

Théia
ISSN : 3074-5527
Publisher : Université Lumière Lyon 2

1 | 2024
Retours d'expérience en édition numérique de sources en histoire et histoire de l'art

Edited by Pierre-Jean Souriac

http://publications-prairial.fr/theia/index.php?id=78

Electronic reference
« Retours d'expérience en édition numérique de sources en histoire et histoire de l'art », *Théia* [Online], Online since 14 avril 2025, connection on 13 mars 2026. URL : http://publications-prairial.fr/theia/index.php?id=78

Copyright
CC BY-NC-SA 4.0

DOI : 10.35562/theia.78

Prairial
Pôle éditorial
Lyon Saint-Étienne

ISSUE CONTENTS

Retours d'expérience en édition numérique de sources en histoire et histoire de l'art

Angela Göbel, Morgane Pica and Pierre-Jean Souriac
Introduction

Ariane Pinche
Des plumes aux pixels : actualité de l'édition scientifique numérique

Olivier Spina
RePPOL : un projet d'édition numérique d'une enquête sur les croyances et les pratiques religieuses dans le Kent (Angleterre, 1543)

Angela Göbel
Faciliter l'édition numérique avec les méthodes de reconnaissance automatique de texte

Maximilian Görmar
La reconnaissance d'entités nommées dans les éditions numériques à l'exemple du récit de voyage du pharmacien Wagener

Christine Chadier
Éditer la recherche et les données de la recherche : conjuguer plateforme de livres et entrepôt de données

Notes et documents

Béatrice Denis
Saragosse lieu de siège, Saragosse pittoresque : les vues de Saragosse de Louis-François Lejeune pour le *Voyage pittoresque et historique en Espagne*

Ellie Douska
« Échos d'un coin du Chili » : l'Amérique latine à portée de main

Adeline Dollet
Victorine Meurent, modèle et artiste. Relecture d'un mythe de la décadence artistique

Remise du diplôme de Doctorat *Honoris Causa* à Orhan Pamuk, écrivain et prix Nobel de Littérature, Université Lumière Lyon 2, 24 juin 2024

Orhan Pamuk
On love and museums

Zeynep Bursa-Millet
Ce qu'une historienne peut dire des romans d'Orhan Pamuk

Laurent Baridon
La spirale inspirée

In Memoriam

Monica Martinat
De la grandeur d'une historienne

Joana Barreto
Hommage à Gaylord Brouhot (1981-2024)

Retours d’expérience en édition numérique de sources en histoire et histoire de l’art

Introduction

Angela Göbel, Morgane Pica and Pierre-Jean Souriac

DOI : 10.35562/theia.83

Copyright
CC BY-NC-SA 4.0

OUTLINE

Le choix de la source
Le choix technique
Le rendu final
Le devenir de ce travail d'édition

TEXT

1 Pour son premier numéro, et dans une démarche autoréflexive, la revue *Théia* souhaite interroger la pratique même de l'édition numérique. Depuis plusieurs années, les éditions *Chrétiens & Sociétés* portées par l'axe Religions & Croyances du LARHRA (Laboratoire de Recherches Historiques Rhône-Alpes, UMR 5190) se sont lancées dans la publication de livres en ligne via les outils de la Science Ouverte, ainsi que dans l'édition numérique de sources visant à valoriser le document imprimé en le traitant comme une base de données insérée à un réseau d'informations. L'Axe de Recherche en Histoire Numérique, quant à lui, encourage l'encodage sémantique des données notamment par le développement de l'environnement de recherche Geovistory qui permet de construire des bases de données tout en les reliant à l'édition numérique de la source textuelle.

2 Sous cet angle, l'édition numérique invite à réfléchir sur les potentialités nouvelles offertes par un tel outil au traitement de l'information historique. La mobilisation des connaissances permise par le croisement d'une source textuelle ou iconographique et des bases de données externes offre aux chercheuses et chercheurs de penser une méthodologie de dépouillement et de mise à disposition des données qui va bien au-delà de la seule lecture du document et que l'outil numérique permet de valoriser par la masse des

informations qu'il peut traiter. Ces possibilités nous rappellent que base de données et sources sont complémentaires pour le travail de recherche en Histoire et en Histoire de l'Art.

3 Une Journée d'Études s'est tenue sur ce sujet le 22 juin 2023, et nous souhaitons publier les communications faites pour l'occasion. Les articles qui vont suivre sont le fruit d'une mise en relation des historiennes et des historiens travaillant sur l'édition numérique pour confronter leurs projets, leurs questionnements, leurs réussites et leurs échecs. Alors que de tels projets voient le jour maintenant depuis de nombreuses années, ces textes apparaissent comme des jalons, entre bilan et prospective. Ils mettent en regard plusieurs entreprises éditoriales pour en mesurer les proximités et les divergences. Plusieurs de ces articles sont des retours d'expérience sur des projets inachevés mais qui sont autant de réflexions sur la manière de s'emparer des humanités numériques dans l'édition de texte. Les auteurs ont accepté un cahier des charges thématique de manière à pouvoir faire émerger des éléments de comparaison. Les articles qui vont suivre répondront tous aux quatre interrogations suivantes :

Le choix de la source

4 Qu'est-ce qui préside au choix de telle ou telle source pour en faire une édition numérique ? Y a-t-il une singularité à ce type d'édition ou les choix se font-ils selon les mêmes critères que pour une édition papier ? Les contraintes législatives et financières jouent-elle un rôle particulier ?

Le choix technique

5 Il s'agira ici de questionner le choix fait par les historiens de privilégier telle ou telle solution technique par rapport au corpus de(s) source(s) sélectionné pour le projet d'édition. Quels sont les critères de décision concernant l'utilisation ou non d'une reconnaissance de texte (HTR/OCR, Transkribus, eScriptorium, Tesseract…), le recours à la XML/TEI, LaTeX ou d'autres langages, ou la reconnaissance d'entités nommées ? Comment les données tirées de la source interagissent-elles avec la source elle-même ?

Le rendu final

6 Cet aspect prolonge les deux précédents et permet de mieux cerner la singularité de l'édition numérique. Les outils techniques permettent des possibilités de visualisation du document et des informations qui lui sont associées. L'historien et l'historien de l'art font ici alors un choix dans le rendu de leur travail : qu'est-ce qui apparaît – ou non – à l'écran ? Comment révéler les informations complémentaires qui viennent éclairer le document grâce à sa mise en réseau avec d'autres bases ? Comment rendre à l'écran des informations thématiques que l'on peut tirer d'une analyse globale du document ? Dans le cas des images, comment intégrer les métadonnées et les choisir afin qu'elles soient interopérables ? Comment lier l'image et son commentaire, ainsi que l'image et son contexte, et permettre de dépasser le simple stade de l'illustration ?

7 Ce choix a-t-il influencé le choix de la méthodologie ? Cet aspect est le plus complexe, mais aussi celui qui est le plus en lien avec les résultats de l'analyse historique. Il s'agit ici de se questionner sur la visualisation d'une ou plusieurs données historiques contenues dans le document.

Le devenir de ce travail d'édition

8 Que faire de ce travail une fois qu'il est achevé ? La question posée ici sera celle de la diffusion du savoir historique, entre outil scientifique et outil de valorisation du savoir historique adressé à un plus large public. Au regard de ces différentes questions, les expériences d'éditions numériques menées dans des cadres différents proposeraient une base d'expériences multiples promettant un dialogue fructueux.

ABSTRACTS

Français
La revue *Théia*, dans son premier numéro, explore la pratique de l'édition numérique. Elle interroge l'utilisation des outils de la Science Ouverte pour publier des livres en ligne et valoriser les sources historiques. Une Journée d'Études a eu lieu le 22 juin 2023, et les articles qui suivent sont le résultat

de discussions entre historiens et historiens de l'art sur leurs projets d'édition numérique. Ces textes visent à comparer différentes approches et à réfléchir sur des questions clés : le choix des sources, les solutions techniques, le rendu final des documents et la diffusion des résultats. L'objectif est de favoriser un dialogue autour des humanités numériques et de leur impact sur la recherche en Histoire et en Histoire de l'Art.

English
In its first issue, the journal Théia explores the practice of digital publishing. It examines the use of Open Science tools to publish books online and enhance historical sources. A Study Day was held on 22 June 2023, and the articles that follow are the result of discussions between historians and art historians on their digital publishing projects. The aim of these texts is to compare different approaches and to reflect on key issues: the choice of sources, technical solutions, the final rendering of documents and the dissemination of results. The aim is to encourage a dialogue on digital humanities and their impact on research in History and Art History.

INDEX

Mots-clés
édition numérique, LARHRA UMR 5190, méthodologie

Keywords
digital edition, LARHRA UMR 5190, methodology

Schlagwortindex
digitale Edition, LARHRA UMR 5190, Methodologie

Palabras claves
publicación digital, LARHRA UMR 5190, metodología

Parole chiave
pubblicazione digitale, LARHRA UMR 5190, metodologia

AUTHORS

Angela Göbel
Université Jean Moulin Lyon 3, LARHRA UMR 5190
IDREF : https://www.idref.fr/28172489X

Morgane Pica
ENS Lyon, LARHRA UMR 5190
IDREF : https://www.idref.fr/281725098

Pierre-Jean Souriac
Université Jean Moulin Lyon 3, LARHRA UMR 5190
IDREF : https://www.idref.fr/111096332
ORCID : http://orcid.org/0000-0001-7856-7523
HAL : https://cv.archives-ouvertes.fr/pierre-jean-souriac
ISNI : http://www.isni.org/0000000041612143
BNF : https://data.bnf.fr/fr/15834889

Des plumes aux pixels : actualité de l'édition scientifique numérique

From writing with quills to pixels: the latest in digital scientific publishing

Ariane Pinche

DOI : 10.35562/theia.102

Copyright
CC BY-NC-SA 4.0

OUTLINE

Édition numérique : définition et héritages méthodologiques
  Édition numérique scientifique : définition
  Édition documentaire contre édition monumentale ?
  Les différentes formes d'édition numérique
Édition numérique : comment, pour qui, nouveaux enjeux ?
  Un standard : XML TEI
  Qui sont les « lecteurs » des éditions numériques ?
  Nouvelles problématiques
Vers un changement d'échelle
  Acquisition automatique de texte
  Les chaînes d'acquisition textuelle

TEXT

1 Les pratiques de l'édition savante [1] sont les héritières d'une méthodologie qui remonte aux premières tentatives de reconstitution d'un texte original ou du moins d'une version canonique pour des textes aux traditions complexes comme la Bible, les œuvres d'Homère, Cicéron, Virgile, Chrétien de Troyes ou encore Shakespeare [2]. Depuis les humanistes du XVI^e^ siècle jusqu'à nos jours, les éditeurs ont mis au point différentes méthodes pour rendre la pluralité textuelle. Ces méthodes vont de la restitution minutieuse d'un texte le plus proche possible de sa source – restitution de la mise en page, des abréviations, etc. – à la reconstitution d'un archétype le plus proche possible de la version de l'auteur. L'éditeur y retrace les relations entre les différentes réalisations du texte depuis

le(s) texte(s) de l'auteur jusqu'aux réalisations accessibles aujourd'hui [3].

2 Depuis près d'un siècle persiste une tension méthodologique entre éditions « conservatrices », qui suivent un témoin issu de la tradition manuscrite, et les éditions « reconstructionnistes », qui cherchent à retrouver le texte original de l'auteur. La méthode reconstructionniste, née, entre autres, des travaux de K. Lachmann [4], s'est établie comme la méthodologie de référence dans les études classiques. Ces éditions opèrent une nette distinction entre le texte conçu comme étant le texte original et les sources manuscrites regardées comme un véhicule imparfait du texte à reconstruire, là où l'éditeur d'une édition conservatrice envisagera le texte du témoin manuscrit comme une réalisation possible du texte [5]. La méthode conservatrice, initiée au début du XX^e^ siècle par J. Bédier, prône l'édition du « meilleur des manuscrits » (qualité du texte, place dans la tradition manuscrite). Elle se prête tout particulièrement à l'édition des textes médiévaux, en permettant de conserver les graphies d'un document qui a réellement circulé, là où la méthode reconstructionniste aboutit à la création d'un texte dont la langue ou les graphies ne sont plus en lien avec des attestations historiques [6]. Dans la continuité de cette approche, dans les années quatre-vingt, B. Cerquiglini défend, dans l'*Éloge de la variante* [7], la place du témoin manuscrit comme représentant d'un état de transmission du texte ayant sa valeur intrinsèque. Il donne alors naissance à la *New Philology* et influence également la philologie génétique [8]. Dans la continuité de ces réflexions, Peter Shillingsburg définit son travail davantage comme la présentation d'un processus textuel, plutôt que comme l'établissement d'un produit final immuable [9]. Enfin, J. Bryant considère que « the only "definitive text" is a multiplicity of texts, or rather, the fluid text » [10], proposant ainsi un point de vue sur le texte opposé à celui de l'école lachmanienne. S'ajoutent, aujourd'hui, au paysage scientifique, les éditions génétiques qui retracent le processus complet d'écriture de l'auteur à travers l'analyse de ses brouillons [11]. Toutefois, si les méthodes changent, les éditions scientifiques visent toutes l'établissement d'un texte fiable et contextualisé, quelle que soit l'alternative choisie.

3 Enfin, malgré le soin apporté à ces éditions, elles sont souvent difficiles à appréhender à cause des contraintes imposées par le

format papier qui entraînent l'utilisation de règles de représentation et/ou de schématisation de l'information [12], en raison du nombre restreint d'informations transmissibles aux lecteurs. En outre, jamais un livre ne pourra contenir dans ces pages les illustrations, l'arrangement des feuillets, les variations de la tradition manuscrite et rester consultable [13]. Enfin, ces productions, figées dans le papier, nécessitent d'être mises à jour, voire refaites au gré des évolutions méthodologiques ou des nouvelles avancées de la recherche. Les éditions numériques ont, pour partie, émergé pour essayer de dépasser ces limitations, permettant de consulter les numérisations des sources, d'offrir des parcours de lecture, de plonger le lecteur dans un réseau de multi-fenêtrages [14] et d'hyperliens pour proposer une lecture augmentée. Ainsi, après avoir dressé un état de l'art de l'édition numérique, nous présenterons ses standards, son lectorat, mais également ses nouveaux enjeux. Enfin, nous analyserons l'impact des récentes innovations technologiques sur les systèmes de production textuelle.

Édition numérique : définition et héritages méthodologiques

4 Le passage des éditions au format numérique n'est pas synonyme d'une révolution totale des méthodologies de l'édition [15]. Au contraire, elles semblent être solidement enracinées dans les débats philologiques du XX^e^ siècle et ont contribué à enrichir les multiples voies envisageables. Certains théoriciens de l'édition numérique considèrent le passage au numérique comme une évolution naturelle de la philologie traditionnelle. Selon H. W. Gabler, elles doivent respecter les mêmes critères d'érudition scientifique pour l'établissement du texte et être mises en œuvre à l'aide d'instruments qui renforcent l'analyse critique du texte (collation, concordance, stemma, etc.) [16]. En outre, la philologie matérialiste tire parti de la mise en ligne des numérisations des documents, de systèmes de multifenêtrage pour la comparaison, et de la contextualisation diachronique et synchronique des œuvres via des systèmes d'hyperliens [17].

Édition numérique scientifique : définition

5 Que signifie éditer numériquement ? Patrick Sahle propose ces quelques critères pour définir les « *digital scholarly editions* » (DSE) que nous reprenons et traduisons pour partie ci-dessous [18] :

- **S pour Scholarly** : une édition doit offrir une critique textuelle, apportant ainsi une valeur supplémentaire par rapport à la simple mise en ligne d'un fac-similé. Une bibliothèque numérique, qui se contente de numériser des documents, sans offrir une analyse critique, ne répond pas à cette définition.
- **D pour Digital** : une édition numérique ne doit pas pouvoir être convertie en une édition imprimée sans une perte substantielle d'informations ou de fonctionnalités. De même, une conversion numérique d'une édition imprimée n'est pas une édition numérique, sauf si elle est enrichie de contenus ou de nouvelles fonctionnalités.
- **E pour Édition** : une édition numérique doit fournir un texte, qu'il s'agisse d'une simple transcription ou d'un texte plus élaboré. Ainsi, les catalogues, index ou bases de données ne sont pas inclus dans cette catégorie.

Enfin, selon J. Carlquist, une édition scientifique numérique de qualité doit s'appuyer sur un encodage complexe et riche, doit inclure un texte interrogeable et des images, et être enrichie de métadonnées appropriées et riches de sens, un apparat critique, des index et un glossaire [19].

Édition documentaire contre édition monumentale ?

6 Les nouvelles possibilités offertes par les éditions numériques ont ravivé le débat entre éditions documentaires et éditions monumentales, pour reprendre les termes de P. Robinson :

> *One cannot know the work without the documents – equally, one cannot understand the documents without a comprehension of the work they instance. From this, a principle appears: a scholarly edition must, so far as it can, illuminate both aspects of the text, both text-as-*

> *work and text-as-document. Traditional print editions have focused more on the first. An evident advantage of digital editions is that they might redress this balance, by including much richer materials for the study of text-as-document than can be achieved in the print medium.* [20]

7 L'objet numérique libéré des contraintes du papier permet une accumulation d'images, de textes et de représentations offrant la possibilité inédite de représenter le texte dans ses réalisations en tant que document, là où le papier représente le texte comme monument. Bien que nombre d'éditions en ligne soient tournées vers le texte comme document, comme les éditions du projet ELEC de l'École nationale des chartes [21], de nombreux critiques ont souligné les failles d'une pratique qui se limiterait uniquement à une représentation imitative de la source. Toutefois, M. Dahlström clame combien il est fallacieux de dire qu'une édition numérique imitative ne découle pas d'une démarche scientifique [22]. La conversion des signes manuscrits en caractères informatiques relève déjà d'une interprétation, car le passage à une représentation normalisée entraîne une réduction de la variété des formes de lettres de la source [23]. Transcrire et encoder sont le fruit d'un processus de sélection impliquant réflexion et méthode : choix de la granularité de l'imitation de la source, représentation du système abréviatif, niveaux de différenciation des allographes, etc. Toutefois, quoiqu'on puisse considérer qu'une édition documentaire est déjà le premier pas vers une édition critique en tant que source primaire [24], il est primordial de dépasser le stade de la mise en ligne d'archives ou d'une accumulation documentaire, sans quoi le risque est de perdre le lecteur dans un amas d'informations dont il ne parviendra pas à faire sens, amenant à une forme de refus d'éditer [25].

8 Certains projets comme le projet *Hyperdonat* ont expérimenté une voie médiane, essayant de concilier l'approche du texte-comme-document et du texte-comme-monument en proposant un texte de référence et la possibilité de parcourir la tradition manuscrite à travers une interface de comparaison des témoins [26]. Le projet d'édition de *Guiron le Courtois*, en s'appuyant sur la modularité des corpus numériques, propose d'établir le texte branche par branche avec des éditions intermédiaires pour chacune d'entre elles, afin d'établir un texte critique qui garde une trace de la surface

linguistique pour chaque branche de la tradition. La réunion des différentes éditions permettra à terme d'arriver à l'édition critique finale de l'œuvre, tout en suivant les innovations du texte au fil de la transmission [27].

9 Ainsi, même si les premières éditions numériques ont surtout favorisé le texte comme document, tandis que les éditions papier ont favorisé le texte comme monument, les progrès techniques de ces dernières années, grâce, entre autres, à la reconnaissance automatique d'écriture (ATR), offrent la possibilité de traiter des corpus de plus en plus vastes, et donc de traiter des traditions manuscrites complètes, tout en établissant un texte de référence.

Les différentes formes d'édition numérique

10 L'une des premières applications du numérique pour la mise en ligne de textes a été la création de bibliothèques numériques, offrant la possibilité de mener des fouilles textuelles. En France, l'Observatoire de la vie littéraire (OBVIL, 2014-…) [28] a mis à disposition plus d'une trentaine de corpus littéraires français structurés en XML TEI [29] sur Github [30], accompagnés d'outils tels que *Dramagraph* pour faciliter l'analyse de la répartition des répliques dans les pièces de théâtre [31]. À l'échelle internationale, le projet *Perseus Digital Library* (1987 – …) [32] donne accès à des centaines d'éditions de textes, ainsi qu'à une interface de consultation, de recherche, et à des URN [33] pérennes pour citer les textes, tout en assurant l'accès aux fichiers sources encodés en XML TEI.

