



HAL
open science

Archiver le web littéraire : défis méthodologiques et conceptuels

Servanne Monjour, Nicolas Sauret

► **To cite this version:**

Servanne Monjour, Nicolas Sauret. Archiver le web littéraire : défis méthodologiques et conceptuels. Colloque ResPaDon. Le web : source et archive, Laurence Favier; Marie-Madeleine Géroutet; Emmanuelle Bermès, Apr 2023, Lille, France. halshs-04059659

HAL Id: halshs-04059659

<https://shs.hal.science/halshs-04059659v1>

Submitted on 5 Apr 2023

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Archiver le web littéraire : Défis méthodologiques et conceptuels

Servanne Monjour et Nicolas Sauret

4 avril 2023

Notre communication va déplacer la réflexion du côté du web littéraire, ou de la littérature dite “nativement numérique”, dont les enjeux d’archivage rejoignent en partie ce dont il a été question lors de ce colloque, mais posent dans le même temps des défis un peu particuliers, en termes autant méthodologiques qu’épistémologiques.

Comme vous le constaterez, notre corpus est un peu différent de celui du projet LIFRANUM, et appelle une approche elle-même distincte, tant sur le plan méthodologique que conceptuel.

Précisons aussi que nous présentons des travaux en cours, des réflexions en cours, en particulier sur l’archivage.

Nous commencerons par jeter un pavé dans la marre avec une affirmation sans doute radicale, mais assez lucide : parler de littérature numérique aujourd’hui, c’est bien souvent parler des œuvres que l’on n’a pas lues, ou du moins pas directement lues. La proposition provocatrice et ironique que lançait Pierre Bayard en 2007 dans un essai qui prônait la “non-lecture”, est le lot quotidien des chercheurs qui ont choisi de s’intéresser à ces corpus fragiles et éphémères.

Que la raison de cette fragilité soit matérielle, institutionnelle ou même esthétique (parce que cette littérature s’inscrit dans un paradigme de la performance, elle est souvent conçue justement pour disparaître), le résultat est le même : la disparition de nos objets de travail est un problème récurrent. Sans l’avoir quantifié précisément, nous estimons perdre chaque année 40% de notre corpus d’étude.

Nous travaillons donc bien souvent sur un ensemble de traces: captures d’écran, enregistrements, données moissonnées automatiquement, stockées et indexées par nos propres soins. Bref, nous travaillons sans en avoir bien toujours conscience, sur une archive de la littérature contemporaine plus que sur la littérature contemporaine elle-même.

Or il apparaît essentiel de réfléchir à la constitution de cette archive dans la mesure où il s’agit ici de consolider notre pratique de recherche et poser les bases d’un travail scientifique de qualité. Et c’est sur ce point que notre démarche rencontre, et parfois se confronte à la démarche institutionnelle des archives du web.

Si l’on insiste sur ce point, c’est notamment en raison d’un problème de légitimité et de légitimation qui touche toute manifestation “numérique” de la littérature – adjectif dont on conservera ici volontairement le manque de précision, pour faire écho au discours médiatique parfois très violent, notamment porté par l’édition traditionnelle.

Mais surtout, ils existent hors du media livresque, c’est-à-dire hors du cadre traditionnel de la publication dont on connaît la fonction légitimante : l’objet livre endosse immédiatement

l'aura de son éditeur que ce soit chez un grand éditeur ou au contraire au sein d'une petite maison d'édition expérimentant des auteurs et des formes. Ces œuvres pénètrent presque automatiquement les portes des institutions patrimoniales (on pense au dépôt légal), mais aussi plus facilement les portes de l'institution académique. Le cercle de la légitimation est bouclé.

Au contraire, pour nos auteurs, le livre n'est plus qu'un moment du texte, au sein d'une pratique d'écriture qui s'est profondément diversifiée en investissant le web, souvent à dessein, dans un but d'exploration de formes inédites. En choisissant de s'écrire et de se rendre publics hors les livres, ces auteurs cherchent certes une forme d'émancipation et transforment complètement la présence sociale de l'écrivain dans notre société, mais échappent également à cette triple légitimation: celle de l'éditeur, de l'institution patrimoniale et de l'institution académique.

