



Le rôle des humanités numériques dans le nouvel espace politique

Pierre Lévy

Publié le 21-01-2019

<http://sens-public.org/article1369.html>



Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA
4.0)

Résumé

Alors que plus de 50% de la population mondiale est connectée à l'Internet, les grandes plateformes, et particulièrement Facebook, ont acquis un énorme pouvoir politique. Cette nouvelle situation nous oblige à repenser le projet d'émancipation des lumières. Je propose dans cet article que les chercheurs en sciences humaines et sociales relèvent ce défi en adoptant et en diffusant de nouvelles normes d'intelligence collective réflexive. Les communs de la connaissance, la science ouverte et la souveraineté des individus sur les données qu'ils produisent font l'unanimité. Mais ces principes incontournables sont encore insuffisants. La puissance de calcul et de communication disponible, combinée à l'utilisation d'IEML (une langue à la sémantique calculable), nous permettent d'envisager une mise en transparence des opérations de création de connaissance, de sens et d'autorité. Je présente ici les grandes orientations stratégiques permettant d'atteindre ces objectifs. Une révolution épistémologique des sciences humaines est à portée de main, et avec elle une nouvelle étape dans l'évolution de la pensée critique.

Abstract

While more than 50% of the world's population is connected to the Internet, the major platforms, and particularly Facebook, have acquired an enormous political power. This new situation forces us to rethink the emancipation project of the enlightenment. In this article, I propose that researchers in the humanities and social sciences take up this challenge by adopting and disseminating new standards of reflexive collective intelligence. Knowledge commons, open science and the sovereignty of individuals over the data they produce are unanimously accepted. But these essential principles are not enough. The available computing and communication power, combined with the use of IEML (a language with computable semantics), enables us to contemplate a transparency of the knowledge, meaning and authority creation processes. I present here the main strategic orientations for reaching these objectives. An epistemological revolution in the human sciences is within reach, and with it a new stage in the evolution of critical thinking.

Mot-clés : Facebook, IEML, épistémologie, humanités numériques, intelligence collective, transparence, pragmatique, jeux de langage, sémantique, linguistique, révolution scientifique, science ouverte, communs de la

connaissance, Lumières, autorité, vérité

Keywords: Facebook, IEML, epistemology, digital humanities, collective intelligence, transparency, pragmatics, language games, semantics, linguistics, scientific revolution, open science, knowledge commons, Lumières, authority, truth

Table des matières

Introduction	5
On a raison de critiquer Facebook	7
La mutation anthropologique	9
La mission historique des plateformes	12
Au-dessus des états	14
La guerre secrète	15
Les lieux communs de l'émancipation	17
Un projet épistémologico-politique	23
IEML : un système de coordonnées sémantiques	27
Vers une explicitation des jeux de communication et de référence	32
L'acte de référence	34
Les niveaux de référence et la question de la vérité	38
Observer l'émergence de l'autorité et de la croyance	41
Conclusion : la mission des humanités numériques	47
Bibliographie	51

Le rôle des humanités numériques dans le nouvel espace politique

Pierre Lévy

Introduction

L'heure est au désenchantement chez les premiers enthousiastes de la communication numérique. Les données privées des utilisateurs sont exploitées sans scrupules par les grandes plateformes. De toutes parts, la propagande politique utilise les techniques de pointe du big data et de l'intelligence artificielle pour manipuler le nouvel espace politique. Les discours de haine sèment les fleurs du mal sur les médias sociaux. Notre démesure aurait-elle attiré sur nos têtes la colère de quelque justice immanente ? Le premier quart de ce texte veut faire la part des choses. À partir de l'exemple de Facebook, je reconnais d'abord les aspects sombres de la situation contemporaine, mais je tente aussi d'équilibrer le tableau. Même s'ils ne sont pas toujours mis en valeur dans le discours public, il existe de nombreux emplois émancipateurs de la communication numérique, en particulier ses usages pédagogiques et savants. D'autre part, puisque l'humanité est connectée à cinquante-cinq pour cent (entre quatre-vingts et quatre-vingt-dix pour cent en Europe et en Amérique du Nord), il n'est pas étonnant que l'Internet contienne désormais, avec les individus et les sociétés qui les portent, toutes les facettes de la nature humaine, y compris les moins flatteuses. Tout le monde ou presque se trouvant plongé dans le nouvel espace de communication, les conflits géopolitiques, les oppositions sociales et les guerres culturelles s'y manifestent inévitablement. Au-delà du constat, j'ai voulu répondre à la question qui hante bien des gens de bonne volonté : comment renouveler et faire progresser l'humanisme et les idéaux d'émancipation des Lumières sur la planète interconnectée du

XXI^e siècle ? La majeure partie de ce texte est consacrée à l’exploration de ce problème et à l’énoncé de quelques solutions, hypothétiques et provisoires, et qui visent avant tout à provoquer le débat. Comme on le verra, je n’ai pas abandonné la perspective de l’intelligence collective¹, mais j’insiste plus aujourd’hui sur sa dimension réflexive et critique. Ma méditation porte non seulement sur le projet – faire croître et perfectionner un pôle des usages savants de l’Internet qui fasse contrepoids à ses emplois aliénants – mais aussi sur la communauté porteuse d’un tel projet : les chercheurs dans le domaine des humanités numériques. Quand les corpus deviennent des *datasets*, quand les méthodes de recherche mettent en œuvre des algorithmes, quand les réseaux de recherche s’organisent (aussi) en ligne et quand d’ailleurs la majorité de la population se sert d’ordinateurs et de smartphones, il est sans doute pléonastique de faire suivre « les humanités » de l’adjectif « numérique ». Dit-on d’Érasme qu’il fut un humaniste imprimé ? Tout comme l’imprimerie a profondément transformé les pratiques savantes, jusqu’à faire naître une communauté scientifique indépendante de l’Église et à supporter le mouvement des Lumières, le monde de la recherche en sciences humaines est gros aujourd’hui d’une profonde transformation, qui pourrait affecter non seulement sa propre identité, mais aussi son rôle politique et culturel. C’est vers ce potentiel que je pointe en utilisant l’expression consacrée d’humanités numériques.

Ce texte porte la marque des passionnantes discussions que j’ai eues avec Louis van Beurden, ingénieur et chercheur en traitement automatique des langues à l’Université de Montréal, qui travaille avec moi à l’implantation du langage IEML. Il prolonge également les dialogues que j’ai eu le bonheur d’entretenir avec Marcello-Vitali Rosati, titulaire de la Chaire de Recherche du Canada en écriture numérique à l’Université de Montréal. Il a enfin bénéficié des échanges stimulants que j’ai eus dans le cadre du réseau de recherche Revues 2.0 et des activités du Centre de Recherche interuniversitaire sur les Humanités Numériques, dirigé par Michael Sinatra.

1. Voir *L’intelligence collective* (Lévy 1994) et plus récemment les articles de blogue *L’intelligence collective, en quelques mots...* (Lévy 2016d) et “Cultiver l’intelligence collective” (Lévy 2016b).

On a raison de critiquer Facebook

Mark Zuckerberg est aujourd’hui l’une des personnes les plus haïes par les élites intellectuelles européennes et américaines. Déjà en 2016, Bernard Stiegler mettait en relation les GAFAM avec l’État islamique, le second n’étant qu’un symptôme de la vraie maladie sociale – la disruption – causée par les premiers (Stiegler 2016). Au chapitre du réquisitoire contre la Silicon Valley en général et Facebook en particulier, il faudrait citer ici de nombreux auteurs dont, dans des genres différents, Eric Sadin (2016) ou Olivier Ertzscheid et Antonio Casilli (2017). Plus récemment, Gérard Wormser (2018) consacre un ouvrage fort convaincant à la dénonciation de Facebook au nom de l’humanisme démocratique et des idéaux des Lumières. Côté américain, la liste est fort longue : rappelons notamment que, dès 2013, la satire d’Evgeni Morozov avait mené la charge (2013). Mais le tollé contre Facebook aux États-Unis atteint une intensité proche de la furie depuis les révélations sur l’usage de données confidentielles par Cambridge Analytica afin de personnaliser les messages de la campagne de Donald Trump sur Facebook. Ajoutons à cela l’action des « usines à trolls » du Kremlin contre la candidature Clinton sur les médias sociaux et voici que Facebook est accusé d’être responsable de la victoire électorale du millionnaire populiste. Que Mark Zuckerberg soit connu pour pencher du côté démocrate et qu’il ait récemment pris la tête d’une campagne pro-immigration² ne change rien à l’affaire. Le récent portrait de Zuckerberg dans le *New Yorker*, haut lieu de l’intelligentsia américaine de la côte Est, ne dissimule pas sa nature : un réquisitoire sans pitié³.

Mais que l’on y prenne garde. Accuser Facebook de l’élection de Donald Trump et du Brexit n’est-ce pas comme si l’on expliquait la montée du fascisme en Europe dans la première moitié du XX^e siècle par la radio, la victoire de Kennedy contre Nixon en 1960 par la télévision, la révolution Khomeiniste dans l’Iran de 1979 par les magnétophones à cassettes, l’élection de Barack Obama en 2012 par le big data⁴ et celle de Jair Bolsonaro au Brésil en 2018 par Whatsapp ? Dans chacun de ces cas, les médias ont certainement joué un rôle en distordant le champ de force politique antérieur et en faisant le jeu de certains acteurs. Attention cependant de ne pas éliminer trop vite les situations politiques, les gens, les émotions et les idées au profit d’une cause

2. Voir en ligne.

3. Voir en ligne.

4. Voir en ligne.

unique. Il semble que, pour certains commentateurs, le « déterminisme technologique » doive être dénoncé chez leurs adversaires, mais qu'ils puissent s'en prévaloir – et mettre les plateformes en position de bouc émissaire – lorsqu'il devient trop douloureux de regarder en face certaines réalités socio-politiques. N'oublions pas que, depuis au moins 2015, à chaque élection, ce sont tous les partis politiques qui montent d'impressionnantes structures de commande stratégique de la communication visant l'influence dans les médias sociaux et que leur action, guidée par les données, utilise invariablement des techniques de microciblage des électeurs.

Au-delà du scandale « Cambridge Analytica » et du rôle des plateformes dans les élections récentes, on peut se demander avec Gérard Wormser comment l'idéal des Lumières et la grande tradition de l'humanisme européen pourraient se reconduire dans des médias sociaux soumis à un impératif implacable de monétisation des données et qui vivent au rythme des rapports trimestriels aux investisseurs. Parce que « l'engagement » est la clé de sa survie économique, la plateforme sert les intérêts privés de ses utilisateurs et leurs satisfactions immédiates. C'est ainsi que le nouvel espace public numérique renforce les réflexes narcissiques d'un individu roi, dépendant aux likes, obsédé par son décompte d'« amis » ou de followers, se mirant dans son petit réseau de pairs et de relais d'opinion, enfermé dans les bulles cognitives d'une machine sociotechnique qui ne lui renvoie à lire, à écouter ou à visionner que ce qu'il aime⁵. Reste-il possible, dans ces conditions, de formuler un quelconque projet historique d'émancipation, capable de dépasser les limites des microcommunautés amicales, familiales, locales ou économiques favorisées par Facebook ? La tête baissée sur leur smartphone, les membres de ces communautés font tourner en boucle les stéréotypes qui renforcent leurs identités éclatées et leurs mémoires courtes sous le regard narquois des experts de l'intoxication, communicants stipendiés, spécialistes du marketing et agents d'influence géopolitiques... Pire, à l'arrière-plan des intoxicateurs professionnels, n'est-ce pas l'architecture sociotechnique illustrée par Facebook qui est elle-même radicalement toxique, puisque son revenu est fondé sur la revente des données produites par les utilisateurs et que ceux-ci produisent d'autant plus de données qu'ils sont psychiquement dépendants de la plateforme ? On sait combien, de ce point de vue, la population adolescente est à risque.

5. L'étude d'Elisabeth Dubois montre cependant que « les bulles cognitives » ne sont pas aussi étanches que le voudrait le slogan anti-GAFAM en vogue (Dubois et Blank 2018). Voir en ligne.

Les droits de la critique sont imprescriptibles, et cela d'autant plus que l'individu critiqué est puissant. Or Mark Zuckerberg incarne précisément une des nouvelles formes de la puissance : une plateforme accueillant deux milliards et deux cents millions d'utilisateurs dans le monde (auxquels il faut ajouter les centaines de millions de WhatsApp et Instagram), dont la capitalisation boursière – malgré ses hauts et ses bas – reste mirobolante et dont la technologie (centres de données, intelligence artificielle, peut-être réseau de satellites) triomphe face à celle des médias classiques. Dans son *Discours sur la première décade de Tite-Live* (livre I, chapitre 3), Machiavel loue les Romains d'avoir sacralisé la fonction tribunitienne. Résumons ainsi le rôle des tribuns : accuser les grands (particulièrement ceux qui sont soupçonnés de viser la royauté) et défendre la plèbe. Avec le grand penseur politique de la Renaissance, j'admire la constitution de la République romaine. Il ne s'agit donc pas ici pour moi de faire taire les accusateurs ni de nier la pertinence de nombre de leurs arguments. Que l'on m'autorise pourtant à poser trois questions. La première, presque en passant, et qu'on ne prendra pas trop au sérieux : de quel côté se trouve la plèbe, dans l'affaire Zuckerberg ? Car ce sont les élites intellectuelles de l'université, du journalisme et de l'édition qui maudissent Facebook et c'est bien la populace de ceux qui n'avaient pas de voix dans l'ancien régime de communication qui peuplent en masse les médias sociaux. D'autre part, et c'est là ma seconde question, plus insistante : l'accusation et la déploration, fortement teintées de ressentiment – si légitimes soient-elles – suffisent-elles pour comprendre ? Car seule une intelligence vaste et systémique de la mutation anthropologique en cours pourra inspirer un idéal politique à la hauteur des enjeux. En effet, après l'analyse critique et la synthèse généralisante, l'exigence de responsabilité intellectuelle m'impose au moins d'esquisser un projet, une direction de pensée et d'action. Ce qui m'amène à la troisième question, que pose Wormser, mais à laquelle son livre ne répond pas : comment reconduire – voire même réinventer – la tradition de l'humanisme philosophique et l'idéal d'émancipation des Lumières dans le nouvel espace politique ?

La mutation anthropologique

Commençons par quelques faits saillants à l'échelle de la planète que ni la colère contre les riches et les puissants ni la nostalgie de la haute culture idéalisée du passé ne doivent faire oublier. Certes, le dernier siècle a été marqué

par des guerres épouvantables, des génocides, des totalitarismes implacables et mille formes d'oppression, coloniales et autres. Pourtant, entre la fin du XIX^e siècle et le début du XXI^e siècle, la population humaine est passée d'un milliard à sept milliards d'individus. Gagnant plus de 20 ans, l'espérance de vie a bondi⁶. La proportion de la population alphabétisée est passée de quinze à plus de quatre-vingts pour cent⁷ et la grande pauvreté a fortement reculé⁸. Je ne prétends ni que « tout va bien » ni qu'il n'existerait plus d'injustice, d'oppression ou de malheur. La liste interminable des maux de notre époque est tout aussi présente à mon esprit qu'à celui de mes lecteurs et je suis comme eux accablé par les nouvelles quotidiennes. Mais je voudrais observer le phénomène Internet à l'échelle de notre espèce dans son ensemble. À comparer avec le XIX^e siècle, nous avons affaire à une nouvelle humanité : urbaine, éduquée, en meilleure santé et dont la moitié féminine a gagné en puissance. Or au passage du XX^e au XXI^e siècle, cette humanité s'est précipitée dans l'espace numérique⁹. Alors qu'un pour cent de la population mondiale était connecté à l'Internet à la fin du XX^e siècle, nous avons dépassé les cinquante-cinq pour cent en 2018. À peine une génération ! Dans dix ou quinze ans, l'Asie et l'Afrique auront rejoint les taux de connexion du reste du monde et nous nous rapprocherons des quatre-vingt-dix pour cent de branchement.

