Les Big Data : pistes de réflexions historiques, éthiques et épistémologiques pour l'appropriation sociale

Ghislaine Azemard, Ben Henda, Henri Hudrisier

RED DE LAS CATEDRAS UNESCO DE COMUNICACIÓN
NETWORK OF UNESCO CHAIRS IN COMMUNICATION
RÉSEAU DES CHAIRES UNESCO EN COMMUNICATION

Universidad Iberoamericana, Mexico City

CONFERENCIA ORBICOM: DATOS ABIERTOS, MEDIOS Y CIUDADANÍA
28-29 de Octubre de 2015
ORBICOM CONFERENCE: OPEN DATA, MEDIA AND CITIZENSHIP
28-29 October 2015
CONFÉRENCE ORBICOM: DONNÉES OUVERTES, MÉDIAS ET CITOYENNETÉ
28-29 octobre 2015

LES BIG-DATA : PISTES DE REFLEXIONS HISTORIQUES, ETHIQUES ET EPISTEMOLOGIQUES POUR L'APPROPRIATION SOCIALE ; RETOURS ET ATTENTES D'EXPERIMENTATIONS EN LEARNING ANALYTICS

Ghislaine AZEMARD, Mokhtar BEN HENDA, Henri HUDRISIER.

Résumé :

Dans ce papier nous visons à explorer des pistes de transformations de la nouvelle technoculture qui nous deviendra indispensable pour nous approprier deux technologies émergentes étroitement associées : les *Big-datas* et les *Datas Analytics*. Nous évoquons notamment l'histoire des big-datas et de l'analyse statistique, mais aussi de leurs avant-coureurs pré-numériques. Pour ce qui est de l'analyse statistique multidimensionnelle nous discutons les questions épistémologiques, mais aussi éthiques provoquées par ce retournement copernicien de l'analyse (discussion Benzécri/Bourdieu). Les questions de l'être techno-numérique et de son rapport d'année en année plus étroit avec l'intelligence humaine, elle aussi bouleversée par sa mise en réseaux collaboratifs, sont rapidement évoquées (Heidegger, Simondon, Derrida et la bibliographie actuelle). Nous concluons en faisant état de l'étroite participation de notre Chaire ITEN-Unesco à la normalisation ISO des TICE et du chantier de normalisation des *Learning Analytics* qui vient de s'ouvrir. Nous présentons aussi notre projet HD Muren (Humanités digitales mettant en synergie participative éducation, recherche et patrimoines dans un contexte culturellement euro-méditerranéen) ; un projet qui s'inscrit dans le cadre à projet d'enseignement-recherche Idéfi-Créatic.

Mots clés :

BIG DATA, MUTATIONS TECHNOLOGIQUES ET ÉPISTHÉMOLOGIQUE, DATA ANALYTICS, HUMANITES NUMÉRIQUES, LEARNING ANALYTICS.

Abstract:

In this paper, we aim to explore ways of transformations of the new techno-culture that will become essential to appropriate two emerging technologies closely related: Big-datas and Datas Analytics. We discuss as well the history of big-datas and statistical analysis but also their digital pre-warning. In terms of multivariate statistical analysis, we discuss the epistemological issues but also ethical Copernican reversal caused by this analysis (discussed Benzécri / Bourdieu). The questions to being techno-digital, closely related with human intelligence, it also upset by its implementation collaborative networks, are quickly raised (Heidegger, Simondon, Derrida and the current bibliography). We conclude by stating the close involvement of our ITEN-UNESCO Chair for ISO standardization of ICT and Learning Analytics standardization project that just opened. We also present our project Muren HD (digital humanities involving crowdsourcing synergy for education, research and heritage in a culturally Euro-Mediterranean context); a project that is part of education-research-project IDEFI Créatic.

Keywords

BIG DATA, TECHNOLOGICAL & EPISTHEMOLOGIC CHANGES, DATA ANALYTICS, DIGITAL HUMANITIES, LEARNING ANALYTICS

---oOo---

1 INTRODUCTION

Dans un essai sur l'Art de la lecture, Homi Bhabha, enseignant en Humanités à Harvard, soutient l'idée que les sciences humaines sont «avant tout les sciences princeps de l'interprétation ». Autant qu'évaluer les évidences linguistiques, orales ou visuelles, les sciences humaines à travers la littérature, les études classiques, les langues modernes, ou la philosophie utilisent l'interprétation pour créer un très vaste univers d'associations, de contextes, de significations et de valeurs. L'interprétation serait donc une activité qui, à travers l'exercice du jugement d'importants travaux (d'art, de littérature, de musique, de sculpture, d'architecture, etc.) génère de la valeur ajoutée conceptuelle, sociale et culturelle. Les sciences humaines nous aident ainsi à devenir non seulement des citoyens au sens politique, social et juridique, mais aussi des citoyens au sens culturel. C'est là que se situe la force réelle des humanités. L'interprétation humaniste joue aussi un rôle. C'est là qu'elle doit intervenir pour faire face au flux débordant de données du monde numérique. Tant que nous enseignons à nos étudiants comment faire de l'interprétation, ceci permet que le flot de faits et d'informations continue à être transformé en connaissances. L'interprétation est la force de médiation qui doit se diffuser à travers toutes les données pour produire et classer les connaissances.

Cette acception de la force interprétative des sciences humaines trouve aujourd'hui sa continuité dans une technique plus ancienne, étroitement associée aux Big-data : les Data Analytics qui constituent la force interprétative des Big data, ce qui bouleverse et redessine les fondements historiques de l'analyse des données et l'interprétation des discours. Les *Big Data* et les Data Analytics ont émergé dans le monde de l'entreprise, mais elles sont en

pleine expansion transdisciplinaire, produisant de nouvelles synergies de convergence et d'interopérabilité entre des domaines scientifiques qui évoluaient jusqu'ici dans une verticalité souvent étanche. A une question portant sur l'apport que peuvent avoir les Big data sur les SHS, Michel Wieviorka, sociologue français et Directeur de la Fondation des Maisons des Sciences de l'Homme, répond **: «** Les *Big Data* induisent une autre manière de travailler, mais elles permettent, ou permettront bientôt, de prévoir des comportements d'achat, de calculer la probabilité pour une personne d'avoir ou de transmettre une maladie génétique, de connaître celle de conduite accidentelle pour des automobilistes, de récidive pour des criminels […]. Cela va modifier toute notre vie collective, transformer par exemple le droit, mettre fin au voile d'ignorance qui permet de traiter les individus comme abstraitement égaux. Nous allons vers une individualisation qui doit beaucoup à ces immenses ensembles de données »[1].

2 LE CONTEXTE TECHNO-HISTORIQUE DES *BIG DATA*

Les big-data, (indissociables aujourd'hui des Data Analytics) constituent une valeur ajoutée considérable pour l'intelligence humaine mais elles n'ont pu se développer que parce qu'existe une informatique en réseau permettant de disposer de capacités de traitement et de prise en compte de volumes de données jusqu'alors impossible à rassembler et surtout à interpréter. Elles interrogent obligatoirement le spécialiste de l'information et de la communication dans la mesure où elles constituent véritablement non seulement une innovation qualitative et quantitative, mais une véritable mutation de notre potentiel d'analyse de certains problèmes.

Des avant-coureurs historiques

Historiquement, l'humanité a déjà connue des mutations de sa capacité à rassembler les données et l'appropriation techno-culturelle qui en a été faite à chaque époque considérée est intéressante à observer pour nous aider à penser notre civilisation en devenir des big-data.

Certes, déjà avec l'écriture la notion de pensée rationnelle pouvait émerger dans la mesure où des communautés de « sages » dispersés dans l'espace géographique et héritant de leurs écrits accumulés par plusieurs générations pouvaient « penser ensemble » un même problème. Il en est résulté une forme particulière de pensée qualifiée de scientifique, décrite et théorisée notamment par les philosophes grecs. C'est globalement en nous inspirant des différents concepts de l'épistémologie antique, réactualisée à la Renaissance après la révolution de l'imprimé, puis perfectionnée par les Lumières et les Encyclopédistes que nous construisons encore aujourd'hui nos raisonnements scientifiques.

