

Avec la *Chanson de Roland*, le numérique pour le bonheur des choristes amateurs et humanistes.

Jacques Ducloy

Université Paris 8, Laboratoire Paragraphe, F-93200 Saint-Denis, France

Mots-clés

Musicien amateur - Chanson de Roland – Époque carolingienne - Wiki sémantique – Manuscrits – Bibliothèque numérique – Édition critique – Semantic MediaWiki – Musique

Résumé

Cet article présente deux bibliothèques numériques (hypertextes) qui ont dans leur public des musiciens amateurs et des mélomanes. Elles ont été créées dans un contexte académique par des chercheurs qui explorent les nouvelles pratiques induites par l'appropriation des outils numériques avancées par les chercheurs. Les outils étudiés sont les wikis sémantiques (sous MediaWiki le moteur de Wikipédia), l'ingénierie XML et les techniques statistico-linguistiques d'exploration de corpus. Ici les chercheurs associés sont des linguistes, des philologues, des archéologues. Il y a bien sûr les musicologues qui peuvent valoriser leurs travaux auprès des choristes ou des mélomanes. La première bibliothèque est une collection d'ouvrages déposés dans une encyclopédie sur la musique. La deuxième est un atelier de travail sur la Chanson de Roland et les légendes de Charlemagne. Les philologues y manipulent des ouvrages très hétérogènes par leur nature (manuscrits, partition, éditions critiques, dictionnaires) ou par leurs approches scientifiques.

Abstract

This article introduces two digital libraries (hypertexts) whose audience includes amateur musicians and music lovers. They were created in an academic context by researchers who explore the new practices induced by the appropriation of advanced digital tools by researchers. The tools studied are: semantic wikis (under MediaWiki, the Wikipedia engine), XML engineering and statistical-linguistic tools for corpus analysis. The associated researchers include linguists, philologists, archaeologists. The musicologists can promote their work to choristers or music lovers. The first library is a collection of books or articles deposited in an encyclopedia on music. The second is a workshop on the Chanson de Roland and the legends of Charlemagne. The philologists handle very heterogeneous works there by their nature (manuscripts, score, critical editions, dictionaries) or by their scientific approaches.

Avant-propos

Cet article fait partie des expérimentations décrites dans cet article. Il a été initialement rédigé sur le wiki « Wicri/Chanson de Roland¹ ». Il est également disponible, toujours en format numérique augmenté, sur le wiki « Wicri/Musique² ».

Les lecteurs y trouveront le même contenu, complété par un glossaire intégré dans une encyclopédie explorable par des liens sémantiques. Ils y trouveront des images actives ou des partitions qu'ils pourront écouter. Les liens sont personnalisés sur chaque wiki (Par exemple, un lien sémantique³ sur « Wicri/Musique » peut être converti en lien interwiki⁴ sur « Wicri/Chanson de Roland »).

De nombreuses notes de bas de page de cet article sont remplacés par des liens dans les versions numériques.

1 Introduction

L'usage du numérique se généralise dans les chorales, au cœur de la pratique musicale en amateur⁵, et dans leurs associations. Deux usages sont généralement privilégiés par les chefs de chœur : l'apprentissage des choristes et la promotion des concerts.

Les sites web sont alors souvent développés par des informaticiens, amateurs non rémunérés, pour l'apprentissage de chanteurs, eux-mêmes amateurs non rémunérés, avec un public souvent composé d'amateurs qui ne payent pas leur entrée. Cette description volontairement caricaturale n'empêche pas que ces sites web soient souvent d'excellente qualité, mais vise seulement à souligner que, dans ce milieu, l'amateur a un statut non négligeable.

L'amateur joue un rôle fondamental dans le monde du numérique. L'aventure Wikipédia montre qu'un ensemble d'amateurs non rémunérés peut effectivement rivaliser avec des institutions centenaires.

Nous avons voulu tester l'applicabilité des « pratiques wikipédiennes » dans un contexte professionnel, pour diffuser des informations scientifiques complexes et souvent très spécialisées. Pour des raisons de formation et d'expérimentation, nous avons cherché des sujets compréhensibles par un large public. La musique est devenue un thème privilégié pour former des scientifiques. Ils mènent alors des expérimentations numériques dans un domaine où ils sont des amateurs.

Les choristes amateurs savent que la qualité d'une bonne interprétation implique naturellement la compréhension du sens de la pièce chantée. Le texte d'une pièce musicale devient alors un cadre

¹ < https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/Article_Musamat_2022 >

² < https://wicri-demo.istex.fr/Wicri/Musique/fr/index.php?title=Article_Musamat_2022 >

³ Un lien qualifié est un lien complété par un attribut pour aller sur une page du même wiki , voir plus loin paragraphe 2.2.

⁴ Un lien dit interwiki pointe vers une page située sur un autre wiki du réseau.

⁵ Sur les versions numériques, le terme amateur est relié aux articles correspondants du dictionnaire Trésor de la langue française.

pour introduire, de façon collaborative des commentaires. Dans les cas simples (comme une chanson d'amour), une explication relativement simple peut suffire.

Mais par exemple pour les chants religieux, des problèmes générationnels se posent. Quatre-vingt pour cent des plus de 75 ans ont chanté *Kyrie eleison* tous les dimanches. Mais depuis le Concile Vatican II, et avec le recul du christianisme, cette invocation n'a plus aucun sens intuitif pour les jeunes générations. Le travail de vulgarisation nécessaire à la compréhension des faits religieux peut devenir conséquent.

Dans cet article, nous allons présenter une expérimentation numérique qui associe des stratégies professionnelles avec des pratiques d'amateur.

Plus précisément, nous décrirons la démarche adoptée par projet Wicri qui allie des technologies d'analyse de corpus avec l'écriture numérique. Puis nous montrerons des applications générales sur un wiki dédié à la musique, et notamment pour les amateurs. Enfin nous présenterons une bibliothèque dédiée à l'univers de la Chanson de Roland où la réédition hypertexte d'une œuvre contemporaine (composée par Gilles Mathieu) permet à l'amateur d'explorer un immense paysage, historique, linguistique et... imaginaire...

2 Avec le projet Wicri, la connaissance scientifique pour la société

Cette étude se déroule dans le cadre du projet Wicri (Wikis pour les **communautés** de la recherche et de l'innovation) que nous allons rapidement présenter.

Notre projet étudie des mécanismes pouvant contribuer à restaurer la position française sur la diffusion numérique de la culture et l'information scientifique vers la société. Rappelons quelques références historiques : 1946, création des bulletins signalétiques du CNRS par Jean Wyart ; 1960, création du CRTLF⁶ par Paul Imbs ; 1974, informatisation de la base Pascal avec Nathalie Dusoulier ; 1989, création de l'INIST.

Depuis 1995, la production du dictionnaire TLF est arrêtée. De même la production des bases Pascal et Francis a été abandonnée. En même temps, Wikipédia a démontré la puissance d'un réseau d'amateurs pour produire des encyclopédies et des dictionnaires. Du côté des bases de données, aux Etats-Unis, la National Library of Medicine a doublé sa production d'analyses bibliographiques.

Nous avons donné des éléments d'analyse sur le décrochage de la position française (Ducloy 2019). Une des raisons est l'inadaptation des chaînes de production tayloriennes basées sur le contrôle a priori face aux pratiques coopératives, basées sur la confiance, et démontrées par l'expérience Wikipédia. Nous pensons que les chercheurs de multiples disciplines dans leurs diversités (ici, la musicologie, mais aussi l'histoire, la linguistique, la philologie) ont intérêt à interagir avec leurs publics potentiels (ici les musiciens et le public amateur).

⁶ Le Centre de recherche pour un Trésor de la langue française est le laboratoire où a été créé le dictionnaire Trésor de la langue française (TLF).

2.1 Le réseau Wicri

Le projet Wicri a démarré en 2008. Notre objectif était donc de sensibiliser les organismes de recherche aux nouvelles pratiques numériques appliquées à la connaissance scientifique, technique ou culturelle. Pour voir comment appliquer cette approche à un milieu professionnel nous avons utilisé MediaWiki, le moteur de Wikipédia.

Dans un contexte régional, nous avons d'abord créé un ensemble de wikis institutionnels pour aider les chercheurs et leurs unités à faire connaître les retombées de leurs recherches. Puis, rapidement, nous avons développé des wikis sur une base thématique, au départ les sciences du génie de l'environnement. Nous avons alors rencontré des difficultés pour construire un matériel de formation destiné à des contributeurs experts dans des thématiques très spécialisées. En effet, l'analyse des résultats obtenus dans une des branches des sciences du vivant demande souvent des connaissances très pointues (par exemple en génomique). Nous avons donc cherché un domaine d'application plus universel pour monter des programmes de formation pour un public élargi. La musique est ainsi devenue un espace d'expérimentation important pour l'équipe Wicri.

Pour naviguer dans le réseau des wikis, les utilisateurs se repèrent avec des logos qui identifient soit une région, soit une thématique. Une carte (reproduite sur tous les wikis) permet de s'orienter (figure 1).



Figure 1 : le réseau Wicri en 2022⁷

L'organisation générale est la suivante :

- La première ligne donne accès à des wikis techniques destinés aux contributeurs. On y trouve au centre un logo qui pointe vers le wiki d'accueil du réseau Wicri.

⁷ Cette carte est active dans la version numérique.

- Les deux lignes suivantes donnent accès à des wikis thématiques (par exemple, la musique avec une clé de sol).
- Les dernières lignes donnent accès à des wiki régionaux.