11 Certaines éditions numériques, tout en respectant les codes de l'édition traditionnelle, proposent des éditions enrichies. C'est le cas du projet la *Queste del saint Graal* (1999 – …) de Christiane Marchello-Nizia et Alexei Lavrentiev [34] à l'origine de la création du portail de la Base de français médiéval (BFM) [35] qui héberge des éditions en XML TEI de textes en ancien français. Grâce au média numérique, la *Queste* met en regard le texte avec sa source, tout en fournissant une transcription imitative avec les abréviations et une transcription normalisée. En outre, le texte est intégralement étiqueté : lemmes et POS (Part-of-Speech [36]) facilitant les recherches lexicales et des analyses textométriques.

12 Influencés par la *New Philology*, certains projets proposent des parcours de lecture où la matérialité des œuvres est centrale. Le projet *The Walt Whitman Archive* (1995– …)[37] ou encore *The Rossetti Archive* (1993-2008)[38] en sont des exemples. Le projet *The Complete Writings and Pictures of Dante Gabriel Rossetti* offre un parcours enrichi de numérisations d'un corpus où l'auteur a non seulement écrit les textes, émis des productions picturales (souvent en amont du texte et en lien étroit avec lui)[39], mais aussi élaboré la conception matérielle des livres, leur donnant une importance cruciale à la compréhension du processus créatif de l'auteur[40]. Plus récemment, le projet *Woolf Online* retrace l'univers mental de Virginia Woolf en mettant à disposition des lecteurs ses brouillons et des collections d'images. Pour les éditions génétiques, le numérique permet d'explorer les brouillons d'auteurs. Dans l'édition de *Frankenstein* de Mary Shelley du projet the *ShelleyGodwin Archive* (2013– …)[41], on peut parcourir les différentes étapes de l'écriture et identifier les mains responsables des corrections manuscrites[42]. En France, le projet *Bovary* (20022009)[43] propose un plan de la création de l'œuvre de Gustave Flaubert, permettant de retracer le processus d'écriture du roman. Ces projets, à travers la création de réseaux d'hyperliens et une documentation abondante, mettent en scène de manière dynamique le processus créatif. Elles promeuvent une textualité numérique ouverte et interactive en opposition au livre statique et fermé[44]. Toutefois, une telle abondance de documents demande un investissement important du lecteur, avec une navigation complexe pouvant entraîner une perte du sens au profit d'une navigation compulsive, comme le décrit A. Mangen :

> *The urge to click can easily become too tempting to resist, if we are cognitively or perceptually stimulated with possibilities that seem more exciting than what we are presently focused on. Knowledge sites have a wealth of potentials that can risk disrupting our phenomenological preoccupation with them, thereby limiting the possibility of hermeneutical reflection.* [45]

13 En outre, les systèmes de renvoi dépendent des interfaces en ligne, ce qui les rend vulnérables aux mises à jour techniques, menaçant ainsi leur maintenabilité.

14 Le passage au format numérique offre donc des opportunités inédites aux éditeurs et permet d'aborder des corpus de manière exhaustive, tout en engageant le lecteur dans des expériences interactives. Toutefois, la vigilance reste de mise pour éviter la prolifération d'informations sans valeur critique et assurer la pérennité des projets dans un environnement numérique en constante évolution.

Édition numérique : comment, pour qui, nouveaux enjeux ?

15 Les éditions numériques découlent d'une histoire décennale ayant forgé une communauté scientifique unie autour de la norme qu'est la TEI (*Text Encoding Initiative*). Non seulement elle a fourni les outils nécessaires à la création d'éditions numériques, mais elle a également été propice à la réflexion et aux échanges sur les questions relatives à l'édition des textes au sein d'une communauté dynamique et ouverte.

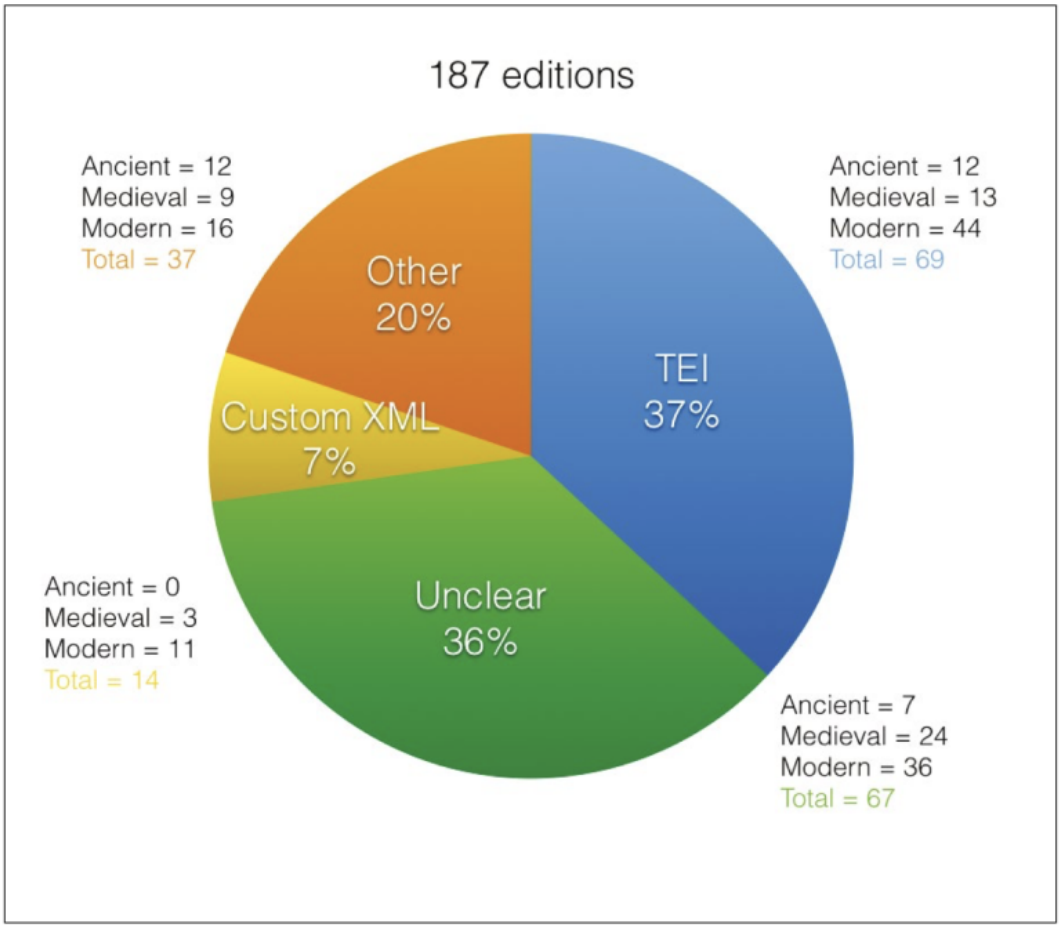
Un standard : XML TEI

16 La TEI a vu le jour en 1987 avec pour objectif de standardiser la représentation informatique des données textuelles et des sources historiques. La même année, lors d'une réunion réunissant des spécialistes des archives, des sciences computationnelles et des sciences des textes à Poughkeepsie (États-Unis), la TEI s'est dotée de *guidelines*[46]. Depuis 2003, elle repose exclusivement sur le langage XML (*eXtensible Markup Language*), qui fonctionne sur un système d'arborescence et d'imbrication strict. Son système de balisage permet d'enrichir le texte source d'informations aussi diverses que la hiérarchisation du texte, les entités nommées et les annotations linguistiques et plus encore, le tout dans un langage lisible à la fois par l'humain et par l'ordinateur. L'objectif de la TEI est de fournir une solution aussi bien aux débutants pour structurer un texte qu'aux experts cherchant une solution pour bâtir un corpus complexe. Elle offre suffisamment de liberté pour permettre à chacun de proposer un encodage adapté à sa source et à ses objectifs de recherche. L'esprit même de la TEI n'est pas de contraindre les éditeurs à un choix technique appliqué de manière mécanique, mais de permettre aux choix épistémologiques de rester centraux dans la modélisation

des données. Cependant, la contrepartie de cette liberté se traduit par des pratiques éditoriales, tout comme la philologie traditionnelle, qui ne sont pas uniformisées.

17 Enfin, bien que la plupart des projets cités dans cet article reposent sur des corpus encodés en XML TEI, il est indéniable que le standard n'est pas encore appliqué de manière systématique. Comme le soulignent G. Franzini, S. Mahony et M. Terras, seulement 37 % des éditions numériques parmi celles répertoriées dans un catalogue de 187 éditions suivent les prescriptions de la TEI [47].

Figure 1 : Répartition des technologies utilisées pour les éditions numériques, illustration issue de G. Franzini, S. Mahony et M. Terras, « A Catalogue of Digital Editions », 2016



18 Ainsi, malgré l'établissement de standards documentés et partagés par une large communauté, les pratiques de l'édition numérique ne

sont pas encore unifiées, ne serait-ce que d'un point de vue technologique.

Qui sont les « lecteurs » des éditions numériques ?

19 Il est relativement complexe de prévoir quels seront les usages et les lecteurs des éditions numériques. Non seulement le lectorat potentiel échappe à toute délimitation disciplinaire, mais la multiplicité de ses usages ouvre encore davantage les possibilités, entraînant même parfois une perte de contrôle de l'éditeur sur l'utilisation de son édition [48]. Bien que l'on qualifie souvent les « utilisateurs » d'éditions numériques sous le terme générique de « lecteurs », le terme même est contestable dans le cas d'une analyse textuelle basée sur du *distant reading* [49]. Voici une tentative de délimitation de trois types principaux d'utilisateurs/lecteurs en fonction du type d'usage du texte [50] :

- Le lecteur-consultant qualitatif, qui adopte une lecture linéaire du corpus similaire au support papier, soit pour lire une œuvre accessible en ligne, soit pour établir un corpus à analyser de manière qualitative.
- Le lecteur-explorateur pour les éditions basées sur du multifenêtrage ou des hyperliens, comme les projets The Rossetti Archive ou encore Woolf Online. Le lecteur y crée son propre parcours dont la complexité peut être déterminée en fonction de son degré d'expertise.
- Le lecteur-utilisateur de données. Ce sont des utilisateurs qui ne consultent pas les interfaces de lecture ou d'exploration, mais qui utilisent le texte brut, structuré, voire enrichi pour faciliter la fouille [51] via une API ou en récupérant directement les fichiers sources pour des études quantitatives ou en faire des données d'entraînement [52]. Le texte est alors traité comme des données [53].

20 L'éventail des utilisations possibles est très large, de la simple consultation à la fouille de texte, exigeant une donnée structurée et enrichie. Le traitement du texte comme donnée implique un passage d'une pratique individuelle visant à produire un objet fini à la production de données réexploitables dans un autre cadre que celui pour lequel elles ont été créées. Ainsi, l'éditeur numérique n'est plus simplement le producteur d'un texte clos sur lui-même, mais il

devient également un modélisateur de données à partager, d'où l'importance de l'utilisation de standards (caractères Unicode, encodage, etc.) pour en faciliter la réutilisation.

Nouvelles problématiques

21 Les éditions numériques demeurent aujourd'hui un ensemble de pratiques multiples, ce qui les rend difficiles à évaluer. En effet, tandis que « les éditions critiques au format imprimé répondent, pour leur part, à des modèles d'évaluation et de validation instaurés dès le XIX^e^ siècle [...]. Les éditions critiques numériques sont pour le moment peu nombreuses à faire l'objet de telles recensions » [54]. Les méthodes classiques d'évaluation ne s'appliquent pas et les éditeurs scientifiques se heurtent à une difficulté de reconnaissance de leur travail [55]. En France, des groupes de travail pour la mise en place de standards émergent, tels que l'ancien consortium Cahier [56], ou encore le consortium ARIANE-HN [57]. À l'international, on peut citer les *Guidelines for Editors of Scholarly Editions* [58] de la *Modern Language Association* qui proposent une grille pour guider les éditeurs sur les informations importantes à fournir sur le corpus et son établissement, ainsi que sur sa mise en ligne et la question de la préservation des données [59].

22 Comme le soulignait Frédéric Duval, la force du numérique réside dans sa modularité : « une édition numérique peut proposer des parcours plus ou moins ouverts, de la consultation libre d'un dépôt d'archives à des parcours commandés par des approches et intérêts divers [60] ». Cette modularité s'étend également au fait que l'œuvre peut être modifiée à l'infini [61]. Ces éditions pourront être maintenues au cours du temps par une succession d'experts [62]. Cependant, cette modularité engendre de nouvelles problématiques. Alors que les éditions scientifiques papier produisent un nouvel objet en guise de mise à jour, l'édition numérique est modifiée, allant parfois jusqu'à remplacer la version précédente, soulevant dès lors la question du *versioning*. Cette problématique a déjà été abordée par la TEI, qui permet la création d'un journal des versions dans l'élément <versioningDesc>, énumérant les dates et les responsabilités des changements, sans toutefois permettre de remonter à une version antérieure. Pour remédier à ce problème, E. Pierrazzo plaide en

faveur d'une gestion du *versioning* au sein des logiciels d'édition numérique, assurant ainsi l'accès aux versions antérieures [63]. Cette gestion du *versioning* peut également être réalisée par l'archivage des fichiers sources.

23 Enfin, si les éditions numériques n'entraînent pas de frais d'impression, les coûts d'hébergement sur un serveur et de maintenance ne doivent pas être sous-estimés. L'utilisation de systèmes informatiques rend ces objets vulnérables à l'obsolescence. Là où un manuscrit se conserve des centaines d'années, la conservation des objets numériques est plus complexe et son obsolescence liée à différents facteurs, tels que l'obsolescence « fonctionnelle » (fichiers endommagés), l'obsolescence de l'entité qui assurait la mise en ligne, ou encore l'obsolescence technologique. Face à ce constat, il nous semble essentiel d'assurer, avant tout, la pérennité des données de l'édition dans des fichiers au format le plus simple et standard possible, comme le XML TEI, dans des dépôts pérennes, et de leur associer des DOI pour faciliter leur accès et leur citabilité (Nakala [64], Zenodo, etc.).

24 Ainsi, l'édition numérique est encore en construction et cherche les critères de sa scientificité. La nouvelle économie dans laquelle elle s'inscrit n'est plus uniquement une économie du texte, mais une économie des données qui demande d'anticiper leur curation et la maîtrise des principes FAIR (*Findable*, *Accessible*, *Interoperable*, *Reusable*) [65].

Vers un changement d'échelle

25 Avec la montée en puissance de l'intelligence artificielle et l'automatisation des chaînes d'acquisition textuelle, une transformation significative se dessine en termes de taille des corpus. Alors que l'édition se limitait autrefois à une œuvre ou à une section cohérente d'un ensemble plus vaste, l'idée d'éditer des œuvres sérielles ou complètes gagne du terrain. Même la très établie TEI s'est adaptée et cherche à fournir un encodage de base comprenant des informations essentielles pour les corpus destinés à une lecture à distance [66]. Cette capacité à accumuler des données textuelles pourrait faciliter la conciliation entre des approches du texte en tant que document et du texte en tant que monument, en permettant

d'acquérir plus de texte plus vite et en facilitant, par exemple, une comparaison automatisée des témoins manuscrits, notamment par le biais d'une constitution semi-automatique du stemma. Cependant, cette nouvelle approche du texte comme données risque d'accentuer le fossé entre les philologues traditionnels et les humanistes numériques :

> As I have argued elsewhere, it is difficult to treat texts responsibly as "data" when much of our data set is inaccurate, whether because of faulty editing or because of the lack of digitization of certain types of texts, particularly those by what we might think of as non-canonical authors. The tensions between such approaches threaten to create splits between digital editing and digital humanities reminiscent of the textual studies wars of the second half of the twentieth century. [67]

26 C'est pourquoi il est important que les spécialistes du texte s'investissent dans ces problématiques afin de contribuer à l'élaboration d'outils numériques (modèles de reconnaissance automatique d'écriture, reconnaissance d'entités nommées, assistance à la collation, alignement de traduction, annotations linguistiques, etc.). L'objectif est de guider la communauté scientifique, qu'elle soit traditionnelle ou spécialisée dans les humanités numériques, pour rendre accessibles un nombre accru de textes, peut-être moins canoniques, tout en respectant les normes scientifiques établies par la communauté, sans laisser des géants commerciaux s'emparer d'une forme de numérisation plus rentable que scientifique des fonds patrimoniaux [68].

Acquisition automatique de texte

27 Grâce à des outils comme *Transkribus* [69] et *eScriptorium* [70], l'acquisition automatique de texte (ATR) permet de travailler sur des documents historiques complexes. Avec des corpus d'entraînement de qualité, les modèles de reconnaissance d'écriture manuscrite (HTR – Handwritten Text Recognition) peuvent atteindre des taux de précision de 92 % à 98 %, voire 99 %, faisant de l'acquisition automatique de texte une tâche résolue du point de vue informatique [71]. Actuellement, l'enjeu principal réside dans la production de données d'entraînement de qualité pour développer

des modèles répondant aux critères scientifiques du monde de la recherche.

28 Pour maximiser la qualité des transcriptions, il est impératif d'utiliser des données d'entraînement de première qualité, en établissant des normes de transcription. Ces normes incluent la distinction ou non des allographes, le traitement des abréviations et la préservation des ligatures, mais aussi le choix d'un set de caractères de préférences unicode pour représenter les sources, choix qui doivent être dictés par des critères scientifiques, la nature des sources et les objectifs scientifiques [72]. Le consortium pour la reconnaissance d'écriture manuscrite des matériaux anciens (CREMMA) a élaboré des réflexions approfondies sur la manière de transcrire les vérités de terrain [73]. Deux documents ont émergé de ces efforts : un guide dédié aux documents médiévaux [74] et un ensemble de recommandations pour les documents modernes [75].

29 D'importants jeux de données ont également été produits grâce à la collaboration entre différents projets tels que CREMMA, FoNDUE (FOrmes Numérisées et Détection Unifiée des Écritures) [76], Gallic(orpor)a [77], HTRomance [78] et HTRogène [79]. Ces corpus, englobant une diversité de sources du Moyen Âge aux imprimés du XX^e^ siècle, respectent les normes de transcription précitées, avec quelques ajustements en fonction de la spécificité des sources traitées. Ces corpus représentent une ressource précieuse pour l'entraînement de modèles d'ATR et ont permis l'entraînement de modèles génériques [80] :

- *Gallicorpora+*, pour les imprimés français du XVI^e^ au XIX^e^ siècle [81] ;
- *CATMuS Medieval*, modèle multilingue pour les manuscrits médiévaux compris entre le X^e^ et le XV^e^ siècle [82] ;
- *HTR-United - Manu McFrench* pour les écritures manuscrites modernes françaises [83].

30 Lors de l'utilisation d'un modèle HTR, il est crucial de prendre en compte la technologie sous-jacente, la compatibilité d'environnement et les scores de précision. Ces scores sont dérivés de la capacité du modèle à prédire une transcription sur un échantillon des données d'entraînement réservées à cette fin. La composition du corpus d'entraînement doit être soigneusement considérée pour interpréter correctement le score. Plus le corpus est homogène, plus les scores

peuvent être élevés, même avec un corpus d'entraînement restreint, car la précision est évaluée sur un échantillon très similaire au corpus d'entraînement. À l'inverse, plus un corpus est hétérogène, plus les scores élevés deviennent difficiles à atteindre, mais cela augmente les capacités du modèle à fonctionner sur un corpus inconnu. Les modèles génériques, tels que ceux mentionnés précédemment, conçus pour être utilisés sur de nouveaux corpus, peuvent également être affinés pour répondre à des besoins spécifiques avec un investissement relativement faible en données d'entraînement [84]. Ainsi, le défi actuel consiste à élaborer des modèles génériques capables de traiter des sources les plus variées possibles, objectif que le projet CATMuS [85] tente de relever en développant des modèles multilingues couvrant une vaste période historique.