---

[1] Philippe Testard-Vaillant (2014). « Interview de Michel Wieviorka : les sciences humaines et sociales à l'ère numérique : 10 janvier 2014 ». Paru dans *CNRS Le Journal* du 02/04/2014 « Big data, la déferlante des Octets ». https://lejournal.cnrs.fr/articles/interview-de-michel-wieviorka-les-sciences-humaines-et-sociales-a-lere-numerique

Comme pour les big-data actuelles, l'accumulation fantastique de données grâce l'écriture bousculait déjà la culture de l'homo sapiens d'avant l'écriture, puis celle de l'intellectuel d'avant l'imprimerie. Le sujet a été mainte fois traité sur le plan historique, mais aussi par nombre de penseurs de notre civilisation de l'ère numérique pour que nous ne nous attardions pas sur ce sujet. Nous le signalons cependant pour mémoire parce que comme pour les big-data, ce qui est en cause avec « le Miracle grec » n'est pas tant seulement l'appropriation démocratique de l'écriture, mais une véritable mutation résultant de la prise de conscience par une élite s'instituant comme « savante », qu'une nouvelle mise en forme scripturale très codifiée, permettait dès lors de « calculer la pensée » en mettant en œuvre une « logique formalisée » qui dépassait de beaucoup la mise en forme culturellement naturelle selon des patterns grammaticaux transmis de façon ancestrale dans toute communauté linguistique. Il ne s'agissait plus seulement d'énoncer des phrases permettant peu ou prou de défendre un point de vue mais d'inventer une logique : un cadre formel de l'argumentation scientifique, par exemple les *syllogismes*[2]. Par là même, il devient possible pour les philosophes antiques de théoriser un mode de *penser autrement* grâce à l'accumulation en masse, jusqu'alors inconnue de descriptions analytiques du monde[3]. La bibliothèque d'Alexandrie est de ce point de vue un véritable avant-coureur des big-data.

La mutation qui lui fait suite est celle de l'imprimerie : son perfectionnement fut constant pendant les 50 ou 100 années qui ont suivies son invention. Ce progrès notamment techno-typographique est indissociable d'une « invention élitaire de l'Humanisme ». Globalement le monde des lettres (mais aussi de la science), s'adapte et invente des modalités de traitement éditorial à même de prendre en compte des quantités exponentielles de données nouvelles. Les exemples que nous avons souvent déjà cités dans d'autres publications[4], sont ceux de la standardisation de la mise en page, ou encore de la standardisation de l'écriture grecque pour rendre accessible l'énorme corpus des écrits grecs dont les manuscrits sont scripturairement très disparates[5]. Ainsi s'invente une nouvelle police unifiée : le « Grec du Roy » : en fait l'entente de tous les éditeurs-imprimeurs s'accordant à prendre pour référence typographique unique une graphie proche de la façon d'écrire à Athènes au temps d'Aristote. On l'a compris ces inventions culturelles sont formellement des avant-coureurs des big-datas : pour faire face à une augmentation exponentielle des données culturelles ou scientifiques on invente, ou on perfectionne, le système d'information antérieur. Cette évolution innovante des débuts de l'âge de l'imprimerie et la mise en place d'une technoculture ad hoc sont très abondamment décrits et très souvent mise en relation de

---

[2] La logique propositionnelle est l'héritière directe des travaux des logiciens antiques.

[3] On peut lire notamment sur ce sujet : HAVELOCK (Eric A.), Aux origines de la civilisation écrite en Occident, Paris, éd. François
Maspéro, 1981.

[4] Par exemple : Henri Hudrisier, Sophia Benzina, Hichem Ismail, Rachid Zghibi, Sihem Zghidi, Laurent Romary, Mokhtar Ben Henda, Loula Abdelrazak, Arnaud Laborderie, Ghislaine Azémard, *La stylistique des notes chez Rétif de la Bretonne : un hypertexte avant le numérique*, in CIDE.17 Livre post-numérique : historique, mutations et perspectives, Actes du 17e colloque international sur le Document Electronique, sous la dir. de Khaldou Zreik, Ghislaine Azémard, Stéphane Chaudiron, Gaétan Darquié, Paris, éd. Europia, 2014.

[5] Ces textes sont écrits dans des alphabets formellement très différents sur plus de mille ans de littérature et de l'Indus à la Sicile.

similarité avec la relativement lente évolution de l'appropriation sociale d'une technoculture de l'informatique. Ainsi, comparer l'évolution de la civilisation de l'imprimerie dans les deux premières générations qui ont vécu son introduction progressive en le mettant en parallèle avec les deux générations qui ont suivi l'invention de l'informatique est très éclairant pour penser les big-data qui perfectionnent l'ère de l'informatique.

L'imprimerie a, certes, permis de fonder une connaissance et une intelligence collective via l'édition, les bibliothèques et les archives patrimoniales, mais l'accomplissement de ce processus a duré des décennies voire des siècles. Les auteurs des Lumières, par exemple, sont devenus des valeurs morales et philosophiques communément partagées par nos sociétés « dites occidentales ». Mais cette transformation des « valeurs humanistes partagées » n'est pas seulement le résultat d'un siècle de publications, d'échanges intellectuels et de partages d'idées. Il a fallu aussi le choc historique de la Révolution française et de son idéalisation mondiale[6]. Il fallait aussi une lente digestion sociale et culturelle des idées de tolérance qui ne pouvaient se populariser dans les sociétés occidentales sans la banalisation des valeurs première de l'agnosticisme ou de l'athéisme : sans que notre société occidentale admette la possibilité que les « incroyants » et les « sans dogmes » soient considérés comme égaux aux croyants dans leur comportement moral même s'ils ne sont pas soumis à l'effroi des punitions divines. Cette lente assimilation culturelle a été intimement liée à des processus éducatifs de masse qui ont permis que ces idées se propagent et viennent à bout de l'obscurantisme d'avant les Lumières. Si nous prenons pour hypothèse une croyance dans le progrès humain il nous semble clair que sans imprimerie, point de tolérance. Et pourtant, il est bien connu que nombre d'intellectuels antiques ont dû pratiquer une certaine forme d'humanisme, de tolérance, voire d'athéisme mais au cœur d'une société dans laquelle l'esclavage supportait une part prépondérante du système productif. Qu'en est-il vraiment du 21^{e} sc. sur cette question éthique ? Notre bonne conscience d'intellectuel risquerait gros à trop fouiller ces questions pourtant étroitement liées, comme le souligne Simondon, à l'être techno-numérique (nous y reviendrons).

L'effort éducatif, muséographique, journalistique et l'organisation de nombreuses expositions régionales, nationales ou universelles constituent en soi, des accumulations jusqu'alors inédites et impensables de corpus, de big-datas contextualisées, rationnellement organisées, scientifiquement ou industriellement classées ou présentées.

La publication imprimée de la quasi-exhaustivité historique des observations astronomiques permettent à Copernic de découvrir son modèle circum-solaire

Revenons sur la mutation de l'imprimé. En tant qu'historienne de « l'âge de l'imprimerie », Elizabeth Eisenstein, met l'accent sur une appropriation spécifique de la technoculture de l'imprimé. Elle souligne que Copernic n'a pu découvrir son modèle circum-solaire que parce qu'il a été le premier astronome à pouvoir rassembler grâce à l'imprimerie de très nombreuses annales d'observations astronomiques, dont n'avaient pas pu bénéficier ses

[6] Idéalisation telle que la présente Michelet largement vulgarisée dans la société française.

prédécesseurs. Laissons Thomas Kuhn[7], cité par Elizabeth Eisenstein, nous exposer ce fantastique exemple, le plus emblématique nous semble-t-il, du saut qualitatif des *Avant-coureurs des Big Data* disponibles avec l'imprimerie :

« Un demi-siècle encore après la mort de Copernic, il n'y avait eu aucun changement qui pût être potentiellement révolutionnaire dans les données accessibles aux astronomes… [Ce n'était pas des lunettes de Galilée[8] que Copernic pouvait pointer vers le ciel, mais de simples viseurs n'augmentant pas l'échelle de vision comme en utilisaient déjà les astronomes de l'Antiquité]… Une étude plus attentive de ces changements pourrait contribuer à expliquer pourquoi les systèmes de cosmographie, de cartographie du globe terrestre, de synchronisation des chronologies, de codification des lois et de compilation de bibliographies furent tous radicalement transformés avant la fin du XVIe siècle. »

Ce qui change radicalement avec l'analyse de Copernic, c'est que s'appuyant sur des quantités phénoménales d'observations astronomiques (y compris bien sûr ce qu'on pouvait considérer à l'époque comme des données erronées), il a su faire émerger une nouvelle vision de l'univers. Les astronomes d'avant Copernic décrivaient des sphères concentriques à la terre dans lesquelles des astres divers décrivaient des orbes en principes circulaires. Ils observaient aussi des astres aux orbes erratiques qui parcourraient des trajectoires comportant des reculs, des inflexions de courbes incompréhensibles jusqu'à ce que les « big data rassemblées par l'imprimerie et analysées par Copernic » lui donne l'idée qu'une logique se dégageait de toute ces orbes atypiques. Leurs apparents reculs et les inflexions de courbes provenaient de la superposition géométrique de plusieurs systèmes de circonvolutions : celui de la Terre et la Lune, celui du Soleil, et celui plus vaste de notre Galaxie qui projetés de façon distincte l'un dans l'autre donne une pleine logique à toutes ces trajectoires d'astres. Avant que les données astronomiques aient été exhaustivement imprimées, les données manuscrites rassemblées par tous les astronomes médiévaux et antiques n'atteignaient pas une masse critique suffisante pour qu'on puisse entrevoir l'interrelation logique qui à permis que s'invente le système copernicien. Ainsi que le signale Copernic lui-même : «le mouvement de la terre seule suffit donc à expliquer un nombre considérable d'irrégularités apparentes dans le ciel[9]».

