# L'édition critique multimodale de sources anciennes

## Une recherche collaborative pour la création de nouveaux outils

Marie-Agnès LUCAS-AVENEL(1), Marie BISSON(2)

(1) Centre Michel de Boüard, CRAHAM, UMR 6273 Université de Caen Normandie, France – CNRS, (2) Maison de la recherche en sciences humaines USR 3486 Université de Caen Normandie, France – CNRS

La situation des études visant à l'édition scientifique numérique des sources anciennes est paradoxale : l'accès en ligne aux textes anciens grâce à la numérisation des collections et des témoins manuscrits ainsi que leur insertion dans des bases de données permettent de démultiplier les modes d'interrogations et les recherches pluridisciplinaires sur ces sources. Pourtant, rares encore sont les projets qui aboutissent à de nouvelles éditions critiques numériques de sources littéraires anciennes et médiévales[1]. Parmi les diverses raisons, on peut invoquer l'aspect chronophage de ce type de recherches (Pierazzo, 2014), qui s'accommode mal d'une programmation de la recherche sur des contrats courts. Néanmoins, la collaboration qui a été mise en place à Caen entre le CRAHAM (UMR 6273), le pôle Document numérique de la Maison de la Recherche en Sciences humaines (USR 3486) de l'Université de Caen Normandie – en lien avec les *consortia* d'Huma-Num et l'Equipex Biblissima, – et les Presses universitaires de Caen (PUC) a permis de mener une réflexion sur le choix et l'utilisation des outils et de proposer de nouvelles pratiques. Celle-ci a abouti à la création d'environnements numériques en XML-TEI outillant la recherche pour la réalisation d'éditions de sources, avec des retombées importantes aussi bien pour l'éditeur scientifique que pour les

[1] Pour les éditions de textes latins, l'un des précurseurs a été Bruno Bureau et le projet « Hyperdonat ». Ajoutons aussi l'édition critique multimodale des *Chroniques latines du Mont Saint-Michel (IX^e^-XII^e^ siècle)*, éd. Pierre Bouet et Olivier Desbordes, Caen, PUC, 2009 https://www.unicaen.fr/services/puc/sources/chroniqueslatines/accueil, et l'édition critique numérique de Cicéron, *De natura deorum,* éd. C. Auvray-Assayas, Caen, PUC, 2019
https://www.unicaen.fr/puc/sources/ciceron/accueil (consultés le 19 juillet 2021).

lecteurs (chercheurs et autres). Dans le cadre de cette collaboration, on mène une recherche à la fois large et innovante appliquée à une source ou à un corpus de sources, sur la manière d'outiller la recherche, selon la « démarche » proposée par Pierre-Yves Buard (Buard, 2015 : p. 26) dans sa définition des humanités numériques[2].

Concernant l'édition critique[3], en particulier celle des textes anciens et médiévaux, il nous semble, en effet, que la réflexion sur le choix et l'utilisation des outils numériques devient partie prenante de la recherche elle-même, dans une approche à la fois méthodologique et théorique, suivant une progression empirique et heuristique[4]. De nombreux chercheurs en sciences humaines – peut-être surtout les éditeurs de sources[5] – se sont interrogés à la fois sur les meilleures méthodes à adopter avec le développement des outils informatiques et sur la finalité du travail éditorial, en s'intéressant en particulier au rapport du lecteur au document et aux modes d'accès aux données. Cette réflexion méthodologique a été facilitée par la mise en place des *consortia* et des programmes scientifiques mentionnés ci-dessus et accompagnée d'une réflexion théorique, donnant lieu à un important travail de définition ou de redéfinition de l'édition « savante » ou « critique » (« scholarly edition »), selon un rapport nouveau de la triade chercheur, éditeur, lecteur au(x) document(s) ou au(x) texte(s)[6]. Un rapide rappel des débats qui divisent les chercheurs sur la préférence à accorder à l'édition critique ou à l'édition documentaire, quand il s'agit de faire une édition numérique, nous amènera à montrer d'abord que faire le choix de l'édition critique numérique incite à défendre la complémentarité des deux types d'édition. Cependant la réalisation d'une édition critique et documentaire numérique, qui doit garder des objectifs scientifiques clairs, n'est possible que dans le cadre d'un partenariat interdisciplinaire que nous présenterons ensuite. Enfin, nous montrerons comment, à côté des avantages nombreux pour le lecteur de l'édition multi-support, le travail d'encodage exerce aussi une influence sur la recherche qui conduit à la production de l'édition critique. On exposera ainsi certains des résultats obtenus dans le domaine de l'ecdotique outillée par ordinateur : on montrera combien la création d'un environnement XML dédié facilite et rend plus sûr le travail qui consiste à établir le texte et son apparat critique et multiplie les possibilités de recherche sur la source et ses témoins à la fois pendant le travail de préparation à la publication, ainsi que pour et après la publication.

[2] « Nous avançons donc que les humanités numériques sont à considérer comme une démarche qui se propose d'outiller informatiquement la recherche en sciences humaines et sociales et d'en diffuser les résultats en prenant soin, dans un mouvement de distanciation, d'examiner comment les outils mis en place et développés sont susceptibles d'influencer le travail de recherche lui-même ».

[3] On pourra se reporter aux travaux d'Apollon, Régnier et Bélisle, 2017. Bien que les textes anciens et médiévaux soient largement pris en considération dans l'ouvrage, on n'y trouvera que très peu d'exemples d'éditions critiques numériques de textes en langues anciennes.

[4] Le pôle Document numérique de la MRSH de Caen s'est doté d'une double mission d'ingénierie de service et de recherche, qui lui permet de répondre au mieux aux besoins des chercheurs en sciences humaines et sociales (SHS) en étant force de propositions au sein de programmes nationaux et internationaux http://www.unicaen.fr/recherche/mrsh/document_numerique.

[5] On utilisera ici le terme « éditeur » au sens d'auteur d'éditions critiques de sources anciennes, mais la remarque vaut aussi – et sans doute *a fortiori* – pour les éditeurs matériels. Bien que le travail de l'un et de l'autre soit distinct, il ne peut être mené à son terme que de manière collaborative et en partenariat aussi avec l'archiviste (ou le bibliothécaire-conservateur) et le développeur. Voir Buard, 2015, en particulier p. 19-21.

[6] On trouvera des définitions des divers types d'éditions dans le volume dirigé par Apollon, Régnier et Bélisle, 2017, en particulier p. 106-107 et le glossaire (p. 351-372) associé aux sources qui ont permis de le constituer. On pourra se reporter aussi à Sahle, 2016.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

On s'appuiera sur l'édition critique de l'*Histoire du Grand Comte Roger et de son frère Robert Guiscard* par Geoffroi Malaterra, composée en latin vers 1100. L'œuvre est une source majeure sur les membres de la famille de Hauteville, partis de Normandie pour s'établir et conquérir l'Italie du Sud entre 1030 et 1098. Le texte latin, qui est divisé en quatre livres et rédigé à la fois en prose et en vers, a été établi grâce à la collation exhaustive des manuscrits primaires (A, B – ou Be, copie de B, quand ce dernier fait défaut –, C, Z, conservés dans diverses bibliothèques de Sicile, Besançon et Barcelone), et deux éditions qui ont marqué son histoire (la *princeps* de Geronimo Zurita [1578] et celle d'Ernesto Pontieri [1927-28]). On a aussi tenu compte des annotations de l'édition de Caruso (Ca) [1723], qui, reproduisant la *princeps*, a relevé des variantes de A issues de feuillets aujourd'hui perdus ; de même, l'exemplaire de la princeps conservé à la bibliothèque universitaire Giacomo Longo de Messine (Me) comporte des marginales du XVIII[e] siècle, signalant des leçons de A. Suite à la redécouverte du manuscrit Z, Gianvito Resta avait montré les manques de l'édition de Pontieri et la nécessité de réaliser à nouveaux frais une édition scientifique (Resta, 1964)[7].

