HAL
open science

Du fragment daté au corpus patrimonialisé

Jessica de Bideran

To cite this version:

Jessica de Bideran. Du fragment daté au corpus patrimonialisé. Etudes digitales, 2016, Le texte à venir, 1, pp.125-142. hal-01855780

HAL Id: hal-01855780

https://hal.science/hal-01855780v1

Submitted on 8 Aug 2018

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.

2016 : « Du fragment daté au corpus patrimonialisé : numérisation et muséalisation de l'article de presse mauriacien », *Études digitales (1)*, p. 125-142.

Du fragment daté au corpus patrimonialisé : numérisation et muséalisation de l'article de presse mauriacien

Jessica de Bideran

Introduction

Portés par des équipes de chercheurs en Lettre, les projets d'éditions numériques de corpus de textes littéraires qui se sont multipliés ces dernières années – édition des manuscrits de Gustave Flaubert[1] ou de Stendhal[2], des archives manuscrites du fonds Jean Paulhan[3], etc. – signent progressivement l'entrée de ces sciences dites « molles » dans l'ère des Humanités digitales (Le Deuff, 2014).

Cherchant à dépasser les limites des éditions papiers, ces expériences s'appuient sur la structure hypertextuelle du Web pour retranscrire, via l'espace infini de l'écran, les réseaux littéraires, les influences et la variété de ces œuvres. Les scientifiques impliqués dans ce processus tentent peu à peu de redéfinir les rapports aux savoirs en œuvrant pour le partage des données issues de la recherche et la valorisation des connaissances produites. Parallèlement, ce nouvel environnement oblige à repenser les modes de financement de ces projets[4], à recomposer les équipes de recherche en multipliant les profils professionnels ainsi qu'à s'insérer dans un paysage national visant à coordonner une production raisonnée de corpus numériques[5]. Nécessairement transdisciplinaires, ces projets supposent pour les membres engagés dans leur mise en œuvre l'apprentissage de savoir-faire, l'acquisition d'une *lingua franca* permettant d'échanger. En ce sens, ces différents travaux de numérisation et de diffusion de corpus littéraires peuvent être considérés comme des espaces de médiation où les équipes engagées dans ces réalisations transforment les objets textuels sur lesquels elles travaillent. Portés par des chercheurs en littérature rompus à l'exercice de décomposition du texte, cette recomposition numérique ne détermine-t-elle pas dès lors une nouvelle dimension symbolique à ces textes ?

À partir d'une description précise du programme de numérisation « Mauriac en ligne – textes

---

[1] Projet porté depuis 2003 par le Centre Flaubert, composante du laboratoire CÉRÉdI (Centre d'études et de recherche Éditer-Interpréter) de l'Université de Rouen. On pourra par exemple consulter le site présentant la transcription des manuscrits de *Madame Bovary*, http://www.bovary.fr/, et l'atelier Bovary, site explicitant la mise en œuvre du projet, http://flaubert.univ-rouen.fr/bovary/atelier/atelier.php (Pages consultées le 23/02/2015).

[2] Projet porté, entre autre, par l'équipe Traverses 19-21 (Centre d'études stendhaliennes et romantiques) de l'Université Stendhal-Grenoble 3, http://manuscrits-de-stendhal.org/ (Page consultée le 23/02/2015).

[3] À travers le projet HyperPaulhan, axe du programme de l'Observatoire de la vie littéraire (OBVIL) qui entend développer toutes les ressources offertes par les applications informatiques pour examiner aussi bien la littérature française du passé que la plus contemporaine, http://obvil.paris-sorbonne.fr/projets/hyperpaulhan (Page consultée le 23/02/2015).

[4] Le projet dont il est question ici fait l'objet d'un financement local provenant de la Région Aquitaine ; cette source financière a, selon nous et comme nous le verrons plus loin, des conséquences sur le statut des textes numérisés. Programme de recherche porté par l'Université Bordeaux Montaigne, celui-ci est en effet mené en partenariat avec le Centre François Mauriac de Malagar grâce à un soutien de la Banque numérique du savoir en Aquitaine

[5] Le programme « Mauriac en ligne » est ainsi membre depuis le 02 novembre 2015 du consortium CAHIER (Corpus d'auteurs pour les humanités : informatisation, édition recherche), consortium de la TGIR (très grande infrastructure de recherche) pour les Humanités numériques portée par le CNRS, Huma-Num. L'objectif principal de ce groupe de travail est de permettre la rencontre de chercheurs, ingénieurs et conservateurs autour de principes communs (libre-accès et réutilisation) ainsi que la pratique de technologies partagées sur les bases de données, les entrepôts de données, etc. Voir : http://cahier.hypotheses.org/ (Page consultée le 20/11/2015)

de presse », nous souhaitons apporter quelques éléments de réponse à cette question en nous appuyant plus spécifiquement sur l'analyse détaillée de la plateforme de consultation actuellement en cours d'expérimentation. Parce qu'il relève à la fois d'une organisation matérielle et d'une proposition de lecture, l'outil exploité illustre en effet des choix techniques mais aussi heuristiques effectués par l'équipe de recherche. La plateforme éditoriale et la méthodologie mise en œuvre amène ainsi à s'interroger sur le statut de ce corpus numériques en reconsidérant la mémoire de l'œuvre journalistique mauriacienne, par-delà les clés de lectures que l'auteur a lui-même souhaité transmettre de son vivant en la figeant dans des rééditions papiers.