Penser les big-data en acceptant la mutation de nos habitus épistémologiques

Si nous nous sommes permis ce bref retour historique sur les « avant-coureurs » des *Big Data* c'est pour montrer que les nouveaux potentiels cognitifs que les *Big Data* induisent, doivent pouvoir être pensés comme une révolution copernicienne des mentalités cognitives numériques. Simultanément, les problèmes éthiques ou de relations entre les humains et les

---

[7] KUHN (Thomas), La révolution copernicienne, Paris, éd. Fayard, 1973, p. 153 ; cité et commenté par EISENSTEIN (Elizabeth L.), La révolution de l'imprimé dans l'Europe des premiers temps modernes, Paris, éd. La Découverte, 1991, pp. 10 et 101

[8] Ces lunettes astronomiques, dites de Galilée (qui nait plus de 20 ans après Copernic) n'apparaitront que plus d'un demi-siècle après la mort de Copernic.

[9] Nicolas Copernic, Commentariolus,

entreprises ou institutions - se posent d'évidence face aux big-data. Il devrait pouvoir se fonder un néo-humanisme, envisagé en fonction des règles de la nouvelle technoculture du numérique et de l'intelligence collective. Soulignons néanmoins que cette dernière question peut sembler orthogonale à la précédente. La morale et l'éthique des big datas est très importante mais nous aurions tendance à penser que cette dernière question est faiblement liées aux questions épistémologiques découlant d'une informatique des big-datas.

Il devient, en effet, largement admis que les *Big Data* nous permettent de faire émerger de nouvelles données, de nouvelles hypothèses, de nouvelles façons de penser des problèmes jusqu'alors inaccessibles à notre entendement. Elles sont déjà une composante principale de tous les secteurs d'activité dans l'économie mondiale. Elles imposent une nouvelle conception des mécanismes fonctionnels des institutions et des entreprises qui ont besoin de comprendre où elles se situent en maturité par rapport à l'usage de grands volumes de données. Cela implique l'examen du niveau de préparation de leurs environnements, de leurs cadres juridiques et réglementaires, de leurs infrastructures technologiques et de leurs capacités à exploiter les *Big Data* comme une nouvelle stratégie d'action. Vues sous cet angle nous pouvons déjà entrevoir qu'en pensant les big-datas avec nos habitus épistémologiques d'antan, nous ne prenons pas la pleine mesure des questions posées par les big-datas d'aujourd'hui et en devenir. Sans nul doute, la nouvelle épistémologie du e-sémantique en construction devra intégrer des éléments d'éthique et de morale (économique, sécuritaire, privacy, prévalence du bien-fondé). C'est en suscitant chez les grands acteurs des big-datas des comportements de ce type que nous resterons dignes de nos prédécesseurs de la Renaissance et que nous serons capables d'ouvrir l'ère d'un Néo-humanisme fondée notamment sur une nouvelle épistémologie intégrant à part entière les big-datas et l'analyse des données.

3 HUMANITES NUMERIQUES & *BIG DATA :* LA CONNAISSANCE DISTRIBUEE

Avec l'émergence de l'informatique et du numérique, il y a eu, comme à chaque nouvelle transition technoculturelle, une réappropriation des mêmes principes fondamentaux de la production et de la diffusion de l'information. Le numérique n'était à ses origines qu'une transformation (informatisation) des procédés pratiqués par l'industrie du calcul puis du texte imprimé avec toutefois des progrès se succédant par *générations technologiques* vers de nouvelles filières technologiques de mise en œuvre et d'usage : la téléphonie, la radio, la TV, la bureautique, la robotique la domotique, les technologies nomades, etc. En revanche, l'informatique et le numérique ont poussé les limites de ce potentiel mobilisateur pour en faire une intelligence collective quasi instantanée grâce aux *Big Data* et l'éventail des technologies associées comme l'informatique dans les nuages (*Cloud Computing*) et les réseaux à très haute vitesse (GRID).

Les *Big Data* du numérique font désormais l'objet d'un large consensus sur leurs capacités à dépasser les approches traditionnelles de gestion et diffusion des données grâce aux critères de « 4V » (volume, vitesse, variété et véracité) et le potentiel qu'elles ont à susciter l'innovation et le progrès dans tous les domaines d'activités. Par l'intelligence collective distribuée et ses capacités à résoudre des problèmes en traitant des données qui ne soient pas la propriété d'une seule personne ou localisées à l'intérieur d'un seul ordinateur, les *Big*

Data font émerger des interactions coordonnées entre un grand nombre d'utilisateurs et leurs *doubles technologiques*.

A lire ce dernier paragraphe qui décrit la doxa d'appropriation sociale des big-data on perçoit combien des slogans marketing comme ce fameux 4V posent un sérieux problème d'éthique et de déontologie techno-épistémologique : *volume* OK, *vitesse* et *variété* bien sûr, mais *véracité* ? Est-ce celle de la logique rationnelle de l'analyse ? Nous y reviendrons plus bas. Celle encore d'une transparence et d'une sincérité et donc d'une véracité du recueil des données ?

L'émergence puis la maturité opérationnelle à large échelle de l'analyse des données : l'acte de naissance des big-datas modernes.

Un des progrès fondateur des big-data nous vient des extraordinaires progrès de l'analyse des données, mathématiquement découvertes au début du 20^{e} siècle mais que l'informatique a véritablement fait muter.

Qu'elles soient multifactorielles, en composantes principales, analyse factorielle, analyse des correspondances multiples multivariées, etc. la plupart de ces techniques participent toutes d'un *retournement copernicien* de *l'intelligence d'analyse*. Comme l'imprimerie pour Copernic ces nouveaux potentiels d'analyse statistique nous permettent de traiter d'énormes corpus ; opportunément la statistique, dans ses fondements mêmes, exige pour être représentative, le traitement de gros corpus. Dans la plupart de ces méthodes, on hiérarchise les dimensions géométriques d'un hyper-espace à N dimensions (N étant le plus petit côté de la matrice d'analyse). L'innovation révolutionnaire de ce type d'analyse automatique (impossible à concevoir de façon opératoire sans ordinateur) tient à ce que l'on évolue d'une stratégie du « choix d'items ou de couples d'items pertinents humainement réalisés » parmi les hypothèses de liens conceptuels (ce que faisait Copernic) à une « heuristique de la pertinence des relations liant des ensembles d'items automatiquement proposée par l'outil d'analyse ».

Les items d'un corpus s'organisent statistiquement en un hyper-espace à N dimensions qui n'est rien d'autre qu'une sorte de « tableau non plus à double entrée mais à N entrées » que constituent les N items d'une problématique. Dès lors, en hiérarchisant par l'analyse les poids relatifs de ces items dans l'hyper-espace on fait apparaître des liens privilégiés entre différents items d'un corpus : on compose des ensembles, on hiérarchise des poids relatifs, des distances entre items, on projette dans des graphes à 2 ou 3 dimensions des nuages de points. Jean-Paul Benzécri (le fondateur de l'analyse factorielle) a eu dans les années 70 des débats très passionnant avec Pierre Bourdieu. Nous citons ici un article de Björn-Olav Dozo rapportant ce débat[10] :

[10] Dans la revue « Contextes » 3/ 2008 : Questions biographiques en littérature » : un article de Björn-Olav Dozo : *Données biographiques et données relationnelles, Notes théoriques pour une utilisation complémentaire des outils quantitatifs*. Consulté sur : https://contextes.revues.org/1933?lang=en

« L'analyse factorielle permet de faire surgir la structure des données, la façon dont chaque variable se situe par rapport aux autres, de manière différentielle et relationnelle. La sociologie structurale de type bourdieusien, et Bourdieu lui-même, en ont fait un outil de représentation puissant, au service de leurs thèses : l'outil permettait de mettre au jour la structure multidimensionnelle et relationnelle du champ étudié.