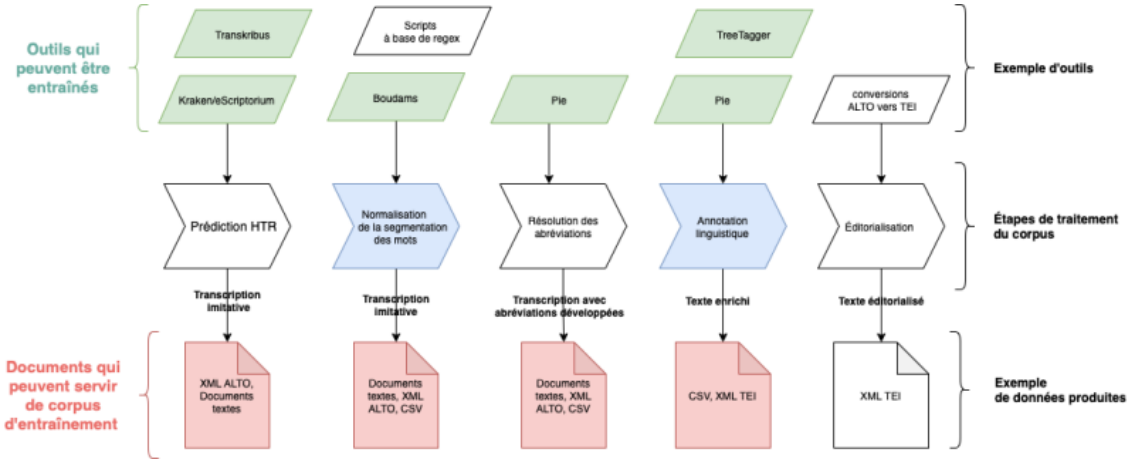
Les chaînes d'acquisition textuelle

31 L'acquisition automatique de texte à partir de sources numérisées a profondément transformé le rôle de l'éditeur dans la production du texte numérique. La capacité d'acquérir des textes à grande échelle, y compris ceux d'origine manuscrite, a conduit les projets de recherche à aborder des corpus de plus en plus vastes. En outre, l'assise actuelle du standard XML TEI facilite l'émergence de protocoles de publication en ligne [86], notamment avec l'apparition de *TEIPublisher* [87] qui s'appuie sur la modélisation des données en TEI. Enfin, la démocratisation d'outils comme *eScriptorium* ou *Trankribus* a également favorisé la mise en place de protocoles pour transformer les XML ALTO de l'ATR en XML TEI en vue de la mise en ligne des corpus [88]. Le projet *DiScholEd* illustre cette dynamique en visant à établir un protocole facilitant la production de textes et d'éditions en ligne à partir de leur numérisation [89]. Ainsi, l'édition en ligne de la *Correspondance de Constance de Salm* [90], dont le texte a été acquis automatiquement, propose un système de navigation, une édition critique et une édition diplomatique, ainsi qu'un travail sur les entités nommées, avec un accès facilité aux index et à des cartes. La flexibilité de cette approche et son adaptabilité à la source permettent de proposer des protocoles communs tout en évitant une standardisation excessive de l'interface qui ne serait pas en accord avec l'interprétation proposée par l'éditeur [91].

32 Les chaînes d'acquisition textuelle, intégrant une étape d'ATR ou non, peuvent être agrémentées d'une série de normalisations et d'enrichissements générés automatiquement [92], comprenant le développement des abréviations, la normalisation de la segmentation des mots, l'annotation linguistique et des entités nommées, l'ajout de métadonnées et proposer un balisage automatique vers la TEI en s'appuyant sur les informations de mise en page produites lors de l'ATR [93] (voir figure 2) pour optimiser la hiérarchisation, l'enrichissement et la fouille du texte.

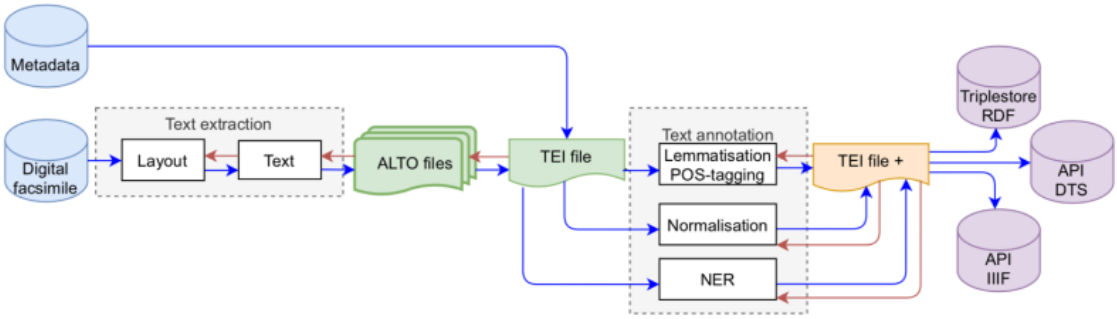
33 Le projet *Gallic(opor)a* [94] illustre ces ambitions, bien que seuls les objectifs liés à l'ATR et à la pré-éditorialisation en TEI aient pu être atteints dans le cadre du financement initial.

Figure 2 : Exemple de protocole, issu de Ariane Pinche, « Guide de transcription pour les manuscrits du X^e^ au XV^e^ siècle », 2022



34 Des initiatives plus récentes, telles que la thèse de Matthias Gille Levenson [95], explorent des chaînes éditoriales semi-automatiques incluant une assistance à la collation. Son logiciel, *teiCollator* [96], décompose le processus en plusieurs étapes, de la transcription du témoin à la réalisation d'une édition critique consultable en PDF via LaTeX.

***Figure 3 : Chaîne de traitement des données dans le projet Gallic(orpor)a,* issue d'Ariane Pinche, Kelly Christensen et Simon Gabay, « Between automatic and manual encoding Towards a generic TEI model for historical prints and manuscripts », *TEI Conference and Members&apos ; Meeting*, Newcastle, 2022.**



35 Ainsi, l'évolution rapide des techniques d'ATR et des chaînes éditoriales numériques redéfinit le paysage de l'édition, plaçant l'éditeur au cœur d'une démarche de traitement de vastes corpus. L'acquisition automatisée permet désormais d'explorer des textes, qu'ils soient originalement imprimés ou manuscrits, à une échelle inédite, qui appelle une évolution des pratiques éditoriales. Les chaînes éditoriales, de plus en plus sophistiquées, englobent des métadonnées variées et des outils automatisés, créant ainsi un écosystème éditorial plus complet. En outre, l'éditeur, désormais confronté à des corpus de grande envergure, est invité à collaborer avec des experts en intelligence artificielle, traitement automatique du langage, et autres domaines connexes. Cette évolution, loin de reléguer la philologie traditionnelle au second plan, offre l'opportunité de construire des méthodes éditoriales solides, tout en explorant des approches de lecture adaptées à divers profils d'utilisateurs. Ainsi, en embrassant ces transformations, l'éditeur peut non seulement relever les défis de l'ère numérique, mais également contribuer activement à l'élaboration d'une édition numérique accessible, riche et en constante évolution.

36 Les évolutions marquantes et les débats philologiques ont façonné l'édition numérique depuis les années quatre-vingt jusqu'à nos jours [97]. Tout au long de cette analyse, nous avons souligné la pertinence de l'édition numérique dans les débats philologiques. Notre examen a mis en lumière les modifications induites par les

innovations techniques, transformant la perception du texte et redéfinissant l'objet éditorial. Cependant, ces avancées n'ont pas été sans susciter de nouvelles difficultés, perturbant des méthodes séculaires et demandant encore aujourd'hui de réfléchir à l'évaluation, la pérennisation, l'accessibilité et la citabilité des éditions numériques. Enfin, nous espérons avoir montré comment le passage au format numérique, suivi de l'émergence des chaînes éditoriales automatisées, représente aujourd'hui une opportunité inédite. Cette évolution permet la constitution de corpus d'une ampleur sans précédent et facilite l'accès aux textes, répondant ainsi aux besoins d'un public toujours plus vaste. Toutefois, cette transition souligne également la nécessité de continuer à adapter nos pratiques et réflexions aux défis posés par ces nouvelles méthodologies, afin de garantir la qualité des éditions numériques dans le paysage académique.

NOTES

1 Notre propos ne traitera pas de l'écriture numérique ni des circuits d'édition commerciale. Édition scientifique désigne, ici, les éditions critiques qui s'intéressent à l'histoire des différents témoins du texte, les éditions génétiques qui s'intéressent aux différentes strates du texte au cours de son processus de création, mais également les éditions qui donnent un contexte à l'œuvre : matérialité du support, illustrations, réception, etc.

2 Jerome McGann, *Radiant textuality: literature after the World Wide Web*, New York, 2001, p. 55.

3 Hans Walter Gabler, « 6. Theorizing the Digital Scholarly Edition », dans *Id.*, *Text Genetics in Literary Modernism and Other Essays*, Cambridge, Open Book Publishers, 2019, p. 121-141.

4 Sotera Fornaro, « Karl Lachmann et sa méthode », *Revue germanique internationale*, 2011, p. 125-138. Paolo Trovato et Michael D. Reeve, *Everything you always wanted to know about Lachmann's method: a non-standard handbook of genealogical textual criticism in the age of post-structuralism, cladistics, and copy-text*, Limena, Libreriauniversitaria.it edizioni, 2014.

5 Richard J. Tarrant, « Classical Latin litterature », dans David C. Greetham (dir.), *Scholarly editing : a guide to research*, New York, The Modern Language Association of America, 1995, p. 96.

6 Richard J. Tarrant, *op. cit.*

7 Bernard Cerquiglini, *Éloge de la variante : histoire critique de la philologie*, Paris, Seuil, 1989.

8 Elena Pierazzo, « 3. Modelling Digital Scholarly Editing: From Plato to Heraclitus », dans Matthew James Driscoll, (dir.), *Digital Scholarly Editing: Theories and Practices*, Cambridge, Open Book Publishers, 2017, p. 41-58.

9 Peter Shillingsburg, « Author, Texts, and Polemics of Textual Criticism », dans *Id.*, *Devils and angels: Textual editing and literary theory*, Charlottesville, 1991, p. 26.

10 John Bryant, « Witness and Access: The Uses of the Fluid Text », *Textual Cultures*, vol. 2 1, 2007, p. 16-42.

11 « La critique génétique envisage l'œuvre littéraire comme processus d'écriture ; […] en conséquence, son objet n'est pas le texte édité et clos, mais la diversité et la multiplicité des traces de ce processus, qui, comme des témoins muets, nous sont livrées dans les fonds d'archives littéraires, notamment des auteurs modernes. » : Bénédicte Vauthier, « Éditer des états textuels variants », *Genesis. Manuscrits – Recherche – Invention*, n° 44, mai 2017, p. 39-55., texte traduit de l'allemand d'Almuth Grésillon, « "Critique génétique" : Probeartikel », Beihefte zu Editio, n° 36, 2013, p. 197.

12 Jerome McGann, *op. cit.*, p. 56.

13 Dans les éditions des œuvres de Tertullien ou l'édition des commentaires de Servius, l'apparat peut occuper jusqu'au trois quarts de la page, rendant la lecture du texte fastidieuse. Voir Tertullien, *Le manteau*, éd. Marie Turcan, Paris, France, Le Cerf, 2007. Paul Wessner (éd.), *Donat, Aeli Donati quod fertur Commentum Terenti : Accedunt Eugraphi commentum et Scholia Bembina*, Leipzig, B. G. Teubner, 1905.

14 « Pour des raisons économiques, le papier est contraint de se limiter à un seul état textuel (*single text editions, one text editions*), qu'il soit ou non reconstruit. Grâce au "multifenêtrage" et à l'hypertexte, le numérique s'est imposé comme le médium approprié et indispensable à l'application des "nouvelles" théories textuelles. » : Frédéric Duval, « Pour des éditions numériques critiques. L'exemple des textes français », *Médiévales*, n° 73,

2017, Le *texte à l'épreuve du numérique*, dossier thématique sous la direction d'Anne Rochebouet, p. 13-30.

15 Pierre Mounier, « Manifeste des *Digital Humanities* », *Journal des anthropologues*, n° 122-123, 2010, p. 447-452.

16 Hans Walter Gabler, *op. cit.*

17 « *It pushes traditional scholarly models of editing and textuality beyond* » : Dino Buzzetti et Jerome McGann, « Critical Editing in a Digital Horizon », dans Lou Burnard, John M. Unsworth et Katherine O'Brien O'Keeffe (dir.), *Electronic textual editing*, New York, Modern Language Association of America, 2006.

18 Patrick Sahle, « 2. What is a Scholarly Digital Edition? », dans Matthew James Driscoll et Elena Pierazzo (dir.) *Digital Scholarly Editing: Theories and Practices*, Cambridge, Open Book Publishers, 2017, p. 19-39.

19 Jonas Carlquist, « Medieval Manuscripts, Hypertext and Reading. Visions of Digital Editions », *Literary and Linguistic Computing*, n° 19 / 1, avril 2004, p. 105-118. G. Franzini, S. Mahony et M. Terras, « A Catalogue of Digital Editions », dans *Matthew James Driscoll et Elena Pierazzo* (dir.), *op.cit.*, p. 161-182.

20 Peter Robinson, « Towards a Theory of Digital Editions », *Variants. The Journal of the European Society for Textual Scholarship*, n°10, 2013, p. 123.

21 « Éditions en ligne de l'École des chartes (ELEC) », consulté le 3 janvier 2024, (http://elec.enc.sorbonne.fr/).

22 Mats Dahlström, « How Reproductive is a Scholarly Edition? », *Literary and Linguistic Computing*, n°19/1, avril 2004, p. 17-33.

23 Peter Robinson et Elizabeth Solopova, « Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue », *The Canterbury Tales Project. Occasional Papers*, juillet 1993. Edition en ligne consultée le 3 janvier 2024 (https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=572872612f33aaae5717c128dc2d73567e539e76)

24 Peter Robinson, « What is a critical digital edition? », *Variants: The Journal of the European Society for Textual Scholarship*, n°1, janvier 2002, p. 43-62.

25 Elena Pierazzo, *Digital scholarly editing: theories, models and methods*, Londres, Routledge, 2015.

26 Ariane Pinche, Bruno Bureau et Christian Nicolas, « Hyperdonat, digital edition project », *TEI Conference and Members' Meeting 2016*,

septembre 2016, Vienne (Autriche) ⟨hal– 01413479⟩. Le projet est resté au stade expérimental.

27 Richard Trachsler et Lino Leonardi, « L'édition critique des romans en prose : le cas de Guiron le Courtois », *Manuel de la philologie de l'édition*, Berlin/Boston, De Gruyter, 2015, p. 44-80.

28 <https://obvil.sorbonne-universite.fr/bibliotheque>.

29 Voir 2.1 *Un standard : XML TEI.*

30 <https://github.com/OBVIL>.

31 Frédéric Glorieux, « Dramagraphie (sur le pamphlet 6 de Franco Moretti) », *J'attends des résultats*, 2016, en ligne et consulté le 8 octobre 2024 (https://doi.org/10.58079/tod3).

32 Gregory Crane, « The Perseus Digital Library and the future of libraries », I*nternational Journal on Digital Libraries*, n° 24, juin 2023, p. 117-128, en ligne consulté le18 décembre 2023 (https://www.perseus.tufts.edu/hopper/ ; https://doi.org/10.1007/s00799-022-00333-2)

33 *Uniform Resource Name*, standard informatique qui permet d'identifier une ressource indépendamment de sa localisation et de son accessibilité par internet, ce qui permet à cet identifiant d'être pérenne.

34 Christiane Marchello-Nizia, Alexey Lavrentiev et Céline Guillot-Barbance, « 6. Édition électronique de la Queste del saint Graal », *Manuel de la philologie de l'édition*, Berlin/Boston, De Gruyter, 2015, p. 155-176 ; Christiane Marchello-Nizia et Alexei Lavrentiev (éd.), *Queste del saint Graal*, Lyon, ENS éditions, 2019. Publié en ligne par la Base de français médiéval, http://catalog.bfm-corpus.org/qgraal_cm, Dernière révision le 2018-11-30.

35 Céline Guillot, Serge Heiden et Alexei Lavrentiev, « Base de français médiéval : une base de référence de sources médiévales ouverte et libre au service de la communauté scientifique », *Diachroniques. Revue de Linguistique. française diachronique*, n° 7, 2018, p. 168.

36 Soit un équivalent de ce que nous appelons la nature grammaticale.

37 Michael Robertson, « The Walt Whitman Archive », *Journal of American History*, vol. 99/3, décembre 2012, p. 1019-1020. Article consulté en ligne le 18 décembre 2023 (https://whitmanarchive.org/about/index.html).

38 Jerome McGann, « Dante Gabriel Rossetti and the Betrayal of Truth », *Victorian Poetry*, vol. 26/4, 1988, p. 339-361. Article consulté en ligne le 18 décembre 2023 (http://www.rossettiarchive.org/index.html).

39 Voir le cas du tableau et du poème : « The Girlhood of Mary Virgin », produits sensiblement à la même époque (1847–1848), en ligne consulté le 25 mars 2024 (http://www.rossettiarchive.org/docs/s40.rap.html).

40 On peut également citer le projet *Blake archive* qui donne accès aux illustrations de ces œuvres « illuminées » pour rendre au mieux la richesse des œuvres de l'auteur (https://www.blakearchive.org).

41 « Shelley-Godwin Archive », site consulté le 20 décembre 2023 (http://shelleygodwinarchive.org/about/). Le projet est le résultat d'un partenariat entre la New York Public Library et le Maryland Institute for Technology in the Humanities, en coopération avec la Bodleian Library d'Oxford, la Huntington Library, la British Library, la Houghton Library et le Victoria and Albert Museum.

42 Exemple de page issu du projet : « Shelley, M. "Frankenstein, MS. Abinger C. 56 », dans The Shelley-Godwin Archive, MS. Abinger c. 56, 2v. Retrieved from http://shelleygodwinarchive.org/sc/oxford/ms_abinger/c56/#/p8.

43 « Édition des manuscrits de Madame Bovary de Flaubert | Transcriptions | Classement génétique », en ligne et consulté le 20 décembre 2023 (https://www.bovary.fr/).

44 Jerome McGann, *op. cit.*, p. 25.

45 Anne Mangen, « Hypertext fiction reading: haptics and immersion », *Journal of Research in Reading*, vol. 31/4, 2008, p. 404-419.

46 Nancy M. Ide et Michael Sperberg-McQueen, « The TEI: History, Goals, and Future », *Computers and the Humanities*, vol. 29/1, 1995, p. 515.

47 Greta Franzini, Simon Mahony et Mélissa Terras, « A Catalogue of Digital Editions », dans Matthew James Driscoll et Elena Pierazzo (dir.), *Digital Scholarly Editing : Theories and Practices*, Cambridge, Open Book Publishers, 2017, p. 161-182.

48 Daniel Apollon et Claire Bélisle, « Le destin de l'appareil critique dans l'édition numérique scientifique », dans Daniel Apollon, Philippe Régnier et Claire Bélisle (dir.), *L'édition critique à l'ère du numérique*, Paris, L'Harmattan, 2017, p. 105.

49 Franco Moretti, *Distant reading*, Londres, Verso, 2013. Le *distant reading* propose d'appréhender le texte de manière quantitative avec des analyses de réseaux, des analyses lexicométrique ou stylométrique.

50 Ioana Galleron et Fatiha Idmhand, « De l'interopérabilité à la réutilisabilité des éditions électroniques », *Humanités numériques*, n° 1, 2020, mis en ligne le 1er janvier 2020, consulté le 10 octobre 2024 (http://journals.openedition.org/revuehn/350 ; DOI : https://doi.org/10.4000/revuehn.350).

51 Identification des entités nommées, annotations morphosyntaxiques.

52 Ariane Pinche, « Exploitations et valorisations des données numériques connexes à l'édition », dans Robert Alessi, Marcello Vitali-Rosati (dir.), *Les éditions critiques numériques : entre tradition et changement de paradigme*, Montréal, Presses de l'Université de Montréal, 2023, p. 133-153.

