, « Can we trust Big Data? Applying philosophy of science to software », *Big Data & Society*, vol. 3, no 2, 2016, p. 1-17.
- TYZLIK-CARVER, Magdalena, « Posthuman Curating and its Biopolitical Executions: The Case of Curating Content », *Executing Practices*, vol. 6, 2017, p. 171-189.
- WACHTER-BOETTCHER, Sara, *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*, New York, W. W. Norton & Cie, 2017.
- WHITEHEAD, Alfred North, *Process and Reality: An essay in Cosmology*, New York, Free Press, 1978 [1929].
- ZŁOTOWSKI, Jakub, Diane PROUDFOOT & Christoph BARTNECK, « More human than human: Does the uncanny curve really matter? », acte de communication du HRI2013 Workshop on Design of Humanlikeness, 2013, p. 7-13. En ligne : <<https://ir.canterbury.ac.nz/handle/10092/8698>> (consulté le 21 février 2020).
- ZOURABICHVILI, François, « Qu'est-ce qu'un devenir, pour Gilles Deleuze? », acte de communication, Lyon, Horlieu Éditions, 1997, p. 1-15. En ligne : <<http://horlieu-editions.com/brochures/zourabichvili-qu-est-ce-qu-un-devenir-pour-gilles-deleuze.pdf>> (consulté le 21 février 2020).

Notes

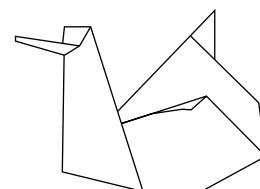
- 1 V. MAYER-SCHÖNBERGER & K. CUKIER, *Big Data: A revolution that will change how we live, work, and think*, Londres, John Murray, 2013 ; R. KITCHIN, « Big Data, new epistemologies and paradigm shifts », *Big Data & Society*, vol. 1, no 1, 2014, p. 1-12.
- 2 C. ANDERSON, Chris, « The end of theory: The data deluge makes the scientific method obsolete », *Wired*, 2008. En ligne : <<https://www.wired.com/2008/06/pb-theory/>> (consulté le 30 janvier 2018).
- 3 M. BONENFANT, F. DUMAIS & G. TRÉPANIÉ-JOBIN (dir.), *Les pratiques transformatrices des espaces socionumériques*, Montréal, Presses de l'Université du Québec, 2017.
- 4 C. ANDERSON, Chris, « The end of theory: The data deluge makes the scientific method obsolete », *loc. cit.*
- 5 *Idem.*
- 6 M. PRENSKY, « H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom », *Innovate*, vol. 5, no 3, 2009, p. 1. En ligne : <<https://nsuworks.nova.edu/innovate/vol5/iss3/1/>>.
- 7 C. ANDERSON, Chris, « The end of theory: The data deluge makes the scientific method obsolete », *loc. cit.*

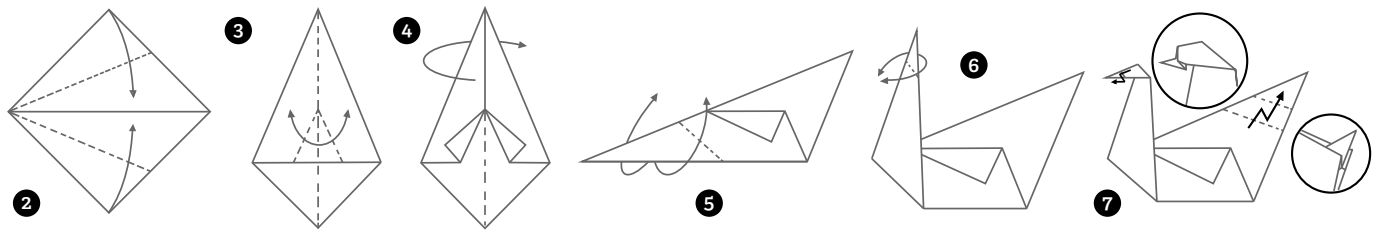
- 8 D. BOYD & K. CRAWFORD, « Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon », *Information, Communication & Society*, vol. 15, no 5, 2012, p. 662-679.
- 9 I. STEADMAN, « Big data and the death of the theorist », *Wired*, 2013. En ligne : <<https://www.wired.co.uk/article/big-data-end-of-theory>> (consulté le 21 février 2020).
- 10 J. DYCHE, « Big Data "Eurekas!" Don't Just Happen », *Harvard Business Review Blog*, 2012. En ligne : <<https://hbr.org/2012/11/eureka-doesnt-just-happen>> (consulté le 21 février 2020).
- 11 D. BOLLIER & C. M. FIRESTONE, *The Promise and Peril of Big Data*, Washington, The Aspen Institute, 2010. En ligne : <https://assets.aspeninstitute.org/content/uploads/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf> (consulté le 21 février 2020) ; L. FLORIDI, « Big Data and Their Epistemological Challenge », *Philosophy & Technology*, vol. 25, no 4, 2012, p. 435-437.
- 12 Canadian Internet Public Policy Interest Clinic, *On the Data Trail: How Detailed Information About You Gets Into the Hands of Organizations With Whom You Have no Relationship*, Rapport, Ottawa, 2006. En ligne : <<https://cippic.ca/sites/default/files/May1-06/DatabrokerReport.pdf>> (consulté le 21 février 2020).
- 13 Cf. E. CACCAMO, *Imaginer les technologies de mémoire totale avec la science-fiction audiovisuelle occidentale (1990- 2016). Étude sémiotique, intermédiaire et technocritique des représentations de la mémoire personnelle*, thèse de doctorat, Montréal, Université du Québec à Montréal, 2017 ; J. METCALF & K. CRAWFORD, « Where are Human Subjects in Big Data Research? The Emerging Ethics Divide », *Big Data & Society*, vol. 3, no 1, 2016, p. 1-14.
- 14 À titre d'exemple, voir R. RAHMAN, « Big Data or Pig Data? », *Realm of the SCENSCI*, 2012. En ligne : <<https://thescensci.com/2012/12/14/big-data-or-pig-data/>> (consulté le 21 février 2020).
- 15 C. DALTON & J. THATCHER, « What Does A Critical Data Studies Look Like, And Why Do We Care? », *Society and Space*, 2014. En ligne : <<https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care>> (consulté le 21 février 2020).
- 16 *Idem*.
- 17 J. SYMONS & R. ALVARADO, « Can we trust Big Data? Applying philosophy of science to software », *Big Data & Society*, vol. 3, no 2, 2016, p. 4.
- 18 Voir par ex. J. RANCIÈRE, *Le destin des images*, Paris, La Fabrique, 2003 ; A. GALLOWAY, « Are some things unrepresentable? », *Theory, Culture & Society*, vol. 28, no 7-8, 2011, p. 85-102.
- 19 J. DYCHE, « Big Data "Eurekas!" Don't Just Happen », *loc. cit.*
- 20 M. PIGLIUCCI, « The end of theory in science? », *Science & Society*, vol. 10, no 6, 2009, p. 534 ; D. HALES, « Lies, Damned Lies and Big Data », *Aid on The Edge of Chaos. Rethinking International Cooperation in a Complex World*, 2013. En ligne : <<https://aidontheedge.wordpress.com/2013/02/01/lies-damned-lies-and-big-data/>> (consulté le 21 février 2020) ; F. MAZZOCCHI, « Could Big Data be the end of theory in science? A few remarks on the epistemology of data-driven science », *Science & Society*, vol. 16, no 10, 2015, p. 1250-1255.
- 21 B. CASSIN, « Intraduisible et mondialisation. Entretien réalisé par Michaël Oustinoff » *Hermès*, vol. 3, no 49, 2007, p. 197-204 ; C. DALTON & J. THATCHER, « What Does A Critical Data Studies Look Like, And Why Do We Care? », *loc. cit.* ; R. KITCHIN & T. P. LAURIAULT, « Small data in the era of Big Data », *Geo Journal*, vol. 80, no 4, 2015, p. 463-475.
- 22 L. SWEENEY, « Discrimination in Online Ad Delivery », 2013. <<https://dataprivacylab.org/projects/onlineads/1071-1.pdf>> (consulté le 21 février 2020).
- 23 N. FRIEDERICI, S. OJANPERÄ & M. GRAHAM, « The impact of connectivity in Africa: Grand visions and the mirage of inclusive digital development », *Electronic Journal of Information Systems in Developing Countries*, vol. 79, no 2, 2017, p. 1-23.

- 24 M. GRAHAM, « Big Data and the end of theory? », *The Guardian*, 2012. En ligne : <<https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory>> (consulté le 21 février 2020).
- 25 J. SCHRADIE, « Big Data not big enough? How the digital divide leaves people out », *Media Shift*, 2013. En ligne : <<http://mediashift.org/2013/07/big-data-not-big-enough-how-digital-divide-leaves-people-out/>> (consulté le 21 février 2020).
- 26 Cf. D. B. RESNIK, « Statistics, ethics, and research: An agenda for education and reform », *Accountability in Research*, vol. 8, 2000, p. 163-188 ; A. E. SHAMOO & D. B. RESNIK, *Responsible Conduct of Research*, New York, Oxford University Press, 2003.
- 27 I. BROWN & C. T. MARSDEN, *Regulating code: Good governance and better regulation in the information age*, Cambridge, The MIT Press, 2013 ; C. J. BENNETT et al., *Transparent lives: Surveillance in Canada*, Edmonton, Athabasca University Press, 2014 ; B. BERENDT, M. BÜCHLER & G. ROCKWELL, « Is it research or is it spying? Thinking-through ethics in Big Data AI and other knowledge sciences », *Künstliche Intelligenz*, vol. 29, no 2, 2015, p. 223-232.
- 28 Cf. M. GARCIA, « Racist in the Machine: The Disturbing Implications of Algorithmic Bias », *World Policy Journal*, vol. 33, no 4, 2016, p. 111-117 et S. WACHTER-BOETTCHER, *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*, New York, W. W. Norton & Cie, 2017.
- 29 Voir par ex. A. HERN, « Google: 100,000 lives a year lost through fear of data-mining », *The Guardian*, 2014. En ligne : <<https://www.theguardian.com/technology/2014/jun/26/google-healthcare-data-mining-larry-page>> (consulté le 21 février 2020).
- 30 J. J. GIBSON, « The Theory of Affordances », dans R. E. Shaw & J. Bransford (dir.), *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, Hillsdale, Lawrence Erlbaum Ass., 1977, p. 67-82.
- 31 T. A. SEBEOK, *Signs: An introduction to semiotics*, Toronto, University of Toronto Press, 2001 [1994], p. 5.
- 32 *Ibid.*, p. 27-28.
- 33 C. S. PEIRCE, *Écrits sur le signe*, trad. de l'anglais par G. Deledalle, Paris, Seuil, 1978, p. 137.
- 34 *Ibid.*, p. 121.
- 35 T. A. SEBEOK, *Signs: An introduction to semiotics*, op. cit., p. 33.
- 36 C. S. PEIRCE, *Écrits sur le signe*, op.cit., p. 139.
- 37 L'abréviation EP suivie du numéro de volume renvoie à C. S. PEIRCE, *Essential Peirce: Selected Philosophical Writings. Vol. 2 (1893-1913)*, N. Houser & C. Kloesel (dir.), Bloomington, Indiana University Press, 1998.
- 38 L'abréviation MS suivie du numéro de section renvoie à C. S. PEIRCE, *Manuscripts in the Houghton Library of Harvard University*, tel que rassemblé et annoté par Richard Robin (*Annotated catalogue of the Papers of Charles S. Peirce*), Amherst, University of Massachusetts Press, 1967.
- 39 F. de SAUSSURE, *Cours de linguistique générale*, publié par C. Bailly & A. Sechehaye avec la collab. d'A. Riedlinger, éd. critique préparée par T. de Mauro, Paris, Payot, 1972 [1916], p. 158-159.
- 40 L'abréviation CP suivie du numéro de section renvoie à C. S. PEIRCE, *The Collected Papers of Charles Sanders Peirce*, C. Hartshorne & P. Weiss (dir.), vols 1-6, Cambridge (MA), Harvard University Press, 1931-1935 ; A. W. Burks (dir.), vols 7-8, même éditeur, 1958.
- 41 A. N. WHITEHEAD, *Process and Reality: An essay in Cosmology*, New York, Free Press, 1978 [1929].
- 42 J. DEELY, « The Green Book: The Impact of Semiotics on Philosophy », communication présentée lors du premier hommage annuel à Oscar Parland, Université d'Helsinki, 2000, p. 12-13. En ligne :

- <http://www.commens.org/sites/default/files/news_attachments/greenbook.pdf> (consulté le 21 février 2020).
- 43 I. BOGOST, « Process versus procedure », communication présentée lors de la quatrième conférence internationale du Whitehead Research Project, Claremont, Californie, 2-4 décembre 2010. En ligne : <<http://bogost.com/downloads/Bogost%20-%20Process%20vs.%20Procedure.pdf>> (consulté le 21 février 2020).
 - 44 G. DELEUZE & F. GUATTARI, *Mille plateaux : Capitalisme et schizophrénie II*, Paris, Minuit, 1980.
 - 45 F. ZOURABICHVILI, « Qu'est-ce qu'un devenir, pour Gilles Deleuze? », acte de communication, Lyon, Horlieu Éditions, 1997, p. 2. En ligne : <<http://horlieu-editions.com/brochures/zourabichvili-qu-est-ce-qu-un-devenir-pour-gilles-deleuze.pdf>> (consulté le 21 février 2020).
 - 46 G. DELEUZE, *Francis Bacon. Logique de la sensation*, Paris, Seuil, 2002 [1981], p. 62.
 - 47 G. CHASLOT *et al.*, *Algotransparency*, 2018. En ligne : <<https://algotransparency.org/>> (consulté le 21 février 2020) ; P. COVINGTON, J. ADAMS & E. SARGIN, « Deep Neural Networks for YouTube Recommendations », actes du 10th ACM Conference on Recommender Systems (RecSys'16), 2016, p. 191-198. DOI : 10.1145/2959100.2959190 (consulté le 21 février 2020) ; P. LEWIS, « Fiction is outperforming reality": How YouTube's algorithm distorts truth », *The Guardian*, 2018. En ligne : <<https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth>> (consulté le 21 février 2020).
 - 48 M. KRAMPEN *et al.* (dir.), *Classics of Semiotics*, New York, Springer, 1987, p. 218.
 - 49 J. FISSETTE, « L'icône, l'hypoicône et la métaphore. L'avancée dans l'hypoicône jusqu'à la limite du non-conceptualisable », *Visual Culture*, no 14, 2009, p. 7-46 ; J. FISSETTE, « La photographie à l'infini. L'apport de la sémiotique de Peirce à l'étude de la photographie dans les médias », *Ocula*, no 15, 2014. En ligne : <<https://www.ocula.it/files/OCULA-15-FISSETTE-La-photographie-a-l-infini.pdf>> (consulté le 21 février 2020).
 - 50 U. ECO, « Pour une reformulation du concept de signe iconique », *Communications*, no 29, 1978, p. 141-191.
 - 51 *Ibid.*, p. 160.
 - 52 P. LORINO, « Vers une théorie pragmatique et sémiotique des outils appliquée aux instruments de gestion », Document de recherche, ESSEC, 2002, p. 8. En ligne : <<http://pflurance.hautetfort.com/list/seminaire-7-l-intelligence-strategique-en-gestion/4140115640.pdf>> (consulté le 21 février 2020).
 - 53 U. ECO, *Lector in fabula. Le rôle du lecteur ou la Coopération interprétative dans les textes narratifs*, trad. de l'italien par M. Bouzaher, Paris, Le Livre de Poche, 1985 [1979], p. 27.
 - 54 S. SILVA & M. KENNEY, « Algorithms, Platforms, and Ethnic Bias: An Integrative Essay », *Phylon*, vol. 55, no 1-2, 2018, p. 9-37 ; M. GARCIA, « Racist in the Machine », *loc. cit.* ; S. WACHTER-BOETTCHER, *Technically Wrong*, *op. cit.*
 - 55 J. FISSETTE, « La photographie à l'infini », *loc. cit.*, p. 18.
 - 56 C. PUSCHMANN & J. BURGESS, « Big Data, Big Questions: Metaphors of Big Data », *International Journal of Communication*, vol. 8, p. 1690-1709. En ligne : <<https://ijoc.org/index.php/ijoc/article/view/2169/1162>> (consulté le 21 février 2020).
 - 57 D. BOYD & K. CRAWFORD, « Critical questions for Big Data », *loc. cit.* ; J. DYCHE, « Big Data "Eurekas!" Don't Just Happen », *loc. cit.* ; M. ANDREJEVIC, « Big Data, Big Questions: The Big Data Divide », *International Journal of Communication*, vol. 8, 2014, p. 1673-1689. En ligne : <<https://ijoc.org/index.php/ijoc/article/view/2161/1163>> (consulté le 21 février 2020) ; J. SYMONS & R. ALVARADO, « Can we trust Big Data? », *loc. cit.*
 - 58 R. KITCHIN, « Big Data, new epistemologies and paradigm shifts », *loc. cit.*, p. 2.
 - 59 L. GITELMAN, « *Raw Data* » *Is an Oxymoron*, Cambridge, The MIT Press, 2013.

- 60 C. PUSCHMANN & J. BURGESS, « Big Data, Big Questions », *loc. cit.*
- 61 ONLINE ETYMOLOGY DICTIONARY, « Data », En ligne : <<https://www.etymonline.com/word/data>> (consulté le 21 février 2020).
- 62 M. BONENFANT *et al.*, « Big Data, médiation symbolique et gouvernementalité », dans E. Broudoux & G. Chartron (dir.), *Big Data-open data: Quelles valeurs, quels enjeux?*, Louvain-la-Neuve, De Boeck Supérieur, 2015, p. 31-41.
- 63 R. KITCHIN, « Big Data, new epistemologies and paradigm shifts », *loc. cit.* ; A. ROUVROY & T. BERNIS, « Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation? », *Réseaux*, vol. 1, no 177, 2013, p. 168-173 ; A. ROUVROY, « Big Data : de nouveaux outils à combiner aux savoirs établis et à encadrer par la délibération publique - Entretien avec Antoinette ROUVROY », *Statistique et société*, vol. 2, no 4, 2014, p. 33-41.
- 64 D. BOYD & K. CRAWFORD, « Critical questions for Big Data », *loc. cit.*
- 65 R. KITCHIN, « Big Data, new epistemologies and paradigm shifts », *loc. cit.*
- 66 Cf. M. LAZZARATO, *Signs and Machines. Capitalism and the Production of Subjectivity*, New York, The MIT Press, 2014 ; T. CHRISTIAENS, « Digital subjectivation and financial markets: Criticizing social studies of finance with Lazzarato », *Big Data & Society*, vol. 3, no 2, 2016, p. 1-15 ; A. ILIADIS & F. RUSSO, « Critical data studies: An introduction », *Big Data & Society*, vol. 3, no 2, 2016, p. 1-7.
- 67 M. LAZZARATO, « La machine », *Transversal*, 2006. En ligne : <<https://transversal.at/transversal/1106/lazzarato/fr>> (consulté le 21 février 2020).
- 68 A. ROUVROY & T. BERNIS, « Gouvernementalité algorithmique et perspectives d'émancipation », *loc. cit.*
- 69 J. ZŁOTOWSKI, D. PROUDFOOT & C. BARTNECK, « More human than human: Does the uncanny curve really matter? », acte de communication du HRI2013 Workshop on Design of Human-likeness, 2013, p. 7-13. En ligne : <<https://ir.canterbury.ac.nz/handle/10092/8698>> (consulté le 21 février 2020) ; M. TYZLIK-CARVER, « Posthuman Curating and its Biopolitical Executions: The Case of Curating Content », *Executing Practices*, vol. 6, 2017, p. 171-189.
- 70 R. KURZWEIL, *Humanité 2.0 : la bible du changement*, trad. de l'anglais par A. Mesmin, Paris, M21 éditions, 2007 ; M. MORE & N. VITA-MORE, *The Transhumanist Reader: Classical and Contemporary Essays on The Science, Technology, and Philosophy of The Human Future*, New York, John Wiley, 2013.





Médias socionumériques alternatifs : étude sémiotique et rhétorique de diaspora*

Emmanuelle CACCAMO

Cygne noir, no 7, 2019 : « Algorithmes »

Résumé

Dans l'usage de la langue, les expressions « réseaux socionumériques » et « médias socionumériques » font directement référence aux plateformes hégémoniques et centralisatrices comme Facebook et Twitter. Il existe pourtant tout un ensemble de médias socionumériques « alternatifs », libres et distribués, moins connus du grand public. Cet article s'intéresse à cette seconde catégorie de technologies issues de l'informatique libriste. En prenant pour cas d'étude diaspora*, ce texte se penche sur la sémiotique et la rhétorique des médias socionumériques libres en les comparant aux médias socionumériques hégémoniques. Par une approche comparative, une première partie de l'article aborde la rhétorique entourant des médias socionumériques eux-mêmes (l'argumentaire qui sous-tend le projet et l'imaginaire symbolique véhiculé), tandis qu'une deuxième partie s'intéresse au design d'interaction de l'interface, c'est-à-dire aux signes mobilisés et aux stratégies rhétoriques employées par les plateformes interactives. L'article montre de quelles manières diaspora* tient un rôle « alternatif » vis-à-vis de Facebook et identifie quelques limites à ce statut.

Pour citer cet article

CACCAMO, Emmanuelle, « Médias socionumériques alternatifs : étude sémiotique et rhétorique de diaspora* », *Cygne noir*, no 7, 2019. En ligne : <<http://www.revuecygnoir.org/numero/article/caccamo-medias-socionumeriques-alternatifs>> (consulté le xx/xx/xxxx).



Cet article de *Revue Cygne noir* est mis à disposition selon les termes de la licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Pas de Modification 2.5 Canada.

MÉDIAS SOCIONUMÉRIQUES ALTERNATIFS : ÉTUDE SÉMIOTIQUE ET RHÉTORIQUE DE DIASPORA*

Lorsque l'on parle de « réseaux socionumériques » ou de « médias socionumériques », les plus hégémoniques d'entre eux, à l'exemple de Facebook et de Twitter, viennent immédiatement à l'esprit. Ces médias numériques connectés, qui incitent les utilisateurs et utilisatrices à créer et à partager des contenus, à collaborer et à interagir¹, présentent néanmoins une grande diversité². Les uns sont le fruit d'entreprises capitalistes et ont fait l'objet d'une quantité phénoménale de travaux universitaires³. Les autres, pour la plupart détachés d'intérêts financiers, émanent de la culture du logiciel libre, ou culture libriste, et intéressent encore trop peu les universitaires. Il existe pourtant plusieurs « médias socionumériques libres », à l'exemple de Friendica, Pleroma, Movim, GNU Social, diaspora*, Mastodon ou Needle.

L'idéologie libriste qui sous-tend leur existence définit de prime abord les médias socionumériques libres comme des médias « alternatifs » aux technologies numériques mercantiles et hégémoniques⁴. En effet, les premiers arguments mis de l'avant par les mouvements libristes renvoient généralement à (1) à la décentralisation (lutte contre l'hégémonie centralisatrice du privé), (2) au fait de partager le code source afin que chacun puisse en reconnaître sa logique et se l'approprier⁵ et (3) à l'idée que les utilisateurs et utilisatrices aient le contrôle sur leurs propres données⁶. L'idéologie libriste s'accompagne dans de nombreux cas d'un refus de la publicité et de la revente des données personnelles des internautes. Le thème des « communs numériques » est également convoqué.

Un certain nombre de médias libres, parfois créés en réaction aux plateformes privées, présentent à première vue des ressemblances en termes de design d'interaction et de fonction avec les médias hégémoniques. Par exemple, du point de vue du design d'interaction, diaspora* (2010) présente des ressemblances avec Facebook, de même que Mastodon (2016) à l'égard de Twitter. Afin de fournir aux internautes une expérience proche de ce à quoi ils sont habitués, plusieurs médias libres mobilisent des éléments de design employés par les médias socionumériques dominants. Or la plupart des médias socionumériques aujourd'hui hégémoniques ont été créés dans un but commercial. Pour cette raison, leur design d'interaction a été pensé pour être « addictif » dans le contexte de l'« économie des données personnelles et de la vie privée »⁷. Dès lors, quelles différences en termes de design d'interaction les médias socionumériques

libres présentent-ils à l'égard de leurs pendants hégémoniques? À quel point leurs éléments de design sont-ils différents?

La question du design fait évidemment partie de la réflexion dans le domaine libriste. En 2016, Tristan Nitot et Nina Cercy écrivent par exemple que le design est le prochain défi du logiciel libre :

Proposer des alternatives techniques est un premier pas essentiel, sans lequel rien n'est possible. Il s'agit désormais de passer un nouveau seuil : faire en sorte que les utilisateurs de ces alternatives ne soient pas seulement un public déjà sensible aux enjeux du logiciel libre et de la décentralisation, mais des utilisateurs lambdas qui utiliseront ces solutions non pas par dépit, mais par choix, parce qu'elles sont aussi ergonomiques et puissantes que les solutions proposées par Google ou Microsoft. Pour cela, se concentrer sur le design et l'ergonomie et réfléchir en termes d'expérience utilisateur est un défi majeur posé au logiciel libre⁸.

Pour donner un autre exemple, l'association française d'éducation populaire Framasoft élabore depuis 2019 un nouveau média appelé Mobilizon, « un outil libre et fédéré pour sortir nos événements de Facebook! »⁹. Le design d'interaction est un aspect très important du projet.

Le présent article s'inscrit dans le cadre d'une réflexion sur la sémiotique et la rhétorique du design d'interaction des médias socionumériques. Plus précisément, cette étude s'intéresse aux singularités sémiotiques et rhétoriques des médias dits « alternatifs ». En quoi leur sémiotique visuelle et leur rhétorique interactive humain-machine¹⁰ diffèrent-elles de celle des médias socionumériques dominants? Dans quelle mesure l'idéologie libriste, qui convoque une dimension éthique, se reflète-t-elle dans le design d'interaction et les stratégies rhétoriques mobilisées dans le dialogue humain-machine? Et en ce sens, en quoi peut-on qualifier les médias socionumériques libres d'« alternatifs »? Pour répondre à ce questionnement, j'analyserai le média socionumérique libre diaspora*, qui représente une alternative à Facebook et qui, selon les versions, compte un grand nombre d'utilisatrices et d'utilisateurs¹¹. D'autres médias alternatifs seront ponctuellement pris en exemple.

D'un point de vue méthodologique, le cadre théorique de cette étude mobilise les travaux en sémiotique visuelle ainsi qu'en rhétorique verbale, visuelle et interactive humain-machine¹². Tout comme les discours verbaux et les images ont recours à des opérations sémiotiques et rhétoriques, les médias socionumériques et leurs interfaces graphiques comportent des énoncés, des signes et des figures rhétoriques ou plus largement des stratégies rhétoriques visant à orienter, à influencer et à persuader les utilisateurs et utilisatrices d'opérer telle ou telle action gestuelle ou tel ou tel raisonnement. D'après Casey Boyle et ses collègues, il est devenu plutôt commun de dire que

les technologies numériques possèdent une dimension rhétorique¹³. Ian Bogost, chercheur en études vidéoludiques, a plus particulièrement forgé un concept que je définirai plus loin et sur lequel je m'appuierai : la « rhétorique procédurale »¹⁴. Celle-ci prend en compte l'interactivité des systèmes numériques dans les études rhétoriques. En sus des registres verbaux, visuels et sonores, elle renvoie aux effets persuasifs reliés au registre de la communication interactive humain-machine.

