HAL
open science

À quoi peut bien servir l'analyse du web ?

Amar Lakel, Olivier Le Deuff

To cite this version:

Amar Lakel, Olivier Le Deuff. À quoi peut bien servir l'analyse du web ?. Les Cahiers du numérique, 2018, Enjeux et apports des recherches en humanités numériques, 13 (3-4), pp.39 à 62. 10.3166/LCN.13.3-4.39-61 . hal-02485370v2

HAL Id: hal-02485370

https://hal.science/hal-02485370v2

Submitted on 25 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A quoi peut bien servir l'analyse du Web ? Les communautés de sites des humanités numériques sur Internet

AMAR LAKEL, MAITRE DE CONFERENCES EN SCIENCES DE L'INFORMATION ET DE LA COMMUNICATION, LABORATOIRE MICA, UNIVERSITE BORDEAUX MONTAIGNE, BATIMENT MSHA, 10 ESPLANADE DES ANTILLES, 33607 PESSAC CEDEX
AMAR.LAKEL@U-BORDEAUX-MONTAIGNE.FR

OLIVIER LE DEUFF, MAITRE DE CONFERENCES EN SCIENCES DE L'INFORMATION ET DE LA COMMUNICATION, LABORATOIRE MICA, UNIVERSITE BORDEAUX MONTAIGNE, BATIMENT MSHA, 10 ESPLANADE DES ANTILLES, 33607 PESSAC CEDEX
OLEDEUFF@GMAIL.COM

DOI:10.3166/LCN.12.1.25-38 © 2016 Lavoisier LCN/DOI (calibri 9)

A quoi peut bien servir l'analyse du Web ? Les communautés de sites des humanités numériques sur Internet

Résumé

Suite au succès des *digital humanities* dans le monde anglo-saxon, les pays francophones ont connu un engouement similaire pour ce qui fut d'abord un désir d'évolution avant de constituer un véritable champ de recherche. Contemporains et acteurs de ce « mouvement scientifique », nous avons développé une réflexion autant théorique que méthodologique qui a abouti au développement d'un logiciel de constitution d'archives numériques enrichies. Le but de cet article est de tester les méthodes d'analyse réseau et de cartographie des humanités numériques francophones qui s'appuient sur une collecte de documents numériques liés par des liens de citations qualifiés avec le logiciel My Web intelligence. L'objectif est de mettre à l'épreuve notre méthodologie qui vise à faire une sociologie numérique des acteurs à travers leurs stratégies d'éditorialisation. La problématique ici est de comprendre comment le développement de publications en ligne, de nature très hétérogène, peut révéler une spécificité du mouvement francophone des humanités numériques au regard d'une sociologie des sciences humaines. Nous proposons ici une interprétation des résultats ainsi qu'une réflexion critique quant à la méthodologie employée. Nous mettons en avant le rôle de la communication web pour comprendre les stratégies des acteurs du domaine étudié.

Mots-clés : humanités numériques, cartographie, crawl, hypertexte, interprétation, méthodologie critique

A quoi peut bien servir l'analyse du Web ? Les communautés de sites des humanités numériques sur Internet

A Louise Merzeau, partie trop vite en pleine fondation des études digitales, nous dédions cette brique.

1. Introduction :

Les *digital humanities*, traduites par *humanités numériques* ou *digitales* constituent un mouvement qui vise à améliorer les potentialités pédagogiques et de recherche des sciences humaines et sociales depuis une dizaine d'années. Si l'expression reste récente, elle tire sa filiation des *humanities computing* (McCarty, 2014) qui caractérisaient cette part des sciences sociales et humaines qui cherchait, dans les méthodes quantitatives puis dans les moyens informatiques, de nouvelles approches pour travailler les corpus traditionnels des sciences humaines et sociales. La petite histoire est souvent réduite, et tend à faire remonter l'origine à l'initiative de Roberto Busa qui va utiliser les moyens d'IBM pour indexer de façon automatique l'œuvre de Saint Thomas d'Aquin. Pourtant l'histoire est vraisemblablement plus longue et plus complexe tant cette histoire des humanités digitales précède probablement les développements de l'informatique (Le Deuff, 2015). Au niveau francophone, les humanités numériques et digitales se sont constituées progressivement grâce à une suite d'évènements fédérateurs. Si dans leur généalogie, les chercheurs des humanités

numériques puisent dans des références similaires au niveau des réalisations antérieures (en effet, les acteurs francophones engagés dans les *humanities computing* font fréquemment référence au développement de la TEI (*Texte Encoding Initiative*) (Burnard, 2015), la communauté francophone des humanités numériques a voulu émerger spécifiquement en partant en quête d'une existence autonome tant institutionnelle que théorique. C'est d'abord un cadre organisationnel francophone qui s'est peu à peu créé autour d'évènements fédérateurs comme les différents ThatCamps (*The Humanities and Technology Camp*). Ce cadre a permis l'émergence d'une association dédiée « Humanistica »[1] et de plusieurs manifestes dont celui de Paris 2010[2]. Cependant, la communauté francophone n'est pas nécessairement une communauté uniforme comme en témoignent les débats sur la traduction de *digital humanities* entre humanités numériques et humanités digitales[3]. Mais c'est surtout l'émergence des grandes institutions (jusqu'au Président de la République française[4]) qui a mis ce champ sous les feux de la rampe au point d'en faire un mot à la mode pour les SHS, ces cinq dernières années.

Le mouvement s'inscrit dans une ouverture et une volonté de renouvellement des méthodes et des manières de faire de la science. Le *ThatCamp* qui revendique une logique non institutionnelle et de non-conférence s'est avéré un évènement fortement moteur d'une communauté imprégnée d'une culture numérique de la contribution ouverte (Thély, Serres, Le Deuff, 2013). Le mouvement francophone met en avant une richesse de métiers comme le montre la typologie des inscrits aux divers *ThatCamps* organisés avec des chercheurs, des enseignants-chercheurs, mais aussi beaucoup d'ingénieurs, des professionnels de l'information comme des bibliothécaires sans oublier un grand nombre d'apprentis chercheurs, de jeunes chercheurs en formation ou en quête de poste[5]. Ce mouvement semble donc emprunter tout autant à la culture académique qu'à la culture numérique. D'une part le sérieux et la rigueur scientifique président à une activité évaluée par les pairs dans un mouvement d'institutionnalisation classique (projet de recherche, revues, laboratoires, instituts, etc.). D'autre part, la logique horizontale qui fait fi des hiérarchies (le

---

[1] Humanistica. Assocation francophone des humanités numériques. http://www.humanisti.ca/

[2] Manifeste des digital humanities.https://tcp.hypotheses.org/318

[3] Voir à cet effet la partie controverses et nomenclatures du premier numéro de la revue Etudes digitales qui présentent plusieurs textes qui reviennent sur le choix entre digital numérique. *Etudes Digitales, le texte à venir*, Classiques Garnier, 2015,1, 1.p.249-269

[4] Discours de M. François Hollande, président de la République, sur l'école des chartes du 9 octobre 2015, in Vie Publique. <http://discours.vie-publique.fr/notices/157002867.html>

[5] L'atelier data du ThatCamp Paris 2015 avait mis en avant ses aspects.

porte-parole est un doctorant[6]), les nouveaux régimes de prises de paroles (spécificité des *ThatCamps*) ainsi que la porosité entre acteurs professionnels, techniques et scientifiques semblent directement inspirés des modèles de sociabilités des acteurs de l'économie digitale(Merzeau, 2012).