La matière que nous travaillons en tant que chercheur · e se façonne directement dans le Web, dans des formes qui échappent souvent à l'idée même de corpus. Nous sommes ainsi amenés à produire nous-mêmes les fonds sur lesquels travailler, mais avec bien évidemment beaucoup de maladresses puisqu'il s'agit de *scraper* le Web, puis d'organiser cette matière souvent mouvante, incomplète, en édition continue nécessitant alors un archivage continu. Cette collecte sauvage, menée à tâtons soulève de nombreuses questions pour lesquelles nous ne sommes pas forcément formés.

Ce sont des questions à la fois méthodologiques et techniques : comment je moissonne, comment je capte, comment je sélectionne, comment je conserve, comment j'indexe ?

Mais elles deviennent très rapidement des questions conceptuelles : qu'est ce que je moissonne, qu'est-ce que je capte, comment je le décris ?

Pour illustrer les difficultés auxquelles nous sommes soumis, et proposer quelques pistes destinées à surmonter celles-ci, nous allons présenter un cas d'étude sur lequel nous travaillons en ce moment : le projet *Musée Marilyn* de l'écrivaine Anne Savelli.

Cas d'étude : la fabrique de *Musée Marilyn* d'Anne Savelli

Musée Marilyn est le 13^e ouvrage d'Anne Savelli, il a été publié en 2022 aux éditions Inculte [[@savelli_musee_2022](#)].

Comme l'indique son titre, le roman se conçoit comme un véritable musée – il s'agit d'une exposition fictive, consacrée à l'actrice Marilyn Monroe, ou plus précisément aux photographies de Marilyn Monroe.

Élément notable: aucune illustration photographique ne figure dans *Musée Marilyn*, qui propose un fin travail d'ekphrasis ou de description de l'image photographique.

Il s'agit autrement dit d'un récit sur la photo, mais sans photo, et vous verrez que ce détail a toute son importance.

Récit-fleuve de 500 pages, *Musée Marilyn* est le fruit de près de huit années de travail de recherche et d'écriture, un vaste chantier qu'Anne Savelli a pris soin de documenter et de publier en ligne, sous de multiples formes: un premier feuilleton, intitulé « Volte-Face », est publié dès 2015 dans la revue littéraire en ligne Remue.net...

un travail d'écriture collectif et les premières pages du livre sont par ailleurs diffusées sur son site personnel (Fenêtre Open Space), qui accueille également d'autres textes de l'autrice.

Surtout, une iconothèque (une collection d'images), qui servira de matière première à l'élaboration du récit, est peu à peu constituée et diffusée, selon une démarche elle-même à caractère archivistique (car il est important de préciser que la littérature numérique est étroitement préoccupée par la question de l'archive, aussi bien pour des raisons poétiques que mémorielles). Pendant toutes ces années, Anne Savelli a amassé une impressionnante collection d'images numériques et numérisées, partagée avec ses lecteurs sur un canal essentiellement numérique.

Cette iconothèque procède surtout d'un investissement des plateformes sociales, d'abord Facebook, puis Twitter et enfin Instagram, que l'autrice utilise tant pour exposer sa collection. . . .

. . . que pour l'enrichir, à l'aide des contributions de ses lecteurs-followers.

Une littérature éditorialisée

À bien des égards, le projet Marilyn d'Anne Savelli (on justifiera dans un instant ce terme "projet") apparaît comme un cas d'école d'une littérature éditorialisée : elle se définit comme un vaste chantier processuel, collectif, disséminé dans l'espace numérique (plateformes littéraires, réseaux sociaux) et non numérique, et potentiellement infini, puisqu'Anne Savelli publie encore, aujourd'hui, sur les réseaux sociaux et sur son site, des contenus à propos de Marilyn.

À ces caractéristiques formelles s'ajoute un brouillage de la frontière entre l'écriture fictionnelle et l'écriture du réel, signe d'un nouveau réalisme littéraire que l'on voit émerger depuis quelque temps et qui ne manque pas de faire débat dans notre champ disciplinaire – mais c'est un autre sujet.

