



La réimpression dans la presse francophone du 19^e siècle: outils numériques et enjeux de mesure de viralité

Guillaume Pinson, Julien Schuh, Pierre-Carl Langlais

► To cite this version:

Guillaume Pinson, Julien Schuh, Pierre-Carl Langlais. La réimpression dans la presse francophone du 19^e siècle: outils numériques et enjeux de mesure de viralité. Bases de données et outils numériques : des révélateurs de l'imprimé et du littéraire, May 2016, Montréal, Canada. <halshs-01485991>

HAL Id: halshs-01485991

<https://halshs.archives-ouvertes.fr/halshs-01485991>

Submitted on 9 Mar 2017

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.

« LA REIMPRESSION DANS LA PRESSE FRANCOPHONE DU 19^E SIECLE : OUTILS NUMERIQUES ET ENJEUX DE MESURE DE VIRALITE »

« Prenez un journal.
Prenez des ciseaux.
[...]
Découpez l'article.
Découpez ensuite avec soin chacun des mots [...].
Agitez doucement.
[...]
Le poème vous ressemblera »
Tristan Tzara, « Pour faire un poème dadaïste »

Introduction : viralité et circulation médiatique

Notre réflexion vise à poser les bases méthodologiques, théoriques et pratiques à l'étude de la « viralité » dans la presse francophone au XIX^e siècle, en faisant l'hypothèse que l'ère du journal transforme en profondeur les modalités d'existence des textes, en valorisant certaines formes, certains genres, certains thèmes pour leurs capacités à être reproduits dans des contextes variés. La mondialisation médiatique, qui connaît une phase de développement accéléré au XIX^e siècle, engendre l'expansion de réseaux d'information interconnectés comme jamais dans l'histoire de l'humanité, inédits dans leur capacité à lier les lecteurs entre eux. Les développements du télégraphe et du chemin de fer dans le premier tiers du siècle, des lignes maritimes à la vapeur, la pose des premiers câbles sous-marins à partir de la fin des années 1850, le développement des agences de presse, l'organisation de plus en plus rigoureuse du travail de l'information, sont autant de transformations qui confirment cette interconnexion. Et dans ce processus se trouvent aussi engagés la littérature, les grands imaginaires sociaux et les sensibilités collectives, qui ne sont pas du tout extérieurs au système médiatique, comme s'ils arrivaient par surcroît. Bien au contraire : tout conduit à une culture médiatique qui va se trouver placée, par le fait même de cette interconnexion, en tension entre standardisation et recherche de l'écart, entre reproduction et innovation, entre le local et l'international ; la littérature a ainsi joué un rôle capital dans le développement du nouvel espace-temps de la modernité médiatique, elle a constitué une forme de traduction sensible de ces tensions, elle-même engagée aussi bien dans la part « industrielle » de l'écriture que dans son versant « autonomiste », un peu en retrait.

Or, il est difficile d'entrer à l'intérieur d'un tel système, de tenter de le comprendre et d'en analyser le fonctionnement. Par où commencer ? Que choisir dans la masse ? De toute évidence sur bien des points les frontières nationales ou régionales, même si elles demeurent historiquement importantes et s'imposent comme telles par les contemporains, doivent être dépassées. Les humanités numériques invitent à les faire éclater (ne serait-ce que par volonté par la suite de mieux y revenir), et à l'origine de cette attention mondialisée aux phénomènes médiatiques il y a sans doute le développement foudroyant ces dernières années des corpus de presse numérisés. Tout comme notre actuelle culture numérique, la culture médiatique est fondée sur le *lien*, et peut-être faut-il commencer par en comprendre le fonctionnement, la valeur, tenter de retourner à hauteur d'expérience des individus, tenter aussi de saisir les circulations et de les mesurer plus finement, plus systématiquement et moins intuitivement. L'une des hypothèses que l'on peut formuler est que cette culture du lien était loin d'échapper aux contemporains, et même qu'au contraire elle constituait un élément central de la valeur des objets médiatiques. Vers le milieu du XIX^e siècle, alors que le journal quotidien modifiait partout en occident la vie sociale, politique, professionnelle, familiale, le lecteur voulait appréhender la place qu'il occupait dans le monde en se « connectant » à ses concitoyens par l'intermédiaire du journal. Une part de son expérience de lecture prenait sans

doute son sens dans le fait qu'il savait avoir en main un journal qui se trouvait lui-même connecté à ce monde ; que ce journal n'était pas un objet clôt dans son texte mais au contraire traversé de textes mis en circulation, échangés, recopiés, réimprimés, sautant de journal en journal. Exemple entre mille de ce premier système médiatique global dans sa dimension perçue au microscope (fig. 1), *La Minerve* de Montréal pouvait ainsi publier dans ses colonnes une petite actualité tirée du *Courrier des États-Unis* de New York, annonçant la publication des *Mystères de Londres* de Paul Féval à Paris dans le *Courrier français*, et son imminente réédition dans la *Semaine littéraire*. Un journal de Montréal se trouvait ainsi connecté à un journal parisien par l'intermédiaire d'un journal francophone new-yorkais. La valeur d'un tel entrefilet ne reposait peut-être pas tant seulement sur le « contenu » de l'annonce d'une publication dérivée de l'immense succès des *Mystères de Paris*, mais aussi sur le fait que le lecteur de Montréal, grâce à son journal, parvenait à se situer au sein du système médiatique, à prendre conscience de sa propre « localisation » à travers l'œuvre collective formée de la liaison des journaux entre eux.