## 1 Édition documentaire ou édition critique numérique ?

Les débats sur la définition et la finalité de l'édition critique numérique ont été marqués par les travaux de Peter Robinson (Robinson, 2002 et Robinson, 2013), en dialogue avec Eggert, 2009 ou Gabler, 2011. La question centrale de ces débats, basée sur les liens entre « work » au sens d'« œuvre[8] », « text » (ou « version », Haugen 2017 : p. 238) et « document » (ou « artefact » : Driscoll, 2010), est celle de la finalité même de l'édition savante numérique : doit-elle, comme le propose Gabler, se concentrer sur l'édition du/des document(s), porteur(s) matériels du texte et témoins de l'œuvre, ou bien sur celle du texte[9] ? On s'accorde aujourd'hui à reconnaître que le document a été trop longtemps laissé pour compte par les auteurs d'éditions critiques. Dans la même veine, Matthew J. Driscoll insiste sur les apports de la « nouvelle philologie » (Driscoll, 2010)[10] et de la réflexion de Bernard Cerquiligni, *Éloge de la variante* (Cerquiligni, 1989), qui souligne l'importance de considérer le caractère mouvant ou instable des textes manuscrits, dont une édition électronique peut rendre compte en donnant accès aux fac-similés de tous les documents. À ces remarques, il faut encore ajouter les réflexions des codicologues et les travaux pionniers de Peter Stokes sur l'*Exon Domesday book*, en matière de codicologie numérique (Stokes et Noël, 2019)[11]. Frédéric Duval a déjà relevé cette différence entre les écoles anglo-saxonne et française (Duval, 2017 : p. 14)[12]. Revenant sur le « triomphe des éditions orientées vers le document »,

[7] Sur l'histoire du texte et les principes méthodologiques suivis pour l'édition critique, voir Geoffroi Malaterra, *Histoire du grand comte Roger…*, Lucas-Avenel (éd.), 2016, p. 89-108.

[8] Pour une définition de ces termes dans des travaux français, Buard, 2015 : p. 77-81 et Duval, 2017 : p. 15-16.

[9] Sur les conséquences de la préférence accordée au document ou à son contenu, en matière d'édition matérielle, pour le choix du modèle de représentation, voir Buard, 2015 : p. 15.

[10] Voir aussi la *Queste del saint Graal*, Christiane Marchello-Nizia et Alexei Lavrentiev (éd.), Lyon, ENS de Lyon, 2019. Publié en ligne par la *Base de français médiéval*, http://catalog.bfm-corpus.org/qgraal_cm, consulté le 19 juillet 2021, « Introduction », p. 7-8.

[11] Voir aussi le site dédié https://www.exondomesday.ac.uk

[12] « Cette terminologie naissante [i. e. « digital scholarly edition »] est déconnectée des traditions ecdotiques française ou néo-lachmanienne, largement ignorées, ce qui rend les débats parfois compliqués à suivre ».

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

il invite les éditeurs à la vigilance vis-à-vis de ce qui pourrait devenir une pratique hégémonique, dont le succès est lié à la « coïncidence des nouvelles possibilités techniques offertes par le numérique avec des théories du texte à la mode, contestant le bien-fondé des éditions idéalistes et reconstructionnistes », « au motif », ajoute-t-il, « qu'elles donnent à lire un texte qui n'a jamais existé » (Duval, 2017 : p. 16-17). Ces théories, explique Frédéric Duval, défendent un « relativisme qui rend caduque la notion d'autorité dont jouirait une version du texte par rapport aux autres », refusent de « privilégier un état textuel (*text*) plutôt qu'un autre » et s'intéressent dès lors aux documents que l'édition papier n'a pas la possibilité de reproduire, dans les limites qui sont les siennes – contrainte d'éditer un seul état textuel (*single text editions, one text editions)* –, et dont peut au contraire s'affranchir l'édition numérique.

Si ces considérations théoriques, rappelées par Frédéric Duval, expliquent en partie le peu de succès des éditions critiques numériques basées sur de multiples témoins, il faut en avancer d'autres qui tiennent plus à la pratique et qui ont été relevées par Elena Pierazzo : il s'agit d'abord du temps très long et de l'investissement important, nécessaires à la mise en œuvre du travail, depuis la transcription, la collation et l'analyse des témoins jusqu'au développement des nombreux outils pour la publication Web[13] ; en outre, se reportant aux travaux de Tara Andrews, elle souligne la difficulté de concevoir des éditions critiques électroniques en l'absence de standards et de consensus entre les communautés de chercheurs (Pierazzo, 2014). Cependant, les travaux sur la *Chronique* de Matthieu d'Édesse (Andrews, 2013) invitent à ne pas opposer les éditions numériques critiques et documentaires et à continuer dans le sens de la complémentarité permise par le numérique. C'est d'ailleurs aussi l'idée qu'a défendue Peter Robinson en 2013 : il refuse l'idée d'une hiérarchie entre *text* et *document* et montre au contraire qu'il serait vain de chercher à comprendre l'un sans étudier l'autre. Frédéric Duval (Duval, 2017 : p. 25), reprenant ces conclusions auxquelles il adhère, cite Peter Robinson (Robinson, 2013 : p. 123)[14] en le traduisant :

> On ne peut pas connaître l'œuvre sans les documents – de même, on ne peut comprendre les documents sans une compréhension de l'œuvre qu'ils génèrent[15]. On peut en tirer le principe suivant : une édition scientifique doit éclairer autant que possible chaque aspect du texte, aussi bien le texte-comme-œuvre que le texte-comme-document. Les éditions imprimées traditionnelles se sont davantage intéressées au premier. Un avantage évident des éditions numériques est qu'elles peuvent rééquilibrer le balancier en incluant un matériau bien plus riche pour l'étude du texte-comme-document qu'on ne peut y parvenir par le moyen de l'impression.

[13] « Without large investment in developing tools that can support the editorial work, it seems unlikely that this type of edition will take off on the Web anytime soon ».

[14] « One cannot know the work without the documents – equally, one cannot understand the documents without a comprehension of the work they instance. From this, a principle appears : a scholarly edition must, so far as it can, illuminate both aspects of the text, both text-as-work and text-as-document. Traditional print editions have focused more on the first. An evident advantage of digital editions is that they might redress this balance, by including much richer materials for the study of text-as-document than can be achieved in the print medium ». L'inclusion d'une édition critique au sein de l'édition documentaire a également été défendue par Tanselle, 1995 : p. 592, puis par Shillingburgh, 1996 : p. 95. Voir aussi Vanhoutte, 2010 : p. 28.

[15] Frédéric Duval donne peut-être une force excessive au verbe « instance » employé par Robinson. Le sens ne serait-il pas plutôt « [l'œuvre qu'ils] présentent » ? Cependant, la traduction de Frédéric Duval souligne l'importance des documents, surtout quand on dispose de ceux des auteurs.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