De manière plus traditionnelle, la rhétorique s'appuie sur des stratégies et différents types d'arguments et use également de figures ou tropes. Rappelons brièvement trois grandes définitions de la rhétorique : (1) la rhétorique renvoie à l'art de bien parler, aux règles de mise en forme de l'énoncé, notamment par l'emploi de figures d'ornementation (dimension esthétique) ; (2) il s'agit également d'un art de convaincre (art d'argumenter) et de persuader, parfois au moyen de tropes qui peuvent surprendre et étonner (dimension argumentative) ; (3) la rhétorique constitue enfin une méthode pour révéler certaines opérations idéologiques qui visent à persuader parfois par des moyens trompeurs (dimension critique)¹⁵. Notons que ce texte s'inscrit en phase avec la dimension normative et morale des études rhétoriques classiques, qui considèrent certaines pratiques rhétoriques comme plus éthiques que d'autres, et qui reconnaissent que certaines énonciations ont pour effet recherché de tromper leur énonciataire.

Deux niveaux seront pris en compte dans la suite du texte. Une première partie aborde la rhétorique accompagnant le projet diaspora*, l'argumentaire qui le sous-tend idéologiquement, et s'intéresse à l'imaginaire symbolique du média. De façon comparée, il s'agit de montrer en quoi ces éléments présentent des différences en regard de Facebook. Une deuxième partie s'intéresse au design d'interaction de l'interface de diaspora*, c'est-à-dire aux signes mobilisés et aux stratégies rhétoriques employées par la plateforme socionumérique. Comparer les stratégies rhétoriques d'interaction humain-machine de diaspora* à celles employées par les médias socionumériques hégémoniques permettra de mieux évaluer la dimension éthique du média libre.

1. Rhétorique et sémiotique entourant le projet diaspora*

Comme la plupart des entreprises capitalistes, les médias socionumériques hégémoniques mobilisent dans leurs stratégies marketing divers éléments rhétoriques en vue de persuader des bienfaits de leur utilisation et véhiculer un univers symbolique particulier¹⁶. Analysant Facebook et Twitter, la chercheuse Valérie Jeanne-Perrier identifie en premier lieu la figure du créateur comme élément rhétorique fondamental¹⁷. À l'exemple de Mark Zuckerberg et de Jack Dorsey, le fondateur d'une plateforme numérique est

masculin, relativement jeune, dynamique, cool, visionnaire et démiurge. Cette figure du fondateur-leader, mobilisant un ensemble d'attributs symboliques, concourt en partie à produire l'imaginaire des médias socionumériques, que d'aucuns qualifient de « révolutionnaires ». Pour désigner cet élément rhétorique, Jeanne-Perrier parle de « mythologie première »¹⁸.

Au contraire, les médias socionumériques libres développent généralement une figure de fondation toute différente : celle de l'association ou de la communauté militante. Le média diaspora* est un exemple paradigmatique. Créé en 2010 par Ilya Zhitomirskiy, Dan Grippi, Max Salzberg et Raphael Sofaer, quatre jeunes programmeurs du NYU's Courant Institute¹⁹, diaspora* est un logiciel libre non commercial qui permet d'échanger des contenus et de rester en contact avec des amis et des connaissances. Surnommé le « *Facebook killer* » (le « tueur de Facebook »), diaspora* n'est pas hébergé sur un serveur unique, mais sur une multiplicité de serveurs ou nœuds (appelés *pods*) de façon à ce qu'il en existe plusieurs versions locales, mais fédérées²⁰. Un nœud est « un serveur faisant tourner le logiciel diaspora*, il est connecté au réseau diaspora*²¹ ». diaspora* fait ainsi partie de la catégorie des « médias socionumériques distribués ». Emboîtant le pas à la campagne de « dégafamisation » du web menée en France par l'association Framasoft depuis 2014, l'organisme québécois FACIL²² a lancé en 2019 un nœud diaspora* sous le nom Parlote FACiLe²³. Cette version québécoise coexiste et communique avec de multiples autres versions, à l'exemple de son équivalent français Framasphère lancé en 2014²⁴. Dans le discours de Parlote FACiLe ou de Framasphère les fondateurs de diaspora* importent peu. À l'inverse, les organismes militants pour une informatique libre qui ont créé des versions locales forment les entités référentes (FACIL et Framasoft en l'occurrence). Ces entités symbolisent la lutte contre l'informatique hégémonique privée et centralisatrice ainsi que le combat en faveur des communs numériques.

On voit donc que, d'un côté, un média socionumérique hégémonique comme Facebook valorise la centralisation – valeur qui se vérifie tant sur le plan technique qu'au regard de la figure fondatrice, individuelle et centrale –, tandis que, de l'autre côté, les médias socionumériques libres valorisent la décentralisation et la collectivisation des activités d'entités différentes et autonomes.

L'argumentaire entourant le projet, le nom du média, le slogan, le logo et l'univers symbolique constituent d'autres éléments fondamentaux à prendre en compte dans l'analyse rhétorique. La rhétorique verbale du projet diaspora* met l'accent sur le contrôle des usagers et usagères sur leurs données et se présente comme « le réseau social où vous gardez le contrôle²⁵ » (ce slogan a été repris par Parlote FACiLe). L'argumentaire se déploie sur un ton didactique autour de trois thèmes fondamentaux reliés notamment

à la protection de la vie privée : (1) la décentralisation : les données des utilisateurs et utilisatrices ne sont pas conservées dans un serveur central et propriétaire ; (2) la liberté : l'utilisateur ou l'utilisatrice n'a pas l'obligation d'utiliser son identité administrative ; et (3) la confidentialité : les utilisateurs et utilisatrices demeurent propriétaires de leurs données, qui ne sont pas revendues, et opèrent les réglages qu'ils souhaitent en matière de confidentialité. Chacun des thèmes est détaillé et expliqué de façon pédagogique sur le site web du projet. Contrairement à Facebook dont le slogan fait appel au pathos (« Facebook vous permet de rester en contact avec les personnes qui comptent dans votre vie »), celui de diaspora* s'appuie sur le logos. Cette différence est fondamentale : sur un plan axiologique, elle traduit la façon dont l'énonciateur considère et reconnaît à juste titre son destinataire. Dit autrement, il est plus éthique de s'adresser à la raison (logos) qu'aux affects (pathos).

Du point de vue de la rhétorique visuelle, diaspora* se veut minimaliste. La plateforme est en ceci comparable à Facebook. Sur le site de présentation du projet diaspora*, on trouve peu d'images, une palette de couleurs sobre (blanc et noir), peu de symboles distinctifs. Cette rhétorique épurée est également reprise sur la plateforme d'interaction comme nous le verrons plus tard (ce qui n'est pas sans faire l'objet d'un débat au sein de la communauté de programmeurs). Hormis, la figure du pissenlit, qui redouble visuellement l'idée de dissémination diasporique (fig. 1) contenue dans le nom du média et le logo en forme d'astérisque (fig. 2), diaspora* n'est pas construit autour d'un récit spécifique ou d'un univers symbolique très élaboré. L'univers symbolique est minimaliste (fig. 3).



Figure 1. Éléments de la charte graphique et figure du pissenlit²⁶.



Figure 2. Logos et éléments graphiques de diaspora*.

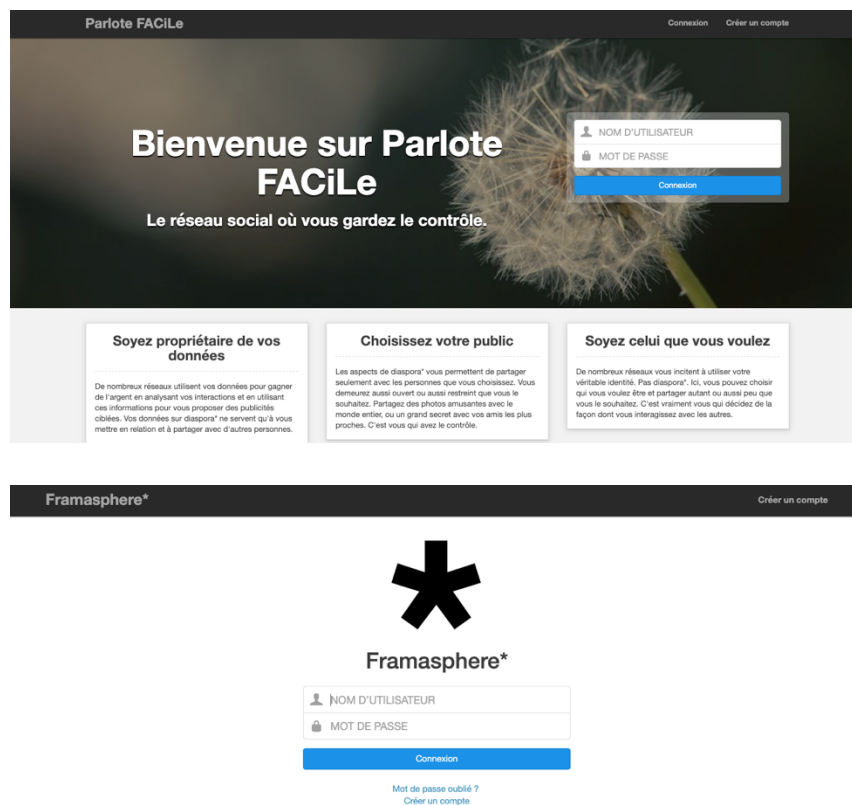


Figure 3. Minimalisme de l'interface de présentation et d'accès à diaspora* par Parlote FACiLe (en haut) et Framasphère (en bas).

Notons cependant qu'effectuer une comparaison entre les médias libres et les médias dominants présente quelques limites. En effet, les médias hégémoniques disposent d'un budget conséquent et d'équipes de salariés spécialisés en communication qui œuvrent à créer un univers marketing reconnaissable et attractif. De plus, ces salariés consacrent normalement l'entièreté de leur temps à diffuser l'idéologie de l'entreprise dans l'espace médiatique. D'après Jeanne-Perrier, les médias socionumériques hégémoniques opèrent un « lobbying sémiotique »²⁷. Par leurs stratégies marketing et la complicité plus ou moins consciente des médias traditionnels et des utilisateurs, ils sont parvenus à disséminer dans l'espace social un ensemble de signes distinctifs et reconnaissables en dehors même des plateformes numériques. Ces signes sont identifiables même pour les individus ne possédant pas de comptes sur les plateformes numériques en question. Les symboles du pouce bleu et blanc ou la punaise rouge (ou goutte inversée) de positionnement de Google Maps, qui circulent à divers endroits en ligne et hors-ligne, sont des exemples bien connus. La punaise de positionnement de Google Maps est par exemple réutilisée par différents groupes ou individus de manière à venir signifier l'idée du lieu ou de la localité.

D'un point de vue rhétorique, ces signes visuels *brevetés* qui renvoient à la charte graphique des marques déposées (Google Maps ou Facebook en l'occurrence) fonctionnent comme des antonomases visuelles. L'antonomase est à l'origine une figure de style verbale par laquelle on remplace un nom commun par un nom propre ou inversement. Cette figure est souvent employée dans l'univers de la consommation ; ainsi on demande souvent un Kleenex lorsqu'on veut un mouchoir. Le terme déposé Kleenex en vient à désigner, par extension, l'ensemble des mouchoirs en papier jetables. À force d'utilisation, certains noms de marques sont lexicalisés : le terme « frigo » dérivé de la marque Frigidaire en est un bon exemple. Le symbole de la punaise rouge agit comme une sorte d'antonomase visuelle lorsqu'il est utilisé en dehors de Google Maps pour signifier à lui seul le concept du positionnement, sans égard volontaire à la marque. Ce symbole acquiert alors une certaine autonomie de sens. Quant au symbole du pouce en l'air en général, il n'a évidemment pas été inventé par Facebook. Facebook s'est approprié un symbole socialement partagé et lui a conféré une fonction interactive dans l'espace numérique. La figure du pouce en l'air circulant dans l'univers social, en ligne et hors-ligne, agit elle aussi comme antonomase visuelle. D'une part, il signifie l'assentiment, mais d'autre part, il réfère aussi, directement ou indirectement, à la marque déposée Facebook. Comme l'écrit Umberto Eco en parlant d'antonomase visuelle, « [c]e mécanisme est basé sur des processus psychologiques d'identification [...], mais le processus d'identification n'est possible que grâce à des artifices rhétoriques qui rendent conventionnellement reconnaissable [...] le cas isolé proposé²⁸ ». Tous ces

signes conventionnellement identifiables, disséminés au point de devenir courants et banals, renforcent et entretiennent la puissance économique et symbolique des médias socionumériques auxquels ils se rattachent, et contribuent à assurer leur domination hégémonique. L'exercice d'un tel « lobbying sémiotique » garantissant à l'image de marque et aux symboles qui lui sont associés une exposition maximale dans l'espace public fonctionne donc, toutes proportions gardées, grâce au concours de trois éléments : (1) un budget marketing conséquent, (2) des stratégies de communication poussées, et (3) la force du nombre d'utilisateurs de la plateforme qui constituent autant de relais de dissémination idéologique. Il est évident que les médias libres peinent à rivaliser avec leurs pendants commerciaux compte tenu de ces facteurs déterminants.

En outre, je me permets de rappeler que plusieurs médias libres se sont constitués en réaction aux médias hégémoniques²⁹. Comme j'ai tâché de le montrer, ces derniers ont œuvré à déployer un univers symbolique reconnaissable et ont popularisé des signes, des éléments et des pratiques de design parfois plus anciens, à l'exemple du mot-clic repris et popularisé par Twitter. En termes sémiotiques, on pourrait dire que certains médias libres, en se constituant, se sont trouvés pris au piège par un répertoire symbolique préexistant et dominant, ou aux prises avec une guerre de symboles visuels et de pratiques de design. En somme, le lobbying sémiotique des plateformes socionumériques hégémoniques « colonise » notre imaginaire³⁰.



Figure 4. Exemple de saynète reliée à l'univers symbolique de Mastodon.

Dans le monde libre, il existe tout de même des médias constitués autour d'univers symboliques forts et singuliers. Alternative libre à Twitter, Mastodon présente par

exemple un univers symbolique jovial, teinté d'humour enfantin, mettant en scène des mastodontes – sorte de gros éléphants aujourd'hui disparus qui tranchent avec la « légèreté » de l'oiseau Twitter (fig. 4)³¹. Sur un ton humoristique, Mastodon a également développé un vocabulaire bien à lui : le « pouet » (« toot » en anglais) correspondant au « tweet » de Twitter.

Jusqu'ici, nous avons abordé différents éléments qui entourent le projet diaspora* : la figure de fondation, l'argumentaire entourant le projet, le nom du média, le slogan, le logo et en partie l'univers symbolique. Intéressons-nous à présent au design interactif des interfaces graphiques de la plateforme socionumérique d'un point de vue rhétorique et sémiotique.

2. Sémiotique et rhétorique procédurale du design d'interaction de diaspora*

Dans cette partie, je vais m'intéresser à la rhétorique interactive humain-machine de l'interface de diaspora* en la comparant avec celle des médias socionumériques hégémoniques. Plus généralement, certains chercheurs qualifient la dimension rhétorique des médias numériques de « procédurale ». Pour Bogost qui forge le concept :

[...] la rhétorique procédurale est la pratique de persuasion par le biais de processus en général et de processus computationnels en particulier. À l'instar de la rhétorique verbale, utile pour l'orateur et le public, et de la rhétorique écrite, utile à l'écrivain et au lecteur, la rhétorique procédurale est utile au programmeur et à l'utilisateur, au *game designer* et au joueur³².

Comme l'écrivent Maude Bonenfant et Thibault Philippette, « [l]a rhétorique procédurale vise à convaincre d'agir au travers de représentations et d'interactions reposant sur un système de règles, plutôt qu'au travers de mots, d'écrits, d'images ou d'images animées³³ ». Par différents mécanismes de persuasion inscrits dans le code source (les stratégies rhétoriques, les règles et les gestes interactifs à opérer par exemple), les usagers et usagères sont incités à se comporter de telle ou telle manière. Le concept de « rhétorique procédurale » doit être conçu comme un concept heuristique ; en complément d'autres régimes sémiotiques (régimes verbal, visuel, audiovisuel, etc.) exploités par les médias numériques, il met en lumière la dimension rhétorique du langage interactif humain-machine. Bien que le concept ait été développé dans le cadre des études vidéoludiques, je montrerai qu'il s'applique à l'objet d'étude de cet article. Dans le domaine des médias socionumériques, le design d'interaction lui-même présente une interaction cadrée, un répertoire d'actions et différents gestes interfacés incitant les usa-

gers à adopter des comportements spécifiques (« liker », téléverser des images, donner des informations personnelles, etc.).

Le design d'interaction d'un média socionumérique reflète en fait un certain nombre de décisions de conception qui ont été prises en amont ; et « [t]oute décision de design pousse potentiellement les usagers dans une direction ou une autre³⁴ ». Ces choix reposent sur différentes stratégies, elles-mêmes définies par différents cadrages économiques, techniques, sociaux et idéologiques. Certaines stratégies peuvent être plus éthiques que d'autres : comme en rhétorique dite classique, l'on peut différencier les stratégies éthiques des stratégies fallacieuses. Les stratégies éthiques respectent notamment le logos et se fondent sur le respect de la personne interlocutrice³⁵. Les stratégies fallacieuses, quant à elles, peuvent être sciemment élaborées pour tromper le destinataire.

En matière de design d'interaction, de la même manière qu'il existe des stratégies fallacieuses en rhétorique classique, il existe en effet des « stratégies obscures » (« *dark strategies* ») et des « modèles interactifs persuasifs fallacieux » (« *dark patterns* »)³⁶. Les stratégies obscures mises en place par de nombreux médias socionumériques hégémoniques renvoient à des méthodes fallacieuses de plus en plus connues visant à exploiter les tendances, les faiblesses et les biais psychologiques des utilisateurs et utilisatrices. Comme nous allons le voir, les médias socionumériques hégémoniques sont très friands de ces stratégies fallacieuses, car elles soutiennent en grande partie leur modèle économique. À l'inverse, on pourrait penser que leurs pendants libres œuvrent plutôt du côté des stratégies éthiques en ce qu'il se revendiquent être une « alternative ». De prime abord, on peut penser que la revendication non commerciale de ces médias alternatifs, la lutte pour la liberté liée aux outils numériques et le contrôle sur les données personnelles qu'ils proposent aux usagers et usagères s'accompagnent nécessairement d'un design adéquat qui traduit ces éléments sur le plan de la rhétorique interactive humain-machine. Nous verrons qu'avec diaspora* cette hypothèse ne se valide qu'en partie sur le plan des stratégies de captation de l'attention, mais qu'elle est confirmée sur le plan des stratégies relatives aux données personnelles.

Dans la suite de ce texte, j'analyserai la rhétorique procédurale de diaspora* par rapport aux pratiques fallacieuses exploitées par les médias socionumériques hégémoniques. J'aborderai différents travaux élaborés en sciences informatiques en nouant un dialogue avec les études sémiotiques et rhétoriques. Seront abordées les stratégies sémio-rhétoriques de captation de l'attention ainsi que celles relatives aux données personnelles. Je présenterai ensuite quelques-unes des stratégies fallacieuses en les liant à des « modèles interactifs persuasifs » à l'éthique douteuse. En dernier lieu, je discuterai du concept de rhétorique procédurale.

2.1 Stratégies sémio-rhétoriques de captation de l'attention

Les méthodes fallacieuses visant à exploiter les tendances, les faiblesses et les biais psychologiques des internautes sont actualisées par différents éléments de design de façon à capter le plus longtemps et le plus souvent possible l'attention des usagers et usagères. Par le biais de différentes procédures rhétoriques, les médias socionumériques cherchent à enraciner des *habitudes* fortes chez les internautes en s'appuyant parfois sur des stratégies de manipulation insidieuses et sournoises³⁷. Les notifications visuelles, sonores et haptiques instantanées et non planifiées³⁸, les « *likes* » et le défilement infini (*infinite scroll*) forment quelques-uns des éléments de design qui œuvrent au cœur des stratégies de création d'habitudes. Dans le cadre des médias socionumériques, l'habitude peut se former à partir d'un signal déclencheur pouvant mener à une action associée à un type de récompense variable³⁹. Pour s'ancrer, une nouvelle habitude doit être répétée à de multiples reprises et générer chez l'internaute une attente de récompense. L'on conviendra par ailleurs que certaines habitudes encouragées par certains médias socionumériques sont plus extrêmes que d'autres, certains designs étant sciemment conçus pour favoriser une dépendance⁴⁰. Dans ce domaine, le symbole de la cloche de notification utilisé par différentes plateformes n'est pas anodin. Relique du behaviorisme, la cloche constitue un élément fondamental de l'étude sur les réflexes conditionnés menée par Pavlov. Elle convoque en un certain sens un imaginaire pavlovien.

Ces éléments aujourd'hui considérés comme pouvant favoriser des habitudes extrêmes, voire malsaines, sont-ils présents sur la plateforme libre diaspora*? L'interface de diaspora* mobilise par exemple le symbole de la cloche de notifications. Sur la plateforme, les notifications parviennent à l'utilisateur en temps réel, donc de manière non planifiée. Ces notifications sont seulement visuelles et non auditives. L'utilisateur ou l'utilisatrice peut néanmoins faire le choix de couper les notifications provenant d'une publication particulière. Sur le symbole de la cloche, les notifications qui s'accumulent sont représentées par défaut par un chiffre dans un cercle rouge. Il est intéressant de noter que les usagers et utilisatrices peuvent modifier le thème de l'interface. En fouillant dans les paramètres, l'utilisateur peut choisir d'activer le thème de couleur de Framasoft. Le cercle de notification devient alors vert (fig. 5). Ce changement est loin d'être anodin, car le rouge, culturellement associé au danger, attire l'attention et augmente la vigilance⁴¹. Du point de vue de l'attention cognitive, une couleur froide, avec une plus faible luminance (sensation visuelle de luminosité d'une couleur), constituerait un signal chromatique moins distrayant et moins alarmant qu'un signal de couleur rouge. La signification de

certaines couleurs repose en grande partie sur une arbitrarité culturelle. Un signal chromatique vert, comme dans l'exemple qui suit, est probablement moins « agressif » pour plusieurs d'entre nous en Occident.

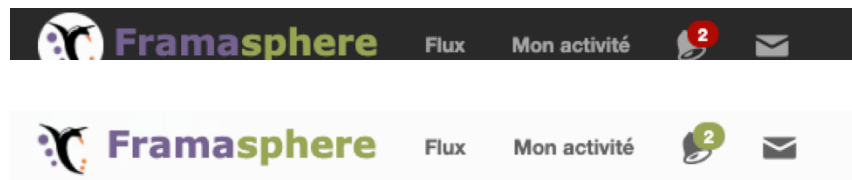


Figure 5. En haut le thème par défaut et en bas le thème Framasoft.

Par ailleurs, on observe que le chiffre de notifications augmente au fur et à mesure que les notifications s'accumulent. Cet élément est-il bien nécessaire? Certaines plateformes ont fait le choix de ne pas utiliser de quantifieur de notifications. Les activités notifiées sont simplement signalées par un marqueur visuel (un point par exemple). Un tel changement de couleur ou d'état du symbole suffit en effet à fournir un indice signalant à l'utilisateur que de nouvelles informations peuvent être consultées. Quant au « like », comme de nombreux médias socionumériques, diaspora* mobilise l'élément de design « J'aime » avec le symbole du cœur. La plateforme quantifie également le nombre de « j'aime » récoltés, de même que le nombre de fois où la publication a été partagée. On peut se demander s'il n'y aurait pas là aussi de meilleures pratiques de design à imaginer afin d'en limiter les effets négatifs⁴². La quantification est-elle fondamentalement nécessaire? De plus, la dimension non planifiée de réception des notifications, qui selon plusieurs études favorise un comportement compulsif, doit aussi être critiquée. Cette dimension est toutefois à nuancer avec diaspora* car (1) les notifications ne sont pas sonores ni haptiques et (2) il n'existe pas d'application pour téléphone – les usagers et usagères doivent passer par un navigateur web⁴³. Ces éléments limitent en un sens la possible distraction et éventuellement les comportements compulsifs alimentés entre autres par les sonneries et les vibrations intempestives liées aux notifications des téléphones.

L'interface de diaspora* convoque enfin le défilement infini. Il faut noter que cet élément de design préexiste aux médias socionumériques. Aussi, il est important d'identifier minimalement (1) le type de plateforme qui mobilise cet élément, si elle est commerciale et emploie un algorithme particulier à cet effet ; et (2) le type de terminal utilisé pour accéder à la plateforme (ordinateur, tablette, téléphone de dernière génération). Relativement aux effets du défilement infini, on peut formuler l'hypothèse selon

laquelle trois facteurs sont susceptibles d'induire un comportement compulsif chez les usagères et usagers :

- a) les contenus présentent peu d'informations textuelles (et donc plus d'images, favorisant ainsi une économie cognitive) ;
- b) les contenus sont filtrés par un algorithme commercial visant à capter l'attention ;
- c) les usagers accèdent aux plateformes en tout temps depuis leur téléphone tactile portable, en faisant défiler le contenu d'un coup de pouce.

Selon cette hypothèse, le défilement infini causerait plus de préjudices lorsque (a), (b) et (c) sont réunis. Bien qu'il faille élaborer des études comparatives sur la question, il me semble de prime abord erroné d'affirmer que l'expérience du défilement infini sur diaspora* est similaire à celles sur Facebook ou Instagram. Le média diaspora* ne mobilise pas d'algorithmes ayant pour visée de capter l'attention des utilisateurs pour revendre leurs données personnelles (facteur b), le flux de publication est organisé en fonction des mots-clés choisis sciemment par l'utilisateur, et il n'existe pas d'application pour téléphone portable (facteur c) (bien qu'il soit possible d'accéder à diaspora* depuis un navigateur web sur téléphone). Ceci dit, le défilement infini pourrait être repensé, en s'inspirant par exemple du design de la page. On peut supposer que le moteur de recherche Google d'Alphabet a conservé le design en forme de pages de manière à ce que les internautes n'aillent pas trop loin dans les résultats. Ce choix de design – qui bénéficie au modèle d'affaire de Google dans l'exemple que je viens de donner – ne pourrait-il pas inciter les internautes à passer moins de temps sur les plateformes des médias socionumériques ?

En somme, eu égard aux éléments mentionnés, diaspora* n'est pas tout à fait exemplaire en matière de captation de l'attention. En matière de rhétorique procédurale, les effets de certains éléments de design, à l'exemple du défilement infini, sont à relativiser, tandis que d'autres mériteraient probablement d'être repensés afin d'ancrer davantage le statut de média alternatif de la plateforme. Le média diaspora* se distingue toutefois *largement* de Facebook en ne proposant pas de pages commerciales qui attirent le chaland, ni de groupes ou de fenêtres de clavardage qui peuvent être très chronophages. Il ne propose pas de pages d'événements non plus. Ce minimalisme a pour effet de ne pas centraliser toutes les informations relatives au quotidien des individus (événements, groupes, etc.) sur une même plateforme. Autrement dit, à l'inverse de Facebook, diaspora* n'est pas pensé pour être un tiers-lieu, après la maison et le travail, comme s'efforcent de l'être les espaces commerciaux depuis les années 1980.