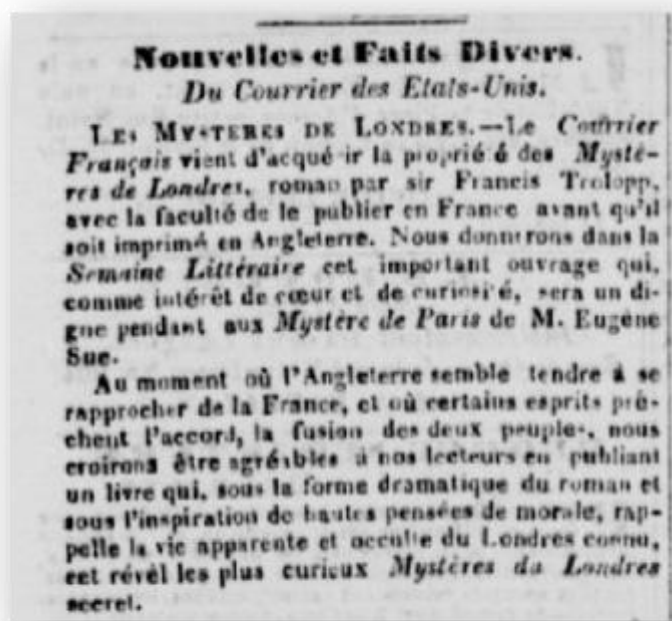


Figure 1. *La Minerve* (Montréal), 4 janvier 1844.

Tel est donc notre projet : en nous situant à une échelle internationale, penser l'internationalisation de la culture médiatique francophone en nous situant sur l'axe nord-atlantique, que l'un des cosignataires de cet article a étudié de façon générale dans un ouvrage récent¹, et en appliquant les méthodologies de mesure de la « viralité » médiatique. Dans le cadre de la phase II du projet Médias 19, qui fait l'objet d'une demande ANR en France, nous souhaitons appliquer en effet la notion de viralité à un corpus de presse de France, de Belgique, du Canada et des États-Unis (New York et La Nouvelle-Orléans) (fig. 2 ; en rouge les villes considérées en priorité).

¹ Guillaume Pinson, *La culture médiatique francophone en Europe et en Amérique du Nord, de 1764 à la veille de la Seconde Guerre mondiale*, Québec, Presses de l'Université Laval, coll. « Cultures québécoises », 2016.

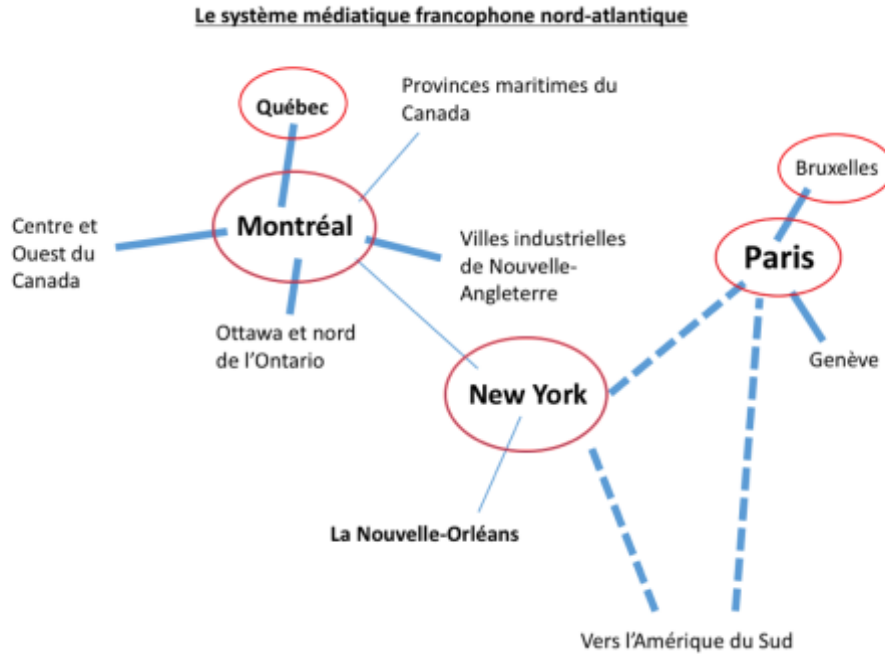


Figure 2. Le système médiatique francophone nord-atlantique.

