



## 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'icongité 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 sémiopragmaticiste 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 sémiose 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 sociales<sup>1</sup>. 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-même<sup>2</sup> » puisque la base de données massives donne à voir des corrélations permettant de tirer des conclusions. Cette logique quantitative<sup>3</sup> 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 »<sup>4</sup>. 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-mêmes<sup>5</sup>. » 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 scientifiques<sup>6</sup> ». 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 peut<sup>7</sup> », 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 réel<sup>8</sup>. 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 marketing<sup>9</sup>. « 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 marché<sup>10</sup> » : les avancées récentes en science des données ont aussi révolutionné le mode de fonctionnement des entreprises<sup>11</sup>. 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'utilisation<sup>12</sup>.

Il est utile de souligner que les conséquences éthiques de tels développements ont été vivement critiquées, non seulement dans les cercles universitaires et scientifiques<sup>13</sup>,

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 britannique-américaine *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 débat<sup>14</sup>.

## 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 problématise 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)<sup>15</sup>. 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'expérience<sup>16</sup>? ».

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 humains<sup>17</sup> ». 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 visualisables<sup>18</sup>. 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 chercher<sup>19</sup> », 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 tendance<sup>20</sup>. 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 qualitative<sup>21</sup>. 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 sociaux<sup>22</sup>. Ceci constitue un « fossé numérique »<sup>23</sup> et produit des angles morts dans les bases de données (« *data shadows* »)<sup>24</sup> qui nuisent à la représentativité des bases de données<sup>25</sup>.

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 statistique<sup>26</sup>. 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 généralisée<sup>27</sup>. 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 jour<sup>28</sup>.

### 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 données<sup>29</sup>. 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 article<sup>30</sup>.

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 religieuse<sup>31</sup>. Elle fait figure de charnière dans le domaine de ce que Thomas Sebeok appelle la « science intégrée de la communication »<sup>32</sup>, 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 » sémiotiques<sup>33</sup>, 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 »<sup>34</sup>. 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 dynamique<sup>35</sup>. 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, légisigne<sup>36</sup>, l'interprétant – rhème, proposition, argument (EP 2, p. 172-173)<sup>37</sup> et la relation entre l'objet et le representamen – icône, index, symbole (MS [R] 491, p. 1-2)<sup>38</sup>. 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 correspondance<sup>39</sup>.

De plus, la sémiotique peircienne implique qu'un signe n'existe qu'*in actu* (CP 5.569)<sup>40</sup>. 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 processuelle<sup>41</sup> 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 modernes<sup>42</sup>.

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*)<sup>43</sup>. Le niveau processuel se caractérise par un flux continu et dynamique d'évènements 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 Guattari<sup>44</sup>, 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*<sup>45</sup>. 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 aspects<sup>46</sup>. 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 internet<sup>47</sup>.

Dans le cas où le présupposé de similarité iconique est erroné, alors le processus peut être qualifié de dégénération du signe (CP 5.73)<sup>48</sup>. 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 numériques<sup>49</sup>. 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 culturelle<sup>50</sup>. En effet, les signes iconiques deviennent signifiants dans le cadre de références culturelles partagées qui informent « les normes scripturales [et autres]<sup>51</sup> ». 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 »<sup>52</sup> 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 « forées » 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 »<sup>53</sup> 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.<sup>54</sup>. 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 généralité<sup>55</sup> ». 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 subjectivité<sup>56</sup>. » 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 matériel<sup>57</sup>. Ces métaphores de matière brute encouragent à se représenter les données comme de la matière naturelle non raffinée<sup>58</sup>.

Cependant, si les imposantes bases de données nous arrivent toujours déjà « cuites » (*baked*)<sup>59</sup>, 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 rareté<sup>60</sup>, 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 transparait dans le terme anglais *data*, lequel a commencé à être utilisé pour parler des données numériques seulement au milieu du XX<sup>e</sup> siècle<sup>61</sup>.

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 symboles<sup>62</sup>. 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 » spécifique<sup>63</sup>, 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 réel<sup>64</sup>. Cet article a porté un regard critique sur la manière dont la « révolution des données »<sup>65</sup> 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 liés 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-machine<sup>66</sup>. Ce devenir signifie non seulement une dépendance grandissante aux technologies numériques pour l'organisation de nos vies, mais aussi un certain « esclavage » ontologique<sup>67</sup>. Certain-e-s chercheur-e-s défendent aussi l'idée selon laquelle la gouvernance algorithmique façonne un type de sujet qui lui est spécifique<sup>68</sup>. Enfin, dans ce cadre d'analyse critique, des études récentes et des travaux de spéculation théorique examinent la fabrique biopolitique de la subjectivité posthumaine<sup>69</sup>.

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 adaptions 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 posthumanistes<sup>70</sup>. 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](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 socio-numériques*, 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 10<sup>th</sup> 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](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://pflurance.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 BERNS « 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. TRÉPANIÉRIE-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](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](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 10<sup>th</sup> 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://pfluenceance.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 "Eureka!" 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 HRIZ2013 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.