2.2 Stratégies sémio-rhétorique relatives aux données personnelles

Les plateformes numériques commerciales et hégémoniques mettent aussi en place des stratégies visant à influencer les usagers à baisser leur garde quant à la divulgation de leurs données personnelles. Dans le contexte actuel où s'observe une croissance de l'économie des données personnelles, l'idée pour bon nombre de plateformes numériques est de mettre à la disposition d'un maximum d'utilisateurs un ou des services leur permettant de récolter un maximum d'informations sur ceux-ci⁴⁴. Ce type de stratégies fallacieuses constitue une sous-catégorie des stratégies obscures : elles portent exclusivement sur la vie privée, on parle alors de « stratégies obscures portant atteinte à la vie privée » (« *privacy dark strategies* »)⁴⁵. Les chercheurs néerlandais Jaap-Henk Hoepman ainsi que Christoph Bösch et ses collègues ont en effet identifié différentes stratégies de design relatives au (non-)respect de la vie privée développées par les plateformes numériques⁴⁶. Hoepman propose une liste de stratégies éthiques tandis que Bösch et ses collègues, à partir des travaux de Hoepman, ont mis à jour une liste de stratégies fallacieuses. Ces stratégies ne renvoient pas toutes au design d'interaction, mais plus largement à la conception générale de la plateforme numérique. Ceci dit, pour fonctionner, certaines de ces stratégies nécessitent la mise en place de « modèles interactifs persuasifs fallacieux » (*dark patterns*) susceptibles d'être reproduits de plateformes en plateformes. Il s'agit de sortes de « tropes » propres au design d'interaction humain-machine dont je vais à présent exposer quelques exemples.

Comme je l'ai mentionné plus tôt, l'une des stratégies fallacieuses employée par certaines plateformes est de *maximiser* la collecte de données personnelles⁴⁷. Par le biais de différents modèles interactifs persuasifs il s'agit de collecter plus d'informations sur les utilisateurs que ce qui est réellement nécessaire pour utiliser la plateforme. Les auteurs citent en exemple un formulaire d'inscription comportant de nombreux champs à remplir, et bien que nombre d'entre eux puissent être facultatifs, leur seule présence incite les utilisateurs à livrer un très grand nombre d'informations personnelles. La stratégie inverse est de limiter la collecte d'information⁴⁸. Par exemple, pour être utilisée, la plateforme diaspora* ne requiert qu'un identifiant (du type monpseudo@framasphe.org), une adresse courriel et un mot de passe. Il est possible d'utiliser diaspora* uniquement avec ces trois données profilaires. Les paramètres offrent cependant la possibilité à qui le souhaite de remplir un profil plus complet : mots-clés reliés aux intérêts personnels (par ex. #cinéma, #politique), photo, prénom et nom, biographie, genre, lieu de résidence, date de naissance. Comme on peut le lire sur la plateforme, « [a]ucune de

ces données n'est obligatoire ; [les usagers sont] libre[s] de les fournir ou pas, qu'elles soient authentiques ou imaginaires ». La question de la liberté est bien au cœur de la plateforme libre. Néanmoins, on observe que les utilisateurs ont une certaine tendance à en dire plus sur eux qu'il n'est requis. Selon Mathieu Gauthier-Pilote de FACIL, en matière de confidentialité, « les données les mieux protégées sont encore celles qui ne sont jamais produites⁴⁹ ». Il va sans dire que la collecte de données effectuée par diaspora* est sans commune mesure avec les stratégies déployées par des plateformes comme Facebook, qui catégorisent et analysent toute interaction des usagers grâce à des algorithmes et des programmes d'apprentissage machine. Néanmoins, afin de limiter la collecte d'informations personnelles, il est légitime de se demander si les plateformes libres ne devraient pas aller jusqu'à éviter de proposer à leurs usagers de remplir des champs profilaires (genre, âge, etc.), et ce, même si ceux-ci sont facultatifs.

Dans le cadre des stratégies fallacieuses, la maximisation n'opère pas seulement en obligeant les usagers, elle les incite à délivrer d'eux-mêmes leurs informations personnelles. Il est toutefois important de préciser que ces « fallacies numériques » ne sont pas forcément mises en place de manière délibérée par les plateformes. Rappelons en effet qu'en logique et rhétorique classique, on fait la différence entre le paralogisme et le sophisme. Le premier est involontaire alors que le deuxième est sciemment produit. Comme le résume Christian Plantin, « [o]n se laisse prendre à ses propres paralogismes, on prend avec les sophismes ; c'est ce qui différencie l'imbécile et la crapule⁵⁰ ». Bien que l'analogie soit délicate à transposer littéralement aux modèles de design d'interaction humain-machine, on observe que certaines plateformes libres mobilisent des stratégies fallacieuses sans forcément en mesurer les conséquences. C'est, semble-t-il, le cas de Communecter, un média socionumérique libre français qui incite ses participant·e·s à délivrer un *très grand nombre d'informations* (biographie, nom, date de naissance, adresse courriel, adresse postale, numéro de téléphone, liens vers les différents profils des médias socionumériques, etc.)⁵¹. Certaines informations comme l'adresse postale sont par défaut publiques, visibles même par les individus n'ayant pas de compte. À l'inverse, sur diaspora*, le profil complet est par défaut privé, ce qui témoigne d'une meilleure pratique en termes de protection de la vie privée. En effet, dans sa liste de pratiques de design éthique, Jaap-Henk Hoepman privilégie non seulement de minimiser la collecte de données, mais aussi de cacher les informations personnelles⁵². Sur un plan rhétorique, il s'agirait d'afficher le moins possible de représentations verbales et visuelles des données personnelles, d'en limiter la traduction graphique de façon à ce que l'identité numérique visible des individus soit la plus minimaliste possible⁵³.

Obscurcir l'accès aux informations sur les pratiques de conservation, le traitement et la protection des données personnelles constitue un deuxième exemple de stratégies

fallacieuses poursuivies par les médias socionumériques identifiées par les chercheurs néerlandais. Cette stratégie de cadrage manipulateur⁵⁴ peut être mise en place grâce à un modèle visant à perdre l'utilisateur soucieux de sa vie privée dans un dédale de pages, et ce, jusqu'au découragement. Il peut aussi être question d'un modèle où la rhétorique de l'interface présente des éléments (un jargon, des boutons, des icônes) volontairement complexes ou ambigus. En termes rhétoriques, on peut alors parler d'abus de forme. Ce type de modèle a été appelé par Tim Jones « Privacy Zuckering », en référence à Mark Zuckerberg, le patron de Facebook, et aux stratégies employées par son entreprise. Selon Bösch et ses collègues, « [a]lors que le fournisseur de service va affirmer que les usagers ont un contrôle total sur les réglages de confidentialité, la présentation, la terminologie et l'expérience utilisateur vont hautement décourager les usagers à apporter des modifications [qui protégeraient leur vie privée]⁵⁵ ».

La stratégie d'obscurcissement peut fonctionner de pair avec celle visant à *tromper* les utilisateurs et utilisatrices grâce à différents signes verbaux et visuels connotant l'idée de sécurité. Bösch et ses collègues donnent l'exemple de certifications du type « norme de chiffrement militaire » (*military grade encryption*) apposées sur la page de l'interface. De telles allégations sont parfois invérifiables et reposent sur une stratégie d'autorité fallacieuse. Dans le cas de l'exemple cité, un imaginaire militaire est mobilisé en vue de persuader de la fiabilité du système. On parle alors de « *cryptowashing* »⁵⁶. Dans le même ordre d'idée, les interfaces utilisent aussi très souvent le symbole visuel du cadenas en vue de rassurer les usagers et de les persuader que leurs données sont bien protégées.

L'illusion de sécurité peut aller de concert avec l'illusion de contrôle pouvant être alimentée par le symbole visuel de l'engrenage que l'on retrouve sur plusieurs plateformes. L'engrenage renvoie à une métaphore mécanique de précision (celle des horloges notamment). Selon le contexte, cette métaphore a pour but de faire rapidement comprendre à l'internaute la fonction de l'engrenage, mais peut participer à donner la fausse impression à l'utilisateur ou à l'utilisatrice d'avoir tout en main pour régler le système. Or, en cliquant sur l'engrenage, on remarque que de nombreuses interfaces mobilisent la technique rhétorique du faux dilemme⁵⁷. Il s'agit par exemple de limiter un réglage à deux options possibles prédéfinies (alors qu'il pourrait y avoir plusieurs autres choix possibles) pour conduire l'internaute à une conclusion forcée. Cette stratégie profite évidemment à l'entreprise qui monétise les données collectées.

Le média diaspora* paraît quant à lui exemplaire en matière d'information et de contrôle sur les données. La plateforme informe les usagers du traitement et de la conservation de leurs données, permet très facilement aux usagers de télécharger leurs données et fournit des explications sur les instances qui peuvent y avoir accès (les ad-

ministrateurs de nœuds). La rubrique d'aide présente un discours fondé sur la franchise. À la question de savoir à quelles informations un administrateur de nœud peut accéder, on peut lire sur la plateforme diaspora* :

En bref : tout. La communication entre les pods est toujours chiffrée (en utilisant SSL et le propre chiffrement d'échange de diaspora*), mais le stockage des données sur les pods n'est pas chiffré. S'il le veut, l'administrateur de la base de données de votre pod (généralement la personne qui gère le pod) peut accéder à toutes les données de votre profil et à tout ce que vous avez publié (comme c'est le cas pour la plupart des sites web qui stockent des données d'utilisateurs). C'est pourquoi nous vous permettons de choisir le pod sur lequel vous vous inscrivez. Vous pouvez ainsi choisir un pod dont l'administrateur vous inspire confiance vis-à-vis de vos données. Utiliser votre propre pod garantit plus de confidentialité parce que vous avez le contrôle de l'accès à la base de données⁵⁸.

À l'égard des stratégies obscures portant atteinte à la vie privée, de nombreuses solutions ont été proposées pour contrer ces fallacies numériques relatives à la confidentialité. Par exemple le projet Privicons fournit des ressources pour un design plus respectueux de la vie privée⁵⁹. Alexandro Acquisti et ses collègues ont quant à eux répertorié un certain nombre de bonnes pratiques visant à contrer les biais cognitifs relatifs à la protection des données personnelles par la mise en place de représentations symboliques plus éclairantes et éthiques⁶⁰.

2.3 Modèles interactifs persuasifs fallacieux (*dark patterns*)

Comme nous venons de le voir à travers différents exemples, les stratégies de conception des plateformes numériques s'accompagnent de nombreux éléments de design et de modèles interactifs persuasifs plus ou moins éthiques.

Plus généralement, lorsqu'ils sont fallacieux, ces modèles interactifs baptisés « *dark patterns* » par Harry Brignull en 2010⁶¹ renvoient à des éléments génériques créés pour inciter les utilisateurs et utilisatrices à réaliser des actions involontaires et non désirées sur l'interface médiatique de façon à ce que ces actions soient bénéfiques pour les affaires de l'entreprise⁶². Brignull et les chercheur·e·s spécialistes du domaine ont documenté un certain nombre de ces modèles interactifs persuasifs fallacieux. Cela va du « Roach motel » au « Color trick » en passant par le « Dirt and Dust »⁶³ pour ne citer que ceux-là.

De la même manière qu'il existe un répertoire de stratégies argumentatives et d'arguments fallacieux dans la rhétorique du discours⁶⁴, il existe dans le cadre de la

rhétorique des médias interactifs une liste de stratégies obscures et de modèles interactifs persuasifs peu éthiques. À tout le moins cette liste est en train de se développer au fur et à mesure que des études se penchent sur la question. Les modèles interactifs persuasifs sont des tropes interactifs qui se reconnaissent par une structure et des effets communs⁶⁵. Pour reprendre la définition classique de Marc Bonhomme : les tropes ou « [l]es figures sont “de rhétorique”, en ce qu’elles constituent des procédés expressifs, investis par les sujets communicants, qui rendent les productions langagières davantage efficaces⁶⁶ ». Les figures peuvent remplir une fonction argumentative lorsqu’elles « [...] contribuent à la persuasion, en agissant sur les capacités de décision des [récepteurs] en vue de changer leurs comportements⁶⁷ ». Sans compter qu’elles peuvent détourner l’attention ou tromper. J’emploie ici le terme persuader pour désigner un processus qui s’adresse plutôt aux sentiments et aux croyances qu’à la raison. La manipulation visée par lesdits modèles interactifs est de persuader par tous les moyens, même irrationnels, en exploitant en particulier les biais cognitifs des individus.

Selon mes observations, diaspora* ne mobilise pas de tels modèles, ce qui concorde avec ses principes de bases. En n’employant pas ces tropes interactifs trompeurs, diaspora* fait preuve d’un souci éthique qui renforce son *ethos* de média alternatif. Néanmoins, chose rare, la plateforme libre fait appel au pathos lorsque l’usager souhaite fermer son compte. Contrairement à de nombreux médias socionumériques, diaspora* ne rend pas la fermeture de compte difficile : dans les paramètres, un bouton rouge « Fermer le compte » est clairement mis en évidence. La clôture du compte s’opère en deux étapes simples. Or, lorsque l’usager clique sur ce bouton, une fenêtre surgit : l’usager est interpellé par une interjection, l’image d’un petit chaton et un court texte explicatif (fig. 6) visent à persuader les individus de ne pas fermer leur compte⁶⁸.



Figure 6. Fenêtre de clôture de compte sur Framasoft.

Si elle en appelle aux sentiments et à une culpabilité (stratégie fondée sur l'affect), cette fenêtre surgissante n'emploie toutefois pas en bonne et due forme la fallacie du « fardeau de confirmation » (*confirmshaming*)⁶⁹. Brignull a en effet identifié un modèle selon lequel certaines plateformes essaient de culpabiliser l'utilisateur ou l'utilisatrice lorsque ceux-ci déclinent une option qui serait bénéfique à l'entreprise ou ferment un compte⁷⁰. Par exemple, en 2019 sur YouTube, l'internaute est fortement incité à se connecter à son compte (ou à en créer un). Une fenêtre surgissante et intempestive lui présente deux boutons cliquables dont les libellés ont été bien choisis : « Se connecter » ou « Abandonner » (fig. 7). Le bouton « Se connecter » est bleu tandis que l'autre est gris,

ce qui entame la présomption d'égalité des choix. Afin de faire disparaître la fenêtre, la personne qui ne souhaite pas se connecter est contrainte de cliquer sur le bouton « Abandonner » (le choix du terme « abandonner » est lourd de sens). Ce type de pratique déplace le fardeau sur l'individu en essayant de susciter un sentiment de culpabilité chez celui-ci⁷¹.

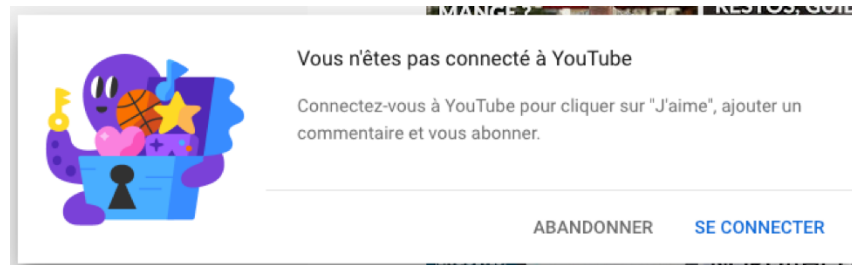


Figure 7. Ex. d'un « fardeau de confirmation » (*Confirmshaming*) employé par YouTube (2019).

Le modèle du « fardeau de confirmation » est cependant ambigu au sens où il peut convoquer un ton humoristique. Dans l'exemple de diaspora*, bien que le modèle s'apparente à ce trope, la plateforme fait d'abord appel aux sentiments des individus pour ensuite s'adresser à la raison des usagers. Aucun jeu rhétorique n'est cependant opéré sur les boutons cliquables comme c'est le cas dudit modèle.

*

Jusqu'ici, nous avons vu qu'en termes de rhétorique diaspora* présente de nombreuses divergences par rapport à Facebook, à l'égard duquel il se présente comme une véritable alternative. Sa figure de fondation ainsi que l'argumentaire entourant son projet tranchent avec ceux des plateformes socionumériques commerciales. En ce qui concerne le design d'interaction, nous avons vu que diaspora* présente matière à amélioration eu égard aux éléments de design de captation de l'attention. Toutefois, la plateforme n'est pas conçue pour centraliser les activités sociales des individus et n'aspire pas à devenir un tiers-lieu. En termes de stratégies rhétoriques liées à la protection de la vie privée, diaspora* met en œuvre de nombreuses pratiques éthiques qui renforcent son *ethos* en tant que média alternatif. Il en va de même au sujet des modèles interactifs persuasifs fallacieux : diaspora* en est pratiquement exempt. Tous ces éléments se retrouvent en partie dans le type de rhétorique procédurale mobilisé.

3. Remarques conclusives sur la rhétorique procédurale et ses liens avec la rhétorique processuelle

En guise d'ouverture conclusive, je souhaiterais me pencher sur la façon dont la rhétorique procédurale et la rhétorique processuelle – que je vais aborder sous peu – mettent au jour certaines limites relatives à la dimension alternative d'un média socionumérique libre comme diaspora*. Cette dernière partie se déplace vers un niveau sociopolitique de la rhétorique.

Rappelons en premier lieu que la rhétorique procédurale repose sur des stratégies plus ou moins éthiques (plus ou moins manipulatrices) et mobilise un certain nombre de modèles interactifs persuasifs ou tropes interactifs. Elle vise à persuader ou à convaincre les utilisateurs et les utilisatrices d'agir de telle ou telle façon au sein de l'interface. Plus généralement, la rhétorique procédurale favorise l'adoption d'habitudes et de réflexes comportementaux (physiques et psychiques).

Il me semble l'on peut rapprocher les effets de la rhétorique procédurale à l'une des remarques formulées par Lev Manovich dans son ouvrage phare, *Le langage des nouveaux médias*. Pour ce dernier, les interfaces informatiques obéissent à une certaine logique, celle d'extérioriser ou de transposer les processus cognitifs dans du code. L'auteur donne l'exemple de l'hypertexte : l'hypertexte consisterait en une extériorisation de la capacité associative mentale des individus. Plus largement cependant, Manovich voit cette logique d'extériorisation comme une « exigence de standardisation de la société de masse moderne⁷² » :

Les sujets doivent être standardisés et les moyens de parvenir à cette fin doivent l'être également. D'où l'objectivation des processus intérieurs, intimes, et leur assimilation à des formes visuelles extérieures facilement manipulables [...] Ce qui était auparavant un processus mental et un état propre à un seul individu devient alors partie intégrante de la sphère publique. Des représentations et des processus intérieurs et imperceptibles furent sortis de la tête des individus et posés à l'extérieur [...] Les médias informatiques interactifs s'adaptent parfaitement à cette tendance à l'extériorisation et à l'objectivation des opérations mentales⁷³.

Pour le dire autrement, le code est fondamentalement idéologique et le web, disons hégémonique, participe à standardiser nos comportements et à limiter nos répertoires d'action. Sur un plan politique, cet aspect soulève des enjeux immenses.

Explorant le concept de rhétorique procédurale, Maude Bonenfant et ses collègues introduisent une nuance : ils proposent de différencier la rhétorique procédurale de ce

qu'ils appellent la rhétorique processuelle⁷⁴. Si la première vise à persuader l'utilisateur ou l'utilisatrice de modifier son comportement par le biais d'une interaction humain-machine, la rhétorique processuelle agit, elle, à un niveau plus profond, à un niveau de conditionnement identitaire et axiologique. La dimension processuelle intervient sur l'identité et l'*ethos* de l'utilisateur ou de l'utilisatrice, sur ses valeurs et ses croyances. Pour reprendre le vocabulaire employé plus tôt dans cet article, par le biais de différents modèles interactifs persuasifs, l'individu intériorise un certain argumentaire, une certaine idéologie, et ce parfois « jusqu'à l'adhésion totale où il n'est plus possible de distinguer ce qui relève de l'individu de ce qui provient d'une autorité extérieure⁷⁵ ». Pour les auteur·e·s, si la rhétorique procédurale est du côté du « faire faire », la rhétorique processuelle est du côté du « faire être ». Autrement dit, au fil du temps, la rhétorique procédurale se double d'une rhétorique processuelle : d'une mise en contact récurrente avec des procédures interactives humain-machine naissent des modes de subjectivation particuliers.

Dans le contexte d'une économie des données personnelles et de la vie privée⁷⁶, la rhétorique processuelle d'un média socionumérique commercial vise à faire adhérer les usagers au discours selon lequel il est impossible aujourd'hui de se passer des plateformes numériques pour entretenir des relations sociales. Ces médias socionumériques véhiculent l'idée qu'ils tiennent une dimension essentiellement sociale, selon une définition assez spécifique du « social »⁷⁷, et qu'ils sont centraux à la vie sociale. On observe ainsi une forme d'industrialisation ou de marchandisation des médiations sociales. Par leurs stratégies rhétoriques processuelles, les plateformes commerciales persuadent les usagers qu'il est *valable* d'utiliser les médias socionumériques, même si certains d'entre eux savent pertinemment et logiquement qu'ils sont surveillés en continu (qu'il s'agisse de surveillance d'État, de surveillance capitaliste, de surveillance interpersonnelle) et que l'on marchandise leurs données personnelles et leur vie privée. Comme on le sait, certaines personnes déclarent naïvement qu'elles « n'ont absolument rien à cacher ». La rhétorique processuelle agit sur les dimensions morale et éthique du sujet qu'elle persuade et transforme. En un sens, elle rend aussi complices les individus d'un système de valeur fondé sur la transparence et la quantification des relations sociales et des actions posées dans un contexte capitaliste. La rhétorique processuelle constitue dans ce contexte un outil au service de l'idéologie capitaliste.

En ce qui concerne les médias socionumériques libres, l'idéologie à la base de leur création est certes très différente de celles des médias commerciaux, mais les notions de rhétorique procédurale et de rhétorique processuelle pointent une limite du caractère véritablement « alternatif » des médias socionumériques libres. Avant d'aller plus loin, il me faut encore préciser la signification du terme « alternatif ». En regard de l'utilisation que j'en ai fait depuis le début de ce texte, le terme connote l'idée d'éthique (*versus* les

méthodes fallacieuses et manipulateurs), mais il se rapporte également aux contre-discours opposés aux discours hégémoniques ainsi qu'à un positionnement politique et idéologique contre la privatisation, la centralisation et la commercialisation des données des usagers et des usagères. Les médias socionumériques alternatifs libristes valorisent généralement une culture numérique que l'on peut qualifier de plus émancipatoire (liberté, décentralisation, contrôle de l'utilisateur, communs numériques, non marchandisation des données, antipublicité). Cette idéologie se retrouve dans une part du code source de diaspora*. Comme l'écrit Mathieu Gauthier-Pilote que j'ai interrogé pour cette étude :

Le caractère non commercial de diaspora* et de Mastodon change déjà *beaucoup* de choses au niveau de l'expérience d'utilisation du simple fait qu'il n'y a pas de place – au niveau de la conception et du design – pour un intermédiaire qui a intérêt à opérer toutes les petites manipulations psychologiques [...] Il y a plus : l'étude du code source de diaspora* et de Mastodon *prouvera* sans l'ombre d'un doute que l'ordre des activités dans le flux d'activités d'une personne n'est pas manipulé, que le système de notification n'est pas abusif et donne le plein contrôle à l'utilisateur ou à l'utilisatrice, qu'il n'y a pas de pistage, de profilage ou d'analyse comportementale qui s'effectue au cœur de la machine, etc.⁷⁸

Le média diaspora* s'oppose à la revente de données personnelles, à la marchandisation de la vie privée, à la publicité et les contenus publiés ne sont pas prévus pour inciter les usagers et les usagères à la consommation⁷⁹.

Pour revenir à l'idée de limite, diaspora* n'en reste pas moins un média socionumérique qui vient avec des effets reliés à ce type de média dans un contexte numérique plus vaste. Mathieu Gauthier-Pilote le conçoit d'ailleurs très bien :

[...] toutes les applications de réseautage social, quelles qu'elles soient, partagent des caractéristiques communes et donc aussi des défauts communs. Elles s'insèrent notamment toutes dans la « culture de l'écran », qui a pris une ampleur ahurissante avec l'avènement de la mobilité. L'usage des appareils portables, tactiles et branchés à Internet a débouché, me semble-t-il, sur une nouvelle médiation de l'instant présent, qui agit notamment comme une distraction permanente dans nos vies. Il est clair que diaspora* et Mastodon ne font rien pour nous sortir de cette situation, hormis promettre un environnement où la publicité n'interfère pas d'une manière aussi envahissante dans toutes nos communications. C'est déjà beaucoup je trouve⁸⁰!

La dimension alternative de diaspora* par rapport à Facebook n'est évidemment pas à remettre en question. On peut néanmoins se demander si les rhétoriques procédurales et processuelles des médias socionumériques libres sont assez radicales et assez différenciées pour persuader les utilisateurs et les utilisatrices d'adopter des modes d'être et d'agir qui ne se conforment plus seulement à ceux que les plateformes hégémoniques cherchent activement à promouvoir. Autrement dit, les plateformes alternatives ne sont-

elles que des répliques non commerciales des grands modèles hégémoniques? Comment, par le biais de stratégies sémiotiques et rhétoriques radicalement distinctes, proposer d'autres normes en vue de « dé-standardiser » nos comportements? Comment, par ce biais, développer, exploiter et maintenir des modes de subjectivation émancipatoires? En somme, comment un média socionumérique peut-il lutter contre sa forme même et les effets pernicioeux de celle-ci⁸¹? Pour reprendre une idée chère aux situationnistes, n'est-il pas difficile de faire une critique de l'aliénation avec des « moyens aliénés »? Réfléchir à la façon dont les médias socionumériques libres pourraient aller plus loin idéologiquement, par le biais de rhétoriques procédurales et processuelles encore plus assumées ou critiques, me semble absolument nécessaire aujourd'hui. Cela constitue une nécessité, car ce n'est que de cette manière que les médias numériques libres peuvent réaliser pleinement le rôle éthique et politique qu'ils se sont eux-mêmes donné pour combattre l'informatique privatrice et le capitalisme numérique. Bien sûr, il ne s'agit pas de faire peser sur ces outils, et sur les communautés de développement informatique qui les portent, l'entière responsabilité de déjouer les travers du capitalisme numérique. Les médias numériques alternatifs proposent déjà de nombreuses solutions bien plus éthiques que leurs pendants propriétaires ou privés, et l'on peut déjà se réjouir de leur simple existence. Mais pour constituer un réel contrepoids à la marchandisation des médiations sociales, les médias socionumériques alternatifs ont besoin de développer des designs d'interaction radicalement distincts de ceux mis en place par les plateformes hégémoniques.

