

Revisiter les bibliothèques numériques pour le patrimoine culturel avec la Chanson de Roland.

Experimenting Digital Libraries for Cultural Heritage with the Song of Roland

Jacques DUCLOY (1), Thierry DAUNOIS (2), Frédérique PÉGUIRON (2), Jean-Pierre THOMESSE (2), Isabelle TURCAN (2)

(1) Paragraphe, Université Paris 8
Jacques.Ducloyl@univ-lorraine.fr

(2) Université de Lorraine

Résumé. Fournir un résumé en français. La taille du résumé peut éventuellement être telle que la première page puisse contenir le résumé et les mots-clés, aussi bien en français qu'en anglais.

Mots-clés. Instruction, police, retrait.

Abstract. Fournir un résumé en anglais.

Keywords. Fournir les mots clés en anglais.

1 Introduction

Qu'est-ce qu'une bibliothèque numérique, au juste ?

Il y a 15 ans Carl Lagoze, un des pionniers des archives ouvertes aux États-Unis, posait cette question dans un article de référence (Lagoze 2005).

Alors qu'il les considérait comme entrées dans « leur adolescence » - nous étions en 2005 -, Carl Lagoze soulignait que leur situation était déjà préoccupante, alors que l'idée infusait que « Google a déjà tout résolu ». Et il insistait notamment sur le fait qu'une bibliothèque n'est pas seulement un endroit où on peut retrouver de l'information pour la consulter¹, mais aussi « des lieux où des personnes se rencontrent pour accéder à un savoir qu'ils partagent et qu'ils échangent ». Reprenant l'idée de David Levy (2000), il reliait déjà les bibliothèques aux communautés dont elles sont l'espace de référence.

Mais, quinze ans plus tard, loin d'être reconnue largement, cette idée de communauté, de la dimension "humaine" des bibliothèques, semble toujours s'effacer derrière la technique. Et pourtant !

Le terme même de bibliothèque devrait nous alerter, dans sa dimension éminemment polysémique. Car une bibliothèque, nous dit le TLF², est à la fois un

¹ Ce qui fait référence à deux problématiques, celle de la recherche d'information et celle de son accessibilité, qui constituent encore aujourd'hui des enjeux forts

² Trésor de la langue française, dictionnaire du CNRS.

lieu, pouvant désigner un espace (de différentes natures : « bâtiment [...] où sont déposées, rangées, cataloguées diverses collections de livres », « cabinet de travail [...] qui renferme une collection de livres », « meuble à rayonnages destiné au rangement et au classement de livres ». Elle peut également désigner la collection de livres elle-même (la bibliothèque de Paul Meyer a permis de constituer le fonds du même nom). Enfin, une « bibliothèque vivante » fait directement référence à l'érudit(e) dont la mémoire est particulièrement remarquable.

La transposition dans un espace numérique de ces différentes fonctions pointe à la fois vers des documents numériques, vers des entrepôts de données et de métadonnées et vers la version moderne des *scriptorium* que sont les *learning centers*.

Carl Lagoze concluait son article en signalant que les bibliothèques numériques ne devaient pas être seulement des endroits où trouver de l'information et y accéder, mais devaient également permettre « d'ajouter de la valeur aux ressources internet », un enrichissement lié à leur contextualisation, à leur mise en relation avec de nouvelles informations et par leur imbrication dans des réseaux de relations - modèles d'usage, savoir communautaire, réseaux sémantiques. Ainsi, disait-il, « La bibliothèque numérique devient alors un espace pour l'information collaborative et l'enrichissement ».

Comment ces éléments peuvent-ils constituer un cadre profitable à une œuvre comme la Chanson de Roland ? Quelles sont les conditions de mise en œuvre d'une telle expérimentation, quel retour d'expérience peut-on en tirer ?