Les aspects institutionnels peuvent ainsi se décliner à plusieurs niveaux :

- **Ville.** Nous avons un seul exemple pour l'instant : Nancy. Toutes les chorales du Grand Nancy peuvent y avoir une page de présentation.
- **Région.** La région est le niveau privilégié pour présenter une université et ses laboratoires. Les chorales pérennes bénéficiant d'une bonne visibilité peuvent présenter leurs manifestations significatives. A l'heure actuelle, suite à des travaux avec l'espace européen de la Grande Région, les entités concernées (par exemple, le Luxembourg) y ont des wikis déployés sur 3 langues (français, anglais, allemand).
- **Pays.** Les régions qui n'ont pas encore de wiki propre sont « hébergées » à ce niveau.
- **Continents.** De même, la plupart des pays sont implantés à ce niveau. Un wiki, créé fin 2022, est dédié à la francophonie. Il héberge une expérimentation sur le Trésor de la langue française (TLF).

Les contributeurs sont inscrits avec un mécanisme de parrainage. Ils peuvent intervenir sur tous les wikis dits communs. Par exemple une chorale présentant un programme sur un wiki régional (exemple Nancy) peut décrire l'œuvre sur le wiki dédié à la musique.

Des wikis peuvent être affectés à des communautés scientifiques (exemple H²PTM pour les colloques de même nom). Ils sont alors gérés par ces communautés qui peuvent ouvrir leur espace à l'ensemble des contributeurs.

Enfin, il est possible d'ouvrir des wikis privés en accès restreint.

2.2 MediaWiki, un atelier flexible pour gérer les données et les connaissances scientifiques dans leurs diversités

Comme déjà mentionné, ces wikis utilisent le moteur MediaWiki (celui de Wikipédia et de la galaxie des sites de la Wikimedia Foundation).

Un point fort de MediaWiki est la puissance de ses mécanismes de programmation. En effet, des modèles disposant de commandes computationnelles sont directement utilisables par les contributeurs. Ceux-ci deviennent alors totalement autonomes pour développer des environnements très spécialisés. Par exemple, sur Wikipédia, l'ensemble des outils dédiés à la géographie administrative (incluant la cartographie) a été réalisée par un réseau de contributeurs.

Ce mécanisme introduit une contrainte. En effet, les contributions doivent s'exprimer avec du *wikitexte*, ce qui implique un rapport familier avec les codes informatiques. Par exemple pour mettre une expression en italique, il faut l'encadrer par deux paires d'apostrophes. Voici comment coder « Roland récite son *mea culpa* ».

```
Roland récite son ''mea culpa''
```

Une fois passées les premières expériences syntaxiques, le contributeur curieux va découvrir un vaste espace de possibilités. L'étape suivante est la maîtrise des appels de modèle. Voici comment coder « La *Neuvième symphonie* de Beethoven est une pièce pour CHŒURS ET ORCHESTRE ».

```
La ''Neuvième symphonie'' de Beethoven est une pièce
pour {{Petites capitales|Chœurs et Orchestre}}
```

Avec un complément de formation, le contributeur pourra lui-même développer de nouveaux modèles, et par exemple réaliser des bandeaux pour naviguer dans un dossier numérique.

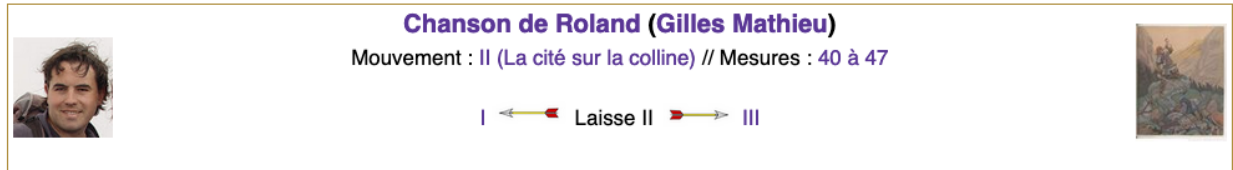


Figure 2 - Bandeau pour naviguer dans les extraits musicaux illustrant les laisses de la Chanson de Roland.

Avec des modèles, le spécialiste d'un domaine d'application, qui se forme au numérique, devient ainsi totalement autonome pour gérer, sur un même espace, des données, même spécialisées, (ici par exemple des partitions) et des articles. Cette publication en est un exemple.

2.3 Semantic MediaWiki, un outil pour organiser les connaissances

Une fois les données rangées, plusieurs mécanismes permettent d'organiser les connaissances sur un wiki.

De façon assez classique pour un documentaliste, les pages peuvent être indexées par des catégories qui vont constituer une arborescence (thésaurus). De façon complémentaire nous utilisons des extensions sémantiques (Semantic MediaWiki) qui permettent une structuration avancée des connaissances. Voici un exemple. Sur la page décrivant les Noces de Figaro, la première phrase est la suivante :

Les Noces de Figaro est un opéra de [Wolfgang Amadeus Mozart](#), sur un livret de [Lorenzo da Ponte](#). Il est inspiré de la comédie de [Beaumarchais](#), *Le Mariage de Figaro*.

Sur Wikipédia, cette phrase serait codée avec des liens simples comme ceci :

```
''''Les Noces de Figaro'''' est un opéra de [[Wolfgang Amadeus Mozart]],
sur un livret de [[Lorenzo da Ponte]].
```

```
Il est inspiré de la comédie de [[Pierre-Augustin Caron de
Beaumarchais|Beaumarchais]], ''[[Le Mariage de Figaro]]''.
```

Sur Wicri, avec les extensions sémantiques, chaque lien peut être qualifié par un attribut. Le wikitexte devient :

```
''''Les Noces de Figaro'''' est un opéra de [[A pour
compositeur::Wolfgang Amadeus Mozart]], sur un livret de [[A pour auteur
de livret::Lorenzo da Ponte]].
```

Il est inspiré de la comédie de [[A pour auteur cité::Pierre-Augustin Caron de Beaumarchais|Beaumarchais]], ''[[Le Mariage de Figaro]]''.

Un attribut comme « A pour auteur de livret » peut être introduit et défini à n'importe quel moment par un contributeur. Il peut servir à naviguer sur une propriété et surtout à exprimer des requêtes. Par exemple, la page Lorenzo da Ponte contient le code suivant :

2.4 Une boîte à outil pour explorer des corpus

En complément des wikis sémantiques, nous utilisons une autre approche technologique : une boîte à outils XML pour développer des systèmes de recherche d'information spécialisés (nommés serveurs d'exploration) sur de grandes masses d'information scientifique. Ces travaux sont déjà anciens, (Ducloy 91) pour les premiers essais sur des corpus de métadonnées homogènes⁸.

Un soutien du projet ISTE⁹ nous a permis d'adapter DILIB (nom de cette bibliothèque XML) au traitement corpus de documents en texte intégral.

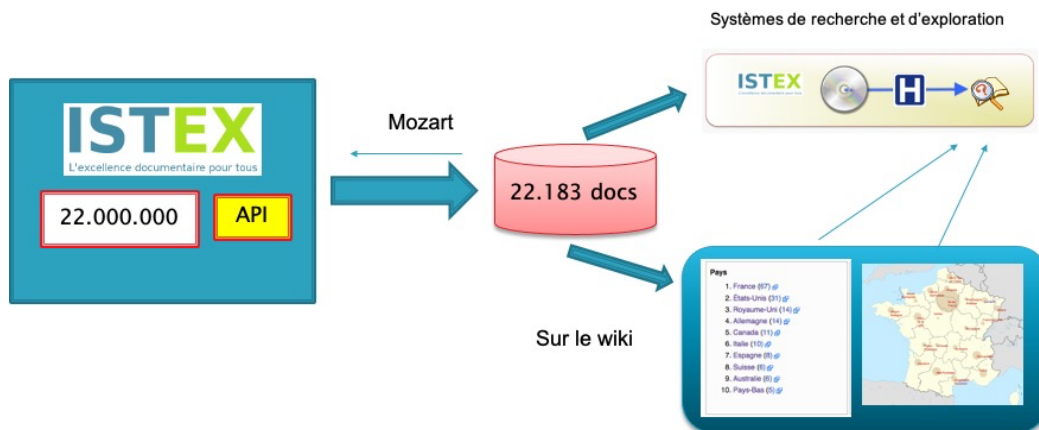


Figure 3 : Serveur d'exploration

Par exemple, au début du projet ISTE⁹, (son réservoir contenait 22 millions de documents), une requête avec le terme « Mozart » a permis d'extraire 22.000 documents. La plateforme DILIB a généré en quelques heures, un système de recherche d'information doté de protocoles statistiques d'exploration. Les wikis sémantiques sont utilisés à la fois comme une interface homme-machine performante, qui permet la navigation dans le corpus, mais également pour héberger les données nécessaires à la curation des données durant le processus.

Sur le wiki « Wicri/Chanson de Roland », deux serveurs sont en cours de développement, un sur la Chanson de Roland et l'autre sur le philologue Paul Meyer¹⁰.

⁸ Pour les spécialistes, notices Pascal en format ISO 2709.

⁹ ISTE⁹ (information scientifique et technique d'excellence) projet issu du programme Investissement d'Avenir.

¹⁰ Voir sur le wiki :

https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/Serveur_d%27exploration_sur_la_Chanson_de_Roland

2.5 Rééditions hypertextes, des articles scientifiques au palais ducal de Nancy

Dans la galaxie Wikipédia, la bibliothèque WikiSource offre plus de 350.000 ouvrages libres et gratuits en français. Un ouvrage peut donc être réédité avec MediaWiki et nous avons également réédité en mode hypertexte de documents de référence. Les extensions sémantiques nous ont permis d'associer ces documents avec un substrat encyclopédique.