Plurimédiatique, disséminé sur plusieurs années et sur plusieurs plateformes, le travail d'Anne Savelli a quelque chose d'insaisissable. Et pourtant, nous avons voulu relever le défi d'en constituer un corpus exhaustif.

Quelle stratégie d'archivage ?

Mais comment construit-on l'archive d'un tel projet relevant d'une logique d'éditorialisation ? Le travail que nous allons désormais présenter semble très simple – il associe une consultation des archives du Web à un moissonnage et archivage du Web vivant, à des fins de constitution et d'analyse d'un corpus – mais nous allons constater combien, du fait de son éditorialisation, le projet d'Anne Savelli engage, à chaque étape de son moissonnage, des questions techniques [et ?] conceptuelles particulièrement épineuses, qui encouragent à reformuler les grands concepts disciplinaires des études littéraires.

1. Consulter les archives de l'Internet et du Web

Du paradigme de l'œuvre à l'approche par projet, de l'archive au corpus situé La première étape de notre travail s'est déroulée dans les archives de l'Internet et du Web de la BNF. La consultation de ces archives nous a conduits à déconstruire un premier concept majeur de la littérature : l'œuvre, issue du paradigme de l'imprimé, auquel nous proposons de substituer la notion d'*approche par projet*, emprunté au vocabulaire de l'art, notamment de l'art contemporain.

De l'œuvre à l'approche par projet La notion de projet met en avant une idée, un concept à atteindre : elle dénote un certain degré d'inachèvement (voire d'inachevable), souligne le travail préparatoire et les moyens nécessaires à l'élaboration d'un produit "final" (même si l'enjeu de cette finalité, aujourd'hui, est discutable), et bien souvent ouvert à la contribution, à la discussion. Elle déplace l'intérêt sur le processus créatif plutôt que sur un artefact final. Désormais, c'est bien le projet qui fait œuvre, et transforme au passage le sens accordé à cette dernière – en particulier en bousculant ses enjeux de stabilité et de clôture : la publication numérique a permis aux écrivains de développer une poétique du chantier, où le lecteur assiste à un nouveau type de performance, laquelle fait partie de l'expérience esthétique.

Non seulement l'instabilité et le caractère inachevé du texte ne sont plus tabous, mais ils font même tout l'intérêt du texte littéraire en ligne : tâtonnements, hésitations, brouillons, ratés... la "poétique" littéraire en contexte de publication numérique revient à son sens premier, un art du *faire*, lequel fait l'objet d'une publication en temps quasi réel. L'approche par projet, en d'autres termes, est la conséquence directe de la littérature éditorialisée. C'est exactement ce qui est en jeu dans ce que l'on appellera donc désormais le projet Marilyn d'Anne Savelli, qui a largement exposé son travail de documentation autour de Marilyn, appelant même ses lecteurs à la construction d'une iconothèque collective, tout en construisant peu à peu récits et métarécits autour de son idée première.

Voyons de plus près comment ce glissement de l'œuvre vers le projet a infléchi dans notre démarche d'archivage, et finalement modifié la dénomination même de notre corpus.

Le projet Marilyn dans les archives institutionnelles La BNF s'est essentiellement concentrée sur la conservation des sites Internet – ce qui constitue déjà un travail en soi titanesque. Le projet Marilyn apparaît donc bien dans ses collections, mais principalement au sein des archives de Remue.net, où se trouve le premier feuillet « Volte-face » dans sa version de 2020 (soit déjà achevée), ou encore dans celles du site personnel d'Anne Savelli, qui a fait l'objet d'un moissonnage depuis 2008. Le projet Marilyn en particulier n'est cependant moissonné qu'à partir de 2018 (une fois), puis en 2021 (à deux reprises) ainsi qu'en 2022 et 2023.

La consultation des archives de l'Internet rend assez bien compte du phénomène d'éditorialisation du site, dont le caractère processuel est souligné par la plateforme permettant une navigation à travers le temps. En revanche, il est parfois compliqué de trouver certains contenus, notamment en raison du renommage de certains onglets par Anne Savelli – et des changements d'url qui en découlent.