Ainsi, dans le deuxième chapitre de *La Distinction* (p. 109-187), intitulé L'espace social et ses transformations, Bourdieu recourt dès le départ à l'analyse factorielle des correspondances multiples pour expliciter sa conception de l'espace social. Quand il explique ce qu'entraîne la création de classes d'individus, il raisonne à partir d'un modèle fondé sur la description d'individus par des variables[11], et se réfère à J.-P. Benzécri, le fondateur et le promoteur de l'analyse factorielle des correspondances en France. Il utilise l'analyse factorielle pour critiquer les dérives qu'entraîne l'usage de certaines catégories en statistique. Il est intéressant de constater qu'il mobilise de la sorte un outil statistique pour mettre en cause certaines catégorisations qui étaient largement utilisées par d'autres statistiques.

…[…]… L'usage qu'il fait de l'ACM (L'analyse des correspondances multiples) montre en quoi les classifications construites a priori présagent déjà d'un découpage du monde qui masque des relations « souterraines », des corrélations entre variables a priori indépendantes mais qui, dans les faits, apparaissent liées, ce que met en évidence l'ACM.
Ce refus de réduire le monde social à des classes préconstruites fut une des grandes prises de position de Pierre Bourdieu à travers son usage des statistiques. …[…]…

L'usage qu'a fait Bourdieu de l'analyse factorielle est donc d'offrir une synthèse d'un travail de recherche, résumé visuel et efficace, qui permet l'appréhension quasi immédiate par le lecteur – après un travail d'interprétation, comme tout graphique – d'un grand nombre de données et surtout des relations qu'elles entretiennent.

Il s'agit là d'un des avantages principaux de la méthode, que Bourdieu a su très bien exploiter : le graphique fait apparaître *une concentration de l'ensemble des possibles* d'un espace social particulier, en soulignant que chaque pratique ne prend sens que par rapport aux autres. »

Nous soulignons dans ce texte le dernier paragraphe qui insiste sur les exigences des méthodes d'interprétation qu'impliquent toutes les analyses statistiques multidimensionnelles.

Quand ce sont des « scientifiques honnêtes et informés » qui utilisent en pleine conscience des méthodes d'analyses statistiques automatiques issues de big-datas, ils s'obligent systématiquement à pratiquer « humainement (et non plus de façon automatique) » une

[11] « […] les individus rassemblés dans une classe qui est construite sous un rapport particulier mais particulièrement déterminant apportent toujours avec eux, outre les propriétés pertinentes qui sont au principe de leur classement, des *propriétés secondaires* qui sont ainsi introduites en contrebande dans le modèle explicatif » Bourdieu (Pierre), *La Distinction*, Paris, éd. Minuit, 1979.., p. 113).

interprétation des graphiques statistiques générés par l'informatique. Malheureusement, on doit regretter que « les vendeurs de big-datas et des analyses automatiques associées » ont très souvent tendance à vendre l'ensemble comme un tout. Dès lors, les utilisateurs scientifiques naïfs (ou nombre de sociologues non naïfs mais *mercenaires* du monde du commerce, des institutions, des grandes entreprises) pensent, ou font semblant de penser, que ces méthodes génèrent automatiquement une analyse « logique et rationnelle » d'une situation.

La réponse à cette question doit être obligatoirement nuancée selon une logique similaire à celle de Pierre Bourdieu il y a quelques décennies.

Si le scientifique (ou l'utilisateur industriel, économique ou institutionnel) cherche à établir des catégories, les hiérarchiser dans leur importance relative, générer ainsi des heuristiques (étymologiquement faciliter le potentiel à découvrir, à trouver, voire à vendre) : dès lors les analyses de ce type sont très pertinentes car elle permettent de proposer des types de proximités auxquelles l'utilisateur n'aurait sans doute pas pu penser tout seul ; de plus il est certain qu'un homme seul ou même un groupe de recherche, aurait dès la mise en œuvre de la recherche, réduit drastiquement le corpus et n'aurait jamais pu exploiter un très vaste corpus comme ceux des big-datas. Si le but de la recherche s'arrête là : établir des catégories, des proximités, des importances relatives, des tendances, l'analyse automatique des données est très pertinente mais exige bien sûr une phase incontournable d'interprétation des données.

Dans certains cas, celui par exemple d'un corpus de profils d'utilisateurs ou de tendances d'achats qui est périodiquement réactualisé en restant très similaire, on peut penser raisonnablement que la réinterprétation systématique chaque mois ne s'impose pas obligatoirement. *Mutatis mutandis*, les mêmes causes produiront sans doute les mêmes effets.

De ce fait, nous nuancerions notre jugement. Certes, dans de très nombreux cas, le sociologue conscient n'a pas tort, car ce que cherche avant tout son employeur, c'est de cibler des catégories d'individus (des clients, des profils d'apprenants, des malades susceptibles de réagir de telle ou telle façon à un médicament, des catégories socio-professionnelle ayant telles ou telles exigences politiques, des types d'individus présentant tels risques de délinquance, des profils en ressources humaines), ou encore des catégories de produits ou services classifiés par rapport à leur usage, mais aussi des relations mixtes entre des services ou produits et des individus utilisateurs, etc. Là où il peut y avoir dérapage déontologique, c'est quand un scientifique conscient, conforte son employeur dans sa naïveté des big-datas qui analysent toutes seules.

Par contre, s'engageant dans le sillage d'un marketing dominant des Big-datas et des Data Analytics qui résolvent tout et analysent « véridiquement » et rationnellement, des dérives graves peuvent, et sont fréquemment signalées et décrites. Par exemple, un banquier, un assureur peut utiliser les big-datas et l'analyse automatique pour codifier l'autorisation d'un emprunt, appliquer tel ratio de risque, évaluer un risque de délinquance dès l'enfance. Autre exemple, des patrons de laboratoire de recherche médicale peuvent écarter

systématiquement tel type de malade (qui risque de ne pas donner des résultats favorables) d'un protocole thérapeutique expérimental. Dans ces situations, la responsabilité déontologique des utilisateurs savants de ces techniques est évidemment mise en cause, parce qu'on est explicitement passé d'une probabilité d'heuristique (une plus grande tendance de trouver telle tendance d'un item ou de qualification du lien entre deux ou plusieurs items) à la « *véracité* » de la description possiblement individuelle de chaque items analysé.

Quantités d'exemples qui concernent la science peuvent être donnés en exemple : l'analyse linguistique, la géologie, l'archéologie, l'histoire, la littérature. Chaque analyse, chaque corpus est un cas d'espèce. Néanmoins celui qui veut s'impliquer dans l'usage de ces nouveaux outils doit effectivement associer dans la construction de son raisonnement une « nouvelle honnêteté scientifique » qui distinguera selon l'usage proposé des heuristiques d'avec les catégorisations rationnelles individualisées.

Pour reprendre quelque uns de nos exemples :

- l'expert linguiste en *traductique* qui analysera automatiquement d'énormes corpus de contextes d'énonciation, agit comme il devrait le faire puisque ça lui permet de proposer « à la volée » une traduction vraisemblable. Par contre, si dans des formations à distance, un linguiste propose de noter automatiquement des exercices, on peut considérer qu'il s'agit là d'une nette dérive déontologique
- le chercheur en littérature peut trouver quantités de situations ou ces types d'analyses sur des très grands corpus soit leur permettrons de découvrir des rapprochements de situations dramatiques, des cooccurrences de personnes,… soit permettrons de repérer des similarités stylistiques inattendues. Il est peu vraisemblable qu'un chercheur en littérature accepte d'analyser uniquement automatiquement des corpus de textes. Par contre un chercheur utilisant la TEI en littérature pourra valablement accepter qu'un premier « *draft* de balisage » d'un corpus de poèmes versifiés soit automatiquement généré par des processus probabilistes[12], quitte à réviser ultérieurement « à la main », ce premier draft qui a toutes les chances de comporter des erreurs. Cependant, le traitement automatique des gros corpus lui aura permis d'accélérer les phases fastidieuses de balisage non approfondi de son corpus. Au-delà, certaines méthodes probabilistes vont même permettre de dégrossir, voire de découvrir des tendances stylistiques, des similarités de situations dramatique, qu'un individu seul n'aurait pas forcément entrevu. Dans toutes ces dernières propositions l'intelligence du chercheur devra être mobilisée in fine pour vérifier la pertinence des indices de découverte.
- l'archéologue, l'historien pourrons grâce à ces méthodes affiner leur connaissance de très vastes corpus, découvrir des échanges commerciaux, culturels, diplomatiques sur de très vastes territoires et dans la diachronie de longues périodes historiques. Il est

[12] Repérer des pieds, des lignes de vers, des ensembles types (quatrains, tercets), voire qualifier des types de rimes (féminines, masculines, riches, pauvre, embrassées ou alternées, etc.)

certain que dans le cas de ces chercheurs en histoire, en littérature, a fortiori en archéologie nous courons peu de risque de tomber sur des utilisateurs naïfs des méthodes d'analyses statistiques automatiques.