Les outils numériques permettent désormais de plonger dans cette culture du lien qui implique de nouvelles formes rhétoriques, qui peuvent émerger de manière spontanée (certains textes circulant sans que leurs auteurs en aient conscience) ou consciente (une forme d'écriture viralisante), avec pour conséquence l'apparition de nouvelles formes de textes journalistiques aux statuts très divers (poèmes, contes, chroniques, listes...), qui partagent la particularité d'être repris, reproduits, découpés, fragmentés dans d'innombrables numéros de périodiques, à des échelles temporelles et géographiques souvent insoupçonnées. Le statut même des textes dans les journaux évolue en conséquence, les questions d'anonymat, de plagiat, de propriété littéraire prenant des dimensions nouvelles devant la possibilité d'une diffusion non prévisible et souvent invisible de fragments textuels. Plusieurs projets de repérage de la diffusion de textes dans la presse anglophone existent, et dans notre cas nous nous inspirons grandement de la méthodologie du projet « Viral Text » de Ryan Cordell, de l'université Northeastern. Cordell a fondé la première phase de son étude sur la presse américaine d'avant la guerre de Sécession, exploitant les bases de données de la presse numérisée à la Bibliothèque du Congrès ; l'algorithme développé permet de repérer, comme une recherche de plagiat, les innombrables réimpressions dans ce corpus, et de détecter les textes viraux ainsi que les réseaux de circulation². Il est en train d'étendre son projet à une dimension internationale, celle de la presse publiée en anglais au XIX^e siècle aux États-Unis, en Grande-Bretagne et en Australie notamment.

Notre projet voudrait transposer ces questions dans le domaine francophone, en analysant les conditions de production et de diffusion de ces textes, en proposant des typologies possibles (au niveau générique, formel, stylistique), entreprises nécessaires en amont d'une entreprise d'automatisation informatique du repérage des formes de viralité dans cette presse. Nous avons déjà effectué quelques plongées expérimentales, avec l'aide d'un jeune chercheur qui a soutenu une thèse très originale au Celsa de Paris-Sorbonne sur la chronique boursière dans la presse française du XIX^e siècle, Pierre-Carl Langlais, que nous avons intégré à l'équipe. Exploitant les outils

² Voir Ryan Cordell, « Reprinting, Circulation, and the Network Author in Antebellum Newspapers », *American Literary History*, vol. 27, Issue 3 Fall 2015, p. 417-445. URL : <http://ryancordell.org/research/reprinting-circulation-and-the-network-author-in-antebellum-newspapers/>

numériques qui nous intéressent, Pierre-Carl a aussi contribué à la réflexion fondamentale que nous effectuons dans ce projet.

Enjeux méthodologiques et techniques

Les bases de données de presse disponibles en ligne permettent déjà de vérifier de manière empirique la pertinence de ce concept de viralité. Le site Retronews (<http://www.retronews.fr>), qui reprend les numérisations de plusieurs titres de presse française de Gallica en rajoutant divers outils de recherche (étiquetage fin du contenu textuel, structuration par article...) ou le moteur de recherche d'Europeana newspaper (<http://www.theeuropeanlibrary.org/tel4/newspapers>) nous ont offert un premier terrain d'expérimentation par la recherche de mots clés récurrents dans les numéros de périodiques numérisés.

L'exemple de la diffusion des « bons mots » ou des phrases célèbres se prête particulièrement bien à des vérifications de ce type. En travaillant sur la circulation du discours anarchisant à la fin du XIX^e siècle, nous avons ainsi pu prendre l'exemple d'une phrase prononcée par Laurent Tailhade, poète pamphlétaire du *Mercure de France*, en réponse à la question d'un journaliste qui l'interrogeait sur l'attentat de Vaillant à la Chambre des Députés le 9 décembre 1893 : « Qu'importent les victimes, si le geste est beau ? » Cette réponse, publiée initialement dans *Le Journal* du 10 décembre, est immédiatement reprise et citée à de nombreuses reprises dans la presse de l'époque : lapidaire, rythmée comme un alexandrin, paradoxale, choquante, elle a tout pour devenir virale (fig. 3). La presse française et internationale la reprend des dizaines de fois ; elle devient un cliché de l'histoire de l'anarchisme. Mais, comme souvent dans ce type de diffusion, cette phrase est déformée, citée de seconde main, traduite maladroitement : la « viralité » implique ces déformations, liées à la décontextualisation de ces fragments textuels. L'information circule sans retour aux sources, de manière rhizomique.

