

# Le nominalisme et la culture : questions posées par les enjeux du numérique

Bruno Bachimont

Ingénieur civil des mines de Nancy en 1985, docteur en informatique de l'université Paris l'université Paris VI en 1990, docteur en épistémologie de l'École polytechnique en 1996, Bruno Bachimont rejoint en 1998 l'Institut national de l'audiovisuel (Ina) où il dirige la recherche. En 2002, il devient enseignant-chercheur à l'université de technologie de Compiègne où il assume la direction à la recherche depuis 2006. Il fut conseiller scientifique à l'Ina de 2002 à 2012. Ses recherches portent sur 3 objets : une philosophie de la technique et du numérique, pour comprendre en quoi nos connaissances et modes de pensée sont constitués, reconfigurés par la technique de manière générale et le numérique en particulier ; une approche du document et de la mémoire, pour articuler les possibles de la technique (ingénierie documentaire, préservation numérique) avec une phénoménologie de la mémoire et du passé ; et enfin les ontologies formelles, pour comprendre comment expliciter et formaliser les cadres conceptuels de notre pensée.

Le numérique est une technologie qui a eu pour conséquence de permettre un traitement technique généralisé et industrialisé des contenus culturels et intellectuels. Cette évolution prend sa source dans plusieurs phénomènes. D'une part, tous les contenus par lesquels nous nous exprimons et communiquons sont désormais pris en charge par un format numérique et peuvent ainsi recevoir un traitement calculatoire. Si on entend par contenu tout objet physique pour lequel on rencontre l'union d'un support matériel de manifestation et une forme sémiotique d'expression, le numérique devient un support universel pour tout ce qui revêt une forme sémiotique faisant signe et sens pour nous : des textes, aux images et aux sons, en passage l'audiovisuel, le numérique n'en finit pas d'intégrer dans sa technicité les contenus de la culture. Mais d'autre part, la nature désormais numérique des contenus permet de les soumettre à des traitements calculatoires et automatiques. Les contenus ne trouvent pas seulement dans le numérique un moyen commode d'inscription et de communication, le numérique permettant comme on le sait des transmissions quasi instantanées et proposant par ailleurs des outils de production et de consultation faciles à mettre en œuvre. Le numérique apporte en outre la calculabilité des contenus et la capacité de les transformer. On peut même dire que les propriétés d'inscription (production, consultation) et de transformation reposent en fait sur la capacité de transformation des contenus, chaque opération étant *in fine* un calcul. Or, contrairement aux supports anciens, les traitements sur le contenu peuvent entièrement être délégués à la machine sans intervention humaine et donc sans médiation interprétative.

Support universel d'inscription, support pour des transformations automatiques, le numérique réifie les contenus culturels pour en faire des objets manufacturés que l'on peut soumettre à des processus automatisés et industriels. Le numérique appartient à cette classe de support permettant la reproductibilité technique. Mais, à la différence de ce concept proposé par Walter Benjamin<sup>1</sup>, on ne parle pas ici seulement de reproductibilité, mais aussi de traitement. Non seulement le contenu peut se reproduire de manière automatisée, mais sa transformation est elle-même programmable.

Le numérique entraîne donc cette conséquence paradoxale de contenus qui peuvent se produire de manière automatisée, se transformer, voire se lire de manière machinique. En effet, la plupart des

---

1. Walter Benjamin, *L'œuvre d'art à l'époque de sa reproductibilité technique : version de 1939*, Folio, 2008.

données produites aujourd'hui sont le fait des machines ; de même, la plupart des contenus sont désormais consultés, indexés, archivés par les machines, pour n'être finalement humainement lus que pour une part infime d'entre eux. Le web est l'exemple paradigmatique de cette situation où l'indexation ne peut être faite que par la machine, mais n'est pas l'exemple unique : c'est désormais le cas de tout entrepôt de données, d'archives numériques, de fonds documentaires où la masse collectée exige désormais une médiation instrumentée.

