



## La ” philologie numérique ” : tentative de définition d’un nouvel objet éditorial

Céline Guillot, Alexei Lavrentiev, Thomas Rainsford, Christiane  
Marchello-Nizia, Serge Heiden

### ► To cite this version:

Céline Guillot, Alexei Lavrentiev, Thomas Rainsford, Christiane Marchello-Nizia, Serge Heiden. La ” philologie numérique ” : tentative de définition d’un nouvel objet éditorial. Richard Trachsler; Frédéric Duval; Lino Leonardi. 27e Congrès international de philologie et de linguistique romanes, Jul 2013, Nancy, France. 2016, Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013). Section 13 : Philologie textuelle et éditoriale. <[http://www.atilf.fr/cilpr2013/actes/section\\_13/Guillot\\_Heiden\\_Lavrentiev\\_Marchello-Nizia\\_Rainsford.pdf](http://www.atilf.fr/cilpr2013/actes/section_13/Guillot_Heiden_Lavrentiev_Marchello-Nizia_Rainsford.pdf)>. <halshs-00846767>

**HAL Id: halshs-00846767**

**<https://halshs.archives-ouvertes.fr/halshs-00846767>**

Submitted on 31 Oct 2013

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

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



Distributed under a Creative Commons Attribution 4.0 International License

# La « philologie numérique » : tentative de définition d'un nouvel objet éditorial

## 1. Introduction : État des lieux de l'édition numérique

L'édition numérique a aujourd'hui plus de vingt ans. C'est en 1990 que Karl Uitti a commencé son travail sur les transcriptions numériques des manuscrits du *Chevalier de la Charrette*<sup>1</sup> ; dans le domaine anglophone, c'est au début des années 90 que Peter Robinson a commencé son travail sur les manuscrits du *Canterbury Tales* (Robinson 2006), et Hoyt Duggan sur *Piers Plowman* (Gifford Fenton / Duggan 2006). Dans la dernière décennie, plusieurs initiatives importantes ont permis l'édition électronique de textes en français médiéval : citons, par exemple, l'édition des textes pour le *Dictionnaire électronique de Chrétien de Troyes* (Ottawa et Nancy, sous la direction de Pierre Kunstmann)<sup>2</sup>, le *Online Froissart* à l'Université de Sheffield (Ainsworth / Croenen 2012), ou bien la collection ELEC (Éditions électroniques de l'École des chartes)<sup>3</sup>.

Cependant, malgré ce développement continu, il semble que le numérique soit encore loin de remplacer les éditions traditionnelles sur papier. Dans un article récent, Dot Porter (2013) s'interroge sur l'usage des ressources numériques par les médiévistes dans leurs pratiques de recherche. Elle montre que le numérique est désormais le moyen préféré pour la consultation des revues, alors que la majorité des médiévistes n'utilise que rarement les éditions électroniques (2013, 22). Deux facteurs expliquent, selon elle, cette situation : d'une part, le petit nombre des éditions électroniques produites, et, d'autre part, une certaine méfiance due à la confusion entre 'éditions numériques' et éditions imprimées 'numérisées' (2013, 13-14). Il importe donc de signaler que par 'philologie numérique' nous n'entendons pas la numérisation des éditions imprimées, telle qu'elle est réalisée dans les grands corpus électroniques (*Base de français médiéval*, FRANTEXT, *Corpus de la littérature médiévale* de Garnier, etc.). Bien qu'incontournables, pour le linguiste, ces ressources ont des limitations importantes du point de vue philologique, la plus sérieuse étant l'élimination systématique de la matière critique (introduction, notes, variantes, glossaires), ce qui oblige le lecteur à se référer à l'édition source. En outre, à cause des droits que l'éditeur se réserve sur le texte, les éditions disponibles à la numérisation ne sont pas toujours les meilleures<sup>4</sup>. Au-delà des corpus, les textes simplement scannés (dans Gallica ou dans Google Books) présentent très souvent une océrisation non vérifiée et un encodage sommaire.