Ces archives sont évidemment précieuses, mais elles ne rendent compte que d'une infime partie du projet – le volet réseaux sociaux étant difficile à cerner depuis la consultation des archives. Certains réseaux sociaux sont certes inclus dans le périmètre de la BNF, mais sur Twitter par exemple, les images du projet marilyn ne sont pas archivées, ce qui transforme le profil en un gruyère très, très troué, dans la mesure où Anne Savelli dispose d'un fil twitter à la fois textuel et visuel, en particulier pour ce projet.

Les archives de l'Internet et du web : la permanence d'un paradigme de l'imprimé ?

Ce que cette consultation révèle, c'est la permanence d'un paradigme (inconscient ?) de la littérature imprimée qui guide la constitution ainsi que la médiation des archives de l'Internet et du Web.

L'archive littéraire de la BNF est ainsi représentative d'une activité d'écriture propre à un auteur, en *son* site (lequel a remplacé l'œuvre, dans une sorte de continuité établie avec l'objet livre). Mais elle ne tient pas suffisamment compte, à notre sens, d'une activité en vérité beaucoup plus éclatée, collective et discursive, qui nécessiterait des archivages plus ciblés et granulaires. De fait, par ses choix d'archivages, la BNF cible une partie bien précise de la production littéraire numérique (elle va en particulier très bien représenter la littérature blogue, l'écriture hypertextuelle et des genres comme l'autoblographie), mais elle passe un peu à côté d'autres pratiques d'écriture (par exemple, la twittérature, qui dans notre cas d'étude est essentielle). De ce point de vue là, c'est tout un pan de la littérature numérique qui en est exclu.

Par ailleurs, si l'interface de consultation des archives du Web sait rendre compte du phénomène d'éditorialisation dans le temps (données temporalisées), elle contraint quelque peu le potentiel d'analyse des chercheurs. En ne collectant que les pages représentées, telles qu'elles s'affichent, l'archivage privilégie la dimension médiatique de la littérature numérique (ce qui convient parfaitement à une approche de l'ordre de la sémiotique du support ou encore des études stylistiques traditionnelles), mais elle ne permet pas un accès aux différentes couches de l'écriture numérique – notamment celle du code, chère aux *critical code studies*, absolument essentielle à l'analyse “seconde génération” de littérature électronique – ou encore à la mise en donnée du texte (à des fins d'analyse de corpus textuel assisté par ordinateur, par exemple), mais également à des approches de l'ordre des platform studies. Seule l'archive du chercheur permet alors de déployer toute la panoplie d'approches disciplinaires propres à l'étude des objets numériques.

vers un corpus *situé* Ce que ce passage par les archives institutionnelles de l'Internet et du Web nous ont finalement enseigné, c'est à reformuler la dénomination même du matériau dont nous avons besoin:

- ni une archive numérique (ce que produit l'institution à des fins de légitimation et de conservation),
- ni un corpus traditionnel (conçu comme un ensemble d'objets littéraires issu de cette archive et lui-même stabilisé),
- mais dans ce que nous proposons d'appeler un *corpus situé*, c'est-à-dire intégrant à la fois un jeu de données (*dataset*), les outils qui ont constitué ce jeu de données, ainsi que la documentation de ces outils et la critique de ce *dataset*. De fait, le corpus situé comprend intrinsèquement la question de recherche qui anime sa constitution, et permet surtout de la rejouer en rendant ses outils manipulables et exécutables. En ce qui nous concerne, les notebooks Jupyter constituent une infrastructure idéale pour construire ces corpus situés, comme nous allons à présent le démontrer.

La question du périmètre d'une œuvre numérique – et par extension de son archive – devient d'autant plus sensible qu'elle va presque nécessairement, à un moment ou à un autre, aller dans le sens d'une approche plutôt qu'une autre, et donc favoriser une définition de la littérature numérique plutôt qu'une autre. Dans l'idéal, une archive de la littérature numérique devrait donc être la plus inclusive possible – par “inclusion”, nous entendons à la fois le type d'objet (le blogue, le réseau social, le site web, le texte, l'image, la vidéo, le podcast), mais également ses différentes couches techniques (son interface, son code, ses métadonnées, son réseau : likes, retweets), de manière à demeurer la moins prescriptive possible.