53 Paul A. Broyles, « Digital Editions and Version Numbering », *Digital Humanities Quarterly*, vol.14/2, juin 2020, article en ligne consulté le 11 octobre 2024 (https://hcommons.org/deposits/objects/hc:38104/datastreams/CONTENT/content) : « Digital editions, as I understand them, are simultaneously scholarly publications and data sources ».

54 Joana Casenave, « La fonction de l'éditeur-auteur dans les éditions critiques numériques », *Humanités numériques*, n° 6, 2022, mis en ligne le 1er décembre 2022, consulté le 14 octobre 2024 (http://journals.openedition.org/revuehn/3241).

55 Peter Robinson, « Electronic editions for everyone », *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*, Openbooks Publisher, 2010, p. 145-163.

56 « Consortium Cahier – Corpus d'auteurs pour les humanités : informatisation, édition, recherche » en ligne, consulté le 15 octobre 2024 (https://cahier.hypotheses.org/).

57 « Ariane », document en ligne consulté le 22 décembre 2023 (https://cst-ariane.huma-num.fr/).

58 « Guidelines for Editors of Scholarly Editions | Modern Language Association », document en ligne consulté le 8 décembre 2023 (https://web.archive.org/web/20190803225011/https://www.mla.org/Re
Reports-and-Other-Documents/Publishing-and-Scholarship/Reports-from-the-MLA-Committee-on-Scholarly-Editions/Guidelines-for-Editors-of-Scholarly-Editions).

59 Philippe Régnier, « Les enjeux de l'édition critique numérique », dans Daniel Apollon, Philippe Régnier, Claire Bélisle (dir.), *L'édition critique à l'ère du numérique*, Paris, L'Harmattan, 2017, p. 79-100.

60 Frédéric Duval, *op. cit.*

61 Jerome McGann, *op. cit.*

62 Peter Robinson, « Where We Are with Electronic Scholarly Editions, and Where We Want to Be », *Jahrbuch für Computerphilologie*, vol. 5/5, 2003, p. 126-146 : « *These editions will not be made or maintained by one person or by one group, but by a community of scholars and readers working together: they will be the work of many and the property of all.* »

63 Elena Pierazzo, *op. cit.*, p. 185-186.

64 Nakala est un entrepôt de données de recherche pour les sciences humaines et sociales et offre des services sur plusieurs étapes du cycle de vie des données de recherche en SHS : sur leur préservation, leur publication et leur réutilisation. Le service est maintenu par Huma-Num.

65 <https://www.go-fair.org/fair-principles/>.

66 Lou Burnard, Christof Schöch et Carolin Odebrecht, « In search of comity: TEI for distant reading », *Journal of the Text Encoding Initiative* [en ligne], n°14, April 2021 – March 2023, mis en ligne le 8 juillet 2021, consulté le 15 octobre 2024 (http://journals.openedition.org/jtei/3500).

67 Amy E. Earhart, « The Digital Edition and the Digital Humanities », *Textual Cultures*, vol. 7 / 1, 2012, p. 18-28.

68 Robert Darnton, A*pologie du livre* : *demain, aujourd'hui, hier*, trad. Jean-François Sené, Paris, Gallimard, 2010, p. 100-109.

69 Philip Kahle, Sebastian Colutto, Günter Hackl, *et ali*, « Transkribus - A Service Platform for Transcription, Recognition and Retrieval of Historical Documents », *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 04, 2017, p. 19-24. En ligne (https://ieeexplore.ieee.org/abstract/document/8270160)

70 Peter A. Stokes, Benjamin Kiessling, Daniel Stökl Ben Ezra, *et al.*, « The eScriptorium VRE for Manuscript Cultures », *Classics@ Journal*, vol.18, 2021, en ligne (https://classics-at.chs.harvard.edu/classics18-stokes-kiessling-stokl-ben-ezra-tissot-gargem/).

71 « *From a computer science point of view, the recognition of handwriting seems to be a resolved task. The latest recognition engines allow for the*

successful recognition of specifically trained hands producing a text as reusable data. », Tobias Hodel, David Schoch, Christa Schneider, *et ali*, « General Models for Handwritten Text Recognition: Feasibility and State-of-the Art. German Kurrent as an Example », *Journal of Open Humanities Data*, juillet 2021, p. 13, en ligne (https://boris.unibe.ch/157474/3/46-597-1-PB.pdf).

72 En effet, comme le remarquait déjà Peter Robinson le passage au format numérique demande un effort de traduction en caractère dédié. Dans la lignée de ces réflexions, on peut également cité le projet TypoRef : Rémi Jimenes, « TypoReF : étudier les matériels d'imprimerie de la Renaissance française », *Bibliothèques Humanistes*, en ligne depuis le 24 mai 2023 (https://bvh.hypotheses.org/9654).

73 Le projet CREMMA est un projet partenaire de l'équipe Scripta (EPHE-PSL) à l'origine du développement de *eScriptorium* et de *Kraken* (Benjamin Kiessling, « Kraken – an Universal Text Recognizer for the Humanities », Abstract of paper 0673 presented at the Digital Humanities Conference 2019 (DH2019), Utrecht, the Netherlands 9-12 July, 2019, en ligne (https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/Z9G2EX)).

74 Ce guide est le fruit d'un séminaire de recherche qui s'est déroulé à l'École nationale des chartes en 2021-2022 : Ariane Pinche, « Guide de transcription pour les manuscrits du X^{e} au XV^{e} siècle », 2022, en ligne (⟨hal-03697382⟩).

75 Alix Chagué, Thibault Clérice et CREMMA, « Règles générales de transcription pour les corpus CREMMA » article en ligne depuis le 23 septembre 2022 (https://gist.github.com/alix-tz/6f89444521bf1cab0522da520f7e4ff4.)

76 Simon Gabay, Pierre Kuenzli, Jean-Luc Flacone et Christophe Charpilloz, « FoNDUE : Documentation, University of Geneva », en ligne depuis 2022 (https://github.com/fonDUE-HTR/Documentation).

77 Benoît Sagot, Laurent Romary, Rachel Badwen, *et ali*, « Gallic(orpor)a: extraction, annotation et diffusion de l'information textuelle et visuelle en diachronie longue », en ligne depuis 2022 et consulté le 16 décembre 2023 (https://gallicorpora.github.io/).

78 « The HTRomance project », en ligne et consulté le 10 décembre 2023 (https://htromance-project.github.io/).

79 « HTRogène », en ligne, consulté le 10 décembre 2023 (https://projet.bibliss ima.fr/fr/appels-projets/projets-retenus/htrogene).

80 Tous les modèles cités ci-dessous ont été entraînés avec Kraken.

81 Ariane Pinche et Simon Gabay, « Segmentation and HTR Model », *Gallicorpora*, 2023, en ligne, consulté le 16 décembre 2023 (https://github.com/Gallicorpora/ Segmentation-and-HTR-Models).

82 Ariane Pinche, Thibault Clérice, Alix Chagué, *et ali*, « CATMuS Medieval », *Zenodo*, en ligne depuis novembre 2023 (https://zenodo.org/records/12743230).

83 Alix Chagué et Thibault Clérice, « HTR-United - Manu McFrench V1 (Manuscripts of Modern and Contemporaneous French) », *Zenodo*, en ligne depuis juin 2022.

84 Environ cinq images contenant une trentaine de lignes suffisent, pour un coût de calcul relativement modéré, ce qui permet d'entraîner le modèle sur une machine personnelle sans nécessiter de carte graphique ou de serveurs dédiés, voir Ariane Pinche, « Generic HTR Models for Medieval Manuscripts. The CREMMALab Project », *Journal of Data Mining & Digital Humanities*, en ligne depuis le 16 octobre 2023 (https://jdmdh.episciences.org/11592).

85 Ariane Pinche, Thibault Clérice, Alix Chagué, *et ali*, « CATMuS-Medieval: Consistent Approaches to Transcribing ManuScripts: A generalized set of guidelines and models for Latin scripts from Middle Ages (8th – 16th century) », *DH2024*, ADHO, Aug 2024, Washington DC, en ligne ⟨hal-04346939⟩.

86 Anne Baillot et Julie Giovacchini, « TEI Models for the Publication of Social Sciences and Humanities Journals: Opportunities, Challenges, and First Steps Toward a Standardized Workflow », *Journal of the Text Encoding Initiative*, en ligne depuis mars 2023 (https://journals.openedition.org/jtei/3419).

87 Amit Kumar, Susan Schreibman, Stewart Arneil, *et ali*, « <teiPublisher>: A Repository Management System for TEI Documents », *Literary and Linguistic Computing*, vol. 20/1, mars 2005, p. 117-132.

88 Hugo Scheithauer, Alix Chagué et Laurent Romary, « From eScriptorium to TEI Publisher », *Brace your digital scholary edition!*, en ligne depuis novembre 2021 (hal-03538115).

89 La chaîne a été mise en place par Floriane Chiffoleau, sous la supervision d'Anne Baillot et de Laurent Romary, avec l'aide d'Alix Chagué et

de Manon Ovide.

90 Exemple : « Letter of Constance de Salm to the King of Prussia Friedrich Wilhelm III. (Dyck, June 24th 1824) ». Encadrement scientifique et technique du projet Anne Baillot Mareike König Floriane Chiffoleau. Réalisation de la chaîne de traitement par Sébastien Biay, <https://dischole.huma-num.fr/exist/apps/dischole/index_cds.html?collection=cds%2Fcorpus>.

91 Edward Vanhoutte, « 5. Defining Electronic Editions: A Historical and Functional Perspective », *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*, Cambridge, Open Book Publishers, 2013, p. 119-144.

92 Des contrôles qualité sont toutefois vivement conseillés.

93 Le projet *SegmOnto* propose un vocabulaire contrôlé pour décrire ces « zones », favorisant la création de jeux de données compatibles, la collaboration entre différents projets, et la mise en place de scripts de transformation, Simon Gabay, Ariane Pinche, Kelly Christensen, *et ali*, « SegmOnto: A Controlled Vocabulary to Describe and Process Digital Facsimiles », 2023, en ligne ⟨hal– 04343404⟩.

94 Ariane Pinche, Kelly Christensen et Simon Gabay, « Between automatic and manual encoding Towards a generic TEI model for historical prints and manuscripts », *TEI 2022 conference: Text as data*, en ligne depuis 2022 ⟨hal– 03780302⟩. Kelly Christensen, Ariane Pinche et Simon Gabay, « Gallic(orpor)a: Traitement des sources textuelles en diachronie longue de Gallica », DataLab de la BnF, Paris, 2022, en ligne ⟨hal– 03716534⟩.

95 Matthias Gille Levenson, *Le Regimiento de los prínçipes et sa glose : étude et édition numérique de la partie sur le gouvernement de la cité en temps de guerre (III, 3)*, thèse de doctorat, École normale supérieure de Lyon, 2023.

96 Matthias Gille Levenson, « TeiCollator : une chaîne de traitement ecdotique semi-automatisée », XXX^{e} *Congrès International de Linguistique et Philologie Romane*, 2022, en ligne (hal– 03715059).

97 Cet exposé s'est efforcé de dresser un vaste panorama des possibilités offertes par les éditions numériques, en présentant une gamme variée de projets allant des pionniers de l'édition numérique à des projets actuels. Afin d'illustrer les cas pratiques de la troisième partie, nous avons mis en avant des initiatives avec lesquelles nous avons collaboré ou échangé, et dont les fichiers XML TEI, ainsi que les scripts et la documentation, sont librement accessibles.

ABSTRACTS

Français

La philologie est une discipline complexe et variée. L'introduction de cet exposé aborde rapidement les diverses pratiques de l'édition savante, mettant en évidence les tensions méthodologiques entre les approches centrées sur le texte et sa transmission et celles davantage orientées vers le document et sa matérialité. À travers un panorama des parcours de lecture proposés par les pionniers de l'édition numérique tels que *Perseus*, *La Queste del saint Graal* ou encore *The Rossetti Archive*, nous examinons les liens qu'ils entretiennent avec les débats de la philologie traditionnelle, ainsi que les réponses que les éditions numériques ont cherché à apporter, de la simple mise à disposition des textes à la création de portails mettant en avant le processus créatif des auteurs. Dans un second temps, nous exposons les critères scientifiques, les standards tels que XML TEI, et les usages des éditions numériques. Cette partie met en lumière les nouveaux enjeux auxquels l'édition numérique doit faire face, notamment dans la transition du texte vers des données pérennes, lisibles et exploitables par les ordinateurs, ouvrant ainsi le texte à des usages inconnus des éditions traditionnelles. Enfin, nous abordons l'impact des innovations technologiques sur les systèmes de production textuelle. L'utilisation de l'intelligence artificielle entraîne des changements méthodologiques dans le domaine de la philologie, notamment en termes de taille de corpus, en facilitant son acquisition et son enrichissement. Ce changement amène également de nouvelles pratiques de traitement du texte avec la mise en place de chaînes d'acquisition plus ou moins automatisées de la numérisation des sources à la mise à disposition d'un texte numérique.

English

Philology is a complex and diverse discipline. This introduction of this article briefly explores various scholarly editing practices, highlighting methodological tensions between approaches centered on the text and its transmission, and those more focused on the document and its materiality. Through a survey of reading pathways offered by digital editing pioneers such as Perseus, La Queste del saint Graal, and The Rossetti Archive, we examine the connections they have with traditional philological debates, as well as the responses digital editions have sought to provide: from simply making texts available to creating portals that showcase authors' creative processes. Next, we delve into the scientific criteria, standards such as XML TEI, and the uses of digital editions. This section sheds light on the new challenges digital editing faces, especially in transitioning text into data that is durable, readable, and usable by computers, thereby opening up the text to uses unknown to traditional editions. Finally, we discuss the impact of technological innovations on textual production systems. The use of artificial intelligence brings about methodological changes in the field of

philology, particularly in terms of corpus size, facilitating its acquisition and enrichment. This shift then introduces new text processing practices with the implementation of more or less automated acquisition pipelines, from digitizing sources to providing digital text.

INDEX

Mots-clés
édition, édition numérique, XML TEI, chaîne éditoriale, reconnaissance automatique de texte

Keywords
edition, digital édition, XML TEI, editorial pipeline, automatic text recognition

AUTHOR

Ariane Pinche
CNRS, CIHAM UMR 5648
IDREF : https://www.idref.fr/253121914
ORCID : http://orcid.org/0000-0002-7843-5050
HAL : https://cv.archives-ouvertes.fr/ariane-pinche
BNF : https://data.bnf.fr/fr/18166519

RePPOL : un projet d'édition numérique d'une enquête sur les croyances et les pratiques religieuses dans le Kent (Angleterre, 1543)

RePPOL, a digital edition of an investigation on religious beliefs and practices in Kent (England, 1543)

Olivier Spina

DOI : 10.35562/theia.114

Copyright
CC BY-NC-SA 4.0

OUTLINE

Introduction
Pourquoi éditer le manuscrit 128 ?
  Une enquête judiciaire
  Un projet aux ambitions scientifiques et pédagogiques
La transcription : grandeur et décadence d'un projet collaboratif
  Le choix de la plateforme TACT
  Le volet pédagogique : demi-échec ou semi-réussite ?
  Un exemple de contrainte : faire coexister deux transcriptions
Les défis de l'édition numérique
  Quelle édition pour quels moyens techniques ?
  La complexité du balisage structurel
  Des balisages sémantiques non automatisés

TEXT

Introduction

1 Cet article propose un retour d'expériences à l'issue du projet *Rethinking the Prebendaries Plot On Line* (RePPOL) mené entre 2019 et 2023 en collaboration avec Aude de Mézerac-Zanetti [1]. Ce projet a été financé par plusieurs institutions : l'I-Site de l'Université de Lille (équivalent d'un labex régional), la Maison Européenne des Sciences Humaines et Sociales de Lille (MESHS), le Labex lyonnais COMOD et

le Laboratoire de Recherche Historique Rhône-Alpes (LARHRA UMR 5190).

2 RePPOL est un projet d'histoire numérique au croisement de la recherche et de l'enseignement visant à transcrire, puis éditer sous format numérique un manuscrit d'environ 400 folios conservé à Cambridge dans le fonds dit de la « Parker Library » du Corpus Christi College (ms. 128). Cette édition entendait s'accompagner d'un traitement scientifique de la source en anglais et en français : introduction, bibliographie, production d'indices, de systèmes d'informations géographiques (SIG), de bases de données et de visualisation de réseaux, ce qui nécessitait d'opérer un balisage complexe du texte transcrit. Depuis décembre 2022, l'édition est accessible à la communauté scientifique selon les principes du FAIR (Findable / Accessible / Interoperable / Reusable)[2], en téléchargement au format LateX ou en format HTML (Hypertext Markup Language). Cette édition est, pour le moment, hébergée sur le site de la MESHS de Lille (https://reppol.meshs.fr/).

3 Si RePPOL a pu être lancé et mené à terme, c'est en grande partie grâce à la Maison Européenne des Sciences Humaines et Sociales de Lille. En 2018, les instances universitaires lilloises ont développé une stratégie de soutien et d'implication directe dans les projets d'humanités numériques portés par les enseignants-chercheurs de leurs établissements, quelles que soient les disciplines. Il s'agissait non seulement de prodiguer des conseils, mais aussi d'aider les chercheurs à amorcer et conduire ces projets, en leur fournissant des moyens humains et matériels. Pour ce faire, la MESHS a recruté en 2019 plusieurs ingénieurs d'études spécifiquement formés aux humanités numériques et dont la mission principale était l'appui à la recherche. Deux d'entre elles ont largement contribué au projet RePPOL : Victoria Le Fourner a chapeauté l'ensemble des aspects techniques en lien avec les humanités numériques (dont la rédaction de l'ODD) et Florence Perret s'est occupée, de son côté, des questions de visualisation de l'édition numérique, particulièrement la charte graphique du projet[3].

4 Avant de porter le projet RePPOL, Aude de Mézerac-Zanetti et moi-même n'avions aucune compétence ou formation particulières en humanités numériques. Dans les premiers temps du projet, trois

choix structuraux ont été faits, souvent inconsciemment, et se sont rétrospectivement révélés être des contraintes tant dans la définition du périmètre scientifique du projet que dans les solutions techniques mobilisées pour le mener à bien. Le premier résidait dans le choix d'éditer un manuscrit d'une taille modeste, comparée aux grands corpus qui sont au cœur de la majeure partie des projets d'humanités numériques. Toutefois, malgré sa taille, la structure relativement complexe du document (pluralité de scripteurs, diversité des pièces juridiques, etc.) complexifia son traitement [4]. Deuxièmement, RePPOL a été pensé comme un projet alliant production d'un savoir académique et formation par la recherche d'étudiants de niveau master. De ce fait, RePPOL se définissait comme une expérience collaborative, ce qui requit d'avoir recours à un nombre important d'outils numériques de natures diverses, mais aussi d'affiner nos réflexions sur nos propres pratiques de recherche. Enfin, le dernier choix était d'ordre financier. En 2018, nous envisagions RePPOL comme un projet d'ampleur réduite, ne nécessitant qu'un budget limité [5]. L'absence de compétences initiales nous rendait peu à même d'évaluer clairement nos besoins techniques, humains et financiers. Cependant, au fil des mois de travail, ce *work in progress* n'a cessé de grossir et ses besoins financiers avec. Sans financements pérennes, nous avons dû à chaque étape du projet faire des choix pour mettre en équation et optimiser du mieux possible ambitions scientifiques, fonds disponibles et supports techniques que ces derniers permettaient. Rétrospectivement, le dépôt d'une demande spécifique auprès d'un organisme national de financement de la recherche nous aurait sans doute poussé à définir au préalable et avec beaucoup plus de précisions nos besoins scientifiques, techniques et financiers.

5 Dès lors, dans un premier temps, seront exposées les ambitions scientifiques et pédagogiques qui ont présidé à la définition des contours du projet RePPOL. Il s'agira ensuite de mettre en lumière les défis et les limites techniques et scientifiques auxquels a été confrontés l'équipe de chercheurs et d'ingénieurs, d'abord pour réaliser la transcription collaborative du texte du manuscrit 128 puis pour élaborer les balisages structurel et sémantique.

Pourquoi éditer le manuscrit 128 ?