Ainsi, plutôt « Hack » que « Yack » (Nowviskie, 2014), le mouvement des humanités numériques semble tirer son innovation d'une tension entre les deux modes de socialisation. Le but de cet article est d'analyser au travers un corpus numérique, la communication digitale francophone sur les concepts d'humanités digitales et d'humanités numériques. Cette recherche s'inscrit dans l'analyse de l'éditorialité des acteurs numérique et dans la sociologie de la communication et des sciences. À travers l'analyse structurale des médias numériques et de leur hypertextualité, notamment par l'utilisation du logiciel My Web Intelligence[7], nous avons voulu analyser les communautés d'acteurs en réseaux à travers leur comportement communicationnel et notamment hypertextuel. Les principaux acteurs des *Digital Humanities* francophones et leurs différents positionnements sur le réseau de citation, tels qu'ils apparaissent dans notre méthodologie, peuvent-ils nous révéler l'émergence et la structuration d'un champ scientifique nouveau ? Le mouvement est relativement récent et lié à l'impulsion de certains acteurs autour d'évènements précis. Y a-t-il une tension entre sociabilité libre et individuelle et sociabilité institutionnelle dans les comportements éditoriaux sur internet ? Ou au contraire sommes-nous dans un processus d'institutionnalisation classique qui voit des initiatives individuelles se contractualiser progressivement dans des investissements de forme communicationnel (Flichy, 2003; Callon et Latour, 1991)? Notre hypothèse majeure est basée sur l'idée que la communication numérique a transformé la mise en visibilité de l'activité scientifique, notamment avec les blogs et carnets de recherche ainsi que les réseaux sociaux dans le prolongement d'un chercheur 2.0 (Gallezot, Le Deuff, 2009). Mais, pour autant, la communication scientifique numérique peut-elle révéler la science en train de se faire ? L'étude d'une communauté d'experts dans des forums libres manifestait un processus de structuration progressive que l'approche quantitative pouvait éclairer (Lakel, Massit-Folléa, 2007). Mais une analyse du web s'attaquant à une « archive » complexe ouvre des perspectives heuristiques nouvelles. Par l'apparition et la disparition d'acteurs, l'analyse des comportements communicationnels et de leur impact en ligne, l'étude peut-elle reposer sur des traces suffisantes pour tenter une sociologie des sciences ?

---

[6] http://www.humanisti.ca/comite/
[7] http://mywebintelligence.net

L'approche quantitative et cartographique sera discutée au regard de l'expérience que les auteurs peuvent avoir de ce champ afin de révéler les concordances et les incohérences entre la carte et le territoire (Korzybski, 2015), du moins dans sa dimension de monde vécu. Dans une première partie, nous reviendrons sur la méthodologie mobilisée pour analyser la structuration du champ des humanités numériques francophones. Nous montrerons, dans une seconde partie, comment les résultats révèlent les tensions d'une progressive institutionnalisation du champ. Enfin, nous reviendrons sur la méthodologie par une discussion critique de l'analyse du web et des limites épistémologiques que émergent de notre étude.

2. Du Corpus Au Reseau : Methodologie Pour Une Comprehension Des Strategies D'acteurs En Situation D'innovation.

2.1 Une brève histoire des humanités numériques francophones

Le mouvement des humanités numériques francophones émerge en tant que tel quelques années après le succès des premières manifestations et écrits anglo-saxons (Schreibman, 2008, Gold, 2012). A l'origine, ce sont des initiatives individuelles qui vont permettre la construction progressive d'une communauté autours de ce concept. Les initiatives autour de Marin Dacos et de Pierre Mounier (Dacos & Mounier, 2015) vont permettre la réalisation des premiers *ThatCamps* ainsi que de la démonstration de la nécessité de compétences nouvelles, d'ingénieries dédiées voire d'infrastructures. Frédéric Clavert va organiser un évènement clé au Luxembourg avec une focale particulière sur l'histoire (Clavert & Noiret, 2013). La thématique est particulièrement importante dans le champ des humanités digitales francophones, plusieurs acteurs en sont issus avec des proximités opportunes notamment avec les sciences historiques économiques. La communauté s'élargit peu à peu, mais un socle de personnes particulièrement actives demeure notamment dans l'animation des ateliers des *ThatCamps* (Les ThatCamps de Paris, ceux de Saint-Malo, de Lyon et de Lausanne) et des manifestations qui vont mentionner humanités digitales ou humanités numériques dans leur titre.

Plusieurs ouvrages ont abordé le sujet ces dernières années avec un prisme qui oscille entre une volonté d'expliquer les mutations voire de les encourager, et des publications qui intègrent une partie critique. Le rôle de l'ingénierie apparaît important dans les humanités numériques francophones, l'étude des présents aux *ThatCamp*s montre que les acteurs ont été ou sont liés à l'ingénierie de la recherche[8]. Thématiquement, le mouvement francophone a attiré une importance communauté d'historiens qui ont trouvé de quoi poursuivre des filiations plus anciennes marquées notamment par le besoin de travailler autour de statistiques et de corpus de données numériques (Clavert, 2015 Lemercier & Zalc, 2010) en renouvelant l'étude des textes par la fouille statistique et l'encodage du type TEI (l'exemple type en est le consortium Cahier[9]). Les sciences de l'information et de la communication ont manifesté également très tôt leur présence et continuent de s'interroger sur leur place dans ce mouvement (Domenget, Bonaccorsi, Carayol, 2016 ; Cormerais & Ali., 2016). Il convient également de mentionner le tournant design des humanités numériques inspiré notamment par les travaux pionniers de Johanna Drucker (Drucker, 2009) et qui ont vu la communauté francophone développer réflexions et productions autour de ces thématiques davantage artistiques et esthétique avec Nicolas Thély (Thély, 2014) dans un premier temps et Stéphane Vial (Vial, 2016) dans un second. Le mouvement apparaît donc complexe et hétérogène en mêlant des acteurs de disciplines différentes, mais qui se retrouvent sur des terrains communs via la culture numérique.

Les humanités numériques francophones ont produit, ces trois dernières années, une abondante littérature réflexive, soit par des textes programmatiques soit par des textes critiques, alors même que de leur propre analyse le mouvement n'a pas une décennie. La prolifération des textes qui tentent de comprendre un mouvement en gestation (Mounier, 2015 ; Berra, 2015, Granjon et maggis, 2016 ; Diminiscu et Wievorka, 2015 ; Ouakrat et Mesengeau, 2015 ; etc.) se rejoignent globalement autour d'un même argumentaire. S'appuyant sur de nombreux manifestes et des rassemblements fondateurs, les auteurs évoquent une « révolution » académique même si Pierre Mounier (Mounier,2015) et Yves Cittons (Citton, 2015) semblent alertés du côté « déjà vu » , ils notent également beaucoup d'opportunisme financier dans la

[8] L'atelier data du ThatCamp Paris 2010 avait mis en avant le phénomène en étudiant le profil des différents inscrits aux ThatCamps. Les ingénieurs d'études ou de recherche y étaient fortement représentés.

[9] Consortium « CAHIER » (Corpus d'auteurs pour les humanités : informatisation, édition, recherche), consortium labellisé par la TGIR Huma-Num en septembre 2011 http://cahier.hypotheses.org/

prolifération des projets. Par conséquent, les critiques même du mouvement appellent de leurs vœux cette « évolution » en y ajoutant une touche de réforme au profit d'un engagement social, critique voire politique.

C'est de cette « révolution » dont nous voulons faire l'étude en resocialisant le discours dans son contexte ancré (Ouakrat et Mesengeaut, 2016) pour voire si véritablement des traces d'agencements nouveaux (Plantin 2014,) viennent bouleverser la science en train de se faire.