Revenons sur la position de Bourdieu dans son rapport aux méthodes d'analyse factorielle de correspondances et celles de l'analyse des correspondances multiples. Cela nous permet de nuancer grandement les réflexions de Pierre Levy sur son blog, tout en reconnaissant par ailleurs la pertinence de la dichotomie relativement triviale qu'il pointe entre les usages industriels et en sciences exactes d'une part et ceux en sciences sociales et usages sociétaux de l'autre malheureusement largement exploité par les usages des big-datas à fin commerciale.

Ainsi Pierre Levy, identifie sur son Blog la situation des *Big Data* et de l'analyse des données à partir de deux aspects critiques importants. Il identifie d'abord deux sources de connaissances qui sont désormais en compétition sur Internet. La première, et la plus prolifique jusqu'à très récemment, est celle des domaines scientifiques et techniques comme la recherche en physique, en médecine ou en astronomie. La deuxième, considérée comme plus récente et de moindre visibilité dans l'univers numérique, est celle des sciences humaines et sociales ou ce qu'on appelle les « humanités numériques ».

L'analyse nous paraît un peu sommaire dans la mesure où, depuis les débuts mêmes de l'analyse des données, des sociologues, et non des moindres, ont su théoriser leur épistémologie d'usage de ces techniques statistiques.

Aujourd'hui, nombres d'études sociologiques des réseaux décrivent combien ces algorithmes non seulement imprègnent nos vies, mais aussi les modifient. Nous créons de données plus que jamais auparavant en utilisant Internet, nos Smartphones, des médias sociaux, en réalisant des transactions commerciales, en utilisant des appareils munis de capteurs multiples[13]. Notre univers quotidien des *Big Data* est le produit d'une infinité de faits, de produits, de livres, de cartes, de conversations, de références, d'opinions, de tendances, de vidéos, de publicités, de sondages, etc. Comme le signale le président exécutif de Google, Eric Schmidt, nous produisons chaque deux jours, la même quantité de données accumulées depuis le début de la civilisation humaine jusqu'à l'an 2003. L'Internet, et en particulier le World Wide Web, est de ce fait un substrat presque idéal pour l'émergence d'une intelligence distribuée qui couvre la planète par l'intégration des connaissances, compétences et intuitions de milliards de personnes à travers des milliards de dispositifs de traitement de données. Cette intelligence distribuée deviendrait de plus en plus puissante à travers un processus d'auto-organisation dans laquelle les personnes et les dispositifs s'échangeraient sélectivement les liens utiles. De nouvelles constructions de données liées

[13] Soulignons à ce propos un progrès en émergence qui peut donner le vertige : l'introduction de caméras d'aide au pilotage dans tous les véhicules neuf à l'horizon de 2020. Ces données seront directement utilisées pour la conduite, mais expédiées aussi sur le Cloud pour les expertises d'assurance en cas d'accident. Pour ceux qui militent pour la réduction des caméras de surveillance urbaines voilà qu'émerge une nouvelle catégorie de données autrement plus dérangeantes pour ce qui est de l'omni-surveillance.

jouent un rôle de plus en plus grand pour qu'émergent de nouvelles idées dans toutes les disciplines. Des méthodes créatives de visualisation des données font aussi fréquemment partie intégrante des processus de création de connaissances.

Comme on le constate, cette doxa doit être, répétons le grandement nuancée et critiquée de façon qui peut rester positive : toujours la posture positive de Bourdieu.

4 DECONSTRUIRE LE CONCEPT DE *BIG DATA* POUR MIEUX LE COMPRENDRE : EN APPREHENDER AUSSI SON « ETRE TECHNIQUE ».

Nous pensons qu'il est important de s'interroger sur les big-data dans une logique déconstructiviste qui constitue une pointe avancée de l'accompagnement philosophique de l'appropriation sociale de l'éclatement du champ communicationnel dès les années 60.

Il nous semble en effet que le concept de déconstruction constitue (de façon apparemment contradictoire) une approche conceptuellement très productive pour donner sens à l'éclatement provoqué par la mondialisation des réseaux, leur éclatement linguistique, leur diversité d'applications quotidiennes banales d'année en année plus large : éducation, domotique, aménagement du territoire, pilotage des véhicules individuels, contrôle médical individualisé, monétique, etc.

Dans tous ces domaines règne à la fois une sensation de maîtrise mondiale automatique de la fragmentation et du possible ré-appariement instantané et de l'autre une « désorientation évidente ». Les outils comme Google, Google translate, Wikipédia, hautement distincts dans leurs approches nous confortent dans la confiance de maitrise. Par contre, dès qu'on se penche pour considérer leur relativité ontologique, cela justifie nos sentiments de désorientation.

Face à ces contradictions, le « bricolage philosophique » est souhaitable. Dépassant le bricolage, et à l'ère de l'intelligence collective, mieux vaudrait encore attaquer ces problématiques en pratiquant systématiquement l'interdisciplinarité. C'est par exemple ce qu'a compris très tôt l'Université technologique de Compiègne en renforçant considérablement l'amalgame de philosophes (comme Bernard Stiegler notamment) dans ses équipes d'enseignement et de recherche.

Bernard Stiegler, par exemple a été et continue d'être un fantastique « passeur » d'une approche heideggérienne de l'être technique face aux êtres vivants, voire pensants. Il est aussi résolument lié à la philosophie de Jacques Derrida qui a été son directeur de thèse. Depuis plus de 25 ans on peut constater qu'il n'a jamais cessé de proposer des analyses de la technoculture en général, et de la technoculture numérique en particulier. Un des auteurs de cette communication doit beaucoup dans ses habitus de raisonnement aux « réflexes conceptuels » transmis sur le tas dans les travaux de préfiguration des NTIC à la BnF ou encore de l'Inathèque.

Heidegger, Simondon et Derrida constituent pour nous, les maillons fondamentaux d'une boite à outil philosophique de base nous permettant d'appréhender avec un relatif recul

philosophique l'irruption des big-datas et des analyses automatiques qui en sont indissociable.

L'être et le temps[14] du fait qu'il est le premier ouvrage d'Heidegger (1927), se lit de façon relativement aisée pour un non spécialiste. Sans obligatoirement tout comprendre un certain nombre de concepts heideggériens nous oblige impérativement à réviser notre vision triviale de la technique, de la modernité et pour ce qui nous concerne de la communication. Vu la date de parution de l'ouvrage[15] on peut s'étonner de la précocité de ses analyses qui pourraient nous apparaître comme spécialement conceptualisées pour avoir une vision critique des TIC.

Pour ne pas risquer de déformer le langage de spécialité de la philosophie et pour renouer avec une pratique de la Chaire ITEN-Unesco qui consiste à publier simultanément en ligne et sur papier des « 100 Notions sur des thématiques des TIC et du multimédia[16] » nous nous contentons ici de citer des extraits du lexique de Heidegger tel qu'il est disponible sur le « portail de la philosophie ».

Ce *Lexique de Heidegger* nous apparaît comme très opératoire dans la mesure où nombre de termes constituent un véritable réservoir de concepts parfaitement en phase avec notre modernité numérique. Nombre d'entrées de ce lexique Heidegger mériteraient qu'on les approfondisse en les rapportant aux questionnements actuels des TIC et notamment aux big-data. Prenons quelques-unes de ces notions qui nous ont parues très opératoires[17] :

Anwesen : « Entrée en présence ». Les choses futures ou passées font à leur manière mouvement dans le présent.
→ *Fondamental pour penser le numérique*

Ereignen, (Appropriation) un de ces termes intraduisibles mais fondamentaux du lexique heideggerien, d'où découlera le concept d'**Ereignis**. En première approximation signifie prendre appui sur *das Eigene* « comme mouvement d'amener une chose à son propre ». Nous ne sommes pas dans le registre notarial de la « propriété » mais plutôt dans celui de l'expression bien française de « remettre en main propre ». Il est nécessaire d'entendre toujours à travers le terme « approprier », non pas qu'une chose devienne la propriété, la possession, mais bien : « amener quelque chose à être ce qu'elle est ».
→*Ce terme serait sans doute fondamental pour interroger de façon approfondie des philosophes sur les questiosn de privacy, d'expropriation indolore et invisible de ce qui nous est propre dans le traitement des masses des données*.

