

Les apports des technologies Web à l'édition critique : l'expérience de l'Ecole des chartes

Gautier Poupeau

► **To cite this version:**

Gautier Poupeau. Les apports des technologies Web à l'édition critique : l'expérience de l'Ecole des chartes. Digital philology and medieval texts, Jan 2006, Arezzo, Italie, 2006. <sic_00137229>

HAL Id: sic_00137229

https://archivesic.ccsd.cnrs.fr/sic_00137229

Submitted on 18 Mar 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les apports des technologies Web à l'édition critique : l'expérience de l'Ecole des chartes

L'Ecole nationale des chartes mène depuis quatre ans une politique de publication électronique qui a abouti à la création d'une collection, Éditions en ligne de l'École des chartes¹ qui rassemble des éditions de sources, des bases de données, des instruments de référence et des recueils d'articles en accès libre et gratuit. Cette collection est complétée par un manuel en ligne de sciences auxiliaires de l'histoire, Theleme, Techniques pour l'historien en ligne, Études, manuels, exercices² constitué de bibliographies et de dossiers documentaires qui abordent toutes les techniques amenant à la synthèse historique à partir d'exemples concrets mis au point par les élèves avec l'aide de l'équipe pédagogique de l'établissement.

Cette expérience a tout d'abord représenté un moyen de mieux faire connaître et partager les compétences de l'École des chartes en matière de traitement et d'exploitation des sources historiques, en particulier l'édition critique de sources. L'édition électronique sur le Web représentait un vecteur de diffusion sans précédent et souple et limitait les moyens logistiques par rapport à la publication sur le support papier qui n'était pas abandonnée pour autant. Ainsi, le résultat d'entreprises de longue haleine comme l'édition de sources pouvait être mis à disposition sur le Web au fur et à mesure de sa réalisation, sans attendre sa mise au point finale. C'est dans cet esprit que la publication sur le Web du *Cartulaire blanc de l'abbaye de Saint-Denis*³, qui comprend 2500 actes, a vu le jour et est éditée et publiée au rythme d'un chapitre par an.

Pour autant, plus nous avançons dans la découvertes des possibilités de ce nouveau support, plus nous apparaissaient tangible les promesses de traitement et d'exploitation de l'information que nous laissaient entrevoir depuis quelques années les bases de données textuelles sur cédéroms. Entre tradition et innovation, une partie des apports des éditions sur le Web nous sont alors apparus. Le premier d'entre eux concerne l'accès à l'information.

Olivier Guyotjeannin, professeur de diplomatique médiévale à l'Ecole des chartes et directeur scientifique de l'édition du *Cartulaire blanc*, résume son sentiment sur l'édition électronique par cette phrase :

« Pour résumer notre sentiment, l'édition électronique change tout et rien. Rien, pour ce qui est de la préparation et de la présentation des éditions, dont la disposition typographique même, de la date à l'annotation colle même délibérément à la typographie des éditions « papier » dont nous suivons le modèle. Tout, en ce qu'elle amène à repenser

1 Éditions en ligne de l'Ecole des chartes, ELEC, <http://elec.enc.sorbonne.fr>

2 Techniques pour l'historien en ligne : études, manuels, exercices, <http://theleme.enc.sorbonne.fr>

3 Olivier Guyotjeannin (dir.), *Le Cartulaire blanc de l'abbaye de Saint-Denis*, ELEC n°3, éd. Ecole nationale des chartes, Paris, [en ligne], <http://elec.enc.sorbonne.fr/cartulaireblanc/>