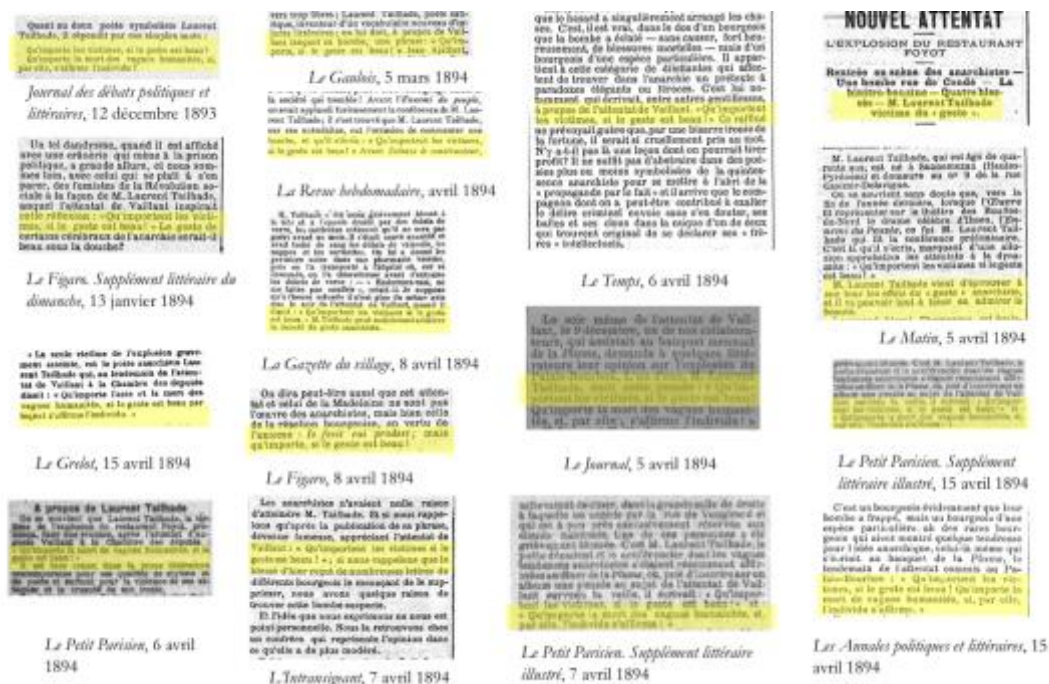


Figure 3. Diffusion virale de la phrase de Laurent Tailhade dans la presse française.

³ Propos recueillis par Paul Brulat, journaliste au *Journal* de Fernand Xau, lors du Banquet Rodin du 9 décembre 1893 et reproduits dans *Le Journal* du 10 décembre 1893, cité dans Gilles Picq, *Laurent Tailhade ou de la provocation considérée comme un art de vivre*, Paris, Maisonneuve et Larose, 2001, p. 341.

On ne peut s'empêcher de comparer ces formes de diffusion et de réappropriation aux formes que prend la diffusion de l'information dans notre culture numérique. Les objets culturels sur internet prennent d'emblée un caractère interactif : ils affichent leur viralité, la possibilité de les rediffuser par le biais de boutons de partage sur les réseaux sociaux. Mais les formes plus anciennes n'étaient pas moins interactives, et leur matérialité conduisait à certaines formes de diffusion, de reproduction, de citation, de remédiation⁴. Plutôt que de twitter une phrase, on la mémorisait pour la citer en société ou la placer dans un article ; plutôt que de partager un LOLCAT, on faisait des albums de chromos. Les contraintes du périodique entraînent l'apparition de formes de textes et d'images adaptés à ces processus de dissémination non contrôlée. Ces nouvelles modalités d'utilisation des textes s'incarnent dans une forme particulièrement parlante : la coupure de presse.

La pratique de la coupure de presse symbolise les nouvelles normes de production et d'utilisation des textes dans les périodiques : courts, décontextualisables, fonctionnant en série. De véritables industries remplissant le rôle des « alertes Google » apparaissent à cette époque. La revue symboliste *La Plume* du 15 février 1892 publie ainsi une réclame pour une entreprise, « Le Courrier de la Presse », qui se charge de découper des articles dans tous les journaux pour les fournir aux écrivains et artistes qui s'abonnent au service (fig. 4) : toute mention de leur nom leur sera communiquée par un bureau de lecteurs professionnels, disposant d'une liste de mots clés à rechercher dans les journaux internationaux qu'ils découpent pour leurs clients. De véritables pratiques de veille médiatique se mettent en place, les textes sélectionnés pouvant faire l'objet d'une réappropriation dans une presse autophage. On assiste à l'émergence d'une culture médiatique de la réappropriation qui transforme les modalités de l'imitation, de la réécriture, de la transtextualité, les conditions de l'originalité...



Figure 4. Page de réclame, *La Plume*, 15 février 1892, n.p.

