Humanités numériques, corpus et sens

Julien Longhi

Questions de communication

31 | 2017
Humanités numériques, corpus et sens

Humanités, numérique : des corpus au sens, du sens aux corpus

Julien Longhi

Édition électronique
URL : https://journals.openedition.org/questionsdecommunication/11039
DOI : 10.4000/questionsdecommunication.11039
ISSN : 2259-8901

Éditeur
Presses universitaires de Lorraine

Édition imprimée
Date de publication : 1 septembre 2017
Pagination : 7-17
ISBN : 9782814303256
ISSN : 1633-5961

Référence électronique
Julien Longhi, « Humanités, numérique : des corpus au sens, du sens aux corpus », *Questions de communication* [En ligne], 31 | 2017, mis en ligne le 01 septembre 2017, consulté le 15 avril 2024. URL : http://journals.openedition.org/questionsdecommunication/11039 ; DOI : https://doi.org/10.4000/questionsdecommunication.11039

> DOSSIER

JULIEN LONGHI
Agora
Université de Cergy-Pontoise
F-95011
julien.longhi@u-cergy.fr

HUMANITÉS, NUMÉRIQUE : DES CORPUS AU SENS, DU SENS AUX CORPUS

L'interaction croissante entre les sciences humaines et sociales (SHS) et des domaines des sciences informatiques ou des sciences et techniques liées au numérique (fouille de données, intelligence artificielle, gestion des connaissances, visualisation de données…) est notable, et devient un enjeu stratégique de plus en plus fort, en lien avec l'ouverture des sources de savoirs, la mise à disposition de textes, documents, données, etc. Parallèlement, le développement d'outils informatiques de plus en plus sophistiqués[1] et accessibles donne aux chercheurs en SHS des nouveaux moyens d'accéder à leurs observables. Malgré l'instabilité que l'on constate derrière les usages et les dénominations, l'expression « humanités numériques » semble s'imposer dans le paysage académique. Dans *Humanités numériques. État des lieux et positionnement de la recherche française dans le contexte international*, Marin Dacos et Pierre Mounier (2014 : 15) s'interrogent justement sur leur définition : « Comment définir les humanités numériques ? Qu'est-ce qui en fait l'unité malgré leur grande diversité ? ». Se fondant sur un aperçu historique, ils proposent plusieurs réponses : « Au plus haut niveau de généralité, on pourrait dire que les humanités numériques désignent un dialogue interdisciplinaire sur la dimension numérique des recherches en sciences humaines et sociales, au niveau des outils, des méthodes, des objets d'études et des modes de communication » (*ibid.*).

[1] De plus en plus d'outils en accès libres, tels *Tropes*, *Iramuteq*, *Hyperbase web*, permettent une analyse fine des corpus de données textuelles.

Ceci est perceptible dans les divers projets et initiatives dont on peut voir des présentations, publications, etc. Dans ces cadres notamment, des « outillages » sont constitués pour répondre à la demande sociale, mais parfois sans réflexion philologique ou herméneutique (Rastier, 2011). Du côté des SHS, la disponibilité de données issues du numérique, l'ouverture croissante des sources d'informations au public (données ouvertes) et la diffusion de contenus et d'informations sur les réseaux donnent accès à des nouvelles sources et formes de connaissances. Si la constitution de corpus numériques semble être la voie privilégiée pour accéder aux sens produits et circulants, s'intégrant ainsi aux sciences de la culture[2] (Rastier, 2004), il reste encore de nombreuses strates d'analyse de ces « humanités numériques », au premier rang desquelles la constitution d'une sémantique du discours, et des pratiques qui informeraient sur les processus de construction du sens dans ces nouveaux espaces. En effet, si, comme le stipule le « Manifeste des *digital humanities* »[3], « le tournant numérique pris par la société modifie et interroge les conditions de production et de diffusion des savoirs », la sémantique devrait jouer un rôle essentiel[4]. Ceci est plus largement valable pour la sémiologie/sémiotique (selon les appellations[5]), puisque la nature des signes composant ces « humanités » est variable et plurielle. Cette perspective sur les SHS, centrée sur la matérialité des signes dans les corpus, doit être intégrée dans une recherche qui s'intéresse au sens produit par une société, une culture, car « la prise en compte de cette hétérogénéité générique est […] le seul moyen d'approcher la complexité de la procédure qui relie un texte à l'interdiscours d'une formation sociale donnée » (Adam, Heidmann, 2009 : 14). Or, si ces nouveaux matériaux sont exploités à des fins commerciales (web 3.0, aspects publicitaires tels les annonces ou recommandations d'achats) ou pour analyser des phénomènes historiques, politiques, sociologiques, etc., ils le

[2] Pour F. Rastier (2001 : en ligne), « les sciences de la culture sont les seules à pouvoir rendre compte du caractère sémiotique de l'univers humain. Pour connaître l'humain par l'homme, elles doivent reconnaître la part qu'il prend dans cette connaissance, non seulement comme destinataire critique de "résultats", mais comme acteur doué d'affects et de responsabilité ».