tout le dispositif de consultation. »⁴

Ce propos résume parfaitement notre démarche de travail de ces dernières années. Dans un premier temps, nous avons fait le choix d'adapter les formes existantes de l'édition de sources. En effet, quel aurait été l'intérêt de remettre en cause des schémas qui ont fait leur preuve et qui sont devenus des règles à l'image des *Monumenta Germaniae Historica*? Nous risquions plutôt de détourner le chercheur obligé de réapprendre à appréhender une édition de sources. Notre but était donc l'appropriation pour les amener en douceur à ce nouveau support, le temps des innovations à tous les niveaux n'étaient tout simplement pas venu et plutôt qu'une révolution, nous souhaitions une évolution. Mais, adapter ne signifie pas imiter et, comme le fait remarquer Olivier Guyotjeannin, ces premières initiatives intégraient une innovation de taille : **la multiplication des points d'accès à l'information** ce qui représente l'apport du support numérique le plus simple à mettre en place. Si on prend l'exemple du Cartulaire blanc, huit types d'accès différents à l'information sont proposés :

- une table chronologique cumulative de tous les actes en ligne ;
- un système de feuilletage reprenant la navigation linéaire du support papier ;
- une table des actes par chapitre ;
- un tableau statistique des actes par auteur ;
- des tableaux généalogiques ;
- des liens hypertextes entre les actes ;
- des introductions permettant de replacer les actes d'un chapitre dans son contexte historique et géographique ;
- la recherche simple en texte intégral.

Les possibilités de l'hypertexte jouent ici un rôle fondamental dans la mise en place de ce réseau d'informations. Dans une autre édition, l'obituaire du Saint-Mont⁵, cette liste est complétée par un *index nominum* et un *index rerum* qui ne sont plus simplement des outils de repérage comme sur le support papier, mais des outils de navigation efficaces offrant autant de parcours de lecture que d'entrées d'index, multipliant ainsi les approches de la source.

Une fois que les principes de fonctionnement de base étaient assimilés, l'étape suivante a consisté à améliorer nos méthodes de travail, en mettant l'accent sur la structuration de l'information qui nous paraissait insuffisante dans nos premières expériences. Le choix de la TEI s'est imposé naturellement. Elle permettait de bénéficier de l'expertise d'une large communauté, d'assurer les impératifs de pérennité et correspondait à pratiquement tous nos besoins, d'autant plus que la mise au point et la

4 Olivier Guyotjeannin et Gautier Poupeau, « Le projet d'édition électronique du Cartulaire blanc de l'abbaye de Saint-Denis et les projets électronique de l'École nationale des chartes », dans *le Médiéviste et l'ordinateur*, n°42, printemps 2003, éd. IRHT, Paris, 2003, [en ligne], http://lemo.irht.cnrs.fr/42/mo42_12.htm, consulté le 7 janvier 2006.

5 Marie-José Gasse-Grandjean, *L'obituaire du Saint-Mont*, ELEC n° 8, éd. École nationale des chartes, Paris, [en ligne], <http://elec.enc.sorbonne.fr/obituairesaintmont/>

maintenance d'une grammaire XML auraient représenté un travail trop considérable pour un établissement de taille modeste comme le notre. Le travail d'assimilation de la TEI a duré plus d'un an, mais a permis à l'établissement d'acquérir une expertise sans précédent dans ce domaine en France. De plus, elle nous ouvre maintenant de nouvelles perspectives.

Dans un manuel intitulé *Conseils pour l'édition de textes médiévaux*⁶, Françoise Vielliard et Olivier Guyotjeannin rappellent que « les potentialités ouvertes par la mise à disposition de bases textuelles informatisées et de corpus numérisés permettront très vite de résoudre le dilemme [entre les tenants de l'édition critique et de l'édition imitative], en juxtaposant commodément plusieurs moyens de prendre connaissance des textes médiévaux : aspect physique des manuscrits et actes originaux, éditions de travail, éditions accessibles à la lecture courante... ». Cette idée de juxtaposer différentes versions d'une même édition est à l'origine d'un travail en cours d'élaboration : *l'édition des chartes de l'abbaye de Corbie* par Laurent Morelle qui démontre tout le potentiel du support numérique. Dans la thèse originale dont est issue cette édition, Laurent Morelle intégrait à la fois les règles de l'édition critique et l'édition de travail lui permettant d'étudier le corpus d'un point de vue paléographique, diplomatique et archivistique. Nous aboutissons à un travail scientifiquement parfait, mais qui devient très rapidement illisible, surtout si l'historien recherche une information précise dans ce corpus. Grâce aux possibilités du XML, nous pouvons encoder dans le fichier source l'ensemble des informations et proposer plusieurs pages Web correspondant aux différentes couches d'informations :