L'analyse empirique de ces pratiques, par la recherche de mots clés dans les bases existantes, ne fonctionne que pour des exemples isolés. Comment analyser ces objets qui apparaissent à une échelle temporelle, géographique, textuelle extrêmement large ? Nous avons envisagé au départ la construction de corpus textuels extrêmement structurés : encodage en TEI, selon un schéma élaboré en interne, avec des textes entièrement relus, nettoyés, étiquetés... C'est le choix opéré par exemple par le projet OBVIL (<http://obvil.paris-sorbonne.fr>). On se heurte cependant rapidement à la question de la possibilité matérielle d'établir une telle base : le corpus représente des millions de pages. D'où le choix d'une autre option : plutôt que de créer un corpus adapté, adapter les corpus numérisés disponibles à nos besoins, en créant des outils permettant de nous approprier des résultats malgré la multitude de formats disponibles, malgré les manques et les erreurs de numérisation.

Ce choix impliquait la question de la mise à disposition des données pour la recherche. Une institution comme la BnF, financée par l'État et l'Union européenne, avec une mission de mise à disposition au public du patrimoine imprimé français, a une attitude paradoxale concernant la

⁴ Jay David Bolter et Richard Grusin, *Remédiation. Understanding New Media*, Boston, MIT Press, 2000.

diffusion des savoirs : Gallica permet d'accéder à beaucoup de documents libres de droit, mais les conditions de leur réutilisation sont souvent très contraignantes, et les utilisateurs ne disposent pas d'API permettant d'interroger les données comme ils le souhaitent, ni de dépôts de données brutes. La notion anglo-saxonne de copyright est souvent utilisée par les institutions pour affirmer une propriété sur les images et données issues du traitement de documents libres de droits⁵ ; le site retronews, établi à partir de numérisations de presse dans Gallica, montre à quel point la volonté (souvent politique) de rentabiliser ces investissements conduit à proposer des sites à moitié libres, avec un contenu premium payant permettant la réutilisation des données ou un accès plus large. Les débats actuels en France autour de la possibilité d'utiliser pour la recherche les métadonnées montrent à quel point l'idée d'une propriété de ces objets est prégnante. Gallica, par conséquent, n'offre pas tous les outils d'appropriation de ses numérisations que l'on pourrait espérer : on peut télécharger les documents un par un, en mode image ou texte si disponible, en cochant à chaque fois une case indiquant la responsabilité légale de l'utilisateur, mais en aucun cas télécharger l'ensemble du contenu d'un périodique par exemple.

Pierre-Carl Langlais a développé plusieurs outils numériques que nous avons utilisés comme base et retravaillés pour contourner ces contraintes. Le premier de ces outils était le logiciel Pyllica, écrit en Python, qui permettait de récupérer automatiquement les textes ou les images de Gallica⁶. La nouvelle formule de Gallica les a rendus obsolètes, et nous avons réécrit le code en intégrant quelques options supplémentaires, permettant par exemple de sélectionner des images de meilleure qualité (Gallica bloque la qualité maximale qu'on peut télécharger gratuitement, alors qu'a été mis en place sur le serveur la norme IIIF permettant l'échange d'images de très haute définition – encore un exemple de paradoxe). Ces outils ont déjà permis de récupérer des collections de périodiques numérisés pour une analyse quantitative.

Mais toutes les institutions ne partagent pas la retenue de la BnF : Europeana newspapers a produit plusieurs API permettant d'interroger les métadonnées de ses collections de journaux européens, et surtout a commencé à mettre en ligne des collections entières de données, au format json, reprenant le contenu textuel des journaux, que l'on peut télécharger directement⁷.

À partir de ces collections de données, Pierre-Carl Langlais a mis au point un autre logiciel en Python, qui permet de repérer les réimpressions de textes dans les corpus de presse⁸. À partir des corpus d'Europeana, on peut faire ressortir les reprises textuelles de manière plus ou moins précise. Le principe utilisé est le même que celui de Cordell : étant donné que la numérisation et l'océrisation de ces journaux produisent beaucoup de scories, il fallait être capable de repérer des reprises malgré les erreurs d'OCR ou les modifications apportées au texte (réécriture, coupures, etc.). En repérant les concentrations de termes identiques, le logiciel permet de présenter des réimpressions possibles. Cet outil n'est pas limité à la presse : en intégrant dans les corpus des ouvrages, on peut repérer les prépublications, et vérifier par exemple que le genre des « souvenirs littéraires » est essentiellement un genre médiatique, qui existe d'abord comme chronique dans la presse, avec les contraintes médiatiques que l'on imagine, avant d'être repris en volume (fig. 5).