[3] Accès : http://tcp.hypotheses.org/318. Le syntagme *humanités numériques* est la traduction de *digital humanities*. Des auteurs comme O. Le Deuff (2015 : en ligne) revendiquent l'adjectif *digital* qui permet de prendre en compte « l'indexation avec un index qui désigne autant ce qu'il faut savoir ». Cette remarque est très juste, pour des raisons institutionnelles néanmoins, nous utiliserons ici *humanités numériques*. Le titre de cette présentation, avec « Humanités, numérique », montre que cet assemblage ne va pas de soi, et une certaine vision est proposée au fil de cet article en articulant le questionnement sur le sens et la problématique des données à travers les corpus.

[4] F. Paquienséguy (2017 : en ligne) et ses collaborateurs proposent également un manifeste tourné vers les sciences de l'information et de la communication : « Nous estimons nécessaire de repenser nos fondamentaux en considérant un spectre élargi et inédit de transformations dont les Humanités Numériques n'appréhendent qu'un versant, à partir de différentes disciplines, centré sur les questions de méthode susceptibles de renouveler les pratiques de recherche en sciences humaines et sociales et de favoriser de nouvelles transdisciplinarités. […] Le numérique n'est plus seulement un ensemble de moyens et supports matériels ou intellectuels, mais un nouvel environnement, un écosystème complexe que nous gagnerons à penser autour du paradigme technologie-société-communication ».

[5] Les deux usages existent, et nous ne pouvons entrer ici dans le détail terminologique. Nous considérons néanmoins que la sémantique s'intègre dans une perspective générale sémiologique/sémiotique, tel que cela avait déjà été proposé par F. de Saussure à propos de la linguistique.

sont moins à des fins scientifiques et citoyennes, avec un regard épistémologique réflexif porté par les disciplines impliquées et de manière conjointe. Les synergies entre chercheurs en SHS et en sciences et technologies de l'information et de la communication par exemple ne doivent pas se limiter à des collaborations utilitaires[6], mais devraient conduire à des questionnements réflexifs à la fois sur les méthodes de recherche, les prérequis des disciplines et les objectifs assignés à la recherche. Il s'agit donc de viser, *via* les synergies engagées au-delà des disciplines, à mutualiser « les savoirs parcellaires […] pour former une configuration répondant à nos attentes, à nos besoins et à nos interrogations cognitives (Morin, 1994 : en ligne). Par exemple, l'apparition de ces nouveaux observables et de ces nouvelles pratiques doit conduire à renouveler l'appréhension du sens tel qu'il se constitue dans les discours, numériques notamment.

Aussi, malgré la multiplicité des travaux sur les humanités numériques, ce projet éditorial garantit une certaine originalité par la prise en compte des facteurs suivants. D'une part, il contribue à ne pas oublier, dans le syntagme *humanités numérique*s, le terme *humanités*, qui renvoie aux textes et aux documents, qu'il s'agit de considérer conjointement de leur point de vue sémiotique et informatique ; d'autre part, pour l'analyse du discours, la sémantique textuelle ou les sciences de l'information et de la communication centrées sur la matérialité linguistique ou sémiotique de ses objets, il montre qu'il y a un rôle à jouer et un nécessaire dialogue interdisciplinaire avec les sciences informatiques ou les sciences de l'information et de la communication. Ce rôle peut être tenu de manière disciplinaire (contributions issues des sciences du langage dans ce dossier), mais aussi, et surtout, par l'échange et le dialogue (contributions relevant de la sémiotique, de la sociologie, de l'épistémologie). L'enjeu est important, puisque, « en se saisissant de toutes les potentialités ouvertes par l'informatique et l'internet, les SHS réussiraient leur "transition numérique", rejoignant ainsi le tournant initié il y a plus de 60 ans par l'industrie et les sciences de la nature » (Bigot, Julliard, Mabi, 2016 : en ligne).