La philologie numérique que nous évoquons implique donc l'édition du texte, à partir de sources manuscrites, et de la matière critique nécessaire à son évaluation et à son utilisation. Cependant, alors qu'il semble exister des conventions bien établies qui guident l'organisation d'une édition imprimée, les pratiques de l'édition électronique restent très diverses, et on observe même une certaine réticence à profiter pleinement des possibilités du numérique. L'édition du formulaire d'Odart Morchesne (Guyotjeannin / Lusignan 2011), par exemple, est une édition multimédia, et pour cette raison la version électronique reste très fidèle aux normes de l'édition imprimée : on affiche un seul texte critique avec des notes en bas d'écran, avec très peu de liens hypertextuels et sans l'image du manuscrit. Les sources XML-TEI de l'édition sont téléchargeables, ce qui donne la liberté à l'utilisateur de l'exploiter avec ses propres outils. En revanche, *The Online Froissart* livre une série de transcriptions de plusieurs manuscrits qui peuvent ainsi être comparés, ainsi que des images des manuscrits originaux. Le glossaire et l'index des noms propres sont complètement réinventés : par défaut, une description de chaque personnage historique s'affiche lorsqu'on clique sur les noms propres du texte et, lorsqu'on sélectionne un autre mode de lecture, chaque mot devient un lien vers sa définition dans le *Dictionnaire du Moyen Français*. Mais cette ressource reste en quelque sorte complémentaire à l'édition imprimée (Ainsworth 2007). Il s'agit de transcriptions de manuscrits et non pas d'une édition critique ou même courante. Les éditeurs ne suggèrent aucun texte de base par défaut et il est nécessaire de sélectionner l'un des quarante témoins avant de pouvoir commencer la lecture.

Dans ce contexte de création de nouveaux objets éditoriaux et d'hétérogénéité relative, nous voudrions apporter des éléments de réflexion sur les principes de l'édition numérique, en nous appuyant sur nos expériences menées dans le cadre d'une collection d'éditions, la « Collection de la Base de français médiéval » (désormais « Collection BFM »), dont les textes phares sont pour l'instant la *Queste del saint Graal* (Marchello-Nizia / Lavrentiev 2013) et le *Psautier*

<sup>1</sup> Voir le site du projet, <<http://lancelot.baylor.edu/about>>.

<sup>2</sup> <<http://www.atilf.fr/dect>>.

<sup>3</sup> <<http://elec.enc.sorbonne.fr>>.

<sup>4</sup> Par exemple, le *Corpus de littérature médiévale* de Garnier reste essentiellement basé sur les textes édités par Champion ; pour les autres textes, le choix d'édition semble porter plutôt sur des éditions dans le domaine public.

*d'Arundel* (voir ici même Pignatelli / Lavrentiev 2014). Les conventions d'édition proposées ici visent à fournir une base concrète aux éditeurs et aux évaluateurs, qui disposeront ainsi d'un cadre d'analyse et de comparaison essentiel à la reconnaissance de la valeur scientifique d'un travail de philologie numérique. En soulignant les points forts et les nouvelles possibilités qu'offre le numérique, nous cherchons à reprendre les débats récurrents dans le champ de l'édition et de la critique des textes afin d'établir un modèle d'édition renouvelé.

## 2. Points forts du numérique

Les atouts du numérique paraissent immédiatement perceptibles à tous : capacités de stockage accrues, accessibilité des données, souplesse d'utilisation, interactivité et caractère dynamique, etc. Dans le but d'illustrer leur apport à l'édition, nous aborderons les thématiques des choix et normes d'édition, des types d'édition (diplomatique, critique, etc.) et de la critique textuelle (étude de la tradition manuscrite et sélection du ou des manuscrits de base).

### 2.1. La transparence et la normalisation des choix d'édition

Le numérique repose sur la définition très précise des choix et procédures d'édition. Dans le domaine français, les normes éditoriales semblent être relativement consensuelles et stables depuis le premier tiers du vingtième siècle et le manuel publié par le groupe de recherche de l'École nationale des chartes (2001-2002) en donne une image très fidèle. L'examen détaillé d'un nombre important d'éditions scientifiques fait pourtant conclure F. Duval (2006) à une certaine hétérogénéité des pratiques. Il est probable que cette hétérogénéité repose sur le pragmatisme qui caractérise la doctrine française. La nécessité affirmée d'adapter les choix d'édition au type de texte qu'on édite, au public et parfois aussi au « goût » de l'éditeur ouvre la porte à des variations qui peuvent être relativement importantes en pratique.