Par exemple, autour de l'innovation scientifique en Lorraine, nous avons réédité le contrat de plan État Région. Puis, pour modéliser des systèmes de recherche, nous avons réédité des appels à communication de congrès prestigieux. Enfin, nous avons réédité des ensembles d'articles scientifiques (sur des wikis associés à des colloques).

Une action régionale autour de la Renaissance en Lorraine nous a amené à rééditer un ouvrage de 1852 sur le Palais ducal de Nancy¹¹. À partir d'un original en mode « image + OCR » sur Gallica, nous avons notamment montré comment transformer en hypertexte une gravure de fin de volume (fig. 4). Elle contenait des liens, matérialisés par des lettres, qui pointaient vers une vingtaine de paragraphes qui eux-mêmes renvoyaient vers des pages du livre.

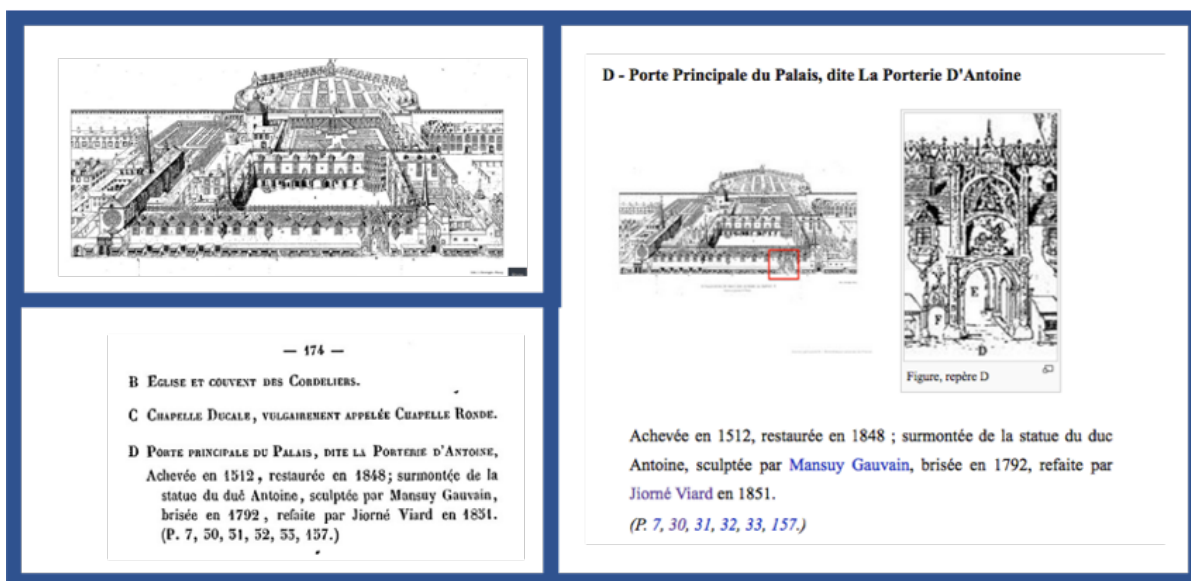


Figure 4. Le Palais ducal : à gauche la gravure et une rubrique (D) avec des renvois ; à droite, le développé de la rubrique D en hypertexte

Cet exemple a inspiré d'autres actions comme le traitement de l'édition de Francisque Michel de La Chanson de Roland (1869), annotée par le philologue Paul Meyer. Cette expérience a été déterminante pour la création ultérieure du wiki Wicri/Chanson de Roland.

¹¹ Le Palais ducal de Nancy (1852) par Henri Lepage. L'ouvrage numérisé est visible ici : [https://wicri-demo.istex.fr/Wicri/Europe/France/GrandEst/Lorraine/Nancy/fr/index.php/Le_Palais_ducal_de_Nancy_\(1852\)_Lepage](https://wicri-demo.istex.fr/Wicri/Europe/France/GrandEst/Lorraine/Nancy/fr/index.php/Le_Palais_ducal_de_Nancy_(1852)_Lepage)

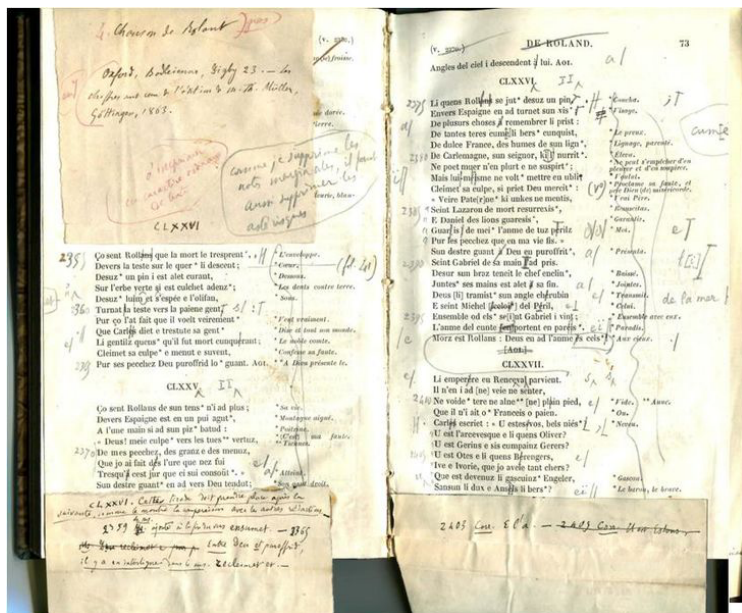


Figure 5 : exemple d'annotation

L'ouvrage ayant été numérisé, le texte et les annotations ont été transcrits en mode wiki¹².

3 Le projet Wicri, une bibliothèque numérique encyclopédique au service des choristes

Les choristes amateurs interprètent des œuvres variées qui éveillent leur curiosité mais leur posent de nombreux problèmes d'interprétation.

Une première barrière est directement liée à la langue et à ses conséquences immédiates sur la prononciation. Ce problème se règle, parfois difficilement, lors des premières répétitions. Parfois un choriste qui parle couramment la langue en question va pouvoir aider ses collègues. Mais dans certaines situations complexes, par exemple, avec les passages en swahili du *Jubilate Deo* de Dan Forest, il ne reste qu'à espérer que personne ne comprendra cette langue dans le public...

Dans des situations apparemment plus simples comme des pièces sacrées en latin, la prononciation peut dépendre de la période à laquelle ces œuvres ont été écrites.

Les pages du wiki Wicri/Musique vont donc contenir des extraits de partition mais surtout des textes et des images pour améliorer l'interprétation et répondre à la curiosité des chanteurs et des auditeurs.

¹² Voir par exemple, sur le wiki, la page 72 de l'ouvrage annoté.
[https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/FPM,_Chanson_de_Roland_\(1869\)_F._Michel,_page_72](https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/FPM,_Chanson_de_Roland_(1869)_F._Michel,_page_72)

3.1 La musique avec LilyPond

Pour la musique, nous utilisons le logiciel de gravure musicale LilyPond. La musique y est codée dans un langage formel dont la syntaxe rappelle, pour les mathématiciens, celle de TeX. Voici par exemple les premières notes du thème « *Au clair de la lune* » en si bémol majeur.



Figure 6. *Au clair de la lune* en Lilypond

Les notes sont codées par des lettres (comme dans la notation anglaise) en démarrant par le « la ». Elles sont suivies éventuellement d'une altération, puis d'une indication de durée. Ainsi « d2 » désigne un ré sur une blanche ; et « bes4 » désigne un si bémol sur une noire.

Ce logiciel est développé par une communauté internationale dans la dynamique des logiciels libres. Il permet de produire des partitions de haute qualité. Mais, comme pour l'écriture dans un wiki, son utilisation demande une sensibilité informatique pour manipuler les commandes.

Dans notre cas, un atout déterminant est son intégration au monde MediaWiki. En effet, une extension permet de l'utiliser de façon coopérative et interactive. Nous avons réédité des articles comportant des phrases musicales.

Par exemple, Jean-Jacques Rousseau a écrit un dictionnaire en deux volumes. Les exemples sont détaillés dans des planches en fin de deuxième tome. Sur Gallica, le lecteur intéressé par le terme « carillon », le lecteur doit feuilleter le tome 1 (en naviguant dans les copies de pages). Il doit ensuite trouver le tome 2 (dans Gallica), y localiser les planches pour tomber sur la partition relative à l'article.

Sur Wicri/Musique, tout est regroupé sur une seule page (Figure 7). En un clic, le lecteur peut consulter dans la souche encyclopédique une biographie d'un horloger cité dans l'article.