Désormais, en effet, quel éditeur de texte songerait à déplorer les potentialités nouvelles offertes par le numérique pour l'étude du document et sa mise à disposition libre et gratuite pour le plus grand nombre ? En donnant accès aux sources utilisées, accompagnées d'un commentaire explicitant le contexte de leur composition et d'une restitution facilitant leur lecture, on permet aux lecteurs de mieux percevoir le travail mené sur les sources, de vérifier et de critiquer le résultat publié. La recherche qui aboutit à une édition critique dans un contexte numérique n'empêche pas celle qui doit être menée sur le document, mais en facilite au contraire le développement dans une démarche visant à la complémentarité plutôt qu'à l'instauration d'une hiérarchie ou d'une primauté de l'édition critique vis-à-vis de l'édition documentaire. Ainsi, comme l'a affirmé Philippe Régnier, « puisque la possibilité matérielle et technique en existe », il devient hautement souhaitable désormais de publier le texte édité tout comme ses sources annotées (Régnier, 2017 : p. 95-96)[16]. Il s'agit donc non pas de demander au lecteur de faire lui-même l'édition critique du texte, mais de lui permettre de circuler en étant guidé à partir du texte édité vers les textes documentaires. Associer ces différentes représentations du texte, c'est ainsi donner au lecteur les clés qui lui permettent de comprendre la démarche de l'éditeur scientifique, voire de la critiquer, et non de l'inviter à faire un travail long et difficile. Il s'agit aussi de ne pas induire chez le lecteur l'idée que le texte qui lui est donné à lire est « pur », qu'il n'a pas souffert des aléas de la transmission ; au contraire, on veut redonner toute leur place aux résultats de la recherche pour la rédaction de l'apparat critique, qui est marginalisé dans les éditions papier, comme l'a noté Elena Pierazzo (Pierazzo, 2014)[17]. La lecture de l'apparat, dans les éditions traditionnelles, est réservée à un petit groupe d'initiés. Au contraire, si l'apparat critique est associé aux documents, il devient, plus encore peut-être que le texte édité, une donnée essentielle qui facilite la circulation dans tous les documents, tandis que réciproquement la lecture des restitutions de chaque témoin peut permettre la compréhension de l'apparat pour les non-initiés.

Dès lors, il ne nous semble pas que l'apparition du numérique doive amener à débattre pour redéfinir la finalité de l'édition critique d'une manière théorique, car la réflexion a déjà été menée depuis longtemps – on pense bien sûr aux travaux de Lachmann, Gaston Paris, Bédier et à ceux de leurs successeurs. L'objectif qui consiste à combiner des éditions critiques et des éditions documentaires grâce aux nouvelles technologies n'implique pas de remettre en question la finalité de l'édition critique. « Le changement de médium n'altèr[e] en rien l'ontologie des textes (Duval, 2017 : p. 19-20[18] ».

[16] « Comment, puisque la possibilité matérielle et technique en existe, procurer l'état restauré que l'on estime le plus probable, sans offrir, aussi, les différents états qui valent par leur existence, et, inversement, comment exhiber et légitimer les états évolutifs sans proposer, aussi, l'hypothétique état matriciel auquel les référer ? Autant sa communauté scientifique et la société attendent du philologue qu'il émette un avis tranché sur le texte et sur ses états, autant elles comprendront de moins en moins – puisque, encore une fois, il n'est plus limité à un livre ou à une série de livres – qu'il leur en impose une série unilatérale et leur refuse un accès direct et intégral à tous les arguments du débat. Faire entrer la littérature du Moyen Âge dans l'ère d'une reproductibilité indéfinie et universelle n'est plus une utopie ».

[17] « Such a format has generated a fair amount of criticism because it gives texts a false impression of stability and purity, and makes it almost impossible to really appreciate the editorial work and the complexity of the tradition ». Voir aussi Shillingburg, 1996 : p. 118.

[18] Les propos de Duval, 2017 : p. 19, sont tout à fait justes : « G. Thomas Tanselle a rappelé avec raison que le changement de médium n'altérait en rien l'ontologie des textes. La question fondamentale demeure le choix entre l'orientation vers l'auteur (ou toute autre autorité) et l'orientation vers le document. De manière secondaire s'y adjoint la question de savoir si le médium (papier ou numérique) doit ou non intervenir dans la théorisation du texte. Pour ma part, il est clair que, si le numérique doit modifier les méthodes d'édition,

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

Le bouleversement, en revanche, concerne la manière dont on accède au texte et dont on le lit, ce qu'a parfaitement rappelé aussi Frédéric Duval en posant la question de la « textualité de l'édition numérique » : « a-t-on encore affaire à un texte ou bien plutôt à un hypertexte ? ». Le numérique offre au chercheur de multiples parcours de lecture, qui viennent compléter le dialogue entre l'auteur et l'œuvre que le livre assure d'une manière différente. Sans doute celle-ci est-elle plus proche de celle qu'a voulue l'auteur, surtout quand on considère les sources littéraires anciennes et médiévales, dans lesquelles les systèmes hypertextuels sont encore peu développés[19]. C'est pourquoi, la réflexion théorique, nous semble-t-il, doit porter non pas sur ce qu'on entend par « texte édité », mais sur la manière dont on veut en assurer la lisibilité et même, en amont de la publication, sur les potentialités du numérique pour faciliter le travail scientifique. Il s'agit donc d'une recherche empirique qui porte sur les moyens et la méthode à employer pour établir le texte et sur la manière de présenter au mieux les résultats dans un contexte numérique, sans pour autant renoncer aux avantages de la publication papier (commodité de lecture, stabilité du texte, pérennité, citabilité, etc.) ; en particulier, la question est de savoir si on peut recommander de bonnes pratiques pour de futurs travaux, tout en autorisant une certaine flexibilité qui permette de prendre en compte les spécificités des textes qu'on a choisi d'éditer et qui préserve la liberté et la créativité du chercheur dans ses choix d'analyse[20].

## 2 L'édition de Geoffroi Malaterra : contexte et objectifs

Les réflexions que nous avons menées sur l'édition savante numérique se sont appuyées sur les travaux réalisés pour l'édition critique de l'œuvre de Malaterra, mais ont largement bénéficié des résultats obtenus auparavant pour l'édition critique double support du *De piscibus*, livre IV de l'*Hortus sanitatis* à partir de données encodées en XML-TEI[21] et d'autres projets du pôle Document numérique de Caen. Ces projets présentent des spécificités, mais ils ont aussi beaucoup d'éléments communs, qui ne sont pas seulement d'ordre méthodologique : les contenus scientifiques se croisent parfois suffisamment pour qu'on mesure très concrètement l'utilité d'assurer l'interopérabilité des données grâce à l'utilisation des standards de la TEI sur lesquels s'accordent désormais les chercheurs en SHS.

---

voire la conception de l'édition, il ne doit pas modifier la conception du texte édité, qui n'a rien à voir avec elle. À dire vrai, les éditeurs ont toujours eu tendance à sous-estimer le fait qu'ils devaient articuler deux textualités différentes, celle du texte édité et celle de leur édition. Le numérique implique évidemment une nouvelle textualité de l'édition, mais ne modifie pas celle du texte édité ».

[19] Cette remarque doit être modérée en fonction du type de texte ou de document – on ne lit pas un récit comme une encyclopédie – et par le développement de l'hypertexte déjà chez les Anciens et au Moyen Âge.

[20] Cette flexibilité a été réclamée par Vanhoutte (2010 : p. 3), qui considère que la construction d'une interface standardisée pour les éditions électroniques, *a priori* commode pour l'utilisateur, serait absurde du point de vue théorique. Il plaide plutôt pour un assemblage de moyens par lesquels l'utilisateur peut interagir avec le texte, le commentaire et le matériel auxiliaire. Il faut, selon lui, que l'interface puisse continuer de dépendre des spécificités de la source et de l'interprétation de l'éditeur et ne pas imposer une interface générale qui risquerait de réduire l'édition électronique à la publication d'un outil offrant un ensemble de fonctions fixes plutôt qu'un outil modèle pour explorer le texte et générer du sens.