Pour une part, cette situation est la suite logique du processus général de grammatisation auquel on assiste depuis les origines de l'Homme, où l'expression de son rapport au monde se manifeste à travers des supports matériels qui permettent d'objectiver sa représentation du monde. Cette objectivation correspond à la construction d'un objet de pensée, une prise de conscience rendue possible par une telle objectivation car présentée à la conscience. C'est ainsi que l'écriture, manifestant la langue via un support spatial permettant le regard et non l'audition, rend visible des structures et rapports inaudibles dans la parole proférée ; c'est ainsi dit-on que la grammaire comme représentation structurée de la langue devient possible, car elle dispose grâce à l'écriture d'un outil d'observation de celle-ci<sup>2</sup>. Mais l'objectivation est aussi la constitution d'un objet matériel manipulable : doté d'un support matériel d'origine technique, il se prête à la manipulation. L'objectivation est alors la constitution d'un ordre objectif, celui du contenu culturel matérialisé, qui peut se transformer et se manipuler par les outils techniques travaillant à même le support. À la fois prise de conscience d'un ordre cognitif et constitution d'un ordre matériel, la grammatisation trouve dans le numérique un aboutissement cohérent de sa dynamique.

Mais pour une autre part, le numérique fait rupture dans le processus de grammatisation. En permettant une manipulation automatique du contenu, le numérique fait sortir le contenu culturel hors du contrôle interprétatif où était soumis jusqu'à présent tout contenu grammatisé. En effet, que ce soit un texte recopié (les copistes du Moyen Âge), un texte annoté, un texte structuré, la manipulation du support était l'effet d'un esprit agissant soumettant à son contrôle interprétatif la transformation effectuée. Ce contrôle pouvait bien sûr être fautif, et la copie s'assortir d'erreurs comme le savent bien les philologues. Progressivement, certaines opérations se sont extraites du contrôle interprétatif, notamment la reproduction qui devint mécanisée avec l'imprimerie comme on le sait, et qui fut au principe même des enregistrements sonores, filmiques et audiovisuels. Mais comme on l'a dit, en passant d'une reproductibilité mécanique à une transformation sortant elle-même du contrôle interprétatif, c'est un autre statut que l'on confère au contenu, dans ce qu'il dit, exprime ou représente.

Cette mutation de statut consiste à transformer le contenu d'expression d'une pensée ou d'une intention en un fait constaté qui se traite désormais à travers des outils de collation, comparaison, transformation. Les machines écrivent pour d'autres machines qui vont les lire. L'inscription ne reflète pas une pensée qu'elle exprime mais le processus qui l'a produite. De même, sa lecture repose sur le constat de sa présence (indexation) et l'analyse de sa structure (inférence), opérations effectuées sur l'inscription comme suite structurée de symboles physiques la composant.

De ce fait, le contenu n'est plus une expression culturelle à interpréter, mais un fait à analyser. Ce qu'on appelle contenu, et la signification qu'on lui attribue, n'est qu'un nom arbitraire et vide de sens auquel il faut substituer une analyse mettant en rapport ce contenu à son contexte de production et à son environnement. De même que le nom que nous donnons en langue naturelle aux objets de notre environnement ne permet en rien de les connaître, de même la compréhension que nous pourrions avoir d'un contenu ne permet plus de l'interpréter et de connaître le fait culturel dont il serait l'expression. De même que la mesure expérimentale est ce qui permet de comprendre le lien entre nos concepts et la nature, l'analyse automatisée des contenus est ce qui permet de comprendre les faits de la culture, au-delà d'une compréhension qu'il faudra qualifier de naïve des contenus.

---

2. Sylvain Auroux, *La révolution technique de la grammatisation*, Éditions Mardaga, coll. Philosophie et langage, 1995.

Ainsi voit-on se mettre en place, à l'instar de la nominalisation de la nature qui s'est produite du XIV<sup>e</sup> siècle jusqu'à la révolution galiléo-cartésienne, une nominalisation de la culture qui bouleverse à la fois notre rapport à la culture et aux représentations que nous en avons.