2.2 Quantifier/qualifier les actes d'éditorialisation numériques : quelles méthodes ?

Notre corpus est constitué de 1800 pages indexées par le moteur de recherche Google[10] et qui répondent à la requête « Humanités (digitales OR numériques) » ou ayant été directement citées par une de ces pages (*broad topic queries*). Chaque *page* web est isolée par une URL qui lui sert d'identifiant unique. Si méthodologiquement cet ensemble de pages web constitue l'échantillon d'une archive du web, nous l'appellerons dans notre projet, un *territoire* sémantique, car il constitue le corpus numérique de discussion sur un ensemble de notions. Par la suite, l'outil My Web intelligence va agréger les données des pages web sur un second identifiant unique : le *domaine d'expression* (703 dans le cadre de cette recherche)[11]. *Cette opération* permet de constituer une nouvelle population de « sites web » (certaines fermes de contenus constituant plusieurs « sites », il existe une différence entre le domaine d'expression et le nom de domaine) définie comme ayant au moins une expression indexée sur la requête. Ainsi, la constitution du corpus permet de travailler sur deux échelles, celle de l'expression et celle de la plateforme éditoriale (sources des traces de l'actant de notre étude).

Par la suite, tant les pages que les domaines vont être enrichis de données d'une façon soit automatique, soit humaine. On trouve parmi les variables qui

[10] L'algorithme de Google reste cependant non accessible tout comme son index.
[11] Nous n'avons garder dans notre étude que les producteurs de contenu en excluant tous les sites de distribution du type libraire, archive ouverte, etc.

qualifient la page, son contenu indexé (Titre*[12], URL*, Expression*, Lemmes*) et la date de publication. Pour les données relationnelles, ne seront relevés que les liens de citations explicites, c'est-à-dire de lien contenu au cœur de l'expression extraite automatiquement de la page web. On évite ainsi de relever aveuglément les liens de la page entière qui n'ont souvent rien avoir avec l'acte communicationnel en lui-même[13]. Ainsi le lien est véritablement une citation explicite dans l'acte d'expression. Viennent ensuite, dans une approche mixte, « les facteurs et les variables sociales ou démographiques qui déterminent, au moins en partie, certains comportements » (Mesengeau et Ouakrat, 2016). C'est au niveau du domaine d'expression que l'on qualifie le type d'émetteur (c'est-à-dire l'identification sociale des responsables éditoriaux). Si la variable est qualifiée humainement, elle repose sur deux dimensions que sont les secteurs d'activité et le niveau d'institutionnalisation. À titre d'exemple, dans cette étude nous distinguerons « chercheur individuel », « association de chercheurs », « laboratoire de recherche », « d'institution administrative », etc. Cette variable est suffisamment ouverte pour permettre le recodage a posteriori. À ce même niveau, on qualifiera le domaine d'expression selon le type de sites éditoriaux par une nomenclature fermée. Cette nomenclature qualifie les types d'économie éditoriale que l'on retrouve sur le web[14]. Ces deux variables visent à inscrire sociotechniquement l'acte d'éditorialisation. Nous avons préféré cette qualification à celle d'auteur qui ne prend en compte qu'une partie de la chaîne éditoriale. Enfin nous avons qualifié la situation géographique des acteurs en fonction d'une position à Paris, en Province, dans l'ère francophone (hors France) ou à l'international (hors ère francophone).

La plateforme My Web Intelligence va également assurer le recodage d'une série de variables automatiquement à la fois pour la recherche, mais aussi pour le classement des documents. Il s'agit, au niveau de la page, de faire différentes mesures d'impact sur les réseaux sociaux et d'en faire un indice synthétique sur une base 100 qui permet de classer les pages ayant reçu le plus d'approbation manifeste sur ces plateformes. Au niveau du domaine d'expression, le nombre de pages indexées dans le projet par domaine montrera l'importance de la

---

[12] Les variables avec une « * » sont des variables constituées automatiquement par la plateforme My Web Intelligence. Pour une discussion technique, nous renvoyant aux codes sources ouverts et discutés sur https://github.com/MyWebIntelligence/MyWebIntelligence

[13] Pour l'extraction total des liens de la page se reporter au Logiciel Hyphe : http://hyphe.medialab.sciences-po.fr/

[14] Site institutionnel, Site thématique, Portail de ressources, Site éditorial, Moteur de recherche, Site e-commerce, Blog, Réseaux sociaux.

thématique pour le domaine. Ces mesures jouent un rôle primordial dans la hiérarchisation du corpus et dans la gestion de l'enquête (Kleinberg, 1999).

2.3 Détecter des communautés éditoriales sur le web

Les données relationnelles qui révèlent les citations qualifiées des pages web traitant des humanités numériques (et digitales) sont la base informationnelle première de l'analyse structurale du réseau de la communication scientifique en ligne. Notre analyse vise à utiliser les méthodes de l'analyse structurale des réseaux (Wasseman et Faust, 1994 ; Scot, 2012), implémentées dans le logiciel Gephi[15]. Afin que le graphe hypertextuel soit en mesure de montrer les relations et les affinités scientifiques, nous avons utilisé des variables secondes toutes recodées à partir des relations pondérées : l'algorithme de visualisation cartographique *Force Atlas* (Jacomy et ali, 2011) et l'algorithme de détection/qualification des communautés *ModularityClass* (Blondel et ali, 2008) pour déceler les communautés de citations du réseau. Nous avons aussi calculé au niveau de chaque domaine le *Page Rank* local (Brin et Page, 1998) (c'est-à-dire un *Page Rank* calculé uniquement au niveau des sites présents dans le corpus) et l'*Authority Rank* qui permet d'évaluer et de classer les domaines selon leur autorité dans le réseau (pour une discussion de cet indicateur voir Kleinberg, 1999).

L'algorithme de *ModularityClass* implémenté par Gephi est sans doute celui qui a le plus concouru au succès de la plateforme d'analyse. Comme tous les algorithmes en analyse structurale des réseaux, ce sont les liens entre les nœuds (en prenant en compte ici la pondération, en l'espèce le nombre de liens entre deux domaines) qui permettent de détecter les nœuds qui sont le plus liés entre eux par rapport au reste du réseau. L'indice de modularité permet de connaître la densité des liens intracommunautaires au regard des liens intercommunautaires. L'algorithme va donc d'une manière itérative (en partant sur une base un nœud égal un réseau) associé les nœuds en communauté de plus en plus large tout en recherchant les communautés les plus denses. Ainsi

[15] https://gephi.org/

en partant de binôme et en associant progressivement les communautés pour former des communautés de plus en plus larges, il est en mesure de créer une échelle à la granularité fine des communautés relationnelles. En fonction de sa problématique, on peut fixer « le grain » pour avoir un nombre de communautés données (12 communautés dans notre étude). Ainsi, chaque domaine d'expression est inclus par un identifiant unique dans une seule communauté. Cette variable est la base d'une étude de la sociabilité au regard des autres variables.

Nous ajoutons à cette catégorisation algorithmique, une méthode visuelle en usant de la cartographie réseau par spatialisation de force. L'algorithme Force atlas et sa version 2 fait partie des *Force Directed Algorithms* qui visent à traduire visuellement la notion de communauté. Sur l'idée que la densité des relations fonde la communauté d'une part (voir algorithme précédent), et que la proximité spatiale manifeste le mieux visuellement les communautés d'autre part, les algorithmes de force s'appuient sur une simulation spatiale qui vise à organiser une force de répulsion et d'attraction des nœuds entre eux (Jacomy et ali, 2011). Les nœuds sont comme des particules chargées qui exercent une force de répulsion vectorielle continue et les liens entre les nœuds, au contraire, forment une force d'attraction entre les nœuds. Au-delà de la myriade d'algorithmes qui implémente cette idée simple, il faut retenir que la visualisation cartographique agrège les nœuds en communautés et met de la distance entre les communautés. Ainsi la position du nœud dans l'espace n'est pas une position en soi, mais une position relative. En effet dans ces cartes de réseaux, ce sont les axes de répulsion et d'attraction, d'éloignement et de proximité qui sont signifiants. Les méthodes visuelles ne sont pas aisées à interpréter d'autant qu'il est nécessaire de bien connaître la signifiance de la carte (son processus de signification) pour l'interpréter. Nous avons pris le parti de ne pas utiliser les cartes comme outil démonstratif, mais comme outil d'exploration de données et de tests préalables de la validité des hypothèses.