L'Internet a distribué le pouvoir de lire, d'écrire, de publier, de critiquer, de classer et d'organiser l'information dans un espace de communication ubiquitaire. Qu'une information se trouve en un point du réseau et la voici partout. Dès lors, si l'échange de messages point à point a toujours lieu, le gros de la communication sociale s'effectue de manière stigmergique. L'étymologie grecque de ce mot étrange explique assez bien son sens : des signes (*stigma*) sont déposés dans l'environnement par l'action ou le travail (*ergon*) de membres d'une collectivité, et ces signes guident en retour – et récursivement – leurs actions¹⁰. Dans l'espace numérique, nous communiquons principalement par l'intermédiaire d'une énorme masse de données commune,

6. Voir Max Roser (2018), en ligne.

7. Voir Max Roser et Esteban Ortiz-Ospina (2017b), en ligne.

8. Voir Max Roser et Esteban Ortiz-Ospina (2017a), en ligne.

9. Sur la notion d'espace numérique voir Vitali-Rosati (2018).

10. « La stigmergie exprime la notion que les actions d'un agent laissent des signes dans l'environnement, signes perçus par lui-même et les autres agents et qui déterminent leurs prochaines actions » voir les pages consacrées au concept sur Wikipédia, en français et en anglais.

masse dont chaque action en ligne transforme le système de relations. Dans son récent ouvrage, Gérard Wormser évoque « le spectre d'une vie humaine asservie aux données » (Wormser 2018, 60). Le fondateur de *Sens Public* a sans doute bien perçu l'ampleur de la mutation, mais il nous faut dépasser l'effroi initial.

Notre intelligence collective prolonge celle des espèces sociales qui nous ont précédées, et particulièrement celle des grands singes. Mais l'usage du langage – et d'autres systèmes symboliques – tout comme la force de nos moyens techniques nous a fait passer du statut d'animal social à celui d'animal politique. Avec le langage sont venus les institutions sociales complexes, la réflexion, la création et l'échange des idées. Proprement humaine, la *Polis* émerge de la symbiose entre des écosystèmes d'idées et les populations de primates parlants qui les entretiennent, s'en nourrissent et s'y réfléchissent. L'évolution des idées et celles des peuplements de Sapiens se déterminent mutuellement. Or le facteur principal de l'évolution des idées réside dans le dispositif matériel de reproduction des symboles. Au cours de l'histoire, les symboles (avec les idées qu'ils portaient) ont été successivement pérennisés par l'écriture, allégés par l'alphabet et le papier, multipliés par l'imprimerie et les médias électriques. Ils sont aujourd'hui numérisés et calculés, c'est-à-dire qu'une foule de robots logiciels – les algorithmes – les enregistrent, les comptent, les traduisent et les transforment automatiquement. En somme, l'évolution culturelle nous a menés au point où les écosystèmes d'idées se manifestent sous l'avatar de données animées par des algorithmes dans un espace virtuel ubiquitaire. Et c'est dans cet espace que se nouent, se maintiennent et se dénouent les liens sociaux, là que se jouent désormais les drames de la *Polis*...

Considérons l'individu du XXI^e siècle dans son intégralité : un corps et un esprit, impossibles à séparer, fonctionnant de concert. Son esprit virtuel se connecte aux nuages (le « cloud ») – aux nœuds centraux du réseau – là où se trouve la puissance de mémoire et de calcul. Son corps actuel se branche aux bords du réseau – le « edge » –, c'est-à-dire aux interfaces, sur les PC, tablettes, smartphones, bracelets-montres enregistrant les signes vitaux, lunettes de réalité virtuelle ou augmentée... Cette part physique interagit avec une multitude d'objets connectés et de robots. Retour à la stigmergie : plongés dans le réseau, la tête dans les nuages et les mains aux interfaces, nous communiquons par l'intermédiaire de la masse océanique des données que nous produisons, transformons et consommons. Et chaque fois que nous ai-

mons, partageons, suivons ou agissons d'une manière ou d'une autre, voici que nous déposons dans la mémoire collective des phéromones électroniques qui déterminent en boucle l'action des autres et entraînent les neurones formels des intelligences artificielles.

La mission historique des plateformes

Le rôle historique des GAFAM (Google, Apple, Facebook, Amazon et Microsoft), des NAFU (Netflix, AirBnB, Tesla et Uber) et des BATX (Baidu, Alibaba, Tencent et Xiaomi) aura été de canaliser le déversement de la population mondiale dans le nouvel espace anthropologique. Sans doute les grandes entreprises du numérique ont-elles été fort actives dans ce processus, mais pourquoi refuser aux multitudes leur pouvoir d'action et leur responsabilité ?

Qu'on me permette une digression sur les origines théoriques de cette dénégation de toute responsabilité active au plus grand nombre. Il est temps de relativiser les critiques acerbes que Theodor Adorno¹¹ avait adressé aux médias et aux industries culturelles au milieu du XX^e siècle : enregistrement musical, cinéma, radio et télévision... Sans doute traumatisé par l'expérience du nazisme, le philosophe allemand exilé aux États-Unis portait un jugement négatif sur la culture populaire de son époque. Il était même allé jusqu'à condamner le jazz parce que ses disques étaient devenus des produits de grande consommation promus par la radio : une sorte de musique industrielle intrinsèquement capitaliste. La postérité intellectuelle d'Adorno – Debord, Baudrillard, Virilio suivis par la longue cohorte des dénonciateurs des médias et de l'abrutissement des foules – est encore aujourd'hui très vivace dans le monde académique et s'applique désormais au médium algorithmique. Cette tradition de la théorie critique (à ne pas confondre avec la pensée critique qui n'est autre que l'usage réfléchi de la raison) pose problème parce qu'elle considère « les masses » comme uniformément passives, victimes d'une aliénation économique et politique que la technique ne peut que redoubler en la réifiant. Au nom de l'émancipation complète qui l'attend dans un futur toujours reporté, la théorie critique refuse de reconnaître au plus grand nombre quelque parcelle d'autonomie et de responsabilité au pré-

11. Voir *La Dialectique de la Raison* rédigé avec Max Horkheimer (Adorno et Horkheimer 1974).

sent. Les seuls véritables acteurs seraient ceux qu'un récit convenu désigne à l'avance comme oppresseurs.

Pour revenir à Facebook, force est de constater que les masses se sont bel et bien précipitées dans les médias sociaux et les nouveaux services en ligne. Mark Zuckerberg, comme d'autres potentats du numérique, a certes surfé sur cette vague en l'exploitant. Dans un récit alternatif – ou peut-être simplement complémentaire – à celui de l'accusation, les GAFAM et consorts ne sont que des truchements, les outils industriels qu'une génération a saisis pour accomplir le plus grand saut de connexion de l'histoire de l'humanité. Dépassées et débordées par le rôle qu'elles ont fini par jouer, les grandes plateformes n'ont-elles pas été portées au pouvoir par leurs utilisateurs ?

Au pouvoir, en effet, puisqu'elles remplissent désormais des fonctions régaliennes, traditionnellement attribuées aux états souverains. Facebook prétend garantir nos identités. Il avertit notre famille et nos amis que nous sommes en sécurité quand arrive un désastre naturel ou une attaque terroriste. Dans une de ses déclarations les plus célèbres, Mark Zuckerberg affirme qu'une des missions de sa plateforme est de s'assurer que les processus électoraux se déroulent d'une manière loyale et ouverte dans les pays démocratiques¹², une profession de foi nécessaire en 2017, mais qui eut été impensable en 2004, année de la fondation de l'entreprise. Facebook n'est pas seul en cause. Google Earth et Google Street View sont maintenant utilisés par plusieurs gouvernements locaux et nationaux comme interface d'affichage du cadastre et parfois comme source d'information primaire. Twitter est devenu un service mondial quasi officiel pour la politique, la diplomatie et les nouvelles puisque tous les chefs d'État, ministres, ambassadeurs, dirigeants de partis et journalistes y ont un compte. Microsoft vend son infrastructure numérique aux écoles publiques de régions ou de pays entiers. Le royaume du Danemark a ouvert une ambassade officielle auprès de la Silicon Valley¹³. Les cryptomonnaies indépendantes des États-nations (comme Bitcoin) connaissent un succès croissant et seront bientôt intégrées aux principales plateformes, tandis que les contrats intelligents¹⁴ commencent à contourner les anciennes méthodes d'authentification des actes et la bureaucratie à base de papier des gouvernements.

12. Voir Heath (2017), en ligne.

13. Voir Baugh (2017), en ligne.

14. Supportés par exemple par *Ethereum*, voir en ligne

En somme, les mégaplateformes ne sont pas seulement les entreprises pourvues de la plus grande capitalisation boursière, elles jouent de fait un rôle indispensable de service public à une échelle géographique et démographique inouïe. Mais il y a plus : les gouvernements et les opinions publiques demandent aux plateformes de filtrer leurs contenus et d'expulser les trolls... mais aussi de garantir la liberté d'expression. La plateforme censure toujours trop... ou pas assez ! Or ce rôle de contrôle moral de l'espace public et de police de la parole était traditionnellement dévolu aux institutions religieuses et aux états souverains. En réclamant du Roi Mark qu'il rende la Justice sur son immense territoire virtuel, nous le confirmons dans sa nouvelle souveraineté. Mais parce que sacré Roi, le voici proche de devenir une victime sacrificielle : les sommets du pouvoir, comme les grands arbres, attirent la foudre.

Au-dessus des états

Les États-nations n'ont pas disparu, ils ont été déplacés et recontextualisés dans un nouveau système de pouvoir qui ne leur laisse plus forcément le dernier mot. Au-dessus des états, voici la nouvelle canopée politique, l'oligopole des grandes plateformes – les seigneurs des données – avec leurs écosystèmes de standards et de logiciels ouverts, leur longue traîne d'entreprises associées et de start-up, leurs milliards d'utilisateurs. On y fabrique les outils cognitifs et la glue émotive qui coordonne le reste de l'industrie. Là se tissent, se confirment et se donnent à lire les liens sociaux. Ici se creusent les labyrinthes mouvants et les chambres d'écho d'une opinion publique dont les médias traditionnels ne sont plus maîtres. L'Église avait perdu de son autorité avec la montée de l'imprimerie. De même, emporté par les vagues de données, le quatrième pouvoir glisse des mains des professeurs et des éditeurs. Les journalistes ne jouent plus désormais qu'un rôle d'amplification et de réentrée de l'information dans un jeu complexe où les médias sociaux et les moteurs de recherche tiennent le rôle principal¹⁵.

Au vu de la situation en Chine, en Russie ou aux États-Unis (où la configuration est plus ambivalente), il est clair que les plateformes ont noué des relations symbiotiques avec les pouvoirs politiques traditionnels. Pour mieux se disputer le contrôle sur les esprits, les appareils militaro-policiers, les services de renseignement et de propagande, les leaders populistes en prise di-

15. Voir Boyd (2018), en ligne.

recte avec les foules, tous se sont rapidement adaptés au nouveau champ de bataille numérique¹⁶. Et pour faire bonne mesure, ajoutons au tableau la montée des mégapoles – ces hubs planétaires des réseaux de transport et de commerce – qui s’engagent dans la course à la ville intelligente. Avec l’aide des géants du numérique, les grandes agglomérations rivalisent d’enthousiasme pour l’Internet des choses, la réalité augmentée, l’intelligence artificielle et la blockchain! C’est donc en s’alliant avec les vieux pouvoirs territoriaux que les seigneurs des données dessinent la configuration politique du XXI^e siècle. Paradoxalement, si l’information numérique est ubiquitaire, la richesse et la puissance sont toujours fortement centrées autour de quelques nœuds d’accumulation, d’échange et de calcul¹⁷. Les nuages de l’Internet retombent aujourd’hui en pluie d’or de part et d’autre du Pacifique. Il n’est plus de pôle géopolitique puissant sans champions algorithmiques¹⁸. La Chine monte, tandis que l’Amérique commence à douter d’une Silicon Valley pourtant toujours en tête de l’innovation. Économiquement conquise par les GAFAM, idéologiquement divisée contre elle-même par une maskirovka insidieuse et multiforme, l’Europe vacille et ses intellectuels désespèrent.

La guerre secrète

Il faut s’y faire : la *Polis* a basculé dans la grande base de données virtuelle de l’Internet. Dès lors, les luttes de pouvoir – toutes les luttes de pouvoir, qu’elles soient économiques, politiques ou culturelles – sont reconduites et compliquées dans le nouvel espace numérique. Sur le terrain glissant des médias sociaux, les camps qui s’affrontent disposent leurs armées de trolls coordonnées en temps réel, équipées de bots dernier cri, renseignées par le big

16. Deux ouvrages récents témoignent bien de la transformation de l’Internet en champ de bataille : *Network Propaganda. Manipulation, Disinformation, and Radicalization in American Politics* de Yochai Benkler, Robert Faris, and Hal Roberts (2018) et *LikeWar : The Weaponization of Social Media* de P. W. Singer and Emerson T. Brooking (2018).

17. Voir l’étude classique de Fernand Braudel – *Civilisation matérielle, économie et capitalisme, XVe- XVIIIe siècle* (1986) – qui montre bien comment les centres se déplacent au cours des siècles : les villes-monde comme Venise, Gênes, Anvers, Amsterdam, Londres ou New York émergent, rayonnent puis déclinent.

18. Comme Vladimir Putin l’a dit avec raison aux jeunes russes dans son discours pour la rentrée scolaire en automne 2017 : « Ceux qui deviendront les leaders dans le domaine de l’intelligence artificielle domineront le monde ». On croirait entendre Laurent Alexandre! Voir Vincent (2017), en ligne.

data et augmentées par le *machine learning*. Dans la guerre civile mondiale qui fait rage, politique intérieure et extérieure inextricablement mêlées, les nouveaux mercenaires sont les influenceurs.

À la biopolitique (Foucault 1975) aujourd'hui généralisée à la biosphère, faut-il ajouter une psychopolitique numérique à base de neuromarketing, de données personnelles et de gamification du contrôle¹⁹ ? Peut-être. Sans doute l'action politique a-t-elle changé d'échelle et d'instruments depuis qu'elle utilise la stimulation dopaminergique et les renforcements narcissiques addictifs de la communication numérique. Mais une communauté politique n'exige-t-elle pas toujours des mécanismes d'identification au groupe et la création d'une résonance émotive et sémantique entre les psychismes des individus ? La politique se fonde nécessairement sur des ressorts psychosociologiques, à l'échelle de la cité comme à celle du parti. D'autre part, les anciens médias unidirectionnels – qui n'étaient pas moins contrôlés par les puissances économiques et politiques que les nouveaux médias interactifs – maîtrisaient déjà d'excellents principes de propagande, encore utilisés avec succès de nos jours : répétition continue, simplicité des mots d'ordre, images mémorables, provocation affective et résonance identitaire. Les tragédies politiques du XX^e siècle n'ont pas eu besoin du numérique pour s'accomplir.