## 1– Le nominalisme

### • Nominalisation de la nature : le nominalisme médiéval

Le nominalisme est cette école de pensée, dont on connaît les héros/hérauts principaux à travers des figures comme Guillaume d'Ockham, Jean Buridan, Jean de Gerson, etc., qui consiste à considérer que les noms généraux de catégories comme « humanité », « animalité », ne renvoient à aucune entité existante. Il n'existe que des entités singulières, les noms généraux comme humanité n'étant alors que des *signes* permettant de se référer à ces dernières de manière collective, mais sans avoir pour autant d'entité associée comme une humanité qui serait une partie ontologique de chaque être humain. Le nominalisme apporte donc une solution *sémiotique* et non plus *ontologique* au problème de traiter les termes généraux.

Le nominalisme s'oppose au réalisme qui considère que de telles entités existent, non pas de manière séparée ou autonome, mais au sein des individus qu'elles désignent et constituent<sup>3</sup>. Le réalisme repose *in fine* sur l'idée que les structures fondamentales du langage renvoient à la structure du réel, la logique du discours étant, correctement étudiée, une logique du réel<sup>4</sup>. La langue est un accès au réel et son intelligibilité est une intelligence de la nature.

La révolution nominaliste a donc rompu le lien ontologique entre la langue et la nature, laissant ainsi la place pour qu'une nouvelle relation s'établisse entre notre concept et la connaissance de la nature : cette relation sera la mesure expérimentale exprimée dans l'idiome mathématique, permettant ainsi à la légalisation formelle d'exprimer la structure ontologique du monde. Ce que Galilée saura exprimer en disant que la philosophie est écrite dans ce grand livre qui se tient constamment ouvert devant les yeux (je veux dire l'Univers), mais elle ne peut se saisir si tout d'abord on ne se saisit point de la langue et si on ignore les caractères dont elle est composée. Cette philosophie, elle est écrite en langue mathématique ; ses caractères sont des triangles, des cercles et autres figures géométriques, sans le moyen desquels il est impossible de saisir humainement quelque parole ; et sans lesquels on ne fait qu'errer vainement dans un labyrinthe obscur.

### • Nominalisation de la culture : la révolution contemporaine du numérique

Le numérique inaugure une semblable révolution. En effet, les outils numériques permettent de constituer de grandes bases de données permettant l'utilisation sur leur globalité d'outils en général de nature statistique où des phénomènes sinon indécélables peuvent être mis en évidence. C'est par exemple l'apparition d'un terme nouveau dans un corpus diachronique de journaux, l'observation épidémiologique d'un virus en suivant les sujets traités par la presse locale, etc. Les outils ainsi mis en œuvre permettent de porter un nouveau regard et d'objectiver, au sens explicité plus haut, de nouveaux phénomènes.

Mais l'innovation ne se limite pas à la capacité accrue de détecter un nouveau fait, de nouveaux phénomènes. Les outils permettent d'étudier l'ensemble des données comme un tout, et de donner des propriétés du corpus qui ne sont le fait d'aucune donnée en particulier. Un exemple particulièrement explicite de cet aspect est donné par les différentes cartographies, notamment du web, qui montrent la façon dont s'organise le corpus, les liens entre ses différentes composantes pour finalement permettre de comprendre la dynamique de l'ensemble comme tout. À ce propos, Lev Manovich a proposé de

---

3. Christophe Erismann, *L'homme commun : la genèse du réalisme ontologique durant le haut Moyen Âge à Paris*, Vrin, 2011.

4. Ruprecht Paqué, *Le statut parisien des Nominalistes*, PUF, 1985.

parler de *Cultural Analytics*<sup>5</sup>, où trois étapes structurent tout travail de ce type : collecte des données, analyses statistiques, visualisation des résultats.

Le contenu documentaire intégré dans le corpus n'est plus considéré comme un contenu, mais comme un fait, une donnée syntaxique et symbolique, indépendamment de toute interprétation qu'on pourrait en faire, que l'on peut alors comparer et mesurer aux autres composantes du corpus. Son intelligibilité n'est plus acquise par la lecture et l'interprétation qu'on peut en faire, mais découle désormais de la contribution qu'il apporte comme fait à l'analyse globale du corpus auquel il appartient.

