



JEAN-GABRIEL GANASCIA, BERTRAND JOUVE, PASCALE KUNTZ
Introduction

Volume 1, n° 1 (2020), p. 5-11.

http://roia.centre-mersenne.org/item?id=ROIA_2020__1_1_5_0

© Association pour la diffusion de la recherche francophone en intelligence artificielle et les auteurs, 2020, certains droits réservés.



Cet article est diffusé sous la licence
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.
<http://creativecommons.org/licenses/by/4.0/>



La Revue Ouverte d'Intelligence Artificielle est membre du
Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org

Introduction

On fait souvent remonter les Humanités Numériques au projet *Index Thomisticus* de Roberto Busa qui visait, dès 1949, à créer automatiquement un index de la *Somme théologique* de Thomas d'Aquin avec le concours d'ordinateurs. On peut aussi trouver leur origine dans les travaux d'Augustus de Morgan qui proposa, dès 1851, une étude quantitative de la fréquence des mots pour caractériser le style des auteurs. Ces applications de l'informatique au domaine des humanités se sont poursuivies dans les années soixante avec la tentative d'Alvar Ellegård de détermination automatique de la paternité des lettres de Junius ou celle de Frederick Mosteller et David L. Wallace d'identification des auteurs des « Federalist Papers ». L'existence, dès 1963, du *Centre for Literary and Linguistic Computing* à Cambridge et d'un groupe de recherche sur l'analyse de textes à l'université de Tübingen, puis la création de la revue *Computer and the Humanities*, en 1966, attestent indubitablement, de l'intérêt précoce porté à l'utilisation du calcul et des ordinateurs dans le champ des humanités. Et, cette tendance se renforce dans les années soixante-dix avec la parution régulière du bulletin de l'*Association for Literary and Linguistic Computing*, et l'organisation de l'ICCH (*International Conference on Computing in the Humanities*). Ultérieurement, apparaissent, dès 1986, *Ansaxnet*, la première liste de discussion pour les humanités et le projet de TEI (*Text Encoding Initiative*) qui établit des directives pour le codage et l'échange de textes. A cela s'ajoute, dans le monde francophone, la tenue récurrente depuis 1992 et tous les deux ans, des JADT (*Journées d'analyses de données textuelles*). Mais, il fallu attendre le tournant du millénaire pour que, sous la pression d'un éditeur, le champ *Humanities and Computing* se transmute en *Digital Humanities* (« Humanités Numériques » en français). Ce nouveau vocable évoque un changement profond consécutif à l'apport de l'informatique qui n'est plus simplement un outil au service des disciplines d'érudition traditionnelles, mais qui aide à forger de nouveaux opérateurs d'interprétation. Cela constitue, de ce fait, une rupture épistémologique dans le champ des humanités qui se font désormais numériques. En pratique, ceci signifie que dans la première époque, celle des *Humanities and Computing*, l'ordinateur aidait à établir des index ou des concordanciers, à identifier des auteurs, avec des statistiques lexicales, ou à faire des publications électroniques, tandis que désormais, avec l'avènement des *Digital Humanities*, on fait appel à de la fouille de textes, à de l'apprentissage machine, supervisé et non-supervisé, à de la visualisation de données, à des analyses sémantiques, par exemple à de la reconnaissance d'entités nommées, à la théorie des graphes etc. pour mieux comprendre et interpréter des textes.