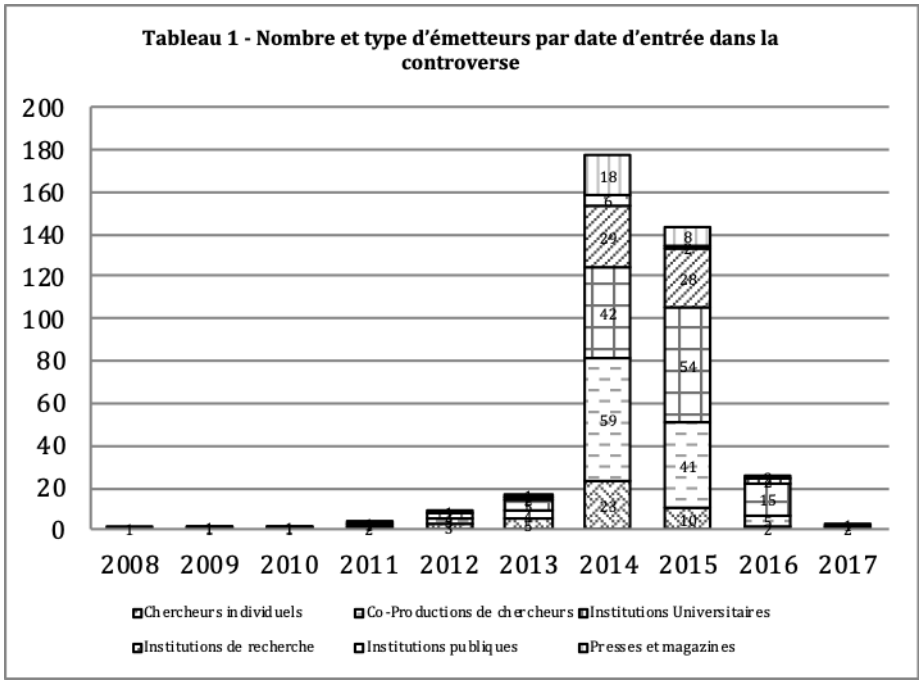
3. Les Humanites Numeriques : Institutionnalisation Ou Effet De Mode ?

Les humanités numériques reçoivent régulièrement des critiques que ce soit dans ses méthodes ou en ce qui concerne le besoin d'y développer autour d'outils une méthodologie et une pensée critique (Berry, 2012). Notre étude

vise donc à distinguer ce qui relève d'un effet de mode des circonstances d'une réelle institutionnalisation. La communication scientifique s'effectue de plus en plus sur le web et a ainsi acquis les caractéristiques propres au marketing digital avec ses avantages et ses inconvénients. Nous souhaitions également vérifier si cette évolution de la communication scientifique avec le web la rend compatible avec des phénomènes viraux et éphémères. L'analyse de la dynamique de la production éditoriale est approchée par les structures de citation.

3.1 L'institutionnalisation rapide des humanités numériques.

Les premiers résultats de notre étude concernent le dénombrement des *types de locuteurs* produisant du contenu sur les mots-clés « humanités numériques » et « humanités digitales ». Il s'agit de qualifier les collectifs sociaux en charge de l'éditorialité des pages répertoriées tel qu'il se présente comme sujet locuteur sur le support médiatique. La question ouverte a abouti, après nettoyage, à 17 modalités que nous avons regroupées en cinq grandes catégories. On constate qu'en dehors de la presse et des revues généralistes (6,1 %), la quasi-totalité des sites qui parlent des humanités numériques provient de sites du monde académique. On peut dire d'emblée que le sujet de réflexion épistémologique et heuristique connaît une très faible ouverture aux acteurs non académiques alors même que les plus hautes instances politiques s'en sont fait l'écho. Au sein des acteurs, les chercheurs parlant à titre personnel représentent 13,1 % des sites auxquels il faut ajouter leur production 29,1 % (sites dont les chercheurs sont à la source collectivement : associations de chercheurs, colloques, revues scientifiques, livres, projets de recherche). Ainsi l'activité issue directement des chercheurs et de leur collaboration libre représente 41,2 % des domaines d'expression. Les humanités numériques sont pour près de la moitié le fait d'initiatives individuelles plus ou moins en marge des grandes institutions. Les institutions universitaires (Universités, bibliothèques et laboratoires de recherche) forment un premier degré d'institutionnalisation que l'on qualifierait d'autonome. Avec 38 % des domaines, ils sont l'autre moitié des sources éditoriales. Enfin les institutions de recherche para- ou interuniversitaires (ANR, MSH, CNRS, etc.) et les institutions publiques sont très minoritaires avec 14,7 %.

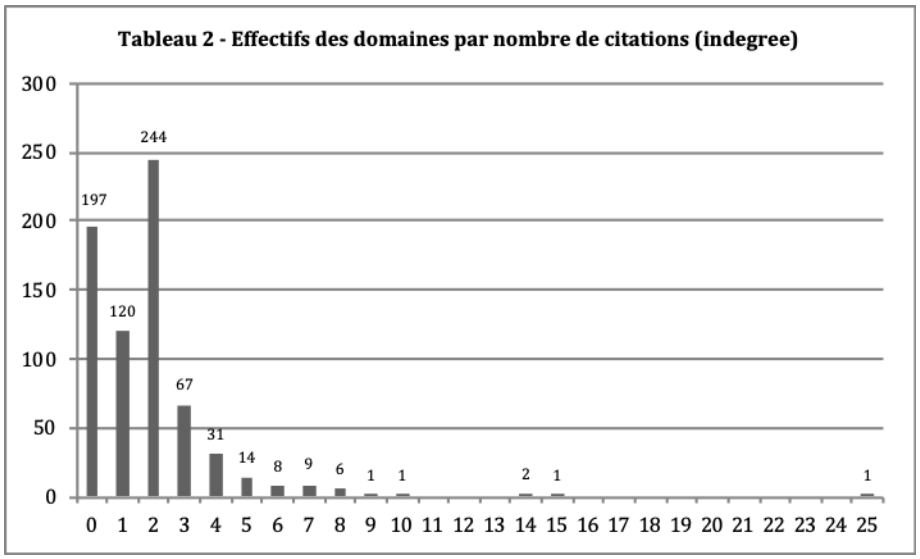


Pour autant, la courbe d'entrée dans la conversation numérique des domaines d'expression (c'est-à-dire la date du premier article publié par site sur le sujet) dans le débat scientifique, année après année, montre un phénomène extrêmement rapide de structuration du champ (Tableau 1). En effet, on constate, après une forte ruée sur la thématique en janvier 2014, une diminution de plus en plus rapide des nouveaux entrants annonçant la fin de la période de constitution du réseau d'acteurs qui ne voit quasiment aucun recrutement de nouveaux locuteurs à partir du quatrième trimestre 2016. On voit donc très nettement le processus d'institutionnalisation se mettre en œuvre par l'apparition tardive des institutions de tutelle fortement présentes pour embrayer le phénomène dans l'espace conversationnel. À la suite des chercheurs individuels et de leurs productions, l'arrivée des institutions universitaires et des institutions de recherche s'effectue très rapidement. Le premier trimestre 2014 voit le phénomène d'expansion/institutionnalisation littéralement exploser. Le positionnement de l'EHESS est notable et présentant un rôle d'ambassadeur du mouvement et porteur d'une crédibilité

académique[16]. Il ne faut pas non plus exclure un opportunisme de mode à un moment où « numérique » devient la clés des financements par projet de la recherche apportant une population de chercheur en réalité peu impliqué dans le mouvement. (Mounier, 2015 ; Citton, 2015)