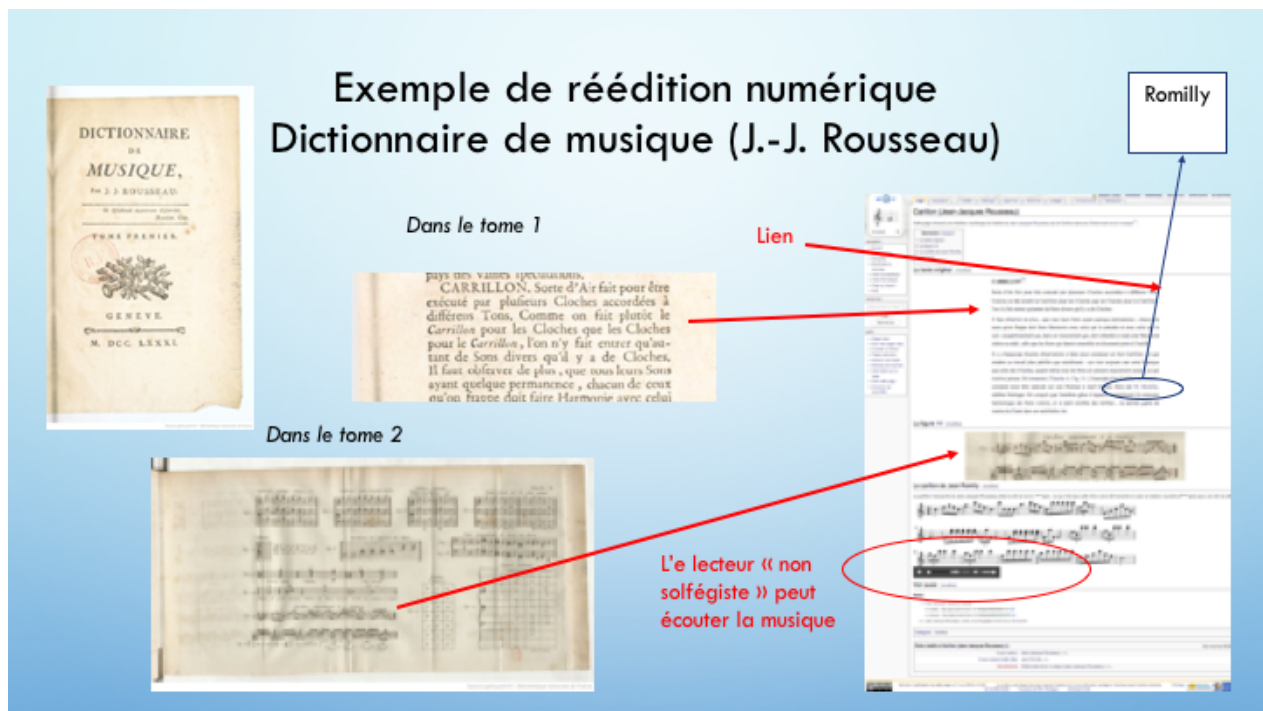


Figure 7 : Diapositive de présentation du dictionnaire de Rousseau

Le lecteur amateur peut être dérouté par une clé de sol en première ligne. En effet, cette notation n'est en effet pratiquement plus utilisée (notamment dans le domaine du chant choral amateur).



Figure 8 : La partition de Jean-Jacques Rousseau

Sur le wiki, la musique est donc retranscrite avec une clé plus actuelle.



Figure 9 - retranscription (avec écoute) de la partition

Et surtout, elle peut être écoutée¹³.

3.2 Apprendre le solfège avec LilyPond en mode wiki

Nous avons vérifié le mode collaboratif, en menant une expérience de découverte du solfège.

En 2021, deux stagiaires (niveau L3 MIASHS) ont rejoint l'équipe Wicri sur 2 projets en parallèle. L'un d'eux, (Léonard Braux) possédait des connaissances solfégiques. Nous lui avons donné pour mission de stage un travail d'adaptation, avec des transcriptions de partitions, de la suite de Gilles Mathieu.

En revanche, la deuxième, Dalila Ladli, travaillait sur l'histoire de l'information scientifique. Elle n'avait pratiquement jamais travaillé sur une partition lors de sa formation initiale en Algérie. Pour lui permettre de suivre le travail de son collègue, sachant qu'elle n'avait pas de difficulté avec la manipulation de langages informatiques, nous avons tenté une expérience d'initiation au solfège.

Pour lui permettre de suivre le travail de son collègue, sachant qu'elle n'avait pas de difficulté avec la manipulation de langages informatiques, nous avons tenté une expérience d'initiation au solfège. Nous avons organisé une transcription de deux chansons populaires, « *Il court, il court le furet* » et « *J'ai du bon tabac* ». Pour la première, Léonard Braux a créé la ligne musicale et nous avons demandé à Dalila Ladli d'aligner les paroles. Pour l'exemple « *Au clair de la lune* » plus haut l'exercice se traduirait par l'ajout de :

```
\addlyrics { Au clair de la lu -- ne, }
```

Puis, avec notre aide, nous lui avons demandé de s'attaquer intégralement à la deuxième.

Cette expérience s'est déroulée sur trois jours, lors de 4 séances d'environ 20 minutes.

Cet exemple montre l'intérêt de ce type d'outil dans une phase d'apprentissage. En effet, il demande en fait de construire une vision abstraite d'une ligne musicale et de ses contraintes rythmiques. Le fait de pouvoir visualiser et écouter immédiatement ce que l'on code est ici déterminant.

¹³ Voir (et écouter) l'article carillon :

[https://wicri-demo.istex.fr/Wicri/Musique/fr/index.php/Carillon_\(Jean-Jacques_Rousseau\)](https://wicri-demo.istex.fr/Wicri/Musique/fr/index.php/Carillon_(Jean-Jacques_Rousseau))

3.3 Les partitions d'apprentissage

Les chorales sont demandeuses d'outils d'apprentissage, voix par voix. Pour une œuvre contemporaine bien imprimée, des solutions très performantes intègrent des mécanismes de reconnaissance automatique de partitions musicales¹⁴.

Cela dit, quand plusieurs lignes musicales sont déjà codées, LilyPond devient très performant pour réaliser des montages. Citons par exemple un passage « difficile à mettre en place » comme les démarrages décalés des voix dans une fugue.

De même, pour des partitions anciennes (difficiles à reconnaître par un robot), LilyPond s'avère très pertinent. Là encore, comme pour l'apprentissage élémentaire cité plus haut, la transcription d'une œuvre complexe permet d'approfondir sa musicalité, et d'améliorer ses compétences musicologiques (alors que la simple copie d'une partition avec un outil WYSIWYG¹⁵ n'implique pas nécessairement d'approfondissement).

Notons enfin que la croissance des communautés de l'information libre sur le Web donne accès à de très nombreuses sources de partitions. Pour des œuvres libres de droit, des chorales commencent à rendre public leurs documents de travail. Ainsi, le site ChoralWiki¹⁶ propose ainsi environ 40.000 partitions (dont 5.000¹⁷ en LilyPond).

Cela dit, sur Wicri/Musique, cette fonction est plutôt la conséquence d'un objectif plus important : faire connaître le contexte musical, historique et linguistique d'une œuvre donnée. La musique ancienne est un domaine particulièrement intéressant de ce point de vue.

3.4 Du côté de la Renaissance et des musiques anciennes

Nous avons voulu montrer comment une infrastructure encyclopédique ouvrait des pistes exploratoires parfois imprévues à partir d'une recherche élémentaire.

Par exemple, le dictionnaire du TLF ouvre des portes parfois inattendues. Voici un exemple : pour le terme « Musique » nous avons trouvé une citation qui a joué un rôle de déclencheur :

- Même sans la musique le *Victimae pascali laudes* est un admirable poème en vers libres ([Gourmont](#), Esthét. lang. fr., 1899, p. 250).

Cette citation nous a conduit à rédiger un article sur la séquence *Victimae pascali laudes*, et une entrée encyclopédique sur Remy de Gourmont. Ces nouvelles pages sont reliées par des liens sémantiques (par exemple « A pour auteur cité » entre l'article musique et Remy de Gourmont).

Nous avons également mené des expérimentations sur la Renaissance où nous avons principalement traité des motets ou des chansons. De telles pièces peuvent être exposées sur une page wiki¹⁸. Voici des exemples.

¹⁴ Nous n'avons pas réalisé d'études comparatives pour en dire plus dans cet article.

¹⁵ Jargon informatique pour désigner le mode utilisé avec Word : *What You See Is What You Get*

¹⁶ <https://www.cpd1.org/>

¹⁷ Chiffre obtenu par la requête « lilypond site:cpd1.org » sous Google.

¹⁸ Complétée éventuellement par des sous-pages par pupitre pour les partitions d'apprentissage.

Nous avons traité *l'Ave verum* de William Byrd, en commençant par la rédaction de pages d'apprentissage. Puis, la rédaction d'une page complémentaire sur *l'Ave verum corpus* grégorien amène à s'interroger sur l'ajout des « *miserere* ». Le contexte historique fait plonger dans les origines de la religion anglicane avec par exemple le rejet du latin. *L'Ave verum* de William Byrd était donc une œuvre chantée dans la clandestinité.

L'analyse de la partition montre que les premiers vers sont séparés par des silences. En effet, ils devaient permettre au chœur de se synchroniser avec le rite de la consécration du pain et du vin. Ils ont donc un sens fondamental dans l'interprétation. Ils sont très visibles dans les partitions anciennes (et pourraient être agrémentées d'un point d'orgue dans une transcription). Nous avons donc choisi d'illustrer cet article avec une iconographie. Elle montre un enfant de chœur tenant une clochette qui devait tinter pendant les silences.



Source gallica.bnf.fr / Bibliothèque nationale de France

Figure 10 : Pour interpréter *Cujus latus perforatum fluxit aqua et sanguine*.

Voici un exemple avec la chanson « *Quand mon mary vient de dehors* » de Roland de Lassus. L'examen des partitions anciennes demande d'explicitier la notion de *musica ficta* (altérations non matérialisées). Son examen a permis de catégoriser deux versions.



Figure 11 - Deux versions avec une altération non signalée dans la première.