Il est impossible de réaliser un état des lieux du développement des Humanités Numériques en France en quelques lignes, mais on pourra par exemple se reporter au rapport de Pierre Mounier et Marin Dacos de 2014⁽¹⁾. Pour autant, il est clair que le développement de l'informatique grand public dès les années 70 a pénétré progressivement tous les secteurs de la recherche scientifique, sans épargner le champ des sciences sociales et les humanités. Les effets sont bien sûr différenciés suivant les « disciplines » mais aucune n'est aujourd'hui à l'écart. Dans le domaine de la linguistique, déjà fortement ouvert aux outils mathématiques, les approches automatisées ont par exemple donné naissance au secteur très actif du « Traitement Automatique des Langues ». La « nouvelle géographie » du début des années 70, qui revendiquait un effort accru de modélisation quantitative, a permis le développement des « Systèmes d'Information Géographique ». Sans porter de jugement de valeur, nous pourrions comme cela relever les liens toujours très étroits et dynamiques qui existent entre la modélisation quantitative avancée et l'histoire économique et sociale ou l'économie mathématique. Les Humanités Numériques d'aujourd'hui sont marquées par une deuxième rupture qui a eu lieu au début des années 2000 avec le « Big Data »⁽²⁾ et la volonté d'utiliser le format numérique pour pouvoir traiter de façon automatisée ou semi-automatisée d'importants volumes de données. La facilité avec laquelle ceux-ci sont produits, peuvent circuler et être traités, autorise par exemple à penser la construction d'une histoire connectée à l'échelle nationale, européenne, voire mondiale. La sociologie des réseaux, quant à elle, se développe à très grande vitesse. Au même moment, ce phénomène « Big Data » a aussi entraîné un renouvellement de l'Intelligence Artificielle⁽³⁾, renouvellement qui a amplifié le mouvement de transformation des sciences humaines et sociales intégrant par exemple des techniques avancées pour l'étude de corpus textuels. Plus récemment encore, l'utilisation de techniques d'analyse topologique des données dans la conception d'algorithmes d'apprentissage machine ouvre de nouveaux horizons pour l'intégration des sciences humaines et sociales en autorisant des données aux structurations plus complexes.

En associant numérique à mathématiques et en l'étendant aux sciences humaines le discours de Claude Levi-Strauss plaidant en 1958 à l'Unesco pour les « mathématiques de l'homme » trace aujourd'hui encore une voie pour l'avenir : « *Les besoins propres aux sciences sociales, les caractères originaux de leur objet imposent aux mathématiciens un effort spécial d'adaptation et d'invention. La collaboration ne saurait être à sens unique. D'un côté, les mathématiques contribueront au progrès des sciences sociales, mais, de l'autre, les exigences propres à ces dernières ouvriront aux mathématiques des perspectives supplémentaires* »⁽⁴⁾. Le numérique permet maintenant aux chercheuses et aux chercheurs en sciences humaines de récolter des données sur des

⁽¹⁾Pierre Mounier et Marin Dacos. « Humanités numériques état des lieux et positionnement de la recherche française dans le contexte international ». Paris : Institut français, 2014.

⁽²⁾Les Big Data à découvert, Mokrane Bouzehoub et Rémy Mosseri (dir.), CNRS éditions, Paris, 2017, 364 p.

⁽³⁾Kersting, K. & Meyer, U. Künstl Intell, From Big Data to Big Artificial Intelligence ? KI - Künstliche Intelligenz 32 (1), 2018.

⁽⁴⁾Levi-Strauss C. (1954). Les mathématiques de l'homme. *Bulletin International des Sciences Sociales*, VI(4), pages 643-653.

nouveaux supports avec de nouveaux moyens d'observation et de considérer des corpus d'une taille inégalée jusqu'alors. Parmi les nombreux défis posés par cette évolution, retenons en deux qui nous semblent paradigmatiques de la nécessité de la fertilisation croisée des disciplines.

Le premier concerne l'éventail des données pouvant être mobilisées pour l'analyse d'un phénomène : comment concilier les différentes échelles d'observation désormais accessibles ? C'est-à-dire en pratique comment combiner les informations extraites des ouvrages numérisés, des traces numériques sur les média sociaux, des entretiens ethnographiques, des monographies, *etc.* Un détour par les sciences du vivant peut-être ici fécond. Associées à l'informatique, ces dernières ont su lancer il y a un peu moins de vingt ans un programme de recherche ambitieux autour de la biologie intégrative qui vise à intégrer dans l'analyse de « l'homme biologique » les données recueillies à différentes échelles, du génomique au métabolique. Ce projet fédérateur pourrait inspirer le développement au cœur des Humanités Numériques d'une « science humaine intégrative », qui par un dialogue interdisciplinaire intense, pourrait proposer des méthodologies innovantes capables de combiner de façon opérationnelle via des artefacts adaptés le « micro » et le « macro », le « quanti » et le « quali » pour affronter la complexité des phénomènes étudiés. Un tel projet semble renouer à la fois avec les fondements de l'intelligence artificielle et avec ses développements les plus récents puisque la question du traitement multi-échelles des informations est centrale dans l'apprentissage profond.