3.2 Les vecteurs de socialisation : entre proximité géographique et proximité sociale

L'analyse structurale du réseau de relation des sites web nous apporte un complément d'information sur la hiérarchie des sites et leur organisation. Le corpus de site web répond parfaitement aux théories propres à la bibliométrie démontrant qu'il existe une courbe logarithmique des citations avec un cœur de réseaux, une zone secondaire et une longue traîne en dispersion (Bradford, 1934). En l'occurrence ici le classement des sites en fonction du nombre de liens entrants au sein du corpus respecte cette courbe (Tableau 2).



[16] L'EHESS fait office depuis sa création d'institution ouverte et à la marge du monde académique. On peut dire a posteriori qu'il n'est pas étonnant qu'un mouvement innovant issu des marges soit accueilli, comme premier moment institutionnel, par un organisme « pont » aux frontières de l'université et du CNRS.

L'application d'un algorithme de détection des communautés (*Modularity Class)* implémenté dans le logiciel Gephi nous a donné dix communautés de plus de 5 sites structurées par leur co-citation. Elles sont pourtant très inégales en taille et couvrent 582 sites sur 703. Ainsi moins d'un tiers des sites ne sont liés à personne ou presque. Cette densité relationnelle est assez remarquable sur le web. On peut classer la taille de ces communautés en tailles décroissantes de 107 membres à 8 membres. Pourtant quand on cherche la corrélation entre ces communautés (Tableau 3) et le type d'acteurs qui la constitue, on constate une corrélation claire, X2 (45, N = 323) = 95,6, p < . 001, V=0,243. L'influence structurante des individus présents sur les réseaux sociaux, les blogs ou les revues en ligne sur la question des humanités numériques (Martin Grandjean, Stéphane Pouyllau, Pierre Mounier, Marin Dacos, etc.) montre que ce sont d'abord les « loups solitaires » avec une présence numérique qui ont été moteur dans la communication scientifique numérique grâce à leurs propres ressources et leur stratégie éditoriale. Cette médiatisation n'est pas nécessairement corrélée à une inclusion au sein des grands projets institutionnels. On repère, à la fois, des réseaux de chercheurs évoluant dans une logique d'association horizontale avec la création d'associations de chercheurs à l'image d'*Humanistica*, de projets de recherche ou de gestion de revue en ligne. Mais on identifie également des institutions universitaires ou de recherche qui suivent leur propre logique de socialisation en développant de grands projets comme la grande infrastructure TGIR-Humanum ou le positionnement du CNRS. Il est possible alors de soutenir l'hypothèse d'une séparation entre deux régimes de socialisation selon les modes d'institutionnalisation de pratiques, qu'ils soient ascendants ou descendants. Les institutions universitaires apparaissent alors comme une zone intermédiaire intégrant les initiatives individuelles tout en se liant parfois avec les grandes institutions de recherche.

Enfin, au regard de la cartographie des sites, nous avons voulu tester si l'ancrage géographique n'est pas en fin de compte un des facteurs majeurs qui rassemble les sites d'une même communauté numérique. Des sites d'institutions proches géographiquement auraient tendance à davantage se citer : les institutions parisiennes avec Huma-num et le CNRS, l'ENS comme pont entre Paris et lyon, Lille avec le réseau des MSH et la proximité de la Bretagne, la communauté suisse avec Humanistica et l'École Polytechnique de Lausanne, et un réseau international autour de l'EHESS. Tout d'abord, il faut constater que le facteur géographique révèle une répartition très inégale des acteurs des humanités numériques distinguant Paris et la Province. La région parisienne apparaît dominante (42,5 %) très fortement connectée à la région lyonnaise via l'ENS (et grenobloise) avec 16,6 % des domaines. La zone

francophone hors France avec 14,3 % forme un second pôle fortement connecté au premier par des institutions parisiennes comme l'EHESS ou encore l'Institut Français. Ainsi devant ce pôle relié à Paris, tout le reste de la France ne représente que 25,1 %. Quel que soit le découpage choisi, il existe bien une corrélation très forte entre la provenance géographique et les communautés en ligne nettement plus marquée que la précédente (X2 [18, N = 323] = 143,4, p < . 001, V=0,348). Ainsi, on peut dire que le pôle Paris se distingue clairement de la province et du pôle international. Même s'il faut remarquer une légère porosité, on constate très clairement trois économies relationnelles, trois petits mondes en fonction de la situation géographique des acteurs.

3.3 Centralité et marges : qui sont les influenceurs des humanités numériques ?

Les sites de projets de recherche, mais surtout les évènements scientifiques comme les *ThatCamps* et les colloques estampillés humanités numériques semblent avoir été des moments structurants dans le réseau tant dans la réalité que sur le numérique en rassemblant les acteurs dans des moments de rendus intenses où chercheurs, associations et laboratoires se comptent et se donnent à voir. Ils semblent avoir formé des ponts entre les communautés de chercheurs. Nous nous sommes alors demandé si le niveau d'institutionnalisation et la localisation géographique avaient des conséquences structurales sur sa place dans le réseau. Force est de constater que ces deux facteurs sont globalement faiblement influents sur les positions structurales numériques. Par un test d'analyse de variance (ANOVA) sur la variable Type d'Émetteur et la variable Localisation, on constate que les deux facteurs n'ont aucune corrélation significative avec le nombre de citations d'une part (*indegree*) et la marginalisation (*eccentricity*). Par contre les variables *Authority* et *Pagerank*, elles, sont fortement corrélées avec respectivement un *F (6, 712)=2 348, p=0,03* et *F (6, 712)=2 807, p-0,01*. Les chercheurs individuels et leur coproduction (notamment les associations de chercheurs) présentent un plus grand indice d'autorité et montre un relationnel plus fourni que les autres catégories, ce qui prouve que l'influence demeure davantage à l'échelon individuel plutôt qu'au niveau institutionnel. Ainsi l'hypothèse d'un mouvement scientifique qui vient de la base et de personnalités fédérant des chercheurs individuels autour d'événements et de projets collaboratifs est là encore corroborée.

4. Des Limites Et Ouvertures De L'approche Cartographique : Discussion Sur L'acces Aux Archives Numeriques.

La corrélation de la représentation numérique de ces réseaux avec la réalité de la vie scientifique mérite d'être discutée. L'enjeu scientifique est donc d'évaluer l'avantage de cette méthodologie au regard des autres formes d'exploration des groupes sociaux et de leurs pratiques et comportements, notamment en sociologie des sciences (Callon et Latour, 1991 ; Callon, 1986) et de l'innovation (Flichy, 2003 ; Alter, 2003). Nous présentons ici un recul critique sur les données au regard de notre expérience participante de la communauté scientifique des humanités numériques francophones.

