



CLASSIQUES  
GARNIER

DEMONET (Marie-Luce), « Lecture archéologique de l'Exemplaire de Bordeaux », *Bulletin de la Société internationale des amis de Montaigne*, 56, 2012 – 2, p. 321-337

DOI : [10.15122/isbn.978-2-8124-3977-3.p.0321](https://doi.org/10.15122/isbn.978-2-8124-3977-3.p.0321)

*La diffusion ou la divulgation de ce document et de son contenu via Internet ou tout autre moyen de communication ne sont pas autorisées hormis dans un cadre privé.*

© 2013. Classiques Garnier, Paris.  
Reproduction et traduction, même partielles, interdites.  
Tous droits réservés pour tous les pays.

## LECTURE ARCHÉOLOGIQUE DE L'EXEMPLAIRE DE BORDEAUX

Le 18 juillet dernier (2012), l'Agence Nationale de la Recherche a retenu le projet MONLOE (« Montaigne à l'œuvre »), déposé par l'équipe des Bibliothèques Virtuelles Humanistes (BVH) du Centre d'Études Supérieures de la Renaissance (CESR) et la section de l'humanisme de l'IRHT, pour un financement de trois ans. Au moment du colloque du centenaire de la SIAM en juin, la sélection n'étant pas connue, il nous a paru préférable d'attendre quelques semaines avant d'en exposer le projet aux membres de la société. Jusqu'ici, le secteur SHS de l'Agence n'avait guère favorisé les projets présentés par l'équipe des BVH<sup>1</sup>.

Dans ses grandes lignes, MONLOE avait été annoncé dès l'été 2011 : il est temps désormais d'en présenter ce qui concerne *EB*, dans le cadre des discussions sur une édition optimale des *Essais*. Il est certain que les débats suscités au sein de la SIAM depuis ces dix dernières années ont été extrêmement stimulants, mettant en évidence les demandes contradictoires des chercheurs : un texte à la fois stratifié et lissé, lisible, définitif et génétique, avec la possibilité de fouiller toutes les versions. Le dernier échange entre André Tournon et Alain Legros (*BSAM* de juillet 2012) montre qu'il est encore nécessaire de préciser ce que l'on

---

1 Sur les six dossiers déposés depuis 2005 (l'équipe des BVH portait trois d'entre eux), aucun n'avait été sélectionné. Seul le projet NAVIDOMASS (2006-2010), proposé en STIC par le laboratoire d'informatique de La Rochelle, avait inclus les BVH comme partenaire dans un programme ANR : il a permis de développer les travaux sur la reconnaissance semi-automatique des lettrines et des caractères anciens, accompagnés d'un protocole d'encodage en XML/TEI « Renaissance et temps modernes ». Les publications numériques et les recherches des BVH depuis 2003 ont été financées par le Ministère de la Recherche, la Région Centre, le CNRS, l'Université de Tours, le Ministère de la culture, des crédits européens FEDER, et même par des bourses Google. Depuis 2012, elles sont soutenues également par l'atelier numérique de la Maison des Sciences de l'Homme Val de Loire (universités d'Orléans et de Tours), le PRES Centre Val de Loire Universités et l'Equipex BIBLISSIMA, dont le CESR est partenaire.

peut faire pour éclairer textuellement les enjeux éditoriaux, eux-mêmes appuis d'interprétations diverses.

MONLOE est la première phase d'un ensemble éditorial multiformat qui devrait nourrir un corpus d'auteur destiné à un public diversifié (chercheurs, étudiants, enseignants, amateurs), avec un portail dédié mais intégré au programme des BVH. Il s'agit d'abord de réunir les principales éditions des *Essais* jusqu'à 1598, l'édition *archéologique* d'EB, tous les ouvrages possédant la signature de Montaigne conservés dans les différentes bibliothèques publiques et privées dans le monde (une centaine)<sup>1</sup>, tous les écrits de Montaigne accessibles et une sélection des sources majeures des *Essais*. Il offrira aussi une reconstitution 3D de la « librairie<sup>2</sup> ». Une version sonore des *Essais* était envisagée, avec la lecture intégrale du texte original en prononciation moderne : à cause du budget alloué, moins élevé que prévu, l'Agence a écarté cet élément.

Le projet permettra de montrer et d'exploiter les ressources existantes et d'expérimenter des procédures nouvelles, avec des méthodologies en développement dans le domaine des humanités numériques : affichage page à page et parallèle de l'image et du texte, corpus textuel balisé permettant plusieurs niveaux d'analyse et de requête, identification des « mains » de Montaigne et visualisation des strates. Les méthodologies élaborées pour la reconnaissance des écritures et le traitement des données textuelles, tout comme les données issues du projet, seront librement diffusées et pourront être réutilisées par d'autres équipes.

## GÉNÉTIQUE, DIACHRONIQUE OU ARCHÉOLOGIQUE ?

Avec Alain Legros, nous avons qualifié de « génétique » le traitement spécifique envisagé pour EB, au moment où le projet commençait à prendre forme, quand une journée d'étude organisée au CESR par

1 Cette partie du projet a été présentée plus longuement lors d'une communication au colloque de la FISIER (« Bibliothèques et collections », Cambridge, 17-19 septembre 2012) : « Reconstituer la "librairie" de Montaigne ». Pour la présentation de l'ensemble, voir le site des BVH ([bvh.univ-tours.fr](http://bvh.univ-tours.fr)). L'avancement des travaux sera présenté dans le carnet de recherche des BVH sur [hypotheses.org](http://hypotheses.org).