Certains aspects des principes d'édition ont par ailleurs été négligés dans la réflexion depuis les origines : par exemple, le découpage des unités lexicales et discursives dans ses relations avec la segmentation médiévale et les usages propres à chaque manuscrit (usage des agglutinations, de la ponctuation ou des pieds de mouche, etc.) est traditionnellement très peu traité dans les introductions des éditeurs. Et ces points ne sont que très peu abordés aussi dans les textes fondateurs de l'édition française (Roques 1926, manuel de l'École des chartes, etc.). Il s'agit pourtant d'un domaine de recherche dont l'intérêt a déjà été démontré (Andrieux-Reix / Monsonégo 1997 et 1998, Lavrentiev 2009) et dont les développements les plus récents de la linguistique diachronique ne font que souligner davantage l'importance pour l'étude du changement linguistique (cf. le rôle des processus de coalescence dans la lexicalisation et la grammaticalisation), comme pour les recherches sur l'évolution des modes d'organisation et de progression textuelle (voir notamment Combettes 2012).

Les technologies du numérique conduisent ainsi à préciser des points qui pouvaient rester implicites, voire relativement flous, dans les éditions traditionnelles. Les procédures actuelles permettent de définir la liste précise des choix opérés à chaque phase et à chaque niveau du codage du texte (en caractères, en unités lexicales, en unités de plus grande taille)<sup>5</sup>, et, éventuellement aussi, d'indiquer le statut et l'identité des personnes qui sont à l'origine de ces choix<sup>6</sup>. Les rôles, les responsabilités et les différents acteurs du document électronique peuvent ainsi être déterminés de façon claire. Les usages les plus répandus parmi les producteurs de ressources numériques tendent aussi à assortir le document électronique d'une documentation technique qui rend compte de la façon dont le texte a été établi et, souvent aussi, de ses possibilités d'usage. Enfin, le principe d'ouverture et d'échange qui caractérise très largement le monde du numérique conduit tout naturellement l'éditeur du texte électronique à rechercher et à appliquer des standards et normes d'édition qui soient aussi larges et partagés que possible.

### 2.2. Des éditions à facettes multiples

L'un des principaux apports du numérique est également qu'il donne à l'éditeur le moyen de dépasser les limites imposées par un type d'édition particulier. Parce qu'elles offrent la possibilité de coder simultanément des choix auparavant considérés comme antagonistes et exclusifs, les nouvelles technologies permettent de réaliser des éditions multi-facettes, à la fois normalisées et diplomatiques par exemple. La délimitation de plusieurs couches de codage, chaque couche correspondant à un niveau de représentation des données primaires, permet au producteur du texte électronique de générer, à partir d'un document unique, plusieurs vues différentes (ou *présentations*, voir infra) du texte. Pour prendre un exemple très concret, il est désormais possible de suivre l'usage du manuscrit dans son emploi des lettres *i* et *j*, *u* et *v*, tout en adaptant les graphies aux normes modernes 'ramistes' dans une couche distincte du document.

---

<sup>5</sup> En particulier, l'usage de balises dont la sémantique doit être clairement définie oblige à détailler ces choix.

<sup>6</sup> Ce qui permet de dire aussi, par exemple, si l'occurrence d'un signe de ponctuation relève de la seule décision de l'éditeur ou si elle reflète l'état du manuscrit.

La souplesse des outils de visualisation permet de restituer en bout de chaîne ces choix multiples en donnant à voir une multiplicité de réalisations du même texte. Les usages dominants tendent à regrouper les choix concordants en reproduisant les types d'édition traditionnels (niveaux normalisé et diplomatique). Mais les contraintes matérielles du papier étant dépassées, l'éventail des possibilités est bien plus important et plusieurs éditions électroniques donnent accès à trois niveaux de visualisation. En outre, la plupart de ces éditions proposent, à côté d'une ou de plusieurs transcriptions du texte, l'image correspondante du manuscrit. Les différents niveaux de transcription peuvent ainsi constituer une progression qui va du niveau le plus fidèle au manuscrit au texte le plus conforme aux normes de l'édition critique traditionnelle. Cette pratique a l'avantage de rendre transparents les choix de l'éditeur.

Ces nouvelles possibilités techniques ouvrent le champ au débat sur la définition d'une typologie des éditions et des usages. Rien n'interdit, par ailleurs, de laisser l'utilisateur faire lui-même les choix qui lui paraissent pertinents ou qui répondent à ses intérêts ou ses perspectives de recherche. Le caractère dynamique et interactif des outils numériques rend cela également possible.