4.1 Le réseau sur le web structuré par le réseau social

L'approche globale de l'analyse réseau nous montre les forces vectorielles qui constituent les structures communautaires en matière d'innovation scientifique. Une première force géographique sépare nettement Pays francophones, Province, Paris (avec une forte prépondérance de cette dernière) et tourne autour des grandes métropoles régionales de recherches. Une seconde force repose sur le processus d'institutionnalisation qui sépare les chercheurs individuels qui se regroupent autour d'associations et d'évènements et les grands projets sur fonds publics autour du CNRS. Enfin, on voit une force centrifuge qui exclut tout ce qui n'est pas membre de la recherche académique vers les marges. On voit ainsi émerger de cette approche une empreinte numérique d'un champ disciplinaire en construction dégageant une topologie d'acteurs en fonction de leur degré d'institutionnalisation, du rapport géographique centre/périphérie et de l'endogamie disciplinaire. L'analyse de cette agora scientifique met en avant une très forte inégalité des actants que ce soit sur le nombre de citations, le nombre de publications ou l'impact dans les réseaux sociaux. Ainsi nous faisons le constat que parmi les modalités de structuration des réseaux sur internet, les enjeux sociaux (Grandjean, 2016) jouent un rôle indiscutable. La communication scientifique serait alors en rien une virtualité de façade et correspondrait à la représentation de jeux stratégiques d'acteur territorialisé ?

Il est possible de voir poindre la dynamique de l'émergence des champs scientifiques avec un long processus d'initiatives isolées de chercheurs

indépendants qui tentent de rallier des confrères et de s'organiser autour d'évènements programmatiques. L'exemple des *ThatCamps* et de l'émergence d'*Humanistica* durant ces évènements est à cet égard emblématique de cette dynamique. Les institutions scientifiques traditionnelles arrivent a posteriori pour légitimer et s'emparer de la thématique en mobilisant les ressources qui manquent aux chercheurs innovants. La recherche est donc affaire d'acteurs-réseaux qui se structurent progressivement par des mises en relation et des investissements de forme (colloque, marquage institutionnel, etc.), ainsi que par la recherche des ressources qui sont le monopole des grandes institutions. Si le phénomène est bien connu des théories de l'acteur-réseau (Callon et Latour, 1991), la nouveauté qu'apporte cette étude repose sur la possibilité d'analyser le phénomène avec une précision accrue par la facilité d'accès aux traces numériques. Enfin, cette analyse montre très clairement que l'on peut interpréter les phénomènes de structuration de la communication numérique en ligne selon le modèle des réseaux invariants d'échelle (Albert & Barabasi, 2002) qui met en avant le concept davantage de cumul dans des réseaux complexes aux relations imprévisibles. On voit ainsi progressivement émerger une structure cœur-périphérie longuement analysée sur le web (Björneborn, 2004). L'acteur qui cumule un avantage en termes de légitimité et de connexion aurait ensuite un avantage cumulatif d'attraction des nouvelles connexions.

4.2 Le web a-t-il une archive ?

La mise en place de cette étude a porté de nombreuses difficultés méthodologiques qui ne lui sont pas propres. Certes, le choix de partir de mots-clés en français permet de dégager une cartographie qui semble cohérente par rapport à notre connaissance du milieu. Nous étions partis de l'hypothèse que la révélation du processus de définition des concepts émergeant en SHS par le discours sur le web nous donnerait accès, par la grande hétérogénéité des actes communicationnels étudiés, à une analyse sociologique de la science en train de se faire (Latour, 1989, Latour et Lemonier, 1994). Et cette hypothèse est validée. Cependant, nous sommes confrontés à la problématique de l'accès aux archives du web. En effet, la communication numérique n'est accessible que par des opérateurs d'indexation. Si nous rappelons que l'analyse du web réside principalement dans l'analyse de la communication sociale des acteurs (qui reste toute de même l'objet essentiel des SIC), il faut ajouter à cela que l'analyse de la communication numérique dépend grandement des institutions d'archivages du web. Ce n'est pas une spécificité de l'analyse du web que de travailler à partir de

corpus archivés. L'histoire, l'archéologie… toutes les sciences du document dépendent des instances d'archivage. Cela implique de pouvoir mener des études sur des temporalités plus longues avec des archives accessibles et plus exhaustives que nous le permettent actuellement *InternetArchive* ou la politique du dépôt légal du web mené conjointement par la BNF et l'INA.

L'approche cartographique du web s'inscrit dans la logique du *Big Data* (Kaplan, 2015) plutôt que dans celle du *small data* qui repose sur des corpus davantage stabilisés et structurés. L'enjeu est cependant de pouvoir bénéficier de *smart data* destinés à la recherche. Il faudrait pour cela imaginer un index du web qui soit public (Lewandowski, 2014). Actuellement, ce travail d'indexation est principalement la propriété de groupes industriels comme Google. En effet, quelles que soient les formes de requêtes de ces « oracles » (direct, par *API* ou par contractualisation), nous ne pouvons obtenir qu'un échantillon infime du corpus concerné avec une modalité d'échantillonnage. Aussi, la réponse obtenue va grandement dépendre du poste informatique et de son historique. Les études digitales nécessitent de problématiser la constitution du corpus sous peine de voir leur conclusion nulle et non avenue (Dimenescu et Wieviorka, 2015). Inversement, l'idée de cadrer le corpus sur une limite de sites préétablis pose un autre problème soit, sur les critères de sélection de ces sites, soit, sur la limite de l'étude dans des champs bien connus (sites institutionnels, journaux, etc.) qui n'étudient donc aucunement la spécificité du web comme espace ouvert de communication. En effet, beaucoup d'études reprennent les champs d'études de la communication institutionnelle pour en étudier la numérisation. Si ces études peuvent mettre à jour des évolutions contemporaines, elles semblent rater, à nos yeux, le cœur même des problématiques des études sur la communication numérique, c'est-à-dire la mutation des modes d'éditorialité par l'extension de la publicisation.

Enfin, même si la communication numérique scientifique sur le web est relativement facile d'accès, elle nécessite un investissement financier et éditorial continu. Ainsi, force est de constater que toute communication qui n'est pas outillée par une équipe éditoriale en activité (qui certes peut se réduire au chercheur blogueur) disparaît du web. Une des conséquences du point précédent, c'est qu'il est très difficile de travailler sur l'histoire du web (les tentatives d'archives du web sont encore aujourd'hui très partielles et les discussions sur sa représentativité sujettes à controverse). Cette limite dans la diachronicité du web limite donc le web à être étudié sur le plan sociologique des sites actifs. Étrange statut que ce document qui ne vit que par l'activité de

ses éditeurs et qui meurt avec la dissolution de cette intention éditoriale. Entre histoire et sociologie, les études numériques sont fortement marquées par la spécificité de son objet d'étude : le web comme corpus de documents vivants. Mesurer le web, c'est donc mesurer avant tout une activité communicationnelle organisée sur une étendue temporelle d'à peine une décennie. Avec l'analyse du web, nous sommes astreints, par la nature même de l'archive, à une sociologie du temps présent de l'activité communicationnelle d'un champ d'acteurs.