[21] Hortus sanitatis *: Livre IV, Les Poissons*, C. Jacquemard, B. Gauvin, M.-A. Lucas-Avenel (éd.), avec la collaboration de C. Février et F. Lecocq, Caen, PUC (Fontes & Paginae) 2013 (https://www.unicaen.fr/puc/sources/depiscibus/accueil, consulté le 19 juillet 2021). L'édition du *De Piscibus* est la première réalisation du programme Ichtya
https://www.craham.cnrs.fr/projet/programmes/ichtya/, consulté le 19 juillet 2021.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

### 2.1 L'édition critique est un produit fini, citable et référençable

Les recherches menées dans le cadre du « projet Malaterra » ont pour finalité d'aboutir à l'édition critique de l'ensemble de l'œuvre. Cependant, elles comprennent deux étapes, dont l'une a été achevée en 2016 et l'autre est en cours de finalisation. La première correspond à la publication de l'édition critique des deux premiers livres. Cette partie du travail n'était pas nativement numérique, contrairement à la seconde, laquelle vise à la publication des deux autres livres. Dès lors que le projet est devenu numérique, il a pris une dimension collaborative associant à l'étude philologique, littéraire et historique une réflexion portant sur la structuration des données dans le respect des standards internationaux et sur les fonctionnalités offertes par les outils numériques pour la valorisation de ces données dans le cadre de la publication. Le programme a été conçu pour offrir une édition critique multi-support (papier et Web) et multimodale, donnant à lire le texte latin, édité suite à l'analyse de ses témoins, et accompagné d'un apparat critique aussi exhaustif que nécessaire, rédigé selon les recommandations de la CUF. Le texte latin est accompagné de notes philologiques, de notes de sources et d'une traduction française, elle-même dotée de notes de commentaires. Le tout est complété par une bibliographie et des index, qui apparaissent sur les deux publications papier et numérique.

En revanche, tandis que l'introduction développée présentant l'auteur, l'œuvre et l'histoire du texte est réservée à la publication papier, la publication en ligne prévoit les fonctionnalités suivantes : choix de différentes visualisations (texte latin et/ou traduction française), développement ou non des notes d'apparat critique, lecture des notes en pop-up ou en bas de page, recherche plein texte, hyperliens permettant une navigation interne à la publication et externe vers les institutions partenaires et assurant l'interactivité de la bibliographie (etc.) ; en outre, – et c'était la principale innovation par rapport au travail réalisé pour l'édition du *De piscibus* –, elle donne à lire plusieurs représentations du texte latin : non seulement le texte édité et les fac-similés des manuscrits et de la princeps, mais aussi des restitutions du texte de chaque témoin cité dans l'apparat critique, réalisées à partir de l'encodage fin de cet apparat. Toutes ces fonctionnalités, dont on détaillera plus loin la nature et les limites, font de la publication en ligne une édition critique qui répond aux exigences de l'école française traditionnelle et qui est enrichie d'une édition documentaire. Le résultat de cette recherche, qui a donné lieu, pour les livres I et II, à une publication papier et à une publication numérique, quasi simultanées en 2016, est un produit fini, citable et référençable[22].

### 2.2 Nature de la collaboration

Il convient de préciser la nature de la collaboration entre éditeur scientifique, ingénieur, conservateur et éditeur matériel, car il s'agit là d'un aspect essentiel pour la recherche actuelle et future en ecdotique. Le rôle dévolu aux institutions de conservation est notamment la numérisation des ouvrages qu'elles conservent. Soit elles assurent la mise en ligne de leurs images sous forme de liens pérennes, voire, dans le meilleur des cas, en respectant le protocole IIIF pour les rendre facilement accessibles

[22] L'importance de la citabilité et de la stabilité de l'édition critique est aussi soulignée par Buard (2015 : p. 72), qui rappelle en outre les avantages de la forme papier ; voir aussi, *ibid.*, p. 98.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

et consultables ; soit – et c'est le cas de l'édition numérique de l'*Histoire du grand comte Roger* – il s'agit pour l'éditeur d'obtenir les droits de mise en ligne des fac-similés des documents sources, dans le but de les valoriser – si possible par un libre accès –, aux côtés des résultats obtenus dans le cadre de la recherche. Le choix de l'éditeur matériel est également fondamental, puisque ce dernier doit être capable de traiter les fichiers XML-TEI pour la publication[23]. Quant à la collaboration avec les ingénieurs, elle accompagne la réflexion ecdotique dès le début des travaux, afin que le chercheur puisse livrer à l'éditeur matériel des fichiers XML-TEI validement encodés. Il revient au chercheur de déterminer l'encodage qu'il désire effectuer et le degré de granularité auquel il souhaite parvenir, en cernant précisément ses besoins en matière d'édition et de recherche. Il est accompagné dans cette réflexion par un ou des ingénieurs en humanités numériques, spécialistes des sources anciennes, de la conception d'environnements XML adaptés aux choix d'encodage, et de la création de sites Web pour visualiser les données. La collaboration pour le développement des programmes de recherche dans le domaine de l'édition savante numérique est en réalité un binôme dans lequel l'un et l'autre mènent ensemble une recherche expérimentale appliquée à un corpus, tout en réfléchissant aux impacts des outils pour la recherche.

Ce sont ces apports à la recherche que nous voudrions à présent montrer en prenant en considération non seulement le produit fini, mais aussi la recherche « en train de se faire » et les expérimentations, caractéristiques de la recherche en humanités numériques, auxquelles peut donner lieu la conception de nouveaux outils.

## 3 L'encodage pour la recherche

L'intérêt de la publication multi-support a déjà été noté : l'édition numérique peut démultiplier le nombre de lecteurs en associant aux données publiées sur le papier les fac-similés des documents et en ajoutant de multiples fonctionnalités (hyperliens, recherches plein texte, choix de parcours de lecture, etc.) qui facilitent la lecture de consultation, complémentaire de la lecture immersive. Parmi ces fonctionnalités, beaucoup sont communes à tous les projets ; certaines sont aussi spécifiques et dépendent des objectifs scientifiques du chercheur, qui sont liés à la nature même de la source. La standardisation et l'étiquetage des données indispensables à leur encodage exigent une rigueur dans leur appellation que le travail traditionnel ne requiert pas, mais qui accroît la lisibilité au sein d'un même programme et l'interopérabilité entre deux programmes complémentaires. Par exemple, la typologie des notes contraint à séparer les informations pour mieux les classer et ainsi mieux les retrouver. L'encodage impacte donc la méthode de travail et les résultats de la recherche.

L'encodage en XML, selon les recommandations de la TEI, offre une série de modules qui définit la structure des textes et une série d'autres modules adaptables à la nature de la source et permettant de décrire et de structurer des données spécifiques avec la finesse suffisante pour en assurer l'extraction puis la publication sur les différents

[23] Cet enjeu a été relevé par Pierazzo, 2014, pour l'édition de Proust ; sur celle-ci, voir André et Pierazzo, 2013. La collection *Fontes & paginae* créée par les PUC pour l'édition des sources, s'appuie sur le principe du *single source publishing*, un flux de textes encodé en XML-TEI, à partir duquel sont produites les éditions multi-supports. Voir Buard, 2015, et l'Infrastructure de recherche Métopes (dir. D. Roux http://www.metopes.fr/index.html, consulté le 19 juillet 2021).

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

supports. Comme toutes les notes sont encodées selon leur type (notes de source, notes philologiques, etc.), elles peuvent être extraites et accompagner le texte à une place dédiée sur le papier ou sur le Web, ce qui permet au lecteur de s'y retrouver facilement et de trier les informations qui l'intéressent. Cette présentation raisonnée de toutes les données structurées sur le volume papier, malgré la complexité de l'annotation, et sur le site Internet dédié, dont l'affichage respecte les choix de l'édition papier, est un résultat particulièrement innovant dans le domaine de l'édition de sources anciennes.

Le travail d'encodage de l'*Histoire* de Malaterra propose un encodage fin de l'apparat critique ; ce travail a été mené à partir de l'étude comparée des témoins pour établir les principes méthodologiques servant de référence au choix des variantes. Il amène à classer les témoins – en particulier les manuscrits – au sein d'une arborescence qui en définit les liens et d'une hiérarchie qui en précise la valeur. Chaque manuscrit a donc été considéré à la fois comme une source indispensable à la restitution du texte de l'archétype, dont la place dans la tradition textuelle et la valeur de témoignage doivent apparaître clairement, et comme un état du texte corrompu, plus ou moins fiable, et nécessitant de subir la critique.