2 En partenariat avec Archéotransfert (R. Vergniewux et P. Mora, Université de Bordeaux-CNRS).

l'ITEM avait soumis des solutions de balisage d'*EB*, à partir de principes génétiques d'édition<sup>1</sup>. Une première proposition avait été élaborée par Laetitia Bontemps (CESR) auprès du consortium TEI (« Text Encoding Initiative »). Lou Burnard (Oxford) et Elena Pierazzo (King's College, Londres) proposaient d'appliquer les principes de l'édition génétique selon la terminologie de l'ITEM, à cette sorte de « brouillon » que représente *EB*. Toutefois les usages établis alors par l'ITEM pour les brouillons d'écrivains ne convenaient pas tout à fait à un document hybride, à la fois imprimé *et* manuscrit. Un groupe de travail coordonné par le consortium CAHIER offrira un cadre collaboratif plus adéquat pour élaborer des méthodes éditoriales sur ces objets complexes, en relation avec les projets internationaux<sup>2</sup>.

Le terme même de « génétique » inscrit déjà une interprétation au sein du travail éditorial. Il suppose résolu de nombreux problèmes qui se posent d'abord au stade de la transcription, laquelle ne peut susciter que des hypothèses sur une éventuelle chronologie de rédaction. Tel qu'il avait été déposé à l'ANR en octobre 2011, le projet mettait en avant une édition « diachronique », afin de représenter les dépôts chronologiques de la composition entre l'imprimé de 1588 et l'exemplaire de travail de 1592. Finalement, c'est plutôt le terme employé par Alain Legros dernièrement qui est le plus adapté à notre démarche, à savoir une édition « archéologique », moins pour nous référer à Foucault que pour mettre l'accent sur l'exploration de ce champ de fouilles, grâce à la lecture minutieuse et à la mise en évidence des dépôts laissés par l'impression et l'écriture.

Ce projet était-il *nécessaire* ? Je ne suis pas une spécialiste de l'édition critique, puisque je n'en ai réalisé qu'une, en collaboration, celle du livre VIII des *Recherches de la France* d'Étienne Pasquier (1996). Outre le travail de normation des textes numérisés pour Frantext<sup>3</sup>, je n'ai

- 
- 1 Manifestation organisée par Bénédicte Vauthier (Université de Berne), janvier 2011 : « Édition critique et génétique de manuscrits du Moyen Âge à nos jours ».
  - 2 MONLOE s'inscrit dans le cadre des missions définies par le consortium « CAHIER » (Corpus d'Auteurs pour les Humanités : Informatisation, Édition, Recherche) de l'infrastructure de recherche « CORPUS » lancée par le ministère de la recherche en juin 2011. Le CESR, au sein de la MSH Val de Loire, est coordinateur du consortium et un atelier « édition de manuscrits » est en projet. Le programme européen COST « Interedition » développe des outils adaptés aux éditions savantes et génétiques en ligne.
  - 3 Clermont-Ferrand, 1991-1997, textes français du XVI<sup>e</sup> siècle fournis à l'INaLF (devenu l'AtiLF) par le laboratoire EQUIL XVI.

réalisé ou supervisé que des éditions sans appareil critique comme celle de Rabelais (1995), des transcriptions, d'abord en html à partir de 1998 à Poitiers, puis progressivement en langage XML à Tours depuis 2007. Le principe même de la transcription, souvent considéré comme évident, est toutefois un objet d'étude et de recherche en soi avant toute ambition d'approcher la dimension créative. Certains membres de la communauté TEI ont consacré à la transcription plusieurs études qui montrent que, même à ce niveau apparemment élémentaire, les recherches sont toujours en cours, modifiées par les possibilités techniques de balisage<sup>1</sup>. Les normes de transcription changent avec les progrès des langages utilisés, de l'encodage établi selon les exigences de communautés scientifiques diverses. Elles sont aussi dépendantes des outils d'interrogation, qui ne permettent pas encore d'effectuer toutes les requêtes souhaitables, alors même qu'un balisage patient a déployé toutes les ressources de l'érudition et de la philologie numérique.

C'est aussi une démarche déontologique que de vouloir livrer *EB* dans ses dépôts, avec ses imperfections, lacunes et leçons incertaines. Elle plaide pour l'accessibilité gratuite des œuvres patrimoniales et envisage comme un devoir moral de ne pas réserver à une élite un document qui est, selon les critères du Ministère de la culture, un « trésor national ». Une façon de se rendre utile à la communauté des lecteurs de Montaigne, surtout de ses lecteurs futurs, et de donner déjà à lire un texte aussi authentique que possible.