4.3 Le statut ambigu de l'hypertextualité

L'hypertexte reste un concept complexe, et qui possède différentes définitions selon les contextes et les auteurs (Ertzscheid, 2002) et qui repose sur une logique de documents interreliés (Clément, 1995). Les liens exploités ne sont pas nécessairement qualifiés sémantiquement. Nous sommes dans une structure qui est celle d'un web qui n'est pas encore pleinement un web de données qui pourrait nous permettre d'attribuer à des concepts et des auteurs des URL afin de pouvoir les inclure dans la cartographie. Les méthodes du TAL permettraient éventuellement de pouvoir plus aisément distinguer des concepts récurrents. La cartographie web souffre d'une ambiguïté de l'acte hypertextuel qui n'est pas normé et qui met sur le même plan des liens parfois différents. L'acte « d'hyperlier » n'est pas toujours pleinement sémantisé du fait d'un usage fort variable quant à l'endroit où est réalisé le lien. Ainsi, l'hypertexte pose, très sérieusement, la question de son interprétation à la fois en ce qui concerne ce qu'en comprend le lecteur (Koszowska-Nowakowska et Renucci, 2011), mais aussi dans ce qu'a voulu signifier l'auteur. Il est opportun de distinguer ce qui relève de la possibilité de cliquer de son interprétation (Bouchardon, 2011). De même, le contexte visuel et éditorial de l'hypertexte (Saemmer, 2015) est difficile à appréhender en matière de dispositifs basés sur du *crawl*. L'intérêt pour des dispositifs permettant de pouvoir collecter et enrichir les données issues de la logique hypertextuelle constitue des solutions opportunes (Szoniecky & Louapre, 2015)

L'originalité de notre approche avec le projet *My web Intelligence* est de traiter des masses de données suffisamment conséquentes pour minimiser l'hétérogénéité des hyperliens en privilégiant l'acte de citation. Travailler sur des grandes masses de données oblige à l'agrégation de ces données en indices globaux. De plus, la seule manière qui soit susceptible, à nos yeux, de contrer la

spécificité de chacune des activités d'hyperliens est dans un premier temps d'intégrer dans la solution *My web intelligence* une solution dite de *readibility* qui extrait le contenu éditorial principal d'une page et ne relève que les hyperliens dans ce corps de texte. Ainsi nous sommes sûrs que les liens que nous relevons sont ceux volontairement émis par l'auteur du contenu central. Ensuite en travaillant avec de grandes masses de données, l'enjeu est de penser qu'au-delà d'un certain seuil (lequel ?), la répétition de liens qualifiés dans les corps des textes manifeste ce que l'on pourrait interpréter comme un lien de référence entre un site et un autre. Or force est de constater que très peu de sites sur un corpus se citent explicitement, car créer un lien hypertexte dans le corps du texte est un acte fort et engageant. En fait, le problème des hyperliens repose là encore plus sur leur absence que sur leur présence.

La cartographie web s'appuie davantage ici sur un travail de communication scientifique plutôt que sur des logiques de publications reconnues. En effet, le choix réalisé ici ne s'inscrit pas dans une approche scientométrique qui serait également opportune pour comprendre quelles sont les références les plus souvent citées. Une approche scientométrique mettrait davantage en avant des auteurs et des publications. Or, ici, il s'agit d'un travail de sociologie des actants qui se mobilisent autour d'une thématique de recherche et qui continuent à communiquer. Ainsi, une étude sur le web est davantage une étude sur la communication et le marketing (ici de la recherche) qu'une enquête sur les pratiques scientifiques. Notre étude nous a montré que c'est bien la présence d'une stratégie éditoriale active qui donne des traces. Et ces traces ne sont pas corrélées à toutes les pratiques réelles de la recherche. Mais comme tous les champs, la recherche est, elle aussi, sujette aux questions d'influence et de réputation, de publicité et de marketing. Le web n'est pas une représentation de notre monde, mais s'avère une représentation de nos stratégies de communication même individuelle et fugace. Pour autant, ces stratégies sont des moments essentiels de la sociologie des acteurs scientifiques.

5. Conclusion :

Notre étude permet d'obtenir une compréhension du mouvement des humanités numériques ou digitales francophones qui met en avant des acteurs et notamment des acteurs-réseau qui participent à la construction d'une forme

de transdiscipline à travers l'étude de leurs actes communicationnels. L'analyse réseaux, et les cartographies qui en sont issues, ne peuvent être totalement comprises sans connaissance externe qui fonde l'interprétation. L'approche proposée ici pour comprendre l'émergence du mouvement des humanités digitales francophones nous incite à proposer à l'avenir une méthode comparative qui mêle différentes approches de corpus et d'analyse afin d'étudier les différentes traces potentielles (méthodes mixtes). Ce travail s'inscrit à double titre dans une approche scientifique qui est celle des sciences de l'information et de la communication dans leur relation historique, méthodologique et épistémologique avec les humanités digitales (Cormerais, Lakel, Le Deuff, Pucheu, 2016). Nous avons voulu montrer que les corpus numériques pouvaient former une soure crédible pour l'analyse sociologique des mouvements scientifiques. Pour cela nous avons dû développer un outil Open Source qui respecte la méthodologique d'enquête en science sociale. Cette innovation technique, à l'état de prototype, mérite d'enrichir le débat sur la méthode[17].

Notre hypothèse était donc la suivante : les humanités numériques sont le fruit d'une hybridation entre la culture numérique et la culture académique. Les chercheurs en SHS ont expérimenté une mutation dans les processus mêmes de structuration de leurs champs comme un grand nombre de « professions du sens » à l'ère du numérique. Notre étude montre qu'il existe encore une fracture entre les centres et les périphéries qu'ils soient institutionnels, géographiques ou professionnels bien au delà de la période initiale des temps de l'innovation. Dans le processus d'innovation scientifique nous avons réussi à démontrer comment la tension centre/périphérie (que ce soit institutionnel ou géographique) structure les conditions de possibilité de la mise en réseau. La question reste pourtant posée : cette tension est-il d'ordre temporelle (du fait des débuts du mouvement), qualitative (propre à un changement sociologique profond dans la communauté des SHS) ou quantitatif (le corpus étant encore partiel). L'application de la méthodologie sur des corpus plus complets avec des index plus importants numériquement, nous offrirait une carte de plus en plus proche du territoire. De plus, la corrélation entre l'acte d'hyperlien et le contenu (analyse sémiotiquement et sémiologiquement) viendrait compléter l'étude en montrant les liens entre processus de production et stratégie d'acteurs. Enfin une étude longitudinale permettrait de clore le débat entre innovation et révolution. Nous envisageons donc deux évolutions pour cette recherche : sur un axe qualitatif, la diversité méthodologique vers des entretiens semi-directifs

[17] Nous nous inscrivons à ce titre entièrement de les tentatives de formaliser la méthodologie des études digitales tel que *TaDIRAH* du DARIAH: http://tadirah.dariah.eu/vocab/index.php

et, sur un axe quantitatif, une augmentation de la capacité de l'outil pour pouvoir travailler avec des corpus plus volumineux. Notre recherche se veut, enfin, un apport aux humanités digitales à deux niveaux. D'une part, en fondant une méthodologie de recherche de l'éditorialité numérique, d'autre part, en apportant un éclairage réflexif sur le mouvement qui dépasse le manifeste ou l'opinion lettrée.

Bibliographie

Albert,R A.-L. Barabasi, Alberto, A. (2002°) "Statistical mechanics of complex networks", *Reviews of Modern Physics* 74, 47–97

Alter, N. (2003), *L'innovation ordinaire*, Paris, P.U.F., 2003.

Berry, David. M. (2012 °) *Understanding Digital Humanities*. Palgrave Macmillan

Björneborn, Lennart. (2004). *Small-world link structures across an academic web space: a library and information science approach*, Phd Thesis, Royal School of library and Information science, Copenhagen

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. (2008 °) "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008, no. 10

Bradford, S.C. « Sources of Information on Specific Subjects ». *Engineering: An Illustrated Weekly Journal* (London), 137, 1934 (26 January), pages 85–86.

Brin, Sergey. and Page, Lawrence. (1998) "The Anatomy of a Large-Scale Hypertextual Web Search Engine*". In: Seventh International World-Wide Web Conference* (WWW 1998), April 14–18, 1998, Brisbane, Australia.

Burnard, Lou. *Qu'est-ce que la Text Encoding Initiative ?* New edition [online]. Marseille : OpenEdition Press, 2015 <http://books.openedition.org/oep/1237>.