- L'édition critique proprement dite avec un texte lisible pour prendre connaissance rapidement des informations intégrant tout l'appareil critique indispensable : regeste, tableau de la tradition, apparat critique et identifications des noms de personnes et/ou de lieux ;
- l'édition génétique propose le texte en précisant ses particularités physiques : résolutions d'abréviations, lettres décorées, lettres allongées, les changements de lignes, les restaurations, des détails de l'acte sous forme d'images particulièrement intéressants pour leurs dispositions ou leurs caractéristiques physiques, les *signa*... et l'édition des mentions dorsales.
- Une note critique de l'acte assez longue replace l'acte dans le contexte historique et/ou propose une discussion sur la datation de l'acte ou tout autre problème d'ordre diplomatique, paléographique ou codicologique.
- La numérisation du recto et du verso de l'acte accompagnée de la possibilité de zoomer grâce à la technologie Flash.

Dans ce cas, **les technologies Web permettent de proposer différentes approches de la source aux chercheurs**⁷. Il est même envisageable que le chercheur puisse configurer lui-même l'affichage en

6 École nationale des chartes, Françoise Vielliard et Olivier Guyotjeannin, coord., *Conseils pour l'édition des textes médiévaux*. I. Fasc. 1 : conseils généraux, éd. du CTHS, Paris, 2001, p. 14.

7 Elena Pierazoo, « Just different layers ? Stylesheets and digital edition methodology », *Digital humanities* 2006, Paris, [en ligne], <https://webcgi.oulu.fi/dh2006/viewabstract.php?id=94>, consulté le 18 mars 2007.

choisissant les informations qu'ils désirent afficher. La possibilité de faire apparaître les différentes couches peut aussi renseigner les chercheurs sur les méthodes employées par l'éditeur scientifique pour aboutir à son édition finale et à sa note critique. Enfin, elle apporte de nouvelles perspectives sur l'élaboration même de l'édition, puisqu'il est ainsi possible d'utiliser un tel système pour aider à l'élaboration de l'édition.

De plus, cette expérience laisse entrevoir les possibilités d'encodage presque infinies du XML. En effet, chaque chercheur pourrait encoder les informations en fonction de son propre intérêt et exploiter les informations issues de cet encodage. Pour les chartes de l'abbaye de Corbie, le codage des résolutions des abréviations permettra d'étudier les différentes abréviations, leurs évolutions à travers le temps... et le codage des mentions dorsales va nous permettre de mettre au point une typologie pour étudier l'évolution des mécanismes d'archivage de l'acte. De même, dans le cadre de l'édition des *Chroniques latines de Saint-Denis*⁸ dont l'édition critique des trois premiers livres est maintenant disponible en ligne, l'encodage des différentes leçons des manuscrits R et P permet d'envisager la juxtaposition du texte mis au point par l'éditeur scientifique, mais aussi celui des deux manuscrits originaux. Dans la même publication, l'encodage fin des citations permet de dresser des listes précises des citations par œuvres et par chapitre accompagné de l'indication d'approximation avec l'original⁹. L'étude de ces citations est alors facilitée. A l'image de ces deux exemples, on peut envisager par exemple des études paléographiques très fines, la TEI permettant de coder les changements de main et même d'encre ou aussi effectuer des études sur la ponctuation originale des textes. Bref, chaque spécialiste pourrait ajouter facilement à l'édition de base les annotations propres à son étude. Cette approche fait, d'ailleurs, l'objet d'une expérience en France par le laboratoire CIHAM de Lyon sur les sermons de Jacques de Voragine¹⁰.