2. Moissonner le Web vivant

Du paradigme de l'auteur à celui du collectif. Du corpus situé à sa généricisation

Assez rapidement, nous nous sommes tournés vers le Web vivant, et plus particulièrement sur les plateformes sociales qui jouent un rôle majeur dans l'élaboration du projet Marilyn, notamment en ce qui concerne sa dimension iconographique : Instagram, Facebook et Twitter. Pour le moment, nous nous sommes concentré sur Twitter, dont l'évolution récente des conditions d'accès donne présentement des sueurs froides aux chercheurs.

Sans même parler de l'accès aux données, les seules fermetures de comptes consécutives aux décisions d'elon musk viennent confirmer l'évanescence de nos corpus de recherche.

Comment appréhender un corpus de *tweets*, *retweets*, de *threads* et de réponses ? Quelle conception de l'autorité se dessine à travers cette écriture profilairé résolument tournée vers la conversation, l'échange, la mise en réseau ? Comment calculer (et qualifier) le degré d'engagement des différents contributeur au projet ?

Sur un plan méthodologique, nous avons opté pour l'utilisation d'un notebook Jupyter qui nous permettait une première approche exploratoire et collaborative, avant de tendre vers une méthode généricisable, c'est-à-dire applicable à d'autres corpus. Les carnets Jupyter combinent des éléments de code exécutables et des éléments d'écriture discursive. Ils articulent ainsi les scripts de collecte et de traitement des données, le résultat de ces scripts, et leur documentation et analyse, ouvrant un espace d'écriture susceptible de produire une herméneutique des données décrite par Stefan Sinclair et Geoffrey Rockwell dans leur ouvrage *Hermeneutica*.

En phase avec les pratiques du champ des humanités numériques, notre corpus et le carnet Jupyter créé pour le collecter et l'analyser sont rendus accessibles sur une forge logicielle, créant les conditions nécessaires à la réappropriation de la méthode.

Dans cette première approche, le notebook a également été idéal pour articuler une collecte et une analyse progressive, de manière justement à cerner et circonscrire pas à pas ce qu'on appelait "le projet marilyn". Cette articulation fine Collecte/Analyse nous semble parfaitement définir la notion de corpus situé.

Plusieurs bibliothèques logicielles sont disponibles pour explorer et collecter des données sur les réseaux sociaux. Pour cette collecte, nous avons choisi la bibliothèque *minet* mise à disposition par le MediaLab de Sciences Po [[@pelle_minet_2019](#)]. Cette bibliothèque hérite d'une part de l'approche résolument sciences humaines du MediaLab, et de son expérience en matière de collecte de données sur le Web vivant. Plus particulièrement, *minet* a le mérite de proposer une méthode de collecte par scrapping, indépendante de l'API de Twitter et de sa politique d'accès pour le moins instable.

Nous avons articulé cet outil avec d'autres outils, comme *pixplot* pour l'exploration et la catégorisation des images. Effectivement, dans ce projet photo-littéraire, les contributions en images et en photographies constituent évidemment un sous-corpus particuliers que l'on analyse en tant que tel.

Ici c'est le résultat de *twitter explorer* (qui propose une analyse de réseau des hashtags ou des profils). Mais tout cela relèvera de l'analyse littéraire, or l'enjeu aujourd'hui est de rendre compte de notre démarche d'archivage et de constitution de corpus, et de la confronter aux pratiques institutionnelles discutées dans ce colloque.

Le long du "fil d'Ariane" que constitue la page du profil Twitter d'@athanorster (Anne Savelli), ou le long des différents fils sémantiques #marilyn ou #AlerteMarilyn, se déploie une écriture plurielle et plurimédiatique, faite de contributions diverses, destinées le

plus souvent au projet, mais parfois simplement agrégées et accueillies au sein du projet au prétexte d'une collision de mots-clés. Cette écriture en mouvement révèle avant tout le caractère collectif du projet Marilyn, orchestré bien entendu par Anne Savelli, malgré tout joué et interprété par une communauté plus ou moins investie dans le temps, généralement bienveillante et joyeuse (on devine que certaines amitiés existent par delà le réseau), tissant des liens, enrichissant le fil de multiples singularités, de points de vue et de captations d'un réel bien plus large que celui d'Anne Savelli.