Avant de déplorer l'arrivée de l'Internet, consultons les enseignements de l'histoire. Hors le mythe ou l'utopie, que l'on nous montre une vie politique réelle qui soit étrangère à la domination, à la polémique, aux scandales, à la calomnie et à la trahison. La bataille pour le pouvoir est partout. Seuls changent les lieux, les valeurs, les régimes et les partis. Pourquoi notre époque échapperait-elle aux invariants anthropologiques du conflit ou de la tromperie ? En analysant les interférences clandestines que des puissances hostiles ont perpétrées dans la vie politique des démocraties par le canal des médias sociaux, on se souviendra de la lucidité d'un Machiavel au matin de l'imprimerie :

La guerre secrète consiste à se mettre dans la confiance d'une ville divisée, à se porter pour médiateur entre les deux partis jusqu'à ce qu'ils en viennent aux armes : et quand l'épée est enfin tirée à donner des secours prudemment dosés au parti le

19. Sur la gamification du contrôle social en Chine, voir l'article « En Chine, un système de "crédit social" règlera la vie des citoyens », ainsi que Denyer (s. d.), en ligne, Botsman (2017), en ligne, Osborne (2015), en ligne, Custer (2015), en ligne.

plus faible, autant dans le but de faire durer la guerre et de les laisser se consumer les uns par les autres, que pour se garder, par un secours trop massif, de révéler son dessein de les opprimer et de les maîtriser tous deux également. Si l'on suit soigneusement cette marche, on arrive presque toujours à son but. (Machiavel 2010, 588)

Les lieux communs de l'émancipation

Résumons nos constats : premièrement, la grande majorité de la population mondiale est – ou sera bientôt – connectée à l'Internet, c'est-à-dire à un appareillage global de communication, de mémoire et de calcul encore inimaginable au début du siècle dernier et qui contient désormais la vie de la cité, de ses cimes scientifiques, artistiques et spirituelles à ses abîmes criminels. Du côté obscur de la force, les pouvoirs économiques, politico-militaires et culturels utilisent sans scrupules le nouveau médium algorithmique afin de combattre leurs adversaires et de séduire, exploiter ou asservir leurs sujets. Du côté lumineux, l'Internet ouvre à une liberté d'expression et à une diversité d'information inconnues à l'âge des médias classiques. Nous commençons tout juste à explorer les possibilités de coordination à grande échelle et d'intelligence collective offertes par l'Internet, de Wikipédia à GitHub, de l'économie collaborative aux villes numériques. Bien que nous ayons tendance à les opposer d'un point de vue moral, le pouvoir d'oppression et de destruction – négatif – et la puissance de création et de pensée – positive – sont deux aspects de la même augmentation cognitive. Et nous-mêmes – à notre corps défendant, et quoique l'obscurité semble toujours venir de l'autre – nourrissons ces deux figures de la force et leur donnons prise, y compris par la servitude volontaire si bien décrite par La Boétie et par la participation à la foule de microprocessus et de rapports sociaux qui nous traversent et que Michel Foucault nous a invités à analyser en détail (Foucault 1975 , 1976). Enfin, aucun système d'intelligence artificielle ne nous protégera contre les « fake news » et les manipulations en tous genres. Le seul remède sérieux aux menées du côté obscur et la meilleure manière d'actualiser les virtualités positives du nouveau médium sont l'éducation des populations et tout particulièrement leur entraînement à la communication réfléchie et à la pensée critique. Or le rayonnement de la pensée critique – et c'est cette hypothèse

qui va orienter la suite de ce texte – dépend à son tour de la qualité de la recherche et de l’enseignement en sciences humaines.

Désormais plongés dans un nouvel espace politique, il nous faut assumer nos responsabilités d’intellectuels et formuler, sinon le programme complet au moins l’orientation fondamentale, d’une nouvelle philosophie des Lumières, d’un humanisme de l’âge numérique qui soit à la hauteur de la puissance des algorithmes et de la mutation anthropologique en cours. Le « nous les intellectuels » de la phrase qui précède englobe au premier chef les chercheurs en sciences humaines, quelle que soit leur situation institutionnelle ou professionnelle. Par exemple : les auteurs, lecteurs et éditeurs des collections d’ouvrages et des revues scientifiques en sciences humaines. Plus largement, l’expression « les intellectuels » désigne l’ensemble des acteurs qui – au-delà de leur stricte pratique professionnelle et de leurs engagements particuliers à court terme – se sentent responsables de la civilisation qui vient et de leur legs aux générations futures. S’il est encore vivant sous ses avatars contemporains, le mouvement des Lumières est d’abord une entreprise de croissance et de diffusion des connaissances. Or ce mouvement ne doit plus « s’adapter » au monde numérique, et encore moins déplorer son avènement, mais plutôt prendre la tête d’un mouvement d’émancipation de l’intelligence collective dans la société mondiale data-centrique.

Commençons par énoncer quatre impératifs qui recueillent aujourd’hui les suffrages des gens bien informés :

1. maximiser les communs de la connaissance ;
2. ouvrir les données, les modèles, les algorithmes et les logiciels ;
3. optimiser la transparence des processus en ligne ;
4. assurer autant que possible la souveraineté pratique et légale des individus sur les données qu’ils produisent ou qui les visent.

Maximiser les communs de la connaissance signifie qu’on ne devrait pas avoir à payer pour accéder à des bases de connaissance. Cela comprend les dictionnaires, encyclopédies et manuels scolaires, les résultats de la recherche, les données produites par les institutions ou les organismes gouvernementaux, les informations nécessaires à la prévention des maladies et à l’amélioration de la santé publique, les bases de données juridiques et les ressources d’apprentissage de toutes sortes. Une bonne part de ce programme est déjà réalisé ou en cours de réalisation. Constatons toutefois que beaucoup de données qui

sont officiellement publiques sont inaccessibles dans la pratique faute d'outils de recherche, de navigation et d'exploitation adéquats. On préfère d'ailleurs maintenant parler de données FAIR (*findable, accessible, interoperable and reusable*) plutôt que de données publiques²⁰. Outre qu'elle favorise directement l'apprentissage et l'accès à l'information, la « mise en commun » promeut la transversalité : le décloisonnement et la pollinisation croisée si utiles à la croissance de l'intelligence collective. Les communautés de création et de recherche peuvent ainsi interconnecter et recombinaison les informations disponibles, ce qui nous amène au principe d'ouverture.

Le tropisme vers l'ouverture a d'abord été expérimenté dans la communauté scientifique (comme la publication rapide dans arXiv.org), le mouvement du logiciel libre, la licence creative commons, Wikipédia, GitHub et bien d'autres entreprises. On ajoute ici au principe de l'accès gratuit la possibilité pour des volontaires de participer de manière collaborative à la croissance des communs. L'ouverture a été adoptée par de nombreux gouvernements et grandes industries. « Ouvert par défaut » devient la norme. Donnons en exemple Microsoft (le M final des GAFAM), qui a fait sa fortune sur la commercialisation de logiciels propriétaires fermés, mais dont les principaux revenus viennent maintenant de services nuagiques, et qui annonce en 2018 que tout son portefeuille de brevets logiciels bascule dans l'open source²¹. Pourquoi ce tropisme vers l'ouverture ? Parce qu'il maximise l'amélioration des biens et des services, favorise la confiance et soutient l'engagement collaboratif. Il peut être appliqué aux formats de données, aux systèmes d'exploitation, aux modèles, aux algorithmes et même au matériel²². L'ouverture s'applique également aux taxonomies, ontologies et architectures d'information puisque l'espace public que nous appelons de nos vœux doit encourager tous ses participants à créer, commenter, catégoriser, évaluer et analyser son contenu en connaissance de cause et avec pleine capacité d'agir.

Enfin, la transparence des acteurs et des processus forme la base de la confiance et conditionne l'authenticité d'un dialogue sans lequel s'étiole la vie démocratique d'une cité. Afin de limiter la corruption et les manipulations – qui s'alimentent évidemment de l'opacité – les gens et les données doivent être retraçables et auditables, y compris dans le cas des administra-

20. Voir le site web de Force 11.

21. Voir Vaughan-Nichols (2018), en ligne.

22. Voir la notice Wikipedia de l'Open Source Hardware.

teurs des plateformes et des responsables des gouvernements. Dans la continuité de la tradition d'émancipation des Lumières, la transparence ne doit pas être unidirectionnelle (les sujets dominés sous le regard panoptique d'une caste dominante qui reste dans l'ombre), mais bel et bien réciproque²³, sans distinction entre dirigeants et dirigés. Comment pourrions-nous viser une intelligence collective réflexive – permettant aux équipes et communautés de toutes tailles d'observer et de comparer leurs activités cognitives – si nous n'affirmions pas la transparence comme une des valeurs principales de notre mouvement ?

À ces arguments de bons sens et d'éthique en faveur de la transparence j'ajoute, pour leur faire contrepoids, deux remarques dont l'une relève du réalisme politique et l'autre de la banalité anthropologique. Premier point, selon la théorie²⁴ et la pratique de Julian Assange, les pouvoirs contemporains ont une dimension « conspirationniste » : leurs réseaux d'échanges d'informations secrètes leur permettent de penser et d'agir à l'insu de leurs adversaires et du public. Pour les vaincre, leurs opposants doivent interrompre ou manipuler leur communication interne et surtout organiser des fuites vers l'extérieur. Illustrée par l'action de Wikileaks et d'autres réseaux de journalistes et de lanceurs d'alerte, la guérilla informationnelle mène une guerre d'attrition contre une conspiration ennemie qu'elle force à épuiser ses ressources pour maintenir la sécurité de son réseau. Le secret devenant de plus en plus coûteux, les pouvoirs adoptent des contre-stratégies qui minimisent l'occultation classique au profit de modes de coordination interne et de stupéfaction de l'adversaire qui sont d'autant plus efficaces qu'elles se déroulent au grand jour. Second point, la pression à la transparence sur les individus et les organisations les plus puissants est un phénomène aisément observable. Les grandes compagnies cotées en bourse doivent publier régulièrement leurs résultats et sont scrutées par une multitude d'analystes et d'activistes. Au sommet des États-nations riches et démocratiques, les présidents, les ministres, voire l'ensemble des membres de la classe politique et journalistique exposent leurs idées et sentiments sur les médias sociaux et vivent dans une sorte de télé-réalité permanente où leurs actes sont passés au crible par le public²⁵. Tandis que

23. Voir Brin (1998).

24. Voir la section *Writings* de la notice Wikipédia consacrée à Julian Assange.

25. La chose n'est pas nouvelle. Agamemnon dit à la scène 5 de l'acte I de l'*Iphigénie* de Racine : « Triste destin des rois ! Esclaves que nous sommes *Et des rigneurs du sort et des discours des hommes ! Nous nous voyons sans cesse assiégés de témoins* et les plus

les faibles et les démunis restent dans l'obscurité et que les vieux pouvoirs maffieux cultivent le secret, la transparence assumée pourrait signaler l'éclat des puissances montantes.

Les nécessités de toutes sortes et le désir d'attention qui valorisent la transparence doivent être tempérés par le respect de la vie privée des citoyens. Luciano Floridi (2014) a bien remarqué que si la transmission d'informations était encore coûteuse au XIX^e et XX^e siècle, c'est au contraire la limitation de la communication qui devient onéreuse dans l'espace numérique. Si bien que l'équilibre entre vie publique et vie privée qui s'était stabilisé durant quelques siècles est remis en question à l'âge des téléphones intelligents, des médias sociaux, de l'Internet des choses, du trafic des données dans les nuages... et de la collusion entre les plateformes et les services de renseignement. Depuis quelques années, l'ancienne problématique de la vie privée fait d'ailleurs place à la nouvelle revendication de la souveraineté des individus sur les données qu'ils produisent et qui les concernent. À qui nos messages sur les médias sociaux sont-ils en fin de compte diffusés ? Savons-nous exactement qui aura le droit d'exploiter nos données, et comment, lorsque nous acceptons les conditions d'utilisation des plateformes ? Pouvons-nous nous prévaloir du droit à l'oubli et demander à Google d'effacer de ses résultats des sites qui diffusent des informations calomnieuses sur notre compte ou qui révèlent des fautes depuis longtemps passées et expiées ? À ces craintes s'ajoutent de nouveaux espoirs. Les informations cruciales qui concernent notre état civil, notre situation financière ou juridique, nos titres de propriété, nos contrats, nos diplômes ou notre santé sont aujourd'hui confisquées par des institutions cloisonnées et opaques datant de l'ère du papier et face auxquelles l'individu se trouve en situation de dépendance. Pourquoi ne pas placer le citoyen du nouvel espace politique au centre et – une fois les informations garanties par une procédure acceptée de tous – lui accorder l'autogestion de ses données personnelles tout en fluidifiant les transferts d'information ? Fini le harcèlement bureaucratique des innombrables formulaires incompatibles issus d'administrations voisines qui réclament les mêmes informations sous d'autres rubriques et les mêmes pièces à fournir pour la millième fois ! C'est toute la promesse de « l'identité souveraine » appuyée sur la chaîne de blocs ou d'autres cryptotechnologies²⁶. Cette identité numérique autogérée nous libérerait du poids de prouver sans

malheureux osent pleurer le moins ! » Je souligne. Racine reprend d'ailleurs ici l'*Iphigénie en Aulide* d'Euripide.

26. Voir (2017), en ligne.

cesse à nouveau que nous sommes nous-mêmes. Elle nous défendrait aussi bien contre l'oppression issue de la fragmentation bureaucratique des vieux états que contre l'exploitation informationnelle des nouvelles plateformes centralisées.

En somme, les nouvelles techniques d'autogestion de l'identité personnelle vont redéfinir la relation privé-public au profit d'une maîtrise directe de ses données par le citoyen qui pourra décider à qui, quand et pour quels usages ses informations cryptocertifiées sont transmises²⁷. Selon le même esprit, mais avec d'autres techniques, Tim Berners-Lee, le concepteur du World Wide Web, a récemment lancé le protocole « Solid » qui a précisément pour objectif de rendre aux individus le contrôle sur leurs données et les applications qu'ils utilisent²⁸. On retrouve le même souci de décentralisation et de réappropriation de leurs données par les utilisateurs dans des projets comme Holochain, Scuttlebutt et bien d'autres. Mais dans la perspective d'intelligence collective réflexive qui est la nôtre, il ne faudra pas seulement accorder aux personnes un contrôle sûr et pratique sur leurs propres données, mais aussi leur renvoyer l'image mouvante que leurs données agrégées et anonymisées dessinent ensemble – selon les communautés et les réseaux sociaux – grâce à des techniques d'analyse et de visualisation enfin démocratisées.

Notons que l'augmentation des communs de la connaissance, la prime à l'open source, l'exigence de transparence des processus et des algorithmes comme la souveraineté des individus sur leurs données font déjà partie de la culture politique en émergence. Même s'il ne sont pas encore réalisés, ces principes sont aujourd'hui largement acceptés et – sauf catastrophe ou régression – s'imposeront de plus en plus dans l'avenir. Sur la base de ces lieux communs, il nous faut maintenant formuler des propositions plus risquées, qui ne font pas encore partie de la doxa mais qui, si elles sont comprises et surtout mises en pratique, nous feront franchir un seuil décisif d'intelligence collective réflexive. Le projet de renouvellement de l'humanisme et de réactivation du programme d'émancipation des Lumières dans le nouvel espace numérique a deux aspects conjoints : épistémologique et politique.