Mais pour exploiter les résultats, il est nécessaire de les visualiser et de les montrer. C'est ainsi que les analyses des bases de données se complètent d'interfaces de visualisation où l'utilisateur est invité à interpréter les résultats sur la base de ce qui est montré. À l'instar des cartographies effectuées, l'interprétation est alors essentiellement perceptive et qualitative, reposant sur notre capacité à donner du sens aux formes montrées.

On s'aperçoit que le nouveau *modus operandi* repose donc sur deux démarches complémentaires fondées sur le refus de l'interprétation directe du contenu comme moyen d'accéder à la compréhension d'un fait de culture ou phénomène socioculturel.

- Premièrement, l'exploitation du fait à travers un corpus travaillé de manière statistique. Le fait n'a plus d'intelligibilité propre, mais seulement via les propriétés globales du corpus auquel il appartient.

- Deuxièmement, les résultats sont montrés à travers des interfaces de visualisation où l'intelligence des phénomènes socioculturels découle d'une interprétation perceptive et qualitative des visualisations montrées.

D'une certaine manière, le rapport linguistique au monde de la culture, la langue et l'expression du contenu étant ce qui permet de le comprendre, est remplacé par un rapport perceptif et qualitatif. Si on a bien un nominalisme qui disqualifie le réalisme linguistique où la langue est un instrument de connaissance, on lui substitue un réalisme de la perception dont la critique reste à faire.

## 2– De nouvelles représentations de la culture

La nominalisation de la culture telle que nous venons de la décrire présente différentes difficultés qui ne la disqualifient pas mais qui invitent à l'approfondir pour qu'elle devienne un paradigme scientifique cohérent et tenable. Le premier ordre de difficultés est d'ordre épistémologique : dans quelle mesure les traitements effectués sur les données permettent-ils de proposer une analyse scientifique du réel culturel ? Nous verrons que les deux piliers de ce nominalisme sont problématiques. Le second ordre renvoie plutôt à des considérations phénoménologiques où la question est de savoir ce que cela veut dire que d'avoir une compréhension de la culture. Nous verrons que la langue possède une dimension inéliminable, et que l'enjeu est bien, derrière les analyses menées, de retrouver une compréhension de la culture comme monde humain, habité par nos alter egos.

### • Un nouveau paradigme scientifique : problèmes épistémologiques

La constitution de ces grandes bases de données ouvertes à l'exploration systématique présente comme on l'a dit deux difficultés épistémologiques notoires : le traitement effectué sur les données lors de leur mise en corpus d'une part, l'interprétation de leur visualisation d'autre part. Ces deux

---

5. Le terme « cultural Analytics » a été inventé par Lev Manovich en 2007.

problèmes sont bien connus mais ne sont pas pour autant faciles à traiter ni explicitement abordés dans le monde des grandes bases de données, ou *big data*.

La première difficulté concerne donc la constitution même du corpus : pour permettre le traitement statistique ou tout simplement l'application d'outils de recherche (fouille de données par exemple), les données sont formatées. Si bien que les informations d'origine, associées à leur format initial, sont souvent oubliées ou transformées. La collecte des textes en fait des grandes bases de mots ou de phrases, mais en perdant la structure textuelle, la mise en forme graphique et d'autres attributs qui sont pourtant porteurs de sens et qui conditionnent par ailleurs le sens des composantes que l'on garde. En faisant des textes des sacs de mots, on extrait bien des informations sur l'emploi des mots, mais quel sens donner à ces extractions si pour cela on a dû « oublier » ou négliger la position du mot dans le texte et son rôle ? De même, l'audiovisuel propose des programmes très structurés : les rassembler pour apprendre des motifs de structure qu'on a effacés en constituant des corpus qui ne sont plus que des successions d'images en perdant l'information structurelle, est-ce bien pertinent ?

Plus profondément, on s'aperçoit que la constitution du corpus contribue à couper le lien entre les données d'une part et la réalité sémantique qu'il désigne ou signifie. Cela signifie que, dans un geste théorique profond, mais discutable, cette dimension sémantique est abandonnée pour retrouver une sémantique des données qui ne sera plus donnée que par le traitement global.