L'édition d'*EB* est le cœur d'un corpus susceptible d'enrichissements ultérieurs. Nous étions tentés depuis longtemps par l'idée d'une restitution au plus près de son état visible et j'avais fait quelques tentatives en traitement de texte à Poitiers, puis au sein du CESR pour la disposition des additions marginales. Le legs « Gilbert de Botton » à la Bibliothèque de l'Université de Cambridge et le colloque de 2008<sup>2</sup> ont été l'élément déclencheur de ce qui correspondait bien à un projet que

---

1 Voir les différentes contributions aux congrès des Digital Humanities, dont les résumés sont en ligne, en particulier celle de Michael Sperberg-McQueen, Claus Huitfeldt, Yves Marcoux (Université du Maryland, 2009). Ces conférences annuelles sont organisées par The Association for Literary and Linguistic Computing, The Association for Computers and the Humanities, The Society for Digital Humanities. Une communauté francophone vient de se constituer.

2 Les actes du colloque de Cambridge viennent de paraître en septembre 2012 dans la collection « Cambridge French Colloquia », édités par Philip Ford.

l'ANR pourrait soutenir, au moment où se mettait en place en France l'infrastructure CORPUS destinée à favoriser la publication en ligne de sources partageables.

L'ambition du projet est aussi d'avancer dans nos connaissances sur l'écrit manuscrit ou imprimé, au XVI<sup>e</sup> siècle, au moment où l'imprimerie semble imposer une nouvelle façon, plus « civile », non seulement de lire, mais aussi de concevoir. Les caractéristiques mêmes du manuscrit chez Montaigne ont longtemps été négligées, alors qu'elles témoignent de ses graphies spontanées ou contrôlées et d'un rapport à la matérialité de l'écriture manuscrite invisible dans l'ouvrage imprimé, pour ressurgir sur *EB*. Si le fait de *voir* l'écriture de Montaigne ne nous donne que de rares indices sur la façon dont il disait, prononçait, « dictait », en revanche l'examen des indices graphiques invite à entrevoir une relation entre l'écrit et l'oral, entre l'écrit et le geste, cet *impetus* du mouvement de la main dont on suppose qu'il prolonge le mouvement de l'esprit en enquête. Quant aux retouches de ponctuation ou de majuscules « de scansion » sur les parties imprimées (A. Tournon), elles suggèrent une intention de fragmenter et de lier en même temps le rythme de la pensée. Encore faut-il pouvoir montrer clairement ces changements, ce que l'encodage en TEI permet désormais de faire depuis que les signes de ponctuation (ou la ponctuation zéro) disposent de leurs balises spécifiques<sup>1</sup>.

#### NUMÉRISER/TRANSCRIRE/VISUALISER

Le traitement des ensembles de livres et de documents constitués autour d'un auteur bénéficie de plusieurs modèles accessibles, certains demandant une adaptation experte, d'autres étant aisément applicables<sup>2</sup>. Suivant la variété des approches, l'accent est mis sur le document reproduit en numérisation patrimoniale, sur un ensemble documentaire (bases de données encyclopédiques) ou sur sa version textuelle

- 
- 1 Propositions de balisage par Alexei Lavrentiev (laboratoire ICAR, ENS-Lyon), acceptées par la TEI en 2010 et intégrées aux Guidelines.
  - 2 Voir les projets Stendhal à Grenoble, *Bovary* à Rouen, *Bouvard et Pécuchet* à Lyon, Descartes à Caen, et bien d'autres corpus d'auteurs hors de France.

(corpus linguistiques). Le principe de la « numérisation patrimoniale » appliqué par les Bibliothèques Virtuelles Humanistes s'étend au projet MONLOE, associé à une exploitation éditoriale, textuelle et graphique. Il part du fac-similé numérique : les cahiers des charges sont établis en concertation avec la Bibliothèque nationale de France et le Ministère de la culture, dans le cadre des pôles associés. Ces normes qui exigent désormais une haute définition sont indispensables pour pouvoir donner une image numérique qui supporte les traitements nécessaires à la reconnaissance de l'écriture et au traitement spécifique des caractères, des encres et ornements typographiques : raison majeure pour laquelle il faudra numériser à nouveau, cette fois directement sur l'original, le précieux exemplaire conservé par la Bibliothèque Municipale de Bordeaux. En attendant cette réalisation, nous nous servons naturellement de sa disponibilité (en lien avec la reproduction de l'édition Villey) sur le site du *Montaigne Project* de Philippe Desan à Chicago<sup>1</sup>.

#### FAC-SIMILÉS

Les BVH ont déjà mis en ligne (2007) le fac-similé de l'exemplaire des *Essais* de 1580 conservé au Musée de Sologne de Romorantin, auquel sera associée la transcription quasi-diplomatique, gérée en même temps que celle de 1582. Les innombrables corrections effectuées sur différents exemplaires de 1580 invitent à proposer un texte unique, avec affichage au choix de l'état de 1580-Romorantin (type 2 pour la page de titre du vol. 1), complété par l'état de 1582 (exemplaire de la Bibliothèque municipale de Bordeaux).

Comme les autres fac-similés, 1588 sera traité par AGORA, logiciel mis au point par le Laboratoire Informatique de Tours. Cet outil paramétrable permet l'identification et l'extraction automatique de différents éléments de contenu pouvant apparaître dans les ouvrages anciens (lettrines, bandeaux, portraits, notes manuscrites, signatures, réclames, zones de texte imprimé, caractères typographiques). Il fournit en sortie un format ALTO qui établit les coordonnées de chacun des éléments de l'image et lie automatiquement l'image au texte. La transcription d'*EB* se fera directement dans un fichier XML structuré avec ALTO.