Une enquête judiciaire

6 Le manuscrit 128 fait partie d'une collection de manuscrits constituée par l'archevêque de Cantorbéry Matthew Parker (1559-1575) dans la seconde partie du XVI^e^ siècle. Il s'agit d'un recueil factice regroupant des documents tous en lien avec Thomas Cranmer (1533-1556), prédécesseur de Parker sur le siège de Cantorbéry et personnage clef de la réformation sous Henri VIII (1509-1547) et Édouard VI (1547-1553), mort en martyr sous Marie I^ère^ (1553-1558) [6]. Pour rappel, dans les premières années de la décennie 1530, l'Angleterre rompt avec les papes et la monarchie ouvre une période de réformes religieuses marquées par l'évangélisme. Les raisons qui ont présidé à l'élaboration et à la conservation du recueil qu'est le ms. 128 nous échappe en bonne partie. Il semble qu'il faisait partie de la documentation amassée dans l'entourage de Matthew Parker afin d'écrire une apologie du premier archevêque de Cantorbéry de la nouvelle Église royale d'Angleterre [7].

7 Le projet ne portait que sur les 250 premiers folios des 400 que comprend le recueil dans son entier. Cette partie a été produite dans le cadre d'une enquête menée en 1543 dans le comté du Kent et que les historiens ont surnommée depuis le XIX^e^ siècle le « complot des prébendiers » (*Prebendaries' Plot*) [8]. Plusieurs clercs (dont des prédicateurs de la cathédrale de Cantorbéry) et des gentlemen du comté ont dénoncé au roi leur ordinaire, l'archevêque Cranmer, l'accusant d'avoir nommé et de protéger des clercs bénéficiers et des prédicateurs hétérodoxes. Ce faisant, le prélat enfreindrait les ordres royaux en matière de religion, particulièrement les Six articles de foi adoptés en 1539, qui imposent la reconnaissance de la présence réelle dans l'eucharistie [9]. Ces dénonciateurs espéraient sans doute qu'Henri VIII destitue l'archevêque. Cependant, la manœuvre échoue. Après quelques jours d'hésitation, le roi considère qu'il s'agit de dénonciations calomnieuses. Il renouvelle sa confiance à Cranmer et, non sans ambiguïté, il lui confie la charge de mener une enquête qui a deux objectifs fort différents. D'une part, ses investigations doivent

porter sur ses propres accusateurs et leurs motivations, et, d'autre part, il doit faire la lumière sur l'existence ou non parmi les fidèles de son diocèse de croyances et de pratiques potentiellement déviantes (c'est-à-dire trop conservatrices ou, au contraire, trop réformées). Cependant, l'issue judiciaire de cette enquête nous reste inconnue ou plutôt, il semble qu'aucun protagoniste du *Prebendaries Plot* n'ait été jugé et encore moins condamné, même si plusieurs personnes mises en cause (dont les premiers clercs dénonciateurs) furent incarcérées préventivement pendant quelques mois.

8 Le manuscrit 128 n'est que l'épave documentaire de cette enquête de 1543. S'il regroupe près d'une centaine de pièces de la procédure (lettres, interrogatoires, *bills* de dénonciation…), les lacunes sont importantes. Par exemple, les interrogatoires de certains protagonistes clefs ne sont pas présents dans notre document. Tout cela pose question quant à la nature de celui-ci. Il n'est pas un registre dans lequel auraient été copiés *a posteriori* les documents ; les différentes *marginalia*, corrections, changements de main… que l'on observe sur de nombreux folios tendent à montrer que l'on a affaire à des sources primaires produites au cœur de la pratique inquisitoriale.

9 Si le ms. 128 est une épave, c'est une belle épave. Plus de 200 individus apparaissent dans ces documents, éventail large et plutôt représentatif de la population du Kent du milieu du XVI^{e} siècle : on y trouve des clercs (du simple prêtre rural au prédicateur issu de l'université) mais aussi des paroissiens (gentlemen ou simples artisans), du monde des villes (principalement Cantorbéry) et de celui des campagnes. Au cours de cette enquête, des Kentois aux croyances et pratiques les plus réformées dénoncent des clercs ou des voisins qu'ils qualifient de tenants de l'évêque de Rome et de ses superstitions (maintien des cérémonies, du purgatoire et des saints, conception traditionnelle de l'eucharistie). En miroir, les témoins les plus conservateurs dénoncent des sacramentaires qui piétinent la religion telle que réformée par Henri VIII, dont les Six articles de 1539 ne sont que la dernière mouture.

10 Ce manuscrit permet, dès lors, une plongée rare dans les croyances et les pratiques liturgiques des sujets d'Henri VIII dans les sociétés locales en transition confessionnelle. On saisit comment la

coexistence religieuse ou, au contraire, les tensions confessionnelles sont vécues et traitées à l'échelle des communautés locales (paroisses, villes, *hundreds*[10]). Il permet également de mettre en évidence le rôle actif que les sujets, y compris les laïcs les plus infimes, ont joué dans les processus de mutations politico-religieuses de l'Angleterre des années 1530-1540.

Un projet aux ambitions scientifiques et pédagogiques

11 Depuis plus d'un siècle, ce manuscrit est connu des historiens par le biais de son *calendar*, une édition très partielle dans le volume 18/2 des *Letters and Papers. Foreign and Domestic. Henry VIII*, de 1902, édition qui a été republiée en ligne sur la plateforme British History Online[11]. Ce *calendar* propose une longue analyse diplomatique du document, en suivant de près sa structure. Toutefois, il résume ou omet certains passages, alors qu'il en édite quasi intégralement d'autres. Ainsi, la connaissance qu'ont eu les historiens de ce manuscrit est indirecte et, en partie, tronquée. Ceci explique sans doute qu'il n'ait été que très partiellement exploité, essentiellement mobilisé comme un réservoir d'exemples ponctuels de croyances hétérodoxes hautes en couleur. Aucun historien ne s'est intéressé aux modes d'élaboration, de composition et de conservation du manuscrit 128, et ne l'a réinscrit ou comparé avec des fonds documentaires contemporains et de nature proche, tels que les papiers Cromwell conservés aux National Archives[12].

12 Le ms. 128 est donc un document paradoxal : considéré comme « unique » parce que rarement contextualisé, cité mais méconnu et dont l'histoire matérielle reste à faire. Aude de Mézerac-Zanetti et moi-même avions tous deux croisé ce document dans nos recherches ; elle, dans le cadre d'études sur les transformations liturgiques durant la réformation henricienne[13], moi, au cours d'un chantier consacré aux dénonciations et aux enquêtes en lien avec la politique d'Henri VIII[14]. Conscients du caractère exceptionnel de ce document et de sa méconnaissance dans la communauté scientifique, nous avons lancé le projet RePPOL pour rendre disponible une édition en open-source utile pour toute étude portant sur l'histoire politique, religieuse ou sociale de l'Angleterre du XVI^{e} siècle. À une

échelle plus large, alors que se multiplient les travaux sur la place du peuple dans les réformes religieuses de la première modernité, ce document éclaire sous un jour nouveau les réformations européennes et leurs conséquences politiques [15].

13 L'une des caractéristiques du projet RePPOL, dès sa définition, a consisté à associer recherche et enseignement. En effet, le volet scientifique du projet a émergé de pratiques pédagogiques pensées par Aude de Mézerac-Zanetti dans le cadre d'un master de civilisation britannique à l'Université de Lille. Dispensant des cours d'initiation aux sources et à la paléographie de l'Angleterre de la première modernité, elle désirait montrer aux étudiants à quoi servaient ces apprentissages techniques. Pour les introduire dans l'atelier de l'historien, elle les fit travailler sur certains passages inédits du ms. 128. Les étudiants s'étant montrés très réceptifs, nous nous sommes demandés si nous ne pouvions pas faire se rencontrer notre envie scientifique de rendre disponible à la communauté des historiens le ms. 128 et celle de former par la recherche des mastérants, en associant ces derniers à la production d'une transcription et d'une édition scientifique de ce document.

La transcription : grandeur et décadence d'un projet collaboratif

Le choix de la plateforme TACT

14 Le projet RePPOL a été grandement facilité par le fait que la bibliothèque universitaire de Cambridge, dépositaire de la Parker Library, a fait numériser et mis en ligne gratuitement chaque folio du ms. 128 en très haute définition selon les standards IIIF (International Image Interoperability Framework) [16]. Ces derniers permettent d'utiliser et de conserver chaque image en qualité optimale, ainsi que les métadonnées qui lui sont liées.

15 La taille modeste de notre projet et son absence de financement récurent (ANR ou autre) rendaient impossible le recours à un opérateur privé intégré de type *kleiolab* (Geovistory) pour prendre en charge la transcription et l'édition du document [17]. Nous n'avons pas non plus envisagé de recourir à un logiciel de transcription assistée

par Intelligence Artificielle, comme *Transkibus* par exemple, puisque l'un des objectifs de RePPOL était de former par la pratique des étudiants à la paléographie anglaise et à l'édition d'un manuscrit. Des considérations d'ordre technique sont également entrées en ligne de compte. La partie du manuscrit qui nous intéresse se limite à environ 250 folios, sur lesquels on a identifié au moins six mains différentes ; de plus, les zones de texte peuvent varier grandement d'un feuillet l'autre. Avec de telles contraintes, l'usage d'IA nous semblait peu rentable et donc peu pertinent. En effet, un modèle algorithmique a besoin d'être entrainé, corrigé et donc éduqué avant d'être pleinement performant [18].

16 Forts de ces constats, nous avons choisi d'utiliser une plateforme en ligne de transcription scientifique et collaborative, TACT (Transcription et annotation de corpus textuels), lancée en 2019 par l'université Grenoble-Alpes. Le responsable d'un projet crée un espace sur la plateforme et y téléverse les images des documents qui doivent être transcrits (ce que nous avons fait pour chaque folio du manuscrit) [19]. Le choix de cette plateforme gratuite s'explique aussi par nos contraintes financières. TACT propose une interface de travail en ligne assez simple. L'intérêt principal de cette plateforme était de permettre à une vingtaine d'étudiants de master lyonnais et lillois de transcrire en parallèle, de façon asynchrone, des folios du texte selon des normes paléographiques définies par les deux porteurs du projet. Le volet « commentaire » présent en face de chaque folio nous permettait également de donner des conseils spécifiques à chaque transcripteur. TACT offrait donc une interface simple pour guider, relire et éventuellement corriger les transcriptions des étudiants [20].

Le volet pédagogique : demi-échec ou semi-réussite ?

17 La plateforme TACT s'est révélée très pertinente pour réaliser la transcription : possibilité de zoomer sur les documents, de créer autant de « bacs à sable » [21] que nécessaire pour former les transcripteurs, de réécrire, retranscrire et relire à l'infini… Cet outil numérique permettait donc de suivre efficacement la progression des étudiants et leur montée en compétences. Encore aujourd'hui, bien

que l'édition RePPOL soit achevée, la plateforme du projet sur TACT continue à fonctionner : elle peut servir d'espace d'initiation à la paléographie des écritures anglaises et à leur transcription, pour qui le souhaite.

18 Cependant, nous avons pu constater assez rapidement que le rythme de transcription des étudiants était assez lent. Ces derniers étaient moins gênés par les dimensions techniques, comme la prise en main de l'outil, parce qu'assez intuitive, que par l'anglais du XVI^{e} siècle et son orthographe non fixée, sa syntaxe erratique et le contenu assez décousu du document. Les transcriptions d'étudiants de Master dans le cadre de cours de paléographie se sont ainsi limitées à une petite dizaine de pages sur les 250, et ont nécessité un grand nombre de relectures. Ainsi, marcher au pas pédagogique ralentissait considérablement le planning élaboré pour le volet recherche de RePPOL. *In fine*, Aude de Mézerac-Zanetti et moi avons produit l'essentiel de la transcription. Dès lors, le choix initial de passer par la plateforme collaborative TACT était moins pertinent.

19 Cette contrainte nous a conduit à faire sensiblement évoluer la dimension pédagogique du projet. RePPOL ne devait plus être un support de travail de transcription paléographique pour une vingtaine d'étudiants. Nous avons intégré pleinement au projet Felipe Goes, un étudiant de Master 1 d'études anglophones de Lille. Dans le cadre de son mémoire dirigé par Aude de Mézerac-Zanetti, il a transcrit une cinquantaine de folios du manuscrit. Il a surtout établi un instrument de recherche en faisant le relevé de la structure du ms. 128. Il a identifié tous les types de documents (questionnaires, lettres, procès-verbaux d'interrogatoire etc.) et il a isolé à l'intérieur chaque paragraphe, afin de faciliter l'étape suivante du projet, exposée en troisième partie de cet article : le balisage structurel.

Un exemple de contrainte : faire coexister deux transcriptions

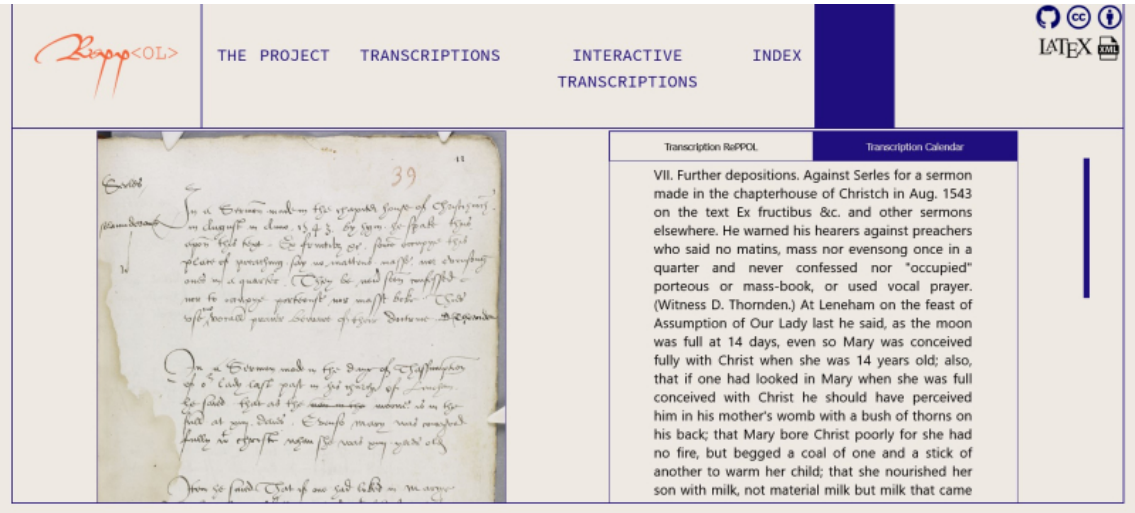
20 Certains choix scientifiques opérés au début du projet se sont révélés être des contraintes très lourdes et peut-être aurions-nous dû les abandonner. Par exemple, nous avions décidé de mettre l'édition partielle publiée en 1902 en regard de celle que nous entendions réaliser. Originellement, il s'agissait, d'un point de vue scientifique, de

permettre une réflexion sur ce qu'est un travail d'édition historique. Quels passages sont résumés ou intégralement donnés par le *Calendar* ? Comment ont été rendus en anglais contemporain certaines expressions Tudor ? Cette plongée dans l'atelier des éditeurs-historiens de la fin du XIX^{e} siècle, s'est révélée finalement secondaire par rapport à l'objectif central du projet : l'édition et le traitement du manuscrit.

21 Or, le choix initial d'utiliser la plateforme TACT s'expliquait également par le fait qu'elle nous permettait de donner corps à cette double transcription. En effet, implémenter dans les commentaires de notre propre transcription celle de 1902, transformait ce *calendar* en une aide à la compréhension du texte pour les étudiants transcripteurs et pouvait parfois jouer le rôle d'une antisèche. Mais cela a nécessité un long et important travail de mise en œuvre technique, puisque le mastérant Felipe Goes a dû appairer chaque folio avec son résumé.

22 Cette édition double a également conduit à modeler l'architecture du site de publication de RePPOL (https://reppol.meshs.fr/). A ainsi été créée une double fenêtre (figure 1), peu pratique pour l'expérience utilisateur bien qu'elle ait nécessité un important travail technique supplémentaire lors de la construction du site.

Figure 1 : Deux transcriptions parallèles sur TACT : le fol. 39, RePPOL et le calendar



23 Si cette ambition scientifique a fortement contraint les choix techniques opérés tout au long du processus, une étude sur le

Calendar est désormais possible et peut être un prolongement futur pour RePPOL.

Les défis de l'édition numérique

24 Pour mener le second chantier, l'édition numérique, nous avons fait le choix d'associer un second étudiant au projet. Le LABEX COMOD a financé un stage de Master 1 de quatre mois, réalisé en 2021-2022 par Adrien Mével, mastérant en humanités numériques à l'Université de Lille [22]. Ce dernier a travaillé, sous la direction de Victoria Le Fourner (maîtresse de stage), à concevoir et mettre en place l'architecture numérique nécessaire à l'édition du manuscrit. Devant l'ampleur de la tâche, les crédits IUF d'Aude de Mézerac-Zanetti ont permis de financer des vacations durant l'année 2022-2023 pour qu'Adrien Mével puisse achever son travail [23].

25 Le projet RePPOL nécessitait d'effectuer un double balisage : un balisage de la structure du manuscrit et du texte, et un balisage sémantique (patronymes, toponymes, dates).

Quelle édition pour quels moyens techniques ?

26 Dès le début du projet, il nous paraissait nécessaire que, à côté de l'édition numérique en HyperText Markup Language (HTML) permettant de présenter en vis-à-vis l'image de chaque folio et sa transcription, il soit possible de consulter et télécharger l'intégralité du texte édité en un seul document interrogeable par mots-clefs. Nous envisagions alors un texte téléchargeable sous un format pdf, respectant au plus près la structure du manuscrit. Or la réalisation technique de cette édition secondaire conjointement, voire à l'intérieur, de l'édition primaire a donné lieu à un débat animé entre les différentes parties impliquées dans le projet pour savoir comment concilier faisabilité technique et accessibilité et lisibilité du texte pour les chercheurs.

27 Compte tenu du temps et des financements limités, les ingénieurs ont proposé trois solutions techniques d'édition. La première était de transformer directement le texte du format HTML de l'édition primaire au format pdf. Cette solution permettait d'imprimer le texte

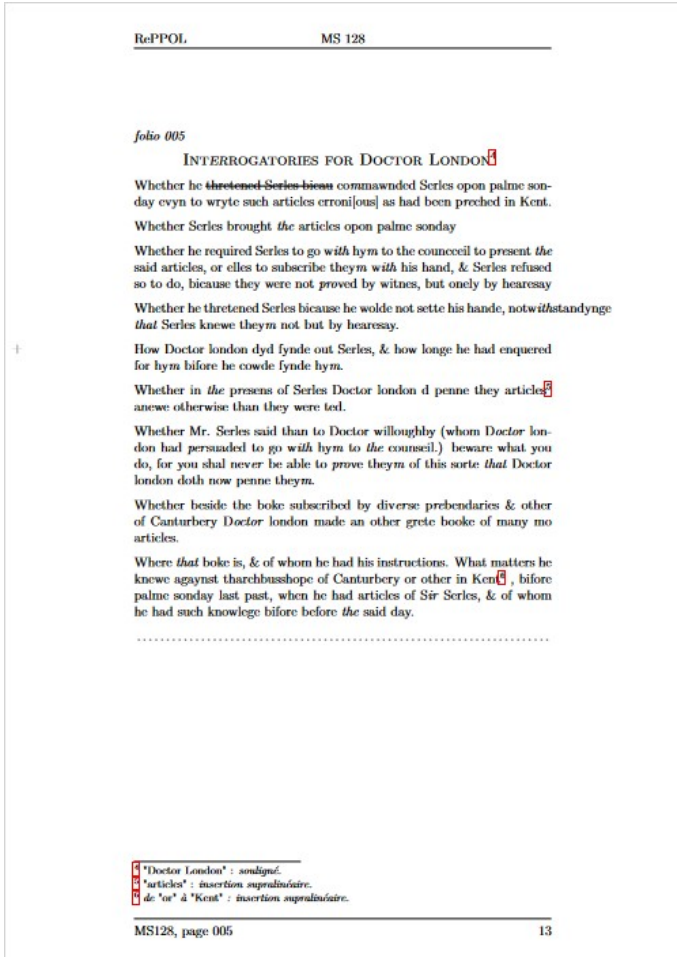
et d'effectuer des recherches lexicales. De plus, elle ne demandait aucun surcroît de travail de type scientifique laissant la réalisation aux mains des seuls ingénieurs. Cependant, le rendu du texte ainsi édité aurait été, tant en ligne que sous forme imprimée, peu lisible et donc peu pratique.