---

[14] Martin Heidegger (trad. Rudolf Boehm et Alphonse De Waelhens), *L'être et le temps*, Paris, Gallimard, 1972 , 324 p.
[15] Cette période de production d'Heidegger présente aussi l'avantage d'avoir été conceptualisée avait l'irruption du Nazisme.
[16] www.100notions.com/
[17] https://fr.wikipedia.org/wiki/Lexique_Heidegger

Bewandtnis, (traduction Martineau « tournure » et traduction Vezin « conjointure »). Ce terme cherche à caractériser l'état ou l'essence d'un ustensile qui ne peut être ustensile ontologiquement à lui tout seul mais qui lui impose de se conjoindre à un autre pour satisfaire à un usage (exemple le bouchon à la bouteille, le bouton à la boutonnière).
→Cette notion peut-être aussi très féconde lorsqu'on la repense en lien avec les big data et les Data Analytics. En effet le traitement des big-data « dénie » en quelque sorte le statut d'intentionnalité préalable des data. Elles sont supposées pouvoir être utiles dans un futur, un temps et dessein qui ne sont pas ceux de ses premiers producteurs.

Dasein (Terme allemand polysémique fondamental d'*Être et Temps* avec pour traduction possible « *être-là* », ou « réalité humaine. Dans la deuxième partie de sa carrière Heidegger écrira *Da-sein* avec césure et trait d'union pour marquer l'évolution de sa conception de l'être, l'homme devenu moins configurateur de monde et plus « berger de l'être ».

Destruktion « Destruction », « Déconstruction », « Désobstruction »
→ Sur ces deux derniers concepts nous voyons bien combien ils traversent la totalité de ce qui nous importe pour penser notre société du post-numérique

Bedeutsamkeit : la significativité désigne la structure ontologique du monde en tant que tel. Le monde est présent comme une totalité de significations toujours-déjà ouverte en fait, à partir de laquelle se donne tout étant intra-mondain. Cette significativité en tant que structure ontologique n'est pas la somme des valeurs mais tout au contraire, une valeur, un rang, une signification particulière ne peut être donnée que dans le cadre d'une significativité d'un monde.
→ Sans prétendre dominer la totalité des notions contenues dans cette définition très spécifique à la discipline philosophique nous percevons bien cependant l'utilité d'un questionnement qui se focaliserait spécifiquement sur les big-data et notamment la pluralité des interprétations sémantiques.

Zeitlichkeit (temporalité) et ***Temporalität***, le temps « comme horizon possible de toute entente de l'être en général ». (L'allemand dispose de deux termes : le Temporel est le temps de l'histoire et des sciences, le Temporal le temps de l'être, à rapprocher d'Historial, l'histoire de l'Être. Le Dasein est à la fois temporel en prenant place dans le temps historique et temporal (traduit par *temporellité* par Vezin) en ce que cette temporalité ou *temporellité*, François Fédier donne quelques éclaircissements sur cette notion complexe de « Temporellité », qu'il définit « comme la manière qu'a l'être humain d'être temporel. Le temps qui est expérimenté se *tempore* au sein de la *temporellité* (se tempore signifie tout simplement déployer sa nature de temps) [...] La temporation est la manière dont le temps se tempore, c'est-à-dire la manière dont le passé est le passé, dont le présent est le présent et le futur le futur [...] c'est toujours au sein d'une temporellité déterminée que nous avons rapport à quoi que ce soit et en particulier aux choses du monde ».
→ La encore sans prétendre aucunement dominer ces notions nous pouvons entrevoir l'importance et l'urgence de débattre entre philosophes et spécialistes de l'infocom de ces notions très pertinentes.

Simondon, et la reprise très spécifique de l'être technique

Soulignons encore la notion d' « être » trop éclatée dans le lexique d'Heidegger ci-dessus mais qui renvoie au terme plus spécifique d'*être technique* travaillé conceptuellement par Simondon[18]. Pour le dire vite, la technique, l' « être technique » est de la pensée humaine mise en conserve qui devient *instanciable* à la demande dans une temporalité qui échappe à son (à ses) inventeur(s). Une roue permettra ainsi à ses utilisateurs de s'en servir sans avoir à refaire le laborieux chemin des innovations successives qui ont permis de dégager la roue primitive (pleine puis à rayons) et qui nous permettent aujourd'hui de disposer de quantité d'objets techniques reliés à ce premier outil : pneumatiques, rouages, rail, roulement à billes, hélice, etc. Simondon a ainsi particulièrement travaillé les questions touchant au rapport entre l'être technique et l'être humain. Ce rapport est celui de la distance obligatoirement prise par l'utilisateur d'un objet technique et son (ou ses) producteur-inventeur(s). Simondon va jusqu'à dire que l'homme se comporte vis à vis de la technique comme avec l'étranger. Il poursuit en soulignant que le travail philosophique que nous aurons à réaliser dans notre rapport avec les êtres techniques est comparable à celui qui avait dû être déployé en son temps pour accepter socialement l'abolition de l'esclavage. Enoncés en 1958 ces propos sont hautement pertinents lorsqu'on les rapporte aux questions posées par les big-data : qu'en est-il de notre rapport à la liberté ? à la transparence du recueil des données ? Au renoncement volontaire de nos droits de producteurs ou d'inventeurs ? Qu'en est-il de notre tolérance personnelle (volontaire ou involontaire) pour bénéficier en retour des facilités d'usage de ces nouveaux dispositifs ?

Derrida et la réinterprétation spécifiquement communicationnelle (grammatologique) de la dé-construction

La philosophie traverse mal les frontières et Derrida a été un des premiers francophones a re- travailler le concept heideggérien de *dé-construction* pour l'appliquer au domaine de la *grammatologie*. Ce deuxième concept, qui lui était plus personnel[19], permettait dès les années 60 de conceptualiser de façon très prospective et d'accompagner philosophiquement la « crise du sens » que la télématique et le multimédia amorçait alors. La renommée de Derrida devint très vite mondiale, s'associant à d'autres concepts comme celui de *rhizome* (Gilles Deleuze et Félix Guattari) ou celui *d'immatériaux* (Jean-François Lyotard). L'exposition des Immatériaux au Centre Pompidou en 1985 consacre d'ailleurs à Paris, le succès d'une mobilisation mondiale des sémiologues, des historiens des sciences, des architectes, des artistes, des muséographes d'arts[20] s'associant pour accompagner éthiquement et esthétiquement l'appropriation de la techno-culture numérique en devenir

---

[18] SIMONDON, Gilbert.. Du mode d'existence des objets techniques, Paris, Aubier 1989 (1958 date de sa thèse du même titre)

[19] A l'exception de Gelb qui fut l'inventeur anglophone du concept mais n'exploita pas sa découverte ailleurs que dans le domaine pointu de l'assyriologie.

[20] Citons entre autre Bruno Latour, Christine Buci-Glucksmann, Daniel Buren, François Châtelet, Hubert Astier Jacques Derrida, Jacques Roubaud, Jean-Claude Passeron, Jean-François Lyotard, Jean-Loup Rivière, Marc Guillaume, Mario Borillo, Michel Butor, Paul Caro, Robert et Sonia Delaunay.

alors très prometteuse mais qui bouleversait les fondements mêmes des anciennes certitudes communicationnelles :
« ... Depuis quelques temps [...] on disait « langage » pour action, mouvement, pensée, réflexion, conscience, inconscient, expérience, affectivité, etc. On tend maintenant à dire « écriture » pour tout cela et pour tout autre chose : pour désigner non seulement les gestes physiques de l'inscription littérale, pictographique ou idéographique, mais aussi la totalité de ce qui la rend possible ; puis aussi, au-delà de la face signifiante, la face signifiée elle-même ; par là, tout ce qui peut donner lieu à une inscription en général, qu'elle soit ou non littérale et même si ce qu'elle distribue dans l'espace est étranger et à l'ordre de la voix : cinématographie, chorégraphie, certes, mais aussi écriture picturale, musicale, sculpturale, etc. On pourrait aussi parler d'écriture athlétique et plus sûrement encore, si l'on songe aux techniques qui gouvernent aujourd'hui ces domaines, d'écriture militaire ou politique. Tout cela pour décrire non seulement le système de notation s'attachant secondairement à ces activités mais l'essence et le contenu de ces activités elles-mêmes. C'est aussi en ce sens que le biologiste parle aujourd'hui d'écriture et de *pro-gramme* à propos des processus les plus élémentaires de l'information dans la cellule vivante. Enfin, qu'il y ait ou non des limites essentielles, tout le champ couvert par le programme cybernétique sera champ d'écriture. À supposer que la théorie de la cybernétique puisse déloger en elle tous les concepts métaphysiques - et jusqu'à ceux d'âme, de vie, de valeur, de choix, de mémoire - qui servaient naguère à opposer la machine à l'homme, elle devra conserver, jusqu'à ce que son appartenance historico-métaphysique se dénonce aussi, la notion d'écriture, de traces, de gramme ou de graphème[21] ». Cette longue citation est pour nous emblématique d'une désorientation culturelle face au progrès des technologies numériques, qui n'a pas, bien sûr, attendu l'arrivée des big-data pour se mobiliser.