Bibliographie

- ACQUISTI, Alexandro *et al.*, « Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online », *ACM Computing Surveys*, vol. 50, no 3, 2017. DOI : 10.1145/3054926.
- BEAUDIN-LECOURS, Andrée, Isabelle DELISLE, Marie-Josée DESROCHERS, Guy GERMAIN, Patrick GIROUX, Laurence LACHAPPELLE-BÉGIN, Claude MARTEL & Jean-Luc TRUSSART, *Guide de l'utilisation pédagogique des médias sociaux*, Cégep de Lévis-Lauzon, Collège de Rosemont, Collège de Bois-de-Boulogne, Université du Québec à Chicoutimi, Performa, Collège de Lanaudière à L'Assomption, 2012.
- BOGOST, Ian, « The Rhetoric of Video Games », dans K. Salen (dir.), *The Ecology of Games: Connecting Youth, Games, and Learning*, Cambridge, The MIT Press, 2008.
- , *Persuasive Game: The Expressive Power of Videogames*, Cambridge, The MIT Press, 2010.
- BOMBARDI-BRAGG, Madeline Renee, *Exploring App Users' Experiences With App Notifications*, Fort Collins, Colorado State University, 2017.
- BONENFANT, Maude & Dominique ARSENAULT, « Dire, faire et être par les jeux vidéo », *Implications philosophiques*, 2016. En ligne <<http://www.implications-philosophiques.org/ethique-et-politique/ethique/dire-faire-etre-ethique-performativite-et-rhetoriques-procedurale-et-processuelle-dans-les-jeux-video/>>
- BONENFANT, Maude & Thibault PHILIPPETTE, « Rhétorique de l'engagement ludique dans des dispositifs de ludification », *Sciences du jeu*, no 10, 2018. En ligne <<http://journals.openedition.org/sdj/1422>>.
- BONHOMME, Marc, « De l'argumentativité des figures de rhétorique », *Argumentation & analyse du discours*, no 2, 2009. En ligne <<https://journals.openedition.org/aad/495>>.
- , *Pragmatique des figures du discours*, Paris, Champion, 2014.
- BÖSCH, Christoph, Benjamin ERB, Frank KARGL, Henning KOPP & Stefan PFATTHEICHER, « Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns », *Proceedings on Privacy Enhancing Technologies*, no 4, 2016, p. 237-254.
- BOYLE, Casey, James J. BROWN Jr. & Steph CERASO, « The Digital: Rhetoric Behind and Beyond the Screen », *Rhetoric Society Quarterly*, vol. 48, no 3, 2018, p. 251-259.
- BRETON, Philippe, *L'argumentation dans la communication*, Paris, La Découverte, 2006.
- BRIGNULL, Harry, « Confirmshaming ». En ligne : <<https://www.darkpatterns.org/types-of-dark-pattern/confirmshaming>> (consulté le 1^{er} juillet 2019).

- , *Dark Patterns*. En ligne <<http://darkpatterns.org/>> (consulté le 1^{er} juillet 2019).
- CANNIZZARO, Sara & Myrdene ANDERSON, « Culture as Habit, Habit as Culture: Instinct, Habituescence, Addiction », dans D. E. West & M. Anderson (dir.), *Consensus on Peirce's Concept of Habit. Before and Beyond Consciousness*, Suisse, Springer, vol. 31, 2016, p. 315-339.
- ECO, Umberto, *La structure absente*, trad. de l'italien par U. Esposito-Torrigiani, Paris, Mercure de France, 1972.
- EEMEREN, Frans H. & Rob GROOTENDORST, *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*, Cambridge, Cambridge University Press, 2003.
- EYAL, Nir & Ryan HOOVER, *Hooked: comment créer un produit ou un service qui ancre des habitudes*, trad. de l'anglais (États-Unis) par P.-M. Deschamps, Paris, Eyrolles, 2018.
- GEORGES, Fanny, « Représentation de soi et identité numérique. Une approche sémiotique et quantitative de l'emprise culturelle du web 2.0 », *Réseaux*, no 154, 2009, p. 165-193.
- GROUPE μ , *Traité du signe visuel*, Paris, Seuil, 1992.
- HILL, Matthew, *Smartphones: The Dark Side*, reportage de la BBC, Londres, 30 min., 2018.
- HOEPMAN, Jaap-Henk, *Privacy Design Strategies (The Little Blue Book)*, s. l., s. e., 2019. En ligne : <<https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf>>.
- HULIN, Thibaud, « Brève histoire de l'identité visuelle du Libre », dans C. Paloque-Berges & C. Masutti (dir.), *Histoire et cultures du libre. Des logiciels partagés aux licences échangées*, Framabook, 2013, p. 227-252.
- JAPPY, Tony, *Introduction to Peircean Visual Semiotics*, Londres, Bloomsbury, coll. « Bloomsbury Advances in Semiotics », 2013.
- JEANNE-PERRIER, Valérie, « Des discours et des signes : pour une évaluation critique de Twitter et de Facebook », *Communication & langages*, vol. 1, no 183, 2015, p. 111-131.
- KASTRENAKES, Jacob, « Facebook is thinking about hiding like counts, too », *The Verge*, 2019. En ligne : <<https://www.theverge.com/2019/9/2/20844574/facebook-hidden-like-count-test-under-consideration-leak>>.
- MAÍZ-ARÉVALO, Carmen, « Getting liked », dans C. R. Hoffmann & W. Bublitz (dir.), *Pragmatics of Social Media*, Berlin/Boston, De Gruyter, 2017, p. 575-606.
- MANOVICH, *Le langage des nouveaux médias*, trad. de l'anglais par R. Crevier, Paris, Les Presses du réel, 2010.
- MEHTA, Ravi & Rui ZHU, « Blue or Red? Exploring the Effect of Color on Cognitive Task Performances », *Science, New Series*, vol. 323, no 5918, 2009, p. 1226-1229.

- NITOT, Tristan & Nina CERCY, *Numérique : reprendre le contrôle*, Lyon, Framasoft, coll. « Framabook », 2016.
- PALOOQUE-BERGES, Camille & Christophe MASUTTI (dir.), *Histoires et cultures du Libre. Des logiciels partagés aux licences échangées*, Framabook, 2013.
- PAQUIN, Louis-Claude, *Comprendre les médias interactifs*, Montréal, Isabelle Quentin éditeur, 2006.
- PEIRCE, Charles S., *Écrits sur le signe*, trad. de l'anglais par Gérard Deledalle, Paris, Seuil, 1977.
- PERELMAN, Chaïm, *L'empire rhétorique. Rhétorique et argumentation*, Paris, Vrin, 1997.
- PLANTIN, Christian, « L'argument du paralogisme », *Hermès*, vol. 1, no 15, 1995, p. 245-262.
- , *Dictionnaire de l'argumentation*, Lyon, ENS Éditions, 2016.
- REBOUL, Olivier, *Introduction à la rhétorique. Théorie et pratique*, Paris, Presses Universitaires de France, 2013.
- ROCHELANDET, Fabrice, *Économie des données personnelles et de la vie privée*, Paris, La Découverte, 2010.
- ROUVROY, Antoinette & Thomas BERNIS, « Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation? », *Réseaux*, vol. 1, no 177, 2013, p. 168-173.
- SAOUTER, Catherine, *Le langage visuel*, Montréal, XYZ, 1998.
- STOYCHEFF, Elizabeth, « Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring », *Journalism & Mass Communication Quarterly*, vol. 93, n° 2, 2016, p. 1-16.
- TURKLE, Sherry, *Seuls ensemble. De plus en plus de technologies, de moins en moins de relations humaines*, trad. de l'anglais par C. Richard, Paris, L'Échappée, 2015.
- WANSINK, Brian, James E. PAINTER & Jill NORTH, « Bottomless Bowls: Why Visual Cues of Portion Size May Influence Intake », *Obesity Research*, vol. 13, no 1, 2005, p. 93-100.

Notes

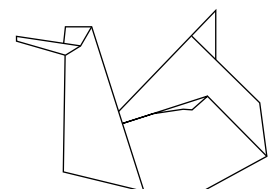
- 1 A. BEAUDIN-LECOURS *et al.*, *Guide de l'utilisation pédagogique des médias sociaux*, Cégep de Lévis-Lauzon, Collège de Rosemont, Collège de Bois-de-Boulogne, Université du Québec à Chicoutimi, Performa, Collège de Lanaudière à L'Assomption, 2012, p. 8.
- 2 Je renvoie par exemple à la carte des médias siconumériques créée par Brian Solis et JESS3, *Conversation Prism*. En ligne : <<https://conversationprism.com/>>.
- 3 La base de données Scopus fait état de plus de 3 400 articles ayant dans le titre le mot « Facebook » (2006-2019) uniquement dans les domaines des sciences sociales et des lettres. Ce résultat ne prend pas en compte les articles pour lesquels le mot Facebook n'apparaît que dans le résumé, ni les autres domaines disciplinaires répertoriés par Scopus.
- 4 Par idéologie libriste, je fais référence au mouvement de Richard Stallman et de ses héritières et héritiers, et non à l'idéologie libérale de l'*open software*. Voir à ce sujet C. PALOQUE-BERGES & C. MASUTTI (dir.), *Histoires et cultures du Libre. Des logiciels partagés aux licences échangées*, Framabook, 2013.
- 5 Un logiciel est libre s'il répond aux quatre libertés énoncées par la Free Software Fondation dès les années 1980. Cf. <<https://www.gnu.org/philosophy/free-sw.fr.html>>.
- 6 Voir par exemple T. NITOT & N. CERCY, *Numérique : reprendre le contrôle*, Lyon, Framasoft, coll. « Framabook », 2016.
- 7 F. ROCHELANDET, *Économie des données personnelles et de la vie privée*, Paris, La Découverte, 2010.
- 8 T. NITOT & N. CERCY, « Proposer des alternatives crédibles », *loc. cit.*, p. 72-73.
- 9 Cf. Mobilizon : <<https://joinmobilizon.org/>>. La campagne de sociofinancement a démarré en mai 2019 et le lancement en version bêta a eu lieu à l'automne 2019.
- 10 L.-C. PAQUIN, *Comprendre les médias interactifs*, Montréal, Isabelle Quentin éditeur, 2006.
- 11 Afin d'avoir une meilleure compréhension des enjeux techniques et idéologiques liés à diaspora*, j'ai également interrogé Mathieu Gauthier-Pilote, président de l'organisme sans but lucratif québécois FACIL (« pour l'appropriation collective de l'informatique libre »). Je le remercie sincèrement pour nos échanges.
- 12 Notamment T. JAPPY, *Introduction to Peircean Visual Semiotics*, Londres, Bloomsbury, coll. « Bloomsbury Advances in Semiotics », 2013 ; C. SAOUTER, *Le langage visuel*, Montréal, XYZ, 1998 ; GROUPE μ , *Traité du signe visuel*, Paris, Seuil, 1992 ; P. BRETON, *L'argumentation dans la communication*, Paris, La Découverte, 2006 ; C. PLANTIN, *Dictionnaire de l'argumentation*, Lyon, ENS Éditions, 2016 ; L.-C. PAQUIN, *Comprendre les médias interactifs*, *op. cit.* ; I. BOGOST, « The Rhetoric of Video Games », dans K. Salen (dir.), *The Ecology of Games: Connecting Youth, Games, and Learning*, Cambridge, The MIT Press, 2008, p. 117-140 ; I. BOGOST, *Persuasive Game: The Expressive Power of Videogames*, Cambridge, The MIT Press, 2010.
- 13 C. BOYLE *et al.*, « The Digital: Rhetoric Behind and Beyond the Screen », *Rhetoric Society Quarterly*, vol. 48, no 3, 2018, p. 255. Je traduis : « It is becoming commonplace that computation is a rhetorical medium and that the infrastructures underlying digital mediation more broadly are rhetorical. »
- 14 I. BOGOST, « The Rhetoric of Video Games », *loc. cit.* ; I. BOGOST, *Persuasive Game: The Expressive Power of Videogames*, *op. cit.*
- 15 Pour ne citer que ceux-là, voir notamment : M. BONHOMME, *Pragmatique des figures du discours*, Paris, Champion, 2014 ; Ph. BRETON, *L'argumentation dans la communication*, *op. cit.* ; C. PERELMAN, *L'empire rhétorique. Rhétorique et argumentation*, Paris, Vrin, 1997 ; O. REBOUL, *Introduction à la rhétorique. Théorie et pratique*, Paris, Presses Universitaires de France, 2013.

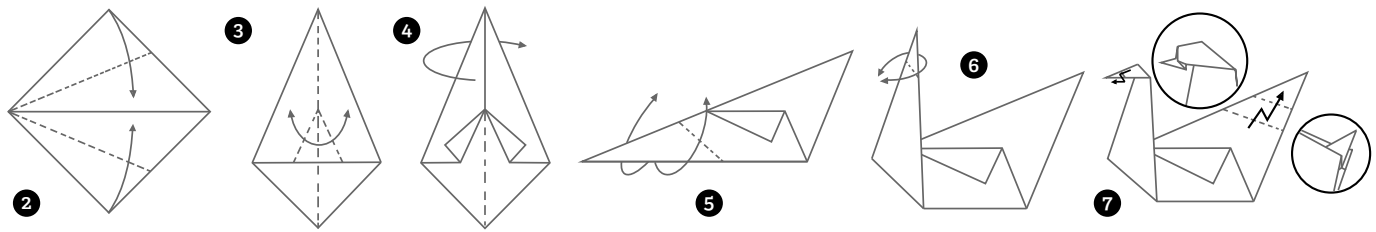
- 16 V. JEANNE-PERRIER, « Des discours et des signes : pour une évaluation critique de Twitter et de Facebook », *Communication & langages*, vol. 1, no 183, 2015, p. 111-131.
- 17 *Ibid.*, p. 114.
- 18 *Ibid.*, p. 121.
- 19 Cf. « diaspora* celebrates one year as a community project », *Diaspora Foundation*, 27 août 2013. En ligne : <<https://blog.diasporafoundation.org/1-diaspora-celebrates-one-year-as-a-community-project>>.
- 20 En septembre 2019, il en existait trois au Canada (tous situés à Montréal). Cf. *Poduptime*. En ligne : <<https://podupti.me/>>.
- 21 Cf. <<https://framasphe.org/help/pods>>.
- 22 « FACIL, pour l'appropriation collective de l'informatique libre (FACIL) est un organisme sans but lucratif dont le mandat est de promouvoir l'adoption, l'usage et la démocratisation de l'informatique libre. » Cf. « Présentation », *FACIL*. En ligne : <<https://facil.qc.ca/presentation>>.
- 23 Cf. <<https://parlote.facil.services/>>.
- 24 Cf. <<https://framasphe.org/>>.
- 25 Cf. <<https://diasporafoundation.org/>>.
- 26 Cf. <<https://wiki.diasporafoundation.org/Branding>>.
- 27 V. JEANNE-PERRIER, « Des discours et des signes : pour une évaluation critique de Twitter et de Facebook », *loc. cit.*, p. 114.
- 28 U. ECO, *La structure absente*, trad. de l'italien par U. Esposito-Torrigiani, Paris, Mercure de France, 1972, p. 241.
- 29 Rappelons-nous que diaspora* est surnommé le « tueur de Facebook », discours assumé par les concepteurs eux-mêmes.
- 30 Je n'opère pas directement dans cet article une comparaison entre la nomenclature verbale de Facebook (amis, mur, profil, partage, flux, etc.) et celle de diaspora*. Ce sujet pourrait faire l'objet d'un autre article.
- 31 Cf. à ce propos T. HULIN, « Brève histoire de l'identité visuelle du Libre », dans C. Paloque-Berges & C. Masutti (dir.), *Histoire et cultures du libre. Des logiciels partagés aux licences échangées*, Framabook, 2013, p. 227-252.
- 32 I. BOGOST, *Persuasive Game: The Expressive Power of Videogames*, *op. cit.*, p. 3. Je traduis : « [...] procedural rhetoric is the practice of persuading through processes in general and computational processes in particular. Just a verbal rhetoric is useful for both the orator and the audience, and just as written rhetoric is useful for both the writer and the reader, so procedural rhetoric is useful for both the programmer and the user, the game designer and the player. »
- 33 M. BONENFANT & T. PHILIPPETTE, « Rhétorique de l'engagement ludique dans des dispositifs de ludification », *loc. cit.*, § 27.
- 34 A. ACQUISTI *et al.*, « Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online », *ACM Computing Surveys*, vol. 50, no 3, 2017, p. 3. Je traduis : « [...] every design decision potentially nudges users in one direction or another. »
- 35 Cf. F. H. EEMEREN & R. GROOTENDORST, *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*, Cambridge, Cambridge University Press, 2003. Il existe un certain nombre de règles de conduite éthique en argumentation : notamment la règle de de bonne foi (ne pas mentir, éviter de cacher des informations et de les déformer) et la règle de responsabilité (avoir conscience des mécanismes du langage et de ses effets potentiels).
- 36 Traduction libre.

- 37 Certains, à l'exemple de Nir Eyal et Ryan Hoover, l'ont très bien compris : la formation d'habitudes est une stratégie économique. Cf. N. EYAL & R. HOOVER, *Hooked : comment créer un produit ou un service qui ancre des habitudes*, trad. de l'anglais (États-Unis) par P.-M. Deschamps, Paris, Eyrolles, 2018.
- 38 M. R. BOMBARDI-BRAGG, *Exploring App Users' Experiences With App Notifications*, Fort Collins, Colorado State University, 2017.
- 39 Nir Eyal et Ryan Hoover discernent par exemple trois types de récompenses : un type de récompense flatte l'égo personnel, un autre type correspond à la reconnaissance d'être bien ancré dans la communauté sociale, une dernière satisfait une quête de ressource poursuivie avec obstination. Cf. N. EYAL & R. HOOVER, *Hooked*, op. cit., p. 95 et sq. Dans le domaine psychologique, certain-e-s chercheur-e-s se sont intéressé-e-s aux récompenses physiologiques en lien avec les notifications numériques non planifiées. Par ex. la chercheuse Madeline Renee Bombardi-Bragg écrit que : « [...] la récompense, biochimiquement régulée par la dopamine, reçue après le stimulus d'une notification a la capacité d'induire une habitude à long terme, c'est-à-dire de créer une tendance comportementale à vérifier les notifications lorsqu'elles arrivent – tendance renforcée lorsque les signaux de notification sont reçus selon un horaire imprévisible [non planifié]. » « [...] the biochemically determinate, dopamine-regulated reward received after a notification stimulus has the ability to induce long-term habits, that is, to create a behavioral tendency to check notifications when they are delivered – a tendency that is reinforced when the notification cues are received on an unpredictable schedule. » (trad. libre) M. R. BOMBARDI-BRAGG, *Exploring App Users' Experiences With App Notifications*, op. cit., p. 30.
- 40 Bien que cela ne soit pas le sujet de cet article, certaines habitudes numériques peuvent dans certains cas s'associer à des comportements extrêmes, par ex. à des comportements de dépendance. Ces habitudes font notamment l'objet de recherches en addictologie (l'addiction étant une forme d'habitude extrême, cf. S. CANNIZZARO & M. ANDERSON, « Culture as Habit, Habit as Culture: Instinct, Habituescence, Addiction », dans D. E. West & M. Anderson (dir.), *Consensus on Peirce's Concept of Habit. Before and Beyond Consciousness*, Suisse, Springer, vol. 31, 2016, p. 315-339). Je renvoie également à l'entrevue de Aza Raskin dans Matthew HILL, *Smartphones: The Dark Side*, reportage de la BBC, Londres, 30 min., 2018. Raskin compare l'expérience du défilement infini à l'expérience du bol de soupe sans fond (voir à ce sujet B. WANSINK et al., « Bottomless Bowls: Why Visual Cues of Portion Size May Influence Intake », *Obesity Research*, vol. 13, no 1, 2005, p. 93-100). Outre cela, les médias socionumériques et les *likes* favoriseraient également la dépression ; voir à ce titre C. MAÍZ-ARÉVALO, « Getting liked », dans C. R. Hoffmann & W. Bublitz (dir.), *Pragmatics of Social Media*, Berlin/Boston, De Gruyter, 2017, p. 575-606.
- 41 Voir par exemple R. MEHTA & R. ZHU, « Blue or Red? Exploring the Effect of Color on Cognitive Task Performances », *Science*, New Series, vol. 323, no 5918, 2009, p. 1226-1229.
- 42 En septembre 2019, Facebook envisageait par exemple de cacher le nombre total de « J'aime » en raison des conséquences néfastes de la quantification. Cf. J. KASTRENAKES, « Facebook is thinking about hiding like counts, too », *The Verge*, 2019. En ligne : <<https://www.theverge.com/2019/9/2/20844574/facebook-hidden-like-count-test-under-consideration-leak>>.
- 43 Cf. <<https://framaspHERE.org/help/miscellaneous>>.
- 44 Sous couvert de gratuité, sans expliciter le modèle d'affaires, les plateformes socionumériques transforment les usagers en « produits » (ou du moins les usagers travaillent gratuitement pour les plateformes).
- 45 C. BÖSCH et al., « Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns », *Proceedings on Privacy Enhancing Technologies*, no 4, 2016, p. 237-254.
- 46 J.-H. HOEPMAN, *Privacy Design Strategies (The Little Blue Book)*, 2019. En ligne : <<https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf>> ; C. BÖSCH et al., « Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns », loc. cit.

- 47 C. BÖSCH *et al.*, « Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns », *loc. cit.*, p. 242.
- 48 J.-H. HOEPMAN, *Privacy Design Strategies (The Little Blue Book)*, *op. cit.*
- 49 Entretien par courriel avec l'auteur, juillet 2019.
- 50 C. PLANTIN, « L'argument du parallogisme », *Hermès*, vol. 1, no 15, 1995, p. 250.
- 51 Cf. <<https://www.communecter.org/>>.
- 52 J.-H. HOEPMAN, *Privacy Design Strategies (The Little Blue Book)*, *op. cit.*, p. 12.
- 53 Pour aller plus loin sur le thème de la représentation de l'identité numérique, dont je ne parlerai pas davantage ici, voir par exemple F. GEORGES, « Représentation de soi et identité numérique. Une approche sémiotique et quantitative de l'emprise culturelle du web 2.0 », *Réseaux*, no 154, 2009, p. 165-193.
- 54 Selon Philippe Breton, le cadrage en argumentation consiste à présenter la réalité d'un certain point de vue (certains aspects sont mis de l'avant tandis que d'autres sont minimisés ou cachés). Cf. P. BRETON, *L'argumentation dans la communication*, *op. cit.*
- 55 C. BÖSCH *et al.*, « Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns », *loc. cit.*, p. 248. Je traduis : « While the service provider will claim that users have full control over their privacy settings, the presentation, terminology and user experience will highly discourage users from making changes. »
- 56 Comme l'écoblanchiment (*greenwashing*), le « cryptowashing » fait référence à une entreprise qui développe une stratégie marketing destinée à faire croire qu'elle est responsable en matière de sécurité et de cryptage, alors qu'elle ne l'est pas réellement. Le néologisme est par exemple utilisé dans le milieu libriste pour dénoncer les pratiques trompeuses.
- 57 C. PLANTIN, « Dilemme », *Dictionnaire de l'argumentation*, Lyon, ENS Éditions, 2016, p. 168.
- 58 Cf. <https://framaspHERE.org/help/account_and_data_management>.
- 59 Voir par ex. : <<http://www.legaltechdesign.com/communication-design/privicons-project/>> ; J.-H. HOEPMAN, « Using icons to summarise privacy policies: an analysis and a proposal », blogue de l'auteur. En ligne : <<https://blog.xot.nl/2016/09/21/using-icons-to-summarise-privacy-policies-an-analysis-and-a-proposal/>>.
- 60 A. ACQUISTI *et al.*, « Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online », *loc. cit.*
- 61 H. BRIGNULL, *Dark Patterns*. En ligne <<http://darkpatterns.org/>> (consulté le 1^{er} juillet 2019).
- 62 C. BÖSCH *et al.*, « Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns », *loc. cit.*, p. 239.
- 63 Le « Roach motel » est un modèle qui rend très difficile la désinscription ou la suppression de compte. Le « Color trick » est un modèle qui exploite nos failles comportementales : dans le cours de l'énoncé interactif, il s'agit d'associer plusieurs fois de suite un bouton d'une couleur (vert par ex.) à la même fonction (souvent à une action jugée positive, par exemple « valider ») pour ensuite utiliser le même bouton coloré pour inciter l'utilisateur à réaliser un autre type d'action favorable à la plateforme (un achat par exemple). Quant au modèle du « Dirt and Dust », il est utilisé sur les interfaces tactiles. Il s'agit d'une sorte de trompe-l'œil interactif : une fausse poussière ou une fausse rayure est représentée sur l'écran pour inciter l'utilisateur à nettoyer son appareil. Ce faisant, l'utilisateur croyant enlever une poussière clique sur l'interface de sa tablette tactile.
- 64 Par exemple, les erreurs logiques, les stratégies fallacieuses fondées sur le cadrage ou la diversion, les stratégies fallacieuses fondées sur l'autorité, les stratégies fallacieuses fondées sur le style ou la forme, les stratégies fondées sur la mobilisation du pathos.

- 65 Louis-Claude Paquin a identifié différentes « figures » reliées à l'interactivité dans son ouvrage *Comprendre les médias interactifs*, op. cit. La liste qu'il a composée s'apparente plutôt à l'équivalent de figures d'ornementation en rhétorique.
- 66 M. BONHOMME, « De l'argumentativité des figures de rhétorique », *Argumentation & analyse du discours*, no 2, 2009, p. 4. En ligne <<https://journals.openedition.org/aad/495>>.
- 67 *Ibid.*, p. 5.
- 68 Il y a là une problématique à laquelle la plupart des médias socionumériques libres sont confrontés : laisser une pleine liberté aux usagers tout en essayant de convaincre ces derniers que les médias libres sont plus avantageux que leurs pendants commerciaux. À cela s'ajoute une autre problématique, celle d'éviter qu'un média libre ne devienne trop hégémonique (centralisateur). C'est par exemple pour cette raison que Framasoft a lancé en septembre 2019 une campagne de « déframatisation » du web : <<https://framablog.org/2019/09/24/deframasoftisons-internet/>>.
- 69 H. BRIGNULL, « Confirmshaming ». En ligne : <<https://www.darkpatterns.org/types-of-dark-pattern/confirmshaming>> (consulté le 1^{er} juillet 2019).
- 70 L'on trouvera de nombreux exemples ici : <<https://confirmshaming.tumblr.com>>.
- 71 On pourrait presque la comparer à une pétition de principe en rhétorique : la prémisse est considérée comme étant acceptée par l'interlocuteur alors que ce n'est pas le cas.
- 72 L. MANOVICH, *Le langage des nouveaux médias*, trad. de l'anglais par R. Crevier, Paris, Les Presses du réel, 2010, p. 147.
- 73 *Ibid.*, p. 147-148.
- 74 M. BONENFANT & D. ARSENAULT, « Dire, faire et être par les jeux vidéo », *Implications philosophiques*, 2016. En ligne : <<http://www.implications-philosophiques.org/ethique-et-politique/ethique/dire-faire-etre-ethique-performativite-et-rhetoriques-procedurale-et-processuelle-dans-les-jeux-video/>> ; M. BONENFANT & T. PHILIPPETTE, « Rhétorique de l'engagement ludique dans des dispositifs de ludification », loc. cit.
- 75 M. BONENFANT & T. PHILIPPETTE, « Rhétorique de l'engagement ludique dans des dispositifs de ludification », loc. cit., § 32.
- 76 F. ROCHELANDET, *Économie des données personnelles et de la vie privée*, op. cit.
- 77 Voir à ce sujet A. MONDOUX, « À propos du social dans les médias sociaux », *Terminal*, no 111, 2012, p. 69-79. En ligne : <<https://journals.openedition.org/terminal/977>>.
- 78 Entrevue par courriel avec l'auteur, juillet 2019.
- 79 diaspora* met aussi un frein à la gouvernementalité algorithmique décrite par Antoinette Rouvroy et Thomas Berns dont je ne parlerai pas plus ici. Voir A. ROUVROY & T. BERNS, « Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation? », *Réseaux*, vol. 1, no 177, 2013, p. 168-173.
- 80 Entrevue par courriel avec l'auteur, juillet 2019.
- 81 Par effets pernicioeux, on pense par exemple à la surveillance interpersonnelle, à l'autocensure ou à l'isolement social. Sur l'autocensure, voir l'étude d'Elizabeth Stoycheff à propos des effets de la surveillance sur l'autocensure et les opinions minoritaires : E. STOYCHEFF, « Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring », *Journalism & Mass Communication Quarterly*, vol. 93, no 2, 2016, p. 1-16. Sur l'isolement, voir S. TURKLE, *Seuls ensemble. De plus en plus de technologies, de moins en moins de relations humaines*, trad. de l'anglais par C. Richard, Paris, L'Échappée, 2015.