### 2.3. Des éditions de variantes multiples

Les débats sur les modalités d'étude de la tradition manuscrite et de classement des manuscrits, sur le choix d'un ou de plusieurs témoins comme manuscrit(s) de base, sur le rôle et l'étendue des variantes sélectionnées ou sur le poids des corrections éditoriales ne sont évidemment pas clos. Dans ce domaine aussi, les évolutions technologiques actuelles apportent un éclairage différent ou permettent de traiter des problèmes anciens au moyen de méthodes en partie renouvelées.

En théorie comme en pratique, ces nouveaux outils sont aussi utiles pour construire des éditions de type bédieriste, qui s'appuient sur un manuscrit de base bien choisi, que sur des éditions davantage reconstructivistes ou critiques. La présence de l'image du manuscrit, ainsi que la possibilité de transposer avec une grande précision ses caractéristiques physiques, tendent à favoriser une transcription fidèle d'un manuscrit particulier. Dans le domaine de la critique textuelle, ces mêmes technologies offrent des solutions pratiques pour encoder, visualiser et comparer différents témoins d'une même œuvre. Elles permettent au littéraire comme au linguiste – toujours désireux de pouvoir comparer des variantes de sens assez proche – d'accéder simultanément à différentes versions du même texte. L'édition de la tradition manuscrite intégrale d'une œuvre, rarement réalisée sur le médium papier<sup>7</sup>, semble désormais possible et ces évolutions pourraient influencer aussi sur les méthodes de reconstruction de l'archétype.

Pour illustrer ces principes très généraux, nous nous appuierons dans la section suivante sur notre projet de « Collection BFM ».

## 3. Principes de philologie numérique de la « Collection BFM »

### 3.1. Choix du manuscrit

Comme on l'a dit plus haut, tous les types d'édition peuvent être réalisés sous forme numérique, qu'il s'agisse d'une édition 'critique' au sens plein, tentant de reconstituer soit le texte original à partir des leçons différentes des manuscrits, soit un texte archétypal prenant en compte les diverses versions qui nous sont parvenues, soit une édition 'normalisée', prenant pour base l'un des manuscrits, qui a été reconnu comme 'bon' selon des critères explicites. Dans tous les cas, l'analyse préalable de tous les manuscrits reste la règle.

Actuellement, c'est l'option dite du 'manuscrit de base' que nous avons privilégiée, d'une part pour l'édition numérique 'prototypique' de la « Collection », celle de la *Queste del saint Graal* réalisée à partir du manuscrit K<sup>8</sup>, et d'autre part pour les éditions actuellement en cours de réalisation (*Vie de saint Alexis*, *Psautier d'Arundel*).

Ce choix repose sur trois raisons. Notre objectif est en premier lieu de reproduire, le plus fidèlement possible, un 'texte usagé', tel qu'il a été mis en circulation, lu et pratiqué par des lettrés contemporains. En outre, après plusieurs décennies de réflexions théoriques sur les notions de réception, de lecture, d'interactions, vouloir restaurer un texte sous une forme n'ayant jamais existé au Moyen Âge paraît problématique. Enfin, pour certains textes, il nous est parvenu de très nombreux manuscrits qui n'ont pas tous été transcrits, et dans de tels cas la stémato-logie se fonde sur la comparaison d'extraits sélectionnés essentiellement sur leur valeur narrative : sans doute avec l'essor de la philologie numérique la situation évoluera-t-elle.

Dans le cas de la *Queste*, choisir le même manuscrit de base que l'éditeur ancien, A. Pauphilet, permettait de

<sup>7</sup> Voir notamment l'édition du *Lai de Lanval* par Rychner (1958) ou le *Recueil des Fabliaux* de Noomen (1983-1998)

<sup>8</sup> À la suite de Pauphilet (1921), nous utiliserons désormais ce sigle pour désigner le manuscrit Lyon, Bibliothèque municipale, P.A. 77.

comparer deux éditions obtenues selon la même pratique du même ‘manuscrit de base’ : pour un texte d’environ 110 000 mots, 20 000 différences.