Le deuxième défi interroge les humanités sur le numérique. Retournons à l'Unesco mais beaucoup plus récemment. Lors du forum de la gouvernance d'Internet qui s'est tenu en novembre 2018 un atelier a été consacré au projet « Software Heritage » qui est la plus grande bibliothèque mondiale des codes sources des programmes informatiques⁽⁵⁾. Leur histoire est courte ; le plus ancien code conservé dans le projet est celui du programme Apollo 11. Mais cette bibliothèque d'Alexandrie contemporaine comporte déjà plus de 5 milliards de fichiers sources. Les questions posées par les analyses de corpus évoquées dans ce numéro spécial se renouvellent sur ces données inédites qui relient en profondeur l'informatique, et en particulier l'intelligence artificielle, et les humanités.

Nous l'avons dit, ces co-constructions en cours au croisement des sciences de l'information et des sciences humaines et sociales ne requièrent plus simplement de l'informatique ou des statistiques comme auparavant, mais de l'intelligence artificielle. C'est en ce sens qu'il nous a semblé important de rassembler quelques exemples de ces interactions. La description de ces expériences doit aider à dégager les orientations que dessinent ces recherches interdisciplinaires, les verrous qui y sont liés mais aussi les dangers à éviter. Comment les outils de l'Intelligence Artificielle peuvent-ils intégrer la connaissance acquise dans le domaine des sciences sociales et des humanités ? L'Intelligence Artificielle peut-elle éviter la dérive d'une utilisation des données numériques qui se développerait sans hypothèses ? De quelles manières ces interactions entre l'Intelligence Artificielle et les humanités sont-elles prometteuses de la production de nouvelles formes de savoirs ?

⁽⁵⁾J.F. Abranatic, D. di Cosmo, S. Zacchinolli (2018). *Building the Universal Archive of Source Code*. Communication ACM.

Entre Intelligence Artificielle et Humanités Numériques il y a un cercle vertueux d'interactions réciproques : l'Intelligence Artificielle transforme le travail des chercheurs dans les disciplines d'érudition traditionnelles que l'on range généralement sous le vocable d'humanités, dans les sciences de l'homme et dans les sciences sociales. En retour, les Humanités Numériques apportent à l'Intelligence Artificielle des problématiques nouvelles.

Dans ce numéro nous illustrons tout particulièrement la première implication mais la deuxième mérite qu'on s'y arrête un peu, au moins en introduction.

L'Intelligence Artificielle vise à reproduire, par simulation informatique, les fonctions cognitives, comme la perception, la compréhension et la décision⁽⁶⁾. Sur certaines tâches de perception, en particulier sur la perception visuelle, la machine atteint parfois des taux de succès proches de ceux de l'humain mais, contrairement à lui, c'est au prix d'un très grand nombre d'exemples sur lesquels il est possible d'« apprendre ». Et même si l'on dispose d'un ensemble d'apprentissage important, la machine peut être trompée devant une image suffisamment bruitée alors que l'humain ne s'y trompera pas. La compréhension doit être entendue comme la traduction d'informations perçues, par exemple d'images ou de paroles enregistrées, ou encore de textes écrits en langage naturel dans un langage formel qui autorise des inférences automatiques et des requêtes complexes. Or, pour de multiples raisons, les résultats de la machine sont bien en deçà de ceux de l'humain. Enfin, les décisions qui doivent être prises en cas d'absence de données ou en temps très court nécessitent de l'intuition que la machine n'a pas. Nous avons donc un saut important à effectuer pour améliorer nos machines et nos algorithmes, saut qui ne sera pas effectué uniquement en améliorant les capacités de calcul et de stockage. Une meilleure compréhension des facultés humaines sera indispensable pour l'accomplir. Or, tant les disciplines d'érudition que les sciences de l'homme et de la société contribuent à cette meilleure compréhension et sont donc susceptibles d'aider à procéder à ce saut conceptuel.