La rédaction d'un apparat critique traditionnel est largement contrainte par des règles très strictes, qui, pour être claires, se veulent le plus souvent minimales, si bien que, dans la très grande majorité des cas, elles sont réductibles à une forme de code indiquant des oppositions simples au niveau du mot (lemme *vs* variante), et ainsi parfaitement adaptées à un système d'encodage. Cet encodage a été fait grâce à l'utilisation du module « textcrit : Critical Apparatus » de la TEI, dans sa version P5, en suivant la méthode de la segmentation parallèle. En effet, cette méthode, qui permet la restitution du texte de chaque témoin, ne nécessite pas de choisir au départ un texte de base : en étant associée à des feuilles de style CSS et de transformations XSLT, elle permet de rendre compte du texte des témoins, de sa valeur et de sa relation aux autres ; la définition du lemme (<lem>) et des variantes (<rdg>), dans le respect des règles de base, ne crée pas de hiérarchie entre les témoins mais les associe dans un même flux textuel par un jeu d'oppositions ou d'identités qui apparaissent clairement à l'écran. Le statut ontologique de chacun des témoins primaires est donc le même ; ils conservent ainsi leur place au centre de la réflexion ecdotique et peuvent faire l'objet d'un questionnement spécifique.

L'environnement utilisé pour l'encodage des données des livres III et IV a bénéficié de la création de nouveaux outils de rédaction et d'analyse. Il s'agissait, en effet, de concevoir un outil qui (a) facilite l'intégration des données et réduise les risques d'erreurs liées à la rédaction de l'apparat critique au moment de l'enregistrement du témoin et de sa variante ; (b) améliore les conditions de travail lors de la réflexion menée pour établir le texte ; (c) permette de générer automatiquement la rédaction de l'apparat critique, tel qu'il doit apparaître sur le volume papier. L'environnement – comprenant le schéma – est diffusé sur le site Web[24] accompagné de sa documentation (Bisson, Cannet et Lucas-Avenel, 2020).

[24] https://www.unicaen.fr/recherche/mrsh/document_numerique/outils/apparat, consulté le 19 juillet 2021.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

### 3.1 L'environnement d'encodage dans XML Mind XML Editor

L'environnement permet d'avoir différentes vues du texte encodé, qui sont autant de restitutions de ce texte et qui sont juxtaposées à la vue arborescente.

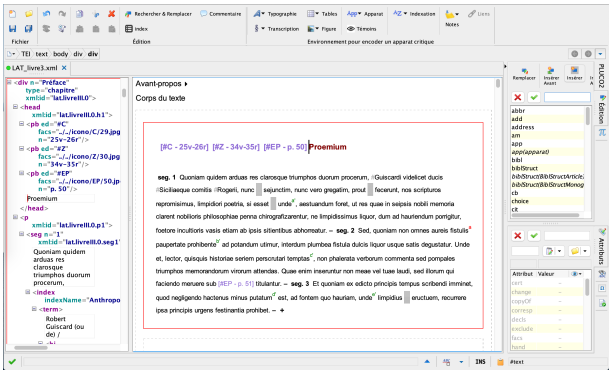



Fig. 1 : Malaterra, III, *Proemium*. Vue arborescente et vue « Transcrire le texte »

Sur la figure 1, la vue arborescente apparaît à gauche, la vue « Transcrire le texte », au centre. Celle-ci donne à voir le texte latin établi avec l'appel des notes de source en rouge (a) et philologiques en vert (a', b', c'). Un onglet « Vue » permet de faire apparaître d'autres présentations du texte, réalisées au moyen de feuilles de styles CSS. Le Pôle Document numérique a conçu quatre vues permanentes pour les étapes de transcription, annotation et indexation du texte. L'environnement d'encodage permet de générer ensuite autant de vues que de témoins déclarés dans les métadonnées du fichier XML (l'outil s'appuie sur l'attribut @xml:id des éléments <witness>).

L'encodage des notes est facilité par l'ajout d'une icône « notes » permettant de sélectionner trois types de notes pour le flux latin (apparat, sources, philologie) ou le type de note commentaire pour le flux français. Les notes de travail ne sont pas destinées à la publication mais permettent à l'éditeur de commenter un travail en cours.

Pour le travail de collation et d'encodage des variantes, le fait de disposer des facsimilés numérisés des manuscrits constitue déjà une véritable révolution. Même si un tel confort ne remplace pas la lecture directe du témoin, elle permet un retour constant à la source jusqu'à l'étape ultime de la publication. Quiconque a collationné des manuscrits sait combien ce travail demande de rigueur dans la lecture, puis dans la notation de la variante et de son témoin. Chaque moment du report d'une variante peut donner lieu à une erreur, depuis le travail en bibliothèque à côté du manuscrit, jusqu'à sa fixation dans la publication. Comme la lecture d'un nouveau témoin occasionne la découverte de nouvelles variantes, qui impose un retour aux témoins lus précédemment et une vérification de ces témoins, la mise à disposition d'une bibliothèque numérique des sources, dématérialisée et délocalisée est une aide considérable.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

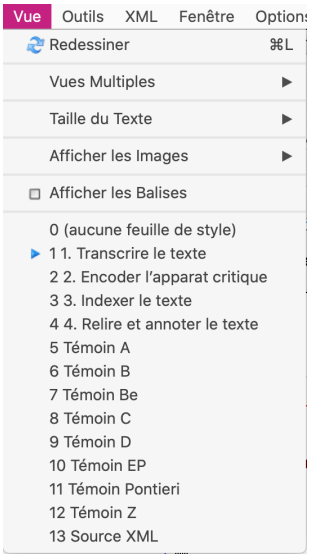



FIG. 2 : Ensemble des vues disponibles pour l'encodage du texte de Malaterra

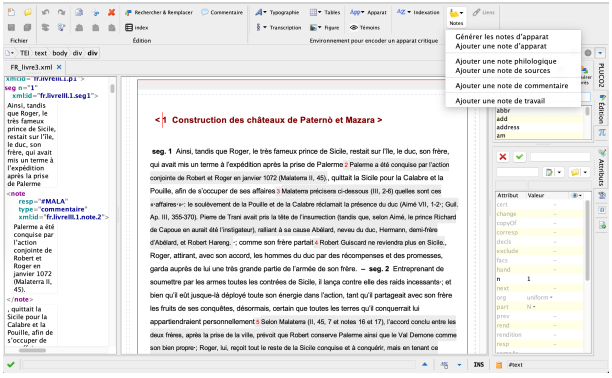



FIG. 3 : Malaterra, III, 1 (traduction), vue « Relecture et annotation »

L'encodage du texte latin et de ses variantes se fait au fur et à mesure de la collation des témoins sur la vue « Encoder l'apparat critique » : pour encoder, par exemple, le manuscrit C (qui, en l'absence de A, est le témoin le plus fiable), on part du texte de Pontieri, déjà numérisé et qui constituait une première transcription ; à la lecture d'une variante, on clique sur l'icône « App », qui fait ensuite apparaître automatiquement un système de balises permettant d'opposer le lemme (en bleu) à ses variantes (en vert) et proposant la liste des témoins préalablement enregistrés[25] : on n'a plus qu'à choisir le témoin que l'on collationne en cliquant sur son sigle. Celui-ci est alors placé

[25] Les témoins sont enregistrés dans les métadonnées du fichier (<teiHeader>), dans des éléments <witness> auxquels sont assignés un attribut @xml:id.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

automatiquement avant la variante dans la vue apparat et, dans la vue avec balises, il apparaît précédé de l'attribut @wit (figure 4).