Associant leurs efforts à ceux de L'École de Francfort[22], mais à bien d'autres chercheurs sémiologues, économistes, sociologues, artistes, muséographes ou architectes on perçoit bien l'importance de cette réflexion multidisciplinaire sur les technologies de l'information et de la communication. Ce sont là, pour nous, des éléments de base, ou plutôt des pistes ouvertes pour affirmer l'urgence d'une mobilisation conjointe et multidisciplinaire entre les philosophes et les spécialistes de l'information et de la communication pour penser les big-data.

5 HUMANITES NUMERIQUES & BIG DATA : EXPLORER LES VOIES DE L'EDUCATION

Le monde de l'éducation s'intéresse particulièrement à l'émergence des Big-data et des Data analytics. Dans cette communication nous rendrons compte de deux aspects de cette question :

[21] DERRIDA (Jacques), De la grammatologie, p. 19
[22] Un courant de pensée qui compte de nombreux intellectuels Theodor W. Adorno (1903-1969), Walter Benjamin (1892-1940), Marcuse (1898-1979), plus tard Habermas. Contraint d'émigrer par le nazisme beaucoup iront aux États-Unis, fondant notamment le Groupe de Palo Alto.

- l'approche institutionnelle et celle des technologues qui s'intéressent bien naturellement à ce qui est directement opératoire : l'analyse et l'exploitation des profils d'apprenants (ou d'autres acteurs de l'enseignement et de la formation)
- - une approche plus expérimentale, liant enseignement, recherche et patrimoine et qui vise à explorer l'analyse proprement dite de productions de patrimoines en mettant en synergie enseignement et recherche : notre projet HD Muren

L'approche institutionnelle prioritaire : exploiter les profils des apprenants

Les institutions d'éducation et de la communauté des technologies éducatives s'intéressent prioritairement à l'exploitation des données sur les profils des acteurs (surtout les apprenants, mais aussi les enseignants et les institutions d'enseignement).

Les organisations éducatives se rendent compte, en effet, qu'il y a un intérêt stratégique à profiter des progrès technologiques apportés par les *Big Data*. De nombreuses expériences sont menées à travers le monde qui presque toutes analysent les modèles de comportement des apprenants en situation d'apprentissage en ligne. Ce nouveau concept de gestion des flux d'apprentissages, est désormais connu sous le nom de *Learning Analytics*. Ce champs d'utilisation des Big-data a pleine légitimité épistémologique à condition d'être mené en pleine connaissance éthique et méthodologique par des institutions qui, de par leur position académique, ne doivent pas (ne devraient pas) transiger avec ces principes déontologiques qui constituent les fondements mêmes de la transmission du savoir. Ces expériences, et déjà aussi ces mises en œuvre en grandeur réelles, nous apportent déjà quantités d'informations très précieuses sur ce que les apprenants apprennent (et comment ils l'apprennent) et symétriquement sur ce que les enseignants enseignent (et comment ils l'enseignent). Les *Data Analytics* en éducation permettraient ainsi (nous permettent déjà) de prendre des décisions plus éclairées sur les programmes d'apprentissage et d'identifier leurs défauts de conception. Rien que de très légitime puisqu'il peut être question d'heuristique : définir une meilleure stratégie pédagogique parmi des infinités de possibles.

Cette communauté de pratiques s'intéresse aussi aux problèmes éthiques ou déontologiques : quels filtres appliquer à ces *Big Data* pour que la sécurité des acteurs - tant personnes physiques qu'institutions scolaires - soit garantie. Elle s'intéresse aussi dans une moindre mesure aux données nouvelles qui peuvent être induites grâce à l'exploitation analytique de l'accumulation exponentielle de ressources d'enseignement (avant tout des profils d'apprentissage et d'enseignement) mais s'intéresse paradoxalement beaucoup moins aux *Big Data* potentiellement originales qui pourraient être induites par l'analyse de masse d'exercices et de productions d'apprenants.

L'analyse de masse des contenus générés non seulement par les cours, mais par les exercices cumulés des étudiants ou des élèves.

C'est là, en effet, l'une des caractéristiques des *Big Data*, celle de pouvoir induire des idées nouvelles en analysant statistiquement des quantités de données souvent banales et redondantes mais quelquefois originales et innovantes. Profitant de la banalisation progressive de l'informatique de pointe, des capacités de traitements linguistiques, de

l'intelligence artificielle, du traitement automatisé des langues, des ontologies et des réseaux sémantiques, etc. il est de plus en plus possible de déployer des technologies avancées d'analyse de données pour détecter de l'originalité et de la valeurs ajoutées dans les ressources numériques.

L'indispensable normalisation du domaine

Rappelons que le potentiel réel des *Data Analytics* n'est pas encore totalement optimisé pour répondre aux promesses escomptées des *Big Data*. La majeure partie de ces secteurs d'activités en sont encore au niveau de l'exploration et de l'appropriation de base. La phase de maturité technologique des technologies du *Big Data*, ne pourra se déclencher avant l'adoption de normes d'interopérabilité du domaine. De par l'interopérabilité qu'elle confère aux données de masse, la normalisation accroît en effet considérablement le potentiel d'utilisation autonome et l'évaluation comparative des technologies des *Big Data*. Elle commence à être pratiquée dans plusieurs domaines d'activités stratégiques. En éducation, les normes du *Learning Analytics* prennent de plus en plus de place dans les processus normatif de l'interopérabilité des systèmes et des dispositifs pédagogiques. Cependant, la démarche normative requiert un processus lent et complexe. Ainsi, les organismes internationaux de normalisation des technologies de l'éducation (notamment l'ISO-IEC JTC1 SC36)[23] se sont déjà saisis de cet enjeu normatif et des normes d'interopérabilité des Learning Analytics sont déjà en chantier.

6 HD-MUREN, UN PROJET AU CROISEMENT DE L'INNOVATION PEDAGOGIQUE ET D'UN RENOUVEAU HUMANISTE NUMERIQUE

C'est précisément sur ce créneau particulier que nous prévoyons d'orienter ultérieurement notre projet HD-MUREN en exploitant des *Big Data* d'apprentissage dans le cadre d'interopérabilité que lui confère la TEI : le cadre de production unique et standardisé, d'HD Muren. C'est aussi parce qu'ils s'inscrivent dans un cadre de pédagogie ouverte[24] que les apports des apprenants de HD-MUREN (mais potentiellement de quantités d'autres projets) sont pris au sérieux.

[23] Plusieurs auteurs de cette contribution sont impliqués comme experts ISO dans cette instance. On consultera utilement les contributions de Jon Mason (Australie) *Data, Data Everywhere : open, linked, interoperable* et de Tore Hoel *An exploration of standardisation options for the new field of learning analytics* à l'Open Forum Initiatives 2015 à l'occasion de la Plénière du SC36 à Rouen en Juin 2015 (dont la Chaire Iten-Unesco était partenaire). Prochainement mis en ligne sur le site de l'AUF, Initiatives 2015.

[24] Cadre de pédagogie ouverte peut (et doit) être compris dans les deux acceptions du terme :
- une méthode d'éducation dite ouverte (telles les pédagogies Montessori, Decroly ou Freinet) par rapport aux méthodes dites fermées ou traditionnelles.
- un cadre similaire tel celui des *open universities* c'est à dire une pédagogie numérique à distance ouverte sur la diversité des utilisateurs : de tout âge, sans qu'il soit nécessaire de passer un examen ; ouvertes sur ceux qui ne peuvent pas fréquenter une université traditionnelle.

L'idée *princeps* du projet HD-MUREN, proposé dans le cadre de la Chaire ITEN-Unesco-Paris 8 et qui s'inscrit dans le cadre plus vaste d'Idéfi-Créatic[25], consiste à utiliser des cohortes volontaires de lycéens ou d'étudiants en première ou deuxième année pour leur faire réaliser de façon participative des corpus numériques à partir de textes littéraires, puis dans une deuxième phase d'humanistes français et arabes. Pendant cette étape initiale, ces lycéens ou étudiants sont impliqués dans des projets d'enseignement-recherche ce qui leur permet d'approfondir leurs compétences de balisage littéraire savant (de poèmes par exemple), d'affiner la description métrique, d'analyser des figures de stylistiques etc. En associant en synergie un enseignant et un chercheur (ou un groupe de chercheurs), l'enseignant peut ensuite impliquer les lycéens ou les jeunes étudiants dans des processus plus complexes par exemple pour le repérage des variantes et des incertitudes d'interprétation dans un *apparat critique*[26] désormais numérique, la description savante de manuscrits et leur mise en parallèle avec leur transcriptions, la mise en parallèle de traductions, etc.