1 <http://artflx.uchicago.edu/cgi-bin/philologic/getobject.pl?c.0:2:1.montaigne>.

## EB EN TEI

Le standard XML/TEI a mis longtemps à être accepté en France pour l'édition des textes et même s'il ne fait pas l'unanimité il n'y a pas pour le moment d'alternative à un protocole international longuement discuté au sein des communautés de la philologie numérique. L'encodage n'est pas entièrement adapté aux traitements offerts par la linguistique de corpus, mais correspond assez bien aux souhaits des philologues et des historiens. Il présente à la fois les avantages et les inconvénients de la standardisation : les choix d'inclusion et de hiérarchisation des balises, tous justifiés, ne correspondent pas à tous les cas qu'offre la réalité des textes, notamment pour l'édition de documents mixtes manuscrits/imprimés. Le langage XML exclut certains chevauchements de balises.

Le protocole « TEI-Renaissance », disponible dans le *Manuel d'encodage* en ligne<sup>1</sup>, est intégré au projet BVH depuis 2007. Les informations complètes sur la nature, les étapes et les auteurs de la transcription sont contenues dans l'en-tête (le « header ») de la version en TEI (téléchargeable) et abrégées dans une notice qui apparaît en info-bulle.

Les métadonnées intégreront aussi des éléments provenant de la TEI. Elles seront produites aux formats standard (MARC, MARC 21 MARC/TEI, DublinCore, RDF), en collaboration avec la Bibliothèque nationale de France et le TGE Adonis, de façon à constituer des entrepôts de métadonnées moissonnables par les principaux agrégateurs et diffuseurs de données patrimoniales : Gallica, Europeana, Isidore. Le modèle préparé avec le logiciel KOHA par les BVH pour la rétroconversion du catalogue des Incunables pourrait être adapté pour constituer le catalogue du portail MONLOE : paramétrable, il a l'avantage de traiter les champs de la notice canonique et de permettre la description normalisée des particularités d'exemplaires.

## L'INTERFACE IMAGE/TEXTE

Le projet doit tenir compte de plusieurs évolutions techniques qui affectent la visualisation : la possibilité d'avoir des écrans très grands, tout autant que la généralisation des écrans très petits (tablettes et smartphones). Dans le premier cas, le lecteur peut afficher en parallèle

1 <http://www.bvh.univ-tours.fr/XML-TEI/index.asp>, 2008, mis à jour en 2009 et 2012.

plusieurs versions et peut toujours revenir au fac-similé face à sa transcription. Le fac-similé est lui-même un « surrogate », un substitut et un simulacre du document original et n'a pas la prétention d'en posséder toutes les qualités. Il ne fait que s'en approcher. Pour les petits écrans, le fac-similé et les transcriptions doivent être redimensionnés pour que ce corpus puisse être portable et convertible au format ePub.

Pour gérer l'ensemble de ses documents (fac-similés, transcriptions, PDF contenant des textes ocrésés à la volée), les BVH ont choisi la plateforme XTF en 2009, ce qui entraîne certaines contraintes d'affichage et limite les fonctions linguistiques avancées<sup>1</sup>. XTF est un système de gestion et d'interrogation des métadonnées et des bases tous formats, développé par l'Université de Berkeley. L'affichage parallèle entre le texte et l'image est en cours d'expérimentation, car il demande d'adapter la double structure du document, image en jpg d'un côté, texte encodé en TEI de l'autre : il existe en effet un conflit fondamental entre la structure physique du document, dont l'unité est la page, et sa structure logique (titre, pièces liminaires, chapitres, phrases), dont l'unité est linguistique (la phrase/le mot). Ces deux « ontologies » (au sens des informaticiens) exigent un redécoupage en unités qui dépendent du résultat que l'on veut obtenir, souvent incompatible avec la segmentation indispensable à la requête linguistique. Un modèle intéressant est celui de la DTA (Deutsches Textarchiv), qui a développé son propre logiciel propriétaire et établi un compromis entre le découpage physique et la structure logique en s'écartant quelque peu des contraintes du langage XML/TEI. La décision de conserver la page et son lignage comme unités physiques de segmentation, alors que toutes les éditions modernes des *Essais* et les corpus linguistiques les évitent, est un choix lourd de conséquences, gérable seulement par des feuilles de style (XSLT) dont l'adaptation est en cours. Le téléchargement de la version en XML permettra de récupérer le fichier selon les règles juridiques de référencement (Licence Creative Commons), réappropriable par quiconque voudrait effectuer une annotation experte, moyennant autorisation, nécessaire à l'édition critique.