Passant du régime papier au régime numérique, ces textes nous semblent par là même passer du régime allographique de la littérature (Genette, 1994) à celui du régime autographique des objets patrimoniaux (Davallon, 2015) ; conséquemment, le dispositif d'écriture et de lecture expérimenté, que l'on pourrait nommer à la suite de Franck Cormerais *mémothèque* (Cormerais, 2015 : 134), amène le chercheur à s'interroger sur la mémoire scientifique que lui-même construit. Cette réflexion s'appuie sur notre participation à ce programme de recherche dans le cadre d'un contrat post-doctoral. Cette implication, au cœur de l'espace de médiation qui se met progressivement en place autour du dispositif, nous donne ainsi l'occasion d'observer finement la mise en œuvre de ce nouveau statut du texte numérique mauriacien.

Offrir une troisième vie aux textes de presse de François Mauriac

En s'appuyant sur le travail de recension mené par ses membres, le Centre Mauriac de l'Université Bordeaux Montaigne porte depuis plusieurs années le projet d'une édition numérique intégrale des textes de presse de François Mauriac. Ce projet souhaite renouveler l'approche du travail journalistique de l'écrivain en numérisant la totalité de ses articles, soit environ 3000 textes publiés entre 1905 et 1970 au sein de 150 publications périodiques. Malgré l'important travail de réédition réalisé durant les dernières décennies du XXème siècle, 560 textes, c'est-à-dire environ 20% du corpus, n'ont en effet jamais été repris en volume et restent difficilement accessibles aux lecteurs. Sur ces 560 textes, 540 sont antérieurs à 1954, date à partir de laquelle les reprises ont été plus systématiques, soit des mains même de Mauriac au sein de ses *Bloc-Notes* et de ses *Mémoires*, soit des mains de chercheurs tel que Jean Touzot.

En donnant accès à l'ensemble de ces textes de presse, conservés physiquement par différentes institutions telles que la BnF ou la bibliothèque universitaire de Bordeaux Montaigne, il s'agit donc de reconstituer un ensemble virtuel[6], complet et homogène, afin de permettre la construction de nouvelles représentations du travail journalistique de François Mauriac. L'ancrage régional du projet, au cœur d'un paysage mauriacien marqué par le Centre François Mauriac de Malagar, ancienne demeure de l'écrivain, influence également les visées communicationnelles puisqu'il s'agit aussi d'offrir à des publics variés – scolaires pour exploitation pédagogique, enseignants en lettre ou en histoire mais aussi publics amateurs – une source incomparable d'informations sur l'histoire et la société françaises du XXème siècle. Le terme de reconstitution n'est ici pas anecdotique ; traditionnellement associée à la représentation d'un événement historique, cette procédure symbolique exploite divers moyens permettant de se remémorer une histoire selon une temporalité complexe car rappelant le passé mais efficace dans un perpétuel présent (Jollet & Massu, 2012, 9).

À la fois diachronie et synchronie, la reconstitution désigne dans le même temps les procédures d'exposition dont bénéficient certaines collections muséographiques, à l'image des dioramas mêlant vestiges et interprétations que l'on retrouve dans les musées archéologiques (Flon, 2012). Cette pratique d'exposition a été plus particulièrement théorisée par Georges-Henri Rivière

[6] Au sens d'actualisable par son affichage en ligne et possédant avec force les qualités – et notamment ici iconiques – des éléments du corpus réel qui, pour sa part n'existe pas de manière tangible.

lors de ces travaux menés dans les années 1960 sur les écomusées. Son acolyte André Desvallées relate ainsi une collecte menée en Aubrac lors du prélèvement d'habitats traditionnels. Plâtre et résine furent en effet largement convoqués pour, littéralement, recréer certaines parties de ces structures déplacées de leur environnement premier pour s'insérer au cœur du futur musée des Arts et Traditions Populaires. Insistant sur le terme « recréation », celui-ci insiste : « *J'utilise ce terme artistique un peu par provocation. Pour être précis […] l'intérieur des Fajoux sont des restitutions : ils ont fait l'objet d'un simple transfert avec démontage et remontage à l'identique. D'autres ensembles […] sont des reconstitutions ; à ces derniers ensembles, pas complets, ont été intégrés des outils et du matériel ancien collecté ailleurs et permettant une présentation synchronique* » (Desvallées, 1989, 187). Ainsi, et selon l'expertise muséologique, la restitution consiste à remonter à l'identique un ensemble patrimonial déplacé de son contexte d'origine, tandis que la reconstitution vise à combler les lacunes du passé par l'ajout d'éléments extérieurs mais cohérents. Elle suppose par conséquent un travail d'interprétation sérieux permettant d'expliciter les traces conservées. Or, ce mouvement, qui suppose d'exhumer de l'oubli des vestiges historiques mais aussi de les enrichir en les recontextualisant, est finalement relativement proche du travail actuellement mené par l'équipe du projet « Mauriac en ligne ».