Il est peu probable qu'on retrouve pourtant cette sémantique : on construira plutôt d'autres dimensions sémantiques des données, jetant une autre interprétabilité sur ces dernières. Mais la difficulté est précisément de savoir quel statut donner au résultat des traitements effectués ? Quelle interprétation en faire ? On rencontre alors le second problème que nous évoquions.

La visualisation des données propose de montrer les résultats sous une forme graphique et qualitative. Si bien que tout utilisateur se retrouve en position de donner un sens à ce qu'il voit. Mais quel statut possède ce sens ainsi attribué ? Par exemple, comment interpréter la proximité spatiale entre des informations montrées sur une carte : la distance reflète-t-elle un artefact de l'algorithme de placement, ou une propriété des données ? Le caractère topologique de la distance reflète-t-il bien également une propriété des données qui auraient alors ces relations topologiques entre elles ? Souvent, le pouvoir de suggestion de cartes ou visualisations montrées se fait au détriment de l'intelligibilité des données : l'interprétation reste un acte d'expertise mobilisant à la fois la connaissance des données traitées et de leur contexte, et la connaissance des outils de traitement et de visualisation.

Trop souvent, la qualité des interfaces est jugée sur leurs performances ergonomiques, non sur la valeur épistémique des conclusions qu'elles permettent de tirer sur les données ainsi montrées. Les visualisations sont des métaphores graphiques, non des modèles, au sens où les propriétés de ce qui est *vu* ne permet de remonter aux propriétés de ce qui est *montré*.

#### • Un nouveau rapport à la culture : problèmes phénoménologiques

Mais le problème n'est pas seulement épistémologique. En coupant les données de leur régime originaire d'interprétation, on les coupe de leur terrain culturel où elles prennent sens. En effet, si l'on s'accorde à considérer que les sciences de la culture étudient le fait humain dans toutes ses dimensions, le traitement ainsi apporté est trop important ou trop sommaire.

En effet, étudier un fait humain, qu'est-ce que cela signifie ? Comme le formule suggestivement Marc Bloch dans son *Apologie pour l'histoire*, le scientifique de la culture traque le fait humain comme l'ogre des histoires enfantines : il flaire l'humanité, là est son gibier. Or l'humanité d'un fait, celle qui fait qu'un fait humain est bien *humain*, consiste en ce que ce dernier est un événement qui survient à un alter ego, qui aurait pu m'arriver et qui par conséquent prend sens pour moi malgré l'altérité de la distance culturelle et temporelle. L'historien et l'anthropologue, pour prendre ces deux figures des scientifiques qui étudient respectivement la distance temporelle et la distance spatio-culturelle, pratiquent cette herméneutique particulière d'une construction théorique de l'Autre à partir

d'un investissement à partir de soi et du présent<sup>6</sup>. Autrement dit, l'intelligibilité du fait culturel repose sur l'empathie qui permet de construire la mise à distance à partir d'une assimilation à ce qui aurait pu m'arriver, et en comprenant que justement, cela ne m'est pas arrivé et ne pourra pas advenir : que ce soit le contexte qui diffère, ma manière de penser qui l'interdit, les limites du pensable et du faisable, l'assimilation problématique au présent et à soi est le moyen privilégié de construire l'altérité et la mise à distance. Comme le dit joliment Antoine Prost, expliquer le terroir médiéval à une personne qui n'a jamais connu à la première personne ce qu'est un terroir est une entreprise sans espoir.

L'investigation du fait humain ne se réduit pas à l'empathie bien sûr, mais la complète par une investigation empirique et logique du fait humain comme fait, en le confrontant et l'analysant à son contexte empirique et à la cohérence qu'il entretient à ce que l'on sait déjà. Des séries statistiques où le fait est encadré et plongé dans des corrélations potentiellement éclairantes, à l'examen du fait dans les causalités matérielles éventuellement multiples ayant permis de le produire, notamment à travers l'étude matérielle des traces qu'il a laissées (ainsi que l'on procède lorsqu'on se livre à une datation au carbone 14 par exemple), le scientifique de la culture recourt lui aussi aux sciences formelles pour interroger la réalité qu'il étudie. Mais toutes les statistiques sur le cours du pain à la veille de la Révolution ne permettront de rien expliquer du tout à quiconque n'a jamais eu faim ni ne sait ce que c'est que manger du pain frais ou cuit au four.