Cette pluralité de l'écriture va dans le sens d'un déplacement ou d'une évolution du concept d'auctorialité dans l'espace numérique. Les plateformes sociales sont effectivement devenues un terrain d'expérimentation littéraire propice à ce déplacement de l'auctorialité, désormais incarné dans un profil – lui-même pris au cœur d'une dynamique d'éditorialisation.

Corpus

Sur Twitter, nous avons procédé à une collecte par tranche (ou carottage), mais plutôt que de fonctionner selon une logique en entonnoir (avec une collecte d'abord très massive dans laquelle nous aurions ensuite effectué un tri), nous avons procédé par petites touches, comme une fouille archéologique qui dévoile progressivement son terrain au fur et à mesure que la poussière est retirée.

Nous avons joué sur 3 facettes de recherche : hashtags, profils, période temporelle, la démarche a consisté à sonder, à catégoriser, puis à sélectionner les paramètres de requêtes constitutifs du projet littéraire. C'est une démarche qui nous a obligé à constamment alterner entre une approche de lecture distante et de lecture rapprochée. C'est ainsi que chaque hypothèse a pu être traduite en requête puis agrégée ou non au corpus complet, toujours de manière différenciée afin de pouvoir identifier les raisons d'un ajout au corpus.

La succession des couches du corpus révèle une stratigraphie qui ne témoigne pas d'un temps qui passe, mais plutôt d'un engagement plus ou moins direct et actif dans le projet.

Ainsi, ce travail de qualification et de collecte "par petite touche" dessine progressivement deux périmètres : celui du corpus (constitué de micro-posts et des images associées), et celui du collectif. En effet, en définissant les contours du corpus, nous avons été obligé de définir ce qu'on entendait par collectif littéraire : Puisque le projet est fondamentalement collectif, qui en sont les membres ? Quelle forme prend ce collectif ? Doit-on établir des degrés de contributions, c'est-à-dire une typologie de l'engagement ?

Pour le dire autrement, les premiers coups de sonde nous fait comprendre que le "projet Marilyn" va bien au delà de "ce qu'a lu, traité ou répondu @athanorster (alias anne savello)". Au contraire, son projet lui échappe, à travers des conversations parallèles et des initiatives indépendantes, reliées entre elles par les hashtags.

Collectif

Nous avons schématisé ce collectif ainsi, en cercles concentriques, témoignant d'une échelle d'engagement : avec les contributeurs directs ou actifs, les contributeurs indirects, c'est-à-dire simplement agrégés par le biais d'un hashtag, d'un retweet ou d'une réponse de contributeur actif, les profils répondants/retweetant, les profils mentionnés enfin.

Mais cette définition du collectif suppose aussi de catégoriser ces différents profils selon une typologie de fonction qu'il soit écrivain, institutionnel, éditeur, chercheur, critique, médiatique, de manière à identifier la stratégie sociale de chaque mention par exemple.

C'est un travail en cours, on ne rentre pas dans le détail, mais ce qui est intéressant de noter, c'est que cette approche de l'archivage nous a en quelque sorte amenés à penser la collecte à l'envers.

On l'a vu ces derniers jours, l'archivage institutionnel cherche à s'emparer sur le web de phénomènes sociétaux, imposant d'emblée un caractère massif aux collectes entreprises. Cette masse empêche d'une part l'exhaustivité et d'autre part une approche fine de la collecte. La stratégie adoptée est alors de constituer une archive "représentative", aussi large que possible, sans préjuger des analyses et des recherches potentielles.

Mais comment être représentatif de la littérature numérique ? Le projet LIFRANUM présenté lundi s'est posé la même question, mais a finalement opté pour une autre stratégie, en se focalisant surtout sur les sites personnels d'auteurs, pour des raisons techniques notamment, délaissant alors toute la littérature de plateformes qui nous intéresse. En ciblant des manifestations littéraires de l'ordre de ce qu'on a appelé "le projet littéraire", nous nous confrontons à un objet qui demande précision et peut-être complétion ?