Ce parti pris linguistique[7] n'a bien sûr pas pour but de placer les différentes disciplines en concurrence, mais plutôt de proposer un angle de présentation de travaux par le biais langagier et en mettant en valeur les corpus. Avec les corpus, c'est la question du sens qui devient centrale : comment le sens se construit-il dans

[6] Dans le cadre des humanités numériques, l'interdisciplinarité entre SHS et sciences et techniques/sciences informatiques, sinon soit l'informatique et les technologies sont des « prestataires » pour SHS, sans que les chercheurs participent au développement des outils, soit les SHS sont une « caution » pour le développement d'outils sans que les chercheurs ne se préoccupent fondamentalement de la nature sémiotique des données par exemple. Le terme *interdisciplinarité* est donc assumé, même si nous sommes conscient des difficultés inhérentes à de telles interactions.

[7] La livraison des dossiers d'*HEL*, coordonnée par G. Bergounioux, B. Colombat et J. Léon (2017), semble complémentaire, car elle traite de manière très approfondie du corpus en linguistique. Elle approfondit donc la dimension épistémologique du recueil et de l'organisation des données textuelles, et thématise moins la dimension sociale, politique, et applicative, de ce que nous englobons ici sous le prisme des humanités numériques, à travers des contributions centrées sur des corpus liés aux événements, controverses, etc.

et par les corpus ? Comment le.a chercheur.e peut-il.elle y accéder ? Comment garantir une objectivation des données et de leur interprétation, en affichant des préoccupations résolument issues des SHS ? Ces questions permettent aussi de renouer avec les premières inspirations des humanités numériques, comme l'indiquent Marin Dacos et Pierre Mounier (2014 : 14) :

> « Parti de la linguistique et des études littéraires, le mouvement des humanités numériques s'étend toujours plus en direction d'autres disciplines. Des disciplines qui relèvent davantage des sciences sociales, comme la sociologie, ou qui se positionnent à la frontière des sciences humaines et sociales, comme la géographie, entrent désormais en interaction avec ce mouvement, en particulier par le biais de la problématique de la représentation des données (*data visualisation*). Ces disciplines, qui mobilisent depuis très longtemps des moyens informatiques, en particulier pour produire des calculs statistiques ou des cartes, ont jusqu'à récemment très peu participé au dialogue à l'intérieur d'un champ fortement dominé par le double paradigme linguistique et historien ».

Nous précisons que cette originalité du prisme linguistique ne constitue en rien une critique à l'encontre d'autres travaux se réclamant des humanités numériques, ni des initiatives qui peuvent exister à diverses échelles. Néanmoins, méthodologiquement, en plaçant le corpus au centre de l'analyse, nous convenons qu'« il ne sera plus nécessaire de sortir du corpus pour comprendre et interpréter ses composants » et que « l'analyse contextualisée ou co-textualisée de chacun des textes se fera grâce à une navigation interne au corpus et non sur la base de ressources extérieures arbitrairement et subitement convoquées » (Mayaffre, 2002 : en ligne).

Un champ controversé ?

Privilégier un prisme linguistique, en se situant dans la continuité de travaux et de projets en sciences de l'information et de la communication, en sociologie, en histoire, pourrait conduire à relativiser les enjeux de territoires pour les objets analysés, et engager des dynamiques autour du découpage des frontières disciplinaires. En effet, nous reconnaissons dans la littérature, initiatives, projets, certaines tensions autour des « humanités numériques », alors même que ce syntagme se voudrait accueillant et heuristique[8]. Comme l'écrit Aurélien Berra (2015 : 613), « les humanités numériques ont une sulfureuse pertinence pour notre temps », dont le scepticisme « s'enracine en partie dans le terreau de nos représentations sociales contrastées de l'informatique. D'un côté, le progrès et une promesse de libération ; de l'autre, la grisaille bureaucratique, le cheval de Troie d'une administration stérile, la menace d'un élément inhumain inoculé dans nos systèmes de savoir » (*ibid.* : 614). L'auteur note d'ailleurs que les différentes étapes autour de cette dénomination (apparition/disparition des revues, des unités de recherche et des institutions transdisciplinaires, des associations professionnelles,

[8] F. Granjon (2016 : en ligne) écrit également que « la région épistémologique censée constituer le soubassement des HN reste pourtant des plus floues. Il n'y a, à ce jour, aucun accord franc sur un répertoire de nécessités, d'attendus, de méthodes ou de concepts qui permettrait de délinéer relativement clairement un espace interdisciplinaire singulier ».

des diplômes et des formations) « montrent la construction d'un milieu, en privilégiant les "premières fois" et les projets encore actifs. Elles pointent vers des questions décisives, comme le partage et la pérennité des données scientifiques, la collaboration entre individus et entre communautés, la formation (*ibid.* : 616).