Évidemment, c'est dans le **domaine de la recherche d'information** que les apports du support numérique sont les plus importants. Pour chacune de nos éditions, il est possible d'effectuer une recherche en texte intégral simple n'intégrant pas de lemmatisation. Un projet de numérisation¹¹ d'éditions de chartiers publiés au XIXe siècle concernant des établissements de l'Ile-de-France commencé en 2005 nous a permis de réfléchir à l'amélioration des possibilités de recherche. Le texte a été encodé en TEI selon les mêmes principes que le Cartulaire blanc de Saint-Denis¹². Grâce au

8 Pascale Bourgain (dir.), *Chroniques latines de Saint-Denis*, ELEC n°13, éd. École nationale des chartes, Paris, [en ligne], <http://elec.enc.sorbonne.fr/chroniqueslatines/>

9 cf. <http://elec.enc.sorbonne.fr/chroniqueslatines/references/>

10 Thesaurus des sermons de Jacques de Voragine, édition d'un corpus de sermons latins médiévaux, <http://www.sermones.net>.

11 Cartulaires numérisés d'Ile-de-France, ELEC, n°11, <http://elec.enc.sorbonne.fr/cartulaires/>

12 Ces principes ont été exposés dans plusieurs articles ou communications dont Gautier Poupeau, « Réflexions sur l'utilisation de la TEI pour coder les sources diplomatiques », dans *le Médiéviste et l'ordinateur*, n°43, 2004, éd. IRHT, Paris, [en ligne], <http://lemo.irht.cnrs.fr/43/43-12.htm>, consulté le 7 janvier 2006.

logiciel eXist, une base de données XML native Open Source, nous avons pu mettre en place différentes interfaces de recherche sur l'ensemble du corpus numérisé et sur l'ensemble des éditions critiques mises en ligne par l'École nationale des chartes et l'Institut de recherche et d'histoire des textes (IRHT)¹³. Nous poursuivons actuellement dans les deux directions pour ajouter du contenu : numérisation et nouvelle édition, sachant bien-sûr que la seconde est la plus fastidieuse et donc la plus lente. Mais, la recherche en texte intégral ne représente qu'une des formes possibles d'interrogations. Chaque acte est accompagné par des métadonnées :

- date ;
- type d'auteur de l'acte ;
- langue utilisée ;
- Région administrative contemporaine concernée.

Nous testons donc actuellement une interface d'interrogation avec eXist qui permettrait d'interroger selon ses critères les dossiers documentaires de thélème, les actes édités (Cartulaire blanc et à terme chartes de l'abbaye de Corbie) et les actes numérisés. Pour ces derniers, nous devons enrichir la numérisation en ajoutant ces informations. Sur le même principe, nous allons faire converger l'ensemble de nos éditions de sources ou instruments de référence¹⁴.

Dans le même ordre d'idée, nous testons trois interfaces de recherche pour améliorer l'exploitation de notre corpus :

- La recherche floue ;
- L'affichage des résultats sous forme de concordances ;
- l'index des formes du corpus.

La recherche floue est utilisée dans le domaine du traitement automatique des langues, en particulier dans la traduction assistée de corpus et de la recherche d'informations. Elle permet de trouver non seulement l'équivalent exact de l'argument de recherche, mais aussi tous les éléments qui présentent une certaine similitude. Le logiciel eXist intègre nativement cette fonctionnalité en se basant sur la distance de Levenshtein¹⁵, algorithme mis au point par Vladimir Levenshtein pour comparer deux chaînes de caractère. Cette fonctionnalité permet de palier à l'absence de lemmatisation et présente l'avantage de proposer l'ensemble des formes présentes dans le corpus. L'interface actuellement en test permet d'afficher tous les formes approchantes d'une forme recherchée, de choisir les formes à

13 Interface d'interrogation du centre de ressources numériques TELMA, <<http://www.cn-telma.fr/corpus/>>, par exemple, pour les Cartulaires numérisés, <<http://www.cn-telma.fr/corpus/cartulaires/>>