La phase de récolement, tout aussi essentielle que laborieuse, s'est ainsi concrétisée par la localisation des différents fonds conservant les textes étudiés en vue de leur numérisation[7]. C'est au cours de cette phase de repérage que le chercheur est amené à s'enfoncer dans le passé (Baudorre, 2011). Le texte de presse est en effet intrinsèquement daté. Écrit pour être lu dans un moment donné et selon un cadre particulier, il est condamné à appartenir au passé presque aussitôt les pages du périodique refermées. Plus d'un siècle après la parution des premiers articles de Mauriac, leur contexte d'écriture et de lecture tout comme leurs lecteurs ont profondément changé. Si, en les réunissant en recueils, l'auteur lui-même a offert une seconde vie à certains de ces textes soigneusement choisis, voire pour quelques-uns retravaillés et légèrement modifiés, numériser les articles de François Mauriac, c'est donc offrir à ces textes une troisième vie. Le choix qui a été fait est en effet celui de revenir à l'état premier de l'article reproduit en mode image.

Alors que la première phase d'expérimentation menée entre 2008 et 2010 sous la direction de Paul Cooke et portant plus spécifiquement sur la production journalistique des années 1937-1938 fit le choix de la seule transcription TEI[8] (Cooke, 2011), la procédure actuellement retenue pour reproduire ces textes est la numérisation des originaux en mode image puis l'extraction du flux textuel à partir d'un système de reconnaissance optique de caractère. Ce choix, qui peut sembler aller à l'encontre des préconisations de structuration des données textuelles délivrées par exemple par le guide des bonnes pratiques Huma-Num, s'est imposé par la volonté d'avancer rapidement dans la numérisation et la transcription des textes malgré nos faibles moyens en ressources humaines[9]. À l'aune des réflexions œcuméniques entourant les Humanités digitales, il convient de s'attarder quelque peu sur cette orientation qui diverge par exemple de celle suivie par de nombreux projets du Consortium CAHIER. Si l'abandon de la transcription TEI résulte effectivement du faible nombre de

---

[7] Deux fonds sont essentiellement concernés : le service commun de documentation de la bibliothèque universitaire de Bordeaux Montaigne conserve d'une part un certain nombre d'exemplaires des périodiques qui nous intéressent ici, soit sous formes d'imprimés, soit sous forme de microfilms. La BnF d'autre part possède une grande majorité de ces revues dont certaines sont par ailleurs déjà accessibles via Gallica. Enfin des fonds comme ceux de la bibliothèque municipale de Bordeaux ou encore de la bibliothèque spécialisée Jacques Doucet permettent de compléter les lacunes éventuelles.

[8] Pour Text Encoding Initiative. Voir en ligne : http://mauriac.ex.ac.uk/

[9] Si l'abandon de la TEI peut quelque peu isoler dans un premier temps ce corpus de l'ensemble des corpus littéraires qui s'établissent actuellement, ce qui est paraît dommageable dans un objectif scientifique d'échange, de cumulation et de reprise des données littéraires ainsi élaborées, soulignons que ce choix technique contribue à redéfinir le statut de ce projet qui sera amené à accueillir d'autres types de données, telles que des photographie ou des films de la famille Mauriac. Voir par exemple sur la TEI : Hudrisier, Henri et Azémard, Ghislaine, 2014 : « La TEI : un collège mondial et un outil commun pour la recherche en littérature », *Les Cahiers de la société française des Sciences de l'information et de la communication*, n°10, 107-126.

personnes ressources susceptibles de s'atteler à cette tâche, il prive *a priori* les futurs chercheurs de données parfaitement balisées et structurées permettant des interrogations et fouilles plus fines du corpus. Toutefois, des solutions techniques existent aujourd'hui pour transformer un document traitement de texte tel que nous l'exploitons lors des phases de transcription vers un format structuré strictement sémantique XML/TEI[10]. Réversible donc, cette orientation nous semble cependant révélatrice du statut accordé à ce corpus numérique. En se concentrant sur la mise à disposition des textes, plus que sur leur structuration, liaison et malléabilité futures permettant par exemple des interrogations statistiques fines, les chercheurs contribuent à façonner un lieu de mémoire qui élève le texte en objet de patrimoine avant d'en faire un objet de recherches.