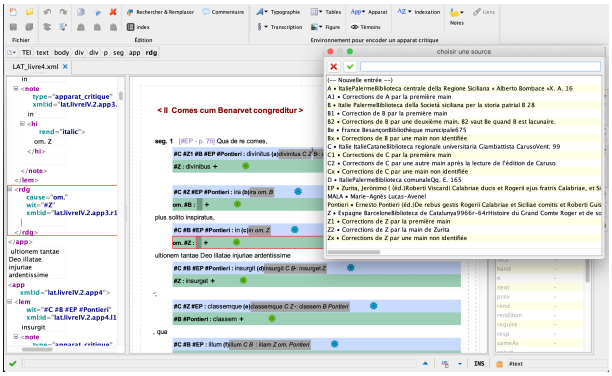


Fig. 4 : Vue « Encoder l'apparat critique » et icône « App »

Les autres témoins peuvent ensuite être ajoutés de la même manière soit dans le lemme, soit à côté du premier témoin qui porte la variante, soit en ajoutant une nouvelle variante. On est ainsi assuré de la validité des données dans l'environnement XML et l'on peut vérifier immédiatement que la variante a été correctement enregistrée en faisant apparaître la restitution de C sur la vue dédiée : sur celle-ci, la variante est mise en relief par un jeu de couleurs programmé par la rédaction de CSS. Ainsi les mots qui ne sont pas concernés par des variantes apparaissent en noir, les mots du témoin adoptés dans le <lem> sont en vert et ceux qui sont encodés comme variantes sont en rouge (voir figure 5).

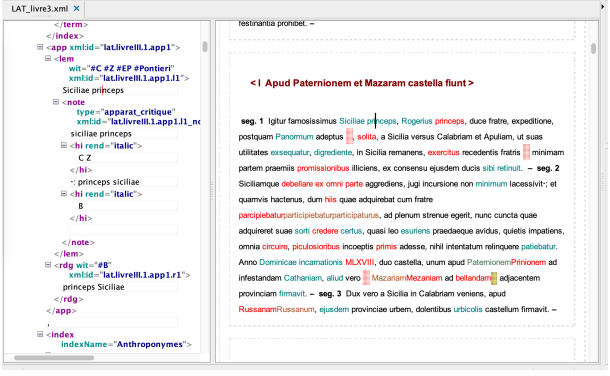


Fig. 5 : Malaterra III, 1, restitution de Z

Cette méthode d'encodage des variantes est beaucoup plus rapide et beaucoup plus sûre. Concernant l'enregistrement des données, il est inutile de faire une transcription de chaque témoin, qui est elle-même source de multiples erreurs : une seule suffit, à partir de laquelle on encode toutes les variantes au fur et à mesure de la collation des témoins et avec le degré de précision souhaité (enregistrement ou non des variantes

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

graphiques, ponctuation, etc.). La lecture du deuxième témoin permet d'entrer un certain nombre de leçons divergentes et de construire une première série d'unités critiques qui seront ensuite complétées lors de la lecture des autres témoins : le sigle de chacun pourra être ajouté soit dans le <lem> soit dans le <rdg> déjà existant ou dans un autre <rdg>, ou bien d'autres unités critiques seront aussi constituées et amèneront à vérifier la lecture du précédent témoin. L'environnement ne contraint pas à faire aussitôt cette vérification, et l'on peut, pendant le travail, rédiger des unités critiques incomplètes ; cette souplesse est commode, sans quoi il faudrait nécessairement collationner tous les témoins en même temps ; mais dès lors qu'un mot est encodé comme un <lem> et associé au témoin qui le porte, l'absence de l'information complémentaire dans les autres témoins engendrera l'apparition d'un blanc dans leur restitution : au moment de la relecture, on évitera ainsi de nombreux oublis (voir *infra*). Les risques d'erreur pendant le travail de collation sont bien connus des éditeurs de texte, dans l'ancrage de l'unité critique – confusion par exemple entre deux *et* placés à peu de distance –, ou dans la notation des variantes. Il convient donc d'insister sur la fiabilité du travail en XML-TEI, qui permet d'identifier de manière pérenne le lieu variant, lui associe le sigle des témoins concernés et évite de revenir sur la notation de la variante, car elle n'est pas liée à un mode de rédaction particulier ; au contraire, elle est annotée de manière standardisée, pérenne et interopérable.

La restitution de chaque témoin, que permet, comme on l'a dit, l'encodage de l'apparat critique suivant la méthode de la segmentation parallèle, est un formidable outil de relecture et de vérification. Associée à l'affichage du fac-similé du manuscrit, elle permet une comparaison efficace de l'un et de l'autre.

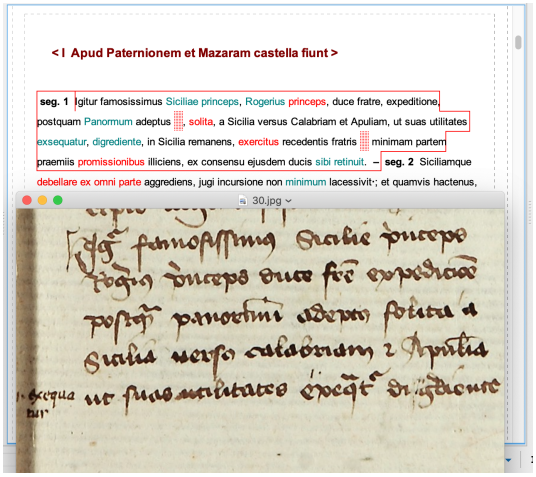



Fig. 6 : Malaterra III, 1, vue simultanée de la vue de Z et du fac-similé

La nature de la restitution dépend de l'encodage adopté. On peut choisir d'encoder toutes les abréviations, la ponctuation et les variantes graphiques et orthographiques de manière à donner à lire une restitution proche du fac-similé. Ce n'est pas le choix qui a été fait pour Malaterra : on n'a conservé, sauf exceptions, que les variantes morphosyntaxiques et lexicales, utiles à l'établissement du texte. Ce choix est bien

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

entendu contestable, et on pourrait souhaiter avoir une restitution qui soit au plus près du document. Elle est conçue cependant comme une aide à l'établissement du texte, c'est-à-dire déjà comme une interprétation du document, qui permet de comparer aisément le témoignage qu'il donne du texte avec celui des autres. Comme c'est ce qui nous intéressait dans cette édition, nous ne sommes pas allées plus loin dans l'encodage, considérant qu'il était nécessaire de fixer des limites à une recherche pour qu'elle puisse être publiée dans un temps qu'on qualifiera de raisonnable. Cependant, il serait possible d'opérer un encodage plus fin et d'obtenir plusieurs restitutions différentes selon la recherche qu'on veut mener. Et il est toujours temps d'affiner l'encodage, car, même si, comme on l'a dit, les fichiers XML publiés constituent un produit fini destiné à être référençables, ils peuvent être téléchargés à partir du site de publication et repris par le lecteur qui voudrait les doter d'un encodage supplémentaire pour mener une autre recherche.

### 3.2 Les différentes visualisations et la comparaison des versions textuelles

Pour accompagner le travail d'établissement et d'analyse de l'éditeur scientifique, différentes visualisations du texte encodé sont proposées. La comparaison des témoins est facilitée par le système des « vues multiples », grâce auquel on peut faire apparaître simultanément les vues colorisées de deux ou plusieurs témoins. Dans le cadre de l'édition de Malaterra, l'attention doit porter, en particulier, sur les leçons de C que Pontieri a trop souvent rejetées et qu'il faut confronter aux autres leçons (figure 7).

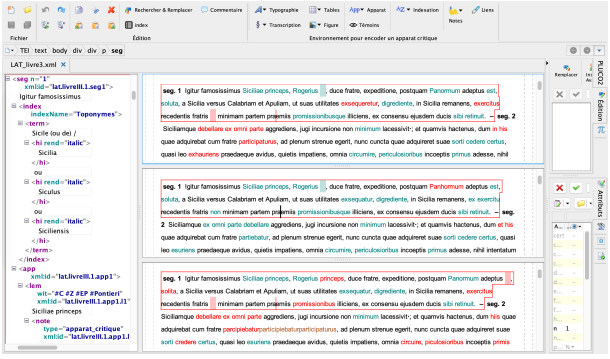


Fig. 7 : « vues multiples » (de haut en bas : restitutions de Pontieri, C et Z)

La présence des couleurs constitue un atout non négligeable par rapport à une simple transcription, car le chercheur sait à quels mots ou groupes de mots il doit prêter une attention particulière. Surtout, la lecture synoptique des variantes dans leur contexte phrastique permet de mieux comprendre l'origine d'une faute ou la logique qui a conduit à une correction de copiste.