Articuler humanités et numérique

Marin Dacos et Pierre Mounier (2014 : 6) montrent bien la complexité sémantique qu'il y a derrière le recours au terme *numérique* : « Le numérique comme instrument de recherche ; le numérique comme outil de communication ; le numérique comme objet de recherche. C'est de ce complexe-là que les humanités numériques se saisissent et c'est pour cette raison qu'elles représentent bien plus qu'un mouvement de mode passager et superficiel, quoi qu'en disent les mauvaises langues ; un véritable mouvement de fond appelé à redéfinir l'ensemble des champs de la recherche en sciences humaines et sociales ». Selon eux, pour la situation française, il y a des acteurs des humanités numériques, « mais aucun n'est structuré en tant que "centre" au sens où nous l'entendons désormais » (*ibid.* : 43). Ceci rejoint le propos de Milad Doueihi (2015 : 704) :

> « Les humanités numériques ne cessent de susciter critiques et interrogations quant à leur statut institutionnel, leur histoire et surtout leur position dans le paysage intellectuel et académique. Cette situation s'explique en partie par la manière dramatisée dont le monde savant vit la conversion numérique de nos sociétés et par la floraison des manifestes annonçant une rupture avec le passé et la naissance de nouvelles méthodes permettant des explorations radicalement inédites des objets culturels. Tout débat sur la cohérence épistémologique des humanités numériques s'appuie sur une narration, celle, plus ou moins précise, de la rencontre des sciences de l'interprétation avec cette nouvelle science qu'est l'informatique. La version canonique de ce grand récit des origines identifie le père Roberto Busa, jésuite et spécialiste de saint Thomas d'Aquin, comme figure fondatrice. Après une rencontre (en 1949) avec Thomas J. Watson, fondateur et patron d'IBM, il obtient la possibilité de travailler sur les ordinateurs de l'entreprise pour la production de son *Index Thomisticus*. Ce sera la première enquête philologique et philosophique assistée par ordinateur. À cette alliance entre informatique et sciences du texte, on a d'abord donné le nom de *Humanities Computing*, bientôt remplacé par celui d'humanités numériques (*Digital Humanities*). Il importe tout d'abord de noter la concentration des premiers travaux et usages de l'informatique autour de la question du texte et des corpus ».

Selon une approche cumulative, interactive et ouverte de la recherche et de ses prérequis, nous considérons que ce dossier, initié d'un point de vue linguistique, mais en interaction avec les sciences de l'information et de la communication, la sémiotique, le traitement automatique, peut apporter des éléments aux connaissances et avancées sur ce sujet. Plus précisément, comme nous l'avons déjà évoqué, notre objectif s'intégrerait dans la perspective des sciences de la culture telles que définies par François Rastier (2004 : en ligne) :

> « Poursuivant un objectif de caractérisation, les sciences de la culture doivent être différentielles et comparées, car une culture ne peut être comprise que d'un point de vue cosmopolitique ou interculturel : pour chacune, c'est l'ensemble des autres cultures contemporaines et passées qui joue

> le rôle de corpus. En effet, une culture n'est pas une totalité : elle se forme, évolue et disparaît dans les échanges et les conflits avec les autres. Aussi, les cultures ne peuvent être décrites que différentiellement, comme les objets culturels qui les composent, au premier chef les langues et les textes ».

Selon nous, cette citation trouve un heureux éclairage avec le passage de l'informatique au numérique, tel qu'il est décrit par Milad Doueihi (2015 : 711) :

> « De l'informatique (qui certes n'a pas entièrement disparu) au numérique, on passe d'une technicité, souvent exagérée et cultivée pour elle-même, mais exigeant une certaine compétence technique, à des usages plus communs, exigeant d'autres compétences : celles que valorise une nouvelle sociabilité en ligne, peuplée de textes, animée par des "partages". Et aujourd'hui, c'est par rapport à cette pratique numérique populaire que les travaux en humanités numériques doivent être aussi pensés ».