On comprend alors pourquoi les études sur les *big data* en font trop ou pas assez : elles vont trop loin, car elles perdent l'humanité des faits qu'elles étudient. Si bien que les résultats brandis sont bien des propriétés des modèles d'analyse, mais ne nous apprennent que fort peu sur la réalité dont les données sont issues. Et donc elles ne vont pas assez loin : pour tenir le programme qu'elles se sont donné, il leur faudrait aller au bout de l'interprétation et comprendre en quoi les analyses menées permettent de retrouver l'humanité des faits étudiés, et ainsi d'avancer dans la connaissance et la compréhension de la culture.

### **3– Conclusion : un programme de recherche inachevé**

Une telle situation n'est pas entièrement nouvelle. On retrouve des débats anciens qui ont agité, avec raison, la communauté scientifique. De l'article fondateur de François Simiand<sup>7</sup> à l'introduction de l'analyse factorielle dans les années 1970 en sciences humaines et sociales<sup>8</sup>, ces questions ont été régulièrement posées. D'une certaine manière, elles ont été résolues en partie quand les nouveaux outils ont atteint leur maturité épistémologique et leur pertinence phénoménologique. Constaté que les sciences du Web et des *big data* en sont encore loin n'implique pas qu'elles ne peuvent les atteindre. Mais leur principal obstacle reste certainement l'obscurantisme philosophique dans lequel elles se complaisent et l'ignorance des enjeux posés par les sciences de la culture. En considérant le fait humain comme un fait mobilisé dans une base de données, ces sciences adoptent de fait un positivisme naïf où le fait collecté est la seule positivité à interroger et d'où il faut partir pour appliquer les modèles de construction scientifique. Il est peu probable que la science de la nature n'ait jamais fonctionné de cette manière malgré les prescriptions de certaines écoles de pensée épistémologiques. Gageons que ce ne soit pas le cas dans les sciences de la culture.

Mais les difficultés rencontrées ne doivent pas nous écarter des fantastiques promesses que ces études nous permettent d'envisager. C'est un nouveau régime de connaissance qu'il nous faut élaborer, à condition de ne pas oublier que derrière la connaissance des faits collectés, c'est la culture des êtres humains qui est ici en jeu.

---

6. Henri-Irénée Marrou, *De la connaissance historique*, Seuil, 1975.

7. François Simiand, « Méthode historique et science sociale », *Revue de synthèse historique*, pp. 129-157, 1903.

8. Jean-Paul Benzécri, *Pratique de l'analyse des données*, Dunod, 1980.



## Références et lectures conseillées par :

Bruno Bachimont

Auroux Sylvain, *La révolution technique de la grammatisation*, Éditions Mardaga, coll. Philosophie et langage, 1995.

Benjamin Walter, *L'œuvre d'art à l'époque de sa reproductibilité technique : version de 1939*, Folio, 2008.

Benzécri Jean-Paul, *Pratique de l'analyse des données*, Dunod, 1980.

Bloch Marc, *Apologie pour l'histoire ou Métier d'historien*, Armand Colin, 1997.

Daston Lorraine et Galison Peter, *Objectivité*, Les Presses du réel, 2012.

Erismann Christophe *L'homme commun : la genèse du réalisme ontologique durant le haut Moyen Âge à Paris*, Vrin, 2011.

Manovich Lev, *Cultural analytics : analysis and visualisation of large cultural data sets*, Software Studies Initiative, 2008.

Marrou Henri-Irénée, *De la connaissance historique*, Seuil, 1975.

Paqué, Ruprecht, *Le statut parisien des Nominalistes*, PUF, 1985.

Prost Antoine, *Douze leçons sur l'histoire*, Seuil, 1996.

Simiand François, « Méthode historique et science sociale », *Revue de synthèse historique*, pp. 129-157.