Dans l'extrait présenté à la figure 8, par exemple, l'examen synoptique des vues de Z et de B – en l'absence du témoignage de C – a permis l'établissement de *cum archiepiscopus… praestolaretur* aux dépens de *cum archiepiscopis… praestolatur et* (ou *praestolabatur*), en facilitant la compréhension des conjonctions de coordination ou de subordination successives (*cum… ut cum… cum… tamen*).

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

Le pôle Document numérique crée en outre un site Web, qui n'est pas celui de l'éditeur matériel mais un site compagnon pour affiner les données de la recherche et visualiser les différentes couches d'encodage. Le site joue également un rôle dans l'établissement des données définitives. Dans le cas d'une édition bilingue avec apparat critique, il donne à voir le texte établi et sa traduction de manière synoptique ; il permet de relire chaque témoin de façon indépendante, dans une vue plus proche de celle qu'aura le récepteur du texte : notes en pop-up ou en bas de page, accès au fac-similé quand celui-ci est disponible sur le Web, etc. Enfin le site Web permet à l'éditeur d'utiliser un moteur de recherche plein texte pour comparer, interroger l'ensemble du corpus, ici les épîtres et les 4 livres[26].



Fig. 8 : Malaterra IV, 28 [29], 1 – vues multiples de haut en bas : vues de Z, B, texte édité

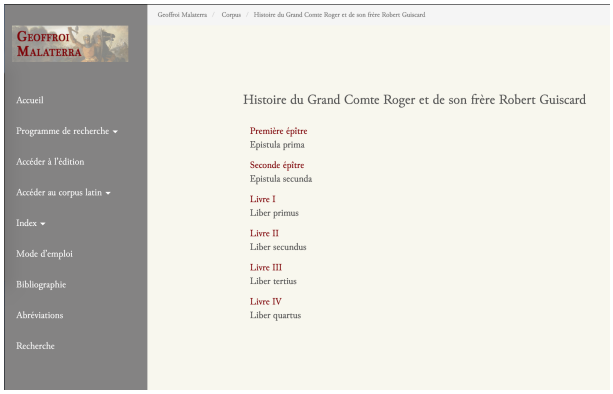



Fig. 9 : Prototype du site Web contenant l'ensemble de l'œuvre de Malaterra

### 3.3 Des outils pour l'exploitation du corpus

Concernant la publication, le même flux de textes encodé en XML-TEI assure la publication multi-support et multimodale selon des principes et une méthodologie définis par P.-Y. Buard dans sa thèse de doctorat, ce qui permet un gain de temps important et évite de perdre des données. On ne reviendra pas sur les différents avantages de

[26] Pour une représentation graphique de l'écosystème de travail de Malaterra, le lecteur pourra se reporter à Bisson et Lucas-Avenel, 2021.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

cette double publication, mais seulement sur les implications de l'encodage XML pour la publication de l'apparat critique. Les unités critiques, rédigées selon les règles de la CUF, apparaissent dans un élément <note type="apparat_critique"> inscrit dans l'élément <lem>. Ces notes sont indispensables à la publication papier ; et, comme elles rassemblent les informations sur les témoins, elles constituent, pour la publication en ligne, un point de départ pour naviguer ensuite dans les restitutions colorisées et les fac-similés. Pour le site qui correspond au volume I, on a voulu laisser au lecteur le choix d'y accéder ou non en dépliant le `[+]` qui suit le lemme.

Pour les livres I et II, les notes d'apparat critique avaient été rédigées avant qu'on décide d'encoder les variantes en TEI pour la publication du volume I, ce qui était ensuite apparu comme un double travail, à la fois long et susceptible de générer des erreurs. Au contraire, pour la préparation du second volume, on a voulu générer automatiquement ces notes par la création d'un outil dédié dont il fallait tester l'efficacité : permettrait-il la génération automatique d'une proportion suffisante de notes ? Quelle serait la part de travail manuel que les notes dont la rédaction n'est pas standardisée rendraient nécessaire ? Impliquerait-il de réfléchir à une nouvelle manière de rédiger les unités critiques ?

L'expérimentation s'est révélée extrêmement concluante. Sur le menu proposé par l'icône « note », il suffit de cliquer sur « générer les notes d'apparat » (voir fig. 3 *supra*). Un nouveau fichier est alors généré, qui n'écrase pas le précédent, dans lequel chaque note critique, rédigée selon les règles de la CUF, est placée dans le <lem> dans un élément <note> de @type apparat. Le lemme apparaît suivi de ses témoins dans l'ordre dans lequel ils ont été classés au moment de l'encodage et il est séparé des variantes par les deux points. Celles-ci sont disposées également suivant l'ordre des <rdg> avec leur(s) témoin(s). L'opposition romain (texte) *vs* italique (intervention de l'éditeur) est respectée, ainsi que le groupement des témoins d'une même famille (on a *ZB* et non *Z B*) ; le sigle EP utilisé pour l'encodage de la princeps est ramené à *ed. pr.*, et le témoignage commun des éditeurs est marqué par *edd.* En outre, le lemme ne fait pas apparaître le nom de l'éditeur, quand ce dernier l'avait adopté : cette information, explicite dans l'encodage, devient implicite dans la note d'apparat. La semi-automatisation de l'encodage décrit plus haut garantit que chaque témoin est écrit sans faute (respect des majuscules pour la notation des témoins, de l'exposant pour les différentes mains, du nom des éditeurs, etc.).

Dans un second temps, celui de l'export PDF ou de la publication sur le site, les notes sont disposées en bloc (au moyen d'une XSL), introduites par une lettre et séparées par une barre verticale. Voici une séquence, fournie par la génération automatique des notes (III, *Proem.* 1) :

> g. fecerunt *C* : facere *Z ed. pr. om. B* | h. scripturos *C B* : scripturas *Z* ed. pr. | i. repromisimus *C B* : repromisisimus $Z^1$ promisisimus *Z* | j. limpidiori poetria *C* : limpidiorum poetria *Z* limpidiora poema *B* limpidior poetria *ed. pr.* limpidioris poetriae Pontieri.

Il existe cependant certaines différences méthodologiques liées aux caractéristiques de l'un et l'autre système (encodage XML-TEI pour la restitution *vs* rédaction de

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

l'apparat critique pour la publication papier) qui empêchent de générer automatiquement un apparat entièrement conforme aux usages. Certaines différences peuvent être modifiées de manière automatique grâce à une programmation : ainsi les majuscules conservées dans les les <lem> et les <rdg> pour la clarté des restitutions sont supprimées dans les notes. De même, les éventuelles marques de ponctuation séparant deux mots d'un même lemme sont supprimées automatiquement pour l'apparat critique qui en refuse l'usage.