Du fétichisme des sémiomarchandises « parlantes » à l'asservissement machinique : le cas des assistants personnels intelligents

Fabien RICHERT

Cygne noir, no 7, 2019 : « Algorithmes »

Résumé

Cet article propose une analyse critique d'une version particulière de l'intelligence artificielle, à savoir les « assistants personnels intelligents ». L'argumentaire s'inscrit dans les travaux récents portant sur le sémiocapitalisme dont la perspective propose des clés de lecture, inspirées notamment par les écrits de Félix Guattari et de Gilles Deleuze, pour aborder les modes de subjectivation dans le capitalisme tardif dominé par les technologies de l'information et de la communication. Les assistants intelligents seront questionnés du point de vue du régime de vérité et des relations singulières qu'ils instaurent avec leurs propriétaires invités à participer avec enthousiasme à l'idéologie consumériste ambiante. Les assistants intelligents seront notamment thématiques comme des sémiomarchandises capables de suggérer des objets, services et expériences sur la base d'un historique « dividuel », du travail aliéné et invisibilisé d'une masse croissante d'info-travailleurs ou plus largement d'un asservissement machinique.

Pour citer cet article

RICHERT, Fabien, « Du fétichisme des sémiomarchandises "parlantes" à l'asservissement machinique : le cas des assistants personnels intelligents », *Cygne noir*, no 7, 2019. En ligne : <http://www.revuecygnoir.org/numero/article/richert-fetichisme-semiomarchandises> (consulté le xx/xx/xxxx).



Cet article de *Revue Cygne noir* est mis à disposition selon les termes de la licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Pas de Modification 2.5 Canada.

DU FÉTICHISME DES SÉMIOMARCHANDISES « PARLANTES » À L'ASSERVISSEMENT MACHINIQUE : LE CAS DES ASSISTANTS PERSONNELS INTELLIGENTS

Introduction

Dans un ouvrage récent, Luc Julia soutient la thèse provocatrice suivante : « [l']intelligence artificielle n'existe pas. Si nous devons garder cet acronyme, l'IA ne doit plus signifier "intelligence artificielle", mais "intelligence augmentée"¹. » Une telle formule péremptoire étonne, surtout lorsque l'on sait que Luc Julia est un ingénieur de renommée internationale ayant fait ses armes chez Apple et Samsung, co-inventeur en chef de l'assistant SIRI (Apple), pionnier de la reconnaissance vocale et promoteur du modèle de la Silicon Valley en France. Son ouvrage, qui brasse anecdotes biographiques et formules didactiques sur l'intelligence artificielle (IA), dresse un argumentaire qui affirme avec vigueur le constat suivant lequel les équipes d'ingénieurs en charge d'imaginer et de développer des systèmes « intelligents » gardent toujours et encore le contrôle sur le fonctionnement et la marge de manœuvre de ces systèmes. Même les techniques dites d'apprentissage profond² (*deep learning*), célébrées par les experts de Google, d'IBM, de Microsoft, d'Amazon ou encore d'Adobe pour la relative autonomie qu'elles octroieraient aux systèmes qualifiés d'« auto-apprenants », restent sous le contrôle exclusif des techniciens et des experts. Si l'on suit Luc Julia jusqu'au bout de sa logique, une machine n'exécute que les tâches pour lesquelles elle a été programmée. Bien plus, ces systèmes « n'inventent rien, ils ne font que suivre des règles, des exemples et des codes, en utilisant les données que nous choisissons pour ces systèmes³ ».

Ce discours se veut rassurant dans un contexte où les risques de l'IA sont souvent dénoncés. À titre d'exemple, une tribune de mise en garde contre des robots tueurs a été signée par des personnalités remarquées, comme Elon Musk, Stephen Hawkins ou encore Steve Wozniak, et déterminées à vouloir encadrer leur développement⁴. Si l'IA est présente dans l'imaginaire socioculturel et dans les programmes de recherche depuis de nombreuses années (dès les années 1950 sous l'influence notable du mathématicien Alan Turing), sa sophistication et son renouveau ont été notamment rendus possibles grâce aux données massives (*Big data*) et à l'amélioration des techniques d'indexation et de stockage des informations. Ce renouvellement ne va pas sans susciter des craintes, elles-mêmes nourries par les productions culturelles de fiction. Or, à la différence des fantasmes inspirés des imaginaires spectaculaires de science-fiction qui

dramatisent notre rapport avec l'intelligence artificielle, les arguments défendus par Luc Julia se présentent comme étant beaucoup moins alarmistes. En effet, à suivre Julia, l'IA doit simplement être considérée comme une sorte d'agent semi-autonome utilisé pour : « faciliter nos décisions, pas pour les prendre à notre place⁵. » Si bien que, dans ce type d'argumentaire, le terme *d'intelligence augmentée* apparaît comme étant plus conforme au mode opératoire des systèmes intelligents voués à aider « les êtres intelligents à avoir plus de capacités et à être meilleurs dans des domaines spécifiques⁶ ».

En donnant à l'intelligence artificielle l'apparence d'un artéfact censé accompagner nos activités et nos gestes quotidiens, Luc Julia assigne finalement à l'IA une fonction managériale d'optimisation et de gestion des ressources de l'individu. Selon cette rhétorique, à partir d'une collecte et d'une analyse toujours plus précise et détaillée des données personnelles, l'intelligence artificielle devrait permettre à l'individu de mieux s'orienter dans le monde social, de se hisser et de se maintenir à la hauteur des attentes d'un système qui valorise les normes de performance et qui thématise l'amélioration de la condition de l'humain à travers une « augmentation » de ses capacités physiques, cognitives ou encore affectives.

L'argumentaire de Julia constitue un argumentaire-type non alarmiste, qui reconduit l'idée que la technologie dépendrait des usages effectués par les individus, lesquels seraient en total contrôle : le problème serait donc moral (bons et mauvais usages). Or, s'arrêter à cette dimension ne permet pas de bien saisir les logiques techniques et leurs enjeux dans un contexte néolibéral. Dans cet article, je propose ainsi d'effectuer une critique de ce type de discours et d'aborder l'IA à partir d'une grille d'analyse marxienne (mode de subjectivation, aliénation, fragmentation, fétichisme, marchandise) qui permettra de mieux la contextualiser dans le capitalisme tardif. Dans le discours de Julia, l'intelligence artificielle est comprise comme un « assistant personnel » dédié plus particulièrement aux marchés de la consommation et des services. C'est cette version de l'IA qui constituera l'objet de mon analyse dans cet article. Cette précision est importante dans la mesure où l'intelligence artificielle est déjà largement utilisée pour ses capacités d'analyse et d'automatisation de certaines tâches dans des secteurs aussi éparés que la physique, la génomique, la santé, les assurances, les télécommunications, etc. Il sera question notamment d'aborder les assistants personnels dits intelligents en mettant en exergue le double renforcement suivant : (1) renforcement d'un rapport de plus en plus étroit et intime des individus avec les dispositifs technologiques ; (2) renforcement de notre rapport hyperpersonnalisé à un monde largement inféodé à des impératifs économiques. Pour aborder ce double renforcement, je procéderai en deux temps. Il s'agira d'abord de questionner le régime de vérité supposé des assistants personnels intelligents et de mieux saisir leurs effets sur les modes de subjectivation. Par la suite, je montrerai

que la puissance suggestive du régime de vérité de ces systèmes intègre une dimension éminemment performative. Je m'appuierai notamment sur l'analyse althussérienne de l'idéologie en insistant sur la notion d'interpellation qui donnera un éclairage singulier des relations que les individus entretiennent avec leurs assistants intelligents. Dans la section suivante, j'envisagerai ces relations dans une perspective plus précisément marxienne. Pour ce faire, j'emploierai le concept de fétichisme de la marchandise pour insister sur le paradoxe que présupposent les systèmes intelligents et leur utilisation : personnalisation de nos rapports avec des marchandises qui finissent par dominer les modalités mêmes de ces rapports. En me servant de certaines analyses de Jean Baudrillard portant sur la société de consommation, il s'agira de montrer que le fétichisme de la marchandise ne revient pas à adorer tel ou tel objet dédié à l'échange et à la consommation, mais à célébrer la logique du système tout entier, ses contradictions et ses tensions toujours déjà intériorisées à différents niveaux dans le monde social. Baudrillard nous invite à un travail réflexif nécessaire de *défétichisation* en ramenant notamment la marchandise à ses conditions réelles de production pour mieux limiter la puissance suggestive de son imaginaire factice. Raison pour laquelle la dernière partie de mon article proposera de revenir sur les nouvelles formes d'exploitation et d'asservissement (dit « machinique ») liées à la confection des assistants intelligents à l'ère de ce que Franco Berardi a proposé d'appeler le sémiocapitalisme.

Alètheia algorithmique

Interopérabilité entre les objets connectés, accompagnement des gestes et des activités, aide à la prise de décisions ou encore fluidification de nos déplacements et de nos usages, autant d'aspects que les applications d'IA sont censées recouvrir. En témoigne par exemple la vision symptomatique d'un futur, en 2040, que Luc Julia appelle de ses vœux :

[Le matin,] vous vous réveillez au moment où c'est le plus approprié pour vous, en fonction de vos cycles de sommeil. Les capteurs qui se trouvent dans votre lit sont en mesure de savoir à quel moment du cycle vous vous trouvez, et vont donc commencer à vous réveiller lorsque c'est le moment adéquat pour vous [...] Vous vous levez et vous allez dans la salle de bain. Votre miroir est plein de capteurs, il prend votre rythme cardiaque et mesure le taux d'hydratation de votre peau. Votre brosse à dent [sic] analyse votre salive, pour savoir quel est le niveau de votre salinité [...] Votre assistant vous suggère de faire une activité sportive, en fonction de votre journée de la veille, de ce que vous avez mangé, de la journée que vous allez passer, et vous propose des exercices en vous indiquant combien de répétitions effectuer [...] Vous partez au travail dans une voiture autonome qui ne vous appartient pas, mais qui passe vous chercher

à la seconde même où vous êtes prêt à partir. Elle s'est déjà adaptée à votre humeur, les couleurs de son habitat sont celles que vous aimez, et elle diffuse la musique que vous souhaitez écouter. [...] L'ensemble de vos objets connectés se calent sur votre emploi du temps. À la maison, pendant votre absence, votre aspirateur fait le ménage et d'autres robots nettoient les vitres. [Au travail, l]es réceptionnistes sont des robots qui vous reconnaissent en effectuant un scan de vos rétines ou de votre peau [etc.]⁷

Voici donc le fond de l'affaire : la dissémination des capteurs dans les objets connectés, dans les différentes pièces de la maison et plus largement dans le mobilier urbain est un prérequis pour permettre une cueillette tous azimuts d'une multitude de données que des systèmes analysent en temps réel. Des algorithmes peuvent être alors programmés pour récolter des informations sur des personnes, formuler des profils-types (citoyens, travailleurs, patients, consommateurs, clients⁸), suggérer des produits et des services adaptés à ces profils ou encore autoriser et interdire l'accès pour certaines demandes (immigration, emploi, prêt bancaire, prestation d'assurance, etc.). De tels usages semblent virtuellement infinis et peuvent *a priori* s'appliquer à n'importe quelles situations.

Le philosophe Éric Sadin a très bien montré quelle était la cible visée par l'IA par-delà toutes ces possibilités d'usages et d'applications : *expertiser le réel*⁹. Si Google, Facebook, Amazon, mais aussi et plus largement les grandes puissances mondiales (États-Unis, Russie, Canada, France, Iran, etc.) sont décidées à investir dans l'intelligence artificielle, c'est précisément en raison de sa puissance d'expertise qui limite le risque et l'incertitude tout en maximisant les gains et les opportunités.

Éric Sadin, dont les analyses ciblent notamment les discours apologétiques vouant un quasi-culte à l'intelligence artificielle, va encore plus loin et propose de voir dans ces nouveaux systèmes le franchissement d'un seuil problématique : non plus seulement le stockage et l'indexation des données massives, mais l'exposition automatique de l'état des choses grâce à l'érection d'« une puissance *alèthéique*, une instance vouée à exposer l'*alètheia*, la vérité¹⁰ ». En se nourrissant d'un corpus infini de données, les systèmes intelligents se voient chargés de révéler des faits autrement opaques à nos consciences. Les gestes, les comportements, les déplacements et les conversations laissent désormais des traces qui sont scrutées, analysées et enregistrées par des systèmes chargés d'établir et de formuler toute une série de diagnostics, de recommandations et de prescriptions.

Qu'il suffise ici d'évoquer le secteur de la santé, déjà partiellement évoqué dans l'avenir fantasmé par Luc Julia, où des capteurs sont chargés d'enregistrer et de mesurer les flux biologiques, d'évaluer la qualité du sommeil, les distances parcourues, de calculer les calories quotidiennes, etc. – autant d'applications qui, par ailleurs, renforcent et banalisent des logiques d'autosurveillance. De telles tendances s'observent également

dans le domaine du travail, mais pour des usages quelque peu différents puisque les systèmes intelligents ne veillent plus à la santé des individus, mais à leur productivité grâce à une taylorisation sans aucune mesure dans l'organisation des tâches (Amazon figure désormais comme un cas d'école : les employés sont munis de bracelets qui chronomètrent le temps d'emballage, la sortie des colis en rayon, etc.).

Dans tous les cas, les traces de l'activité des individus, ou de leurs états physiologiques, ont valeur de vérité puisqu'elles constitueraient une preuve irréfutable d'un état de choses à partir duquel les systèmes intelligents s'autorisent à établir une série de diagnostics. Raison pour laquelle Éric Sadin voit dans la puissance d'expertise déployée par les applications d'intelligence artificielle l'occasion d'instaurer un *régime de vérité* imparable et difficilement réfutable dans la mesure où la collecte des données, la sophistication des équations et des algorithmes impliquent la certitude que les résultats produits par la machine se conforment au mieux à l'état de ce qui est. Sadin associe à ce régime de vérité quatre axiomes cardinaux : « la collecte informationnelle, l'analyse en temps réel, la détection de corrélations significatives et l'interprétation automatisée des phénomènes¹¹. » Le régime de vérité dont parle Éric Sadin constitue finalement le rêve des promoteurs des données massives et de l'intelligence artificielle qui fantasment un réel qui parlerait par et pour lui-même grâce à une collecte toujours plus fine d'informations qui résistent toujours et encore à nos propres facultés de perception. Les coordonnées existentielles de chaque objet ou événement du monde social peuvent *a priori* être traduites en données, relativement aux capacités des outils de captation, de stockage et de traitement des données massives. La vérité repose alors sur la puissance affirmative des données qui refléteraient fidèlement un état de choses.

Évidemment, la confiance accordée à la mise en donnée du monde, aux chiffres et aux « data », s'inscrit dans une tradition plus lointaine que le juriste Alain Supiot fait remonter au moins jusqu'à l'époque où les échanges marchands avaient fini par se généraliser au courant du XIX^e siècle sur une partie de la surface du monde, et avec eux, la tendance à vouloir tout quantifier, comparer et mesurer¹². Quelque temps plus tard, les premières machines de calcul étaient alors vite devenues des outils techniques incontournables pour accompagner le mouvement de quantification du monde et l'instauration d'une administration hyper-rationalisée des choses – conformément aux principes directeurs de la cybernétique que l'on retrouve, sous un vocable quelque peu différent, dans les discours des promoteurs de l'IA (diminution de l'entropie, automatisation de l'organisation sociale, fétichisme de l'information).

L'intelligence artificielle trouve donc sa place dans cette longue trajectoire qui voit dans la quantification intégrale de la vie le meilleur moyen d'en contrôler tous les aspects. Antoinette Rouvroy et Thomas Berns ont proposé d'appeler « gouvernementalité

algorithmique »¹³ la tendance à vouloir modéliser une réalité sociale a-normative au plus proche des singularités individuelles grâce à des techniques avancées de forage de données (*data mining*). Or, Rouvroy et Berns ont bien montré que ce mode de gouvernementalité tend paradoxalement à reproduire, sous des aspects inédits, les normes les plus immanentes au monde social (à commencer par celles du marché qui semblent dorénavant naturaliser des modèles de performance à laquelle chacun devrait se conformer), ou, de manière bien plus problématique encore, à disqualifier par avance toutes tentatives de remises en cause de cette réalité.

La perfection présupposée de l'*alètheia* algorithmique et les difficultés de sa critique consistent en ceci qu'elle s'est construite sur la base d'une expulsion de la subjectivité du fait d'une automatisation intégrale des processus intrinsèques aux systèmes intelligents. Selon la vulgate ambiante, l'humain est essentiellement faillible, capable de prendre des décisions irrationnelles, de faire des erreurs et d'introduire des biais et des jugements là où les systèmes intelligents s'activent sans considération aucune pour la complexité et la singularité des affaires humaines. Cette vision pessimiste de l'humain, présentée de manière emphatique dans les discours de certains prophètes transhumanistes, était déjà observable chez les premiers cybernéticiens (Norbert Wiener en tête¹⁴) qui envisageaient la fragilité, la précarité, la finitude et plus généralement l'indétermination de l'existence comme une insulte devant le génie (technicien) dont l'humain était capable – vision articulée par ailleurs à ce que Nicolas Ledévédec a très justement proposé d'appeler « une anthropologie de la déficience »¹⁵.

Selon cette vision caricaturale et hautement problématique, il n'y a donc pas lieu de s'opposer aux décisions rendues par la machine, sinon c'est qu'elle a été mal programmée ou qu'elle sert des desseins contraires à l'éthique ambiante. L'éthique sert par ailleurs de catégorie « fourre-tout »¹⁶ (essentiellement rattachée au thème du respect de la vie privée) et de garantie quasi nulle d'une prise de responsabilité des ingénieurs qui imaginent des systèmes intelligents pour le compte de leur employeur ; des systèmes, donc, toujours déjà travaillés par des déterminations économiques et sociales.

À la puissance suggestive du régime de vérité des systèmes intelligents s'en ajoute une autre, tout aussi importante, et qui concerne leur dimension *performative*. La vérité, suivant les profondes analyses de Foucault¹⁷, résulte toujours d'un rapport de force, voire d'une forme de violence qui s'exerce sur des états de choses à travers le jeu d'une série de dispositifs de savoir normalisant qui déterminent en retour un cadre interprétatif et des modalités d'action. En d'autres termes, la vérité affecte l'existence des individus et leurs subjectivités puisqu'elle informe le sens des activités, des pratiques, des rituels et des obligations sociales chargés de produire des conduites et des comportements normalisés. Les régimes de vérité influencent les modes de subjectivation qui donnent à

l'individu des repères existentiels pour mieux se comprendre lui-même et déterminer sa place dans l'ordre du monde social. Les systèmes intelligents, en tant qu'ils constituent eux-mêmes un type particulier de dispositif normalisant, prescrivent, en fonction des situations et des contextes, un répertoire d'actions et des clés de lecture pour faciliter l'orientation des individus dans la complexité croissante de la modernité. Expertiser et prescrire : voici donc le double rôle assigné à ces systèmes que Sadin n'hésite pas à qualifier de « techno-idéologique »¹⁸ puisqu'ils impriment un mode de rationalité jusque dans les plis les plus intimes de l'existence.

Mais l'une des nouveautés, et non des moindres, c'est que ces systèmes sont dorénavant capables de nous parler et de nous répondre grâce à des techniques de reconnaissance vocale, aspect que j'aborderai dans la section suivante.

Interpellation et subjectivation

Dans une publicité pour son « assistant personnel », Google nous propose une scène familiale désormais bien connue : une mère s'adresse à une petite boîte pour organiser l'emploi du temps de sa journée ; le père l'utilise pour contrôler des objets à distance et diffuser de la musique dans toutes les pièces de la maison ; la fille a besoin d'une traduction pour ses devoirs et le fils peut lui poser toutes les questions qu'il souhaite (sur les étoiles, les animaux, etc.)¹⁹.

Ces objets « assistent » la vie quotidienne de leurs propriétaires suivant un rituel qui, au regard du succès des dispositifs comme Google Home ou Amazon Echo, a fini par se banaliser : prononciation d'une série de mots-clés (respectivement « Ok, Google » ou « Alexa ») pour activer la commande vocale et démarrer un type particulier de conversation (entendre plutôt : boucles rétroactives humain-machine). Des marchandises se mettent donc maintenant à parler, répondent aux injonctions de leurs propriétaires qui désirent par exemple connaître la météo, écouter de la musique, suivre l'état du trafic routier, en apprendre plus sur une chaîne de montagnes ou encore contrôler à distance d'autres objets connectés dans la maison (volet roulant, éclairage, thermostat, ordinateur).

La fascination est telle que l'on ne s'étonne même plus d'apprendre que les conversations privées enregistrées par ces boîtes intelligentes branchées aux bases de données de leurs constructeurs peuvent être écoutées et analysées par certains de leurs employés. Shoshana Zuboff a très justement proposé d'associer Amazon Echo ou encore Google Home à des chevaux de Troie qui font de l'assistance et de la personnalisation des arguments marketing qui voilent leur réelle mission : monétiser la vie de leurs pro-

priétaires à partir d'un enregistrement continu des conversations et des requêtes²⁰. Il y a dans les assistants personnels imaginés par Google ou Amazon comme une forme de représentation idéale de l'intelligence artificielle qui invite les individus à se livrer à une performance idéologiquement orientée et cadrée.

Sur la nature suggestive et performative de l'idéologie, le philosophe Louis Althusser propose une analyse à partir de laquelle il me semble tout à fait fécond de partir pour éclairer la compréhension des relations entre l'individu et ses technogadgets parlants. Suivant la tentative althussérienne de renouveler le schéma marxiste classique qui articule l'infrastructure (les modes de production) à la superstructure (l'idéologie), on pourrait d'abord être tenté de chercher à y positionner les assistants personnels – ce qui par ailleurs pourrait être pertinent du point de vue des conséquences de l'automatisation aussi bien dans l'organisation de la production que dans la psyché individuelle et collective. Mais c'est à partir d'un autre aspect des observations althussériennes qu'il me semble plus prometteur de partir : l'interpellation.

Dans son texte « Idéologie et Appareil Idéologique d'Etat (AIE) »²¹, Althusser montre que l'idéologie, celle qui participe au maintien de l'ordre et à la reproduction des forces productives dans les sociétés capitalistes, prend une portée existentielle : l'idéologie « recrute » quotidiennement ses sujets et conditionne un certain mode d'être et d'agir. Althusser y développe l'idée que la puissance de suggestion de l'idéologie repose sur une mécanique de pouvoir qui fonctionne à un niveau quasi-inconscient : l'interpellation. Althusser lui consacre la formule idéale du « Hé, vous, là-bas!²² » qui résume à elle seule ce qu'il entend par interpellation : l'interjection « Hé » lance autoritairement un message pour capter l'attention ; « vous » et « là-bas » constituent les embrayeurs²³ qui déterminent un sujet spécifique et sa situation spatiale. Ces embrayeurs constituent des points ou des marqueurs de subjectivation. L'interpellation présuppose alors la capacité pour le sujet de s'identifier à un « vous » avec lequel il entretient une relation existentielle puisqu'il assure sa reconnaissance au sujet de l'énoncé pris à parti par le sujet de l'énonciation. Althusser dérive cette situation archétypale de la procédure du contrôle policier qui authentifie un sujet et sa légitimité aux yeux de l'appareil d'État.

Althusser montre donc que le dispositif d'interpellation amorce un processus de subjectivation et participe à la constitution des individus en sujets sociaux. Si la procédure d'interpellation est omniprésente et imparable, c'est parce qu'elle se présente d'emblée comme un fait de langage – c'est-à-dire non pas seulement en fonction de la langue comme institution sociale qui détermine et fixe l'organisation d'un système de signes et ses renvois à des idées distinctes²⁴, mais bien comme cette faculté humaine de production et d'émission de messages aux effets variables dans le champ social.

Prise dans toute son extension, l'idéologie se présente comme une forme d'appel auquel chacun est sommé de répondre par une série de comportements, de dispositions et d'attitudes jugées licites. Suivant les milieux considérés, l'idéologie déroule un processus de subjectivation et une procédure d'identification durant laquelle le sujet sera interpellé en tant que père, mère, fils, ouvrier, exploiteur, coupable, etc.

L'application de l'analyse althussérienne de l'idéologie au cas concret des assistants personnels dits intelligents et de leurs effets apparaît particulièrement éclairante pour étudier les modalités suivant lesquelles les individus interpellent et subjectifient à leur tour ces systèmes. Il est possible de remarquer que, d'un point de vue strictement pratique, l'individu se positionne normalement toujours d'abord comme le sujet de l'énonciation en interpellant la machine intelligente, alors sujet de l'énoncé, pour activer l'échange. L'individu, sujet de l'énonciation, est censé contrôler intégralement le sujet de l'énoncé. Mais l'interpellation, si elle semble être à l'initiative de l'individu, ne limite en rien son mécanisme et sa puissance suggestive. Car en posant ces questions, aussi variées soient-elles, l'individu mène une conversation qui est constellée de marqueurs de subjectivation²⁵, raison pour laquelle d'ailleurs il semble si difficile de ne pas projeter d'intentions ou de traits de personnalité à l'assistant – quand, par ailleurs, les constructeurs programment des voix féminines et masculines singulières et prosodiques.

Tout se passe comme si l'individu se trouvait dans une forme d'auto-interpellation, sorte de curieuse situation où il se parle finalement à lui-même dans un monologue tout de même enrichi par les signaux envoyés par la machine. Dans son soliloque « augmenté », l'individu s'essaie à poser des questions, inspirées en fonction de sa situation sociale et de ses intérêts propres, à un système qui les analyse en temps réel pour formuler une réponse la plus précise et individualisée possible.

Le mécanisme d'interpellation 3.0 n'assujettit plus seulement l'individu aux normes fixées par les structures sociales en vue d'assurer la production et la reproduction des modes de vie en société, mais à des *profils* singuliers raffinés en vertu d'une analyse exhaustive de l'historique individuel (mais aussi collectif) des consommateurs, lequel influence en retour la subjectivité propre de l'individu.

En effet, ces systèmes intelligents se parent en outre d'une fonction spéculaire : l'individu se réfléchit lui-même dans l'objet connecté suivant une boucle de subjectivation qui lui donne la garantie, finalement illusoire, d'être le seul maître à bord – raison pour laquelle d'ailleurs on associe aux assistants intelligents le rôle intime du confident qui prodigue des conseils à la demande.