En effet, les humanités numériques modifient la texture des textes et, si les genres textuels traditionnels ont bien été décrits, il s'agit aussi de renouveler les cadres de l'analyse, avec des observables aux frontières difficilement discernables et au renouvellement perpétuel. Mais, comme l'a justement indiqué Olivier Le Deuff (2015 : en ligne), il est nécessaire de s'imprégner des méthodes et travaux développés préalablement sur les humanités non numériques : « Il ne faut négliger le fait que beaucoup des questions actuelles ont été déjà abordées par les siècles passés au cours des différents âges de l'information »[9]. Pour interroger la question du sens en corpus, il faut donc à la fois poser les rapports entre théorie et pratique des corpus (Garric, Longhi, 2009) et aborder l'hétérogénéité complexe des données fournies par le numérique (Garric, Longhi, 2012).

Accéder aux humanités par la « texture » du texte

Par son originalité disciplinaire (à l'interface de la linguistique, des sciences de l'information et de la communication, de l'informatique), ce dossier se situe dans une dynamique qui souhaite investir le terrain du numérique. Depuis janvier 2013, l'École polytechnique fédérale de Lausanne a mis en place un atelier des doctorants en humanités numériques appelé « Atelier pour doctorants impliqués en humanités digitales ». Les chercheurs lausannois « réfléchissent à la matérialité des supports et montrent que les divers objets culturels sont des constructions situées dans le temps, l'espace, et les sociétés. Les mutations technologiques récentes favorisent également selon eux les mises en réseaux et les reconfigurations disciplinaires »[10]. Ce mouvement prouve l'acuité de notre problématique dans le contexte international. Ce projet étant piloté par des chercheur.e.s en sciences dites « dures », il diffèrera probablement du nôtre dans la mesure où nous insistons sur la dimension interprétative et herméneutique du travail et que nous centrons notre problématique sur des enjeux sociopolitiques concrets et applicatifs (controverses, réformes, événements politiques).

[9] L'auteur précise que « les humanités digitales s'inscrivent dans une lignée qui est celle d'une lecture organisée qui sort de la linéarité du texte et du document. Dès lors, nous voulons montrer qu'il existe une histoire plus longue des humanités digitales qui va au-delà de la petite histoire ».

[10] Accès : http://dhlyon.hypotheses.org/47.

Des corpus spécifiques

Une autre spécificité de ce dossier est l'attention portée aux données issues du web 2.0 ou des réseaux sociaux. Ces (nouvelles) formes de discours constituent autant de nouvelles sources d'information. Jean-Gabriel Ganascia (2015 : 630) aide à mieux mesurer l'importance de ces nouvelles données, notamment au regard des *big data* :

> « Aujourd'hui, les messages publiés sur Twitter, pourtant limités chacun à cent quarante caractères, représentent 7 To [téraoctets] par jour, soit une demi-BNF [Bibliothèque nationale de France] quotidienne. Le radiotélescope Murchison Widefield Array en Australie produit 7 000 To par minute de données brutes, c'est-à-dire cinq cents BNF. Après traitement et filtrage, il reste 50 To par jour, c'est-à-dire un peu plus de trois BNF. Selon les prospectivistes contemporains, le Web représentera environ 7,9 Zo en 2015, soit un peu plus d'un demi-milliard de BNF. Et cette croissance-là n'est pas prête de s'arrêter. Ces quelques chiffres aident à se faire une idée grossière des caractéristiques quantitatives de ce que l'on appelle les *big data* ».

Tout aussi intéressante est sa caractérisation de ces données, et la pertinence pour des collaborations qui mettront les SHS au cœur des études :

> « Mais le fait d'emmagasiner de grandes masses d'information n'est pas leur seule caractéristique. On les définit souvent par la formule des "3 V" – pour *volume*, *vélocité* et *variété*. Le volume, c'est-à-dire la quantité proprement dite, oscille entre le téraoctet et le pétaoctet (Po = 1015 octets = 1 000 To), à savoir, entre un dixième de BNF et cent BNF. La vélocité renvoie au fait que cette masse de données se renouvelle en permanence. Enfin, les données sont variées au sens où elles sont hétérogènes : elles peuvent contenir du texte, des images, des sons, etc. Le texte lui-même peut intégrer différentes langues, divers systèmes d'abréviations, etc. Mesurées à l'aune des "grosses données", les études littéraires portent sur des volumes bien faibles. L'intégralité des corpus des auteurs classiques, même en considérant les brouillons, les œuvres critiques, les articles de journaux et les correspondances, se trouve incluse dans une petite partie de la BNF. Quant à la vélocité et à la variabilité, elles apparaissent quasi nulles dans le cas des données littéraires » (*ibid.* : 630-631).