### 3.2. Règles de transcription

S’agissant d’une édition électronique, il convient de distinguer la *représentation* des données pertinentes pour l’analyse du texte, qui se fait à l’aide de balises appropriées, de la *présentation* (ou visualisation), qui peut varier selon le support de l’édition (grand écran d’un ordinateur, tablette, liseuse ou papier) et selon les options de personnalisation que le lecteur peut choisir. Dans cette section, nous tâcherons d’exposer nos choix de représentation des données sans entrer dans les aspects techniques du balisage et en indiquant des possibilités de visualisations qui ne sont pas exclusifs.

Le tableau suivant donne une vue synthétique des caractéristiques de chacune des versions éditoriales définies dans l’édition de la *Queste del saint Graal*<sup>9</sup>:

Type de données	Version courante	Version diplomatique	Version fac-similaire
Abréviations médiévales	résolution tacite	résolution avec mise en italique des caractères ajoutés	reproduction de la marque d’abréviation
<i>x</i> final pouvant correspondre à <i>us</i>	maintien de <i>x</i>		
<i>u/v</i> ; <i>i/j</i> (voyelle/consonne)	<i>u, i</i> = voyelles <i>v, j</i> = consonnes	identique au manuscrit (pas de point suscrit)	
Séries de jambages	interprétation		
Distinction <i>ci/ti, ct/tt</i>	interprétation		
Diacritiques modernes : <i>e/é/ë, c/ç</i>	usage de diacritiques dans des cas précis	absence totale de diacritiques	absence de diacritiques modernes
Diacritiques médiévaux <i>í, ý</i>	suppression		maintien
Variante calligraphiques : <i>s</i> ‘long’ ; <i>r</i> ‘rond’ ; <i>n</i> et <i>m</i> finaux, etc.	neutralisation		maintien
Ligatures	séparation		maintien partiel
Grandes initiales (lettrines)	un nouveau paragraphe à chaque lettrine, la taille et la couleur ne sont pas prises en compte		représentation de la taille et de la couleur de l’initiale
Ponctuation « syntaxique »	- un point pour un point du manuscrit suivi d’une majuscule - une virgule pour un point suivi d’une minuscule - .’ pour un <i>comma</i>		maintien des marques médiévales

<sup>9</sup> Une spécification précise de la transcription diplomatique pratiquée dans les projets d’édition de sources associés à la Base de français médiéval peut être consultée dans Lavrentiev (2008, 13) et Marchello-Nizia / Lavrentiev (2013, Introduction, 43-44).

Ponctuation du discours direct	- usage de guillemets, de tirets et de ponctuation forte au début et à la fin du discours direct et de virgules autour des incises - ?? pour un point d'interrogation médiéval - ajout d'un point d'interrogation si nécessaire	- mêmes règles que pour la ponctuation « syntaxique », aucune ponctuation supplémentaire n'est introduite - ? pour un point d'interrogation médiéval	maintien des marques médiévales, aucune ponctuation supplémentaire
Segmentations particulières (agglutinations et déglutinations)	normalisation en général maintien dans les locutions en cours de figement		maintien total
Corrections sribales	texte corrigé	texte corrigé	maintien des marques de correction

3.2.1. Version 'courante'. – Une fois posé le principe d'une fidélité maximale au manuscrit de base, nous suivons dans la version courante les pratiques habituelles concernant la transcription des mots (lettres 'ramistes', signes diacritiques, résolution des abréviations, déglutination de certains groupes de mots). Mais nous innovons un peu en ce qui concerne la ponctuation (tous les signes du copiste et son usage des majuscules sont respectés, les interventions éditoriales sont très limitées et spécifiées en Introduction), les corrections (les auto-corrrections sribales sont indiquées, les corrections éditoriales très limitées toutes listées en Introduction) et le marquage du discours direct (balisé et surligné).

3.2.2. Version diplomatique. – La transcription diplomatique se distingue de la version courante par une représentation encore plus proche de la source primaire. Dans sa visualisation, la transcription diplomatique peut se rapprocher d'une certaine tradition d'édition de sources documentaires et historiques, ce qui permet d'éviter les difficultés de lecture et d'interprétation. Du point de vue linguistique, cette transcription peut être caractérisée comme 'graphématique' (par opposition à 'allographétique') même si les graphèmes des abréviations n'y sont pas représentés directement.