En donnant l'apparence de simplement assister l'individu dans ses tâches quotidiennes, de lui donner de l'information pratique ou de simplement satisfaire sa curiosité passagère, l'intelligence artificielle, à tout le moins dans les versions proposées par Google

et Amazon, instaure une temporalité éminemment profitable. À cet égard, Google Home ou Alexa se présentent comme des stratégies supplémentaires du capitalisme numérique d'investir la sphère privée en soumettant le temps passé avec la machine à des logiques de valorisation marchande. Sadin y voit très justement l'occasion inespérée pour les entreprises de développer une « relation client hyperpersonnalisée et ininterrompue²⁶ ». À la différence des publicités et autres messages promotionnels qui saturent l'attention des consommateurs, parfois avec violence, le mécanisme d'interpellation des systèmes intelligents apparaît comme doux, bienveillant et à l'écoute de leurs propriétaires immergés dans l'idéologie consumériste dont ils n'ont plus (ou trop peu) conscience.

Pourtant, à rester trop près de la conception althusserienne de l'idéologie et de son fonctionnement, on prend le risque de surestimer le rôle des opérations linguistiques dans le cadrage des modes de subjectivation dominants à l'ère de l'intelligence artificielle. Ayant très bien perçu l'influence de la cybernétique dans les sociétés modernes, Gilles Deleuze et Félix Guattari avaient proposé la notion de « mots d'ordre » pour insister sur la dimension performative du langage (suivant une relecture des thèses de John Austin²⁷) et non plus seulement sur ses aspects informatifs ou communicatifs : « L'information n'est que le strict minimum nécessaire à l'émission, transmission et observation des ordres en tant que commandements. Il faut être juste assez informé pour ne pas confondre Au feu avec Au jeu²⁸ ! » Les mots d'ordre saturent le champ social par répétition et redondance, et engagent systématiquement des actes liés à des obligations sociales (promettre, questionner, réclamer, etc.). Les journaux et les nouvelles, par exemple, « procèdent par redondance, en tant qu'ils nous disent ce qu'il "faut" penser, retenir, attendre, etc.²⁹ ».

Deleuze, dans son « Post-scriptum sur les sociétés de contrôle »³⁰, avait par ailleurs fait des mots d'ordre le trait saillant des « sociétés de contrôle » caractérisées par une sophistication, une miniaturisation et une systématisation des mécanismes de surveillance des comportements dans le monde social. Le terme même de « mots » d'ordre semble encore trop restrictif. Les ordres se miniaturisent et ne passent même plus nécessairement par le langage verbal : le simple regard du modèle sur l'affiche publicitaire suffit à susciter un acte d'achat, la voix autoritaire du président à la télévision facilite une adhésion à son programme politique, l'alerte sonore des téléphones intelligents provoque presque automatiquement la gestuelle compulsive de leurs utilisateurs. Partant d'une telle perspective, il faut bien voir derrière les suggestions, sollicitations et préconisations des systèmes intelligents toute l'efficacité d'une idéologie performative qui suscite des actes auxquels se conformer (acheter, bouger, méditer, voyager, etc.).

L'image, caricaturale à bien des égards, du paradoxe suivant lequel l'anthropomorphisation des machines s'accompagne tendanciellement d'une robotisation des humains

eux-mêmes trouve peut-être alors un appui majeur dans la démonstration deleuzo-guattarienne. En plus de participer avec enthousiasme à des jeux de question-réponse avec des outils utilisés pour enregistrer leurs moindres faits et gestes, les consommateurs intériorisent, par ricochet, tout un système d'ordre et de commandement conforme à l'idéologie dominante. Toute la question reste par ailleurs de savoir si la docilité apparente des systèmes intelligents ne renvoyait pas en creux à celle des utilisateurs eux-mêmes appelés à intérioriser des instructions claires et rigoureuses à l'attention de leurs technogadgets. Ce questionnement est d'autant plus important si l'on prend, avec sérieux, toute la mesure du futur pronostiqué par Luc Julia qui anticipe une multiplication sans fin des « assistants personnels » nécessaires à la garantie d'un confort domestique petit-bourgeois. Ce confort, Alain Damasio, écrivain de science-fiction, a très justement proposé de le qualifier de « technococon »³¹ pour souligner la relation ambiguë qui se tisse entre les individus et une partie de la technologie moderne chargée de s'assurer de notre protection, de notre sécurité et de notre bien-être dans un monde social où la pauvreté, les crises sanitaires et écologiques gagnent du terrain.

Dans la section suivante, je propose de réfléchir à ce technococon et aux relations qu'il induit en le prenant globalement à partir du point de vue de l'analyse marxienne du fétichisme de la marchandise. Ce rapprochement aura comme intérêt de réaffirmer le fait que les « assistants intelligents » répondent d'abord et avant tout à des impératifs marchands qui n'ont rien de tout à fait nouveau. En outre, il me permettra par la suite d'élargir l'analyse aux nouvelles formes d'exploitation que la fabrication de ces objets présupposent.

Fétichisme de la marchandise

Le minimalisme qui prévaut dans le design des systèmes intelligents reflète l'aseptisation des environnements collaboratifs de « *startups* » enthousiastes ou la propreté glaciale des lieux de vie et de travail de la Silicon Valley. Observer un objet comme Google Home ou Alexa, c'est contempler l'absence d'aspérité, un lissage des surfaces, un arrondi de formes généralement tubulaires – lesquelles évoquent étrangement la fameuse tour de garde intégrée au dispositif architectural (panoptique) analysé par Michel Foucault et qui permet de voir sans être vu et d'être vu sans rien voir³².

L'apparence des assistants intelligents exprime une sorte de fermeture qui rappelle de manière lancinante que le monde des marchandises est en premier lieu un monde caractérisé par une force disjonctive de séparation du sujet avec l'objet. La sobriété impeccable du « *packaging* » des assistants intelligents suppose un travail actif de

conjuraison des traces de production et de ses conditions. À suivre Fredric Jameson, cette conjuraison figure comme la condition de possibilité de la société consumériste puisqu'elle soulage la culpabilité des consommateurs qui oublient, consciemment ou inconsciemment, le travail incorporé dans leurs objets³³.

Cet aveuglement constant à l'égard des conditions mêmes de la production qui s'effacent derrière l'effarante sophistication des marchandises se présente comme la conséquence de ce que Karl Marx avait proposé d'appeler « fétichisme de la marchandise », concept sur lequel je propose de m'arrêter quelque temps pour mieux situer la place des assistants intelligents. Le fétichisme ne se comprend pas seulement comme une sorte d'adoration et de subjugation à l'égard d'objets auxquels sont attribués des qualités et des pouvoirs surnaturels. Pour Marx, est nommé « fétichisme de la marchandise » le point à partir duquel une société aliène son pouvoir dans ses propres objets dédiés à l'échange et à la consommation jusqu'à ce qu'ils finissent par diriger et dominer les rapports sociaux³⁴. Dans ces sociétés, les hommes et les femmes n'ont plus conscience du fait que leurs propres activités sociales (à commencer par les rapports de production qui ne peuvent être que sociaux) prennent une apparence d'objets qui doivent circuler et s'échanger sur des marchés dédiés. Les rapports humains prennent la forme de choses tandis que les choses, en voilant le caractère éminemment social de leur production, semblent décider d'elles-mêmes les modalités de ces rapports.

À conclure un peu trop hâtivement, le fétichisme de la marchandise serait alors considéré comme source d'« illusion » et de « fausse conscience » pour les individus vivant dans les sociétés capitalistes qui ne parviennent plus à déterminer le caractère social masqué par les marchandises. Du côté de l'économie bourgeoise, le fétichisme de la marchandise apparaît effectivement comme une illusion à partir du moment où les formes fétichisées que sont l'argent, la valeur et la marchandise sont légitimées comme étant naturelles puisqu'elles relèvent d'une nécessité transhistorique. Du côté marxiste, le fétichisme de la marchandise résonne avec la fausse conscience dont sont victimes les masses prolétaires qui sont appelées à devenir les véritables sujets de l'histoire. Or, dans un cas comme dans l'autre, le fétichisme de la marchandise ne doit pas simplement se comprendre comme les moyens dont disposent les sociétés capitalistes et ses dirigeants pour produire de fausses représentations et leurrer les masses.

Jean Baudrillard, dont l'analyse va maintenant retenir mon attention, critique justement l'interprétation à la fois libérale et marxiste du fétichisme de la marchandise. Ce dernier se comprendrait principalement comme une mystification de la conscience devant la puissance des marchandises auxquelles sont incorporées des valeurs « "artificielles", libidinales ou de prestige³⁵ ». Tout comme Marx s'en était justifié, le fétichisme de la marchandise n'est pas un phénomène simple à expliquer et s'arrêter aux signifi-

tions les plus évidentes bloquent l'accès à une compréhension avancée d'un tel processus : « Renvoyer tous les problèmes du "fétichisme" aux mécanismes super-structuraux de la "fausse conscience", c'est s'ôter toute chance d'analyser le véritable *procès de travail idéologique*³⁶. » Par un détour éclairant à l'étymologie même du terme *fétiche*, Baudrillard s'attache à démontrer le glissement de sens qui s'est opéré à la faveur d'une conception limitée aux propriétés illusives surnaturelles investies dans les objets. Il est vrai que, dans la société de consommation où l'écoulement des marchandises est d'une nécessité vitale pour la perpétuation du système capitaliste, l'association d'un objet à un imaginaire arbitraire capable de susciter l'acte d'achat évoque le recours fréquent à ce premier niveau de définition. Pourtant, insiste Baudrillard, avant d'évoquer la fascination qu'exercent les marchandises sur celles et ceux qui les contemplent, le fétiche doit d'abord être référé à « une *fabrication*, un artefact, un travail d'apparences et de *signes*³⁷ ». En outre, le fétiche présuppose « un travail culturel de signes³⁸ » qui utilise le trucage, l'artificiel, la feintise, le faux ou encore, pour reprendre un autre concept de Baudrillard, le simulacre.

Il faut donc ramener la définition du fétichisme au niveau de la fabrication des signes qui, en s'incorporant dans l'objet, effacent toute trace de cette fabrication et semblent devenir inatteignables. Autrement dit, il faut en venir à un travail de dé-fétichisation du terme lui-même en l'inscrivant dans le cadre d'une critique de l'économie politique du signe :

[P]artout les objets sont donnés et reçus comme dispensateurs de forces (bonheur, santé, sécurité, prestige) : cette substance magique partout répandue fait oublier que ce sont d'abord des signes, un code généralisé de signes, un code totalement arbitraire (*faictice*, « fétiche ») de différences, et que c'est de là, et pas du tout de leur valeur d'usage, ni de leurs « vertus » infuses, que vient la fascination qu'ils exercent³⁹.

En partant des catégories saussuriennes du signe (Signifiant/Signifié en lien avec le rapport Valeur d'échange/Valeur d'usage), Baudrillard tient à revenir à l'analyse de Marx qui montre que le fétichisme ne s'applique pas à la marchandise en tant qu'objet porteur de valeurs arbitraires, mais en tant que matrice du système capitaliste qui compte le travail abstrait et la valeur (capitaliste) comme ses « abstractions réelles » les plus fondamentales. Le fétichisme est moins à corréler au signifié (valeur d'usage) qu'au signifiant (valeur d'échange) qui implique un système codé, différentiel et stable. Si bien que, dans le fétichisme, c'est d'abord une passion des codes avant d'être une passion des valeurs⁴⁰.

Le fétichisme de la marchandise se réfléchit comme un processus de généralisation d'un code structurel inscrit dans la logique de valorisation capitalistique. Dès lors,

l'adoration ou la *sacralisation* ne porte pas sur un objet particulier, mais sur la totalité du système exprimée par la marchandise en tant que « forme cellulaire économique⁴¹ » du capitalisme.

Ce détour par le fétichisme de la marchandise nous place tout de même face à une situation encore pour le moment indécidable quant à la place qu'occupent les assistants intelligents dans un tel système : faut-il les voir simplement comme de nouveaux objets qui viennent rejoindre un marché *high-tech* saturé par une offre de techno-gadgets déjà considérable ou bien faut-il les considérer plus sérieusement comme des marchandises qui signalent qu'un seuil a été franchi dans le processus de renversement de la réalité? L'analyse marxienne, croisée avec celle de Baudrillard, nous permet déjà de prendre position quant à ce questionnement : Google Home, Amazon Echo et consorts illustrent sans peine l'idée suivant laquelle, dans les sociétés capitalistes les marchandises donnent l'étrange impression qu'elles mènent leur propre existence et dirigent la vie des individus chargés de les fabriquer et de les échanger. De la même manière, la dévotion de plus en plus présente dans l'espace médiatique à l'égard de l'intelligence artificielle et de ses capacités *a priori* salvatrices rappelle singulièrement l'analogie de l'aliénation religieuse évoquée par Marx pour comprendre comment des créations humaines peuvent finir par diriger la destinée du monde social.

En ignorant ou oubliant le caractère foncièrement social de la production humaine commence donc alors un processus de renversement non pas de la représentation de la réalité, mais de la réalité elle-même : les rapports sociaux deviennent des rapports entre des choses puisque les individus ne se rencontrent sur le marché que pour échanger des marchandises dont les conditions d'échange et de production échappent au contrôle de l'activité humaine. Dans ces formes paroxystiques, ce processus de renversement de la réalité a vu le couronnement de ce que Jean Baudrillard qualifiait de « simulacre »⁴² (suivant lequel le réel a tout simplement disparu) ou ce que Guy Debord avait proposé de nommer « société de spectacle »⁴³ qui consacre l'image comme la forme ultime de la marchandise. Le régime spectaculaire de la marchandise se caractérise par une disjonction toujours plus forte entre la valeur d'usage et la valeur d'échange, ou dans un sens plus baudrillardien, entre le signe et son référent⁴⁴. Je propose dans la section suivante de questionner plus particulièrement cette disjonction à travers la mise en lumière des formes d'exploitation inédites présumées par la confection des assistants intelligents. Cette dernière section me donnera l'occasion d'introduire les notions de sémiocapitalisme et de *digital labor*.

Asservissement machinique et exploitation du « *digital labor* »

Associer les nouvelles technologies à un imaginaire de liberté et d'émancipation tend à masquer l'activité « d'un prolétariat nomade » fortement internationalisé et qui ne compte généralement pour rien aux yeux du capital⁴⁵. La fabrication des technologies nécessaires au maintien et à l'extension du sémiocapitalisme implique une grande masse d'ouvriers employée dans les zones du monde où le coût du travail est avantageux pour les fabricants. Bien évidemment, ce constat s'observe également dans la fabrication des biens matériels, mais s'en trouve renforcé dans le domaine des nouvelles technologies et du numérique. Des entreprises comme Apple ou Microsoft ont su par exemple tirer profit d'une division des tâches à l'international : ici, les travailleurs fabriquent le code, les logiciels, imaginent le contenu des campagnes marketing, l'habillage des magasins ; là-bas, les ouvriers extraient les matières premières ; ailleurs, ils assemblent les matériaux sur des chaînes de montage partiellement robotisées.

Cependant, la disjonction entre la marchandise (notamment son apparence, design, marketing) et ses conditions réelles de production semble être poussée à son intensité maximale avec les assistants intelligents dont la puissance *aléthique* repose notamment sur un niveau d'exploitation tout à fait inédit. L'ingénierie complexe des assistants intelligents, la miniaturisation et le perfectionnement virtuellement illimité de leurs composants (puces, processeurs, blocs mémoires, etc.) ne seraient pas grand-chose sans la collecte et le traitement de masses de données diversifiées qui donnent à l'intelligence artificielle autant d'occasions d'expertiser les aspects du monde social numérisé. Elle ne serait pas grand-chose non plus sans l'exploitation d'une masse considérable de petites mains, des « *clickworkers* » (travailleurs du clic)⁴⁶ chargés de réaliser les tâches répétitives et abrutissantes (entourer des objets sur une image, catégoriser des données, répondre à des questionnaires, traduire des fiches produits, identifier des contenus violents et pornographiques sur Facebook ou Twitter, etc.). Pour les assistants intelligents, les travailleurs du clic sont notamment enrôlés pour lire des phrases à voix haute afin d'améliorer les outils de reconnaissance vocale, l'« analyse de sentiments » (*sentiment analysis* ou *opinion mining*) ou plus généralement le « traitement automatique de la langue naturelle » (*Natural Language Processing*). Ils accompagnent donc les propriétaires des assistants intelligents, qui au fil de leurs usages finissent également par entraîner l'intelligence artificielle à être plus précise dans l'interpellation de leurs désirs et besoins. Ces objets techniques concentrent de fait les différents traits caractéristiques que Gary Genosko assigne à la sémiomarchandise⁴⁷ : (1) production massive de segments sémiotiques intimement liés la subjectivité des producteurs comme des consommateurs ; (2) connexion permanente au réseau informatique à l'échelle mon-

diale ; (3) fusion et indistinction des séquences temporelles liées au travail, au loisir ou au repos. De ce point de vue, les assistants intelligents figurent comme une marchandise supplémentaire au sein de ce que Genosko, à la suite de Franco Berardi, a proposé de nommer *sémiocapitalisme*. Dorénavant, la valorisation du capital fonctionnerait également à partir de la capacité créative, la communication et la personnalité des travailleurs investis dans la fabrication de marchandises immatérielles ou « sémiomarchandise » (logiciels, jeux vidéo, téléphones intelligents, etc., qui condensent des fragments sémiotiques les plus divers dans un artéfact dédié à la vente et la consommation). En m'appuyant sur les concepts deleuzo-guattariens de décodage (dissolution des facteurs extra-économiques et symboliques sans intérêt pour le capital) et d'axiomatisation (soumission des activités sociales au capital comme axiome-maître), j'avais proposé ailleurs⁴⁸ de définir le sémiocapitalisme comme le seuil à partir duquel l'appropriation de la *sémiose* individuelle et collective est devenue nécessaire pour ouvrir de nouvelles sphères de valorisation marchande – une *sémiose* rendue immanente à ces processus de valorisation, c'est-à-dire inextricablement liée au capital. D'un point de vue plus spécifiquement peircien, le processus sémiosique (relation triadique entre un objet, un représentamen et son interprétant) tourne en rond et n'accroche du réel que les éléments nécessaires qui pourront être profitables au sens économique du terme.

Les travailleurs du clic incarnent finalement la force de travail rêvée par le capital : absence totale de sécurité sociale, aucune dépense nécessaire pour les outils de production et les espaces de travail (utilisation des ordinateurs et/ou téléphones personnels achetés par les travailleurs eux-mêmes), réalisation des tâches à n'importe quelle heure du jour ou de la nuit pour un salaire qui oscille entre 3 et 5 dollars de l'heure⁴⁹. Cette classe de travailleurs est employée pour réaliser des tâches très précises qui visent à « entraîner » l'intelligence artificielle. Cette curieuse mise en abyme où l'on « apprend la machine à apprendre » passe notamment par des plateformes développées par Figure Eight ou Amazon Mechanical Turk qui tirent profit de la précarisation croissante de toute une masse d'individu précarisés prêts à occuper leurs temps pour mener un travail que le sociologue Antonio Casilli qualifie de « *tâcheronnisé et datafé* »⁵⁰ et qu'il rattache avec Dominique Cardon à l'idée d'un « *digital labor* »⁵¹. Cette forme d'exploitation, rendue notamment possible par les développements informatiques, résulte surtout d'une conjonction croissante entre l'externalisation et la fragmentation dans la division sociale du travail moderne.

Dans les années 1990, Fredric Jameson avait déjà proposé de revoir les conséquences du mode de domination du capitalisme moderne tardif : non plus seulement l'aliénation de l'individu (isolement, solitude, anomie), mais sa fragmentation (*burn-out*, autodestruction)⁵². Dans cette optique, il n'y a même plus lieu de continuer à parler

d'individu, mais, conformément aux analyses de Günther Anders et plus tard de Gilles Deleuze, de « dividual » ; conséquence d'une transformation du sujet moderne en « une pluralité de fonctions séparées⁵³ », réduit à des échantillons et des données au sein des sociétés de contrôle⁵⁴. Les dividuals, dans leurs rapports à la totalité du système, sont compris comme autant de « pièces composantes intrinsèques, des "entrées" et des "sorties", des *feed-back* ou des récurrences⁵⁵ ». Le *digital labor* exprime ce que Félix Guattari qualifiait d'asservissement machinique, lequel se manifeste par un déclassement de l'humain relégué à une position subalterne, pièce ou rouage de la machine asservi au fonctionnement d'ensemble. À l'image des masses de prolétaires soumises aux lignes d'assemblage et à leurs cadences infernales, s'ajoute celle, plus moderne, des individus aliénés à leurs postes de travail informatisés.

Félix Guattari avait bien vu que cet asservissement appuyé par des dispositifs technologiques de toute nature, avait atteint un seuil critique : modelage systématique des affects et des désirs à une échelle moléculaire et dont l'effet principal consiste dans le déclenchement, l'activation ou le déclic qui produit une répétition de gestes, d'attitudes et de comportements nécessaires au maintien du système et à sa reproduction.

Pour insister justement sur la dimension quasi-inconsciente de ces processus, Guattari avait proposé dans les années 1980 de caractériser une série de signes qui fonctionnent en deçà des systèmes de représentation et d'interprétation : il parlait alors des « sémiotiques a-signifiantes ». À suivre Guattari, il n'y a plus vraiment lieu de parler de signes, mais plutôt de points-signes, de signes partiels, particules, ou a-signifiants, c'est-à-dire des signes qui ne fournissent pas nécessairement de contenu sémantique et qui constituent des « entités qui sont passées en deçà des coordonnées de temps, d'espace et d'existence⁵⁶ ».

Si l'analyse du rôle et du fonctionnement de ces sémiotiques ont trouvé un nouveau souffle dans les travaux de Gary Genosko⁵⁷ ou de Maurizio Lazzarato⁵⁸ qui analysent les mutations du capitalisme moderne tardif, c'est bien parce que Guattari, avec l'idée d'a-signification, avait réussi à établir un parallèle éclairant entre la mise en marche automatisée du code informatique, des signaux électriques, des algorithmes et les fonctionnements psychiques eux-mêmes. Les signes-particules sont peut-être difficiles à identifier dans le monde social, en revanche leurs effets le sont moins. L'assistant personnel forme un exemple paradigmatique : activation de l'objet par commande vocale, déclenchement de protocoles de connexion à la base de données, activation des algorithmes qui analysent les questions, formulation de suggestions, enregistrement de l'historique de conversation, etc. Parallèlement à toutes ces procédures, la mise en marche de la machine et toute l'expérience interpelle l'individu, génère des affects et des passions, module des charges libidinales (érotisation des voix féminines ou masculines),

contrôle et régule des comportements et des attitudes, etc. Cet exercice de pensée vise à démontrer que toutes les procédures de mise en marche de la machine, articulées à des sémiotiques a-signifiantes très hétérogènes, excèdent de loin la simple conversation humain-machine. Pour Guattari, présupposer l'existence d'un tel régime d'a-signification signale le besoin pressant de s'affranchir de la séparation sujet-objet pour mieux saisir les rapports de présupposition, de continuité et d'influence entre les flux sémiotiques et les flux matériels. Fidèle aux enseignements guattariens, Lazzarato avait très justement observé que les sémiotiques a-signifiantes « constituent le point de focalisation d'une proto-énonciation et d'une proto-subjectivité [dans la mesure où elles] suggèrent, permettent, sollicitent, provoquent, encouragent et empêchent certaines actions, pensées, affects ou en promeuvent d'autres⁵⁹ ».

Mais chez Guattari, la particularité des sémiotiques a-signifiantes ne repose pas tant sur les effets d'automatisation qu'elles rendent possibles (aussi bien dans la psyché que dans les appareils techniques), mais dans leur capacité à établir des nouvelles connexions avec « les flux matériels les plus déterritorialisés⁶⁰ ». Les sémiotiques a-signifiantes invitent à une méditation réflexive portant sur les interactions plus ou moins lointaines, plus ou moins directes, qui peuvent s'établir entre ces flux matériels. Dans le cas des assistants personnels, ce sont des flux sonores (langue, intonation, rythme, etc.) qui activent des flux électriques circulant à travers des matériaux et des composants diversifiés (plastique, cuivre, lithium, tantale, etc.) des assistants personnels, lesquels déterritorisent ces flux sonores, analysés en temps réels par des protocoles algorithmiques et enregistrés dans les bases de données des constructeurs. D'un point de vue guattarien, les sémiotiques a-signifiantes articulent également des flux de désir, financiers, énergétiques constellés de points-signes qui forment des points d'intersection, des carrefours à l'origine de l'ouverture ou de la fermeture de tout un univers de possible. Le régime d'a-signification se présentait pour Guattari comme un point de passage vers l'altérité radicale, tout à fait capable d'enrichir la subjectivité, et de découvrir des territoires existentiels et des univers de valeurs inédits. À l'inverse, un tel régime est aussi susceptible, comme le montrent notamment les théoriciens du sémiocapitalisme, de renforcer la superficialité d'un univers de valeur restreint et stéréotypé.

À cet égard, Guattari n'avait cessé de questionner les raisons pour lesquelles les « immenses potentialités processuelles portées par toutes ces révolutions informatique, télématique, robotique, bureautique, biotechnologique... n'aboutissent encore, jusqu'à présent, qu'à un renforcement des systèmes antérieurs d'aliénation, à une mass-médiatisation oppressive, à des politiques consensuelles infantilisantes⁶¹ ». Cette question guattarienne, toujours aussi brûlante, s'articule notamment à ses propres analyses portant sur un « capitalisme mondial intégré » dont la puissance est inséparable de sa

propension à s'incorporer à tous les niveaux du champ social, y compris au domaine des assistants vocaux qui nous intéresse ici.

En fin de compte, si le capitalisme est parvenu à si bien s'adapter au fil des temps, ce n'est pas seulement à travers toute l'efficacité de ses opérations par lesquelles il est parvenu à s'étendre, mais aussi grâce à sa capacité à s'intégrer aux modes de vie les plus singuliers. De ce point de vue, le travail du clic (*clicworking*), nécessaire pour permettre aux assistants intelligents de fonctionner, figure comme une nouvelle forme d'asservissement mécanique (ou suivant la formule consacrée : « *humans-in-the-loop* ») et participe à l'histoire tristement cynique du sémiocapitalisme : les travailleurs du clic entraînent finalement une intelligence artificielle qui finira, à terme, par complètement les remplacer.

Conclusion

Il y a déjà plus d'un siècle, le philosophe Georg Lukács observait chez les individus la généralisation d'une sorte d'« attitude contemplative », sorte d'état dans lequel les individus, soumis à une division d'un travail rationalisé et standardisé, ne peuvent plus que regarder un monde d'objets qui ne leur appartiennent plus. L'individu se désactive et s'incorpore comme une « partie mécanisée dans un système mécanique qu'il trouve devant lui, achevé et fonctionnant dans une totale indépendance par rapport à lui⁶² ». Tout au long de cet article, j'ai cherché à inscrire l'intelligence artificielle dans la filiation de cette critique du monde social qui voit dans le développement de la rationalité calculatrice et froide des sociétés capitalistes un mouvement de désenchantement du monde social. L'intelligence artificielle, telle qu'elle se développe majoritairement, participe d'une vision d'un monde assujéti à des règles finalement arbitraires et aux impératifs de la valorisation marchande. Confrontés à l'*alètheia* algorithmique qui donne l'apparence rassurante d'une saisie du réel et de la vérité de l'étant, les individus se laissent porter par les recommandations et les sollicitations des systèmes intelligents comme autant d'invitations à profiter des ressources virtuellement infinies de la société consumériste. Dorénavant les marchandises parlent et assistent leurs propriétaires qui y contemplent le reflet de leur profil individuel ; voilà donc le signe d'un fétichisme aveuglant offrant une territorialité résiduelle pour mieux conjurer l'angoisse de la perte de sens qui cerne le sujet de part en part. Il n'y a pas de raison de voir dans l'anthropomorphisation de machines plus à l'écoute, programmées pour se soucier du bien-être d'une certaine partie de la population qui y a accès, une alternative souhaitable. À l'instar de l'actuel phénomène de ludification qui vise à intégrer des mécaniques de jeu dans tous les aspects du monde

vécu (sport, tâches domestiques, relations sociales et surtout travail), l'anthropomorphisation des machines participe d'une même tendance infantile à vouloir produire des points de référence existentiels pour dompter l'angoisse d'un monde social qui cultive le vide⁶³, la superficialité et, dans un sens plus baudrillardien, le simulacre ; évidemment ces deux phénomènes ne sont pas choisis au hasard et partagent un fond commun, à savoir celui de produire une disposition d'esprit adaptée aussi bien à la productivité qu'à la consommation en tenant les individus dans un domaine restreint de jouissance par la production d'affects joyeux⁶⁴.