28 La seconde option consistait à produire une édition du texte en LateX [24] en adoptant une transcription ultra-diplomatique (en faisant un choix entre les traditions française et anglaise d'édition) [25]. La version obtenue aurait alors été publiable et lisible, mais cela requérait un temps de travail important de la part des ingénieurs. En effet, en sus de l'édition-miroir conçue en HTML, il leur fallait produire une seconde édition du document, cette fois en Extensible Markup Language (XML).

29 La dernière solution, qui avait les faveurs des ingénieurs, était de produire une seule édition, en HTML. L'un des intérêts était de faire de RePPOL un délivrable unique, une édition en ligne, facilitant ainsi les citations lorsque d'autres chercheurs l'utiliseraient. D'un point de vue technique, cette option offrait l'avantage de permettre aux ingénieurs de concentrer leurs efforts sur une seule réalisation. Cependant, à nos yeux, cela limitait les possibilités de consultation et de travail à partir de l'édition puisque l'utilisateur aurait dû être connecté en permanence au site.

30 In fine, Aude de Mézerac-Zanetti et moi-même avons choisi la seconde option, qui nous paraissait plus pertinente scientifiquement, et surtout plus proche d'une expérience utilisateur classique de chercheur (figure 2). Ce choix nous a conduit à devoir mettre de côté, faute de temps et de moyens humains, certains aspects du projet que nous ambitionnions de réaliser (index croisés, transcriptions interactives du texte [26]…).

Figure 2 : Une page tirée de l'édition ultra-diplomatique téléchargeable, élaborée en LateX



RePPOL MS 128

folio 005

INTERROGATORIES FOR DOCTOR LONDON[1]

Whether he ~~thretened Serles bicau~~ commawnded Serles opon palme sonday cvyn to wryte such articles erroni[ous] as had been preched in Kent.

Whether Serles brought *the* articles opon palme sonday

Whether he required Serles to go *with* hym to the counceil to present *the* said articles, or elles to subscribe the*ym* *with* his hand, & Serles refused so to do, bicause they were not *proved* by witnes, but onely by hearesay

Whether he thretened Serles bicause he wolde not sette his hande, not*with*standynge *that* Serles knewe the*ym* not but by hearesay.

How Doctor london dyd fynde out Serles, & how longe he had enquered for hym bifore he cowde fynde hym.

Whether in *the* presens of Serles Doctor london d penne they articles[2] anewe otherwise than they were ted.

Whether Mr. Serles said than to Doctor willoughby (whom *Doctor* london had persuaded to go *with* hym to *the* counseil.) beware what you do, for you shal never be able to *prove* the*ym* of this sorte *that* Doctor london doth now penne the*ym*.

Whether beside the boke subscribed by diverse prebendaries & other of Canturbery *Doctor* london made an other grete booke of many mo articles.

Where *that* boke is, & of whom he had his instructions. What matters he knewe agaynst tharchbusshope of Canturbery or other in Kent[3] , bifore palme sonday last past, when he had articles of Sir Serles, & of whom he had such knowlege bifore before *the* said day.

. . . . . .

[1] "Doctor London" : *souligné.*
[2] "articles" : *insertion supralinéaire.*
[3] *de* "or" *à* "Kent" : *insertion supralinéaire.*

MS128, page 005 13

31 Cependant, ce choix remettait en question l'utilisation de TACT pour la seconde partie du projet. Cette plateforme n'autorise que des extractions au format JavaScript Object Notation (JSON) qui permet certes de produire une visualisation en HTML mais qui ne peut être converti en XML. Il était, dès lors, impossible d'exporter à partir des transcriptions faites des différents folios dans TACT un fichier XML final propre pour en faire ensuite une visualisation en HTML (et donc de l'édition miroir).

La complexité du balisage structurel

32 Produire une édition mettant en miroir chaque folio du manuscrit avec sa transcription nous paraissait avoir deux intérêts distincts. Tout d'abord, le site de RePPOL peut être utilisé comme un outil

pédagogique dans le cadre d'une formation à la paléographie anglaise. Ensuite, d'un point de vue scientifique, l'édition rend ainsi compte de la nature composite et de la matérialité particulière du ms. 128. Il ne s'agit pas d'un grand texte unifié : il regroupe une grande diversité de documents et plusieurs, d'entre eux, sont composés d'éléments très particuliers. Cette édition rend possible l'étude matérielle du manuscrit (analyse des mains des différents scripteurs, des ajouts et des suppressions au cours du processus d'interrogatoire ou a posteriori) mais aussi une analyse génétique (marques et annotations marginales, qui peuvent, par exemple, avoir servi à composer des actes d'accusation ou *indictments*, réponse à une question numérotée, ajout postérieur à une réponse). Dès lors, le balisage structurel du manuscrit a consisté à encoder chaque saut de ligne, rature, ajout, changement de types de documents ou encore *marginalia*.

33 L'édition miroir respectant la structure du manuscrit nécessitait de passer notre transcription à un format XML. Pour faire ce balisage structurel, il fallait établir un ODD (One Document Does it all) de XML TEI (Text Encoding Initiative) pour créer un schéma TEI spécifique au ms. 128. Un ODD est un document TEI contenant la documentation d'un modèle et, en son sein, un élément schemaSpec qui sert de racine aux spécifications techniques. Il permet de construire un schéma formel XML pérenne et surtout standard. Cet ODD a été élaboré par Victoria Le Fourner et par Adrien Mével grâce au logiciel Oxygen XML editor.

34 Si l'élaboration d'un ODD est une tâche relativement classique pour un ingénieur ou un chercheur rompu aux humanités numériques, celle-ci doit toujours se faire en étroite collaboration avec les porteurs scientifiques d'un projet. Pour pouvoir dialoguer avec les ingénieurs parties prenantes et se faire comprendre d'eux, il nous a fallu appréhender ce qu'étaient des patrons (ou macros), des types de données (*data types*), des classes et sous-classes, des schémas… Au cours de ce processus, nous avons compris que la collaboration avec des ingénieurs impliquait, de notre part, de clarifier et systématiser nos objectifs et nos pratiques de recherche sur ce manuscrit. En effet, si, d'un point de vue purement technique, le modèle XML est élaboré par des ingénieurs, il est toujours le reflet d'une analyse précise des documents (structure, contenu). Ce sont donc les

porteurs scientifiques qui produisent ou sélectionnent les informations à modéliser conformément aux besoins du projet et doivent être en capacité de pré-visualiser le rendu final.

35 Le modèle finalement choisi permet d'avoir une approche topologique ou géographique du manuscrit 128, complètement absente de l'édition de 1902. Il s'agissait de rendre visibles les différentes strates d'élaboration du manuscrit, c'est-à-dire de différencier la première écriture des annotations marginales, témoignant de lectures postérieures et donc des usages possibles de ce manuscrit par les Anglais des années 1540 [27]. Pour ce faire, nous avons utilisé les possibilités de la TEI, en mobilisant à l'échelle de plusieurs folios successifs, mais aussi à l'échelle de chaque feuillet les balises <surface>, <zone> et <line> pour rendre la diversité des documents composant le ms. 128 et le caractère composite de certains de ces documents. Ainsi a été construit un complexe système zonal d'encadrés, emboîtés l'un dans l'autre et de couleurs différentes selon la nature des différents documents, le vert renvoyant par exemple à une section du manuscrit portant des accusations (figure 3). Ce système permettait, entre autres, de séparer à l'intérieur d'un même document synthétisant plusieurs accusations, celles qui visaient tel individu et celles qui concernaient tel autre. Pour les interrogatoires, était différenciée chaque question, ce qui permettait le cas échéant de faire des renvois vers les réponses apportées par les divers témoins et suspects en d'autres endroits du manuscrit. Mais ce zonage interprétatif devait coexister avec un zonage strictement matériel, permettant de conserver les différents paragraphes ainsi que les annotations ou les numérotations ajoutées en 1543 (figure 4). La complexité du manuscrit et les exigences scientifiques que nous avions expliquent que la pose des balises structurelles ne pouvait se faire de façon automatisée, ce qui demanda un important travail de la part des ingénieurs concepteurs de l'encodage.

Figure 3 : Schéma d'encodage de la structure du manuscrit (réalisé par Adrien Mével, supervision Victoria Le Fourner)

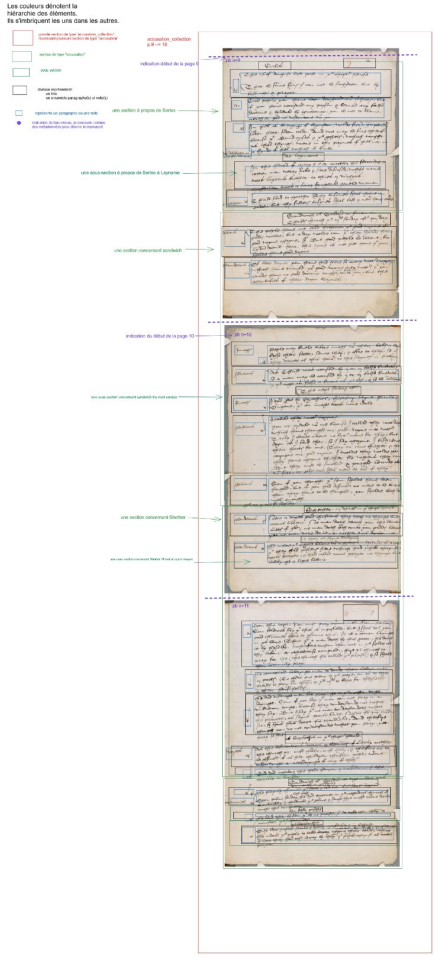
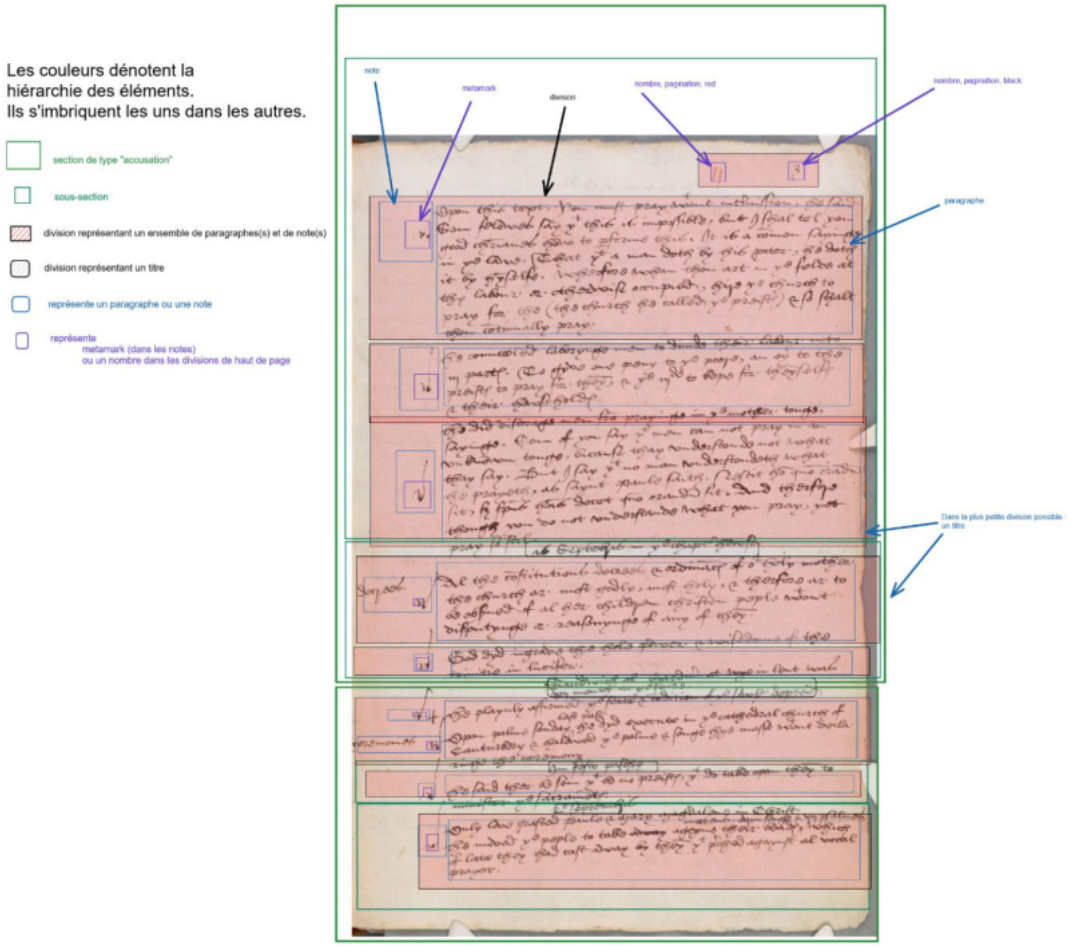


Figure 4 : Le balisage structurel : un système de boîtes (réalisation Adrien Mével sous la supervision de Victoria Le Fourner)



Des balisages sémantiques non automatisés

36 Il nous paraissait indispensable de produire des index (lieux, personnes, dates) afin de permettre aux chercheurs une utilisation de ce manuscrit qui ne soit pas seulement linéaire. Or construire des index d'entités nommées (patronymes et toponymes) nécessitait un travail de balisage sémantique qui paraissait, au premier abord, relativement simple, tant à nous-mêmes qu'aux ingénieurs. Une fois le texte transcrit, les ingénieurs comptaient poser des balises par reconnaissance automatique des termes. Pour évaluer la pertinence de la méthode, nous avons fait des tests sur quelques folios. Cette procédure s'est alors révélée mal adaptée, car elle nécessitait

d'importantes relectures, corrections et ajouts. Ce balisage se révéla très chronophage, en raison de l'orthographe très mouvante des noms (d'autant qu'il y a plusieurs scripteurs) et la difficulté à saisir de manière automatisée les périphrases ou les épithètes qualifiant telle ou telle personne ou lieu. Par exemple, le patronyme de Cranmer est très rarement utilisé dans le manuscrit, mais on le trouve fréquemment désigné par les termes *tharchbishop* (« l'archevêque »), *His grace* (« Sa Grâce ») ou encore *his lordship* (« Sa Seigneurie »), voire simplement par *him* (« lui »)… Il a donc fallu baliser, identifier puis nettoyer les index générés automatiquement. Cette méthode mettait à mal l'élaboration même de l'ODD dont l'une des contraintes fortes est la nécessaire absence d'ambiguïté. Il faut toujours être totalement explicite tant pour le programmateur que pour le programme, ce qui bannit tout flou ou approximation. Or ceux-ci peuvent être, pour des chercheurs, des hésitations salutaires, des ambiguïtés qui dans l'avenir pourront se révéler riches de sens, ou encore des points de suspension qui demandent à être éclairés ultérieurement.

37 Pour faciliter notre travail d'identification des individus et des lieux, et construire ces index, nous avons utilisé des bases de données en ligne spécialisées, particulièrement celle du *Clergy of the Church of England Database* (CCED)[28] et celle de l'*Oxford Dictionary of National Biography* (ODNB)[29]. En utilisant les numéros d'identification attribués par ces bases de données aux différents individus qu'elles répertorient, les entrées d'index de RePPOL peuvent renvoyer directement l'utilisateur vers ces sites (figure 5). Pour les toponymes identifiés (villes, paroisses ou autres), nous renvoyons, le cas échant, au CCED, mais nous utilisons également les données GPS pour localiser ces lieux sur OpenStreetMap (figure 6)[30]. L'encodage ainsi produit doit permettre, à l'avenir, de produire à partir de l'édition RePPOL des visualisations géographiques ou des analyses de réseaux.

Figure 5 : Exemple de balisage sémantique pour l'indexation des noms propres, ici William Orphen

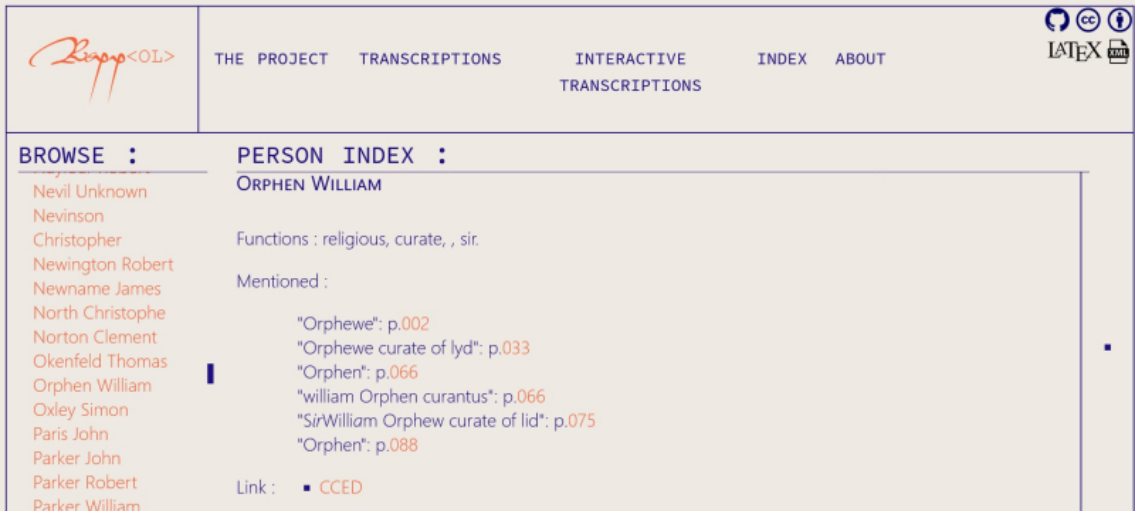
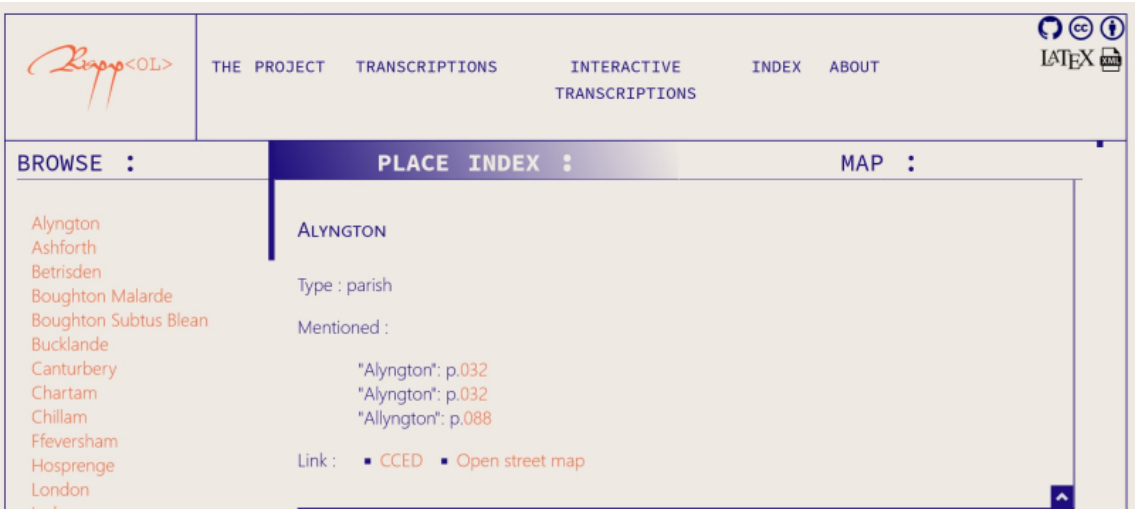


Figure 6 : Exemple de balisage sémantique pour l'indexation des toponymes, ici Alyngton



38 L'index qui nous apparaissait comme le plus important pour nos recherches était un index thématique. Celui-ci aurait permis de localiser dans le manuscrit les différentes occurrences de telle ou telle pratique ou croyance religieuse (purgatoire, images, Vierge…), ou encore la mobilisation de telle ou telle notion politique (sédition…). Or, ce balisage se révéla très ardu à opérer. Une nouvelle fois, les ingénieurs ont proposé un balisage automatique par la sélection de quelques mots qui nous paraissaient pertinents

(« image », « statue », « sédition »…). En phase de test, l'échec fut encore plus marqué que pour le balisage sémantique. En effet, ce n'était pas tant l'occurrence d'un mot qui était significative ou intéressante que l'usage qui en était fait et sa position dans son contexte d'utilisation. Techniquement, le seul balisage thématique viable ne pouvait être correctement fait que par Aude de Mézerac-Zanetti et moi-même car il relevait plus d'une interprétation du texte que d'un simple acte technique neutre. Or cela ne pouvait être fait qu'au prix d'un très long travail de discussion, d'harmonisation et de systématisation. Ce projet d'index et avec lui l'encodage thématique ont donc été abandonnés.