14 Les interfaces de recherche de Telma offrent actuellement la possibilité d'effectuer des recherches par critères sur un corpus, par exemple, pour les cartulaires numérisés, <<http://www.cn-telma.fr/corpus/cartulaires/rechercheCriteres/>>, mais nous étudions la possibilité d'effectuer ces recherches sur les corpus qui partagent les mêmes critères

15 Levenshtein distance, dans Wikipedia, [en ligne], http://en.wikipedia.org/wiki/Levenshtein_Distance, consulté le 7 janvier 2006.

rechercher¹⁶ et d'afficher les résultats de cette recherche¹⁷.

L'affichage des résultats sous forme de concordances ou KWIC permet de visualiser les résultats de la forme recherchée dans le contexte du texte¹⁸. Cela permet d'étudier les différentes utilisations d'un mot et donc ses différents sens. A partir de cette fonctionnalité, nous entrevoyons la possibilité de mettre au point des tables de coocurrences pour effectuer des études de sémantique historique. Enfin, l'index des formes du corpus permet d'entrevoir les possibilités en matière de statistiques textuelles¹⁹, mais aussi dans des outils d'aide à la lemmatisation. Ces différentes interfaces et possibilités d'interrogation sont disponibles sur le site Web du centre de ressources numériques TELMA porté par l'École nationale des chartes et l'Institut de recherche et d'histoire des textes dont la vocation est d'apporter aide et assistance technique à la communauté des chercheurs dans la mise en ligne de sources historiques ou d'entreprises de repérage et de description des sources, en assurant.

Le support numérique impose, dans un premier temps, la mise en place d'une réflexion sur les pratiques, les buts et les méthodes de l'édition critique. Rapidement, elle amène à repenser un certain nombre de règles ou de schémas d'édition considérés comme acquis, tout en apportant des nouvelles possibilités et en créant de nouvelles règles. L'encodage en XML est à cet égard particulièrement significatif. Ces apports se situent aussi bien au niveau du gain de diffusion du travail scientifique en lui-même, que des possibilités d'accès, de présentation et de traitements de l'information. De plus, avec la mise en place de systèmes basés sur les Web services, à l'image de Open archive initiative, on pourra bientôt envisager l'interrogation commune des différents projets qui voient le jour à travers le monde.

16 <http://www.cn->

[telma.fr/corpus/cartulaireblanc/pageNum\(1\)/fullText.xq?session=nouvelle&_corpus_cartulaireblanc=on&mot=abbas&typeRecherche=floue&tailleContexte=100&nbResPage=15&rueil=on&tremblay=on](http://www.cn-telma.fr/corpus/cartulaireblanc/pageNum(1)/fullText.xq?session=nouvelle&_corpus_cartulaireblanc=on&mot=abbas&typeRecherche=floue&tailleContexte=100&nbResPage=15&rueil=on&tremblay=on)

17 [http://www.cn-telma.fr/corpus/cartulaireblanc/pageNum\(1\)/fuzzySuite.xq?abbas=on&abbate=on&abbati=on](http://www.cn-telma.fr/corpus/cartulaireblanc/pageNum(1)/fuzzySuite.xq?abbas=on&abbate=on&abbati=on)

18 <http://www.cn->

[telma.fr/corpus/cartulaireblanc/pageNum\(1\)/fullText.xq?_corpus_cartulaireblanc=on&session=nouvelle&mot=abbas&rueil=on&tremblay=on](http://www.cn-telma.fr/corpus/cartulaireblanc/pageNum(1)/fullText.xq?_corpus_cartulaireblanc=on&session=nouvelle&mot=abbas&rueil=on&tremblay=on)

19 Par exemple, l'index de formes du chapitre Tremblay du Cartulaire blanc, <<http://www.cn-telma.fr/corpus/cartulaireblanc/indexFormes?ouvrage=tremblay>>