Sadin voit dans l'IA l'occasion de nous déposséder de notre capacité à décider collectivement de notre destin. Dans le même état d'esprit, Alain Damasio y voit plutôt le simulacre d'un surplus de contrôle et de confort privé, corrélatif d'une perte de contrôle sur le monde vécu lui-même. Au sortir de cette analyse, je privilégie l'analyse plus ancienne de Guattari qui remarque que l'individu, dans les sociétés modernes, n'est plus préparé à entretenir un rapport avec la finitude, la précarité, l'altérité ou même le cosmos, mais est seulement entraîné à s'approprier des objets préparés et mis de l'avant par le système. L'imperméabilité des rapports individuels et collectifs à toute une multitude de registres existentiels, sous couvert d'un rapport hyper-rationalisé au monde vécu, est le signe paradoxal d'une infantilisation de la subjectivité capitaliste fermée à la précarité et à la finitude. C'est en tout cas dans ce sens que j'ai tenté d'entrevoir l'IA, en partant de l'exemple des assistants personnels intelligents, lesquels poursuivent avec force le processus de renversement du monde social qui confie dorénavant une partie de son destin à des marchandises parlantes.

Bibliographie

- ALTHUSSER, Louis, *Positions*, Paris, Éditions sociales, 1976.
- ANDERS, Günther, *L'obsolescence de l'homme. Sur l'âme à l'époque de la deuxième révolution industrielle*, trad. de l'allemand par C. David, Paris, L'encyclopédie des nuisances/lvrea, 2002 [1956].
- AUSTIN, John L., *Quand dire c'est faire*, trad. de l'anglais, intro. et commentaire par G. Lane, Paris, Seuil, coll. « L'Ordre philosophique », 1970 [1962].
- BAUDRILLARD, Jean, *Pour une critique de l'économie politique du signe*, Paris, Gallimard, 1972.
- , *Simulacres et simulation*, Paris, Galilée, coll. « Débats », 1981.
- BERARDI, Franco, *The Soul at Work: From Alienation to Autonomy*, Los Angeles, Semiotext(e), 2009.
- CARDON, Dominique & Antonio CASILLI, *Qu'est-ce que le Digital Labor?*, Bry-sur-Marne, INA, 2015.
- CASILLI, Antonio, *En attendant les robots. Enquête sur le travail du clic*, Paris, Seuil, 2019.
- CASTORIADIS, Cornelius, *L'institution imaginaire de la société*, Paris, Seuil, 1975.
- DAMASIO, Alain, *Les Furtifs*, Clamart, La Volte, 2019.
- DEBORD, Guy, *La société du spectacle*, Paris, Gallimard, coll. « Folio », 1992 [1967].
- DELEUZE, Gilles, « Post-scriptum sur les sociétés de contrôle », *L'autre journal*, no 1, 1990, p. 2-7.
- DELEUZE, Gilles & Félix GUATTARI, *Capitalisme et Schizophrénie, 2 : Mille Plateaux*, Paris, Minuit, coll. « Critique », 1980.
- DYER-WITHEFORD, Nick, *Cyber-Proletariat: Global Labour in the Digital Vortex*, Londres, Pluto Press, 2015.
- DYER-WITHEFORD, Nick, Atle Mikkola KJOSEN & James STEINHOFF, *Inhuman Power: Artificial Intelligence and the Future of Capitalism*, Londres, Pluto Press, 2019.
- ECO, Umberto, *Les limites de l'interprétation*, trad. de l'italien par M. Bouzaher, Paris, Grasset & Fasquelle, 1992.
- FOUCAULT, Michel, *Surveiller et punir*, Paris, Gallimard, 1975.
- GABRIEL, Markus, *Pourquoi la pensée humaine est inégalable*, Paris, JC Lattès, 2019.
- GENOSKO, Gary, *Critical Semiotics: Theory, from Information to Affect*, New York, Bloomsbury Academic, 2016.

- GIBSON, Adam & Josh PATTERSON, *Deep Learning: A Practitioner's Approach*, Sebastopol, O'Reilly, 2017.
- GUATTARI, Félix, *Cartographies schizoanalytiques*, Paris, Galilée, 1989.
- , *La Révolution moléculaire*, Paris, Les Prairies ordinaires, coll. « Essais », 2012 [1977].
- , *Lignes de fuite : pour un autre monde de possibles*, La Tour-d'Aigues, l'Aube, coll. « Aube poche essai », 2014 [2011].
- HARDT, Michael & Antonio NEGRI, *Multitude : Guerre et démocratie à l'âge de l'empire*, trad. de l'anglais par N. Guilhot, Paris, La Découverte, 2004.
- JAKOBSON, Roman, *Essais de linguistique générale, I : Les fondations du langage*, trad. de l'anglais et préfacé par N. Ruwet, Paris, Minuit, 2003 [1963]
- JAMESON, Fredric, *Le postmodernisme ou la logique culturelle du capitalisme tardif*, trad. de l'anglais (États-Unis) par F. Nevoltry, Paris, Beaux-Arts de Paris, 2018 [1991].
- JULIA, Luc, *L'intelligence artificielle n'existe pas*, Paris, First, 2019.
- KURZWEIL, Ray, *Humanité 2.0 : La Bible du changement*, trad. de l'anglais (États-Unis) par A. Mesmin, Paris, M21, 2007.
- LAZZARATO, Maurizio, *Signs and Machines. Capitalism and the Production of Subjectivity*, Los Angeles, Semiotext(e), 2014.
- LE DÉVÉDEC, Nicolas, *La société de l'amélioration. Le renversement de la perfectibilité humaine, de l'humanisme des lumières à l'humain augmenté*, Montréal, Liber, 2015.
- LORDON, Frédéric, *Capitalisme, désir et servitude. Marx et Spinoza*, Paris, La Fabrique, 2010.
- LUKÀCS, Georg, *Histoire et conscience de classe. Essais de dialectique marxiste*, trad. de l'allemand par K. Axelos et J. Bois, préface de K. Axelos, Paris, Minuit, 1976 [1923].
- MARX, Karl, *Le Capital : Livre 1*, trad. de l'allemand par J. Roy, éd. de M. Rubel, Paris, Gallimard, coll. « Folio essais », 1968 [1867].
- MICROSOFT, *Intelligence artificielle : guide de survie (Comprendre, raisonner et interagir autrement avec l'IA)*, Issy-Les-Moulineaux, 2018. En ligne : <<https://info.microsoft.com/rs/157-GQE-382/images/FR-CNTNT-eBook-MicrosoftLivreBlancGuidedesurviedellIntelligenceArtificielle.pdf>>.
- ORGANISATION INTERNATIONALE DU TRAVAIL (OIT), *Digital labour platforms and the future of work: Towards decent work in the online world*, Genève, 2018. En ligne : <https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_645337.pdf>.

- RICHERT, Fabien, « Le Big Data à l'ère du sémiocapitalisme : décodage et axiomatisation généralisés du champ social », dans A. Mondoux & M. Ménard (dir.), *Big Data et société : Industrialisation des médiations symboliques*, Montréal, Presses de l'Université du Québec, 2018, p. 166-186.
- ROUVROY, Antoinette & Thomas BERNIS, « Gouvernamentalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation? », *Réseaux*, vol. 1, no 177, 2013, p. 168-173.
- SADIN, Éric, *La Vie algorithmique. Critique de la raison numérique*, Paris, L'Échappée, 2015.
- , *L'Intelligence artificielle ou l'enjeu du siècle : Anatomie d'un antihumanisme radical*, Paris, L'Échappée, 2018.
- SAUSSURE, Ferdinand de, *Cours de linguistique générale*, éd. de C. Bally & A. Sechehaye avec la collab. d'A. Riedlinger, préface de J.-D. Urbain, Paris, Payot, coll. « Petite bibliothèque Payot / Classiques », 2016 [1916].
- SUPIOT, Alain, *La gouvernance par les nombres*, Paris, Fayard.
- WIENER, Norbert, *Cybernétique et société : l'usage humain des êtres humains*, trad. de l'anglais par P.-Y. Mistoulon, Paris, UGE, « 10/18 », 1962 [1954].
- ZUBOFF, Shoshana, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New-York, PublicAffairs, 2019.

Notes

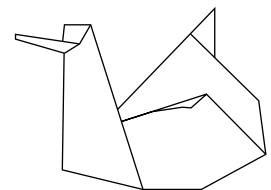
- 1 L. JULIA, *L'intelligence artificielle n'existe pas*, Paris, First, 2019, p. 55.
- 2 Les techniques de *deep learning* constituent une spécialisation des techniques d'apprentissage machine (*machine learning*), elles-mêmes rapportées au champ plus vaste de l'intelligence artificielle. Les techniques de *machine learning* sont généralement utilisées pour « extraire de l'information à partir de données brutes afin de les représenter dans une sorte de modèle ». A. GIBSON & J. PATTERSON, *Deep Learning: A Practitioner's Approach*, Sebastopol, O'Reilly, 2017, p. 1 (trad. libre). Le *deep learning* est caractérisé par un niveau de sophistication supplémentaire dans la réalisation de ces tâches.
- 3 L. JULIA, *L'intelligence artificielle n'existe pas*, op. cit., p. 64.
- 4 Voir la lettre ouverte publiée sur le site Internet du Future of Life Institute (<<https://futureoflife.org/ai-open-letter/>>). La lettre est signée par plus de 8 000 personnes dont Elon Musk, Stephen Hawking, Nick Bostrom ou encore Steve Wozniak.
- 5 L. JULIA, *L'intelligence artificielle n'existe pas*, op. cit., p. 65.
- 6 *Idem*.
- 7 *Ibid.*, p. 87-93.
- 8 Je renvoie notamment à la lecture du rapport *Intelligence artificielle : guide de survie (Comprendre, raisonner et interagir autrement avec l'IA)* publié en 2018 par Microsoft France Issy-Les-

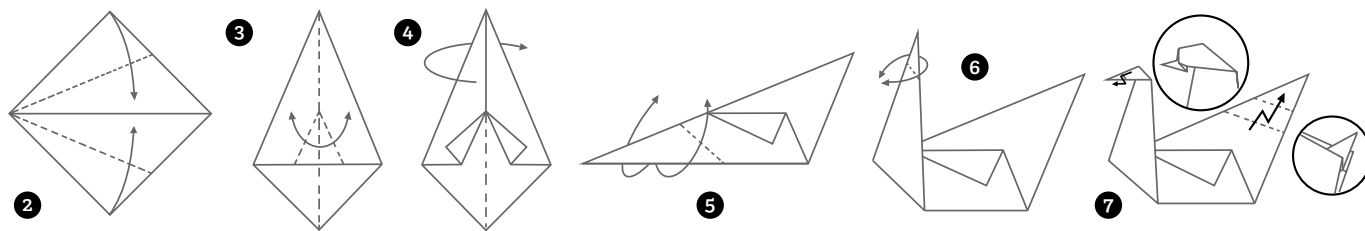
- Moulineaux disponible en ligne : <<https://info.microsoft.com/rs/157-GQE-382/images/FR-CNTNT-eBook-MicrosoftLivreblancGuidedesurviedelIntelligenceArtificielle.pdf>>. Ce rapport présente, avec une certaine clarté, les attendus du développement de l'intelligence artificielle qui se révèle être un levier imparable pour établir de nouveaux modèles d'affaires suivant les principes guides d'une économie comportementale (corrélation entre les états émotionnels, affectifs, culturels et sociaux des individus et leurs décisions économiques).
- 9 É. SADIN, *L'Intelligence artificielle ou l'enjeu du siècle : Anatomie d'un antihumanisme radical*, Paris, L'Échappée, 2018.
 - 10 *Ibid.*, p. 13.
 - 11 É. SADIN, *La Vie algorithmique. Critique de la raison numérique*, Paris, L'Échappée, 2015, p. 28.
 - 12 A. SUPIOT, *La gouvernance par les nombres*, Paris, Fayard, p. 125.
 - 13 A. ROUVROY & T. BERNIS, « Gouvernamentalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation? », *Réseaux*, vol. 1, no 177, 2013, p. 168-173.
 - 14 Profondément marqué par la Shoah et par les désastres du nucléaires, Norbert Wiener avait déclaré sentencieusement que nous nous ne serions que « des naufragés sur une planète vouée à la mort ». N. WIENER, *Cybernétique et société : l'usage humain des êtres humains*, trad. de l'anglais par P.-Y. Mistoulon, Paris, UGE, « 10/18 », 1962 [1954], p. 43. Pour les cybernéticiens, la fabrication d'une machine intelligence capable de gouverner et de contrôler les turbulences du monde (l'entropie) et les erreurs humaines était devenue la seule option envisageable.
 - 15 Le Dévédec montre bien qu'à la fragilité humaine répond une perfectibilité technique source de toutes les fascinations : « L'aspiration immémoriale à la perfectibilité technique tient donc, pour les transhumanistes à la nature imparfaite de l'être humain. Prenant appui sur une véritable anthropologie de la déficience, les représentants de cette conviction conçoivent en effet l'homme comme un être inadapté à son environnement, au corps lacunaire, et à l'existence originellement presque inimaginable tant elle serait défailante. » N. LE DÉVÉDEC, *La société de l'amélioration. Le renversement de la perfectibilité humaine, de l'humanisme des lumières à l'humain augmenté*, Montréal, Liber, 2015, p. 204. Le philosophe Günter Anders avait proposé de voir dans ce fossé irréconcilable entre l'humain et la machine les traits caractéristiques de la « honte prométhéenne » : « la honte qui s'empare de l'homme devant l'humiliante qualité des choses qu'il a lui-même fabriquées. » G. ANDERS, *L'obsolescence de l'homme. Sur l'âme à l'époque de la deuxième révolution industrielle*, trad. de l'allemand par C. David, Paris, L'encyclopédie des nuisances/Ivrea, 2002 [1956], p. 37.
 - 16 É. SADIN, *L'Intelligence artificielle ou l'enjeu du siècle*, op. cit., p. 23.
 - 17 Je pense notamment aux analyses de Foucault portant sur les rapports entre la vérité et la norme dans les sociétés disciplinaires, M. FOUCAULT, *Surveiller et punir*, Paris, Gallimard, 1975.
 - 18 É. SADIN, *L'Intelligence artificielle ou l'enjeu du siècle*, op. cit., p. 56.
 - 19 Je m'appuie notamment sur le visionnage d'une publicité officielle de Google Home présentée lors de la conférence annuelle de Google en 2016 : <<https://www.youtube.com/watch?v=OsXedJqlaWE>>.
 - 20 S. ZUBOFF, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New-York, PublicAffairs, 2019, p. 246.
 - 21 L. ALTHUSSER, *Positions*, Paris, Éditions sociales, 1976.
 - 22 Pour une définition plus large : « Nous suggérons alors que l'idéologie "agit " ou "fonctionne" de telle sorte qu'elle "recrute" des sujets parmi les individus (elle les recrute tous), ou "transforme" les individus en sujets (elle les transforme tous) par cette opération très précise que nous appelons l'interpellation, qu'on peut se représenter sur le type même de la plus banale opération policière (ou non) de tous les jours : "hé, vous, là-bas!". Si nous supposons que la scène théorique imaginée

- se passe dans la rue, l'individu interpellé se retourne. Par cette simple conversion physique de 180 degrés, il devient sujet. Pourquoi? Parce qu'il a reconnu que l'interpellation s'adressait "bien" à lui, et que "c'était bien lui" qui était interpellé (et pas un autre). » *Ibid.*, p. 113-114.
- 23 Ce terme est une traduction proposée par Nicolas Ruwet de l'anglais « *shifter* » que le linguiste Roman Jakobson emprunte lui-même à Otto Jespersen. Il s'agit d'une classe de mots qui donne à l'émetteur et au destinataire la possibilité de se manifester dans un énoncé et de préciser les paramètres spatio-temporels coextensifs à la situation de l'énonciation. Voir R. JAKOBSON, *Essais de linguistique générale, I : Les fondations du langage*, trad. de l'anglais et préfacé par N. Ruwet, Paris, Minuit, 2003 [1963].
- 24 F. de SAUSSURE, *Cours de linguistique générale*, éd. de C. Bally & A. Sechehaye avec la collab. d'A. Riedlinger, préface de J.-D. Urbain, Paris, Payot, coll. « Petite bibliothèque Payot / Classiques », 2016 [1916], p. 75.
- 25 Pour saisir l'importance et la redondance de ces marqueurs, voir la liste complète des commandes vocales que Google est capable de traiter sur <<https://www.cnet.com/how-to/google-home-complete-list-of-commands-so-far/>>.
- 26 É. SADIN, *L'Intelligence artificielle ou l'enjeu du siècle*, op. cit., p. 92.
- 27 J. AUSTIN, *Quand dire c'est faire*, Paris, Seuil, coll. « Essais », 1970 [1962].
- 28 G. DELEUZE & F. GUATTARI, *Capitalisme et Schizophrénie, 2 : Mille Plateaux*, Paris, Minuit, coll. « Critique », 1980, p. 96.
- 29 *Ibid.*, p. 100.
- 30 G. DELEUZE, « Post-scriptum sur les sociétés de contrôle », *L'autre journal*, no 1, 1990, p. 2-7.
- 31 Alain Damasio est un auteur de science-fiction français à succès, remarqué notamment par la singularité de son style et ses thèmes influencés par les analyses de Michel Foucault et de Gilles Deleuze. Dans son ouvrage le plus récent, Damasio imagine des paysages urbains entièrement privatisés et soumis à de grandes multinationales qui s'appuient sur des technologies de contrôle et de traçabilité sans commune mesure (matérialisées par des bagues ultraconnectées). Damasio revient à plusieurs reprises sur l'idée de technococon tressé par des technologies qui ne cessent pas d'assister les individus dans tous les aspects de leur monde vécu, y compris sur le plan des affects : « [Les gens acceptent] un monde qui prenne soin de nos esprits et de nos corps stressés, qui nous protège et nous choie, nous aide et corrige nos erreurs, qui nous filtre l'environnement et ses dangers. Un monde qui s'efforce d'aménager un technococon pour notre bien-être. L'intelligence ambiante pourvoit à ça. Elle nous écoute et elle nous répond. Elle courbe cette bulle autour de nos solitudes. Elle la tapisse d'objets et d'interfaces cools. Bien sûr, elle en profite pour nous espionner jusqu'au slip et pour nous manipuler jusqu'à la moelle! Mais au moins, elle s'occupe de nous, ce que plus personne ne fait vraiment... C'est un cercle vicieux. » A. DAMASIO, *Les Furtifs*, Clamart, La Volte, 2019, p. 277.
- 32 M. FOUCAULT, *Surveiller et punir*, op. cit.
- 33 F. JAMESON, *Le Postmodernisme ou la logique culturelle du capitalisme tardif*, trad. de l'anglais (États-Unis) par F. Novoltry, Paris, Beaux-Arts de Paris, 2018 [1991], p. 436.
- 34 « [L]a forme valeur et le rapport de valeur des produits du travail n'ont absolument rien à faire avec leur nature physique. C'est seulement un rapport social déterminé des hommes entre eux qui revêt ici pour eux la forme fantastique d'un rapport des choses entre elles. » K. MARX, *Le Capital : Livre I*, trad. de l'allemand par J. Roy, éd. de M. Rubel, Paris, Gallimard, coll. « Folio essais », 1968 [1867], p. 154. En d'autres termes, l'énigme de la marchandise repose sur le fait que sa valeur, au lieu de se présenter comme un rapport social, apparaît d'abord comme leur propriété naturelle.
- 35 J. BAUDRILLARD, *Pour une critique de l'économie politique du signe*, Paris, Gallimard, 1972, p. 97.
- 36 *Idem.*
- 37 *Ibid.*, p. 99.

- 38 *Idem.*
- 39 *Ibid.*, p. 100.
- 40 *Idem.*
- 41 K. MARX, *Le Capital*, *op. cit.*, p. 96.
- 42 Baudrillard avait bien montré que le simulacre n'a pas remplacé le réel, mais bien plutôt qu'il s'est évanoui, dissipé derrière le factice et l'artifice. À cet égard, Disneyland figure comme le résultat d'un monde totalement inversé, persuadé pourtant de ne pas l'être : « Disneyland est là pour cacher que c'est le pays "réel", toute l'Amérique "réelle", qui est Disneyland (un peu comme les prisons sont là pour cacher que c'est le social tout entier, dans son omniprésence banale, qui est carcérale). Disneyland est posé comme imaginaire afin de faire croire que le reste est réel, alors que tout Los Angeles et l'Amérique qui l'entoure ne sont déjà plus réels, mais de l'ordre de l'hyperréel et de la simulation. Il ne s'agit plus d'une représentation fausse de la réalité (l'idéologie), il s'agit de cacher que le réel n'est plus le réel, et donc de sauver le principe de réalité. L'imaginaire de Disneyland n'est ni vrai ni faux, c'est une machine de dissuasion mise en scène pour régénérer en contre-champ une fiction du réel. D'où la débilite de cet imaginaire, sa dégénérescence infantile. » J. BAUDRILLARD, *Simulacres et simulation*, Paris, Galilée, coll. « Débats », 1981, p. 25-26.
- 43 G. DEBORD, *La société du spectacle*, Paris, Gallimard, coll. « Folio », 1992 [1967].
- 44 J. BAUDRILLARD, *Pour une critique de l'économie politique du signe*, *op. cit.*
- 45 N. DYER-WITHEFORD, *Cyber-Proletariat: Global Labour in the Digital Vortex*, Londres, Pluto Press, 2015.
- 46 N. DYER-WITHEFORD et al., *Inhuman Power: Artificial Intelligence and the Future of Capitalism*, Londres, Pluto Press, 2019.
- 47 G. GENOSKO, *Critical Semiotics: Theory, from Information to Affect*, New York, Bloomsbury Academic, 2016, p. 105.
- 48 F. RICHERT, « Le Big Data à l'ère du sémiocapitalisme : décodage et axiomatisation généralisés du champ social », dans A. Mondoux & M. Ménard, *Big Data et société : Industrialisation des médiations symboliques*, Montréal, Presses de l'Université du Québec, 2018, p. 166-186.
- 49 Voir notamment le rapport : OIT, *Digital labour platforms and the future of work: Towards decent work in the online world*, Genève, 2018. En ligne : <https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_645337.pdf>.
- 50 A. CASILLI, *En attendant les robots. Enquête sur le travail du clic*, Paris, Seuil, 2019, p. 14.
- 51 D. CARDON & A. CASILLI, *Qu'est-ce que le Digital Labor?*, Bry-sur-Marne, INA, 2015.
- 52 F. JAMESON, *Le postmodernisme ou la logique culturelle du capitalisme tardif*, *op. cit.*, p. 54.
- 53 G. ANDERS, *L'obsolescence de l'homme. Sur l'âme à l'époque de la deuxième révolution industrielle*, *op. cit.*, p. 160-161.
- 54 G. DELEUZE, « Post-scriptum sur les sociétés de contrôle », *loc. cit.*
- 55 G. DELEUZE & F. GUATTARI, *Capitalisme et Schizophrénie, 2 : Mille Plateaux*, *op. cit.*, p. 573.
- 56 F. GUATTARI, *La Révolution moléculaire*, Paris, Les Prairies ordinaires, coll. « Essais », 2012 [1977], p. 412.
- 57 G. GENOSKO, *Critical Semiotics: Theory, from Information to Affect*, *op. cit.*
- 58 M. LAZZARATO, *Signs and Machines. Capitalism and the Production of Subjectivity*, Los Angeles, Semiotext(e), 2014.
- 59 *Ibid.*, p. 96-97. Je traduis : « That signs (machines, objects, diagrams, etc.) constitute the focal points of proto-enunciation and proto-subjectivity means that they suggest, enable, solicit, instigate, encourage, and prevent certain actions, thoughts, affects or promote others. »
- 60 F. GUATTARI, *La Révolution moléculaire*, *op. cit.*, p. 452.

- 61 F. GUATTARI, *Cartographies schizoanalytiques*, Paris, Galilée, 1989, p. 22.
- 62 G. LUKÁCS, *Histoire et conscience de classe. Essais de dialectique marxiste*, Paris, Minuit, 1976 [1923], p. 117.
- 63 Outre sa capacité présumée à stimuler la productivité et la consommation, la ludification est souvent présentée comme une solution de rechange pour lutter contre la *perte de sens* qui caractérise la vie moderne (notamment dans le travail). L'utilisation d'indicateurs chiffrés, d'objectifs quantifiés et de systèmes de gratification/récompense donne au sujet les moyens de conjurer le sentiment d'absurdité ressenti dans la réalisation de certaines tâches quotidiennes. Cette analyse semble également convenir aux systèmes intelligents programmés pour accompagner certaines activités quotidiennes. Dans les deux cas, l'égo du sujet est sauf : il a toujours l'impression d'être le seul maître à bord et de contrôler intégralement sa vie grâce à des techniques de management et des outils technologiques prévus pour lui.
- 64 F. LORDON, *Capitalisme, désir et servitude. Marx et Spinoza*, Paris, La Fabrique, 2010, p. 142.





Les aléas de la subjectivité sous le sémiocapitalisme : Félix Guattari et le potentiel politique des hybrides humains-machines

Gary GENOSKO

Cygne noir, no 7, 2019 : « Algorithmes »

Résumé

Félix Guattari a développé une approche sémiotique mettant en évidence l'idée suivant laquelle le capitalisme usine des subjectivités spécifiques aux différents âges de la vie. Ce texte poursuit cette réflexion portant sur les modes de production de la subjectivité à l'ère de l'informatique planétaire. Il rappelle dans un premier temps l'influence de l'approche de Guattari dans les travaux du philosophe Franco Berardi, lequel a forgé le concept de « sémiocapitalisme » pour désigner une forme de capitalisme opérant sémiotiquement. L'idée même de sémiocapitalisme invite à mener des réflexions sur les nouvelles formes d'exploitation cognitives liées à un travail immatériel qui tend à brouiller la frontière entre le travail et les loisirs et à faire de la subjectivité des travailleurs la matière première des « sémio-marchandises ». Ces réflexions donnent alors l'occasion d'interroger la subjectivité humaine d'un point de vue machinique, c'est-à-dire comme un processus fait de coupures et de connexions capables aussi bien de reproduire que de subvertir la spirale de l'auto-élaboration capitaliste. Ce texte discute aussi du fait que l'accentuation tendancielle de l'hybridation humain-machine n'apparaît pas forcément comme le moyen par lequel le sémiocapitalisme renforce ses capacités de saisie des modes de subjectivation, mais comme l'occasion inespérée de s'émanciper des modalités actuelles de l'assujettissement capitaliste et de développer une proto-altérité machinique.