La première version a été éditée en 1566 par Adrian Le Roy et Robert Ballard. Le texte est truculent et évoque le comportement agressif du vieux mari, jaloux, vis-à-vis de sa jeune épouse. La deuxième est un *contrafactum* qui a été éditée en 1594 par un éditeur protestant, Paul Marceau. Dans un langage moralisateur, elle recommande le travail domestique d'une jeune femme qui a eu la chance d'épouser un honnête bourgeois qui passe son temps dehors. La première version doit être chantée par un musicien professionnel qui doit résoudre les *musica ficta*. En revanche la deuxième peut être chantée par des amateurs¹⁹.

3.5 Irish Mass

Pour les partitions plus conséquentes, l'approche hypertexte permet un large développement de l'ensemble « livret, partitions, apprentissage ». Nous avons réalisé une première expérience avec la messe irlandaise (*Irish Mass*) de Gilles Mathieu.

Cette pièce reprend les mouvements traditionnels d'une messe musicale chrétienne (*Kyrie, Gloria, ...*) mais en version irlandaise. Ainsi le *Kyrie* devient « A Thiarna ». L'ensemble des partitions vocales est transposé et devient un hypertexte à 3 niveaux : une page wiki d'introduction, une page par mouvement, et, pour chaque mouvement, une page par pupitre. Des pages complémentaires donnent des informations complémentaires sur la prononciation ou sur le contexte politique irlandais.

Or Gilles Mathieu a également composé une suite pour orchestre et chœurs SATB (un oratorio profane) sur la Chanson De Roland. Nous avons donc lancé une bibliothèque numérique expérimentale pouvant accueillir des œuvres aussi particulières que l'ouvrage annoté par Paul Meyer ou l'oratorio de Gilles Mathieu.

4 La Chanson de Roland

Le 15 août 778, de retour d'Espagne, Charlemagne perd son arrière-garde, tombée, à titre de représailles, sous le feu des 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

¹⁹ Cette explication a été donnée par Jacques Barbier lors des journées Musamat.

l'entourage de Charlemagne, dont celle de Roland, préfet de la Marche de Bretagne²⁰. Ce fait d'armes a inspiré des cantilènes, des récits et une chanson de geste, la *Chanson de Roland*. Ce poème épique a été déclamé dans toute l'Europe par des jongleurs et des troubadours. Quelques manuscrits ont survécu et font l'objet d'une abondante production littéraire depuis le XIX^e siècle.

4.1 Deux stages d'étudiants qui révèlent un gigantesque univers numérique

Un manuscrit fondamental a été découvert, en 1837 à la bibliothèque Bodléienne d'Oxford, par Francisque Michel. Celui-ci a donc publié une édition critique qui a été annotée par Paul Meyer (et cité plus haut). Or ce manuscrit est depuis une dizaine d'années disponible sur le web (et notamment sur Wikimedia Commons).

En 2021, nous avons donc saisi l'opportunité de deux stages d'étudiant pour lancer une première étude sur la réédition complémentaire de l'oratorio de Gilles Mathieu avec l'ouvrage du fonds Paul Meyer, en utilisant le manuscrit comme élément médiateur.

En fait, ces deux stages d'étudiants ont fait émerger un vaste champ d'investigation dans le domaine des humanités et bibliothèques numériques.

A ce niveau, il est utile de préciser que les porteurs du projet Wicri, n'avaient aucune érudition dans la culture historique et éditoriale du Moyen Age (de vagues souvenirs de cours élémentaires suivis distraitement). Nous avons donc abordé ce sujet avec des comportements d'amateurs qui explorent un paysage historique.

4.2 Un corpus riche et varié

Les événements liés à la Chanson de Roland (et aux légendes de Charlemagne) se situent à la fin du VIII^e siècle. Une des premières bibliographies a été rédigée en 830 (environ) par Eginhard en latin.

La Chanson de Roland est parfois datée du XI^e siècle (première croisade). Les copies disponibles (une dizaine de manuscrits) datent du XII^e siècle où elle a été maintes fois avec différents niveaux de langue, et des traductions (par exemple en allemand). Elle fait partie d'un ensemble de 27 chansons de gestes nommé « le cycle de Charlemagne ». La plupart des personnages de la Chanson de Roland se retrouvent dans les autres chansons. Ainsi leurs histoires se complètent et s'expliquent.

Ces légendes se sont développées pendant plusieurs siècles et ont donné lieu à des récits parfois très populaires (dans la bibliothèque bleue par exemple). En musique un poème du XVI^e siècle, *Orlando furioso*, a été une source d'inspirations pour des compositeurs comme Lully, Vivaldi, Haendel ou Charpentier.

La base potentielle de notre bibliothèque numérique est donc un ensemble de centaines de manuscrits qui couvrent plusieurs siècles de littérature. Plus précisément quelques manuscrits doivent être traités dans leur intégralité (par exemple celui d'Oxford). Les plus importants doivent être signalés avec quelques extraits explicités.

²⁰ La marche de Bretagne, créée au VIII^e siècle par la réunion de plusieurs comtés francs, se composait du Rennais, du Nantais et du Vannetais, ainsi que d'une partie du Maine.

Puis, au XIX^e siècle, notamment dans la dynamique de l'École des Chartes, tous ces manuscrits ont été transcrit, traduits, étudiés et commentés (autour de Francisque Michel, Léon Gautier, Paul Meyer, Gaston Paris). Cette activité s'est poursuivie au XX^e siècle avec par exemple Joseph Bédier et la suite de la revue Romania.

Cette thématique donne lieu à une importante activité de recherche. Par exemple, nous avons développé un serveur d'exploration à partir de 3000 articles sur ISTEEX (dont près de 200 publications du Royaume Uni, puis 175 des Etats-Unis). Une recherche sur Google Scholar donne 30.000 articles ou citations.

Les sujets d'investigations sont donc très variés, historiques, linguistiques, éditoriaux, musicaux ou... légendaires.

4.3 Une gigantesque bibliothèque hypertexte

En pratique, une chanson de geste est un ensemble de vers regroupés en couplets (généralement appelés laisses).

Plus précisément, un manuscrit est un ensemble de vers qui sont, d'une part, répartis sur un ensemble de feuillets recto verso, et d'autre part, regroupés en laisses. Une laisse contient des vers en assonance et commence, en principe, par une lettrine. Dans le manuscrit d'Oxford, une laisse se terminent généralement par une mention mystérieuse « [Aoi] ». Une laisse peut être à cheval sur 2 pages.

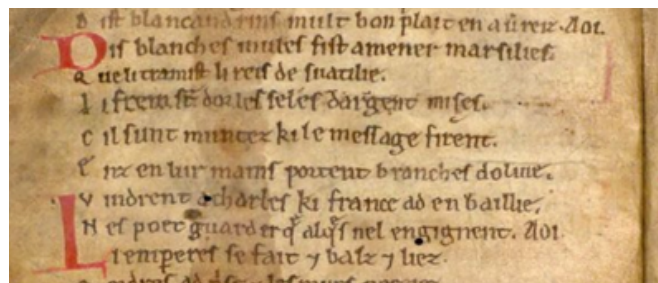


Figure 12 : un enchaînement de 3 laisses, 2 lettrines (D et L) et 2 mentions Aoi en fin de ligne

L'ouvrage de Francisque Meyer suit rigoureusement, pensions-nous, la composition du manuscrit. Il est organisé autour des laisses qui sont traitées dans leur intégralité.

De son côté, Gilles Mathieu a organisé son oratorio autour de quelques laisses (une cinquantaine) dans lesquelles il a extrait quelques vers. De plus ces vers sont regroupés dans une dizaine de mouvements.

Au début du stage, nous avons donc imaginé de faire une association « relativement simple » entre les éléments suivants :

- les phrases musicales de l'oratorio ;
- les laisses correspondantes dans le manuscrit ;

- et leurs transcriptions dans l'ouvrage de Francisque Michel.

Or Gilles Mathieu était parti d'une autre transcription (Léon Gautier). Les amateurs, que nous étions, ont démarré le travail sans rencontrer de problèmes particuliers pour les premières laisses. Et puis rapidement nous avons été confrontés à des problèmes d'alignement.

En effet, dès que l'on cherche à aligner plusieurs ouvrages primaires (transcription) avec un manuscrit, les divergences de numérotation deviennent omniprésentes. Ainsi la dernière laisse est numérotée CCXCI chez Roland Bédier, CCXCIII chez Edmund Stengel, CCXCVI chez Francisque Michel et CCXCVII chez Léon Gautier. En effet, le contenu de certaines laisses pose des problèmes d'interprétation.

Par exemple le feuillet 43 verso ne contient ni lettrine, ni mention AOI, mais est visiblement une charnière dans le récit. Il contient un vers de transition entre la partie dédiée à la mort de Roland et celle qui est dédiée au retour de Charlemagne (figure 13).

Morz est Rollant, Deus en ad l'anme es cels.²¹

Or elle ne contient ni lettrine, ni mention AOI. En revanche un point apparaît sur le manuscrit.

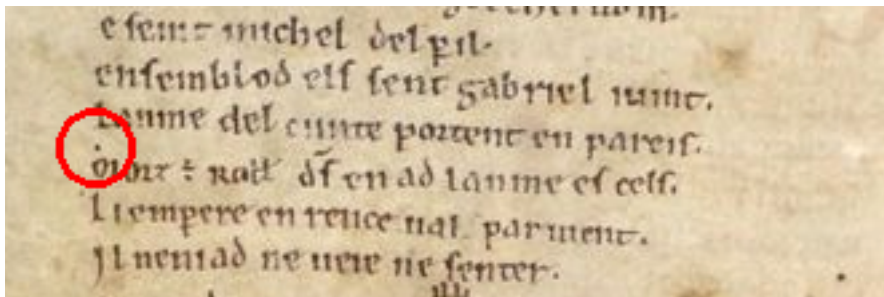


Figure 13 - ... Roland est mort ...