D'ailleurs, Marin Dacos et Pierre Mounier (2014 : 13) confirment que « les médias sociaux (blogs, wikis, réseaux sociaux et outils de micro-blogging) sont de plus en plus massivement utilisés par les équipes de recherche. Des pratiques de communication relevant auparavant de la littérature grise ou de la conversation informelle prennent progressivement un autre statut et contribuent à la conduite même des projets de recherche. Ces pratiques s'accompagnent en outre d'initiatives heureuses, dont ce volume rendra compte :

> « La TEI[11] [Text Encoding Initiative], mais aussi d'autres initiatives de même nature créent progressivement des outils, des méthodes et des espaces partagés entre plusieurs disciplines. Elles ouvrent la possibilité que se développent des recherches concrètes sur, non pas les outils informatiques dans telle ou telle discipline, mais sur les usages des technologies numériques dans la recherche en sciences humaines, dans sa diversité même. Des problématiques partagées émergent alors, sur les pratiques d'encodage de l'information, sur la structuration, la diffusion et l'archivage des corpus ».

[11] « La *Text Encoding Initiative* (abrégé en TEI, en français "initiative pour l'encodage du texte") est une communauté académique internationale dans le champ des humanités numériques visant à définir des recommandations pour l'encodage de documents textuels ». Accès : https://fr.wikipedia.org/wiki/Text_Encoding_Initiative.

Ce renouvellement de l'espace médiatique « pose un problème de représentation de la société quand un grand nombre d'internautes s'exprime sur le web en produisant une archive complexe à localiser, analyser, conserver et interpréter. Ainsi le recueil et la constitution des données deviennent-ils une des priorités pour des chercheurs (Cormerais, Le Deuff, Lakel, Pucheu, 2016 : en ligne). Les travaux menés dans le cadre du projet CoMeRe[12] ont,

> « de 2014 à 2016, créé un noyau de corpus de communication médiée par les réseaux (*Computer Mediated Communication* – CMC) en français. Chaque corpus rassemble un ensemble de conversations intervenant sur la Toile et les réseaux. Nous nous intéressons à une variété de systèmes de communication synchrone ou asynchrone, mono ou multimodaux (éventuellement) : blogues, tweets, SMS / textos, courriels, clavardage, forums, conférence en ligne, mondes synthétiques, etc. ».

Le travail minutieux d'encodage des données et leur balisage avec des métadonnées spécifiques ont contribué à la mise en ligne de corpus librement accessibles et mis en forme selon les standards adoptés à l'échelle européenne. Ainsi le recueil et la constitution des corpus doivent-ils prendre en compte les spécificités textuelles, génériques et discursives des matériaux considérés, car ces spécificités en déterminent les régimes d'interprétation (par exemple, à propos du *tweet* politique, voir Longhi, 2013). Cette attention particulière à la nature des métadonnées et à l'interaction entre données et métadonnées rend nécessaire la contribution active de la linguistique aux avancées dans le domaine des humanités numériques.

Le dossier se situe donc dans un champ de recherche déjà ouvert, en pleine mutation : l'ancrage interdisciplinaire et la confrontation réelle des démarches et des recherches devraient lui conférer une originalité dans le paysage scientifique.

Articuler la théorie et la pratique des humanités numériques

Le dossier s'ouvre par trois articles qui questionnent les humanités numériques du point de vue des sciences informatiques ou cognitives pour l'un, de la sémiotique pour le second, et de la philologie numérique pour le troisième. Les trois adoptent une démarche réflexive et épistémologique afin de mettre en évidence les grands enjeux scientifiques, théoriques, mais aussi technologiques, des humanités numériques. Ainsi Jean-Guy Meunier, dans « Humanités numériques et modélisation scientifique », s'interroge-t-il sur les limites des modèles formels, matériels et conceptuels, pour leur application dans le domaine des humanités : la conséquence de ces limites est que les humanités numériques créent une passerelle originale entre les sciences et l'herméneutique[13], construisant un point de vue différent sur

[12] Accès : https://corpuscomere.wordpress.com

[13] Même si ces deux approches ne sont pas à opposer par principe, l'auteur montre que du point de vue historique et épistémologique, des orientations parfois concurrentes tendent à éloigner les sciences (expérimentales par exemple) des sciences de l'interprétation.