⁵ Pierre-Carl Langlais, « L'inverse du piratage, c'est le copyfraud, et on n'en parle pas », Rue89, 14 octobre 2012, URL : <http://rue89.nouvelobs.com/blog/les-coulisses-de-wikipedia/2012/10/14/linverse-du-piratage-cest-le-copyfraud-et-personne-nen-parle-228658>

⁶ On trouvera une description du logiciel à cette adresse : URL : <http://prelia.hypotheses.org/441> Le logiciel est disponible sur GitHub, URL : <https://github.com/Dorialexander/Pyllica>

⁷ URL : <http://data.theeuropeanlibrary.org/download/newspapers-by-country/README.html>

⁸ Pierre-Carl Langlais, « Mining XIXth Century Periodicals : an Experimental Toolbox », URL : http://www.dhbenelux.org/wp-content/uploads/2016/05/55_Langlais_FinalAbstract_DHBenelux2016_long.pdf

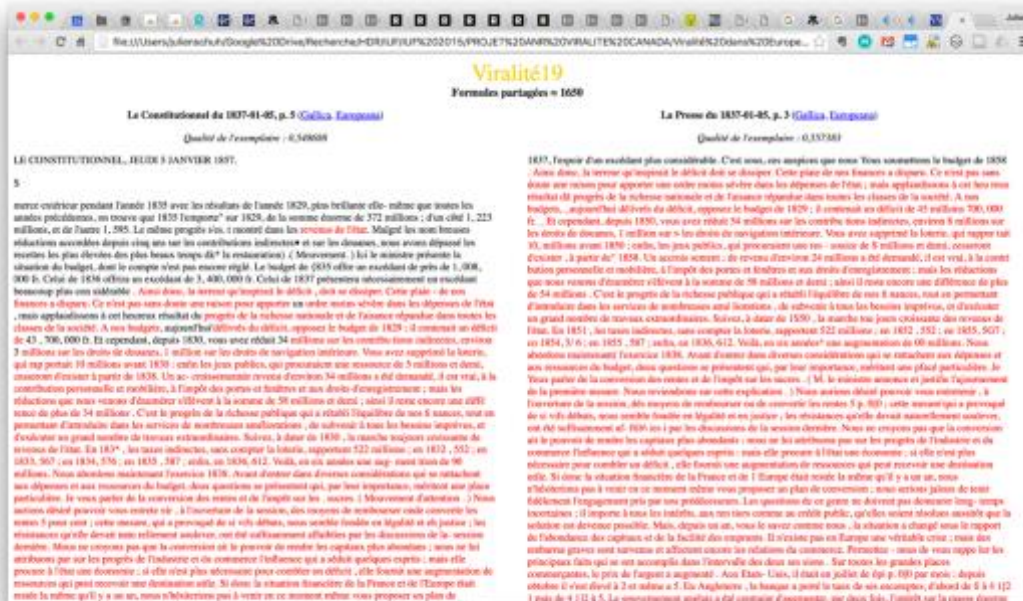


Figure 5. Repérage de fragments textuels identiques dans la presse.

À la croisée de la théorie et de la pratique

Ce type d'outils est développé par plusieurs équipes. Les chercheurs du projet OBVIL ont ainsi développé des outils de repérage de réutilisation de textes dans des corpus restreints et déjà encodés, en reprenant les catégories de Genette concernant l'intertextualité et la transtextualité pour repérer deux phénomènes : « text reuse » and « approximate citations⁹ ». Notre approche se voudrait plus expérimentale : plutôt que de poser des types de réappropriation textuelle et d'interroger un corpus déjà construit, nous souhaiterions développer une démarche permettant de faire apparaître des phénomènes auxquels nous n'avions pas forcément pensé. Il faut imaginer le développement d'outils numériques et d'outils méthodologiques de manière conjointe : les hypothèses que nous pouvons faire sur la notion de viralité, de réimpression, sur les rythmes de reproduction de textes dans la presse, sur leur réappropriation, dépendent étroitement des corpus disponibles, de la précision des outils utilisés, des choix de calibrage des algorithmes.

La question de la granularité de l'analyse est ici essentielle : peut-on repérer des mots viraux ? des tournures de phrase virales ? des configurations grammaticales ? des configurations rhétoriques (l'articulation de plusieurs types de paragraphes entre eux par exemple) ? Autant de questions qui ne trouveront de réponse que par des analyses de corpus très larges. Une fois que ces objets seront délimités, on pourra envisager de calculer l'indice de reproductibilité d'un objet discursif, à partir des caractéristiques qui accentuent ou diminuent ses chances de reproduction : quels objets font l'objet d'une réimpression, et pourquoi ? Quelles caractéristiques ont-ils en commun ? quels sont les rythmes de la réimpression, leur amplitude ? Pour cela, nous envisageons également d'articuler les phénomènes de réimpression avec d'autres bases de données en cours de constitution, comme des bases de personnel de presse, de lieux de sociabilité, de lieux de publication... Les relations entre périodiques, l'importance de tel ou tel écrivain dans le champ littéraire, doivent sans doute être prises en compte dans l'analyse de la diffusion virale des textes ; ces éléments doivent faire partie des métadonnées à intégrer dans nos corpus.