Pour en revenir à la procédure exploitée, les prises de vues, réalisées au format TIFF en vue de leur conservation sont ensuite recomposées en un fichier PDF qui subit un traitement de reconnaissance de caractères[11] afin d'en extraire le texte. L'usager du site a donc accès à un fac-similé de l'article publié dans la presse à la fois iconique, via le PDF, et textuel, via la transcription. Pour autant, il ne s'agit pas d'un simple transfert à l'identique du format papier au format numérique. Le texte supporte en effet une sorte de démontage et de remontage puisque, conforment au projet éditorial, seul l'article de François Mauriac se voit valorisé. Billet publié en Une du *Temps présent* ou chronique de plusieurs pages éditée au sein de la *Revue hebdomadaire,* un repérage quasi chirurgical s'opère pour n'exposer que l'œuvre journalistique mauriacienne. L'intégrité de la page étant toutefois conservée, ces textes sont donnés à lire tels qu'ils sont apparus à leurs premiers lecteurs, dégagés, pour ceux qui furent repris en volume, des strates successives apportées par les éditeurs. Ce choix de rendre disponible le corpus en mode image est à la fois foncièrement lié à la plateforme de consultation retenue tout comme au désir de donner à lire le texte dans sa mise en page primaire. Celle-ci représente en effet une source de renseignements particulièrement pertinente pour la presse qui use de titres, photographies ou dessins enrichissant la lecture. En même temps, cet accent mis sur l'aspect visuel du texte nous semble être le signe de l'empilement des valeurs attribuées à ce corpus littéraire. La mise en forme de ces textes numériques et les possibilités d'appropriation construites illustrent parfaitement cette complexité.

Enchâsser les textes de presse de François Mauriac

Une fois les phases de récolement et de numérisation réalisées[12], l'objectif de mettre à la disposition de différents publics la totalité de ces articles, soit entre 6000 et 7000 pages, suppose d'établir la plateforme qui accueillera ces données numériques. Parallèlement au chantier opérationnel qui consiste en définitive à identifier et à rassembler une collection virtuelle de textes qui n'aura une existence que sous format numérique, la recherche est effectivement orientée vers la conception d'un dispositif d'édition répondant à des pratiques spécifiques d'écriture et de lecture. La réflexion menée s'est appuyée sur un certain nombre de principes logistiques et scientifiques.

---

[10] Citons par exemple Odette, un logiciel libre développé par Frédéric Glorieux qui exploite toutes les fonctionnalités avancées d'un logiciel de bureautique (tables des matières, styles, index, etc.) pour les retranscrire en balises normalisées. Voir : Frédéric Glorieux, 2015 : « Le traitement de textes (odt) pour produire des documents structurés (XML/TEI) – Odette », in *J'attends des résultats*, [en ligne], <http://resultats.hypotheses.org/267> (Page consultée le 20/11/2015). Bien que non nécessaire à ce stade de notre travail, nous réfléchissons donc à exploiter ces outils permettant de contribuer à la communauté TEI afin de rendre notre corpus exploitable par d'autres chercheurs.

[11] Cette démarche est finalement relativement proche de la conservation préventive pratiquée par les musées. Il s'agit en effet de baliser une série d'actions menées en faveur des collections – ici notre corpus textuel – dans le but de prévenir les dégradations et de prolonger la durée de vie des documents numériques qui le composent. Le format TIFF, non dégradable, est ainsi pensé comme le support de sauvegarde qui se voit « enfermé » dans des réserves, autrement dit les serveurs informatiques de l'Université Bordeaux Montaigne, tandis que l'édition de PDF – fichiers moins lourds en terme de poids et sur lesquels la reconnaissance de caractères est pratiquée – permet de proposer aux publics des documents de substitution pour la communication en ligne.

[12] À ce jour, seuls les fonds de la bibliothèque universitaire de Bordeaux Montaigne ont été numérisés. Une convention avec la BnF est en cours d'élaboration afin de pouvoir rapidement recevoir les fichiers déjà numérisés par cet établissement.

Ainsi, l'éventualité de réaliser un développement informatique *ex nihilo* a rapidement était mise de côté. Rechercher une solution *open source* existante et disposant dans ses fonctionnalités courantes de la description des données selon un schéma interopérable mais aussi de la gestion de contributeurs, s'avère en effet plus pertinente, tant, d'ailleurs, du point de vue financier que de la déontologie professionnelle. La réversibilité du choix logiciel effectué à un temps t de l'histoire informatique est assurément un élément important, sinon crucial. Utiliser des solutions *open source*, c'est anticiper un éventuel changement de technologie et s'assurer de pouvoir migrer les données et métadonnées stockées sur une nouvelle plateforme. Une communauté de développeurs dont l'implication décline, des orientations technologiques qui éclatent, une solution aux fonctionnalités plus appropriées qui émerge, autant de raisons qui peuvent expliquer le besoin de modifier le support logiciel. Pour cela, le stockage des données doit être assuré par un système de gestion de base de données. De même, les capacités et la facilité d'intégrer des extensions au logiciel de base sont des éléments déterminants pour cette sélection. L'observation de sites réalisés dans des domaines similaires permet enfin de valider ou d'invalider les différents supports envisagés.