27. Voir par exemple (2018), en ligne.

28. Voir (2018), en ligne.

Un projet épistémologico-politique

Premièrement, c'est un projet épistémologique. Les sciences humaines sont aujourd'hui balkanisées en disciplines et divisées en théories qui souvent ne s'entendent même pas sur leurs désaccords. Nous sommes ici fort loin du découpage des disciplines selon des lignes objectives de niveaux de complexité (physique, chimie, biologie...) qui prévaut dans les sciences de la nature et de l'accord général des sciences « dures » sur leurs méthodes et leurs catégories organisatrices. De plus, en sciences humaines, malgré quelques exceptions, le degré de formalisation des modèles et de testabilité des théories reste bien faible. Tout cela rappelle la situation qui prévalait dans l'alchimie avant l'émergence de la chimie moderne à la fin du XVIII^e siècle. Mais les conditions sont mûres pour une révolution scientifique. Les archives sont progressivement numérisées et la quasi-totalité des activités humaines produit un flot montant de données. Pour traiter ces données, nous disposons d'algorithmes d'analyse et de visualisation efficaces ainsi que d'une puissance de calcul gigantesque. Enfin, les modes de communication du XXI^e siècle ouvrent aux équipes de chercheurs de nouvelles avenues de dialogue, de partage de corpus et d'outils, de collaboration internationale et de publication. Pour peu que nous parvenions à transformer ces conditions favorables en mutation qualitative, les connaissances de la société humaine à son propre sujet – selon l'éventail complet des échelles : des petits groupes à l'espèce – deviendraient plus fiables, plus abondantes, mieux diffusées et mises au service du développement humain de manière plus pertinente²⁹.

Deuxièmement, il s'agit d'un projet politique. Je ne vise pas ici quelque engagement direct dans la vie politique des nations, dans les affrontements géopolitiques ou dans les guerres culturelles car, depuis quelques siècles, des prises de parti diverses et vite datées ont malheureusement contribué à discréditer l'entreprise scientifique dans son ensemble (qu'il suffise ici d'évoquer la biologie au XX^e siècle, qui fut raciste dans l'Allemagne nazie et anti-darwinienne dans la Russie stalinienne. Passons charitablement sur les engagements souvent douteux de philosophes et de chercheurs en sciences humaines). Les travaux dans le domaine des humanités pourront évidemment obéir à des finalités politiques et intervenir dans l'actualité, mais mon objectif dans ce texte est d'indiquer une orientation générale à l'échelle du siècle qui vient,

29. Je renvoie à mon livre *La Sphère Sémantique* (2011), notamment le chapitre 5 : « Pour une mutation épistémologique des sciences de l'homme ».

une feuille de route qui puisse mobiliser de manière durable l'ensemble des chercheurs dans la perspective d'un humanisme renouvelé, indépendamment des contingences de l'actualité politique. Je me situe dans une perspective universaliste de longue durée qui vise à mettre en lumière les effets d'autorité. Cette vision s'alimente notamment des analyses de Michel Foucault sur les épistémès ou régimes de vérité³⁰ et du récent ouvrage de Marcello Vitali-Rosati sur l'éditorialisation (2018). L'autorité que nous accordons à certaines connaissances et aux spécialistes qui les produisent, tout comme les croyances que nous entretenons au sujet de nos sociétés et des places qu'y occupent les un·e·s et les autres sont des éléments déterminants de l'ordre politique. Or ces régimes de vérité et ces hiérarchies d'autorité émergent de jeux symboliques conventionnels, eux-mêmes arrimés à des systèmes de médias³¹. Par exemple, l'arrivée de l'imprimerie en Europe a donné du pouvoir aux éditeurs et aux lecteurs tout en renforçant l'autorité de textes canoniques soigneusement révisés, traduits et « publiés ». Déclinait parallèlement le régime traditionnel d'autorité de l'Église, qui s'appuyait sur la transmission orale et manuscrite. C'est ce qu'illustrent le triomphe de l'humanisme au XVI^e siècle, l'aventure de la réforme luthérienne et le décollage de la communauté scientifique moderne. Bien conscient de la solidarité entre les modes de connaissance, les régimes d'autorité et le système des médias d'une époque, Marcello Vitali-Rosati a montré que le basculement de la communication sociale dans l'Internet provoque aujourd'hui des effets d'autorité – d'ailleurs complexes et différenciés selon les plateformes et les algorithmes – qui remettent en question les hiérarchies savantes et les procédures de légitimation des connaissances qui s'étaient stabilisées depuis quelques siècles. Par la distinction et l'explicitation des processus d'éditorialisation – c'est-à-dire les mécanismes distribués de production de l'autorité et de la croyance – qui se déploient dans le médium algorithmique, notre projet émancipateur contribuera à la création d'un espace politique favorable à la pensée critique comme à l'autonomie des individus et des communautés.

30. Voir *Les mots et les choses* (Foucault 1966) et *L'archéologie du savoir* Gallimard (Foucault 1969).

31. Voir notamment les travaux de Milman Parry, Eric Havelock, Harold Innis, Marshall McLuhan, Walter Ong, Jack Goody, Elisabeth Eisenstein et Frances Yates, auteurs que j'ai repris et cité dans mes livres *Les technologies de l'Intelligence* (1990) et *La Sphère Sémantique* (2011).

Comme les poissons dans l'eau, l'espèce humaine vit dans l'élément de la signification et la question du sens se trouve au cœur des sciences de l'Homme. Je fais l'hypothèse que le nouvel espace numérique va nous permettre de formaliser et d'étudier « la sphère sémantique » à une échelle et à un degré de précision que les technologies cognitives antérieures ne permettaient pas. La condition cruciale de ce programme épistémologico-politique est la transparence, puisque cette qualité supporte aussi bien la formalisation nécessaire au calcul que la réflexivité critique propre à l'humanisme philosophique. Cependant il ne s'agit plus dans ce cas de la transparence ordinaire sur laquelle on s'accorde sans peine, mais d'une transparence radicale qui vise les mécanismes moléculaires de production de sens : la sémantique linguistique, l'interprétation en contexte, les opérations de référence, les émergences coordonnées de l'autorité et de la croyance.

Afin de ne pas nous égarer dans quelque utopie irréalisable au parfum totalitaire, reconnaissons immédiatement les limites d'une telle entreprise. Michael Polanyi, l'un des plus grands épistémologues du XX^e siècle, a montré dans ses ouvrages³² que la connaissance scientifique se construisait dans un va-et-vient entre connaissance explicite (claire, formelle, décontextualisée) et connaissance tacite (opaque, intuitive, incarnée et contextuelle), va-et-vient qui fait écho à la dialectique classique de la théorie et de la pratique. Même dans les domaines scientifiques les plus rigoureux, aussi loin que l'on s'efforce de pousser le processus d'explicitation de la connaissance, il reste toujours une part inexplicable, un fond d'expérience personnelle, de tradition et d'évidence concrète qui ne se laisse pas ramener à la raison universelle. Non seulement la partie tacite de la connaissance ne peut être éliminée complètement, mais c'est elle qui fonde la partie claire et formelle, un peu comme la conscience discursive émerge des myriades de processus opaques de l'organisme. Cela ne veut pas dire que les efforts pour augmenter la dimension universelle et transparente du savoir soient vains – tout progrès de la modélisation est bon à prendre – mais plutôt que ces efforts ne seront jamais couronnés d'un succès définitif : la réserve d'implicite se reconstitue sans cesse et l'inconnu s'étend toujours de nouveau hors de portée de nos théories et de nos instruments d'observation. Reconnaissons nos limites et retenons que « tout » ne peut être explicité et encore moins théorisé.

32. Notamment *Personal Knowledge, Towards a Post-Critical Philosophy* (Polanyi 1958).

Faisons maintenant un pas de plus dans la voie de l'humilité intellectuelle avec Karl Popper³³. Même la partie de la connaissance qui a été formalisée ne nous offre pas d'appui inébranlable. En effet, aucune théorie ou modèle causal ne peut être prouvé absolument vrai. Il est impossible de démontrer la vérité d'une théorie parce qu'une même série d'observations – ou des résultats statistiques donnés – peuvent toujours être expliqués par plusieurs hypothèses causales. On peut seulement démontrer la fausseté d'une théorie lorsqu'elle échoue à passer un test empirique. Il importe de rappeler cette précaution épistémologique à l'heure où triomphent les méthodes statistiques à grande échelle (*big data, machine learning, etc.*). Les études statistiques peuvent suggérer des patterns causaux ou des règles générales, elles ne peuvent les établir absolument. Par ailleurs, l'identification et le découpage des objets comptés dans les méthodes quantitatives supposent nécessairement l'accommodation à une échelle pertinente et le choix d'un système de catégorisation, éléments rarement explicités qui sont forcément conventionnels et dépendants des contextes pratiques et des finalités des chercheurs.

En somme, notre programme épistémologique pour l'avenir des sciences humaines vise résolument à éclairer les mécanismes de production de signification en utilisant toutes les ressources du nouveau médium algorithmique, mais sans entretenir d'illusion sur la possibilité d'une transparence totale et sans prétendre parvenir à une seule théorie vraie. Il s'agit plutôt d'ouvrir la voie à l'exploration critique d'un univers de sens inépuisablement complexe. Pour ce qui est de la dimension politique, une des principales conditions de l'autonomie est la connaissance de soi. Dans cet esprit, le programme de l'intelligence collective réflexive fait converger asymptotiquement les objets et les sujets des sciences humaines. Les communautés analyseront leurs propres données et celles des autres au moyen d'outils partagés et de procédures explicites, selon une multitude d'hypothèses et de récits organisateurs toujours ouverts à la critique. Quant aux individus – chercheurs, auteurs, curateurs, évaluateurs – ils participeront de plusieurs communautés selon des rôles variés et contribueront de manière stigmergique à l'accroissement des connaissances.

Produire de la connaissance est évidemment un cas particulier de production de sens. Or dans les Sciences de l'Homme et de la Société la connaissance a précisément pour objets les conditions, les mécanismes causaux et les pro-

33. Voir *Objective Knowledge : An Evolutionary Approach* (Popper 1972).

cessus créatifs de la production de sens. Pour devenir les éclaireurs d'une civilisation de l'intelligence collective réflexive, les chercheurs en sciences humaines doivent donner l'exemple – dans leur pratique ! – d'une production de sens qui s'offre à la connaissance de la manière la plus transparente possible. Parce que leurs résultats et leurs processus de recherche peuvent devenir la matière première de nouvelles recherches, ils ont vocation à produire les données qui se prêteront le mieux à l'analyse et à l'interprétation. Quelle méthode sera la plus propre à rendre aussi transparente que possible la production des connaissances en sciences humaines ? Puisque nous nous intéressons à la production de la signification et de la connaissance, nos objets sont des actes. C'est pourquoi la pragmatique m'offre un cadre théorique pertinent. J'utiliserai les idées d'Austin (1975), de Wittgenstein (1993, 1953) et certaines notions très générales de théorie des jeux. Soulignons d'autre part que les actes de production de sens peuvent se ramener à des opérations de lecture et d'écriture – ces deux termes étant pris au sens le plus général – et cela d'autant plus qu'ils sont accomplis et captés dans le médium algorithmique. Lorsque je parlerai d'écriture ou de codage dans les lignes qui suivent je sous-entendrai l'opération symétrique de lecture ou de décodage que vise en principe l'écriture.

Suivant la distinction commode proposée par Austin, découpons l'acte symbolique en trois couches : locutoire, illocutoire et perlocutoire. La production de sens commence au niveau locutoire avec un énoncé linguistique, l'écriture d'un texte. Elle se poursuit au niveau illocutoire par la conformité de l'énonciation ou de l'écriture à certaines règles sociales, l'engagement effectif dans un « jeu de langage » conventionnel et l'immersion dans un contexte. Dans le cas des humanités, c'est à ce niveau que s'accomplissent les opérations de référence, de raisonnement et de dialogue, ici que l'on pointe vers un interlocuteur, un auteur, un corpus ou un ensemble de données. L'acte s'achève au niveau perlocutoire en produisant des effets dont les principaux, dans notre cas, sont la croyance et l'autorité. Examinons maintenant comment ces trois dimensions de la création de sens peuvent être formalisées.

IEML : un système de coordonnées sémantiques

Le niveau locutoire concerne cette partie de la sémantique qui dépend du dictionnaire et de la grammaire d'une langue. Le sens est ici proprement lin-

guistique. Or – aussi excessive que puisse paraître cette position – un texte en langue naturelle ne constitue pas la formalisation, ou l’expression transparente, de son sens. On sait, en effet, que les ordinateurs ne peuvent pas comprendre ou traduire un texte en langue naturelle simplement à partir d’une grammaire et d’un dictionnaire (le “code” de la langue) car le même texte en langue naturelle peut correspondre à plusieurs structures grammaticales sous-jacentes et le sens des mots est ambigu. On voit bien en consultant un dictionnaire que de nombreux mots ont plusieurs sens. Les humains interprètent – ou désambigüisent – le texte en s’aidant du contexte, c’est-à-dire en injectant leur compréhension du jeu de langage ou du genre discursif pratiqué par l’énonciateur, en reconnaissant l’environnement social et matériel de l’énonciation, en identifiant les référents vers lesquels pointent les groupes de mots, etc. Autrement dit nous décodons un texte en langue naturelle grâce à une précompréhension, le plus souvent intuitive, de la situation et de ses enjeux. Comme il est impossible de demander à un ordinateur de suivre cette voie, le traitement automatique des langues utilise des méthodes statistiques dont le résultat doit toujours être revu par un humain. Dans le but de formaliser la sémantique linguistique, c’est-à-dire la dimension locutoire du sens, j’ai inventé une langue univoque, IEML (Information Economy MetaLanguage) dont les textes sont des formalisations de leur sémantique : les mots n’ont qu’un seul sens et la structure grammaticale des textes est entièrement déterminée. Chaque texte IEML distinct correspond à un concept distinct et le sens linguistique d’un texte IEML peut être déduit automatiquement de sa lecture. C’est pourquoi un texte IEML est appelé un USL (pour *Uniform Semantic Locator*). Un algorithme détermine le sens linguistique d’un USL et calcule les relations sémantiques entre les USL uniquement à partir du dictionnaire et de la grammaire d’IEML, sans avoir besoin de faire appel à des statistiques sur d’énormes corpus. Dans la pratique, on utilisera IEML non pour écrire des livres ou des articles, mais comme un instrument de balise sémantique de textes rédigés en langue naturelle, comme un outil de catégorisation des données et comme un métalangage documentaire permettant de décrire formellement (donc de manière transparente et critique) les jeux de langage, les rôles de communication et les modalités de référence qui contribuent à la production du sens.