2 Un projet autour de la Chanson de Roland

Le 15 août 778, de retour d'Espagne, Charlemagne perd son arrière-garde, attaquée, à titre de représailles, par les troupes des seigneurs Basques dont il a attaqué les possessions. Lors de la bataille de Roncevaux, l'arrière-garde est écrasée, provoquant la mort de nombreux braves de l'entourage de Charlemagne, dont celle de Roland, préfet de la Marche de Bretagne. On peut imaginer, mais la tradition orale en a perdu la trace, que ce fait d'armes ait été l'objet de chansons de geste et d'épopées, qui ont circulé, au gré de l'errance des troubadours, de seigneurie en seigneurie. Quoi qu'il en soit, la légende de Roland (avec par exemple la trahison de Ganelon, le son du cor, ou l'épée Durandal qui brise le rocher) refait surface et s'inscrit matériellement sur parchemin au XII^e siècle, les Basques ayant, pour des raisons probablement opportunistes, laissé la place dans le récit aux Sarrasins.

2.1 Un corpus riche et varié

De la *Chanson de Roland* et de ses transcriptions médiévales, on connaît aujourd'hui sept versions, et trois fragments. La version considérée comme la plus ancienne et la plus proche d'un hypothétique "texte initial" est le manuscrit conservé à la Bibliothèque Bodléienne d'Oxford (Digby, 23, f. 1r-72r). Communément daté du deuxième quart du XII^e siècle, ce manuscrit a suscité plusieurs dizaines d'éditions modernes, depuis le début du XIX^e siècle, a été traduit dans de nombreuses langues, et été l'objet de plusieurs centaines d'études³.

Une analyse même sommaire des versions manuscrites de la chanson de geste permet immédiatement de comprendre la situation. Là où le manuscrit d'Oxford compte 4002 vers répartis en 291 laisses (couplets), la version Venise 4 - datée du XIII^e siècle - en compte 6011, pour 419 laisses, la version de Châteauroux 8201 vers et 449 laisses, le manuscrit Venise 7 rassemble 8395 vers organisés en 445

³ La consultation de la bibliographie proposée sur le site arlima.net est éclairante sur la richesse et la diversité des écrits sur et autour de la Chanson de Roland.

laisses. Les manuscrits de Paris, Cambridge et Lyon, pour leur part, comptent respectivement 6828, 5695 et 2932 vers, répartis en 375, 354 et 216 laisses.

Cette sérieuse diversité pose donc d'entrée de jeu la question de l'alignement des textes. Sans même parler des études, dont nous avons déjà signalé le grand nombre, on comprend aisément qu'avec la Chanson de Roland et ses éditions modernes, on dispose d'un corpus riche à la fois en volume et en complexité.

Mais la création artistique autour des exploits et de la mort du "neveu" de Charlemagne ne s'arrête pas à sa mise en vers ou à ses traductions : elle a également pris la forme de diverses mises en musique, avec là aussi de nombreuses productions au fil des siècles.

2.2 Une expérience pilote en 2014

La bibliothèque universitaire du Campus Lettres et sciences humaines de l'université de Lorraine à Nancy dispose d'un fonds Paul Meyer. Celui-ci, diplômé de l'École des Chartes, philologue et romaniste, spécialiste de littérature romane, a notamment travaillé à la Bibliothèque nationale. Élu au Collège de France en 1876, il prend la direction de l'École des Chartes en 1882. À sa mort, en 1917, il choisit de léguer sa bibliothèque à l'université de Strasbourg, mais, celle-ci étant soumise aux mouvements de frontières que l'Alsace et la Moselle connaissent depuis 1870, c'est la bibliothèque de l'université de Nancy qui est chargée de l'accueillir, par mesure de précaution. C'est ainsi qu'un fonds Paul Meyer, composé de 4222 titres de monographies et d'environ 7700 brochures, tirés-à-art et petites publications.

Dans ce fonds figurent plusieurs éditions de la *Chanson de Roland*, dont certaines sont annotées de la main de Paul Meyer. Nous pouvons montrer qu'il s'agissait là d'un travail préparatoire à la publication d'ouvrages⁴.

En 2014, saisissant l'opportunité d'un stage, Isabelle Turcan confiait à l'un de ses étudiants de la filière "Métiers du livre" la tâche d'explorer et d'analyser l'édition de Francisque Michel de 1869 annotée par Paul Meyer. En effet, sur sept pages du recueil, on retrouve des notes, des indications d'édition, des paperolles... Le travail de l'étudiant a consisté à rééditer sur un site web les sept pages concernées, avec trois principales présentations du texte : le texte initial de Francisque Michel ; le texte avec les annotations de Paul Meyer ; le texte tel qu'il apparaîtrait une fois appliquées les modifications indiquées par Paul Meyer.