Les travaux scientifiques autour de la communication écrite par exemple, depuis l'analyse de corpus d'auteurs jusqu'aux nouvelles formes d'écritures, apportent à l'Intelligence Artificielle des concepts, méthodes et outils qui participent la formalisation de certaines formes d'architectures du discours et donc à la simulation de certains aspects de la pensée. Sur ce registre, la composante sociale se révèle un élément important dans les échanges interhumains et se révélera, assurément un enjeu de demain en Intelligence Artificielle. En effet, s'il est plus facile de tester et d'étudier des colonies d'insectes et si ces études peuvent nourrir des travaux très intéressants sur les phénomènes de coopération, elles ne dispensent pas de l'analyse des collectifs humains. Déployer des techniques d'Intelligence Artificielle qui soient adaptées aux sociétés humaines plutôt que de demander aux sociétés de s'adapter aux artifices que nous lui construirons semble de plus en plus nécessaire. La compréhension de l'organisation sociale, l'analyse des attentes sociétales, sont des problématiques que la sociologie, et plus généralement les Sciences Humaines et Sociales, ont investies depuis longtemps

⁽⁶⁾ Intelligence artificielle – état de l'art et perspectives pour la France – 2019 (p 40).

et pour lesquelles elles possèdent des compétences spécifiques. Que serait le développement de la voiture autonome sans analyse des usages ? Mieux que cela, quels sont les paramètres pertinents à prendre en compte pour réaliser plus rapidement une bonne inférence en temps réel et qui permettrait de reconnaître plus rapidement une scène de violence, ou d'adapter quasi-instantanément l'action d'un robot face à une situation inhabituelle ? Une partie de la réponse réside-t-elle dans notre capacité à intégrer dans nos algorithmes des considérations subjectives, qui dépendent du contexte de l'action, dans toute sa profondeur sociale et historique ? La fuite en avant vers la donnée ne serait-elle pas à la fois ce qui a permis le bond en avant de l'Intelligence Artificielle mais aussi son nouveau verrou ? Une Intelligence Artificielle au service de l'humain et de la société doit intégrer des connaissances sur l'être humain et sur la société - cœur de métier des Sciences Humaines et Sociales. Inversement, ces disciplines doivent analyser l'impact de l'Intelligence Artificielle sur l'humain et la société, donnant ainsi à ceux-ci les capacités à décider des développements à venir. Bien sûr, les techniques de l'Intelligence Artificielle ont investi ces sciences, et elles pourront donc difficilement faire valoir une objectivité en la matière.

Comme nous l'avons précédemment indiqué, les travaux décrits dans les articles sélectionnés pour ce numéro spécial ne portent pas tant sur l'apport des Humanités Numériques à l'Intelligence Artificielle que sur l'utilisation de l'Intelligence Artificielle par les Humanités Numériques. Chacun des articles apporte un éclairage sur les liens qui peuvent être tissés entre ces deux champs. La richesse de ces liens s'exprime à la fois dans la variété des recherches effectuées et dans la diversité des origines académiques des différents contributeurs et institutions.

L'article de Étienne Cuvelier, Sébastien de Valeriola et Céline Engelbeen est une illustration de l'apport de l'apprentissage pour les recherches en histoire médiévale. Plus précisément, leur problématique initiale porte sur l'identification des sources variées utilisées par les encyclopédistes médiévaux. Et, en se focalisant sur une des trois parties de l'une des encyclopédies les plus importantes du Moyen Âge, le *Speculum Maius*, rédigée au XIII^e siècle par Vincent de Beauvais, les auteurs ont exploré une méthode d'identification automatique basée sur des comparaisons, via une métrique adaptée, entre les notices du *Speculum* et les sources potentielles préalablement identifiées. Les ordres de grandeur permettent de saisir la nécessité du recours au traitement informatisé : plus de 13 000 notices et plus de 60 sources potentielles. En sus de l'évaluation classique des performances de l'approche proposée en terme d'erreur d'identification, l'analyse des résultats soulève des questions qui renvoient à la fois aux choix effectués dans l'algorithme mis en œuvre et à l'incomplétude des sources ou à des erreurs dont l'origine pose de nouvelles questions (erreurs introduites volontairement ou involontairement par l'auteur ou bien différences de versions entre celles de l'auteur et celles référencées dans le corpus étudié ?) qui peuvent enrichir le champ historiographique de l'encyclopédisme médiéval.