À cet égard, le logiciel libre Omeka[13] apparaît comme une solution spécialisée dans la gestion d'archives numériques et fortement ancré dans le paysage universitaire, à l'image du site créé par l'université de Clermont-Ferrand et présentant les carnets géologiques de Philippe Glangeaud[14], ou encore de 1886[15], la bibliothèque numérique des fonds patrimoniaux du service commun de documentation de Bordeaux Montaigne. Autorisant la mise en place d'une communauté qui alimente la base de données selon différents rôles et donc selon différents droits, permettant de regrouper les données collectées par thèmes et sous-thèmes, proposant un moteur de recherche puissant et appuyant la description des métadonnées sur le format international Dublin Core, cet outil propose de fait les quatre principales fonctionnalités recherchées par le comité de pilotage. Développé par le *Center Roy Rosenzweig for History and New Media* de l'université George Mason, celui-ci propose enfin de nombreuses fonctionnalités supplémentaires telle que la création à partir de ces mêmes documents d'expositions virtuelles, système de valorisation documentaire devenu un classique des productions numériques de lieux de mémoire tels que les musées, les bibliothèques ou les archives. Cependant, comme tout outil pré-existant, celui-ci suppose des contraintes qui sont liées aux choix pris par ses éditeurs ; il est par exemple impossible dans Omeka d'attribuer plusieurs collections à un objet ce qui limite l'usage des collections thématiques permettant d'organiser les textes. L'intégration des extensions ménage toutefois de nombreux accès aux données.

Le corpus numérique peut ainsi être livré en son entier, au fil des documents qui se voient structurés par ordre chronologique ou alphabétique en fonction du souhait de l'usager. Des éléments de ce corpus sont ensuite consultables au sein de parcours thématiques mis en avant par le comité scientifique[16] ou à partir d'un nuage de mots-clés visualisable en page d'accueil, créant ainsi des sortes d'anthologies plus ou moins vastes. Une attention particulière a également été apportée aux

---

[13] Pour une comparaison de ses fonctionnalités avec Drupal par exemple, on pourra consulter : Bugnon Nicolas, 2013 : *Bases de données en sciences humaines : Création et pérennisation*, Maîtrise universitaire ès lettres en informatique et méthodes mathématiques, Université de Lausanne, [en ligne] : <http://www.unil.ch/files/live//sites/sli/files/shared/bdsh.pdf>

[14] http://bibliotheque.clermont-universite.fr/glangeaud/ (Page consultée le 23/02/2015)

[15] http://1886.u-bordeaux3.fr/ Pour une introduction plus détaillée aux fonctionnalités d'Omeka, nous renvoyons à la page française de cet outil sur le wiki collaboratif des bibliothécaires (http://www.bibliopedia.fr/wiki/Omeka) ainsi qu'à sa présentation officielle (http://omeka.org/) (Pages consultées le 23/02/2015). Voir également le travail mené par Sylvain Machefert, ingénieur d'étude à la bibliothèque universitaire de Bordeaux qui a en parti accompagné l'équipe projet dans ses choix ; cf. Machefert Sylvain, 2015 : « Quelle place pour les bibliothèques dans les digital humanities ? », in Le Deuff Olivier (dir.), *Le temps des humanités digitales*, Paris : Fyp éditions, 157-162.

[16] Un comité scientifique élargi a en effet été associé au projet dès l'émergence des premières expérimentations en 2008 ; plusieurs dizaines de chercheurs, spécialisés en littérature mais aussi en histoire et en sciences de l'information et de la documentation sont ainsi sollicités par le comité de pilotage, équipe plus retreinte qui gère pour sa part les différentes étapes opérationnelles du programme.

articles évoquant l'Aquitaine puisque ces derniers sont tous géo-référencés et accessibles via une cartographie de la région générée par une Google Map. Si ces accès codifiés suggèrent des lectures relativement passives, l'usager étant fortement guidé dans sa découverte par les critères éditoriaux effectués en amont par les chercheurs, l'internaute a tout loisir de développer une démarche active. L'exploration du corpus peut effectivement être faite à partir du moteur généré par Omeka qui autorise les recherches plein-texte et interroge l'ensemble des métadonnées comme les éditeurs ou la typologie des textes.

Investigation poussée ou butinage plus ou moins orienté, un large panel d'appropriations est par conséquent proposé à l'internaute. Qu'il suive toutefois les anthologies proposées par les experts ou une recherche plus personnelle, l'usager accède à l'article en pointant sur la vignette du substitut iconique. Le visuel de l'article, qui représente en réalité une sorte de miniaturisation du PDF, est en effet valorisé aux deux moments clés de l'appropriation ; celui tout d'abord de la sélection du texte au cœur d'une longue liste d'objets ; celui ensuite de la lecture puisque, par défaut, c'est le mode image qui apparaît au centre de la visionneuse qui couvre l'écran. L'internaute est ensuite libre de consulter l'article en mode texte seul, l'écrit venant se positionner au centre de l'écran, ou en regard du substitut iconique. Le texte lisible sur la version publique est le résultat d'une extraction de « *l'océrisation* » mise en œuvre sur les fichiers images et d'une transcription de ces caractères selon un langage wiki au sein de la partie technique du site. Cette nouvelle inscription textuelle résulte du désir exprimé par le comité de pilotage d'augmenter l'expérience de lecture par une série d'annotations et de références facilitant l'appropriation cognitive de chaque texte. S'appuyant sur une syntaxe et un jeu de balises simplifiés, l'architecture wiki exploitée[17] transforme dans le même temps la plateforme de gestion documentaire en plateforme de rédaction et d'édition.