À cette occasion, nous avons décidé d'assurer en parallèle la réédition de l'ensemble de l'ouvrage. L'étudiant ayant travaillé sur sept pages, nous nous sommes chargés de cent quinze autres pages, non annotées. Et nous avons profité de l'occasion pour effectuer une expérimentation : en annotant sémantiquement les variantes des noms de Charlemagne et de Roland, nous avons pu construire un système d'information sur les variantes (liste, nombre de pages sur lesquelles chacune est utilisée...). L'ensemble de cette expérimentation a été menée avec un seul document de référence : l'édition de 1869 de la Chanson de Roland par Francisque Michel.

2.3 Une bibliothèque numérique sur la Chanson de Roland

Depuis le printemps 2021, un projet de plus grande envergure se met en place. L'idée est de voir comment explorer et exploiter le corpus décrit précédemment, dans toutes ses dimensions et dans toute sa complexité. En effet, grâce à la

⁴ *Recueil d'anciens textes bas-latins, provençaux et français, accompagnés de deux glossaires*, publié en 1874.

numérisation d'un nombre croissant de documents, à la mise en ligne de ressources, il est désormais possible d'accéder à plusieurs de ces sources.

Tels des copistes de l'époque médiévale dans un scriptorium, ou les membres d'une société savante occupant la salle de travail d'une bibliothèque, l'ambition est de pouvoir travailler sur ces textes, d'observer leurs différences et leurs rapprochements.

Grâce à un travail musical effectué précédemment sur *'Irisb Mass* (messe irlandaise) de Gilles Mathieu, nous avons découvert que ce compositeur avait également composé un oratorio profane sur la base du manuscrit d'Oxford et de sa transcription par Paul Gautier, ce qui a donné l'idée d'effectuer le rapprochement numérique de la partition et de la transcription du manuscrit.

La structure apparente du manuscrit d'Oxford - et que l'on peut aisément imaginer en consultant l'une ou l'autre des transcriptions - s'organise autour de "laises" - des couplets - rassemblant un nombre variable de vers. Chacune contient des vers en assonance, et commence habituellement par une lettrine. Dans le manuscrit d'Oxford, elles se terminent généralement par une mention mystérieuse, sur laquelle aucune explication n'est acceptée largement : [Aoi].

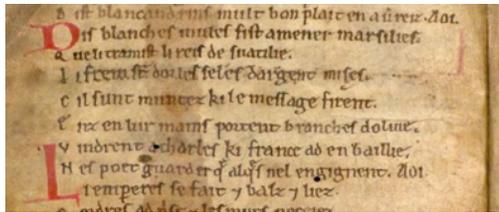


Figure 1. Un enchaînement de 3 laisses, 2 lettrines (D et L) et 2 mentions Aoi en fin de ligne

Lorsque l'on commence à vouloir aligner les textes des manuscrits et leurs transcriptions, on constate rapidement des divergences dans la numérotation des laisses. Ainsi, la dernière laisse du texte est numérotée CCXCI chez Roland Bédier, CCXCIII chez Edmund Stengel, CCXCVI chez Francisque Michel et CCXCVII chez Léon Gautier.

En effet, certains philologues se réfèrent à la différenciation des laisses à l'aide des lettrines et des marques [Aoi] telle qu'elle est dans le manuscrit d'Oxford. D'autres considèrent que le copiste a fait des erreurs qu'ils cherchent à rectifier. Le feuillet 43 verso est exemplaire de ce point de vue car ne contient ni lettrine, ni mention [Aoi]. En revanche, il contient un vers qui marque une charnière essentielle entre deux parties de l'épopée : la mort de Roland.

Morz est Rollant, Deus en ad l'anme es cels.
Roland est mort ; Dieu a son âme dans les cieux.

Le manuscrit contient curieusement un point en guise de lettrine.