IEML est basé sur les grands acquis de la linguistique du siècle dernier. Selon Ferdinand de Saussure (1985), père de la linguistique contemporaine, les symboles linguistiques sont constitués de deux parties, le signifiant (une image

acoustique ou visuelle) et le signifié (un concept ou une catégorie abstraite). Le rapport entre les deux parties du symbole est conventionnel ou arbitraire. Pour illustrer ce dernier point, il est facile de constater que le même signifié correspond à des signifiants distincts dans différentes langues : cheval, cavallo, horse, pferd, etc. Saussure a également montré que le plan du signifiant, ou la phonologie des langues, était basé sur un système de différences entre les sons, chaque langue ayant sa propre liste de phonèmes et surtout sa propre manière de disposer les seuils de passage entre deux phonèmes dans le continuum sonore. Un phonème n'existe pas de manière isolée, en dehors d'un éventail de variations, un peu comme les notes de musique n'existent que par rapport à un système musical. De la même manière, les signifiés ne sont pas des atomes de sens se suffisant à eux-mêmes, mais correspondent à des positions dans des systèmes de différences : les paradigmes. Ainsi, les noms de couleurs s'opposent, se distinguent et se nuancent dans un paradigme des couleurs qui varie avec les langues. Les verbes relatifs aux transactions commerciales (acheter, vendre, prêter, rendre, louer, etc.) marquent des oppositions et des différences qui s'expliquent mutuellement sur le fond invariant du même genre d'action. La sémantique linguistique ne s'ancre donc pas dans des réalités naturelles fixes et indépendantes, mais dans un processus de comparaison, d'opposition, de différenciation et de renvois entre signifiés au sein d'une grille systémique bouclée sur elle-même, comme le sens d'un mot dans le dictionnaire est défini par d'autres mots qui, eux-mêmes, etc. Résumons le moment saussurien de la linguistique en disant que les langues sont conventionnelles et différentielles. L'autre grand moment de la linguistique du XX^e siècle est dû à Noam Chomsky. Pour le professeur du MIT, la capacité linguistique est un trait génétiquement déterminé de l'espèce humaine. Les langues, malgré leur diversité et leur évolution continue, partagent toutes la même « grammaire universelle » correspondant à cette habileté linguistique innée. Cette théorie expliquerait pourquoi les enfants apprennent si vite et si spontanément à parler, sans qu'on ait besoin de leur donner des leçons de grammaire (Chomsky 2000). Chomsky a exposé une version formelle – d'ailleurs contestée et plusieurs fois révisée – de la grammaire universelle (1957). La découverte scientifique la plus précieuse de Chomsky est probablement sa théorie des langages réguliers : il a démontré qu'il existait une correspondance entre l'algèbre et la syntaxe formelle : la langue est donc en principe un objet calculable, au moins sur un plan syntaxique (Chomsky et Schützenberger 1963).

Pour qu'une langue puisse être manipulée facilement par les ordinateurs, c'est-à-dire calculable, il faut qu'elle soit un langage régulier au sens de Chomsky : une sorte de code mathématique. Or les langues naturelles ne sont évidemment pas des langages réguliers. Les langages réguliers effectivement utilisés aujourd'hui sont des langages de programmation. Mais la « sémantique » des langages de programmation n'est autre que l'exécution des opérations qu'ils commandent. Aucun d'eux n'approche la capacité expressive d'une langue naturelle, qui permet de parler de tout et de rien et d'accomplir bien d'autres actes illocutoires que de donner des instructions à une machine. C'est pourquoi je me suis posé le problème qui suit : comment concevoir un langage régulier capable de tout dire – comme les langues naturelles – et dont les ordinateurs ne décoderaient pas seulement la structure syntaxique, mais aussi la sémantique ? Pour résoudre ce problème, je me suis souvenu des leçons de Saussure. Puisque les langues sont conventionnelles, rien n'interdit d'en construire une dont l'arrangement des signifiants soit de type « langage régulier ». Un langage régulier possède une syntaxe calculable, or la syntaxe régit les éléments signifiants de la langue, les phonèmes et leurs enchaînements, à plusieurs niveaux de complexité emboîtés. Puisqu'aussi bien les signifiants que les signifiés doivent être organisés par un système de différences, rien n'interdit non plus de donner – par convention – à ce langage régulier un système de différences des signifiés qui soit parallèle à celui des signifiants. Voici finalement le principe de résolution de mon problème : dans un langage régulier dont le système de différences des signifiés coïncide avec celui des signifiants, non seulement la syntaxe, mais également la sémantique est calculable. C'est précisément le cas d'IEML, qui est donc une langue à la sémantique calculable !

Si le principe général qui fonde la calculabilité sémantique d'IEML est facile à exposer, sa construction effective, qui m'a pris une vingtaine d'années, est évidemment plus complexe. Ce n'est pas ici le lieu de décrire exactement la grammaire et le dictionnaire d'IEML. Je me contenterai de donner les grandes lignes de sa structure linguistique. Le dictionnaire d'IEML contient des unités de sens élémentaires, les « sèmes » (ou morphèmes), qui sont organisés en niveaux. Les sèmes d'un certain niveau sont régulièrement construits avec des sèmes du niveau immédiatement inférieur et entretiennent avec eux une relation de composition sémantique. Les sèmes d'un même niveau s'interdéfinissent dans des paradigmes, qui sont des regroupements de sèmes de la même famille sémantique comme les couleurs, les rôles sociaux, les fonc-

tions techniques ou les compétences. Les sèmes d'un même paradigme ont une partie constante qui indique leur appartenance à une famille sémantique donnée et une partie variable qui décrit un système de différences. La grammaire d'IEML permet de construire des mots par combinaison de sèmes, des phrases par combinaison de mots et des super-phrases par combinaison de phrases. Le mot permet d'exprimer des catégories, des sujets, des topiques ou des thèmes. La phrase décrit des événements, des faits ou des objets complexes. Finalement, la super-phrase énonce des relations entre faits, des hypothèses causales, des théories ou des récits. Le mot, la phrase et la super-phrase en IEML correspondent, sur le plan de la sémantique linguistique, à la priméité, à la secondarité et à la tertiarité de Charles S. Peirce³⁴.

IEML se distingue des autres métalangages documentaires parce qu'il est univoque, mais sans imposer d'ontologie ou de classification a priori et parce que, malgré son univocité, il possède néanmoins la souplesse et les possibilités expressives d'une langue naturelle. L'indexation, la fouille et l'exploration des données utilisées par les sciences humaines seraient améliorées si IEML était utilisé comme système de métadonnées sémantiques. Aujourd'hui, les moteurs de recherche organisent leurs bases de données en indexant les documents à partir d'occurrences de chaînes de caractères, c'est-à-dire sur le plan purement morphologique ou signifiant de la langue. Par contraste, IEML propose un code sémantique où les variations de forme entre les mots correspondent à leurs variations de sens. Les mots-clés et les résumés en IEML ne sont pas des chaînes de caractères évoquant des sons, mais des adresses conceptuelles univoques, des points dans un système de coordonnées cohérent et calculable. En somme, IEML permet d'indexer les documents et les données par leur sémantique et cela à toutes les étapes de l'éditorialisation : écriture, publication, bibliothéconomie, lecture, commentaire, appréciation... Sur cette base, des algorithmes pourraient tisser automatiquement des liens sémantiques entre les données. Imaginons également la plus-value de connaissance qu'amènerait un moteur de recherche capable de distinguer – et de croiser ! – les sujets (mots), les faits (phrases) et les théories (super-phrases).

L'adoption de systèmes de coordonnées mathématiques et d'unités de mesure communes pour l'espace et le temps a joué un rôle capital dans les avancées scientifiques et techniques des trois derniers siècles, comme pour le perfec-

34. Voir *Écrits sur le signe* (1978) et *The Essential Peirce : Selected Philosophical Writings* (1992).

tionnement des possibilités d'action concertées de nos sociétés. Un système de coordonnées sémantique ferait franchir aux sciences humaines une étape décisive et augmenterait notre potentiel d'intelligence collective. Ayant passé un seuil de scientificité grâce au codage et au calcul du sens linguistique, et pour cela plus sûres d'elles-mêmes, les sciences humaines pourraient établir une relation mieux équilibrée avec les sciences de la nature et les spécialités du génie. Au sein des sciences humaines, les disciplines et théories distinctes pourraient dialoguer, voire collaborer plus facilement. Multipliée par les dimensions spatiale et temporelle, la dimension sémantique nous ouvrirait un nouvel univers, cohérent et computable, quoiqu'inépuisablement complexe. Projetées sur le fond du système de coordonnées sémantique, les données feraient émerger des formes et des mouvements précisément observables, des objets et des processus reproductibles. Et si les données enregistraient la mémoire d'une communauté – et particulièrement celle d'une communauté de chercheurs – l'image émergente représenterait une bonne part de ses activités cognitives. Dès lors, une boucle de rétroaction plus fluide entre les écosystèmes d'idées et les communautés qui les entretiennent nous rapprocherait de l'idéal d'une intelligence collective réflexive.

Vers une explicitation des jeux de communication et de référence

L'adoption d'un système de coordonnées sémantique nous fournirait la base d'une nouvelle transparence et d'une réflexivité critique mieux aiguisée, mais la production de sens ne se réduit pas à l'acte locutoire. C'est pourquoi il nous faut maintenant examiner la part proprement pragmatique de la signification et envisager les méthodes qui nous permettraient de l'éclairer en utilisant les nouvelles ressources du médium algorithmique. Commençons par rappeler ce que le niveau illocutoire ajoute au sens locutoire. Le sens locutoire est celui de l'énoncé, nous avons vu qu'il dépend des rapports entre signifiés et signifiants, des systèmes de différences au niveau du son comme au niveau du sens et enfin de la grammaire de la langue utilisée. À ce stade, la sémantique est interne à la langue. Par contraste, le sens illocutoire est celui de l'énonciation. Débordant la langue, il concerne un acte intentionnel dans un contexte spatio-temporel et social particulier. L'énonciation pointe vers une ou des références, elle parle de quelque chose. De plus, elle prend place dans une conversation ou plus

généralement dans un jeu social conventionnel, qui implique des règles, des rôles et des acteurs distincts. Enfin, le sens d'une énonciation dépend de la situation où elle intervient, avec en particulier son passé et la mémoire qu'en ont les interlocuteurs. On notera que la même langue (suffisante en principe pour comprendre le sens d'un énoncé) peut être utilisée dans des jeux sociaux très différents : par exemple dans une salle de danse, dans un tribunal, sur une scène de théâtre, dans une assemblée législative, dans une église, dans une salle de classe, en famille, etc. Pour comprendre le sens d'une énonciation, la connaissance de la langue ne suffit donc pas, il faut aussi connaître les règles du jeu auquel se livrent les interlocuteurs, la situation de la partie, le rôle de l'énonciateur, celui ou ceux des interlocuteurs auxquels il s'adresse, etc.

Nous nous intéressons ici aux usages savants de l'Internet, ce qui limite l'éventail des jeux de langage possibles. Mais même à l'intérieur du jeu social « recherche scientifique », il existe encore une foule de sous-jeux avec leurs règles spéciales, qui dépendent de la discipline, de l'institution, du lieu, du moment et qui se déroulent généralement en phases successives. Par exemple, la recherche subventionnée suit plusieurs étapes : publication d'un programme par un organisme public (avec ses finalités et ses critères), rédaction de demandes de subvention par les chercheurs, constitution de jury ou de comités d'évaluation, obtention d'une subvention par certains candidats, mise en œuvre des recherches, publications, etc. L'obtention d'un doctorat, le déroulement d'un postdoc, la publication d'un numéro thématique d'une revue, l'organisation d'un colloque, l'attribution d'un prix scientifique ou l'élection à une académie obéissent de la même manière à des phases bien réglées. La tenue d'un blogue de chercheur ou la communication dans des médias sociaux comme Research Gate, Academia, CiteULike et autres ne sont pas moins ritualisées et organisées en cycles. Chacun de ces jeux scientifiques a ses règles générales, mais celles-ci varient avec les pays, les programmes gouvernementaux, les revues scientifiques, les associations qui organisent les colloques, les laboratoires où se concoctent les thèses, les médias sociaux où s'échangent les idées, les articles et les données. Chacun des actes illocutoires accomplis par les chercheurs (conseil à un jeune chercheur, évaluation, publication, discussion, abonnement, lecture, citation...) prend sens en fonction du type de jeu auquel il participe, de la phase du jeu où il intervient et de règles particulières dépendant des circonstances. En fin de compte, l'ensemble varié de ces jeux savants s'entrecroise pour former l'activité scientifique dans son ensemble. Gardons en tête que notre but est de mettre en transparence et de formali-

ser autant que possible les actes illocutoires des chercheurs. C'est pourquoi il serait fort utile de disposer d'une liste ouverte des jeux scientifiques en usage, de leurs phases, des types d'actes illocutoires qui les composent et des critères de validité de ces actes.

S'agit-il d'une demande de subvention, d'un article soumis à une revue, d'un compte rendu critique, d'une discussion dans un comité de rédaction, d'un live tweet rapportant un débat dans un colloque... ? Malgré leur variété et la diversité des jeux dans lesquels ils interviennent, les actes illocutoires possèdent le plus souvent deux dimensions capitales : la communication et la référence. La communication s'organise entre deux pôles fondamentaux, le destinataire et le destinataire, dont chacun peut d'ailleurs être pluriel et qui se déclinent en une multitude de rôles complémentaires (comme évaluateur et évalué dans un programme de subvention) ou symétriques (comme deux auteurs dans un débat public). Parmi ces rôles de communication mentionnons : éditeur, auteur, lecteur, curateur, passeur, commentateur, cochercheur, enseignant, étudiant, citant, cité, etc. Pour chaque acte illocutoire on essaiera donc de formaliser non seulement l'identité des interlocuteurs, leur place (échangeable) de destinataire et de destinataire, mais également le rôle qu'ils assument dans le processus général d'éditorialisation, conçu ici comme conversation scientifique au sens le plus large.

L'acte de référence

Intéressons-nous maintenant à la dimension référentielle de l'illocution. Dans les usages savants de l'Internet, l'acte de référence consiste essentiellement à pointer vers un ensemble de données. Toujours selon notre finalité de mise en transparence formelle, la référence peut être décomposée en deux parties : (1) l'adressage des données et (2) l'explicitation de la relation entre les données adressées et les métadonnées sémantiques qui servent à les qualifier. Commençons par le premier point. La pratique de la citation d'une œuvre savante – article ou livre – est ancienne dans la communauté scientifique et a donné lieu à une codification bien établie. On s'accorde pour citer précisément le ou les auteurs, le titre de l'œuvre, l'éditeur (journal ou maison d'édition), éventuellement les personnes ayant dirigé la publication (sous la direction de), le lieu de publication, la date, voire la ou les pages précises dans un ouvrage imprimé. Ces règles de citation ont été établies à l'âge de

l'imprimerie afin de permettre de retrouver les textes dans une bibliothèque, mais elles sont en voie de réactualisation dans l'espace numérique. On utilise maintenant de préférence un DOI (*Digital Object Identifier*) pour référencer les œuvres ou les ensembles de données (*datasets*), et un identifiant ORCID (*Open Research and Contributor ID*) pour désigner les auteurs de manière sûre. Des outils open source tels que Zotero (développé à l'initiative de Dan Cohen de l'Université George Mason) permettent de gérer et d'échanger de manière pratique des bibliographies et des ensembles de références scientifiques en ligne. L'outil d'écriture savante Stylo (logiciel développé à l'initiative de Marcello Vitali-Rosati à l'Université de Montréal) est construit à partir d'un réseau de modules open source qui génère un texte interopérable et facilement réutilisable. Surtout, Stylo prend acte de la transformation de la référence scientifique, qui passe d'un exercice de style graphique hérité de l'imprimerie à la mise en œuvre de pointeurs capables d'aller chercher automatiquement les données dans l'espace numérique. En fin de compte, à supposer que les revues savantes se transforment en sites Web et que le contenu des bibliothèques appartenant au domaine public soit numérisé et mis en accès libre, l'adressage des œuvres ou des données se ramène à une URL.

De même qu'il existe des rôles de communication (auteur, lecteur, évaluateur, commentateur, curateur, etc.), il existe des rôles de citation ou de référence, qu'il faut pouvoir distinguer. En effet, ce n'est pas la même chose de pointer vers des sources primaires – le corpus sur lequel travaille le chercheur –, de citer des travaux antérieurs sur le même corpus, de mentionner un article pour reconnaître une dette intellectuelle ou bien encore pour marquer sa désapprobation³⁵. Or les mesures de l'autorité aujourd'hui en usage se contentent de compter le nombre de citations sans prendre en compte le mode qualitatif de référence.