Pour autant, « Mauriac en ligne » ne peut être considéré comme un site collaboratif car seuls les membres du comité scientifique possédant un compte contributeur peuvent annoter les textes. L'extension wiki se voit ici détournée pour répondre à un besoin de contextualisation documentaire additionnel à la seule attribution de métadonnées descriptives. Ainsi, la possibilité pour chaque internaute d'accéder à la fenêtre de transcription de l'article lu, proposée par Omeka sous chaque document lors de l'intégration du wiki, a été volontairement supprimée de la version publique du site. À la différence donc du plus célèbre des wiki, Wikipedia qui parie sur la masse de rédacteurs et sur l'autorégulation des échanges lors de la rédaction de contenu (Mabillot, 2012), le savoir mis en avant ici demeure celui de l'expert qui a pris le temps de décortiquer le texte et de l'enrichir pour accompagner le visiteur-lecteur dans sa découverte. Conséquemment, la plateforme devient un espace hybride qui enferme à l'intérieur d'une seule structure les fonctions essentielles de préservation, d'étude informationnelle et de communication de la collection virtuelle patiemment rassemblée. Déposés, examinés et montrés via la base de données qui assure le fonctionnement du site, ces textes sont comme enchâssés dans cet espace numérique[18]. Extirpés du purgatoire et désormais exposés au culte des chercheurs, ils deviennent par-là même porteurs de nouvelles valeurs.

Muséaliser les textes de presse de François Mauriac

Traditionnellement assignées aux professionnels du patrimoine œuvrant à la valorisation d'œuvres historiques ou artistiques[19], les missions de préservation, recherche et communication sont

---

[17] Cette dernière se fonde sur l'utilisation de l'extension Scripto proposée par Omeka.

[18] Précisons toutefois que les données brutes issues de la chaîne de numérisation (TIFF/PDF/ODT) sont en réalité conservées de façon autonome par rapport au système d'éditorialisation ici présenté.

[19] « *Aujourd'hui, selon les statuts de l'ICOM, adoptés lors de la 21e Conférence générale à Vienne (Autriche) en 2007, un musée est une institution permanente sans but lucratif au service de la société et de son développement ouverte au public, qui acquiert, conserve, étudie, expose et transmet le patrimoine matériel et immatériel de l'humanité et de son environnement à des fins d'études, d'éducation et de délectation.* » http://icom.museum/la-vision/definition-du-musee/L/2/ (Page consultée le 24/02/2015).

désormais assurées par les chercheurs engagés dans ce programme. Celles-ci se traduisent par une série d'étapes qui va de l'extraction du texte primaire de son milieu d'origine vers une présentation publique en ligne après une contextualisation de celui-ci afin de lui redonner du sens. Chacune de ses activités se déroule dans un univers numérique : commencent d'abord la numérisation des articles à l'aide de scanners et le traitement OCR qui permet de leur restituer une textualité ; vient ensuite la thésaurisation de ces fichiers numériques sur les serveurs informatiques gérés par la DSI de l'Université Bordeaux Montaigne ; les objets documentaires sont alors finement examinés et ces études se matérialisent lors des phases d'indexation, de transcription et d'annotation des articles produites directement au sein de la base de données ; enfin, cette description documentaire qui a pour cadre de réalisation la partie technique du site se concrétise sur la version publique par les divers parcours et cheminements proposés à l'internaute. La base de données constituée pour accueillir, gérer et présenter cette collection virtuelle dans son intégralité fonctionne *in fine* comme un cybermusée[20].

En accompagnant la lecture des textes de commentaires indispensables à une bonne compréhension, d'illustrations qui les replacent dans leur contexte, en définissant un certain nombre de cheminements et d'itinéraires de découverte, celui-ci expose en effet les substituts numériques des articles de presse de François Mauriac afin que ces derniers documentent le travail journalistique de l'écrivain. Parallèlement, cette mise en scène digitale implique un changement de statut de ces objets ; devenus *musealia,* ils ont été *de facto* patrimonialisés puisque, comme le rappelle François Mairesse, « *tout ce qui est muséalisé est patrimonialisé* » (Desvallées et Mairesse, 2011, 254). En déplaçant l'article de l'espace tangible à l'espace numérique et en déployant ses substituts iconiques et textuels sur l'écran informatique, la numérisation et la diffusion en ligne de ces écrits participent à une redéfinition des frontières entre passé, présent et futur en travaillant la mémoire de l'écrivain. Cette transformation numérique ordonne en effet le processus de patrimonialisation de ses textes qui sont retrouvés, étudiés, contextualisés, exposés et conservés. Au-delà de ces gestes spécifiques, la patrimonialisation se caractérise plus particulièrement par une rupture mémorielle (Davallon, 2006).