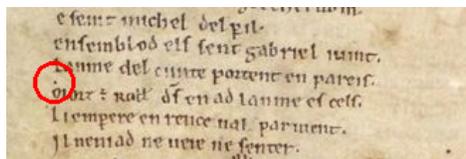


Figure 2. *Un simple point, avant le vers charnière.*

Bédier et Gautier considèrent ce vers comme le début d'une nouvelle laisse. Michel en fait la fin de la précédente et Stengel propose une version sans changement de laisse (et donc avec un décalage dans la numérotation).

Ainsi, et c'est ce qui motive cet article, ce travail s'est avéré à la fois plus complexe et plus riche que nous l'imaginions.

3 Expérimentation, difficultés rencontrées et solutions retenues

3.1 Une organisation à définir

Nous avons déjà eu l'occasion de procéder à des rééditions numériques, et n'avons donc pas été surpris par la première question qui se pose lorsque l'on s'attaque à ce type d'exercice : celle du choix d'une structure éditoriale et informationnelle. En effet, depuis pratiquement 2200 ans, l'organisation classique des codex, puis des livres, longtemps reprise pour les fichiers numériques simples est celle d'un assemblage de feuillets, dans lequel on tourne des pages, avec la possibilité de feuilleter.

Cette organisation avait succédé à près de 300 ans durant lesquels le format classique était celui de la page, qu'il s'agisse de tablettes d'argile ou de papyrus.

1985 marque un premier tournant, avec l'apparition du format SGML, suivi, 10 ans plus tard, par XML, qui se caractérisent par une architecture arborescente. Mais le changement n'est en général pas perceptible pour les utilisateurs.

En parallèle, et au rythme des innovations technologiques, une nouvelle approche se développe, celle de l'hypertexte, dans laquelle « l'utilisateur navigue d'information en information par un jeu de liens d'associations entre les îlots d'informations » (Vignaux 2001). Cette nouvelle structuration s'articule autour de blocs de textes liés entre eux de manière non séquentielle, qui modifie fondamentalement le "parcours" de lecture : de linéaire il devient non-linéaire.

Dans un hypertexte, donc, l'unité n'est plus la classique page : elle est choisie - négociée ! - au cas par cas. À l'inverse, une page donnée peut contenir très peu d'information, ou, à l'inverse, l'équivalent d'un livre entier. De façon très prosaïque, le premier choix est celui du découpage du texte : quelle est l'unité élémentaire la plus pratique à manipuler ? Ici, la laisse nous a semblé être la bonne unité.

3.2 Modéliser l'arborescence des sources

Le deuxième questionnement qui intervient très rapidement, c'est celui de la façon de traiter l'arborescence des sources. Et le phénomène décrit par Carl Lagoze trouve ici une illustration remarquable : certes, depuis 20 ans, nombre de ces sources ont été numérisées, et sont désormais trouvables et accessibles sur le web. Mais cela ne forme pas une bibliothèque pour autant, du fait de l'hétérogénéité des formats et des protocoles sous lesquels elles sont disponibles.

Ainsi, le manuscrit d'Oxford est accessible - en format photo - dans son intégralité sur Wikimedia Commons, avec une organisation séquentielle. Pour le manuscrit de Châteauroux, seule la première page est accessible avec un fac-similé de bonne qualité à l'IRHT, mais les autres pages, accessibles via le site des bibliothèques de Châteauroux, sont encombrées par une inscription de propriété. En fait presque chaque manuscrit (Venise, Cambridge) dépend de son propre service de visualisation. Il n'est pas trivial d'atteindre à un traitement identique des différents manuscrits.

De façon parallèle, pour les sources du XIXe et du XXe siècles, trois principales sources permettent d'accéder au texte : Gallica, Internet Archive et WikiSource. La qualité de la numérisation et la performance des logiciels d'océrisation employés varient sensiblement entre Gallica et Internet Archive. Sur WikiSource, les documents sont relativement faciles à récupérer en texte intégrale, compte tenu du fait qu'ils ont déjà fait l'objet d'un travail par les contributeurs (humains !) du site.