En revanche, d'autres éléments ne peuvent actuellement être générés automatiquement, comme les omissions et les additions : l'encodage XML interdisant l'implicite ne peut se contenter de l'information « vero *om. B* ». Il impose qu'on indique explicitement que les autres témoins donnent à lire le terme *vero*. De même, l'addition indiquée dans un <rdg> peut être opposée à un blanc qu'il faut aussi indiquer dans le <lem> (voir figure 10) :

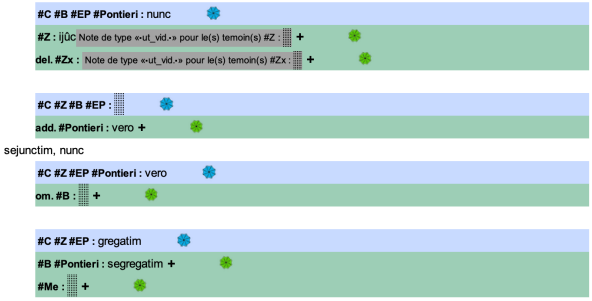



Fig. 10 : Encodage de III, *proem.* 1 : « nunc sejunctim, nunc vero gregatim »

En réclamant une génération automatique, on obtient les unités critiques suivantes :

> b. nunc *C B* : ijuc *ut. vid. Z del.* $Z^x$ | c. *CZB* : vero *add. Pontieri* | d. vero *C Z : om. B.* | e. gregatim *C Z* : segregatim *B Pontieri om. Me.*

Il convient donc de supprimer les témoins du lemme dans le cas d'une addition (note c.) et d'une omission (note d.), pour ne conserver que « vero *add. Pontieri* » et « vero *om. B* ». On reconnaîtra cependant que, quand le travail consiste à supprimer des caractères et non à en ajouter, les risques d'erreurs restent peu élevés. Un autre type d'encodage – au moins pour les additions – permettrait d'éviter d'avoir à faire cette correction. On pourrait en effet encoder comme suit :

```
<app>
   <lem wit="#C #Z #B #EP">nunc</lem>
   <rdg cause="add."wit="#Pontieri">nunc vero</rdg>
</app>
```

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

La génération automatique de la note donnerait le résultat suivant : « nunc *C Z B EP* : nunc vero *Pontieri.* » Cette notation est claire et elle est également admise par la CUF. Cependant, le fait de placer *nunc* dans le <lem> en fait un terme variant – ce qu'il nous a semblé préférable d'éviter.

Enfin, comme l'arborescence XML interdit les chevauchements entre deux balises <app>, on ne peut pas faire apparaître le lien entre les deux vero successifs sans réécrire dans le <lem> et le <rdg> tous les mots qui les séparent et les faire ainsi apparaître comme variants. On ne peut donc générer automatiquement une note du type : « nunc… nunc vero *C Z* : nunc vero… nunc *B* nunc vero… nunc vero *Pontieri* », qui permettrait de rendre plus lisible la logique de la faute ; mais cette notation ne serait pas nécessairement plus claire étant donné qu'il faudrait y insérer les variantes du premier *nunc* propres à Z et écrire : « nunc… nunc vero *C* : ijûc… nunc vero *Z ut vid.* nunc vero *corr.* $Z^{x}$ *ut vid.* nunc vero… nunc *B* nunc vero… nunc vero *Pontieri* ». Une unité critique rédigée de manière trop complexe risque de détourner le lecteur de la source, si bien qu'il nous semble qu'il faut privilégier la rédaction la plus simple qui éclaire le lecteur sur le contenu de chaque témoin. Si l'éditeur estime qu'il est important de faire apparaître nettement les liens entre les mots concernés par la variante, par exemple pour justifier le choix du lemme, il peut l'expliciter dans une note philologique.

Les transpositions au-delà de deux mots consécutifs, ainsi que les omissions longues ou les lacunes risquent également d'engendrer des chevauchements importants (par exemple l'omission de tous les chapitres en vers dans le manuscrit B). Ils sont encodés en TEI par l'utilisation des balises <lacunaStart> et <lacunaEnd>, qui évitent les chevauchements et permettent de faire apparaître les lacunes longues dans les restitutions pour le Web. En revanche pour la publication papier, ces informations doivent être rédigées dans une note critique.

Il convient donc de stabiliser le plus possible le texte et les variantes, avant de générer les notes d'apparat, afin de ne plus avoir à revenir sur les corrections opérées dans les notes critiques rédigées. On peut en outre choisir de supprimer certains éléments que l'on a jugés utiles d'encoder, mais qui peuvent alourdir inutilement l'apparat critique pour la version papier selon les recommandations des éditeurs (Huygens, 2001 : p. 63-64).

Enfin, il reste à traiter la question des troncatures. L'utilité ou non des troncatures est un exemple de choix qui divise actuellement les éditeurs de texte. On a fait le choix de les utiliser pour le volume I, dans le respect des recommandations de la CUF. En conformité avec le volume I, ces troncatures seront apportées ultérieurement aux autres volumes au moment de la publication, même si leur encodage ne présente en soi aucun intérêt scientifique et que leur automatisation est impossible, étant donné qu'elles se font au cas par cas. En outre, l'utilisation des troncatures, qui certes présente le double avantage de réduire la longueur de l'apparat critique et de faire apparaître plus rapidement le lieu variant, est en soit contestable, car il n'est pas sûr qu'elle facilite la lecture des variantes.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

Quels que soient les choix qu'on aura faits, les critères de lisibilité et de cohérence interne restent premiers, si bien qu'ils doivent être explicités dans l'introduction de l'édition. Les quelques cas présentés n'ont pas pour but de proposer un cadre enfermant sur ce type d'aspects, mais seulement d'indiquer quelles sont les potentialités et les limites de l'encodage dans la génération des notes d'apparat critique, car celles-ci peuvent influencer les choix finaux. Il s'agissait aussi de proposer un modèle général associant recherche et ingénierie de recherche dans la réflexion pour la conception et l'utilisation de nouveaux outils : les objectifs sont, d'une part, d'aboutir à une édition de texte établie scientifiquement en répondant à un ensemble de règles éprouvées depuis longtemps par la communauté des spécialistes d'ecdotique et, d'autre part, de mieux valoriser le travail de recherche en les associant aux sources désormais accessibles à un plus large public.

## 4 Conclusions

L'encodage du texte en XML-TEI présente un caractère innovant pour l'éditeur matériel, auquel il permet la publication multi-support et multimodale à partir des mêmes flux textuels (latin et français). Pour le chercheur, l'innovation réside moins dans la finalité du travail – il s'agit toujours d'établir le texte latin et son apparat critique –, que dans les modalités de sa réalisation, qui ont un impact sur les résultats et sur leur accès. L'apprentissage de l'encodage en XML-TEI nécessite un certain investissement, mais dont on fera profiter d'autres projets d'édition numérique utilisant les mêmes standards. Pour le chercheur, d'abord, le numérique peut permettre un gain de temps et de fiabilité, c'est aussi la possibilité de disposer d'outils de comparaison plus efficaces pour l'établissement du texte. Les restitutions des témoins obtenues à partir de l'encodage des variantes sont bien plus que des transcriptions des documents : elles sont déjà une interprétation de celui-ci qui, en étant liées au texte édité et aux autres témoins du texte, facilitent le repérage et l'analyse des variantes. Pour le lecteur, ensuite, les avantages de l'encodage XML-TEI ont déjà été souvent soulignés et ce sont de fait les plus visibles : on insistera sur le fait qu'il peut désormais mieux comprendre le processus sous-jacent à l'édition critique et que, devenant acteur dans l'acquisition des informations, il peut vérifier les données, les critiquer et les poursuivre par de nouvelles recherches.

Il est donc essentiel de soutenir les programmes de recherche visant à l'édition critique numérique de sources anciennes, accompagnées – si possible – d'une édition papier, étant donné l'importance, pour la recherche en littérature, en linguistique, en histoire et plus largement dans tous les domaines des Humanités, de disposer de sources valides, bien référencées et faciles d'accès. Il revient aux chercheurs de s'investir dans le développement des standards internationaux et à leurs tutelles de soutenir les initiatives dans ce domaine. Le développement actuel de nouveaux outils, en accès libre, accompagné de la réflexion et de la documentation nécessaires à la mise en place de bonnes pratiques par les chercheurs rassemblés dans les *consortia*, offre un contexte favorable pour le lancement de projets innovants sur des sources qui restent massivement inédites ou sur d'autres qui seraient avantageusement rééditées et/ou traduites en français et dans les autres langues modernes.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352