Callon M. (1986), « Eléments pour une sociologie de la traduction », *Année sociologique*, n° 36, 1986.

Callon M., Latour, B. (1991), *La science telle qu'elle se fait, anthologie de la sociologie des sciences de langue anglaise*, Paris, La Découverte, 1991.

Clavert, Frédéric. et Noiret Serge. (2013), *L'histoire contemporaine à l'ère numerique/Contemporary History in the Digital Age*, Bruxelles, P.I.E.-Peter Lang S.A

Clavert, Frédéric. "L'apport du numérique aux sciences historiques : exemple d'une analyse computationnelle des archives Werner". *Pierre Werner et l'Europe : pensée,*

action, enseignements/Pierre Werner and Europe : His Approach, Action and Legacy, PIE Peter Lang, 2015

Clément J. (1995). « Du texte à l'hypertexte : vers une épistémologie de la discursivité hypertextuelle », dans Balpe J.-P., Lelu A., Saleh I. (dir) *Hypertextes et hypermédias : Réalisations, Outils, Méthodes*, Hermès, Paris.

Cormerais, Franck, Olivier Le Deuff, Amar Lakel, and David Pucheu. (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*, no. 8

Dacos, Marin., & MOUNIER, Pierre. (2015). *Humanités numériques* (Rapport) (p. 9782354761097). Institut français. Consulté à l'adresse https://hal.archives-ouvertes.fr/hal-01228945/document

Drucker, Johanna. *SpecLab: Digital Aesthetics and Projects in Speculative Computing*. University of Chicago Press, 2009.

Koszowska-Nowakowska, P, Renucci, F. "L'hypertexte : approches expérimentale et herméneutique", *Les Cahiers du numérique*, 2011/3 (Vol. 7), p. 71-91.

Le Deuff, O.. (2015). *Les humanités digitales précèdent-elle le numérique*? Iste éditions. Consulté à l'adresse http://archivesic.ccsd.cnrs.fr/sic_01220978/document

Domenget, J.-C., Bonaccorsi, J., & Carayol, V. (2016). Introduction au dossier « Humanités numériques et SIC ». *Revue française des sciences de l'information et de la communication*, (8). Consulté à l'adresse https://rfsic.revues.org/1984

Ertzscheid, O. (2002). *Le livre, le lien, le lieu : les enjeux cognitifs et stylistiques de l'organisation hypertextuelle*, Thèse de doctorat soutenue en octobre 2002 à l'Université Toulouse Le Mirail.

Flichy, P. (2003), *L'innovation technique : récents développements en sciences sociales, vers une nouvelle théorie de l'innovation*, Paris, La Découverte, 2003.

Gallezot, G, Le Deuff, O « Chercheurs 2.0 ? », *Les Cahiers du Numérique*. « Enjeux actuels de la communication scientifique ». Paris Hermes-Lavoisier, vol.5, n° 2, p. 15-31

Grandjean, M. (2016). "A social network analysis of Twitter: Mapping the digital humanities community", *Cogent Arts & Humanities*, 2016, 3: 1171458.

Gold, Matthew K. *Debates in the Digital Humanities*. University of Minnesota Press, 2012.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. (2014 °) 'ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.' *PLOS ONE* 9, no. 6

Kaplan, F (2015). « A Map for Big Data Research in Digital Humanities. » *Frontiers in Digital Humanities*, 2, no. 1 (2015). doi:10.3389/fdigh.2015.00001.

Korzybski, A (2015). *Une carte n'est pas le territoire : Prolégomènes aux systèmes non-aristotéliciens et à la sémantique générale*. Paris : Editions de l'Eclat

Koszowska-Nowakowska, P., & Renucci, F. (2012). "L'hypertexte : approches expérimentale et herméneutique". *Les Cahiers du numérique*, Vol. 7(3), 71-91.

Kleinberg, J. M. (1999). « Authoritative sources in a hyperlinked environment », *Journal of the ACM (JACM)*, 46(5), 604–632.

Lakel A., Massit-Folléa F., « Société civile et gouvernance de l'internet : la construction d'une légitimité ambiguë », *Hermès, La Revue* 1/2007 (n° 47), p. 167-176

Latour B., Lemonnier, P (1994), *De la préhistoire aux missiles balistiques, l'intelligence sociale des techniques*, Paris, La Découverte

Latour, B. (1989) *La Science en action*, traduit de l'anglais par Michel Biezunski ; texte révisé par l'auteur, Paris, La Découverte, « Textes à l'appui. Série Anthropologie des sciences et des techniques »

McCarty, W. (2014). *Humanities Computing*. Reprint. Basingstoke: Palgrave Macmillan, 2014.

Ouakrat, Alan, and Julien Mésangeau. (2016) "Resocialiser les traces d'activités numériques : une proposition qualitative pour les SIC." *Revue française des sciences de l'information et de la communication*, no. 8, 2016). doi:10.4000/rfsic.1795.

Merzeau, Louise. "La médiation identitaire." Revue française des sciences de l'information et de la communication, no. 1 (July 6, 2012). https://doi.org/10.4000/rfsic.193

Milard, B (2014), « Quelles sociabilités derrière les références bibliographiques ? Citations et relations sociales », *Socio-logos*, 8 | 2013, http://socio-logos.revues.org/2802

Nowviskie, B (2014) "On the Origin of 'Hack' and 'Yack'", *Journal of Dgital Humanities*, Vol.3 N° 2, http://journalofdigitalhumanities.org/3-2/on-the-origin-of-hack-and-yack-by-bethany-nowviskie/

Lewandowski, Dirk. "Why We Need an Independent Index of the Web." arXiv :14,052,212 [Cs], May 9, (2014 °). http://arxiv.org/abs/1405.2212 .

Plantin, J.C "Les chercheurs en SHS rêvent-ils de code informatique ? Trois questions préliminaires à l'apprentissage du code", in Le Deuff O. (dir.), *Le temps des humanités digitales*, FYP éditions, Limoges, 2014

Saemmer, A (2015 ° "Hypertexte et narrativité", *Critique* 2015/8 [n° 819-820], p. 637–652

Schreibman, Susan, Ray Siemens, and John Unsworth. *A Companion to Digital Humanities*. Illustrated edition. Wiley-Blackwell – an imprint of John Wiley & Sons Ltd – , 2008.

Scott, J. (2012). *Social Network Analysis* (3 edition). Los Angeles: SAGE Publications Ltd.

Szoniecky, S. & Louâpre, M. [2015]. Outillages numériques pour les humanités : cartographier des réseaux d'influences. Présenté à ISKO — Magreb 2015 : *Organisation de la Connaissance dans la perspective des Humanités Numériques : recherches et applications*. Consulté à l'adresse https://hal.archives-ouvertes.fr/hal-01220142/document

Thély, N., Serres, A., Le Deuff, O. [2014] "Le THATCamp comme nouvelle forme de communication scientifique ?" In *THATCamp Saint-Malo 2013 : Non actes de la non conférence*. La Non-Collection. Paris : Éditions de la Maison des sciences de l'homme < http://books.openedition.org/editionsmsh/2183.>

Thély, Nicolas. (2014) "En deçà des images. Logique informatique et recherches en esthétique" in *Le Temps des humanités digitales. La mutation des sciences humaines et sociales.* Dir. Olivier Le Deuff, éditions FYP, Paris

Vial, Stéphane. (2016) "Le tournant design des humanités numériques." *Revue française des sciences de l'information et de la communication*, no. 8, 2016). doi:10.4000/rfsic.1828.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications* (1 edition). Cambridge; New York: Cambridge University Press.