Les médiévistes ont des interprétations divergentes. 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). Chaque philologue travaille donc sur une modélisation du manuscrit qui lui est propre. Un manuscrit engendre donc un hypergraphe qui se démultiplie pour chaque édition critique ou transcription.

Or le manuscrit d'Oxford est le plus prestigieux d'un ensemble d'un dizaine d'autres manuscrits qui engendrent les mêmes démultiplications. Nous avons mentionné plus haut les diverses chansons qui font écho à la Chanson de Roland et qui elles aussi font l'objet de plusieurs manuscrits et de multiples analyses.

La simple navigation au sein des interprétations des couplets des légendes de Charlemagne demande déjà la réalisation éditoriale d'un vaste hypertexte. Mais, pour répondre aux besoins des amateurs il faut encore aller plus loin... En effet, une fois compris le contexte d'un couplet, comment savoir comment prononcer et interpréter les simples mots d'un vers donné ?

²¹ Roland est mort ; Dieu a son âme dans les cieux.

La réponse est partiellement donnée dans les notes, glossaires et tables des matières des éditions critiques. L'édition critique de Léon Gautier (1872) est édifiante de ce point de vue. Elle est constituée de 2 volumes de plus de 500 pages.

- Les 297 laisses (avec leur traduction) occupent 325 pages du premier tome.
- Cette partie est précédée d'une introduction historique de 200 pages. Pour être compréhensible par un lecteur du XXI^e siècle, chaque page doit être enrichie par une dizaine de renvois dans un partie encyclopédiques (un hypertexte alphabétique).
- Plus d'un millier de notes explicatives (vers par vers) sont reportées dans le deuxième tome sur 275 pages. Pour devenir lisible en numérique par un amateur contemporain cet ensemble doit être « hypertextualisé », vers par vers.
- Un glossaire de plusieurs milliers d'entrées est réparti sur 200 pages. Il convient d'en faire un hypertexte pour chaque mot du manuscrit.
- Une table des matières de 40 pages doit également subir une transformation hypertexte.

Faut-il entreprendre un tel travail de façon exhaustive et le répéter pour chaque édition critique ?

- Pour des professionnels de l'édition une réponse par oui à cette question implique le lancement d'un projet de dizaines d'hommes années.
- En revanche, les chercheurs et les amateurs peuvent s'attaquer à un sujet en intégrant leurs actions dans un paysage hypertexte suffisant pour interpréter un problème rencontré, mais très partiel dans la complétude du wiki.

La perspective d'un concert autour de l'oratorio de Gilles Mathieu offre précisément un ensemble de problèmes pertinents pour avancer sur l'architecture numérique de la bibliothèque.

4.4 La perspective d'un concert pour organiser le chantier hypertexte

L'oratorio de Gilles Mathieu ne porte que sur un peu plus d'une centaine de vers dans un seul manuscrit. Il rend humainement possible une expérience en vraie grandeur, déjà significative, de cette construction hypertexte.

Cette suite musicale a déjà été interprétée par la Chorale Universitaire de Nancy en 2009. L'ensemble des partitions de l'auteur est donc suffisant pour qu'un concert soit possible.

La question posée à ce niveau est donc : comment améliorer les conditions de travail pour un nouveau projet musical d'un ensemble amateur ?

4.4.1 Construite la colonne vertébrale du manuscrit d'Oxford

Le premier objectif est de donner au choriste des partitions de travail à partir desquelles il doit pouvoir consulter les vers originaux dans les laisses du manuscrit. Pour écrire les transcriptions musicales avec LilyPond, nous avons donc découpé les mouvements de l'oratorio en ensembles de quelques vers contenus dans une seule laisse du manuscrit.

Nous disposons d'un traitement complet avec l'édition annotée par Paul Meyer où toutes les laisses ont été repérées.

En rappelant que Gilles Mathieu s'était appuyé sur une autre édition critique (Léon Gautier), nous avons dû passer par une étape de « production industrielle » : une analyse complète du manuscrit pour trouver une organisation (en fait une nouvelle numérotation) compatible avec celles des diverses transcriptions. Nous avons notamment vérifié l'alignement possible avec les travaux de Joseph Bédier et d'Edmund Stengel dans cette étape.

Ce travail de renumérotation offre des retombées déjà intéressantes dans la dimension philologique avec, par exemple, l'alignement entre les transcriptions de Francisque Michel et celle de Léon Gautier. La collaboration entre des actions de recherche et des exercices d'amateur se relève ici bénéfique.

Mais la perspective d'un concert permet d'aller encore plus loin.

4.4.2 Dialogue avec le compositeur

Le paramétrage implicite de MediaWiki ouvre des espaces de dialogue que nous avons utilisé pour dialoguer avec le compositeur. Gilles Mathieu, est également informaticien à l'INSERM et il applique dans son activité musicale les recommandations du libre accès à l'information scientifique. Les conditions sont donc particulièrement favorables.

Voici un exemple. Dans le début de l'œuvre, Marsile, l'émir de Saragosse, demande conseil à ses pairs. Il lui est répondu d'envoyer une délégation vers Charlemagne. Cela se traduit en musique par l'interjection « *Mandez Carlun* ». Cette réponse est formulée par le pupitre alto. Compte tenu du rôle fondamental de cette interjection, nous avons interrogé le compositeur sur la consigne *mezzo forte* qui paraissait un peu faible. Il nous a répondu qu'il s'agissait de la réponse d'un conseiller à son roi. Elle ne pouvait être que suggérée, et ici, d'autant plus qu'il s'agit de piéger Charlemagne.

La compréhension du contexte est donc fondamentale pour l'interprétation.

Cela dit, une nouvelle contrainte est introduite. En effet, le compositeur peut répondre favorablement à une demande. Nous avons un exemple avec les mules sur lesquelles les conseillers de Marsile sont allés négocier avec Charlemagne. Elles étaient blanches dans le manuscrit par signe de respect. Cet adjectif avait été oublié dans l'oratorio. Il a été rajouté, avec modification d'une séquence musicale. En généralisant ces pratiques, il faut être capable de produire (imprimer) très rapidement un nouvel ensemble de partitions. Nous avons donc entrepris la transcription de toutes les partitions en LilyPond. Ce travail est en cours.

4.4.3 Pour répondre à la curiosité des choristes et spectateurs

Au-delà des exercices et des éléments de compréhension nécessaires à l'interprétation, le thème de l'oratorio ouvre un immense paysage culturel et musical. Voici quelques exemples.

Il nous semblait prioritaire d'offrir une version pédagogique complète. Le site étant en accès public, nous sommes contraints de rééditer des ouvrages en accès public. Nous avons choisi une version dite « grand public » publiée par Léon Gautier (Gautier 1895) et destinée selon lui : *à l'adresse du vrai peuple, des femmes et des enfants*. Les versions précédentes étaient d'abord destinées à ses collègues et élèves de l'école des Chartes, puis aux classes de seconde. Cette version populaire était

aussi destinée à un public qui chantait en latin dans les églises le dimanche et qui avait une culture littéraire totalement différente d'un choriste amateur en 2023.

Il faut donc développer la partie encyclopédique, en fonction des difficultés rencontrées, et donc dans une logique globale difficilement planifiable.

Cet ouvrage pédagogique insère également des laisses qui viennent d'autres manuscrits que celui d'Oxford. Ceci nous a conduit à les introduire dans l'hypertexte.

De plus, l'ouvrage pédagogique de Léon Gautier contient 40 pages d'introduction historique et 45 pages dites d'éclaircissements qui citent par exemple toutes les Chansons de Geste du cycle de Charlemagne et un grand nombre d'éditions critiques. Il faut donc entreprendre un grand nombre de signalements documentaires (avec des liens sur des versions en ligne).

Les mouvements sont découpés en séquences musicales limitées pour être alignés sur les pages dédiées aux laisses du manuscrit. Celles-ci sont aussi utilisables par les philologues ou les linguistes. Chaque terme est décrit dans des glossaires qu'il faut synthétiser. Les besoins des choristes amènent à traiter en priorité les mots qui peuvent introduire des faux sens en interprétation.

Enfin, pour clore cette liste d'exemples, une des questions souvent posées est : la Chanson de Roland a-t-elle inspiré d'autres poésies ou d'autres œuvres musicales ?

Du côté des poésies, nous avons réédité par exemple deux chapitres de la *Légende des siècles* de Victor Hugo.

Pour la musique, nous avons donné des extraits d'un opéra de Mermet (*Roland à Roncevaux*) qui a été composé en 1864. Le dictionnaire des opéras (1881) en décrit ainsi le style : *appels fréquents de trompettes, une marche guerrière donnent le ton général de l'ouvrage*. Plus complexe à commenter, le poème *Orlando Furioso* a inspiré de multiples opéras dont Roland est bien le héros mais qui n'ont pas grand-chose à voir avec le *Roman de Roncevaux*.