Par exemple, nos collectes sur twitter approche tout juste les 1000 fragments, loin des données massives permettant de convaincre l'intervention d'une institution patrimoniale, malgré la pertinence du corpus pour témoigner du fait littéraire numérique.

On voit bien qu'avec une approche comme la nôtre, l'enjeu n'est plus de chercher la représentativité d'un phénomène, mais plutôt de mettre à jour ses multiples manifestations et d'identifier ainsi les frontières de son périmètre complet.

Nous découvrons ce territoire par petites avancées. Pour reprendre la métaphore de la fouille archéologique, il ne s'agit pas creuser à la pelleuse (de manière représentative), mais plutôt d'épousseter centimètre par centimètre notre terrain et de repousser ainsi le périmètre de notre archive pour tendre vers l'exhaustivité du projet littéraire.

Pour autant, même si ce terrain est limité, aux dimensions maîtrisables disons, son éparpillement dans le flux de Twitter ou d'Instagram impose une approche algorithmique de collecte, et donc la mise en place d'une méthode dédiée.

Les corpus situés s'inscrivent ainsi dans un entre-deux : trop massifs, fragmentés et disséminés pour être facilement appréhendable sans un outillage numérique d'automatisation ; mais dans le même temps bien trop minces pour espérer jouer dans la même cour que les grands corpus patrimoniaux (ces bibliothèques dans lesquelles les chercheurs vont réaliser un travail de *distant reading*).

Nous discutons entre nous d'un terme alternatif pour caractériser notre corpus. Nous nous demandons s'il serait opportun de distinguer le corpus situé, qui embarque la question de recherche, et ce qu'on a tenté d'appeler "une archive raisonnée". Raisonnée, car elle embarquerait la cartographie documentée de son périmètre et conserverait une série de potentialités d'analyse – et tout autant de questions de recherche possibles.

Archive raisonnée et · ou corpus situé, notre approche met le doigt dans tous les cas sur la nécessité de construire de nouvelles synergies entre les chercheurs, les institutions patrimoniales, mais aussi les terrains eux-mêmes, c'est-à-dire dans notre cas les auteurs et contributeurs de projets littéraires, qui semblent avoir développé un "tropisme" archivistique: une tendance à collectionner, à amasser des images, des extraits de textes. Ces collections (bibliothèques, iconothèques) existent depuis longtemps chez les écrivains,

mais qui font l'objet, grâce au web, d'une activité de publication inédite, sur laquelle il y a sans doute beaucoup à dire.

C'est d'ailleurs Anne Savelli elle-même qui aura le dernier mot sur ce projet. Alors que nous l'avions sollicitée pour un entretien l'an dernier, à un moment où *Musée Marilyn* était encore à l'état d'épreuve, Anne Savelli nous a rencontrés en apportant avec elle un drôle d'objet : un classeur cartonné à levier qui renferme, dans son jus, la collection de coupures de presse, de photographies et de cartes postales de Marilyn Monroe, une collection de plusieurs centaines d'images, qu'Anne Savelli a composé il y a maintenant 40 ans lorsqu'elle était adolescente. . .

L'existence de cette iconothèque pré-numérique nous rappelle que nous sommes déjà dans une ère post-internet, dans laquelle les objets numériques ne sont qu'une des nombreuses potentialités de publication (au sens premier du terme, rendre public) offertes aux écrivains. De la même manière, l'objet imprimé incarne un état du texte, mais certainement pas sa fin.

Archive du fait numérique = archive également "analogique", comprenant des objets non-nativement-numériques, ou non-numériques. . . Où l'on peut se plaire à rêver d'un corpus situé qui réintégrerait des archives plus "traditionnelles", ou encore des objets plus inattendus embrassant la diversité des incarnations médiatiques de la littérature contemporaines : des livres, des publications web, des podcasts, des morceaux de murs ou des tricots (exemples réels !).