Or, à travers les rééditions en volumes de ces propres textes, parfois modifiés et amendés, François Mauriac, en « *bon gestionnaire de ses écrits* » (Baudorre, 2011, 85), a lui-même tenté de verrouiller l'appréhension de son œuvre. Construisant ainsi sa propre mémoire via les nombreux recueils publiés de son vivant, l'écrivain a fixé pour la postérité son œuvre littéraire qui se matérialise par le caractère fini du volume édité (Audet, 2014). Cette mémoire allographique, idéale, créée et promue par l'auteur lui-même est bousculée par la muséalisation des substituts numériques. Érigés en objets patrimoniaux, ils suscitent une reconstruction historique qui peut aller à l'encontre de la mémoire que l'écrivain souhaitait voir perdurer. Et ceci est particulièrement vrai pour les textes exhumés de l'oubli car si la mémoire sélectionne et délaisse, l'histoire tente de reconstruire ce qui n'existe plus. Pour ne prendre qu'un exemple de cette tension permanente qui persiste entre mémoire et histoire, reprenons cet inventaire à la Prévert de Paul Cooke qui travailla entre 2008 et 2009 sur les textes publiés entre 1937 et 1938. Pour le seul mois de janvier 1937, l'universitaire énumère en ces termes les grands écarts journalistiques effectués par François Mauriac : « …*une conférence consacrée à Mozart […] ; deux billets parus dans l'hebdomadaire dominicain Sept […] dont la majorité des lecteurs seraient donc plutôt des catholiques de centre-gauche ; une chronique à la Une du Figaro, un des grands quotidiens parisien de l'entre-deux-guerres […] ; un court hommage à l'aviateur Jean Mermoz qui […] paraît à la Une des Croix-de-feu, l'un des plus célèbres mouvements nationalistes de l'entre-deux-guerres […] ; un article littéraire en Une du Temps, grand quotidien d'information qui jouissait d'une excellente réputation, un peu comme le monde de nos jours ; et, finalement, un long compte rendu d'une pièce de théâtre de Gabriel Marcel paru dans un*

[20] En suivant la définition du musée numérique, ou cybermusée, ou musée en ligne, proposée par le Dictionnaire encyclopédique de muséologie : « *catégorie de musées dont les collections ainsi que les modes de communication (exposition et médiation) sont essentiellement composées de manière numérique et exploitées, à partir de base de données sur ordinateurs, généralement via Internet.* » (Desvallées et Mairesse, 2011, 631).

hebdomadaire d'extrême-droite caractérisé par l'antiparlementarisme, l'anticommunisme, l'antisémitisme, l'anglophobie et un soutien politique marqué pour les régimes de Mussolini et du général Franco » (Cooke, 2011, 75).

Du point de vue herméneutique, cela signifie que l'équipe scientifique impliquée dans le programme ne tente pas de faire perdurer la mémoire mise en scène par l'académicien mais aspire au contraire à restaurer un corpus complet et fiable afin de mieux connaitre le travail journalistique de François Mauriac. En restaurant la complexité de cette collection textuelle, il s'agit de faciliter la mise en œuvre d'études inédites sur l'écrivain et son environnement, non seulement littéraire, mais aussi sociétal, politique ou religieux. Cependant, l'abandon de la TEI et la mise à disposition de l'image signent le passage du scriptuaire à l'indiciel, c'est-à-dire du texte au substitut iconique de l'article ; mouvement qui imprime une valeur avant tout patrimoniale à ce corpus numérique, désormais objet du passé dans le présent. En exposant et en valorisant la matérialité physique de ces objets textuels numériques, la plateforme de diffusion – et bien que les choix éditoriaux de l'équipe projet soient en cours de construction – devient le lieu d'une mémoire autographique et numérique dont la valeur se métamorphose (Kaplan, 2009 : 9). Cette valeur, non plus seulement artistique et littéraire mais désormais aussi patrimoniale, est dissociée de l'objet texte papier car la mémoire informationnelle qui est élaborée par les chercheurs lors de leur travail de description et d'annotation demeure dans l'espace virtuel du Web. Dans l'état actuel du travail mené par l'équipe, la forte intégration de ce projet dans un paysage Mauriacien aquitain ainsi, par exemple, que certains choix éditoriaux formulés par l'équipe comme la création d'une collection thématique « Aquitaine » ou l'accès géographique aux textes évoquant la région natale de l'écrivain, donnent finalement à voir la construction d'un patrimoine immatériel territorial, une mémoire régionale.

Conclusion

Exhumés de l'oubli et extraits de leur environnement naturel, les textes de presse de François Mauriac sont, à travers le programme que nous venons de présenter, en cours d'intégration dans un corps virtuel qui s'érige en ensemble homogène grâce au dispositif numérique construit. Articulant ce mouvement de patrimonialisation qui va de la redécouverte à la transmission, le système élaboré par l'équipe emboîte en effet dans un même environnement digital conservation, diffusion et consultation. En créant un écosystème de connaissances hypertextuelles, il réorganise par-là même l'appréhension info-communicationnelle de ces documents. Extrayant, lors des phases d'indexation et d'annotation, les informations littéraires mais aussi historiques, politiques ou sociétales dont sont porteurs ces textes, les chercheurs engagés dans ce projet proposent de nouvelles formes d'appréhension de ces articles. Ces manipulations s'inscrivent dans des pratiques de médiation documentaire déjà connues et se rapprochent du travail que réalisent depuis maintenant plus de deux siècles les établissements culturels gardiens de la mémoire collective que sont les musées. L'expérience que ces derniers possèdent en termes de valorisation et de mise en scène patrimoniales influence d'ailleurs la mise en espace de ces ressources documentaires, comme l'attestent les collections thématiques ou l'usage de l'expression « exposition virtuelle » proposée par le module « Exhibit » d'Omeka.