3.3 Gestion des manuscrits

De façon évidente, la priorité a été donnée à la gestion des sources primaires (manuscrits), leurs transcriptions et leurs traductions. En effet, la plupart des articles plus récents contiennent des références à ces documents, souvent sous la forme de numéro de vers ou de numéro de laisse (ni les uns ni les autres n'existant dans les manuscrits).

À l'occasion d'un stage, un alignement a été tenté entre la version de Francisque Michel (1869) et le manuscrit d'Oxford. De premiers travaux d'alignement entre l'oratorio de Gilles Mathieu et le manuscrit. Dans les deux cas, des divergences ont été observées, qui n'ont pas été solutionnées, bien au contraire, en faisant appel à la version de Léon Bédier.

En même temps, l'exploration des sources a mis en évidence un ouvrage d'Edmund Stengel dans lequel la pagination suit le découpage en laisses du manuscrit d'Oxford.



Figure 3. Le haut du feuillet 41 aligné entre le manuscrit et l'ouvrage de Stengel.

Nous avons donc décidé de gérer les manuscrits en nous appuyant sur les laisses avec notre propre numérotation.

3.4 Le choix de l'outil : un hypertexte collaboratif capable de supporter des traitements complexes

Plus la complexité et la richesse du corpus apparaissait, plus il est devenu évident que l'outil à utiliser devait permettre de naviguer dans un hypertexte de façon collaborative. Le choix s'est rapidement porté sur la technologie des wikis sémantiques (Semantic MediaWiki), que nous avons déjà eu l'occasion d'expérimenter avec le projet Wicri. Concrètement un wiki « Wicri/Chanson de Roland » a été créé au sein du réseau.

La gestion du manuscrit d'Oxford est basée sur une première structure hypertexte basée sur les feuillets. A chaque feuillet est associé une page wiki qui est généralement organisée en 3 parties :

- Pour le recto, l'association entre la fac-similé de la page du manuscrit et la transcription de Stengel (les liens sur les images sont actifs et permettent des navigations parallèles) ;
- même chose pour le verso ;

- la liste des laisses (avec des liens) avec notre une numérotation.

Pour chaque laisse, une page wiki permet de retrouver le ou les feuillets dans lesquels elle est contenue.

Elle contient également un ensemble d'informations permettant de confronter les points de vue (notamment diverses transcriptions ou traductions).

3.5 Les interactions entre un oratorio profane, les manuscrits et les traductions

Pratiquement tous les documents offrant des analyses du manuscrit d'Oxford sont organisés autour des laisses. Les pages wikis affectées à des laisses jouent donc un rôle essentiel, constituant la colonne vertébrale de l'hypertexte associé à un manuscrit.

Pour constituer son oratorio, Gilles Mathieu s'est appuyé sur la transcription de Léon Gautier pour structurer son oratorio. Il a opéré un découpage du texte et de l'histoire en dix mouvements, proches de la mise en chapitre de l'ouvrage (exemple : La mort de Roland, ou la cité sur la colline qui correspond au conseil tenu par Marsile à Saragosse). Il a ensuite sélectionné quelques vers significatifs pour les mettre en musique. Cette musique donne un éclairage particulier aux couplets concernés. La réédition de l'oratorio va donc contenir des liens vers les laisses concernées (avec parfois un décalage de numérotation).

Réciproquement, dans chaque laisse concernée, le thème musical est explicité par une ligne mélodique.

La même approche est utilisée pour les traductions et transcriptions. Les textes sont souvent dupliqués pour être plus facilement accessibles aux lecteurs. Mais, de fait, un ouvrage donné est découpé en environ 300 fragments.

Ainsi, les pages associées aux laisses gèrent les interactions, avec les contenus suivants :

- des liens vers le ou les feuillets concernés,
- la liste des numérotations dans les divers ouvrages,
- quelques versions significatives, éventuellement avec les notes associées,
- s'il y a lieu, un extrait musical.

Un blog, installé sur le wiki, et intitulé « dialogue avec un compositeur », permet d'échanger avec Gilles Mathieu sur ses choix musicaux ou sa perception de l'épopée.

4 Bibliographie

Lagoze, C. (2005)

Levy, D. (2000), Digital Libraries and the Problem of Purpose, *Bulletin of the American Society for Information Science*, 26 (6), 2000.