Par ce procédé, nous élargissons le champ de l'interprétation (propre aux sciences humaines) en l'associant à une activité parallèle de production participative de ressources patrimoniales numériques en milieu scolaire qui renforcerait le processus d'interprétation et d'acquisition de valeurs humanistes chez les apprenants. Notre approche se fonde sur l'idée que l'acte pédagogique ne doit pas se contenter d'être uniquement un vecteur de transmission de savoirs, mais doit être aussi un incubateur, mais surtout une *fabrique* de valeurs citoyennes et civiques susceptibles même, dans certains cas, d'élargir l'objet enseigné. La production de données en Humanités serait ainsi dédoublée, voire aussi augmentée d'un processus interprétatif proactif permettant à l'apprenant d'acquérir de manière subliminale des valeurs citoyennes et humanistes véhiculées par les données qu'il traite.

Pour tester et expérimenter ces assertions et hypothèses, notre projet de recherche a une double articulation. Au-delà de la façade des compétences technologiques à expérimenter dans un cadre scolaire, il vise à évaluer chez des jeunes apprenants non seulement leurs pratiques numériques, mais leur capacité à s'approprier des ressources humanistes « ouvertes », notamment des contenus savants, littéraires ou philosophiques.

Notre idée de production numérique participative (*crowdsourcing)* vise à répondre à l'enjeu stratégique de pouvoir disposer en libre accès d'un réservoir de ressources numériques et culturelles massivement ignorées par les acteurs marchands du Net (notamment Google). On se rend compte en effet que si des acteurs importants comme Gallica mettent à disposition de nombreuses ressources culturelles, celles-ci sont souvent soumises à beaucoup de restrictions dues aux droits d'auteurs, aux volumes des données prohibitifs pour les réseaux à bas débits, aux restrictions concernant l'impression et la sauvegarde, etc.

[25] idefi-creatic.net

[26] Cette notion rarement étudiée dès le niveau du secondaire et pourtant facilement compréhensible par des digital natives dans la mesure où de multiples dispositifs logiciels (dans les jeux, dans les réseaux sociaux) utilisent des mécanismes logiques et hypertextuels similaires.

Par la structuration numérique des œuvres humanistes en employant les techniques de la *Text Encoding Initiative* (TEI), nous inscrivons le projet dans la dynamique internationale de l'*Open Source* et *Open Data* et de l'accès universel aux savoirs.

Notre démarche nous paraît légitime parce qu'elle nous semble s'inscrire dans une logique de mobilisation savante similaire à certains aspects de la mutation humaniste de la Renaissance et conforme aux Humanités numériques aujourd'hui. Pour nous (et bien d'autres chercheurs), l'Humanisme numérique n'est pas seulement un type de contenu, ni même seulement une posture morale et philosophique (la médiation numérique de la croyance en l'Homme). C'est aussi une méthode : permettre une circulation mondiale et interopérable de corpus de documents parce qu'ils sont XMLisés, normalisés, structurés et balisés selon des schémas normalisés donc en TEI. Cette détermination de méthode est explicitement inscrite dans les « Principes de Poughkeepsie » qui sont très largement repris dans le « Manifeste des Digital Humanities » (Dacos, 2010). Elle est aussi inscrite dans un long processus historique itératif bien connu depuis que l'on avait commencé à parler du codage binaire de l'information et de la révolution numérique, de l'explosion informationnelle et de l'info-obésité. Croire en « l'homme » c'est être certain qu'à chaque nouvelle étape de ce processus itératif, à chacun des virages technologiques qui ont marqué des changements successifs dans le paysage informationnel mondial, une mobilisation générale s'est toujours mise en place pour inventer, innover, implémenter et adapter des solutions et des méthodes permettant de maîtriser et d'apprivoiser les flux débordants de données.

7 BIBLIOGRAPHIE

BABINET, G. (2015). *Big Data, penser l'homme et le monde autrement.* Le Passeur Editeur.

BENZEKRI (J.-P.), *L'Analyse des données*, Paris, Dunod, 1976.

BOURDIEU, P., *La Distinction*, Paris, éd. Minuit, 1979

COINTOT, J.-C., & EYCHENNE, Y. (2014). *La Révolution Big Data : Les données au coeur de la transformation de l'entreprise*. Paris, Dunod.

COPERNIC, N., *Commentariolus*, et RHETICUS, J., *Narratio prima*, in Introductions à l'astronomie de Copernic (trad., intro. et commentaire H. Hugonnard-Roche, E. Rosen et J.-P. Verdet, éd. Albert Blanchard, Paris, 1975).

CUKIER, K., & MAYER-SCHOENBERGER, V. (s. d.). *Big Data : La révolution des données est en marche.* Robert Laffont/bouquins/segher.

DERRIDA, J. (1967). *De la grammatologie*. Paris : Editions de Minuit.

DOZO, B.O., *Données biographiques et données relationnelles, Notes théoriques pour une utilisation complémentaire des outils quantitatifs*.in la revue « Contextes » 3/ 2008 : Questions biographiques en littérature consulté sur :

https://contextes.revues.org/1933?lang=en

EISENSTEIN (E. L.), *La révolution de l'imprimé à l'aube de l'Europe moderne*, Paris, La Découverte, 1991.

GELB, J. I., *A study of writing, the foundations of grammatology,* Chicago, The University of Chicago Press, 1952, p. 28. Traduction française : Pour une théorie de l'écriture, Paris, éd. Flammarion, Collection « idées et recherches, 1973, rééd. 1992.

HAVELOCK, E. A. (1981). *Aux origines de la civilisation écrite en Occident*. Paris : La Découverte.

HEIDEGGER, M. (2014). *Phénoménologie de l'intuition et de l'expression : théorie de la formation des concepts philosophiques*. (G. É. scientifique Fagniez, Trad.). Paris, France : Gallimard, impr. 2014.

HUDRISIER, H., AZEMARD, G., BEN HENDA, M., DIWERSY, S., LEHMANS, A., LIQUETE, V., ROMARY, L., *Synergie enseignement-recherche pour l'aménagement numérique structuré (TEI) de patrimoines littéraires multilingues et multiculturels*, in Colloque COSSI, Université de Montréal, Juin 2015, (à paraître)

LEVY P. (2014). *Collective intelligence, big data and IEML*. (s. d.). Billet de blog du 14 novembre 2014. Consulté à l'adresse http://pierrelevyblog.com/2014/11/14/collective-intelligence-big-data-and-ieml/

MICHELET (Jules) *Histoire de la Révolution française* (tome I, 1847; tome II [1789-1791], 1847; tome III [1790-1791], 1849; tome IV [1792], 1850 ; tome V [1792-1793], 1851 ; tomes VI et VII [1793-1794], 1853. Paris, Réédition Gallimard, coll. Bibliothèque de la Pléiade, 1939 (2 tomes)

SAPIN, C., (2014, 4 janvier 4). *Quand les politiques se mettent au ''Big Data''*. L'Opinion. Consulté le 20 janv. 2015. http://www.lopinion.fr/4-janvier-2015/quand-politiques-se-mettent-big-data-19954

SIMONDON, G., *Du mode d'existence des objets techniques*, Paris, Aubier, 1958; dernière réédition corrigée et augmentée, Flammarion, 2012

STIEGLER, B., *La Technique et le temps*, tome 1 : *La Faute d'Épiméthée*, 1994 (ISBN 2718604409)

STIEGLER, B., *La Technique et le temps*, tome 2 : *La Désorientation*, 1996 (ISBN 2718604689)

STIEGLER, B., *Échographies de la télévision*, *entretiens filmés avec Jacques Derrida*, 1996 (ISBN 2718604808)

STIEGLER, B., *La Technique et le Temps*, tome 3 : *Le Temps du cinéma et la Question du mal-être*, 2001 (ISBN 2718605634

TESTARD-VAILLANT P. (2014). « Interview de Michel Wieviorka : les sciences humaines et sociales à l'ère numérique : 10 janvier 2014 ». Paru dans CNRS Le Journal du 02/04/2014 *Big dtata, la déféerlante des Octets.* https://lejournal.cnrs.fr/articles/interview-de-michel-wieviorka-les-sciences-humaines-et-sociales-a-lere-numerique