*EB* est un document « profond », avec ses additions, ses variantes, ses ratures, surcharges et lacunes. Alain Legros a d'abord publié le chapitre

---

1 <http://xtf.cdlib.org/>

« Des prières » (I, 56)<sup>1</sup> puis transcrit tout dernièrement « De la force de l'imagination » (I, 21) pour tester les difficultés d'une édition étendue du document, en insistant sur sa dimension forcément lacunaire<sup>2</sup>. Or une édition en XML/TEI peut rendre manifestes ces lacunes, tout en tenant compte de la nécessité de recourir à l'édition de 1595 que le lecteur peut choisir de voir, ou non. Il faudra donc adapter le « TEI viewer » déjà opérationnel sur le site des BVH<sup>3</sup>, et construire un outil compatible avec XTF mais plus adapté aux fonctionnalités requises, notamment à l'édition multistrate. S'il est envisageable de donner en visualisation alternative les textes qui ont surtout des variantes l'un par rapport au précédent (comme l'édition de 1582), il faut trouver une autre solution pour les états très différents comme les *Essais* de 1580 par rapport à 1588, 1595 et à plus forte raison pour *EB*. Plusieurs outils éditoriaux ou linguistiques (corpus parallèles) sont en cours de développement pour permettre de visualiser différentes versions d'un même texte sans excéder les limites de la lisibilité<sup>4</sup>.

#### L'« ORTOGRAPHE<sup>5</sup> » ET LA VARIATION GRAPHIQUE DANS *EB*

Les études sur l'orthographe de Montaigne, menées de façon ciblée par André Tournon sur la ponctuation et les majuscules, et sur les pratiques des typographes par Nina Catach après Charles Beaulieux, pourront être prolongées ou renouvelées par la transcription du plus grand nombre possible des manuscrits autographes de Montaigne. En effet, beaucoup d'études se fondent avant tout sur les éditions imprimées, dont l'auteur écrit lui-même qu'elles n'ont pas suivi ses demandes. Il sera possible d'exploiter ultérieurement ses usages graphiques, avant

1 Genève, Droz, 2003.

2 L'article qu'il publie ici-même en donne un aperçu.

3 Voir l'édition en ligne des *Nouvelles créations* de Bonaventure Des Périers.

4 Frédéric Glorieux et Vincent Jolivet, informaticiens partenaires du projet MONLOE, ont développé une librairie logicielle robuste et modulaire, qui accélère le développement d'interfaces spécifiques dans un langage très standard (PHP) et offrira une solution différente de la plateforme XTF tout en étant compatible avec elle.

5 Graphie de Montaigne dans les *Essais* de 1588 (f° 425<sup>r</sup>°).

toute intention et pratique d'auteur, dans les arrêts, notes du Beuther et notes du Nicole Gilles, puis dans les autres notes de lecture et les lettres (évolution de *-eur* vers *-ur*, *oë/oi*, *ste/cete*. . . .), où les traces d'occitanisme tendent à disparaître, et dans ces témoignages remarquables que sont les additions d'*EB* et les autres autographes conservés<sup>1</sup>.

Pourquoi ne pas moderniser complètement le texte ? L'établissement d'un état de transcription quasi-diplomatique (maintien de la distribution *i/j*, *u/v* et des caractères tildés) offre plusieurs avantages : outre les recherches sur l'écrit, qui actuellement intéressent de plus en plus les linguistes<sup>2</sup>, et sur les manuscrits, c'est également l'écrit comme trace éventuelle de l'oral qu'il convient de conserver, même s'il apparaît que l'écrit, et surtout l'écrit imprimé, impose au français du XVI<sup>e</sup> siècle une lecture du texte non seulement indépendante de l'oral mais *régissant* l'oral. André Tournon a proposé un lien entre la ponctuation, les majuscules de scansion et un enregistrement greffier des témoignages judiciaires. Il convient donc d'étendre le principe de reproduction quasi-diplomatique à tous les graphèmes et non seulement aux ponctèmes. Mes propres recherches vont aussi dans le sens d'une influence rétroactive de l'écrit sur l'oral, l'écrit juridique – manuscrit, puis imprimé – imposant ses usages à une oralité qui n'a plus rien de spontané. L'idée que Montaigne se sert de ce double mouvement oral => écrit et aussi écrit => oral « intérieur » (qui sonne « au-dedans de nos oreilles ») reste à faire<sup>3</sup>. Le choix d'une version de base quasi-diplomatique, qui ne s'impose pas pour tous les auteurs ni pour tous les textes, est nécessaire dans le cas d'*EB* à un alignement précis image/texte et aux travaux d'histoire de la langue écrite, à la prise en compte de l'hétérographie, à une indexation lexicale fiable, à l'efficacité des moteurs de recherche et des outils statistiques.

À l'université de Chicago le projet ARTFL (Mark Olsen) offre son logiciel PhiloLogic pour permettre la recherche partiellement lemmatisée dans ses bases de textes français, et dans un premier temps cette procédure sera mise en place dans le projet MONLOE, comme

1 Voir A. Legros, *Montaigne manuscrit*, Paris, Classiques Garnier, 2010.

2 Colloque SHESL/HTL, Paris, janvier 2013 : « Écriture(s) et représentations du langage et des langues ».

3 Voir notre article à paraître : « Parler comme les livres : la parole civile en France à la Renaissance », dans les actes du colloque d'Urbino, org. N. Panichi, 2011, « La civile conversation ».

pour les BVH. PhiloLogic, outil puissant qui permet de contourner la variation graphique, est déjà utilisé par la base Epistemon : avec lui, il ne serait pas nécessaire de régulariser les *il/j* et *u/v*, mais le détildage préalable est indispensable. Toutefois, les principes éditoriaux de la base textuelle Epistemon sont de pouvoir afficher les deux versions (originale ou quasi-diplomatique/ régularisée ou patrimoniale) en appliquant les feuilles de style du moteur XTF<sup>1</sup>. PhiloLogic, qui traite aussi les requêtes dans le *Montaigne Project*, ne cherche que dans la version régularisée sans tenir compte du balisage TEI (noms propres en particulier). La recherche s'effectue à l'aide de tables de correspondance, de dictionnaires, d'expressions régulières, et n'est possible que sur un état du texte à la fois. La ponctuation n'est pas prise en compte : or elle est très importante pour la mise en évidence des retouches effectuées par Montaigne sur EB<sup>2</sup>.