L'analyse de ce programme permet, *in fine*, d'analyser la fabrique d'un patrimoine écrit, fabrique aujourd'hui influencée par les techniques et pratiques numériques d'édition et de lecture. Plus largement, cette nouvelle documentarisation, qui illustre la troisième vie des articles de presse de François Mauriac, participe à une redéfinition des rapports unissant le fragment daté du journaliste à l'œuvre intemporelle de l'écrivain, entre possible trahison et impossible fidélité à sa mémoire. Car au-delà de cette muséalisation, l'objectif de ce programme de recherche est finalement de remettre ces « *ressources patrimoniales au cœur d'une activité de production – quand bien même les produits de cette activité ne seraient pas pérennes* » (Merzeau, 2014, 63). Pour ne pas laisser ces objets patrimoniaux inertes, il reste désormais aux mauriaciens de s'emparer de cet outil pour construire de

nouvelles connaissances. Ces opérations qui ont engendré, par la numérisation d'un corpus, par la description de données et par la mise à disposition de ces textes, la construction d'un patrimoine numérique ne constituent qu'une « *approche « restreinte » des humanités digitales ; […] qu'un premier niveau* » (Cormerais, 2015 : 138) d'une démarche plus globale et transdisciplinaire dont l'objectif, qui reste à construire, est de produire des interprétations scientifiques inédites.

Bibliographie :

Audet, René, 2014 : « Le long, le bref et le truchement numérique », in *Le temps des humanités digitales*, sous la dir. d'Olivier Le Deuff, Limoges : fyp, p.89-102.

Baudorre Philippe, 2011 : « De Mauriac à Mauriac, la seconde vie des textes de presse », *Nouveaux Cahiers François Mauriac*, n°19, p.83-104.

Bugnon Nicolas, 2013 : *Bases de données en sciences humaines : Création et pérennisation*, Maîtrise universitaire ès lettres en informatique et méthodes mathématiques, Université de Lausanne, [en ligne], <http://www.unil.ch/files/live//sites/sli/files/shared/bdsh.pdf>

Cooke, Paul, 2011 : « Vers une édition électronique de l'œuvre journalistique de François Mauriac (1937-38) : premiers bilans », *Nouveaux Cahiers François Mauriac*, n°19, p.73-82.

Cormerais Franck, 2014, « Humanités digitales et (ré)organisation du savoir », in *Le temps des humanités digitales*, sous la dir. d'Olivier Le Deuff, Limoges : fyp, p. 129-142.

Davallon Jean, 2015 : « Mémoire et patrimoine : pour une approche des régimes de patrimonialisation », in *Mémoire et nouveaux patrimoines* [en ligne], Marseille : OpenEdition Press, <http://books.openedition.org/oep/444>.

Davallon Jean, 2006 : *Le don du patrimoine, une approche communicationnelle de la patrimonialisation*, Paris : Lavoisier – Hermes science.

Desvallées André et François Mairesse (dir.), 2011 : *Dictionnaire encyclopédique de muséologie*, Paris : A. Colin.

Desvallées André, 1989 : « Collecte en Aubrac », in *Cours de Muséologie selon Georges-Henri Rivière*, Paris : Dunod, p.185-192.

Flon Émilie, 2012 : *Les mises en scène du patrimoine : savoir, fiction et médiation*, Paris : Hermès/Lavoisier.

Genette Gérard, 1994 : *L'œuvre de l'art, Immanence et transcendance*, Paris: Poétique/Seuil.

Frédéric Glorieux, 2015 : « Le traitement de textes (odt) pour produire des documents structurés (XML/TEI) – Odette », in *J'attends des résultats*, [en ligne], <http://resultats.hypotheses.org/267>

Hudrisier, Henri et Azémard, Ghislaine, 2014 : « La TEI : un collège mondial et un outil commun pour la recherche en littérature », *Les Cahiers de la société française des Sciences de l'information et de la communication*, n°10, p.107-126.

Jollet, Etienne et Claude Massu (dir.), 2012 : *Les images du monument de la Renaissance à nos jours*, Aix-en-Provence : Presses Universitaires de Provences.

Kaplan Frédéric, 2012 : *La métamorphose des objets*, 2e édition, Limoges : fyp.

Le Deuff, Olivier (dir.), 2014 : *Le temps des humanités digitales*, Limoges : fyp.

Mabillot Vincent, 2012 : Le wiki : un dispositif d'écriture émergente publique et coopérative ?, *Communications et Langages*, n°174, p.69-84.

Machefert Sylvain, 2015 : « Quelle place pour les bibliothèques dans les digital humanities ? », in *Le temps des humanités digitales*, sous la dir. d'Olivier Le Deuff, Limoges : fyp, p. 157-162.

Merzeau Louise, 2014, Mémoire numérique : entre éditorialisation et grammatisation, *Les cahiers de la SFSIC*, n°10, p.56-65.