L'article de Maria Papadopoulou et Christophe Roche montre quant à lui l'intérêt des *ontoterminologies* dans le domaine des humanités pour aider à définir une terminologie et la structure des concepts associés. Ils illustrent leur propos avec une étude des

vêtements de la Grèce Antique en étudiant comment les experts du domaine définissent et nomment les différents vêtements ou leurs parties. Il est alors possible de définir un ensemble de caractéristiques essentielles, de primitives, sur lesquelles s'accordent les experts et à partir desquelles l'ontoterminologie sera construite. Les auteurs présentent la plateforme TEDI comme exemple de logiciel pouvant être utilisé pour accompagner la démarche.

L'article d'Aurélien Benel fait un historique des premiers travaux conduits dans les années soixante-dix, en particulier en France, sur la formalisation et la simulation des raisonnements en archéologie. Cela fait écho aux recherches actuelles sur le web sémantique, ce qui se conçoit aisément car ces travaux se fondaient sur l'intelligence artificielle sémantique qui naquit à cette époque avec la représentation sémantique des connaissances au moyen de réseaux sémantiques, de cadre de données (frame), de prototypes, etc. Toutefois, selon Aurélien Bénel, et c'est ce qui fait, outre l'intérêt de son travail historique, l'originalité de sa contribution, ces travaux ne visaient pas simplement à une représentation du sens en vue de le simuler, autrement dit à une sémantique, mais aussi à la détermination de la cohérence interne de systèmes de signes, autrement dit à une sémiotique. C'est cette tension entre sémantique et sémiotique et l'apport de cette dernière, directement lié aux Humanités Numériques, que met en évidence cet article.

L'article de Emmanuelle Bermès et Eleonora Moiraghi ouvre plus largement la question du recours à l'Intelligence Artificielle, et plus généralement aux technologies numériques, dans les différentes activités (collecte, description, classement, stockage et conservation, information et communication) d'une institution patrimoniale – ici la Bibliothèque nationale de France. Après une présentation des différents champs mis en œuvre, les autrices se focalisent sur trois projets portant sur des corpus différents : (i) le devenir du patrimoine numérisé en ligne à travers les archives sur Internet de la Grande Guerre, (ii) l'analyse des traces d'usages associées aux logs de connexion de la bibliothèque numérique Gallica, et (iii) le développement d'une indexation sémantique via des techniques d'apprentissage profond dans des documents iconographiques de collections de la BnF. L'article est ensuite complété par une présentation du projet Corpus, initié en 2016, dont l'objectif est de « *construire un service de fourniture de corpus permettant la fouille de textes et de données à destination de la recherche* ». La présentation du projet rend compte non seulement des différentes dimensions explorées par les Humanités Numériques mais aussi du dialogue interdisciplinaire en marche dans un contexte opérationnel porté par une grande institution.

En sus des membres du comité éditorial qui ont contribué à ce numéro, nous tenons à remercier les relectrices et relecteurs supplémentaires pour leur implication :

- Nathalie Aussenac-Gilles, IRIT, CNRS.
- Florent Hautefeuille, TRACE, Université Toulouse Jean Jaurès.
- Alexandre Guilbaud, IMJ-PRG, Sorbonne Université.
- Frédéric Kaplan, EPFL, Lausanne, Suisse.
- Matthieu Roche, TETIS, CIRAD Montpellier.

Introduction

- Thierry Poibeau, LATTICE, CNRS.
- Sylvia Piccini, CNR, Pise, Italie.

Jean-Gabriel Ganascia

LIP6, Sorbonne Université

Bertrand Jouve

LISST, CNRS, Université Toulouse Jean-Jaurès

Pascale Kuntz

LS2N, Université de Nantes