On peut dire que la transcription diplomatique présente un compromis raisonnable entre, d'une part, la précision dans la représentation des données primaires et, d'autre part, la lisibilité et la facilité de production et de traitement. Elle permet de savoir pour chaque caractère s'il provient de la source primaire ou s'il a été ajouté par l'éditeur. Cela permet notamment d'utiliser ce type de transcription dans les outils d'alignement fin (au niveau du caractère) entre le texte et l'image, ce qui peut servir à des recherches paléographiques et au développement de technologies de reconnaissance optique de caractères<sup>10</sup>. L'absence des diacritiques et des lettres 'ramistes' met en évidence les ambiguïtés graphiques de l'écriture médiévale, tandis que le respect des majuscules et de la ponctuation originales révèle la structuration du texte perçue par les lecteurs à l'époque de sa production. La transcription diplomatique peut être saisie comme une 'couche de balisage' de la version courante qui demande relativement peu d'effort supplémentaire de la part de l'éditeur.

3.2.3. Version facsimilaire (allographétique). – Cette transcription représente tous les traits potentiellement significatifs du système graphique de la source primaire.

Dans l'état actuel des technologies disponibles, la production de la transcription allographétique demande un investissement considérable de la part du transcripateur et alourdit le traitement informatique. Mais des échantillons restreints sont souvent suffisants pour les recherches qui s'intéressent aux distinctions fines des variantes de caractères. Pour ces raisons, la transcription allographétique est toujours facultative dans nos projets éditoriaux et se limite à quelques folios pour des textes longs (par exemple, 2 folios dans l'édition de la *Queste del saint Graal*).

<sup>10</sup> La transcription diplomatique de notre édition de la *Queste del saint Graal* a ainsi largement été utilisée dans le projet ANR ORIFLAMMS (resp. D. Stutzmann, <http://oriflamms.hypotheses.org>) qui vise à établir une ontologie des écritures médiévales basée sur un corpus de transcriptions alignées aux images de manuscrits.

### 3.3. Usages et accessibilité

3.3.1. Composants de l'édition. – Les éditions de la « Collection BFM » sont dotées de l'apparat habituel : Introduction, offrant les rubriques traditionnelles, Index des noms propres, Glossaire sélectif et Bibliographie. S'y ajoutent les images du manuscrit et, dans certains cas, une traduction du texte en français moderne. Ces éléments sont organisés et mis en ligne dans un portail en accès libre et gratuit.

3.3.2. Outils de visualisation et d'analyse – Les nouvelles technologies permettent désormais un accès médié à un volume de données toujours plus riche et plus grand et l'essor des corpus numériques influe sur les méthodologies de recherche, sur les attentes et le rapport des chercheurs au texte. Les éditions numériques bénéficient de ces évolutions.

Tout comme la *Base de français médiéval*, la « Collection BFM » est rendue accessible au public grâce à la plateforme TXM<sup>11</sup>. Cette plateforme logicielle est dotée d'outils de lecture et de navigation hypertextuelle permettant la visualisation dynamique des différents niveaux de représentation des éditions définis plus haut (affichage côte à côte de deux versions du texte au choix : de l'image du manuscrit et d'un niveau de transcription, ou de la traduction etc.)<sup>12</sup>. Elle intègre par ailleurs une gamme d'outils de recherche et d'analyse très développés (listes de mots et fréquences, concordances, listes de mots cooccurrents, mots statistiquement spécifiques d'un sous-corpus, etc.)<sup>13</sup>. Les possibilités de recherche s'accroissent également grâce à l'annotation réalisée au sein des textes édités (étiquetage morphologique des mots, encodage des passages en discours direct).

3.3.3. Accès libre aux sources. – Il convient de décider dès le départ de la politique d'usage et de rediffusion des données produites. Pour tout projet financé sur des fonds publics, il nous semble indispensable non seulement d'offrir un accès le plus direct possible à la lecture des éditions mais également de garantir un accès ouvert aux données sources (transcription du texte, annotations philologiques, linguistiques, etc.) dans un format standard reconnu internationalement (par exemple, XML-TEI) et sous une licence type reconnue internationalement et applicable par les législations des différents pays d'origine de l'accès aux éditions (par exemple, *Creative commons*)<sup>14</sup>. Ce principe doit être respecté même en cas d'hébergement chez un éditeur privé à qui l'auteur de l'édition n'est en aucun cas obligé de céder les droits d'une façon exclusive (pratique qui freine fortement la constitution de corpus de recherche).