Pour citer cet article

GENOSKO, Gary, « Les aléas de la subjectivité sous le sémiocapitalisme : Félix Guattari et le potentiel politique des hybrides humains-machines », trad. de l'anglais par S. Levesque & F. Richert, *Cygne noir*, no 7, 2019. En ligne : <http://www.revuecygne noir.org/numero/article/genosko-semiocapitalisme-guattari> (consulté le xx/xx/xxxx).



Cet article de *Revue Cygne noir* est mis à disposition selon les termes de la licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Pas de Modification 2.5 Canada.

LES ALÉAS DE LA SUBJECTIVITÉ SOUS LE SÉMIOCAPITALISME : FÉLIX GUATTARI ET LE POTENTIEL POLITIQUE DES HYBRIDES HUMAINS-MACHINES *

À la fin des années 1980, alors que l'informatisation s'accroît et qu'émerge la société connectée sous l'égide de l'infocapitalisme, une tendance liée à la production de la subjectivité a été identifiée de manière critique par Félix Guattari. Celui-ci a défini la subjectivité comme le produit croisé de la génération du savoir et de sa dissémination par le truchement des industries culturelles à l'intérieur d'un cadre plus large où la sensibilité et la sociabilité sont raffinées et renforcées, de connivence avec les mécanismes économiques en place, par des normes juridiques et des innovations technoscientifiques¹. La perspective sémiotique de Guattari a permis d'avancer l'idée que le capitalisme usinait des subjectivités spécifiques aux différents âges de la vie. À travers un processus de domestication opaque, le capitalisme chercherait tout particulièrement à conformer un certain nombre d'attributs singuliers infrapersonnels apparemment anodins aux modèles dominants de l'identité sociale. Les efforts qu'a fournis Guattari pour catégoriser et historiciser les types de subjectivité émergents à cette époque – sujets sériels, ou employés salariés ; sujets précaires, sans emploi ou peinant à trouver de l'emploi ; sujets élites, dirigeants d'entreprises ou appartenant à des familles fortunées – ont ouvert la voie à une interrogation plus fine en ce qui a trait aux modalités de la subjectivation à l'âge de l'informatique planétaire et aux effets de celle-ci sur la structuration des classes sociales. Guattari a bien vu qu'une forme de travail hautement cognitif et communicationnel allait être de plus en plus valorisée².

À partir des analyses de Guattari, Franco Berardi a proposé le terme « sémiocapitalisme » pour désigner une forme de capitalisme opérant sémiotiquement. Au-delà des efforts déployés par Guattari pour définir les subjectivités inhérentes au capitalisme mondial intégré (CMI)³, Berardi a poursuivi l'interrogation amorcée par celui-ci à l'effet que la valeur devait être repensée en dehors d'un principe général d'équivalence⁴. Berardi a développé sa théorie à l'époque de la bulle Internet, dans la deuxième moitié de la décennie 1990. Il a tâché de décrire un processus de génération de plus-value qui

* Ce texte est la traduction d'une conférence donnée par l'auteur sous le titre « The Vagaries of Capitalist Subjectivity » dans le cadre de la séance « Politics, Economy and Culture in the Society of Affect: Is There any Future for Global Capitalism? » du colloque *The Cultural Typhoon* tenu à l'Université Ryokoku de Kyoto, les 23 et 24 juin 2018. La présente traduction française a été réalisée par Simon Levesque et Fabien Richert.

serait disjoint à la fois de la main-d'œuvre et des lieux de production au sens industriel usité de ces termes. En rapprochant deux phénomènes cooccurents, Berardi a théorisé pour la première fois ce qu'il a appelé l'info-travail. Ce concept désigne : (1) la fragmentation du travail, désormais exécuté par une main-d'œuvre dominée par un phylum machinique globalisé, et (2) la capacité du réseau numérique de susciter une main-d'œuvre hautement qualifiée à n'importe quelle heure et depuis n'importe quel point nodal du réseau, et ce, en vue essentiellement de combiner et de recombinaer des fragments sémiotiques quelconques. Parce que les travailleurs du secteur informatique ont eu tendance à brouiller la frontière entre le travail et les loisirs, faisant souvent de leurs activités professionnelles une affaire personnelle, le cycle de production dans ce secteur a cessé de correspondre aux heures de travail usitées pour correspondre toujours davantage au moment exact de la demande. La mise en relation de la pulvérisation du travail fourni par la main-d'œuvre (répartie sur toute la planète), d'une part, et de l'atomisation du temps (en paquets liés à la demande), de l'autre, constitue le fondement même du sémiocapitalisme selon Berardi. Et c'est aussi ce qui détermine son caractère inhumain selon lui. Berardi a vu dans l'exploitation du travail cognitif une forme de subjectivation machinique liée à l'Internet et à la téléphonie cellulaire. En un certain sens, les téléphones portables auraient contribué à réaliser « le rêve du capital : absorber un maximum d'atomes de temps au moment exact où le cycle productif le demande. De cette façon, les [info-]travailleurs offrent leur journée entière au capital, mais ne sont payés que pour le moment durant lequel leur temps est cellularisé⁵ ».

À l'époque de la bulle Internet et de la subjectivité informatiquement machinée, l'auto-entrepreneuriat cognitif marque un recul : non seulement les auto-entrepreneurs exploitent-ils leurs propres capacités intellectuelles afin d'accroître la subjectivation, mais ce faisant ils accroissent aussi l'exploitation de leur propre force de travail. Maurizio Lazzarato a suggéré que l'auto-entrepreneuriat rendait l'individu à la fois maître et esclave, c'est-à-dire à la fois sujet d'énonciation (sujet produit par le discours) et sujet d'énoncé (première personne grammaticale)⁶. L'auto-organisation et l'auto-exploitation seraient ainsi fondamentalement intriquées, et cette intrication serait une caractéristique majeure du travail immatériel.

Les propositions de Berardi et de Lazzarato partagent une même hypothèse : « le sémio-capital est un capital-flux qui se coagule en artéfacts sémiotiques sans pour autant se matérialiser⁷. » Ici, la description berardienne de la production sémiocapitaliste et l'hypothèse de Lazzarato sur le travail immatériel forment une vision commune d'une subjectivité capitaliste fortement usinée. Selon cette vision, un relais humain-machine existe entre le réseau et les travailleurs autonomes, et ce relais forme un nœud dans un processus de communication où les capacités créatives infrapersonnelles sont exploi-

tées à la demande. L'hypothèse du travail immatériel, qui concerne à la fois le processus de production et le produit du travail, doit être comprise comme une tendance et non comme un absolu. D'après Berardi, la production sémiocapitaliste a comblé le fossé existant entre l'exécution et l'innovation : elle a différé le moment de la matérialisation en priorisant la production de simulations pouvant être actualisées matériellement dans le futur. Des signes, des syntaxes et des codes visuels sont manipulés et combinés virtuellement ; au nom de l'innovation constante, une exigence de flexibilité et de fluidité est imposée aux auto-entrepreneurs du numérique par l'entremise de formats réticulaires – logiciels, plateformes ou modèles informatiques – qui cadrent, capturent et commandent des fragments sémiotiques dans l'exploitation des productions virtuelles réparties dans le réseau. La main-d'œuvre, qui se trouve précarisée parce que pulvérisée, n'interagit plus que machiniquement avec un processus de production sémiotisé largement dématérialisé⁸.

Mais il demeure une forme de travail qui fait exception : la coagulation, qui désigne une médiation graduée de l'informe, une sémiotique d'actualisation centrifuge, tranquille et visqueuse, ayant la particularité de préserver une part d'inactualité dans le processus de production. De manière générale, on peut dire que les « sémioproduits » – les marchandises issues de la production sémiocapitaliste telle que définie par Berardi – coagulent. Évidemment, la coagulation est un terme connoté qui rappelle l'explication marxienne de la valeur-travail coagulée dans les marchandises. Or, la coagulation du travail comme substance de la valeur se réconcilie difficilement avec l'atomisation du temps de travail, qui s'avère en fait assez proche du non-travail. À l'ère des Mechanical Turks d'Amazon⁹, il est de moins en moins évident que le temps de travail se traduise en unités temporelles objectives. La différence entre la valeur et le temps de production devient intangible : « le travail crée de la valeur par des voies qui sont de plus en plus difficilement rattachables à des quantas temporels », bien que dans certaines marchandises le temps de travail coagulé apparaisse encore comme un « paramètre mineur », écrivent Adam Arvidsson et Elanor Colleoni¹⁰.

Marx a brillamment théorisé les rapports de continuité entre, d'une part, la sublimation – la volatilisation de la solidité bourgeoise comme destin de classe, d'après le projet communiste – et, de l'autre, la coagulation – la quantité de temps de travail coagulé dans une marchandise considérée comme une valeur d'échange. Cependant, les théories récentes sur le sémiocapitalisme font apparaître plus clairement à quel point le temps de travail est devenu problématique : le temps de travail moyen nécessaire est-il connaissable ? La valeur peut-elle être mesurée ? Le travail humain est-il seulement cohérent ? Après tout, le sémiocapitalisme veut contraindre la subjectivité à s'énoncer de plus en plus machiniquement, subsumant toute activité cognitive sous l'égide d'une

désignation unique de la valeur ramenant toutes les formes de valorisation à la plus-value. En tant qu'opérateur sémiotique, le capitalisme cherche à canaliser l'implosion de la valeur pour en faire le matériau brut de la subjectivité : ce faisant, il empoisonne ainsi l'invention machinique, il opacifie les passages transversaux et fait régresser la polyvalence énonciative en une norme sémiologique unique. Ainsi désormais, défend Lazzarato, un produit immatériel se réduit à « un processus immédiat de valorisation » lorsque considéré en tant que simple tâche cognitive, à l'exemple de cliquer, naviguer ou fournir des termes de recherche dans le but d'analyser et de classer des valeurs¹¹.

Quelles sont les conséquences de penser quelque chose d'aussi intime, intérieur, central et centralisant que la subjectivité humaine (l'individualité, la personne) dans un vocabulaire d'usines et de machines? Quelle place reste-t-il pour la subjectivité lorsque celle-ci ne se présente plus comme le noyau dur donnant sens à la conscience, mais comme une simple pièce mécanique prise dans un grand assemblage de composantes? Dans *L'Anti-Œdipe*, Gilles Deleuze et Félix Guattari caractérisent la subjectivité comme un « résidu », un « appendice », ou encore une « pièce dans la machine ». La machine déplace la subjectivité depuis le centre vers la périphérie, elle crée « un étrange sujet, sans identité fixe [...] naissant des états qu'il consomme¹² ». Un sujet résulte d'une sorte de coupure, il est un résidu. Comme l'écrivent Deleuze et Guattari, « [l]a troisième coupure de la machine désirante est la coupure-reste ou résidu, qui produit un sujet à côté de la machine, pièce adjacente à la machine¹³ ». Un tel sujet se trouve dans une position de contiguïté, il est une pièce parmi les pièces, et une pièce elle-même composée d'autres pièces, sans identité personnelle spécifique.

La subjectivité ainsi conçue est un processus machinique fait de coupures et de connexions ; elle n'est pas un point, elle n'est pas intérieure, elle ne correspond ni à l'esprit ni à une personne. La subjectivité est extérieure, projetée par centrifugation, hors du centre consacré de l'individualité présumée souveraine et de l'ego. Elle existe au-delà des actes d'un ego transcendantal constitué, doué d'autorité et pouvant prétendre à la vérité ; elle est étrangère à la condition d'être pensant sanctifié par la philosophie cartésienne. À la place, des sensations existent là où la pensée s'imaginait régner en maître, et avec elles s'installent l'incertitude, l'imperfection et les extensions actives et passives de la subjectivation.

S'écarter des philosophies concevant un sujet intérieur comme essence unique ou comme état de conscience unificateur implique une forme de sous-traitance, une externalisation des processus de subjectivation en un champ plus vaste, économique et mondialisé. Étant entendu que les subjectivations ne peuvent être séparées des relations économiques auxquelles elles sont indexées, certaines fonctions spécifiques de ces relations sont inévitablement et littéralement sujettes à la sous-traitance externalisée :

des pièces de la machine seront produites *ailleurs*. La subjectivité sous-traitée, résultant de prises de décision automatisées, générées algorithmiquement, est le degré extrême et final de la subjectivité capitaliste à l'ère de l'intelligence artificielle.

Les variétés de la subjectivité se rattachant au modèle sémiocapitaliste esquissées jusqu'ici n'épuisent pas les processus de subjectivation résultant d'une poussée hors de l'identité normalement contrainte par des marqueurs sociologiques et philosophiques définis. Ultimement, si l'on en croit Guattari, Berardi et Lazzarato, la subjectivité n'est pas condamnée à reproduire la spirale de l'auto-élaboration capitaliste. Alors que l'interface entre les humains et les machines continue de se complexifier, les métissages humains-machines pourraient constituer une force émancipatoire de désidentification insufflant la possibilité d'une énonciation a-subjective – qui ne serait pas un fétichisme, mais une nouvelle force d'expression protosubjective, une nouvelle proto-altérité machinique. L'émergence d'hybrides humains-machines nés du métissage du « travail vivant » (le travail des humains) et du « travail mort » (le travail des machines), qui conjoindraient uniment la création et la reproduction, représente une occasion de s'émanciper des modalités actuelles de l'assujettissement capitaliste déterminant des rôles et des fonctions aliénantes (participant d'une formation identitaire aliénante) pour les travailleurs¹⁴.

Berardi a décrit le processus de déssubjectivation inhérent à l'inhabilité des info-travailleurs à s'organiser à travers les zones de précarité. Dans le réseau numérique, écrit-il, les travailleurs sont « dépourvus de toute consistance individuelle¹⁵ » et, comme tels, ils ont du mal à continuer d'exister : « on ne trouve plus de travailleurs individués, mais seulement des atomes fractals de temps abstraits et dépersonnalisés mis à disposition dans la Net-sphère¹⁶. » En tant que résidus du processus de valorisation capitaliste fonctionnellement intégrés dans le réseau, les travailleurs sont inversement recomposables en vue d'atteindre une interopérabilité maximale. On peut penser que de telles manipulations de simulacres mèneront à l'épuisement, mais l'épuisement annonce aussi la possibilité d'un retrait, d'un ralentissement et d'une passivité radicale orientée en vue de l'éclatement de la bulle du travail¹⁷. Le sujet énergique et héroïquement résistant de la mythologie activiste n'est plus, et avec lui l'identité du travailleur industriel moderne a disparu. Le vocabulaire de la coagulation est maintenant remplacé par une coalescence compensatoire non-organisée, qui forme ce que l'on pourrait appeler *un communisme transversal fondé dans le retrait*. Cette nouvelle formation politique s'oppose par exemple à la volonté active de créer des coalitions motivées par un but bien défini, à l'exemple des différents groupes de protestation anti-Trump aujourd'hui, qui tentent de s'organiser tant bien que mal. Dans la théorie berardienne, les vestiges des pratiques autonomistes du refus et du retrait, qui s'appuient sur des décennies d'expérimentations

médiatiques, avec les radios pirates et les webtélés tactiques, peuvent être reconfigurés au regard du renversement introduit par Lazzarato : ce qui ressemble à une dépossession – de l'individu, du travail et du lien vivant – est une occasion de surpasser les formations identitaires asservissantes.

Il va sans dire que, suivant cette pensée, ni les individus ni les classes ne sont plus *sujets* de l'histoire puisqu'un agent de transformation a été introduit, qui prend la forme d'humains désobjectivisés et de machines protosubjectives. On ne devrait pas déplorer la perte d'un sujet ou d'un collectif prédéterminé, mais porter une attention plus soutenue aux forces libérées par le déplacement de l'anthropocentrisme et au passage vers une autonomisation de l'énonciation machinique. Les perspectives de formation de nouveaux modes de subjectivation hautement hétérogènes restent ouvertes aux possibilités les plus progressives que présente le vaste phylum des machines. Là où, par exemple, une *correspondance formelle* était présupposée entre le réseau distribué de production socioéconomique et l'organisation de formes résistantes dans l'articulation de la multitude, il faut désormais envisager une séparation et un déplacement des luttes et de la subjectivation capitaliste coévolutive à la structure même du réseau¹⁸. Ce n'est qu'en brisant ce miroir présumé entre, d'une part, les processus de prise de décisions politiques à travers lesquels la multitude acquiert son existence et se gouverne et, d'autre part, les modèles de la dynamique économique du réseau que pourront véritablement apparaître de nouveaux modèles de subjectivité postcapitaliste. Cela dit, il est évident que la rémanence résiduelle et la contiguïté du sujet ne garantissent pas pleinement son échappée de la contrainte imposée par les systèmes d'identification capitalistes dominants, non plus qu'elles n'assurent son potentiel d'exploration transversale dans le métissage et l'interfaçage possiblement limité avec un phylum machinique pluraliste¹⁹.

En dernière analyse, quatre caractéristiques générales peuvent être énoncées en ce qui a trait au potentiel politique des hybrides humains-machines dans la reconfiguration possible de la subjectivité à l'ère sémiocapitaliste :

(1) *La subjectivation est un processus à travers lequel des identités contingentes sont continuellement formées.* Le hasard et la non-fixité rendent difficiles la normalisation et l'asservissement ; même la sous-traitance ne suffit pas au recentrement permanent de la subjectivité, car elle est continuellement déplacée par ses engagements machiniques.

(2) *La subjectivation est un processus de déplacement.* Mais ce déplacement n'est pas un retour du refoulé puisqu'il élabore son propre diagramme en continu. Un diagramme est le croquis projectif d'un chemin déroulant parmi les composantes participant au processus de subjectivation. Il peut s'avérer aussi imprévisible que l'imagination, mais pour cette même raison les cohérences recherchées à travers lui, en tant que bornes intermédiaires dans le processus de subjectivation, peuvent échapper à l'univers limité

et hautement protégé des figures normatives de la subjectivité capitaliste (célébrités, chefs, guerriers, héros). Le déplacement implique l'engendrement de formes de singularité non seulement incompatibles avec le système, mais aussi incomparables entre elles.

(3) *La subjectivation de type pathique et non discursive esquivé les surcodages dominants de la subjectivité capitaliste.* La contiguïté du sujet machinique passe de l'effet à l'affect. En s'hybridant avec les machines, la subjectivation transfère et partage sa contiguïté dans le réseau. Ainsi, les points de référence extrinsèques généraux, comme l'espace, le temps, l'énergie et le profit peuvent à leur tour être déplacés au bénéfice des coordonnées intensives non linéaires de l'appréhension pathique ou encore d'autres univers de valeur.

(4) *La contiguïté ne suffira pas à nous sauver.* Inspiré par Pierre Lévy, Guattari a considéré les machines à travers le concept d'interface. Son objectif était d'« essayer d'abattre le rideau de fer ontologique entre l'être et les choses²⁰ ». Ainsi a-t-il proposé une réanimation des machines. En effet, Guattari a défendu une conception proto-subjective et « animiste » des machines, ouverte à l'assemblage avec d'autres composantes d'un environnement machinique plus vaste. Ce pluralisme ontologique des machines ne se réduisant pas à la contiguïté que présentent les objets partiels dans leur relation au corps humain, et Guattari ne se satisfaisait pas des assemblages allopoïétiques de composantes non vivantes, il a voulu réintégrer dans son modèle conceptuel machinique des caractéristiques autopoïétiques (propres aux relations entre des êtres biologiques). Ce faisant, il a souligné l'importance du préfixe « proto » dans « proto-subjectivité », imaginant les composantes proto-subjectives et proto-biologiques d'une forme de vie machinique auto-affirmative et auto-créative, qui aurait une appréhension pathique immédiate du monde de l'expérience.

En somme, Guattari n'a pas adopté le vocabulaire de l'interface pour réaffirmer la primauté de la structure. Il a au contraire rejeté la conception structuraliste du sujet défini comme flottant entre deux signifiants (linguistiques). Plus pragmatique que la plupart de ses contemporains, mais aussi plus audacieux conceptuellement, Guattari pensait que la subjectivation machinique pourrait s'auto-engendrer créativement au moyen de connexions hypertextuelles²¹. Sur cette base théorique, l'hybridation humain-machine qu'il a imaginée avait essentiellement pour visée de refonder la subjectivation hors des contraintes et des habitudes sociales imposées par le capitalisme.

Bibliographie

- ARVIDSSON, Adam & Elanor COLLEONI, « Value in Informational Capitalism and on the Internet », *The Information Society*, vol. 28, no 3, 2012, p. 135-150.
- BERARDI, Franco, « Schizo-Economy », trad. de l'italien par M. Goddard, *Substance*, vol. 36, no 1, 2006, p. 75-85.
- , *Félix Guattari: Thought, Friendship, and Visionary Cartography*, trad. de l'italien par G. Mecchia & C. J. Stivale, Londres, Palgrave MacMillan, 2008.
- , *The Soul at Work: From Alienation to Autonomy*, trad. de l'italien par F. Cadel & G. Mecchia, préface de J. Smith, Los Angeles, Semiotext(e), 2009.
- , *After the Future*, éd. établie par G. Genosko & N. Thoburn, trad. de l'italien par A. Bove et al., Oakland, AK Press, 2011.
- DELEUZE, Gilles & Félix GUATTARI, *L'Anti-Œdipe*, Paris, Minuit, 1972.
- GUATTARI, Félix, *Les trois écologies*, Paris, Galilée, coll. « L'espace critique », 1989.
- , « À propos des machines », *Qu'est-ce que l'écologie?*, textes présentés par S. Nadaud, Paris, Lignes & IMEC, 2012, p. 111-129.
- , « Les systèmes d'interface machine », *Qu'est-ce que l'écologie?*, textes présentés par S. Nadaud, Paris, Lignes & IMEC, 2012, p. 285-287.
- LAZZARATO, Maurizio, « Immaterial Labour », trad. du français par P. Colilli & E. Emory, dans P. Virno & M. Hardt (dir.), *Radical Thought in Italy: A potential Politics*, éd. anglaise établie par S. Buckley, M. Hardt & B. Massumi, Minneapolis, University of Minneapolis Press, 1996, p. 133-147.
- , « La Machine », *Transversal texts*, no 10, 2006, non paginé. En ligne : <<https://transversal.at/transversal/1106/lazzarato/fr>>.
- , « Strategies of the Political Entrepreneur », trad. par T. S. Murphy, *Substance*, vol. 36, no 1, 2007, p. 87-97.
- , *Signs and Machine: Capitalism and the Production of Subjectivity*, trad. du français par J. D. Jordan, Los Angeles, Semiotext(e), 2014.
- NEGRI, Antonio & Michael HARDT, *Multitude*, New York, Penguin, 2004.

Notes

- 1 F. GUATTARI, *Les trois écologies*, Paris, Galilée, coll. « L'espace critique », 1989, p. 40-42.
- 2 En particulier, Guattari envisageait le renforcement d'une subjectivité de classe moyenne capable de remplacer la sérialisation de la classe ouvrière qui avait eu cours jusqu'alors : « On tentera d'embourgeoiser au maximum la subjectivité ouvrière. » *Ibid.*, p. 63.

- 3 Voir *ibid.*, p. 40-42.
- 4 « De nouvelles bourses de valeur, de nouvelles délibérations collectives donnant leur chance aux entreprises plus individuelles, les plus singulières [...] » *Ibid.*, p. 67.
- 5 F. BERARDI, *The Soul at Work: From Alienation to Autonomy*, trad. de l'italien par F. Cadel & G. Mecchia, préface de J. Smith, Los Angeles, Semiotext(e), 2009, p. 90.
- 6 « La transformation du salarié en "capital humain", en entrepreneur de soi-même, telle qu'elle est mise en place par les techniques de management contemporaines, est l'accomplissement du processus de subjectivation et du processus d'exploitation, puisque, ici, c'est le même individu qui se dédouble. D'une part, il porte la subjectivation au paroxysme, puisqu'il implique dans toutes ses activités les ressources "immatérielles" et "cognitives" de son "soi", et d'autre part, il porte à identification subjectivation et exploitation, puisqu'il est à la fois patron de lui-même et esclave de lui-même, capitaliste et prolétaire, sujet d'énonciation et sujet d'énoncé. » M. LAZZARATO, « La Machine », *Transversal texts*, no 10, 2006, non paginé. En ligne : <<https://transversal.at/transversal/1106/lazzarato/fr>>. Voir aussi : M. LAZZARATO, « Strategies of the Political Entrepreneur », trad. par T. S. Murphy, *Substance*, vol. 36, no 1, 2007, p. 87-97.
- 7 F. BERARDI, « Schizo-Economy », trad. de l'italien par M. Goddard, *Substance*, vol. 36, no 1, 2006, p. 76.
- 8 F. BERARDI, *Félix Guattari: Thought, Friendship, and Visionary Cartography*, trad. de l'italien par G. Mecchia & C. J. Stivale, Londres, Palgrave MacMillan, 2008, p. 20.
- 9 Un « mechanical turk » est un info-travailleur qui, pour le compte d'Amazon, effectue à distance des tâches cognitives relativement simples (mais qu'un humain fait encore mieux qu'une machine) en échange d'une (faible) rémunération. Voir en ligne : <<https://www.mturk.com/worker>>.
- 10 A. ARVIDSSON & E. COLLEONI, « Value in Informational Capitalism and on the Internet », *The Information Society*, vol. 28, no 3, 2012, p. 140.
- 11 Voir M. LAZZARATO, « Immaterial Labour », trad. du français par P. Colilli & E. Emory, dans P. Virno & M. Hardt (dir.), *Radical Thought in Italy: A potential Politics*, éd. anglaise établie par S. Buckley, M. Hardt & B. Massumi, Minneapolis, University of Minneapolis Press, 1996, p. 144.
- 12 G. DELEUZE & F. GUATTARI, *L'Anti-Œdipe*, Paris, Minuit, 1972, p. 23.
- 13 *Ibid.*, p. 48.
- 14 M. LAZZARATO, *Signs and Machine: Capitalism and the Production of Subjectivity*, trad. du français par J. D. Jordan, Los Angeles, Semiotext(e), 2014, p. 120-121.
- 15 F. BERARDI, *After the Future*, éd. établie par G. Genosko & N. Thoburn, trad. de l'italien par A. Bove et al., Oakland, AK Press, 2011, p. 129.
- 16 *Ibid.*, p. 129.
- 17 *Ibid.*, p. 138.
- 18 A. NEGRI & M. HARDT, *Multitude*, New York, Penguin, 2004, p. 82-88.
- 19 Dans une nouvelle ère machinique marquée par le « pluralisme ontologique », Guattari entrevoit la possibilité d'une « réinvention permanente de la condition humaine » au moyen de nouvelles interfaces machiniques. Voir F. GUATTARI, « Les systèmes d'interface machinique », *Qu'est-ce que l'écophilosophie?*, textes présentés par S. Nadaud, Paris, Lignes & IMEC, 2012, p. 285-287.
- 20 F. GUATTARI, « À propos des machines », *Qu'est-ce que l'écophilosophie?*, op. cit., p. 112.
- 21 « La position autopoïétique et "hypertextuelle" de la machine possède une potentialité pragmatique, elle permet d'avoir une attitude créationniste, de composition machinique, face à ce rideau de fer ontologique qui sépare le sujet, d'un côté, et les choses, de l'autre. » *Ibid.*, p. 129.