Des outils linguistiques spécifiques sont développés dans le projet pour régulariser et effectuer des requêtes sur les *Essais* et les autres textes, notamment les ouvrages de la « librairie » : en partenariat avec le CESR, Marie-Hélène Lay (FORELL, Université de Poitiers) a développé Varialog, générateur de requêtes étendues, qui inclut un module de désabréviation et de dissimilation des textes imprimés de la Renaissance, et permet d'effectuer des requêtes à partir des graphies modernes sans se préoccuper des variantes<sup>3</sup>.

Les tests menés sur un corpus rabelaisien serviront à améliorer les résultats sur le corpus MONLOE, en collaboration avec les travaux du laboratoire ICAR (ENS-Lyon) sur la base du français médiéval (BFM), qui inclut un lemmatiseur-étiqueteur pour l'ancien français. Le moteur d'indexation linguistique Corpus Workbench (CWB) de l'université de Stuttgart est particulièrement adapté à l'exploration de corpus étiquetés et lemmatisés : c'est la librairie d'indexation choisie par le projet Textométrie (Serge Heiden, ICAR).

- 
- 1 Epistemon est l'ensemble des textes (et non des fac-similés) publiés par les BVH, interrogables par des outils linguistiques.
  - 2 Elle ne l'est pas non plus actuellement avec l'outil développé par Gilles Souvay (AtiLF) à partir de la base du moyen français. Le système de règles est très performant mais, établi à partir d'un état de langue plus ancien, il produit actuellement trop de « bruit » pour les textes du XVI<sup>e</sup> siècle.
  - 3 Varialog est développé à l'aide d'une bourse Google, obtenue par le CESR en 2010.

## PRINCIPES DE TRANSCRIPTION

Les transcriptions quasi-diplomatiques des *Essais* de 1580 et 1588 ont déjà été réalisées dans leur état brut en double saisie par un prestataire extérieur, et corrigées par comparaison automatique et application d'un correcteur « orthographique » spécifique. La base de transcription d'*EB* est l'exemplaire de 1588 mis en ligne par l'Université de Virginie (Gordon Collection) en attendant d'autres numérisations de qualité (BM de Bordeaux, BnF), pour relever d'éventuelles variantes. La saisie manuelle des parties manuscrites s'impose, en recourant toutefois aux transcriptions de l'édition municipale à titre de témoin. L'affichage optionnel consistera à choisir entre 1588 seul, « 1588-EB », et « 1588-EB-1595 ».

Le principe de la transcription de base s'approche le plus exactement possible d'un exemplaire original, publié en regard. Il ne s'agit pas d'élaborer d'emblée un appareil critique complet, mais de fournir un texte quasi-diplomatique intégral, avant régularisation. Les principes essentiels en sont :

- le respect de la mise en page, du lignage et de la segmentation
- l'absence d'adjonction d'alinéas, de guillemets, de tirets
- l'absence d'intervention sur l'usage des majuscules
- le respect des graphies et de la ponctuation, sans dissimiler ni désabrégier, ni corriger
- l'utilisation de caractères Unicode, MUFI et PICA<sup>1</sup> si nécessaire.

Ensuite, l'application de l'encodage en XML-TEI P. 5 selon le protocole TEI-Renaissance permet de régulariser le texte sans perdre la trace du texte de base :

- la version régularisée est produite *après* la transcription de base grâce à des procédures semi-automatiques de désabrégiation et de dissimulation (pour *il/j* et *u/v*) ; elle s'affiche grâce à

---

1 MUFI : Medieval Unicode Font Initiative. PICA : Projet d'Inventaire des Caractères typographiques Anciens (MSH Val de Loire et MRSH de Caen), visant à inventorier tous les caractères latins utilisés dans les imprimés du xv<sup>e</sup> au xvii<sup>e</sup> siècles, à en normaliser le codage et à faire dessiner une police de caractères libre de droits à destination des chercheurs et éditeurs.

- une feuille de style qui offre les deux versions selon la balise `<choice>` qui pilote les balises `<orig>`, `<reg>`, `<sic>` et `<corr>`
- les corrections de coquilles ou d'erreurs manifestes sont encodées en TEI à l'aide d'une autre édition mentionnée dans la balise, ou de probabilités graphiques
  - l'encodage gère les problèmes de numérotation des pages et des chapitres
  - l'encodage des citations et des langues employées par Montaigne (balises `<quote>` et `<foreign>`) permet de filtrer, de trier et d'extraire les passages rédigés en langues autres que le français, et de constituer un répertoire des citations.