⁹ Jean-Gabriel Ganascia (LIP6), Pierre Glaudes (CELFF XVI-XXI), Andrea Del Lungo (ALITHILA), « [Automatic Detection of Reuses and Citations in Literary Texts](#) », *Literary and Linguistic Computing*, 2014, doi: 10.1093/lc/fqu020

Un des grands problèmes de la construction de ces bases de textes numérisés est en effet celui de la standardisation et des choix de grilles de métadonnées que l'on applique sur des objets par nature très complexes¹⁰. À terme, nous souhaiterions développer un format permettant de nous réappropriier les bases textuelles de différentes institutions pour les mettre en forme selon un schéma XML pérenne sans perdre cette complexité. Pour créer ce type de base de données, on doit combiner plusieurs niveaux de métadonnées : METS (décrivant les données bibliographiques des objets et leur structuration interne), ALTO (le contenu textuel repéré par rapport à l'image de page), TEI (structuration fine du texte) et MODS (permettant de décrire la hiérarchie interne des articles, etc.) — processus désigné par les initiales OLR¹¹. Europeana Newspapers a proposé un profil XML pérenne pour les périodiques, ENMAP¹² : Ce schéma permettra des recherches très fines (repérages des genres d'articles, des signatures, des titres...) ; mais pour que cette solution soit envisageable, nous réfléchissons au développement d'outils d'étiquetage automatisé des corpus par le biais de la reconnaissance de topics ou de repérages de régularités (mise en page, etc.). Pierre-Carl Langlais se propose de « sémantiser » les textes à partir des différentes dimensions textuelles (lexicales, grammaticales, formules récurrentes, etc.) pour reconnaître automatiquement des genres, voire observer l'émergence de genres nouveaux, de catégories de textes dont nous ne soupçonnons pas l'existence. On pourrait alors repérer des configurations discursives multiples, sans réduire la complexité des objets par un étiquetage par métadonnées en amont.

La viralité à la conscience des contemporains ?

Alors que nous progresserons dans ces travaux, il sera nécessaire de s'interroger sur la culture de la viralité, qui devrait permettre par exemple d'en éclairer les fonctions et les valeurs sur l'axe atlantique. De vraies questions surgissent : quelles sont les nouvelles qui traversaient l'Atlantique, quels microrécits collectifs pouvaient unir les communautés de lecteur des deux côtés de l'océan ? et inversement qu'est-ce qui ne « passait pas », était par exemple viral en Europe mais ne traversait pas l'Atlantique ? On a tendance à imaginer la viralité transatlantique dans le sens Est-Ouest, mais y-a-t-il un flux viral dans le sens inverse ? et qu'en est-il de l'axe nord-sud, Québec-La Nouvelle-Orléans ? Peut-on établir certaines grandes scissions chronologiques de ces déplacements, les croiser aux développements technologiques (le téléphone à partir de la fin du siècle par exemple, si notre projet remonte jusque-là) et à la professionnalisation du journalisme à travers par exemple l'importance croissante du rôle des correspondants de presse ?

Un autre enjeu « culturel » de la viralité sera de tenter d'en saisir les modes d'appropriation par les contemporains. Les travaux en *big data* engendrent nécessairement l'appel à une réflexion sur les expériences de réception ; lorsque nous présentons ce projet, on nous demande souvent si nous ne sommes pas en train de construire une sorte d'artefact quantitatif, du moins de repérer un phénomène certes tangible mais invisible aux lecteurs de l'époque, dont on peut difficilement apprécier la valeur, voire qui n'en avait pas beaucoup aux yeux des contemporains. Ces réticences ne nous semblent pas constituer une réfutation majeure de notre projet ; d'abord parce que ce sont d'abord notre propre réception, notre propre regard sur les objets médiatiques qui est peut-être largement faussé par les lunettes des mots-clefs et des noms d'auteurs connus, lesquels nous font entrer avec certains biais dans la culture médiatique du passé ; ensuite parce que la moindre fréquentation de la presse ancienne fait immédiatement surgir cette immense marquerie de petits textes en circulation, de ces « paragraphes mobiles » qu'a très bien repérés Will Slauter dès la seconde moitié du XVIII^e siècle dans la presse européenne, mais jugés sans valeur par notre souci

¹⁰ Voir <http://dragonfly.hypotheses.org/909>

¹¹ Il existe des guides méthodologiques https://www.google.fr/url?sa=t&rcct=j&q=&esrc=s&source=web&cd=2&ved=0CCsQFjABahUKEwjwvXD4NDHhAhXIbhQKHcAwDhg&url=http%3A%2F%2Fwww.digitizationguidelines.gov%2Fguidelines%2Fdigitize-technical.html&ei=3O_iVfDoGsjdUcDhuMAB&usq=AFQjCNHX_ijna3W6k4Efo-ch55mXqZB3Bw&sig2=y7II.dsJt4HhXpJnBNwVKIA&cad=rjt