Il est temps de s'interroger sur la relation entre les données (représentées par des URL) et les métadonnées sémantiques. Mais qu'est-ce qu'une métadonnée, dans la pratique ? Ce fut longtemps la description codée d'un document ou d'un artefact pour faciliter la recherche ultérieure dans un centre de documentation, une bibliothèque ou un musée. Cette description alimentait un système de fiches, le plus souvent en carton, permettant de retrouver le do-

35. Evelyne Broudoux (2017) a récemment publié un article pour faire le point sur cette question, voir en ligne.

cument sur une étagère (grâce à une cote) en fonction non seulement de l’auteur et du titre, mais aussi de la discipline, du sujet et du contenu du document représentés par des mots-clés normalisés et organisés en thésaurii. Ces systèmes d’organisation de la mémoire, évidemment conventionnels, sont fort variables. Ils dépendaient – et dépendent toujours – des institutions et des pays. Tout allait bien tant que les langages documentaires et les systèmes d’indexation s’appliquaient à des collections contenues par un bâtiment ou gérées par une seule institution.

Avec la numérisation des archives et l’interconnexion des bases de données, l’incompatibilité et la rigidité des systèmes d’indexation hérités de l’ère de l’imprimerie devient criante. La gestion d’un système de métadonnées pour la gestion d’un corpus documentaire et la constitution et le suivi de collections ont longtemps été une activité réservée à des professionnels des sciences de l’information. Mais la pratique de la curation de données a maintenant largement débordé le cercle étroit des documentalistes et des bibliothécaires. Des millions de non-spécialistes organisent des collections de données en utilisant des applications spécialisés comme Pocket, Scoopit, Diigo, etc.³⁶ De plus, la plupart des blogues et des médias sociaux permettent à leurs utilisateurs d’ajouter des tags ou des hashtags lorsqu’ils publient un message quelconque. Nous pouvons utiliser ces balises sémantiques pour nous définir (intérêts, compétences...) ou pour pointer vers les publications d’autres usagers, comme lorsque nous incluons un lien (une URL) et un ou plusieurs hashtags (donc des métadonnées sémantiques) dans un tweet afin de qualifier le document signalé. Les pratiques contemporaines de curation distribuée génèrent plusieurs problèmes. Premièrement, il existe un grand nombre de tags différents, y compris dans la même langue, pour exprimer un concept identique. Deuxièmement, les tags exprimant le même sujet sont différents selon les langues. Troisièmement, le même tag peut vouloir dire plusieurs choses complètement distinctes (comme par exemple #GPL qui peut signifier *general public licence...* ou *global player league*). Enfin, par opposition aux méthodes professionnelles des spécialistes de la documentation, on ne connaît pas la procédure qui a déterminé un usager à choisir telle indexation sémantique d’un document plutôt qu’une autre. La communauté schema.org soutenu par Google et Microsoft, tente de répondre à certains de ces pro-

36. Voir mon article de blogue sur la curation collaborative de données (Lévy 2016a) et mon article sur les compétences nécessaires à la pratique éclairée de la curation (Lévy 2016c).

blèmes en proposant un système d'indexation normalisé du Web. Mais son vocabulaire – orienté par des objectifs commerciaux – reste limité et il est certainement insuffisant pour baliser la recherche en sciences humaines. Il est d'autre part monolingue anglais et se présente sous la forme d'une ontologie relativement rigide. L'adoption des USL (*uniform semantic locators*) codés en IEML mettrait entre les mains du public des « super-hashtags » qui résoudraient les ambiguïtés, élargiraient les possibilités d'expression, soutiendraient la traduction multilingue automatique et ouvriraient l'Internet à une véritable fouille sémantique fondé sur un calcul des proximités de sens. Mais IEML ne peut résoudre les problèmes liés à la relation entre le contenu de l'URL et l'USL qui le catégorise. C'est toujours une opération de curation particulière, un acte de référence, qui décide que tel ensemble de données doit être étiqueté ainsi.

Distinguons ici deux types de métadonnées : les métadonnées « objectives » (auteur, date, langue du document, droits, format de fichier, etc.) qui sont en gros celles du Dublin core et les métadonnées « sémantiques » qui détaillent le sujet, le contenu et les angles d'analyse possibles d'un document ou d'un dataset. Prenant les métadonnées objectives pour acquises, concentrons notre analyse sur les métadonnées sémantiques, qui dépendent de l'interprétation d'un curateur. Aujourd'hui, une recherche dans le domaine des humanités numériques passe inévitablement par la construction et l'aménagement d'une base de données. Les principales étapes de cette construction sont : (1) la sélection d'un sous-ensemble pertinent dans la grande base de données virtuelle de l'Internet ; (2) l'édition et le nettoyage des données sélectionnées ; (3) l'enrichissement de la base de données au moyen de métadonnées soigneusement déterminées, qui vont orienter le calcul des algorithmes et autoriser les réponses aux questions de recherche ; (4) le choix des visualisations ou du rendu des résultats. Les métadonnées sémantiques jouent donc un rôle structurant dans les méthodologies de recherche et dépendent étroitement de la perspective des chercheurs. C'est pourquoi il est impossible de déterminer à l'avance – et une bonne fois pour toutes – les points de vue, les problèmes ou les théories qui vont éclairer ou utiliser un ensemble de données. Même si l'indexation sémantique des documents réalisée par des professionnels selon une approche normalisée a été historiquement fort utile, nous devons poser en principe la liberté de la catégorisation sémantique. Il n'y aura pas de méthode unique pour catégoriser les documents ou les datasets puisque cela reviendrait à brider la créativité herméneutique des savants et à stéri-

liser la recherche en sciences humaines. Concernant les données communes, on acceptera la multiplicité des points de vue et des catégorisations sémantiques, d'autant plus que si les catégories sont exprimées en IEML, elles sont commensurables et leurs relations se calculent automatiquement. Mais afin de rendre l'opération de référence transparente, on identifiera la méthode de catégorisation utilisée – le lien entre URL et USL – et on privilégiera les méthodes formelles et reproductibles. Si la catégorisation sémantique du document est exécutée par un algorithme – par exemple un algorithme qui génère automatiquement un résumé IEML du document – celui-ci devra être ouvert et en accès libre.

Les niveaux de référence et la question de la vérité

Pour clore cette analyse, je voudrais distinguer entre trois niveaux de la relation entre les locutions (métadonnées sémantiques) et leur références (données). Ces niveaux sont déterminés par l'échelle de complexité grammaticale de la locution : mot, phrase ou super-phrase. S'il s'agit simplement d'indiquer le sujet dont parle un document, on utilise comme métadonnée sémantique un mot en IEML, c'est-à-dire l'équivalent d'un mot-clé ou d'un hashtag. À ce niveau, qui correspond à la priméité de Peirce, la catégorisation est de l'ordre de la sensation. L'auteur de la référence n'affirme pas une vérité, mais plutôt une pertinence : la réponse d'un objet à un intérêt, ou l'éveil d'une attention particulière devant un phénomène. Cette saillance originaire du champ sémantique possède une face subjective – un point de vue – et une face objective – une qualité du document – les deux aspects émergeant simultanément. Dans la pratique, beaucoup de métadonnées sémantiques de type « mot-clé » seront mis en œuvre – peut-être successivement – par les auteurs d'un document, par des curateurs ou des éditeurs et par des algorithmes de catégorisation. L'agent et la procédure de catégorisation devront être explicités autant que possible.

Si la métadonnée sémantique évoque un événement ou un objet complexe, on utilise une phrase en IEML. Ce niveau grammatical correspond à la secondarité de Peirce et au monde de « tout ce qui arrive » du *Tractatus Logico-Philosophicus* de Wittgenstein. La phrase IEML est un ensemble de relations grammaticales entre des mots. Considérée comme une proposition logique, elle décrit un état de choses, c'est-à-dire un ensemble de relations

entre saillances sémantiques dans le document ou le dataset. Il est possible en principe de déterminer la vérité de la proposition par l'observation rigoureuse de l'état de choses (le document, les données, le corpus...). Dans notre optique de mise en transparence des opérations de connaissance, il faudra évidemment expliciter les procédures ou les appareils permettant de vérifier les propositions logiques, qu'il s'agisse de capteurs, d'instruments de mesure, d'algorithmes statistiques ou autres. Une proposition vraie est un fait. Grâce à des règles relevant de la logique des prédicats (du type : « tous les X sont des Y », « aucun A n'est un B », etc.) de nouveaux faits peuvent être déduits automatiquement des faits observés. (Ce que l'on appelle aujourd'hui une ontologie dans le monde des *linked data* n'est autre qu'un ensemble cohérent de telles règles pour décrire un domaine selon une certaine perspective pratique). Enfin, la logique des propositions permet de déduire des propositions complexes vraies à partir des faits. En somme, une fois vérifiées par une procédure explicite et reproductible, les métadonnées de type « phrase » permettent d'alimenter le raisonnement automatique de moteurs d'inférence.

Par opposition avec les états de choses ou les faits, qui sont observables, on ne peut observer une théorie : elle est toujours hypothétique. Selon l'épistémologie de Popper, ou bien une théorie a déjà été réfutée, ou bien elle ne l'a pas encore été. Mais l'entreprise humaine de construction de sens a besoin de théories – aussi inobservables et invérifiables soient-elles – pour réunir les faits en un tout cohérent, pour donner une interprétation causale à leurs rapports, permettre tant bien que mal la prévision des événements à venir et même pour coordonner les croyances et les actes des communautés. Dans certaines sciences humaines, par exemple en histoire, on parlera plutôt de récit que de théories. On rapprochera les remarques qui précèdent des idées de Hume, selon qui la causalité, bien que nécessaire à notre compréhension du monde, n'est pas un trait de la réalité objective (une qualité primaire), mais une construction de l'esprit humain (une qualité secondaire)³⁷. Reprenant la critique humienne de la causalité dans sa construction métaphysique, Kant fait de la causalité une des catégories a priori de notre entendement, comme l'espace et le temps sont des formes a priori de notre sensibilité³⁸. De nouveau, la causalité est une structure du sujet, un aspect indispensable

37. Voir par exemple *Traité sur l'entendement humain*, 1.3.14.12.

38. Voir dans la *Critique de la raison pure*, l'exposition des concepts purs de l'entendement qui prend place dans l'analytique des concepts.

de notre manière de connaître. Nietzsche³⁹ radicalise la critique de la causalité accomplie par la tradition philosophique des Lumières en considérant les causes, les effets et les théories ou les récits qui les entrelacent comme des fictions utiles. Dans le contexte philosophique qui vient d’être rappelé, lorsque, dans l’opération de référence, la locution exprime un ensemble de relations entre des faits, c’est-à-dire une théorie ou un récit, on utilisera une super-phrased en IEMML. La super-phrased appartient au domaine de la tercedité de Peirce ou à la classe des conjectures selon Popper et le lien référentiel qu’elle entretient avec les données n’est, à proprement parler, pas vérifiable. On peut néanmoins évaluer la validité d’une théorie selon divers critères et notamment se demander si les données la corroborent, si elle se contredit, si elle minimise le nombre de concepts utilisés (le fameux rasoir d’Occam) et si elle rend compte d’un maximum de faits (pouvoir explicatif).

Résumons notre analyse des trois niveaux de la référence. Même à nous en tenir à des jeux de langages scientifiques, le domaine proprement logique du vrai et du faux est limité à la relation entre une proposition et un état de chose. Dans notre description d’une pratique scientifique idéale, la référence logique correspond à une phrase IEMML connectée à un ensemble de données et comportant l’explicitation de la procédure de vérification. En revanche, la simple catégorisation des thèmes présents dans un dataset n’est jamais fautive, puisqu’elle correspond à une saillance originelle ou à l’évidence d’une pertinence. À ce niveau comme ailleurs, l’arbitraire sera contrôlé par l’explicitation de la procédure de catégorisation. Quant à l’énoncé d’une théorie – qui rend compte de l’enchaînement des faits –, il n’est jamais vrai, puisqu’il ne s’agit par définition que d’une hypothèse. Il existe néanmoins plusieurs critères pour apprécier la valeur épistémique d’une théorie ou d’un récit. Dans un jeu de langage scientifique, les concepts sont jugés sur leur pertinence, les faits sur leur vérité, les théories sur leur élégance.

Est-il besoin de souligner qu’il existe bien d’autres jeux de langages acceptables que ceux de l’épistémologie scientifique la plus stricte ? Une grande partie des meilleurs travaux dans les disciplines des humanités relèvent largement de la rhétorique, c’est-à-dire de l’immense domaine des faits probables, des récits vraisemblables et des décisions souhaitables. Dans leur exercice même, les jeux de langage littéraires répondent plutôt à des contraintes for-

39. Par exemple dans le paragraphe 112 du *Gai savoir*, mais aussi dans bien d’autres passages de son œuvre.

nelles sur la locution, selon une grande variété de genres et de styles, et leurs récits peuvent – par exemple – maximiser le nombre de concepts ou d’actants et suggérer plusieurs systèmes de relations entre les événements rapportés. Centré sur une mise en récit des nouvelles, le journalisme représente encore un autre jeu de langage que celui de la science ou de la littérature, un jeu dont les responsabilités politiques sont évidentes. Les données, les méthodes, les outils d’analyse et de visualisation du meilleur journalisme étant très proches de celles des humanités numériques, l’industrie des nouvelles gagnerait à se rapprocher de la pratique scientifique et à rendre ses opérations de référence les plus transparentes possible.

Observer l’émergence de l’autorité et de la croyance

Abordons pour finir la dimension perlocutoire de la production de sens à savoir, dans le cas de la communauté scientifique, la production d’autorité. Il est important de comprendre comment se fabrique l’autorité dans la communauté scientifique parce que cette autorité détermine une bonne partie des croyances – et par voie de conséquence des pratiques – qui ont cours dans nos sociétés. Dans le cas des sciences de l’Homme, ces croyances concernent l’humanité en général, l’histoire, les phénomènes sociaux, la vie de la cité, les goûts esthétiques et littéraires, etc. Les grilles d’analyse conceptuelles, la détermination des objets pertinents et des sujets importants, la vérité des faits historiques et sociaux, les théories et les récits qui rendent compte des phénomènes politiques et culturels, tout cela est largement produit aujourd’hui par les chercheurs en sciences humaines, puis enseigné à la majorité de la population à l’école, absorbé à l’université avec un grain de sel critique par les futurs cadres, propagé par les journalistes influents et les hommes politiques, mis en œuvre par les hauts fonctionnaires et les dirigeants d’entreprise. C’est dire le lien inextricable qui noue les dimensions épistémiques et politiques... tout comme la nécessité de leur autonomie respective. Une science simplement soumise au pouvoir politique perdrait sa crédibilité avec son autorité. Symétriquement, une cité dirigée par les savants ferait fi de la liberté des citoyens, de la diversité de leurs choix éthiques comme de la multiplicité des sphères d’existence. Il reste qu’une grande partie de nos repères cognitifs sont donnés par les sciences humaines, et cela d’autant plus que les sociétés s’alphabétisent et se scolarisent. Certaines parties de la population acceptent ou rejettent telle ou telle croyance – avec les pratiques et

les politiques qui s'ensuivent – mais elles le font le plus souvent au sein de paradigmes ou d'univers de pensée communs, dans lesquels les oppositions se répondent et font système. Or ces paradigmes ont le plus souvent été élaborés par la communauté des savants, comme ils l'étaient avant l'imprimerie par les clercs et les scribes.