Pour l'édition de 1595 les montaignistes disposent déjà d'une transcription tout à fait estimable (mais désabrégée et dissimulée), offerte depuis une dizaine d'années en ligne sur le site « Trismegiste », dont l'auteur n'a pas pu être identifié. Cette version régularisée servira de base à la diffusion pour le grand public. On y appliquera le balisage des coquilles, celui des noms propres avec leur régularisation selon des thésaurus d'autorités, et un pointeur vers les sources thésauriques et la base prosopographique BUDE développée à l'IRHT (par Marie-Elisabeth Boutrou, section de l'humanisme)<sup>1</sup>.

Faut-il abandonner la balise `<text>` pourtant au cœur de l'encodage du livre pour lui préférer `<document>`, plus adapté aux brouillons qui peuvent prendre des formes complexes ? Après discussions et tests, il est apparu que la mise en page est avant tout générée par celle de l'édition de 1588, le modèle restant le livre avec ses pages. Il a donc été décidé de garder les schémas des « msDesc » (« manuscript description ») de la TEI qui valent pour les manuscrits et imprimés anciens, de façon à éviter de passer par la notion plus générale de document et d'archive. Bien entendu, les différences entre texte imprimé et texte manuscrit seront encodées avec précision, désignant par des « attributs » et des « types » les différents modes d'intervention manuscrite de l'auteur (rature avec la balise `<del>`, surcharge, double surcharge, changement d'encre, etc.). Elles apparaîtront avec des éléments visuels adaptés (couleurs, italiques, etc.) et des indications d'emplacement.

Plusieurs questions restent encore à débattre avec les membres du consortium TEI : faut-il placer l'affichage des additions `<add>` dans

1 BUDE : Base Unique de Documentation Encyclopédique, <http://bude.irht.cnrs.fr>.

ou hors du texte imprimé ? Comment incorporer les additions dans les additions ? Les marques d'insertion et les commentaires peuvent être balisés par <ge :metamark>, de même qu'il sera nécessaire d'historiciser les interventions (étapes, phases, campagnes d'additions-corrrections ou de retouches). Les balises <surface> et <zone> serviront à encoder des données spatiales en ajoutant un @type dans <zone> et peut-être inclure les coordonnées dans <line>. Les retouches de ponctuation demanderont un travail d'orfèvre, mais le résultat en vaudra la peine car elles pourront être distinguées selon leur typologie (point remplacé par virgule, etc.). Le résultat devra être lisible et exploitable, l'utilisateur ayant le choix entre deux types d'affichage, diachronique ou lissé.

La temporalité reconstituée de l'écriture sera marquée par la restitution de différentes interventions manuscrites qui seront notées dans le balisage, de même que <zone> permettra de délimiter la géométrie des additions. Cette transcription à l'identique des parties autographes demande le concours de spécialistes des écritures, de l'écriture de Montaigne, et des pratiques éditoriales de l'époque.

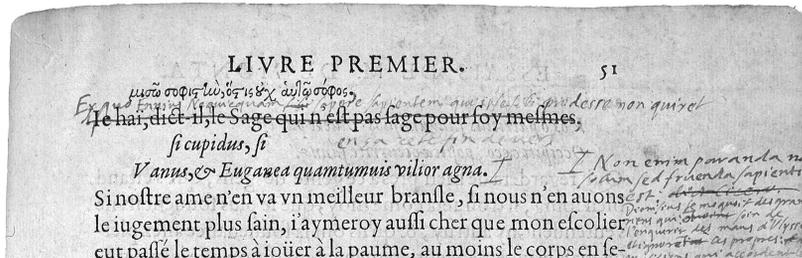
Pendant le premier semestre du projet, les partenaires commenceront par éditer quelques chapitres d'intérêt particulier : I, 21 (« De la force de l'imagination »), I, 56 (« Des prières »), II, 37 (« De la ressemblance des enfants aux pères »), III, 8 (« De l'art de conférer »), III, 9 (« Des coches »), l'idée étant de mettre au point des modèles à la fois stables et évolutifs. À partir de la transcription quasi-diplomatique de 1588 (effectuée) et celle des parties manuscrites (à faire), la désabréviation et la dissimilation ne seront appliquées qu'à la couche ultime du texte régularisé. Le marquage des couches imprimées antérieures (1580-82), préparé pour les éditions de celle-ci, sera offert parallèlement à *EB*, de façon à en offrir l'affichage correspondant à un menu<sup>1</sup>.

Encoder *EB* n'est pas une tâche simple. Observons la transcription de ce haut de page (I, 25, 51r) :

μισῶ σοφιστην', ὅστις οὐχ ἑαυτῷ σοφος.,  
*Ex quo Ennius Ne quicquam filii sapere sapientem qui ipse sibi prodesse non quiret*  
 Je haï, diet-il, le Sage qui n'est pas sage pour soy mesmes.,  
*si cupidus, si en ça cete fin de vers*

1 L'encodage exhaustif des variantes de toutes les éditions dans un même fichier, difficile techniquement à cause des chevauchements de balises, risque de toute manière de rencontrer des obstacles cognitifs de lisibilité.

*Vanus, & Euganea quantumuis vilior agna. I*      *I Non enim paranda nobis  
solum sed fruenda sapientia  
est :*



Michel de Montaigne, *Les Essais*, Paris, A. L'Angelier, 1588,  
Exemplaire dit «de Bordeaux», bibliothèque municipale de Bordeaux,  
folio 51r. Cote S 1238 Rés. C.