5 Bilan et perspectives

Le projet Wicri, nous l'avons vu, a été créé en 2008 en s'appuyant sur l'expérience de Wikipédia. Quinze ans plus tard les éléments de croissance de la Wikimedia Foundation sont donnés par le graphe ci-dessous :

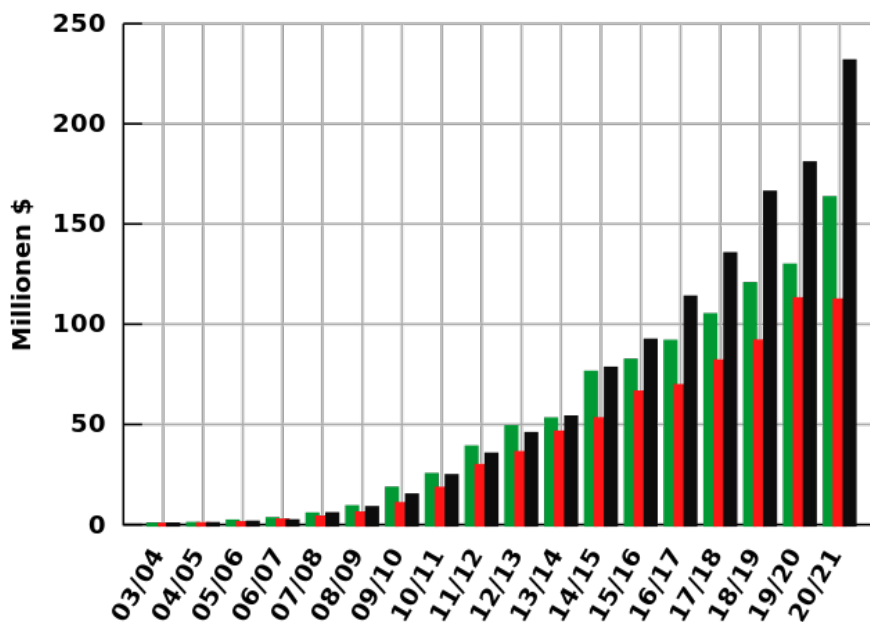


Figure 14. La croissance du chiffre d'affaires de la Wikimedia Foundation

Avec un chiffre d'affaires qui dépasse maintenant les 150 millions de dollars son profil se rapproche alors des GAFAM où l'anonymat des contributions génère des dérives de plus en plus inquiétantes. La faisabilité d'une diversification de l'offre de connaissance scientifique est donc particulièrement d'actualité.

De même, l'analyse des malaises mis en lumière avec les mouvements sociaux liés à la réforme des retraites montre l'importance de la recherche de sens dans le travail. L'importation dans les unités de soutien à la recherche de pratiques ayant fait leurs preuves dans le monde des amateurs est-elle bénéfique pour les agents tout en étant plus efficace en termes de productivité ?

5.1 Aspects techniques

Au sein des service français de soutien à la recherche, le projet Wicri défend des pratiques d'amateurs, là où de nombreux cadres mettent en avant le mot « professionnalisme » (zéro défaut), utilisé parfois comme bouclier face aux craintes de l'innovation. Le projet ne bénéficie donc que d'un soutien institutionnel très limité. Les moyens humains sont actuellement limités à un retraité, et parfois, quelques stagiaires.

Avec des moyens très réduits, les résultats concrets démontrent la puissance technique l'approche « wikis sémantiques + ingénierie XML ». Le réseau contient actuellement 150 wikis. Voici quelques chiffres pour 4 d'entre eux :

Tableau 1 – Indices de production sur les wikis (janvier 2022).

	Pages wiki ²²	Avec contenu ²³	Modifications	Sémantique ²⁴
Musique	4 292	1 398	10 734	51 341
<i>Chanson de Roland</i>	8 039	2 456	37 178	35 581
Histoire de l'IST	1 843	456	3 798	36 303
Santé	3 527	629	2 250	28 774

Ces chiffres révèlent une importante productivité.

Cela dit, la volumétrie visée par un tel site est considérable. Par exemple, le glossaire d'une édition critique de Léon Gautier représente environ 3 000 entrées. Son traitement demande la création d'autant de pages wiki. Un ouvrage de synthèse comme celui de Joseph Duggan contient 2940 pages en trois volumes. Sa réédition en hypertexte demanderait environ 10 000 pages wiki.

La wiki Wicri/Chanson de Roland est une bonne base de travail mais pour devenir un site de référence, il faut viser au moins 50 000 pages. La perspective, même hypothétique d'un concert devient un cadre précieux pour focaliser les contributions sur un premier ensemble de 15 000 pages.

5.2 Aspects rédactionnels et institutionnels

Nous mettons à la disposition des amateurs et des professionnels de culture diverse un gigantesque hypertexte où des livres sont découpés et réassemblés dans un espace commun. Comment rédiger et s'organiser dans un tel ensemble ?

5.2.1 Lire, écrire, naviguer dans une bibliothèque hypertexte

Depuis l'invention du rouleau de papyrus, les auteurs et les lecteurs sont invités à agir de façon linéaire. Il y a près de 3 millénaires, le codex complète le texte par des dispositifs d'accès rapide à partir de numéros de pages (table des matières, glossaire etc.). Depuis moins de 2 décennies le PDF généralise la lecture d'ouvrages toujours conçus de façon linéaire.

Avec les wikis les lecteurs et les auteurs plongent dans le monde encore peu connu de l'hypertexte. Au départ les mécanismes de navigation paraissent plus simples quand le numéro de page est remplacé par un lien. Mais ici les glossaires, ou même les simples paragraphes (les laisses), sont partagés entre un grand nombre d'ouvrages. La notion même de document disparaît, comme dans les pages wikis, où pour une laisse donnée on donne différentes traductions ou des notes de bas de pages extraites de plusieurs éditions critiques.

Les aspects collectifs apportent une nouvelle strate de complexité. Dans les bureaux de l'École des Chartes, chaque philologue ou historien disposait d'une bibliothèque contenant des ouvrages annotés et de tiroirs où ils déposaient des notes manuscrites. Sur le wiki, tout se passe un peu

²² Toutes les pages y sont comptées y compris les modèles.

²³ Ne sont comptées ici que les pages de l'espace principal ayant au moins quelques lignes.

²⁴ Cette colonne donne le nombre de liens hypertextes catégorisés ou sémantiques.

comme s'ils mettaient tout en commun sur une immense table. Mais leur érudition leur permet de se frayer un chemin.

Dans le réseau Wicri, il est également possible d'écrire en quasi simultanéité dans plusieurs wikis (Wicri/Chanson de Roland et Wicri/Musique par exemple). Cette possibilité ouvre de nouveaux horizons et aussi des nouvelles difficultés pour se repérer.

Le public amateur a besoin d'être guidé et il faut lui aménager des chemins balisés de difficulté croissante. En fait, un des défis explorés par le projet Wicri/Chanson de Roland est de voir comment la rédaction d'articles pour les amateurs peut amener les professionnels à approfondir leurs travaux de recherche.

Une forte partie du travail rédactionnel peut être mené par des amateurs, éventuellement incompetents au départ, mais qui vont pouvoir se former au fur et à mesure des actions. Cette situation est proche de celle des chorales d'amateurs, avec un chef amateur. Attention, encore une fois, amateur ne veut pas dire incompetent. Le chef de chœur doit avoir le même niveau de compétence que celui d'un professionnel pour que la chorale puisse interpréter des œuvres complexes. Les apprentissages seront plus rapides et les interprétations seront meilleures si la chorale dispose d'un noyau de passionnés qui se forment en solfège ou en technique vocale.

Les mêmes problèmes se posent dans les infrastructures numériques où la présence d'experts de haut niveau est absolument indispensable dans la vie du projet. Les performances seront nettement améliorées si les amateurs suivent des chemins de formations pour devenir à leur tour experts de l'écriture numérique ; comme les chorales savent s'appuyer un noyau de chanteurs qui acquièrent des compétences professionnelles.

5.2.2 Gérer l'incomplétude

L'approche wiki permet de diffuser très rapidement des premiers résultats, même inachevés. L'intérêt très clair : des non-spécialistes de la technologie bénéficient ainsi d'un substrat concret sur lequel ils peuvent immédiatement travailler. Avec quelques séances de formation, un expert de la musicologie, un linguiste, ou un médiéviste, peut rapidement contribuer sur un sujet à n'importe quelle page de n'importe quel ouvrage de n'importe quelle bibliothèque.

Le revers de cette médaille est la gestion de l'incomplétude qui devient un problème omniprésent. Dans le wiki sur la Chanson de Roland, la volumétrie est déjà consistante. Une amélioration minime sur le contenu des laisses (qui demanderait par exemple 2 minutes par action) peut se traduire par des dizaines d'heures de travail. Cela dit, Wikipédia rencontre des problèmes analogues et sait les traiter en organisant des chantiers (ou en programmant des robots). Le même type d'approche doit pouvoir se dégager ici.

5.2.3 Gérer l'interdisciplinarité et la cohérence

Nous avons souvent évoqué la cohabitation de plusieurs publics. Celle-ci est un exemple des contraintes qui amènent à revisiter l'activité de rédaction. Avant le codex, l'écriture dans un rouleau est une activité linéaire réalisée par une personne. Dans un hypertexte, il faut penser écriture parallèle (comme pour une fugue en musique). De plus l'intégration de modules de code (comme pour l'exemple donné plus haut avec la requête sémantique #ASK) introduit des algorithmes dans l'équipe rédactionnelle.

Dans le réseau Wicri, une telle équipe peut travailler simultanément dans plusieurs « bibliothèques numériques ». Un même sujet peut être abordé avec plusieurs points de vue. Par exemple, dans le manuscrit de Paris de la Chanson de Roland, un miracle se produit après la mise en terre des preux. Des arbres actuellement nommés cormiers brisent les pierres tombales. Sur Wicri/Musique, comme pour cette version, une simple note de bas de page suffit (avec des liens pour aller plus loin). Sur Wicri/Bois, il faut situer cette plante dans les classifications du vivant. Sur Wicri/Chanson de Roland, il est possible de développer une petite étude lexicographique.