De nos jours, l'autorité scientifique, et la croyance qui en présente la face subjective émergent d'une multitude de jugements individuels et collectifs qui portent principalement sur les auteurs et leurs publications. Les publications peuvent être considérées comme la dimension objective et réifiée – le noyau solide – de la conversation scientifique. Souvenons-nous que les premières revues savantes du XVII^e siècle se contentaient d'imprimer les plus copiées des correspondances manuscrites entre philosophes. Si, à l'origine, la publication servait à mettre les conversations entre savants à la portée de la communauté, elle est ensuite devenue le support normal et officiel de ces conversations. Dans le prolongement de son origine dialogique, la littérature scientifique ne comprend pas seulement des articles publiés dans des revues scientifiques imprimées et examinés par les pairs selon les normes classiques du XX^e siècle. Il faut l'entendre en un sens plus large. Elle inclut aussi les publications en ligne sur le modèle de plus en plus suivi de arXiv.org ou d'autres sites de publication ouverte qui accélèrent la circulation des idées en permettant la publication avec une évaluation initiale minimale et qui réservent l'évaluation approfondie des articles déjà publiés à une discussion ultérieure par la communauté des chercheurs concernés. On peut également considérer comme une nouvelle forme de publication la construction de bases de données qui se fédèrent progressivement en une immense réserve de communs scientifiques. Les réseaux de conversation scientifique passent également aujourd'hui par les blogues de chercheurs, les médias sociaux basés sur l'archivage collectif et le partage d'articles, sans oublier les médias sociaux généralistes et les communications qui se font dans les soutenances de thèse, colloques, séminaires, camps d'été et ateliers de toutes sortes.

Quant aux auteurs des publications qui viennent d'être d'évoquées, ils appartiennent à la communauté scientifique par l'intermédiaire d'universités, instituts, centres de recherche, laboratoires, académies, équipes, réseaux, sociétés savantes, comités de rédaction et conseils scientifiques de toutes sortes. Ces institutions, le plus souvent administrées par les chercheurs eux-mêmes (en principe selon les normes de la liberté académique et de l'autonomie universitaire), embauchent leur personnel et équipent la recherche. Le finance-

ment, public ou privé, de ces institutions et de leurs programmes de recherche constitue bien entendu une condition *sine qua non* de leur fonctionnement, d'où le lien inévitable de la communauté académique avec les gouvernements ou avec des entreprises privées investissant dans la recherche. Avec le nombre et la qualité de ses publications et des citations dont il est l'objet, la position institutionnelle d'un chercheur forme évidemment une part importante de son autorité.

Au cœur des procédures de légitimation et de construction de l'autorité se trouvent des pratiques argumentatives, qui sont mises en œuvre aussi bien dans les publications elles-mêmes que dans les décisions de publication et dans les procédures qui décident de l'avancement institutionnel des chercheurs. L'argumentation vise à influencer le jugement des pairs (et des sources de financement) en faveur de personnes, de programmes de recherche, de théories ou de paradigmes. La palette des formes argumentatives est assez large. Elles comprennent notamment : la mise en valeur de l'originalité, la citation de publications scientifiques et d'auteurs déjà reconnus, la référence précise à des corpus, la description de procédures méthodologiques reproductibles, la visualisation de données ou de simulations informatiques, la promesse d'applications (a) utiles selon des normes politiques ou (b) rémunératrices sur le marché. Les arguments prennent souvent la forme du raisonnement hypothético-déductif ou de l'induction statistique, mais – dans la grande tradition de la rhétorique – ils s'appuient aussi sur des « lieux communs » politiques et culturels afin de promouvoir la vraisemblance de leurs conclusions, tout cela sur le fond des savoir-faire implicites partagés par les chercheurs du fait de leur formation et de leurs pratiques. Ajoutons que les styles argumentatifs admis peuvent être assez différents d'une discipline à l'autre.

La communauté scientifique est officiellement régie par la cooptation et le jugement par les pairs, ce qui souligne sa dimension autoréférentielle et autopoïétique. Elle ne reçoit pas ses vérités d'une instance extérieure ou d'une autorité transcendante. Certes, ses membres appartiennent à leur époque, ils sont traversés par les contradictions politiques de leur société, ils reçoivent leurs problèmes de la culture où ils évoluent et visent à améliorer le sort de leurs contemporains. Mais la communauté scientifique se distingue des autres institutions sociales par son autonomie cognitive de principe. Au cours d'un processus collectif, évolutif et circulaire de validation, les jugements qui déterminent le sort des institutions scientifiques, de leurs membres et de leurs publications sont déterminés en retour par le monde académique, les savants

et leur littérature scientifique. Et ces jugements, comme les critères d'évaluation qui les fondent, ne portent pas seulement sur la vérité et l'erreur, mais aussi sur l'original ou le redondant, le laborieux ou l'élégant, le possible ou l'impossible, le fécond ou l'infécond, le prometteur ou le sans avenir, le mis à jour ou le retardataire... Loin d'être infaillibles, les consensus de la communauté scientifique s'apparentent aux moments d'un apprentissage collectif ouvert ou d'un voyage d'exploration. Les inerties bureaucratiques ou conservatrices, les phénomènes de clique et les emballements grégaires sont monnaie courante. La connaissance scientifique avance par essais et erreurs, et cette marche tâtonnante inclut parfois la réalisation difficile que certaines décisions ont mené à des impasses et qu'il faut revenir en arrière. Dans le cas des sciences humaines, ce processus complexe de production de connaissances contribue à la construction des phénomènes sociohistoriques pertinents et des théories explicatives crédibles sur la scène culturelle : de nouveau, en dessinant l'arrière-plan des processus cognitifs qui ont cours dans une société, il joue un rôle politique considérable.

Le plongement de la communauté scientifique dans l'espace numérique a eu des effets indéniablement positifs. J'ai déjà souligné la facilitation de l'accès aux publications, aux corpus de donnée, aux outils algorithmiques et à la puissance de calcul, auxquels il faut ajouter la multiplication des contacts transversaux et des possibilités de collaboration. Mais on observe aussi un aplatissage unidimensionnel des formes d'évaluation – et donc de production d'autorité – par les grandes plateformes, Google en tête. Pour le dire simplement, afin d'être consacré, un auteur ou une publication doit arriver dans la première page des résultats d'une requête, qu'il s'agisse du moteur de recherche pour le grand public ou de Google Scholar, spécialisé dans les publications scientifiques. Certes, Google n'est pas le seul responsable de l'uniformisation quantitative de la production d'autorité. Le classement de Shangai des établissements universitaires ou le calcul du facteur d'impact (nombre de citations divisé par le nombre d'articles publiés depuis deux ans) des revues scientifiques ne peuvent être attribués aux géants du numérique. La montée des méthodes de gestion quantitatives de la science accompagne l'augmentation des dépenses de recherche, la multiplication du nombre de chercheurs, le gonflement des effectifs d'étudiants dans l'enseignement supérieur, la concurrence internationale exacerbée entre centres de recherche et d'enseignement et, en somme, la croissance, l'industrialisation et la mondialisation du secteur. Mais une fois reconnu le contexte général d'un management

de la recherche de plus en plus quantitatif et uniformisant, la distorsion introduite par Google dans le processus d'éditorialisation de la littérature scientifique doit encore nous inquiéter, et cela pour plusieurs raisons. D'abord, ses finalités sont avant tout marchandes. Cela n'est pas un mal en soi, mais il n'y a pas de raison pour que l'autorité scientifique soit mesurée par une entreprise dont les finalités, les valeurs et les modes d'arbitrage ne sont justement pas celles de la science. Ensuite, ses algorithmes sont opaques, protégés par le secret commercial, et donc aux antipodes de la démarche transparente et critique qui devrait réguler l'autorité scientifique. Enfin, le rôle unique de Google dans la communication mondiale en fait le point de passage obligé et la référence par défaut des personnes à la recherche d'information (y compris dans le monde académique!) et cette position de monopole n'est équilibrée par aucun contrepoids sérieux.

Si l'autorité scientifique est calculée par Google, cela signifie que nos croyances sont et seront de plus en plus déterminées par ses algorithmes. La plateforme utilise les résultats du jeu complexe de la validation des connaissances dans la communauté scientifique et – non sans biais de toutes sortes – les projette violemment sur la première page de résultats de Google Scholar. Au-delà du cas Google, voulons-nous transférer aux grandes plateformes commerciales, en plus de leur domination économique et de leur puissance politique de fait, le pouvoir symbolique que l'Église possédait naguère ou que l'Université exerce aujourd'hui? Or le problème ne vient pas du calcul par des algorithmes, mais du fait que ces algorithmes sont opaques, qu'ils sont alimentés par des données grossièrement quantitatives et qu'ils ne permettent pas (et d'ailleurs ne visent pas) une transcendance réflexive de l'intelligence collective des chercheurs. Entendons-nous bien, Google rend d'énormes services et ce sont notamment ses réalisations qui nous permettent d'imaginer les étapes suivantes de l'évolution du médium algorithmique : si cela – que fait Google et qui était encore impensable à la fin du XX^e siècle – est désormais possible, imaginons ce qui pourrait l'être dans une ou deux générations, à condition que nous avancions fermement sur la voie de l'intelligence collective réflexive!

Après avoir analysé le processus de production d'autorité dans la communauté scientifique, ses enjeux politiques et son détournement par Google (ou par les plateformes qui vont lui succéder), nous pouvons maintenant reprendre le cours de notre exposé sur la mise en transparence des opérations de production de connaissance dans l'espace numérique. En principe, les choses

sont simples : l'autorité n'est autre que l'effet perlocutoire des actes locutoires de communication et de référence que j'ai analysé plus haut. Cet effet perlocutoire agrège les opérations de lecture et d'écriture qui peuvent se détailler en évaluation, publication, citation, lecture commentée, dialogue, etc. L'état de la connaissance à un moment donné forme l'environnement commun où se transige la communication scientifique. Selon le modèle de communication stigmergique esquissé plus haut, l'autorité et la croyance résultent d'une multitude distribuée d'actes illocutoires de la part des chercheurs (et de leurs lecteurs) et elle guide en retour les jeux de communication et de référence auxquels ils se livrent. Le calcul des effets perlocutoire des actes des chercheurs ne devrait pas poser de problème technique particulier puisque les médias sociaux contemporains et les grandes plateformes sont déjà capables d'analyser avec une grande précision tous les aspects de notre engagement en ligne.

J'ai fait miroiter dans ce texte une machine abstraite de lecture-écriture collaborative capable de générer de l'intelligence collective réflexive à partir de l'engagement de ses utilisateurs-chercheurs. Le niveau locutoire de cette machine est tissé par IEML, une langue à la sémantique calculable. C'est un système de coordonnées de l'univers du sens, une sorte d'échiquier cosmique dont chaque case est un concept distinct. Les adresses de la sphère sémantique (les USLs) servent à décrire les données, mais aussi les règles et les modalités de la référence aux données ainsi que les jeux et les rôles de communication entre chercheurs. Les chercheurs, les données et les actes de production de sens entrent en scène au niveau illocutoire. À cet étage, les chercheurs sont adressés par un identifiant de type ORCID et les données par des URLs. Nos savants communiquent en endossant divers rôles selon une variété de jeux de langage et, ce faisant, ils pointent vers les données qu'ils consultent, produisent, analysent et catégorisent. Enfin, le niveau perlocutoire enregistre les conséquences sociales des actes de lecture et d'écriture accomplis au niveau illocutoire. À ce dernier étage de notre architecture virtuelle, les chercheurs composent des fonctions d'autorité à partir de variables telles que : la pondération des types d'engagements mesurés (publication, citation, lecture, etc.), les ensembles de données qui importent, les jeux de communication et de référence pertinents, les communautés de chercheurs concernés... Sur le fond du système de coordonnées sémantique commun, ces fonctions d'autorité génèrent des paysages hiérarchisés de données, de travaux, de chercheurs et de conversations. Chaque fonction d'autorité distincte émet un reflet différent

de l'intelligence collective. La production d'autorité scientifique n'est donc ni écrasée ni masquée par des algorithmes opaques, uniformes et fondés sur des données purement quantitatives. Au contraire, les algorithmes sont transparents, ils sont aussi variés que les fonctions d'autorité imaginables et ils sont fondés non seulement sur des comptages et des mesures d'engagement, mais aussi sur des données qualitatives : sémantique, règles, rôles, communautés... Dès lors, une boucle réflexive peut s'établir entre les actes de production de sens et l'exploration reproductible des conséquences de ces actes, dans le va-et-vient entre l'activité scientifique et la découverte progressive de ses effets d'autorité et de croyance.

Conclusion : la mission des humanités numériques

Ce petit manifeste ne propose ni standard, ni protocole, ni format de fichier, ni application, ni plateforme. Ma seule recommandation un peu technique est l'utilisation d'IEMML, une langue compatible a priori avec n'importe quel item de la liste qui précède. La machine abstraite de lecture-écriture ou, si l'on préfère, l'architecture d'intelligence collective réflexive évoquée dans ce texte est d'abord une idée orientatrice, une direction d'évolution qui s'accommodera de la multiplicité et de l'évolution des outils. Il n'est pas nécessaire que toutes ses fonctions soient implantées pour commencer à la mettre en œuvre. On peut envisager une approche progressive, pas à pas, fondée sur l'expérimentation limitée et le test.

Le modèle systémique et multicouche de la production de sens qui est au cœur de cette architecture a été expliqué en s'appuyant sur l'exemple de la communauté des chercheurs en sciences humaines. Mais ce modèle peut s'appliquer aussi bien à d'autres types de communautés et de jeux de langages. J'ai privilégié les Humanités parce que la production de sens est leur objet même et que leur méthode est précisément la pensée critique : un usage dialogique et réflexif de la raison. Les humanités jouent d'autre part un rôle politique capital puisqu'elles éduquent les dirigeants de nos sociétés, déterminent les phénomènes sociaux dignes d'attention, garantissent les réalités historiques et enfin parce qu'elles ancrent la confiance des citoyens dans certaines théories explicatives et certains récits plutôt que d'autres. En œuvrant à augmenter les capacités cognitives, la coordination, la communication et la réflexivité de la communauté de recherche en sciences humaines on travaille

du même coup au développement de la pensée critique pour la cité dans son ensemble.