3.3.4. Traçabilité philologique. – Les éditions électroniques sont souvent évolutives : des corrections ou enrichissements, ainsi que des changements de technologie et d'hébergement peuvent survenir régulièrement. Pour cette raison et afin d'assurer la citabilité au sein des publications scientifiques et la reproductibilité des résultats de recherche, il est indispensable de mettre en place un système de versionnage et il est souhaitable de doter chaque édition d'un identifiant pérenne (de type DOI). L'exposition des métadonnées de l'édition dans un entrepôt OAI-PMH permet de la rendre accessible à travers des archives ouvertes ou bibliothèques virtuelles (telles que Gallica ou Europeana). En ce qui concerne la BFM, toutes les métadonnées des textes publiés sur son portail entreront à moyen terme dans un entrepôt OAI-PMH du Consortium CAHIER <<http://www.cahier.paris-sorbonne.fr>>.

### Conclusion

Les réflexions, les propositions et les réalisations présentées dans cet article sont issues d'un travail bien avancé, mais loin d'être terminé. En les publiant en l'état, nous espérons promouvoir, au-delà de telle ou telle réalisation particulière, la mise en place d'une infrastructure internationale d'édition ouverte pour les textes médiévaux.

Par ailleurs, à terme, il n'est pas impossible que les méthodologies et les pratiques du numérique amènent à réévaluer celles des éditions papier et qu'elles les influencent.

UMR 5191 ICAR

Céline GUILLOT

UMR 5191 ICAR

Alexei LAVRENTIEV

Université d'Oxford

Thomas RAINSFORD

UMR 5191 ICAR

Christiane MARCHELLO-NIZIA

<sup>11</sup> Heiden / Magué / Pincemin (2010).

<sup>12</sup> Voir notamment notre prototype : <http://txm.bfm-corpus.org/txm>.

<sup>13</sup> Les fonctionnalités de la plateforme TXM et de notre portail sont décrites en détail dans Bertrand *et al.* (2012) et dans Heiden *et al.* (2013)

<sup>14</sup> La plupart des ressources de la BFM sont accessibles sous la forme de documents PDF et de fichiers sources XML-TEI comportant une licence « CC BY-NC-SA 3.0 France » (attribution, pas d'utilisation commerciale, partage dans les mêmes conditions), <http://creativecommons.org/licenses/by-nc-sa/3.0/fr>.