39 Dans les premiers temps du projet, nous avions envisagé de confier une partie de ce balisage (dates, noms, parties saillantes de la structure) aux étudiants transcripteurs. La plateforme TACT offre, en effet, une interface relativement simple de balisage, avec un système de boutons configurables. Cependant, les étudiants avaient la plus grande difficulté à faire en même temps le travail de transcription et celui de balisage. Ils ne pouvaient acquérir simultanément les compétences nécessaires à l'une et à l'autre. Oublis et erreurs étaient nombreux. Nous avons donc demandé aux étudiants de se concentrer sur la seule transcription du texte. Une autre raison justifiait cet abandon. Le système de boutons de TACT est adapté à un balisage sémantique ou formel très basique, mais il ne l'est pas pour un document complexe comme l'est le ms. 128. Nous avons donc adopté un système hybride : l'essentiel du balisage général de la « structure » du manuscrit (sauts de lignes, notes marginales, ajouts au texte…) s'est fait dans TACT, mais sans utiliser les boutons préprogrammés par la plateforme. Aude de Mézerac-Zanetti et moi-même avons, lors des transcriptions et des relectures, posé les balises en agissant directement sur le code source de chaque page (figure 7). Cette procédure de codage était beaucoup plus efficace, mais aussi beaucoup plus fastidieuse. Le reste du balisage (essentiellement le balisage sémantique nécessaire aux index) s'est faite en dehors de TACT à partir de l'XML.

Figure 7 : Un exemple de page de code-source sur TACT, permettant d'encoder directement les balises



40 Cet article entendait exposer les principales questions qui sont apparues durant l'élaboration et la conduite de ce projet hybride associant enseignement et recherche qu'est RePPOL. Nous sommes revenus sur les problèmes de conception auxquels nous avons été confrontés ainsi que sur les difficultés à articuler versants technologique et académique. Les outils numériques les plus adaptés aux objectifs pédagogiques du projet n'étaient, par exemple, pas nécessairement les plus pertinents pour la production de savoir scientifique.

41 Codiriger RePPOL nous a certes permis d'acquérir les bases théoriques et pratiques indispensables à la participation active au sein d'un projet d'humanités numériques, tant pour le volet scientifique que pour le volet technique, mais cela nous a également conduit à nourrir une réflexion méthodologique quant à la construction et la publication des données dans le cadre de telles recherches. Les quatre années du projet nous ont permis de réfléchir, à nouveaux frais, à la façon de travailler collectivement à la construction du savoir.

42 Tenir ensemble les versants scientifiques et techniques du projet nécessitait de construire un dialogue serré entre les différents membres de l'équipe. Les ingénieurs nous posaient des questions, nous proposaient des solutions mais ils avaient parfois du mal à

percevoir en quoi leurs instruments contraignaient notre recherche et la façon de produire l'édition scientifique finale. En retour, il était difficile de dialoguer avec les ingénieurs qui nous entouraient, non pas tant pour faire comprendre ce que nous voulions, mais pour exprimer nos désirs conformément à ce qu'il était techniquement possible de faire tant en termes de temps passé (est-ce que telle demande va prendre quelques minutes ou plusieurs heures à être réalisée ?) que technologiquement (qu'est-ce qui est faisable et avec quels instruments ?).

43 Le problème est désormais celui de la pérennité du site hébergeant RePPOL, ou, en tout cas, de son enrichissement. Adrien Mével a achevé son stage au début de l'année 2023 et est désormais employé par la plateforme Cairn. Quant à la MESHS, elle a changé de politique au cours de l'année 2022 : redevenant un simple financeur de projets, l'institution s'est séparée de ses ingénieurs, dont les deux personnes qui participaient depuis le départ au projet RePPOL. Dès lors, nos compétences techniques étant trop limitées, nous nous rapprochons d'Huma-Num pour faire migrer la plateforme de la MESHS.

NOTES

1 Maîtresse de conférences en civilisation britannique à l'Université de Lille, laboratoire CECILLE (ULR 4074).

2 En français, l'acronyme est rendu par Faciles à trouver / Accessibles / Interopérables / Réutilisables.

3 Victoria Le Fourner était alors ingénieure « en charge du traitement des données scientifiques » au CNRS et Florence Perret ingénieure « en charge du traitement des données scientifiques » à la MESHS.

4 On peut citer par exemple le projet *Lettere d'Artista. Roma e l'Europa 1750-1780 / Lettres d'Artiste. Rome et l'Europe 1750-1850*, qui embrasse plus de 6 000 lettres différentes, ou le projet *e-NDP : Notre Dame et son cloître*, une édition numérique collaborative de 26 registres du chapitre de Notre-Dame des XIV^{e} et XV^{e} siècles, soit plus de 14 000 pages.

5 Le financement initial a été obtenu par Aude de Mézerac-Zanetti dans le cadre d'appel à projet Flash de l'i-site Université Lille-Nord-Europe.

6 Diarmaid McCulloch, *Thomas Cranmer* : A *Life*, New Haven, Yale University Press, 1996.

7 Sur la constitution de la bibliothèque de l'archevêque Parker : Nancy B. Bjorklund, « Parker's Purpose for his manuscripts. Matthew Parker in the context of his early career and sixteenth-century church reform », dans Joyce Tally Lionarons (dir.), *Old English Literature in its Manuscripts Context*, Morgantown, West Virginia University Press, 2004, p. 217-241; Timothy Graham, « Matthew Parker's manuscripts: an Elizabethan library and its use », dans Elisabeth Leedham-Green (dir.), *The Cambridge History of Libraries in Britain and Ireland. Vol. 1, To 1640*, Cambridge, Cambridge University Press, 2006, p. 322-344.

8 Michael L. Zell, « The Prebendaries' Plot of 1543: a Reconsideration », *Journal of Ecclesiastical History*, vol. 27/1, 1976, p. 241-253; Eamon Duffy, *The Stripping of the Altars: Traditional Religion in England, 1400-1580*, New Haven, Yale University Press, 1992, p. 433-447; Diarmaid MacCulloch, *Thomas Cranmer*, *op. cit.*, p. 297 et seq. et Ethan H. Shagan, *Popular Politics and the English Reformation*, Cambridge, Cambridge University Press, 2003, p. 197-232.

9 *An Acte abolishing diversity in opynions*, 31 Henry VIII, cap. 14, juin 1539, édité dans John Raithby (éd.), *The Statutes of the Realm*, Londres, HMSO, 1817, vol. 3, p. 739.

10 Les *hundreds* sont les principales subdivisions des comtés.

11 James Gairdner et R.H. Brodie (éd.), *Letters and Papers. Foreign and Domestic. Henry VIII, vol.18, part 2*, Londres, HMSO, 1902, item n° 546, titré "Cranmer and the Heretics of Kent".

12 Il s'agit d'une partie des archives et de la correspondance de Thomas Cromwell, principal ministre d'Henri VIII dans les années 1530, et qui sont aujourd'hui conservées dans les State Papers (SP 1) aux National Archives. Au sein de ce fonds, on trouve de nombreux documents d'enquêtes portant pour certaines sur des croyances et pratiques religieuses.

13 Aude De Mézerac-Zanetti, « La fabrique du croire : les évolutions liturgiques sous le règne d'Henri VIII », dans Monique Cottret et Caroline Galland (dir.), *Croire ou ne pas croire*, Paris, Kimé, 2013, p. 155-170 ; *Id.*, « A Reappraisal of Liturgical Continuity in the Mid-Sixteenth Century: Henrician Innovations and the First Books of Common Prayer », *Revue française de civilisation britannique*, vol. 22/1, 2017 (en ligne).

14 Olivier Spina, « Le dégel de la Parole. Quête d'uniformité et dissensus religieux sous le règne d'Henri VIII », dans Florence Alazard et al. (dir.), *Dissensus. Pratiques et représentations de la diversité des opinions*, Paris, Champion, 2016, p. 63-81 ; *Id.*, « *Till that I know your further plesure what I shuld do in this matter*. Arbitraire et arbitrage royaux autour de la trahison par les mots (Angleterre 1534-1540) », dans Benjamin Deruelle et Michel Hébert (dir.), *Arbitraire et arbitrages : les zones grises du pouvoir (XII^e^ – XVIII^e^ siècles)*, Lille, Presses Universitaires du Septentrion, à paraître en 2024.

15 Ethan Shagan, *Popular Politics and the English Reformation*, Cambridge, Cambridge University Press, 2003. On ne peut que citer l'œuvre ancienne mais fondatrice Robert W. Scribner, *For the Sake of Simple Folk. Popular Propaganda for the German Reformation*, Cambridge, Cambridge University Press, 1981.

16 Il s'agit d'un cadre technique standardisé qui entend permettre l'interopérabilité des images sur le web, leur utilisation et leur diffusion.

17 Des contacts avaient été pris, mais très rapidement, le budget nécessaire fut rédhibitoire.

18 Il n'existe pas encore à notre connaissance de modèle algorithmique adapté aux écritures pragmatiques de l'Angleterre du XVI^e^ siècle.

19 L'espace de RePPOL est toujours consultable à l'adresse https://tact.demarre-shs.fr/project/31

20 Lors de la création du projet, le responsable peut fixer un nombre de relecture minimale avant acceptation de la publication. De même, le nombre de contributeurs inscrits n'est pas limité.

21 Un bac à sable ou *sandbox* est, sur une plateforme, un environnement qui a été isolé afin de permettre de simuler les conditions réelles de fonctionnement et d'utilisation de celle-ci. Dans le cadre de la formation des étudiants, ces espaces de test et d'observation ont servi à les entraîner à la transcription.

22 RePPOL est intégré à l'axe « L'État et les religions » du LABEX COMOD.

23 Aude de Mézerac-Zanetti est membre junior de l'IUF pour la période 2022-2027.

24 Ce système de commandes permet de composer et mettre en page un document en suivant avec la plus grande rigueur des normes typographiques prédéfinies.

25 La transcription ultra-diplomatique s'efforce de rendre compte de manière aussi mimétique que possible des formes et de la disposition spatiale d'un texte.

26 Le site de RePPOL propose la transcription interactive de quelques folios. Il s'agit des pages-tests que nous avions effectués, https://reppol.meshs.fr/reppol_trancriptioninteractive.html

27 Pour un exemple alors novateur de l'utilisation de la TEI : Julie André et Elena Pierazzo, « Le codage en TEI des brouillons de Proust : vers l'édition numérique », *Genesis. Manuscrits. Recherche. Invention*, vol. 36, 2013, p. 155-161.

28 http://www.theclergydatabase.org.uk/

29 https://www.oxforddnb.com/

30 OpenStreetMap est un projet de cartographie en ligne collaboratif s'appuyant sur les coordonnées GPS.

ABSTRACTS

Français
Cet article présente un retour d'expériences sur le projet d'histoire numérique *Rethinking the Prebendaries Plot On Line* (RePPOL) mené entre 2019 et 2023 en collaboration avec Aude de Mézerac-Zanetti (Université de Lille) et des ingénieures de la MESHS de Lille.
RePPOL est un projet au croisement de la recherche et de l'enseignement. Son premier objectif était de rendre disponible aux chercheurs travaillant sur le règne d'Henri VIII d'Angleterre, un document unique : le manuscrit 128 conservé dans la Parker Library (Corpus Christi College, Cambridge). Il s'agit d'un dossier d'enquête menée en 1543 par l'archevêque de Cantorbéry sur les pratiques et les croyances religieuses hétérodoxes d'un certain nombre d'habitants du Kent. Mais RePPOL se voulait un projet collaboratif associant aux travaux de transcription et d'édition numérique plusieurs étudiants de master, dans une démarche de formation par la recherche (paléographie anglaise, encodage…).
Cet article met en lumière les principaux défis scientifiques et techniques rencontrés au cours de RePPOL : choix des outils technologiques les plus pertinents pour une transcription collaborative, dialogue et arbitrages entre ingénieurs et chercheurs quant à la forme du délivrable, prise en compte des contraintes d'argent et de temps pour la définition même du contenu du projet.
L'édition numérique finale est aujourd'hui consultable (https://reppol.meshs.fr/).

English
This article is a feedback on the digital history project “Rethinking the Prebendaries Plot On Line” (RePPOL), led between 2019 and 2023 with Aude de Mézerac-Zanetti (University of Lille), and with the support of engineers from MESHS Lille.
RePPOL is a project at the crossroads of research and teaching. Its first objective was to make a unique document available to academics working on the reign of Henry VIII of England: manuscript 128 preserved in the Parker Library (Corpus Christi College, Cambridge). This is a record of an investigation carried out in 1543 by the Archbishop of Canterbury into the heterodox religious practices and beliefs of two hundred inhabitants of Kent. Moreover, RePPOL was intended to be a collaborative project involving postgraduate students in the transcription and digital editing work, as part of an education-through-research approach (early modern palaeography, encoding, etc.).
This article highlights the main scientific and technical challenges encountered during the RePPOL project, including the choice of the most appropriate technological tools for collaborative transcription; dialogue and arbitration between engineers and researchers on the form of the deliverable; and consideration of financial and time constraints when defining the content of the project.
The final digital edition is now available for consultation (https://reppol.meshs.fr/).

INDEX

Mots-clés
édition numérique, formation par la recherche, enquête, humanités numériques, Angleterre, Réforme, Henri VIII

Keywords
digital humanities, digital edition, investigation, early modern England, Reformation, Henry VIII, education-through-research

AUTHOR

Olivier Spina
Université Lumière Lyon 2, LARHRA UMR 5190
olivier.spina[at]univ-lyon2.fr
IDREF : https://www.idref.fr/164274421
HAL : https://cv.archives-ouvertes.fr/olivier-spina
ISNI : http://www.isni.org/0000000381848777
BNF : https://data.bnf.fr/fr/16542183

Faciliter l'édition numérique avec les méthodes de reconnaissance automatique de texte

L'application de l'outil Transkribus dans le projet « Grand Tour digital »

Facilitating digital publishing with automatic text recognition methods. The application of the Transkribus tool in the 'Grand Tour digital' project.

Angela Göbel

DOI : 10.35562/theia.129

Copyright
CC BY-NC-SA 4.0

OUTLINE

Introduction
Le projet « Grand Tour digital »
  Enjeux et buts du projet
  Corpus des sources
La reconnaissance automatique des récits de voyage avec l'outil Transkribus
À propos de l'application de Transkribus et le workflow élaboré
  Créer un premier modèle de reconnaissance de textes
  Évaluer la transcription automatique
À propos des possibilités et limites de Transkribus

TEXT

Introduction

1 L'utilisation des méthodes de reconnaissance automatique dans les projets d'éditions notamment numériques gagne de plus en plus en importance dans les sciences humaines et sociales. À partir du projet d'édition en cours « Grand Tour digital », cet article propose un retour d'expérience : il souhaite mettre en avant les méthodes mobilisées et discuter les possibilités et les limites qui sont

progressivement apparues [1]. Réalisé à la Herzog August Bibliothek à Wolfenbüttel (HAB) en Allemagne, ce projet a pour but de numériser, d'exploiter et de visualiser des témoignages de voyage de formation ou du *Grand Tour* à l'aide des méthodes d'édition semi-automatiques.

2 Traitant un corpus d'une vingtaine de récits de voyage de l'époque moderne qui sont principalement conservés dans les fonds de la bibliothèque, le projet recourt à la plateforme de reconnaissance de texte et d'écriture manuscrite automatique Transkribus. Il apparaît que si cet outil est en mesure de faire face à un certain nombre de défis, la qualité de la numérisation dépend néanmoins fortement de l'état des manuscrits.

Le projet « Grand Tour digital »

Enjeux et buts du projet

3 À l'époque moderne, les voyages éducatifs tels que le *Grand Tour* étaient une étape cruciale dans le développement personnel des membres des élites nobles et bourgeoises. Leur signification particulière, jusqu'à présent peu étudiée dans la recherche sur les témoignages personnels (*Selbstzeugnisforschung*), réside dans le fait qu'ils capturaient et évaluaient moins les expériences en tant qu'altérité, mais mettaient plutôt en avant la réussite de la gestion des expériences dans des environnements étrangers. Ce type de témoignages personnels fait l'objet du projet d'édition « Grand Tour digital. *Digitalisierung, Erschließung und Visualisierung frühneuzeitlicher Selbstzeugnisse von Bildungsreisen unter Anwendung teilautomatisierter Editionsverfahren* [2] », soutenu financièrement par la *Deutsche Forschungsgemeinschaft* (DFG), de 2022 à 2025.

4 Employée dans le nom du projet, la notion de « Grand Tour » est utilisée dans un sens large. Par conséquent, le corpus de sources ne se limite pas uniquement aux récits du *Grand Tour* classique qui emmenait les voyageurs en France et surtout en Italie, mais inclut également d'autres voyages, tels les tournées à cheval des nobles-patriciens (*adlig-patriziche Kavarlierstour*), les pérégrinations académiques étudiantes (*studentische peregrinatio academica*) et les voyages savants professoraux, comme ceux distingués par

Mathis Leibeteseder [3]. Ces distinctions permettent de classer ces témoignages personnels rendus disponibles à travers leurs numérisations. Les textes édités seront mis à disposition au cours du projet sur le portail des témoignages personnels (*Selbstzeugnisportal*) de la HAB [4]. Le travail est documenté dans un blog accompagnant le projet [5], et une offre ultérieure pour la réutilisation du modèle HTR, des textes et des données de recherche sera également disponible via le serveur GitLab [6] de la HAB.

Corpus des sources

5 Au cœur du projet se trouvent 21 journaux de voyage, principalement rédigés en allemand entre les années 1550 et 1770 (pour un total d'environ 10 300 pages). Ils proposent des récits de voyage dans toute l'Europe et jusqu'à l'Empire ottoman et le Proche-Orient (Alep, Jérusalem). Ces journaux sont conservés principalement à la HAB à l'exception de deux manuscrits qui se trouvent aux Archives d'État de Basse-Saxe (Niedersächsisches Landesarchiv) à Wolfenbüttel.

6 Le projet s'intéresse particulièrement à cinq textes, dont l'un est accompagné de trois transcriptions, qui seront transcrits de manière partiellement automatisée grâce au logiciel de reconnaissance d'écriture manuscrite Transkribus, puis encodés en TEI-XML. L'utilisation de la reconnaissance d'entités nommées (*Named Entity Recognition*, REN) permet leur exploration, et leur visualisation est réalisée en combinant le texte avec l'itinéraire du voyage. Cette approche vise à expérimenter le développement d'un processus éditorial novateur.

7 Ces cinq récits ont été rédigés par les personnes suivantes :

- Barthold von Gadenstedt (HAB Cod. Guelf. 67.6 Extrav.)
- Le pharmacien Wagener (HAB Cod. Guelf. 267.1 Extrav.)
- Christian August de Schleswig-Holstein-Norburg (HAB Cod. Guelf. 221 Extrav.)
- Ernst Ferdinand et Heinrich Ferdinand de Brunswick-Wolfenbüttel-Bevern [7] (HAB Cod. Guelf. 149.14 Extrav.)
- Ludwig Rudolph de Brunswick-Wolfenbüttel (HAB Cod. Guelf. 89 Blank.)

8 S'ajoutent à cela 13 récits de voyage supplémentaires, qui sont numérisés. Leurs métadonnées sont explorées et seront mises à

disposition via la base de données des manuscrits de la bibliothèque [8].