les formes d'interprétation possibles du numérique. Dans « Signifiant et significatif. Réflexions épistémologiques sur la sémiotique et l'analyse des données », Dario Compagno discute des potentialités de l'analyse des données pour la sémiotique. En présentant des exemples d'analyses sémiotiques assistées par ordinateur, il défend l'idée selon laquelle les statistiques peuvent devenir des alliées à l'analyse du sens, sans nuire à la qualité de l'interprétation : il en vient ainsi à proposer des méthodes quali-quantitatives d'enquête, qui sont plus qu'une juxtaposition de méthodes quantitatives et qualitatives traditionnelles, puisque capables de retrouver leur sens et leur dimension sociale sans aller au-delà des données. Cette perspective sémiotique, qui s'intègre de manière plus générale aux sciences de l'information et de la communication, est également à l'œuvre, mais dans une voie plus philologique, dans l'article de Xavier-Laurent Salvador, « Indexer des documents "du dedans" : comment répondre à la question du lieu de la donnée ? ». En effet, l'auteur exploite la problématique de la structuration et du format des documents pour aborder des sujets complexes relatifs à leur interprétation.

Les deux articles suivants s'inscrivent dans ces questionnements sur le sens et l'interprétation et abordent des corpus touchant à la politique. Dans « Ennahdha sur Facebook. Étude sémantique du discours officiel d'un parti », Asma Zamiti et Mathieu Valette étudient un corpus de pages Facebook consacrées à la communication politique en Tunisie, par une méthode différentielle en sémantique de corpus qui consiste à multiplier les points de vue sur le corpus de manière à faire émerger les singularités ; et ce, notamment à propos du terrorisme tel qu'il est traité à la fois par le parti islamo-conservateur Ennahdha et par son fondateur Rached Ghannouchi, sur leurs pages Facebook officielles. Ces préoccupations sont prolongées dans « L'analyse quantitative des médias sociaux, une alternative aux enquêtes déclaratives ? La mesure de popularité des personnalités politiques sur Twitter » par Julien Boyadjian et Julien Velcin, qui tentent d'identifier les logiques sociales de production des opinions politiques sur Twitter à partir des résultats du projet ImagiWeb[14]. Ainsi présentent-ils l'élaboration des algorithmes permettant d'analyser un très grand nombre de messages politiques, et la comparaison de ces informations à des données de sondages d'opinion, afin de mieux saisir les relations (ou l'absence de relations) entre les dynamiques d'opinions en ligne et hors ligne.

Toujours à propos des messages produits sur Twitter et de leur utilité pour l'analyse de phénomènes sociaux, les deux derniers articles traitent des controverses ou événements politico-médiatiques ou sociaux-politiques (polémique sur la gestation pour autrui, réforme du statut des intermittents du spectacle) appréhendés sur Twitter. Dans « Outils notionnels pour l'analyse des controverses » Agata Jackiewicz propose une grille d'analyse destinée à guider l'interprétation des traces langagières des controverses, telles qu'elles se manifestent sur la plateforme Twitter, l'enjeu

[14] ImagiWeb est un projet de recherche interdisciplinaire réunissant des chercheurs en sciences sociales et en informatique. Son objectif était de combiner une approche informatique et sociologique du web 2.0 afin d'analyser les logiques sociales de production et de circulation des messages politiques sur l'internet, en particulier sur le réseau social Twitter.

étant de pouvoir appréhender des situations polémiques dans leur complexité inhérente. Enfin, Dalia Saigh, Boris Borzic, Abdulhafiz Alkhouli et Julien Longhi s'intéressent aux réactions concernant la réforme du statut des intermittents du spectacle sur Twitter (accord du 22 mars 2014) dans l'article « Contribution linguistique à une classification automatique des communautés de sens et à leur analyse. La controverse sur le statut des intermittents du spectacle » : par le croisement d'analyses linguistiques et informatiques, ils illustrent la pertinence d'une recherche fondée sur l'analyse des *tweets* pour rendre compte d'enjeux sociaux et politiques, en croisant des méthodes distinctes (textométrie, fouille de données) mais complémentaires, et qui s'enrichissent mutuellement (la textométrie fournissant par exemple des hypothèses pour la classification automatique, qui est en retour affinée par des analyses fines issues de la linguistique).