¹² Décrit dans un document disponible à cette adresse : URL : <http://www.europeana-newspapers.eu/wp-content/uploads/2014/08/D5.2-Europeana-Newspapers-METS-ALTO-Profile-ENMAP-DRAFT.pdf>

de repérer des écarts, de la nouveauté¹³ ; enfin parce que le XIX^e siècle a lui-même développé une réflexion sur la « viralité » et qu'il a été très tôt frappé par la puissance de la reproductibilité médiatique. Les scènes à cet égard sont innombrables, souvent traduites par la figure du « coupeur de journal », comme le présente par exemple Edmond Texier en 1854 dans *Paris-journaliste* : « le coupeur de journal [est] une des chevilles principales, un des rouages importants de la grande machine de la publicité, [...] chargé de faucher dans le champ de l'actualité tout ce qui lui paraît neuf ou intéressant. D'autres écrivent, lui colle¹⁴. » Et encore en 1924, André Baillon, dans un récit de souvenirs fictionnalisés, raconte la petite scène suivante à propos du secrétaire de rédaction :

À grands coups de ciseaux, M. Sinet découpe dans un journal peu à peu les deux colonnes d'articles :

- Que faites-vous là, M. Sinet ?
- Moi ? j'écris une étude¹⁵.

Le collage hante l'imaginaire, au fond toute la théorie du discours social de Marc Angenot repose sur l'étude de ces vastes mouvements de « contamination » des discours entre eux, au sein desquels la pratique de la réimpression apparaît comme une manifestation « technologique » associée¹⁶. Et l'on pourrait aller chercher aussi dans la philosophie et la sociologie contemporaines la part de réversibilité que contient cette frénésie de la réimpression, celle qui fait que l'on considère ces circulations de formes et d'idées dans leurs effets sur la « contagion » mentale (la terminologie « virale » n'est pas loin), menant par exemple aux *Lois de l'imitation* que Tarde publie en 1890 et qui contient de nombreuses pages sur les effets du journal dans la constitution de vastes communautés de lecteurs dématérialisées.

Pour conclure, nous pouvons simplement revenir sur le terme même de viralité, et ce qu'il implique. Les discussions sur ce projet ont également fait émerger une méfiance terminologique du côté européen, méfiance que les collègues américains partagent moins : le caractère morbide de la notion de « viralité » conduit à l'écarter comme un phénomène relevant d'une maladie de la communication, dont l'exemple type serait le même à l'âge d'internet, photographie de chat ou article au titre tapageur.

Le soupçon vient finalement des modalités mêmes de production et de diffusion de ces objets textuels (et graphiques) : l'idée d'une diffusion non hiérarchique, selon un modèle décentralisé, va contre nos manières habituelles de réfléchir à la diffusion de la culture (parce que notre modèle de la culture est celui d'une culture élitiste, validée et diffusée par un certain nombre d'institutions comme l'école, les musées, les bibliothèques...). L'intérêt de ce type de base de données provient précisément de l'effacement (partiel mais réel) de la hiérarchie culturelle traditionnelle par la construction d'un corpus non hiérarchisé, aussi bien du point de vue du choix des objets (on numérise en masse, indépendamment des considérations de valeur esthétique, historique...) que de celui de leur mise en forme (l'écrasement des données matérielles, certes problématique, est cependant une chance pour mettre sur le même plan des discours qui habituellement ne sont pas mis en communication, parce qu'ils sont tous réduits à des suites de chaînes de caractères).

On pourrait préférer le terme d'« objets culturels récurrents » (images, termes, modèles narratifs ou rhétoriques) pour désigner ces phénomènes, objets qu'il faudra analyser en prenant en compte les conditions de leur reproduction, aussi bien à un niveau matériel (quelles techniques de reproduction sont utilisées et comment l'évolution des techniques conditionne les réemplois),

¹³ Will Slauter, « Le paragraphe mobile : circulation et transformation des informations dans le monde atlantique du XVIII^e siècle », *Annales. Histoire, sciences sociales*, vol. 76, n° 2 (2012), p. 363-389.

¹⁴ Edmond Texier, *Paris-Journaliste, par les auteurs des Mémoires de Bilboquet*, Série *Les Petits Paris*, Paris, Taride, 1854, p. 19-20.

¹⁵ André Baillon, *Par fil spécial. Journal d'un secrétaire de rédaction*, Paris, Rieder et Cie, 1924, p. 68-69.

¹⁶ Marc Angenot, *1889 : un état du discours social*, Montréal / Longueuil, Éditions du Préambule, 1989, réédition sur Médias 19, URL : <http://www.medias19.org/index.php?id=11003>

rhétorique (quelles formes prennent ces objets), historique (quels rythmes, quelles scansions, quelles relations avec les événements historiques), intellectuel (quel contexte permet la réinterprétation ou la réévaluation de certains objets), légal (quelles lois autorisent ou limitent ces réappropriations) qu'économique (quel est le coût de ces transferts) — autant d'éléments qu'il faudra réussir à intégrer aux bases de données, autant d'éléments qui forment une culture en soi, une culture de la remédiation qui est essentiellement médiatique et qui est peut-être tout simplement l'autre nom de la modernité.

Guillaume Pinson (U. Laval) et Julien Schuh (U. Paris Nanterre, IUF), avec la participation de
Pierre-Carl Langlais (CELSA)