Cette possibilité est naturellement très intéressante. Elle pose rapidement des problèmes de cohérence entre des versions voisines sur différents wikis. Dans d'autres registres, comme la promotion de la recherche, les changements administratifs telle que la fusion de deux universités peuvent avoir des conséquences redoutables par leur nombre d'interventions.

Cette publication, dans laquelle nous faisons des manipulations sémantiques, demande une gymnastique permanente pour transférer un exemple avec des ontologies différentes.

La maîtrise de la cohérence, ou la simple nomenclature des noms de page demande en fait une solide culture numérique.

5.2.4 Formation

Autrement dit, les premières contributions sur un wiki donnent un sentiment de simplicité parfois trompeur. Pour prendre une comparaison, une première leçon de piano permet de jouer le début de « au clair de lune ». En revanche, il faut des années de formation pour interpréter une œuvre complexe (ou pour composer un opéra).

Cela dit, les wikis offrent de nombreux outils permettant d'assurer une assistance aux contributeurs débutants. Nous avons de nombreux exemples d'encadrement de stagiaires de formation assez variées. En quelques jours, ils sont capables de réaliser des contributions simples (comme par exemple corriger des documents bruts de procédé de reconnaissance de caractères (OCR)). En quelques semaines ils sont autonomes pour des actions d'indexation en utilisant une ontologie préexistante. En quelques mois ils sont autonomes pour concevoir et réaliser des applications comme une revue numérique à condition de pouvoir bénéficier d'une assistance de haut niveau technique.

Cette assistance doit être déployée pour chaque spécialité. Par exemple, en musique, nous avons donné un exemple avec le logiciel LilyPond où il est possible de recopier une ligne musicale au bout de quelques jours. En revanche, le traitement d'une œuvre complète comme Irish Mass ou L'oratorio de Gilles Mathieu demande une solide formation solfégique et une formation approfondie à l'utilisation de l'outil.

5.3 Aspects institutionnels

Comme rappelé en début de cette section, nous étudions la faisabilité technique d'une solution alternative à Wikipédia. Sa mise en œuvre demanderait naturellement le déploiement d'un vaste réseau de coopérations et un soutien logistique. Le simple besoin d'assistance des chercheurs demande par exemple la présence de learning centers dans les bibliothèques universitaire avec une solide (en termes de compétences et non de charge de travail) assistance technique.

Cela dit, une telle mise en œuvre ne demande pas de nouveaux recrutements mais plutôt une mutation des activités. Les approches que nous proposons sont naturellement créatrices de grandes bibliothèques par des réseaux de coopération qui améliorent la qualité et la productivité des tâches courantes.

Pour le projet Wicri, la simple migration de l'URL du projet (Loria, puis INPL, et maintenant istex) montre que nous n'avons pas encore réussi à convaincre au niveau lorrain.

Nous avons connu le même type de blocage avec l'approche boîte à outils SGML sur laquelle nous avons publié en 1991 (Ducloy 1991). Sur son site la BnF se réfère à la mutation de la Bibliothèque du Congrès en 2001²⁵, comme point de départ de ses travaux avec cette approche.

Concernant, l'approche Semantic MediaWiki, nous avons publié en 2010 au niveau international (Ducloy 2010). Dans le monde des musées, aux USA, les Smithsonian institutions bénéficient d'une expérience de quelques années (Shieh 2022). En France, la BnF publie sur l'utilisation de ce type d'approche (Boulet 2022)²⁶. L'EHESS commence visiblement à intégrer une démarche voisine de celle que nous utilisons ici (Duménieu 2022).

Les réticences au changement sont de nature assez différente. Dans le cas de l'ingénierie XML, les services informatiques des unités de soutien à la recherche avaient sous-estimé les difficultés liées aux traitements des données bibliothéconomiques (norme ISO 2709) ou au passage au contenu dans les publications scientifiques. Ils restaient donc relativement bloqués sur une utilisation dogmatique des SGBD relationnels. Mais ce changement d'ingénierie concernait assez peu les utilisateurs.

Sur un wiki, comme Wikipédia, les contributions sont libres au départ, basées sur la confiance, et éventuellement modérées ensuite. Les institutions publiques ont généralement des protocoles basés sur la validation a priori. Ceci est particulièrement vrai dans les services opérés par les universités ou les pratiques de management sont donc bouleversées. Pire encore, l'expérience Wicri montre qu'un chercheur de toute discipline (y compris la musicologie) peut devenir totalement autonome en s'appropriant une maîtrise avancée du numérique (comme les alchimistes sont devenus chimistes en maîtrisant les systèmes d'équations différentielles). Dans un système où les luttes internes sont omniprésentes, cette indépendance est souvent très mal perçue.

Nous avons probablement commis des erreurs psychologiques (ou politiques) en montrant que nous pouvions mieux faire sur des applications existantes. Avec la démonstration sur la Chanson de Roland nous espérons être plus convaincants en nous attaquant à un problème original.

6 Conclusion

Pour le TLF la première définition de l'amateurisme est : *Qualité de celui, celle qui manifeste un goût de prédilection pour quelque chose (notamment pour l'art).*

²⁵ <https://www.bnf.fr/fr/marc-en-xml>, consulté le 30 janvier 2013

²⁶ Il s'agit en fait d'une solution de production associant MediaWiki et WikiBase.

Dans le monde du numérique, de 1965 à 1970, les enseignements informatiques étaient pratiquement inexistantes. De très grandes réalisations comme le dictionnaire TLF, le réseau Cyclades ou la création des bases Pascal et Francis ont été réalisées par des ingénieurs généralistes qui n'avaient jamais vu un ordinateur pendant leur formation. Ils étaient, au départ, totalement ignorants des technologies qu'ils devaient utiliser mais qu'ils ont maîtrisé grâce au même type de passion que celle des amateurs en musique.

Dans le cadre d'un programme de recherche sur les humanités numériques, nous avons lancé un chantier autour de la musique. Il est ouvert aux musicologues et aux musiciens amateurs. Au-delà de la musique, pour les choristes curieux, le plaisir de chanter ensemble peut se démultiplier par la découverte d'un immense paysage culturel humaniste entrouvert par un œuvre musicale.

De même les musiciens, amateurs ou professionnels, peuvent découvrir le plaisir de concevoir ensemble une réalisation numérique.

Remerciements

Nous remercions vivement Gilles Mathieu pour sa coopération constante sur le projet. Merci aux valeureux stagiaires Dalila Ladli et Léonard Braux qui ont défriché le terrain. Merci aux équipes techniques de l'INIST, à sa direction et aux instances d'ISTEX pour l'hébergement du réseau Wicri. Merci aux groupes de travail Wicri pour leur soutien amical. Merci à Jean-Pierre Thomesse et à Thierry Daunois pour leurs relectures.

7 Bibliographie

- BOULET, Vincent (2022). How to build an «Identifiers' policy»: the BnF use case. *JLIS*. it, 2022, vol. 13, no 1, p. 177-184.
< <https://www.jlis.it/index.php/jlis/article/download/429/422> >
- BUHR J., DEGEN C. (1977). PASCAL: Une base de données multidisciplinaire son utilisation en physique atomique et moléculaire et physique des fluides et des plasmas. *Journal de Physique Colloques*, 1977, 38 (C3), pp.C3-249-C3-251.
< <https://hal.archives-ouvertes.fr/jpa-00217115> >
- DUCLOY Jacques, CHARPENTIER Patricia., FRANÇOIS Claire, GRIVEL Luc (1991) - "Une boîte à outils pour le traitement de l'information scientifique et technique", *Génie logiciel et systèmes experts*, n° 25, pp 80-90, Paris.
- DUCLOY, Jacques, Thierry DAUNOIS, Muriel FOULONNEAU, Alice HERMANN, Jean-Claude LAMIREL, Stéphane SIRE, Jean-Pierre THOMESSE, Christine VANOIRBEEK (2010). *Metadata for WICRI, a Network of Semantic Wikis for Communities in Research and Innovation*, DC 2010, Pittsburgh.
- DUCLOY, Jacques, et al. (2019). Systèmes d'information encyclopédiques édités par les scientifiques, *Revue ouverte d'ingénierie des systèmes d'information*, 1, 2019
- DUGGAN, Joseph (2005). *La Chanson de Roland. The Song of Roland. The French Corpus*, Joseph J. Duggan, General Editor, Turnhout, Brepols, 2005
- DUMENIEU Bertrand, et al. (2022). Un wiki sémantique pour l'édition scientifique d'une correspondance du xix^e siècle, *Humanités numériques*, vol. 6, 2002

< <https://doi.org/10.4000/revuehn.3203> >

GAUTIER, Léon (1895). *La chanson de Roland, Traduction, précédée d'une introduction et accompagné ed'un commentaire, par Léon Gautier*. 22. éd. Edition populaire, illustrée par Olivier Merson, Chiffart, Ferat et Zier.

< https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/La_Chanson_de_Roland/L%C3%A9on_Gautier/%C3%89dition_populaire/1895 >

MARTIN, Robert, et al. (1969). Le trésor de la langue française et la méthode lexicographique, in: *Langue française*, n°2, 1969

SHIEH, Jakie. (2022). Smithsonian Libraries and Archives & Wikidata: Using Linked Open Data to Connect Smithsonian Information.

< <https://blog.library.si.edu/blog/2022/01/19/smithsonian-libraries-and-archives-wikidata-using-linked-open-data-to-connect-smithsonian-information> >