Conclusion

On l'aura compris, en réunissant des chercheurs en linguistique (sémantique de corpus, analyse du discours), informatique, sociologie, science politique, sciences de l'information et de la communication, ou sémiotique, ce dossier entend ouvrir des pistes dans le champ des humanités numériques en plaçant la question du sens et le concept de corpus au cœur des préoccupations de ce champ. L'émergence du sens, saisie à travers la représentation de la structure ou des spécificités des données, par des procédures différentielles, par des grilles notionnelles, par un retour philologique, herméneutique ou sémiotique sur la textualité numérique, confère un point de vue signifiant à ce qui constitue la matérialité des humanités numériques, et éclaire les différents modes d'interprétation des humanités numériques.

Références

Adam J.-M., Heidmann U., 2009, *Le Texte littéraire. Pour une approche interdisciplinaire*, Louvain-la-Neuve, Academia/Bruylant.

Bergounioux G., Colombat B., Léon J., éds, 2017, « Analyse et exploitation des données de corpus linguistiques », *Dossiers d'HEL*, 11. Accès : http://htl.linguist.univ-paris-diderot.fr/hel/dossiers/numero11.

Berra A., 2015, « Pour une histoire des humanités numériques », *Critique*, 819-820, pp. 613-626.

Bigot J.-E., Julliard V., Mabi C., 2016, « Humanités numériques et analyse des controverses au regard des SIC », *Revue française des sciences de l'information et de la communication*, 8. Accès : http://rfsic.revues.org/1783. DOI : 10.4000/rfsic.1783.

Cormerais F., Le Deuff O., Lakel A., Pucheu D., 2016, « Les SIC à l'épreuve du digital et des Humanités : des origines, des concepts, des méthodes et des outils », *Revue française des*

sciences de l'information et de la communication, 8. Accès : http://rfsic.revues.org/1820. DOI : 10.4000/rfsic.1820.

Dacos M., Mounier P., 2014, *Humanités numériques. État des lieux et positionnement de la recherche française dans le contexte international*, Paris, Institut français/ministère des Affaires étrangères pour l'action culturelle. Accès : http://www.institutfrancais.com/sites/default/files/if_humanites-numeriques.pdf.

Doueihi M., 2015, « Quelles humanités numériques ? », *Critique*, 819-820, pp. 704-711.

Ganascia J.-G., 2015, « Les *big data* dans les humanités », *Critique*, 819-820, pp. 627-636.

Garric N., Longhi J., dirs, 2009, *L'Analyse linguistique des corpus discursifs : des théories aux pratiques, des pratiques aux théories*, Clermont-Ferrand, Presses universitaires Blaise Pascal.

Garric N., Longhi J., coords, 2012, « L'analyse de corpus face à l'hétérogénéité des données », *Langages*, 187.

Granjon F., 2016, « Présentation du dossier », *Variations. Revue internationale de théorie critique*, 19. Accès : http://variations.revues.org/726.

Le Deuff O., 2015, « Les humanités digitales précèdent-elle le numérique ? », pp. 421-432, *in* : Saleh I., Carayol V., Leleu-Merviel S. *et al., H2PTM 15. Le numérique à l'ère de l'internet des objets, de l'hypertexte à l'hyper-objet*, Londres, Iste. Accès (version auteur) : https://archivesic.ccsd.cnrs.fr/sic_01220978.

Longhi J., éd., 2012, « L'énonciation et les voix du discours », *Tranel*, 56.

Longhi J., 2013, « Essai de caractérisation du tweet politique », *L'Information grammaticale*, 136, pp. 25-32.

Mayaffre D., 2002, « Les corpus *réflexifs* : entre architextualité et hypertextualité », *Corpus*, 1. Accès : http://corpus.revues.org/11.

Morin E, 1994, « Sur l'interdisciplinarité ». Accès : http://ciret-transdisciplinarity.org/bulletin/b2c2.php.

Paquienséguy F., 2017, « Manifeste pour un positionnement des Sciences de l'Information Communication (SIC) vis-à-vis des *Digital Studies* (DS) et autres mutations du Numérique », *Revue française des sciences de l'information et de la communication*, 10. Accès : http://rfsic.revues.org/2630.

Rastier F., 2001, « Sémiotique et sciences de la culture », *Texto !* Accès : http://www.revue-texto.net/Inedits/Rastier/Rastier_Semiotique.html.

Rastier F., 2004, « Doxa et lexique en corpus – pour une sémantique des idéologies », *Texto !* Accès : http://www.revue-texto.net/Inedits/Rastier/Rastier_Doxa.html.

Rastier F., 2011, *La Mesure et le grain. Sémantique de corpus*, Paris, H. Champion.