## 6. Références bibliographiques

- Ainsworth, Peter (ed.), 2007. *Chroniques. Livre III, Le manuscrit Saint-Vincent de Besançon, Bibliothèque municipale, ms. 865*, par Jean Froissart, avec une étude codicologique par Godfried Croenen, Genève, Droz.
- Ainsworth, Peter / Croenen, Godfried (ed.), 2012. *The Online Froissart, version 1.4*, Sheffield, HRIOline. <<http://www.hrionline.ac.uk/onlinefroissart>>.
- Andrieux-Reix, Nelly / Monsonégo, Simone, 1997. « Ecrire des phrases au Moyen Âge. Matériaux et premières réflexions pour une étude des segments graphiques observés dans les manuscrits français médiévaux », *Romania* 115, 289-336.
- Andrieux-Reix, Nelly / Monsonégo, Simone, 1998. « Les unités graphiques du français médiéval : mots et syntagmes, des représentations mouvantes et problématiques », *Langue française* 119, 30-51.
- Bertrand, Lauranne / Lavrentiev, Alexei / Pincemin, Bénédicte / Guillot, Céline / Heiden, Serge, 2012. *Tutoriel TXM pour la BFM. Version 1.0*, Lyon, ENS de Lyon. <[http://txm.bfm-corpus.org/files/Tutoriel\\_TXM\\_BFM\\_V1.pdf](http://txm.bfm-corpus.org/files/Tutoriel_TXM_BFM_V1.pdf)>.
- Combettes, Bernard, 2012. « Linguistique textuelle et diachronie », in: *Actes du 3e Congrès Mondial de Linguistique Française, EDP Sciences* ([www.linguistiquefrancaise.org](http://www.linguistiquefrancaise.org)), [<http://dx.doi.org/10.1051/shsconf/20120100344>].
- Duval, Frédéric, 2006. « La philologie française, pragmatique avant tout ? », in: Duval, Frédéric (ed.), *Pratiques philologiques en Europe*, Paris, Ecole des Chartes, 115-150.
- Gifford Fenton, Eileen / Duggan, Hoyt N., 2006. « Effective Methods of Producing Machine-Readable Text from Manuscript and Print Sources », in: Unsworth, John / O'Brien O'Keefe, Katherine / Burnard, Lou (ed.). *Electronic Textual Editing*, New York, N.Y., Modern Language Association of America, 241-253.
- Groupe de recherches La civilisation de l'écrit au Moyen âge (ed.), 2001-2002. *Conseils pour l'édition des textes médiévaux*, Paris, Comité des travaux historiques et scientifiques, Ecole nationale des chartes, 3 vol.
- Guyotjeannin, Olivier / Lusignan, Serge (ed.), 2011. *Le formulaire d'Odart Morchesne*, Paris, École nationale des chartes. <<http://elec.enc.sorbonne.fr/morchesne/>>.
- Koschwitz, Eduard, 1920. *Les plus anciens monuments de la langue française publiés pour les cours universitaires*, Leipzig, O.R. Resland, 9<sup>e</sup> éd.
- Heiden, Serge / Decorde, Matthieu / Bertrand, Lauranne, 2013. *Manuel de TXM. Version 0.7*, Lyon, ENS de Lyon. <<http://txm.sourceforge.net/doc/manual/manual.xhtml>>.
- Heiden, Serge / Magué, Jean-Philippe / Pincemin, Bénédicte, 2010. TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. In I. C. Sergio Bolasco (Ed.), *Proc. of 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010* (Vol. 2, p. 1021-1032). *Edizioni Universitarie di Lettere Economia Diritto*, Roma, Italy.
- Lavrentiev, Alexei, 2008. *Manuel d'encodage XML-TEI étendu des transcriptions de manuscrits dans le projet BFM-Manuscrits. Version 2.1*, Lyon, ENS de Lyon. <[http://ccfm.ens-lyon.fr/IMG/pdf/BFM-Mss\\_Encodage-XML.pdf](http://ccfm.ens-lyon.fr/IMG/pdf/BFM-Mss_Encodage-XML.pdf)>.
- Lavrentiev, Alexei, 2009. *Tendances de la ponctuation dans les manuscrits et incunables français en prose, du XIIIe au XVe siècle*. Thèse de doctorat en sciences du langage, Lyon, Ecole normale supérieure Lettres et sciences humaines. <<http://tel.archives-ouvertes.fr/tel-00494914/>>.
- Marchello-Nizia, Christiane / Lavrentiev, Alexei, 2013. *Queste del saint Graal. Édition numérique interactive du manuscrit de Lyon (Bibliothèque municipale, P.A. 77)*, Lyon, ENS de Lyon. <<http://txm.bfm-corpus.org/txm>>.
- Noomen, Willem / Nico van den Boogaard (ed.), 1983-1998. *Nouveau Recueil complet des fabliaux*, 10 vol., Assen, Van Gorcum.
- Pauphilet, Alfred, 1921. *La tradition manuscrite et l'établissement du texte de la Queste del Saint Graal, attribuée à Gautier Map*, Paris, Champion.
- Pignatelli, Cinzia / Lavrentiev, Alexei, 2014. *Le Psautier d'Arundel : une nouvelle édition*, in : Lemaréchal, Alain / Koch, Peter / Swiggers, Pierre (ed.), *Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013). Section 13 : Philologie textuelle et éditoriale*, Nancy, ATILF.
- Porter, Dot, 2013. « Medievalists and the Scholarly Digital Edition », *The Annual of the Association for Documentary Editing* 34, 1-26. <<http://www.scholarlyediting.org/2013/essays/essay.porter.html>>.
- Robinson, Peter, 2006. « The Canterbury Tales and other Medieval Texts », in: Unsworth, John / O'Brien O'Keefe, Katherine / Burnard, Lou (ed.). *Electronic Textual Editing*, New York, N.Y., Modern Language Association of America, 74-93.
- Roques, Mario, 1926. « Etablissement de règles pratiques pour l'édition des anciens textes français et provençaux », *Romania* 52, 243-249.

Rychner, Jean (ed.), 1958, Marie de France, *Le lai de Lanval, texte critique et édition diplomatique des quatre manuscrits français*, Genève, Droz, Paris, Minard.