Mais l'architecture virtuelle proposée ne prend pas la communauté des chercheurs en sciences humaines comme un donné intangible et anhistorique auquel il s'agirait seulement d'offrir de nouvelles méthodes : elle en dessine une métamorphose possible. Souvenons-nous de ce qui s'est passé depuis la Renaissance. En quelques siècles, la pratique de l'écriture et de la lecture s'est étendue au-delà de la communauté des clercs et des lettrés pour englober la majorité de la population. Prenant appui sur la communication imprimée, une communauté scientifique indépendante de l'Église a émergé, pourvue de règles distinctes de production de l'autorité. Au sein même du christianisme latin, la Réforme a accompli une révolution qui relève plus de l'ecclésiologie que de la doctrine théologique. En effet, ce sont d'abord les formes de la communauté qui ont changé, avec notamment le rejet de la monarchie romaine, la fin de la distinction formelle entre les clercs et les fidèles, l'abandon du monachisme (autre forme de séparation) et la liberté de lecture et d'interprétation du texte sacré, désormais accessible en langues vernaculaires. Puis des perspectives de salut orientées vers ce monde-ci – droits humains, nouvelles idéologies à prétentions scientifiques – ont fait concurrence aux religions révélées... À son tour, l'émergence du médium algorithmique aura probablement des effets à long terme sur la structure des communautés et sur les formes de leurs croyances. Comment l'Université va-t-elle évoluer ? Souvenons-nous que cette institution a été créée par l'Église. D'ailleurs, avant les temps modernes, quelles que soient les cultures, les institutions de haut savoir étaient d'essence religieuse. Dans la continuité de leur passé, les institutions académiques contemporaines – quoiqu'elles soient déchristianisées et ne se soutiennent plus de la révélation – ont gardé une fonction de production et de diffusion de la croyance. Supposons maintenant que les usages savants de l'Internet, avec leurs principes d'ouverture, de transparence et d'intelligence collective réflexive se répandent. Ces usages ne sont évidemment pas limités aux membres de la communauté académique. Chaque collectivité (qu'il s'agisse d'une ville, d'une association, d'une entreprise...) produit des données et aura bientôt les moyens techniques de les analyser. Les humanités numériques ont justement pour vocation d'accompagner ces collectivités sur leur chemin d'apprentissage. Par conséquent, il est probable que les bonnes pratiques de la recherche en sciences humaines – avec l'autorité qui en émane – vont se répandre bien au-delà des limites contemporaines de la communauté

académique, un peu comme s'est estompée la distinction entre les clercs et les fidèles au moment de la Réforme... Et pour filer la métaphore, cette fois-ci, le texte de référence n'est plus la Bible, le Coran ou quelque canon sacré mais l'hypertexte en expansion dans lequel l'humanité réfléchit, avec son existence, sa création de sens.

Dans le schéma qui vient d'être évoqué, la communauté des humanités numériques déborde les frontières institutionnelles de l'université pour rassembler un pôle de pensée critique éclairant autant que possible la cité numérique mondiale en émergence et faisant reculer, par leur pratique même, l'emprise du côté obscur évoquée au début de ce texte. Loin d'être monolithique, ce pôle des humanités est animé d'une multitude de conversations entre chercheurs qui produisent et interprètent des données. Plutôt que des « vérités scientifiques » dogmatiques, la communauté des humanistes co-construit et diffuse des croyances. Mais puisque ces croyances sont justifiées par la transparence de leur processus de production et que leur autorité peut être examinée, ce sont des connaissances. En respectant certaines formes, en forçant ses participants à expliciter leur démarche et à qualifier leurs références, le nouvel humanisme promeut une lenteur positive de la communication et de la pensée, une réflexivité critique qui s'oppose à la viralité réflexe et aux usages toxiques des médias sociaux. Tout ceci sans entretenir l'illusion d'une perfection inatteignable. Comme l'entretien d'un jardin n'est jamais achevé, le côté obscur n'est jamais définitivement éliminé. Il resurgit de manière inattendue, encore et toujours, même du sein des groupes qui s'en croient les mieux protégés. Ne demandons pas aux humanités plus qu'elles ne peuvent donner. Il est déjà suffisamment difficile d'élargir le champ des connaissances et de tenir en respect l'ignorance et le mépris.

Une dernière remarque, concernant le rôle technique des humanités numériques. L'intelligence collective humaine est une vaste entreprise de création de sens, action et interprétation mêlées. De l'invention de l'écriture à celle des ordinateurs, outils, médias et codes augmentent régulièrement la fécondité de nos écosystèmes d'idées. Or c'est aujourd'hui dans le domaine de l'intelligence artificielle que se concentre de la manière la plus vive la fabrication de notre outillage culturel. Dès que des procédures de décision sont formalisées, elles animent des moteurs logiques. Dès que des données s'accumulent, elles entraînent des algorithmes. Des centres de calcul aux interfaces de l'Internet des choses, l'intelligence artificielle (IA) est à l'œuvre. Ses résultats sont ou seront

utilisés dans presque tous les aspects de la vie quotidienne⁴⁰. Vu la charge fantasmatique de l'IA et son impact croissant sur les différentes facettes de l'activité humaine – de la médecine à l'art militaire en passant par l'administration des entreprises et la gestion des médias sociaux – de nombreuses voix se sont élevées pour réclamer son usage éthique, et cette exigence a été reprise par les chercheurs et les industriels du domaine. Ce que l'on entend par une éthique de l'intelligence artificielle se ramène généralement aux deux points qui suivent : interdire les applications qui pourraient blesser directement les humains (les robots tueurs autonomes étant le cas le plus souvent évoqué) et corriger les biais (racistes, sexistes, etc.) qui viennent des données d'entraînement. Mais aussi nécessaires que soient les précautions contre les usages malins et les biais de l'IA, sa dimension éthique peut-elle être réduite à ces aspects défensifs de surface ? Est-ce qu'une éthique ne devrait pas viser aussi des fins positives ? Opposés à juste titre à la vision parfois caressée par Google d'une intelligence artificielle générale (capable de tout faire) et autonome, les industriels les plus avisés du secteur affirment que le but de l'IA telle qu'ils la pratiquent n'est pas de remplacer les humains, mais d'augmenter leurs capacités cognitives – et en particulier leurs capacités de décision – aussi bien sur un plan personnel que collectif. Fort bien ! Mais, dans une perspective éthique, ne faut-il pas privilégier l'augmentation de la dimension critique, ou réflexive, de l'intelligence ? Et pour toucher enfin à la racine du problème, une intelligence artificielle éthiquement orientée par cette finalité ne devrait-elle pas justement refléter, et donc prendre en compte, la création collective de sens qui est au cœur de l'intelligence humaine ? Or cette création de sens est précisément ce qui occupe les humanités. Si ces dernières ont pour mission de cultiver la pensée critique et de rendre l'intelligence collective réflexive, elles doivent se mêler d'ingénierie et travailler à l'orientation pratique tout comme à l'enrichissement théorique de l'intelligence artificielle. Elles suivraient en cela l'exemple des grands imprimeurs et architectes de la Renaissance – des ingénieurs, voire des industriels – qui étaient aussi des humanistes : des artistes, des savants versés dans les langues et les grands textes de la tradition.

40. Mais l'intelligence artificielle ne prendra évidemment jamais le « pouvoir », voir mon article de blogue “L'intelligence artificielle va-t-elle prendre le pouvoir ?” (Lévy 2018).

Bibliographie

Adorno, Theodor, et Max Horkheimer. 1974. *La dialectique de la raison : fragments philosophiques*. Collection Tel. Paris : Gallimard.

Austin, J. L. 1975. *How to do things with words*. 2d ed.. A Harvard paperback ; HP 88. Cambridge : Harvard University Press.

Baugh, Peter. 2017. « “Techplomacy” : Denmark’s ambassador to Silicon Valley ». *Politico*, juillet. <https://www.politico.eu/article/denmark-silicon-valley-tech-ambassador-casper-klynge/>.

Benkler, Yochai, Robert Faris, et Hal Roberts. 2018. *Network Propaganda : Manipulation, Disinformation, and Radicalization in American Politics*. Oxford : Oxford University Press.

Berners-Lee, Tim. 2018. « One Small Step for the Web... » *Inrupt*. <https://www.inrupt.com/blog/one-small-step-for-the-web>.

Botsman, Rachel. 2017. « Big data meets Big Brother as China moves to rate its citizens ». *Wired UK*, octobre. <https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>.

Boyd, Danah. 2018. « Media Manipulation, Strategic Amplification, and Responsible Journalism ». *Data & Society : Points*. <https://points.datasociety.net/media-manipulation-strategic-amplification-and-responsible-journalism-95f4d611f462>.

Braudel, Fernand. 1986. *Civilisation matérielle, économie et capitalisme, XVe-XVIIIe siècle, Tome 1 : Les Structures du Quotidien*. Paris : Armand Colin.

Brin, David. 1998. *The Transparent Society*. New York : Perseus Books. https://www.amazon.fr/Transparent-Society-David-Brin/dp/B002VYT1WQ/ref=sr_1_1?s=books&ie=UTF8&qid=1543541191&sr=1-1&keywords=The+transparent+Society.

Broudoux, Evelyne. 2017. « Autorité scientifique et épistémique à l’épreuve de la mesure des citations ». *Études de communication - Langages, information, médiations*, n 48 (juin) : 177-98. <https://doi.org/10.4000/edc.6841>.

Chomsky, Noam. 1957. *Syntactic Structures*. Mouton & Co. https://en.wikipedia.org/w/index.php?title=Syntactic_Structures&oldid=865266707.

———. 2000. *New horizons in the study of language and mind*. Cambridge England; New York, Cambridge; New York : Cambridge University Press. <https://doi.org/10.1017/CBO9780511811937>.

Chomsky, N, et M.P Schützenberger. 1963. « The Algebraic Theory of Context-Free Languages ». In *Computer Programming and Formal Languages*, édité par P. Braffort et D. Hirschberg, 118-61. Amsterdam : North Holland.

Coyote, Noam. 2018. « From Adam and Eve to Facebook, an history of Privacy ». *Medium*. <https://medium.com/beam-mw/from-adam-and-eve-to-facebook-an-history-of-privacy-42b97de43b6b>.

Custer, C. 2015. « China’s “citizen scores” credit system isn’t as Orwellian as the ACLU thinks...yet ». *Tech in Asia*, octobre. <https://www.techinasia.com/china-citizen-scores-credit-system-orwellian>.

Denyer, Simon. s. d. « China’s plan to organize its society relies on “big data” to rate everyone ». *Washington Post*. Consulté le 30 novembre 2018. https://www.washingtonpost.com/world/asia_pacific/chinas-plan-to-organize-its-whole-society-around-big-data-a-rating-for-everyone/2016/10/20/1cd0dd9c-9516-11e6-ae9d-0030ac1899cd_story.html.

Dubois, Elizabeth, et Grant Blank. 2018. « The echo chamber is overstated : the moderating effect of political interest and diverse media ». *Information, Communication & Society* 21 (5) : 729-45. <https://doi.org/10.1080/1369118X.2018.1428656>.

Duffy, Kim Hamilton. 2017. « The Time for Self-Sovereign Identity is Now ». *Learning Machine*. <https://medium.com/learning-machine-blog/the-time-for-self-sovereign-identity-is-now-222aab97041b>.

Ertzscheid, Olivier, et Antonio Casilli. 2017. *L’appétit des géants : Pouvoir des algorithmes, ambitions des plateformes*. Caen : C&F Editions.

Floridi, Luciano. 2014. *The 4th revolution : how the infosphere is reshaping human reality*. First edition. New York ; Oxford : Oxford University Press.

Foucault, Michel. 1966. *Les mots et les choses : une archéologie des sciences humaines*. Bibliothèque des sciences humaines. Paris : Gallimard.

- . 1969. *L'Archéologie du savoir*. Paris : Gallimard.
- . 1975. *Surveiller et punir : naissance de la prison*. Première édition. Bibliothèque des Histoires. Paris : Gallimard. <http://www.gallimard.fr/Catalogue/GALLIMARD/Bibliotheque-des-Histoires/Surveiller-et-punir>.
- . 1976. *Histoire de la sexualité*. Bibliothèque des histoires. Paris : Gallimard.
- Heath, Alex. 2017. « Mark Zuckerberg says Facebook will release the Russia-linked ads to investigators ». *Business Insider*. <https://www.businessinsider.com/mark-zuckerberg-says-facebook-will-release-the-russia-ads-to-investigators-2017-9>.
- Lévy, Pierre. 1990. *Les technologies de l'intelligence*. Sciences et société. Paris : La Découverte.
- . 1994. *L'Intelligence collective : pour une anthropologie du cyberspace*. Paris : Editions La Découverte.
- . 2011. *La sphère sémantique. Computation, cognition, économie de l'information*. Paris-Londres : Hermès-Lavoisier.
- . 2016a. « La curation collaborative de données ». *Pierre Levy's Blog*. <https://pierrelevyblog.com/2016/03/11/la-curation-collaborative-de-donnees/>.
- . 2016b. « Cultiver l'intelligence collective ». *Pierre Levy's Blog*. <https://pierrelevyblog.com/2016/03/30/cultiver-lintelligence-collective/>.
- . 2016c. « La littéracie en curation de données ». *Pierre Levy's Blog*. <https://pierrelevyblog.com/2016/03/16/la-literacie-en-curation-de-donnees/>.
- . 2016d. « L'intelligence collective, en quelques mots... ». *Pierre Levy's Blog*. <https://pierrelevyblog.com/2016/03/03/lintelligence-collective-en-quelques-mots/>.
- . 2018. « L'intelligence artificielle va-t-elle prendre le pouvoir ? » *Pierre Levy's Blog*. <https://pierrelevyblog.com/2018/09/06/lintelligence-artificielle-va-t-elle-prendre-le-pouvoir/>.
- Machiavel. 2010. *Discours sur la première décade de Tite-Live*. La Pléiade.

- Morozov, Evgeny. 2013. *To Save Everything, Click Here : The Folly of Technological Solutionism*. First Trade Paper Edition edition. PublicAffairs.
- Osborne, Samuel. 2015. « China has made obedience to the State a game The Independent ». *Independant*. <https://www.independent.co.uk/news/world/asia/china-has-made-obedience-to-the-state-a-game-a6783841.html>.
- Peirce, Charles S. 1992. *The essential Peirce : selected philosophical writings*. Bloomington : Indiana University Press.
- Peirce, Charles S, et Gérard Deledalle. 1978. *Écrits sur le signe*. Paris : Éditions du Seuil.
- Polanyi, Michael. 1958. *Personal knowledge : towards a post-critical philosophy*. Chicago : University of Chicago Press.
- Popper, Karl R. Sir. 1972. *Objective knowledge ; an evolutionary approach*. Oxford : Clarendon Press.
- Roser, Max. 2018. « Life Expectancy ». *Our World in Data*. <https://ourworldindata.org/life-expectancy>.
- Roser, Max, et Esteban Ortiz-Ospina. 2017a. « Global Extreme Poverty ». *Our World in Data*. <https://ourworldindata.org/extreme-poverty>.
- . 2017b. « Global Rise of Education ». *Our World in Data*. <https://ourworldindata.org/global-rise-of-education>.
- Sadin, Eric. 2016. *Silicolonisation du Monde*. Paris : L'échappée éditions.
- Saussure, Ferdinand de. 1985. *Cours de linguistique générale*. Paris : Payot.
- Singer, P. W., et T. Brooking Emerson. 2018. *Like War : The Weaponization of Social Media*. Boston : Houghton Mifflin Harcourt.
- Stiegler, Bernard. 2016. *Dans la disruption : Comment ne pas devenir fou ?* Paris : Les liens qui libèrent éditions.
- Vaughan-Nichols, Steven J. 2018. « Microsoft open-sources its patent portfolio ». *ZDNet*. <https://www.zdnet.com/article/microsoft-open-sources-its-entire-patent-portfolio/>.
- Vincent, James. 2017. « Putin says the nation that leads in AI “will be the ruler of the world” ». *The Verge*. <https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world>.

Vitali-Rosati, Marcello. 2018. *On Editorialization : Structuring Space and Authority in the Digital Age*. Theory on demand 26. Amsterdam : Institute of Network Cultures. <http://networkcultures.org/blog/publication/tod-26-on-editorialization-structuring-space-and-authority-in-the-digital-age/>.

Wittgenstein, Ludwig. 1953. *Philosophical investigations*. Oxford : Blackwell.

———. 1993. *Tractatus logico-philosophicus*. Bibliothèque de philosophie (Gallimard (Firme)). Paris : Gallimard.

Wormser, Gérard. 2018. *Facebook*. Les atelier de Sens public. <http://ateliers.sens-public.org/facebook>.