Les éditeurs ont depuis longtemps suppléé aux lacunes dues au masicotage malencontreux en recourant à l'édition de 1595, comme dans les fins de ligne en marge droite. Nous ferons de même, en nous servant principalement des balises <gap> et <supplied>, à l'exception des cas où l'on peut raisonnablement avoir un doute et recourir à d'autres restitutions (voir l'édition d'A. Tournon). Ce qui donne par exemple pour « nobis » :

no<supplied source="#ed1595" reason="massicot">bis</supplied>

Naturellement, l'affichage en html montre le mot complet nobis, dont la partie restituée sera d'une autre couleur. Lorsque l'on passe la souris sur le mot, une info-bulle précise « texte original : no », l'encodage complet étant consultable à partir du fichier XML/TEI.

Il faut aussi distinguer ce qui relève de la « méta-écriture » comme les marques d'insertion (ici, après « agna ») et les indications d'emplacement. Alain Legros a relevé ces différents « guidons », selon la terminologie des typographes, utilisés par Montaigne pour en établir une typologie qui devra être intégrée à Unicode.

## LA RECONNAISSANCE DES ÉCRITURES

Le projet comprend un autre volet important : la comparaison de l'écriture de Montaigne avec celles de ses contemporains, à des fins d'authentification et d'étude des « mains » d'humanistes. Dans le cas d'*EB*, l'identification ne fait pas de doute mais l'échantillonnage obtenu à partir de la saisie intégrale des parties manuscrites permettra d'améliorer l'authentification d'autres exemplaires annotés qui ne portent pas la signature de Montaigne. La Reconnaissance Graphique, sous-domaine de l'analyse d'image de document, se focalise sur les arrangements de symboles utilisés dans les domaines très divers du document. On peut essayer de rechercher des « signatures formelles », sans opposer les méthodes globales (statistiques) et locales (spécifiques aux écritures). Certaines études portent sur les caractéristiques locales concernant les degrés, les concavités et les orientations des formes, la distribution des orientations de contour, pour établir la différence entre scripteurs. Les approches fractales ont permis de définir des outils pour les authentifier. Des mesures de complexité ont aussi été introduites, pour caractériser les différents styles indépendamment du contenu du texte et même de l'alphabet utilisé.

L'identification de l'écriture de Montaigne semble facile, mais l'usage montre qu'il n'en est rien : non seulement il faut distinguer entre sa main « française », sa main « latine » ou sa main « grecque » pour les abréviations et les ligatures, mais plusieurs époques doivent être prises en considération. Si aucune machine ne remplace l'intuition et l'analyse de l'expert, et notamment l'acuité d'Alain Legros, elle peut aider à faire le tri dans une banque de données et à proposer des similarités.

Si la numérisation de tous les livres de la « librairie » constitue une étape ultérieure, le fait de pouvoir disposer déjà du « Lucrèce<sup>1</sup> », d'*EB* et des transcriptions d'Alain Legros est un avantage important : le fonds qui associe les transcriptions expertes aux images numériques permettra de constituer une base de graphies « authentiques » datables,

---

1 Exemplaire de la bibliothèque de l'Université de Cambridge, en ligne sur le site des bvh. univ-tours.fr.

indispensable à la collation ultérieure d'autres fragments autographes ou supposés tels, de lettres ou d'ouvrages annotés sans *ex libris*. Même s'il est difficile de parvenir à une certitude, il faut aussi tenir compte des aspects matériels (qualité du papier, encre, plume), et cette base sera un outil de référence.

L'édition « étendue » d'*EB*, c'est aussi l'édition minutieuse de toutes les variantes autographes de Montaigne sur tous les exemplaires de 1588, et les repérages ne sont pas encore terminés. Si deux corrections autographes de la date placée à la fin du « Au lecteur » sont déjà connues (*EB* et exemplaire Lambiotte à Bordeaux), deux autres sont à ajouter : Jean Balsamo en a vérifié une sur un exemplaire privé, et j'ai pu dernièrement signaler celle de l'un des exemplaires conservés à la Bibliothèque Royale de Bruxelles<sup>1</sup>. Toutes les quatre sont différentes et il est possible de conjecturer l'ordre dans lequel Montaigne a commencé cette besogne ingrate de corriger tous les exemplaires, avant de s'arrêter sans doute assez vite<sup>2</sup>. Le résultat de ces conjectures pourra être intégré à l'encodage du texte en dépliant les différentes versions, jusqu'à ce que d'autres chercheurs le contestent ou le complètent.

Marie-Luce DEMONET  
 Université François-Rabelais,  
 Centre d'études supérieures de la  
 Renaissance, Tours,  
 Institut universitaire de France  
 Avec la collaboration  
 d'Alain LEGROS  
 et Marie-Elisabeth BOUTROUE

1 Bibliothèque Royale de Bruxelles, Fonds Solvay FS IX 80. 6628.

2 Voir l'article à paraître (par Alain Legros et Marie-Luce Demonet), « Montaigne à sa plume : quatre variantes autographes d'une correction de date dans l'avis "Au lecteur" des *Essais* de 1588 ».