La reconnaissance automatique des récits de voyage avec l'outil Transkribus

9 La quantité de projets éditoriaux utilisant des logiciels de reconnaissance manuscrite, dont Transkribus, a augmenté de manière significative ces dernières années [9]. L'application de méthodes d'intelligence artificielle offre une nouvelle approche pour interagir avec et traiter des sources historiques, qu'elles soient imprimées ou manuscrites.

10 Transkribus est une plateforme dédiée à la transcription automatique, à l'analyse d'images et à la reconnaissance de structures de documents historiques grâce à l'intelligence artificielle. Le programme a été développé à la suite de deux projets consécutifs à l'Université d'Innsbruck : « tranScriptorium », de 2013 à 2015, et « READ » (*Recognition and Enrichment of Archival Documents*), de 2016 à 2019. Pour l'utilisateur, il est possible d'utiliser Transkribus soit via l'application téléchargeable sur le bureau, soit en ligne avec Transkribus Lite. L'inscription et l'utilisation sont gratuites, mais pour la transcription automatisée, l'utilisateur dispose d'un quota spécifique de « crédits ». Après leur épuisement, il peut en acheter de nouveaux dans différents packs.

11 Transkribus propose de nombreux modèles d'intelligence artificielle (IA) disponibles gratuitement. En janvier 2024, on en compte 138 créés pour des sources manuscrites et imprimées de différentes périodes, ainsi que pour de nombreuses langues européennes et extra-européennes. Pour ce projet, comment choisir le modèle qui convient à cette écriture manuscrite ? Quels critères sont importants, voire convaincants, pour la sélection du modèle de base ? Et pourquoi créer un nouveau modèle pour les écritures de la période moderne, alors que plusieurs modèles sont déjà disponibles [10] ?

12 Transkribus fait face à divers défis lors de la transcription automatisée de récits de voyage de l'époque moderne, tels que celui

de Wagener, mais ces défis sont également applicables de manière générale aux témoignages personnels de cette époque :

- Aspect de l'écriture et lisibilité : la qualité de la transcription dépend fortement de la forme de l'écriture, qui elle-même varie en fonction de la version originale ou d'une copie, ainsi que de l'état de conservation de la source.
- Écriture individuelle : chaque individu a une écriture unique. Les fautes d'orthographe, les ratures, les ajouts, ainsi que l'encre étalée, pressée ou décolorée compliquent la reconnaissance automatisée.
- Mise en page variée : les récits de voyage de l'époque moderne présentent souvent une mise en page variée, incluant le texte principal, des annotations marginales, des tableaux, des listes, des schémas ainsi que des esquisses de taille différente. S'ajoutent de manière répétée des compléments et corrections ultérieurs directement dans le corpus du texte, qui sont soit ajoutés en marge du texte, soit insérés entre les lignes.
- Polices et tailles de caractères variées : un texte peut contenir différentes polices et tailles de caractères, comprenant notamment la cursive allemande, l'écriture latine, ou d'autres encore.
- Utilisation de différentes langues : l'inclusion de différentes langues, comme des copies d'inscriptions ou des citations, représente un défi supplémentaire et peut demander au modèle d'être entraîné sur plusieurs langues.
- Exigences de contenu : les récits de voyage de l'époque moderne contiennent souvent un grand nombre de noms propres (personnes, lieux, œuvres d'art, titres de livres, etc.), des indications de date, des mesures et des caractères spéciaux qui doivent être précisément identifiés et transcrits.

13 Dans le cadre du projet sera développé un modèle entraîné à ces exigences qui sera partagé en accès libre à son achèvement.

À propos de l'application de Transkribus et le work-flow élaboré

Créer un premier modèle de reconnaissance de textes

14 L'objectif central du projet « Grand Tour digital » est la mise au point expérimentale d'un processus éditorial novateur en utilisant Transkribus. Le logiciel lui-même fournit de nombreuses instructions utiles pour l'édition de texte [11]. En outre, la recherche produit de plus en plus de rapports d'expérience, de discussions scientifiques et de conseils pratiques sur cet outil [12].

15 Lors d'une première étape, après avoir transcrit manuellement une vingtaine des pages du journal de voyage de Wagener dans Transkribus, on a pu partir des pages saisies et d'un modèle déjà existant pour permettre l'entraînement d'un nouveau modèle. Pour sa création, nous avons choisi le modèle existant « Transkribus German handwriting M1 » avec une faible erreur ou taux d'erreur de caractères (CER) [13] de 4,70 % comme modèle de base. Ce modèle a été réentraîné à partir de la transcription manuelle des premières pages (p. 7 à 26).

16 Une fois entraîné, ce modèle a été appliqué aux autres pages du manuscrit et ajusté dans des étapes alternées de transcription partiellement automatisée, de correction et de nouvel entraînement du modèle. L'idée centrale derrière ces étapes de travail était d'adapter le programme aux particularités de l'écriture de Wagener afin de réduire progressivement le taux d'erreur dans la reconnaissance manuscrite au cours de la transcription du texte. Dans cette perspective, il est important de comprendre les caractéristiques spécifiques du manuscrit qui a composé cette première base de départ. En ce qui concerne l'indexation du texte partiellement automatisée dans Transkribus, le manuscrit présente généralement une écriture claire avec peu de ratures ou d'ajouts. Parfois, l'encre a traversé le papier, ce qui rend la lecture du

manuscrit difficile tant pour l'homme que pour la machine. À certains endroits de son récit de voyage, Wagener a également ajouté des notes sur de petits morceaux de papier collés à l'intérieur de son journal. La langue du texte est principalement l'allemand, mais on trouve à plusieurs reprises des citations transcrites d'inscriptions en français et en latin, surtout à la fin du récit [14].

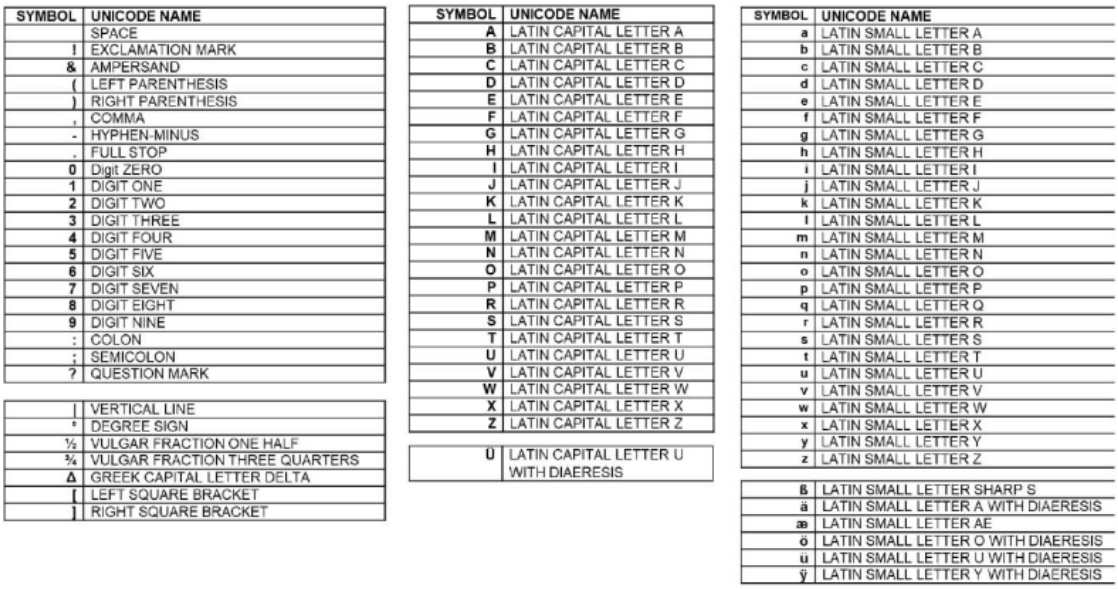
17 Pour au mieux réussir la transcription automatisée dans Transkribus, le texte a été préparé de manière à minimiser autant que possible les erreurs de détection des champs de texte. À cette fin, le marquage des champs de texte peut soit être parcouru automatiquement et corrigé par la suite, soit être créé directement manuellement. Selon le manuscrit, l'une ou l'autre option est recommandée pour travailler de manière aussi efficace que possible. Dans le cas du manuscrit de Wagener, le marquage automatique des champs de texte a été effectué pour les pages restantes du manuscrit, puis, lors de l'examen ultérieur, de petites corrections ont été apportées en cas d'erreurs dans le marquage du texte principal et des notes en marge ainsi que des lignes manquantes ont été ajoutées. Ce travail se fait plus difficile lorsque les champs de texte ou même les mots se chevauchent et se superposent, mais il est également possible de corriger le tout manuellement dans ce cas.

18 Lors de l'entraînement d'un nouveau modèle, Transkribus distingue entre les pages d'entraînement et de validation. Les pages d'entraînement lors du premier passage (p. 7 à 21) étaient celles où le modèle était formé et les pages de validation (p. 22 à 26) étaient celles où le modèle était automatiquement vérifié et le taux d'erreur calculé [15]. Au cours de cette première itération, l'ensemble d'entraînement, associé au modèle de base avait atteint un taux d'erreur de 2,41 %. Les pages de validation, idéalement représentatives des particularités du manuscrit, avaient un taux d'erreur de 11 %. Afin d'optimiser les résultats sur les pages de validation lors des ajustements ultérieurs du modèle, des pages de manuscrit non consécutives ont été sélectionnées à intervalles réguliers (par exemple, par intervalles de cinq pages). Les 86 pages restant à traiter ont été lues automatiquement par Transkribus par tranches d'environ 20 pages chacune, puis corrigées manuellement et réentraînées avec les pages précédentes. En cas de pages fortement variables, il est également possible – à l'instar de la suggestion et de la

documentation de Jacob Möhrke dans son rapport d'atelier – de trier les pages de manuscrit en fonction de leur qualité textuelle et de retirer du jeu d'entraînement les pages qualifiées de « non utilisables[16] ». Ce processus itératif visait à optimiser progressivement le modèle de transcription.

19 Le logiciel Transkribus propose lui-même une vue d'ensemble du jeu de caractères (*character set*) entraîné pour chaque modèle formé. Pour le manuscrit de Wagener, les caractères suivants ont fait sujet de l'entraînement :

Fig. 1. Vue d'ensemble des signes entraînés dans Transkribus à l'aide des premières pages du manuscrit Wagener.



SYMBOL	UNICODE NAME
	SPACE
!	EXCLAMATION MARK
&	AMPERSAND
(	LEFT PARENTHESIS
)	RIGHT PARENTHESIS
,	COMMA
-	HYPHEN-MINUS
.	FULL STOP
0	Digit ZERO
1	DIGIT ONE
2	DIGIT TWO
3	DIGIT THREE
4	DIGIT FOUR
5	DIGIT FIVE
6	DIGIT SIX
7	DIGIT SEVEN
8	DIGIT EIGHT
9	DIGIT NINE
:	COLON
;	SEMICOLON
?	QUESTION MARK

|	VERTICAL LINE
°	DEGREE SIGN
½	VULGAR FRACTION ONE HALF
¾	VULGAR FRACTION THREE QUARTERS
Δ	GREEK CAPITAL LETTER DELTA
[	LEFT SQUARE BRACKET
]	RIGHT SQUARE BRACKET

SYMBOL	UNICODE NAME
A	LATIN CAPITAL LETTER A
B	LATIN CAPITAL LETTER B
C	LATIN CAPITAL LETTER C
D	LATIN CAPITAL LETTER D
E	LATIN CAPITAL LETTER E
F	LATIN CAPITAL LETTER F
G	LATIN CAPITAL LETTER G
H	LATIN CAPITAL LETTER H
I	LATIN CAPITAL LETTER I
J	LATIN CAPITAL LETTER J
K	LATIN CAPITAL LETTER K
L	LATIN CAPITAL LETTER L
M	LATIN CAPITAL LETTER M
N	LATIN CAPITAL LETTER N
O	LATIN CAPITAL LETTER O
P	LATIN CAPITAL LETTER P
R	LATIN CAPITAL LETTER R
S	LATIN CAPITAL LETTER S
T	LATIN CAPITAL LETTER T
U	LATIN CAPITAL LETTER U
V	LATIN CAPITAL LETTER V
W	LATIN CAPITAL LETTER W
X	LATIN CAPITAL LETTER X
Z	LATIN CAPITAL LETTER Z

Ü	LATIN CAPITAL LETTER U WITH DIAERESIS

SYMBOL	UNICODE NAME
a	LATIN SMALL LETTER A
b	LATIN SMALL LETTER B
c	LATIN SMALL LETTER C
d	LATIN SMALL LETTER D
e	LATIN SMALL LETTER E
f	LATIN SMALL LETTER F
g	LATIN SMALL LETTER G
h	LATIN SMALL LETTER H
i	LATIN SMALL LETTER I
j	LATIN SMALL LETTER J
k	LATIN SMALL LETTER K
l	LATIN SMALL LETTER L
m	LATIN SMALL LETTER M
n	LATIN SMALL LETTER N
o	LATIN SMALL LETTER O
p	LATIN SMALL LETTER P
q	LATIN SMALL LETTER Q
r	LATIN SMALL LETTER R
s	LATIN SMALL LETTER S
t	LATIN SMALL LETTER T
u	LATIN SMALL LETTER U
v	LATIN SMALL LETTER V
w	LATIN SMALL LETTER W
x	LATIN SMALL LETTER X
y	LATIN SMALL LETTER Y
z	LATIN SMALL LETTER Z

ß	LATIN SMALL LETTER SHARP S
ä	LATIN SMALL LETTER A WITH DIAERESIS
æ	LATIN SMALL LETTER AE
ö	LATIN SMALL LETTER O WITH DIAERESIS
ü	LATIN SMALL LETTER U WITH DIAERESIS
ÿ	LATIN SMALL LETTER Y WITH DIAERESIS

20 Cette vue d'ensemble ne couvre pas tous les caractères possibles de l'écriture cursive latine, mais se limite aux seuls caractères utilisés dans le manuscrit. Cependant, il n'est pas tout à fait évident de savoir si cette liste inclut également des caractères du modèle de base sélectionné, dans notre cas « Transkribus German Handwriting M1 », ou si elle se limite exclusivement aux premières 26 pages du manuscrit de Wagener. Malheureusement, dans la présentation du modèle, il n'existe pas de référence explicite au modèle de base et des informations approfondies sur ce dernier font défaut, rendant une analyse plus approfondie et une meilleure compréhension de la procédure difficile.

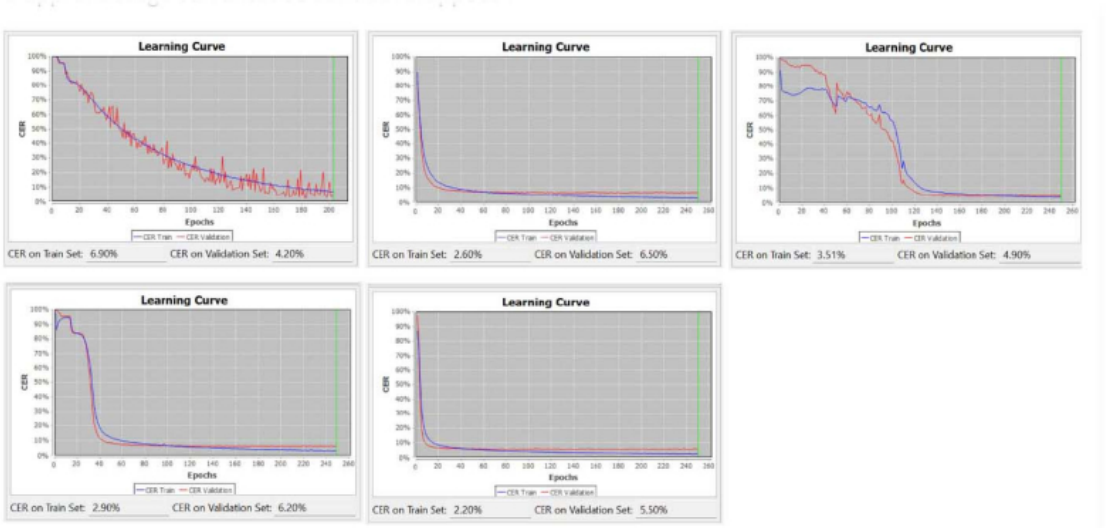
21 Le jeu de caractères généré lors de la première formation avec Transkribus a été élargi et complété avec des caractères manquants au cours des formations et ajustements ultérieurs :

Fig. 2. Vue d'ensemble des signes rajoutés lors des entrainements supplémentaires du modèle HTR.

SYMBOL	UNICODE NAME
‘	APOSTROPHE
+	PLUS SIGN
/	SOLIDUS
Q	LATIN CAPITAL LETTER Q
Y	LATIN CAPITAL LETTER Y
Æ	LATIN CAPITAL LETTER AE
Ö	LATIN CAPITAL LETTER O QITH DIAERESIS
à	LATIN SMALL LETTER A WITH GRAVE
è	LATIN SMALL LETTER E WITH GRAVE
ō	LATIN SMALL LETTER O WITH MACRON
œ	LATIN SMALL LIGATURE OE
ū	LATIN SMALL LETTER U WITH MACRON
Ɔ	LATIN CAPITAL LETTER OPEN O
ƚƀ	L B BAR SYMBOL
△	WHITE UP-POINTING TRIANGLE
⏉	ALCHEMICAL SYMBOL FOR VITRIOL

22 En résumé, lors de la formation du modèle sur l'exemple du manuscrit Wagener, les courbes d'apprentissage suivantes se sont développées :

Fig. 3. Vue des cinq courbes d'apprentissage réalisées lors des entraînements du modèle au manuscrit Wagener. De gauche en haut à droite en bas chaque courbe représente un entrainement.



23 Cette optimisation du modèle fonctionne-t-elle ? D'après les courbes d'apprentissage, une amélioration réelle n'a pas été totale, voire elle a échoué, mais pour quelles raisons ? Les courbes d'apprentissage démontrent une fois de plus que la qualité et le succès ou l'échec de la reconnaissance manuscrite dépendent fortement des pages et soulignent également qu'une sélection liée la qualité de certaines pages pour l'entraînement du modèle serait avantageuse. Une amélioration progressive du modèle n'est pas clairement perceptible dans le cas de l'écriture manuscrite de Wagener – le taux d'erreur, en particulier pour les pages de validation, fluctue entre 4 et 6 %. C'est encore nettement trop élevé dans tous les cas. Cependant, pour les pages d'entraînement, elle diminue de 7 % à environ 2 %. À la lumière de ces premiers résultats, deux questions se posent pour le traitement continu du texte : comment le résultat de l'entraînement du modèle diffère-t-il du résultat de l'application du modèle ? Y a-t-il un meilleur apprentissage en répétant l'entraînement de l'écriture manuscrite en éliminant les pages « inutilisables », ou altère-t-on progressivement le résultat global de cette manière ?

24 Au regard de ces questions, il est intéressant de savoir s'il peut y avoir une amélioration au cours du traitement ultérieur des autres manuscrits avec ce même modèle et si cette amélioration dépendra

de l'inclusion ou de l'exclusion de certaines pages de manuscrit. En réalité, le développement des courbes d'apprentissage pour le premier manuscrit n'a pas toujours été optimal et offre encore suffisamment d'espace pour une amélioration dans le cadre du travail continu de ce projet.

Évaluer la transcription automatique

25 Finalement, qu'est-ce qui a bien fonctionné, qu'est-ce qui n'a pas fonctionné ? Quels problèmes ont été identifiés et quelles solutions Transkribus propose-t-il ? Des évaluations comparables de transcriptions automatisées ont déjà été réalisées dans le cadre d'autres projets et les contributions scientifiques qui en ont résulté soulignent les opportunités de succès ainsi que les difficultés, telles que la dépendance à l'égard de la forme de l'écriture ou le risque de surapprentissage d'un modèle [17].

26 Les *insights* tirés de divers projets et manuscrits sur lesquels Transkribus a été testé montrent clairement que même avec des modèles bien entraînés, une précision de reconnaissance manuscrite à 100 % ne peut jamais être atteinte par une IA : « La reconnaissance automatique de l'écriture manuscrite conduira à un taux d'erreur même avec des modèles bien entraînés. Ainsi, aucun texte philologiquement irréprochable ne sera généré sans correction manuelle ultérieure (post-traitement) [18] ».

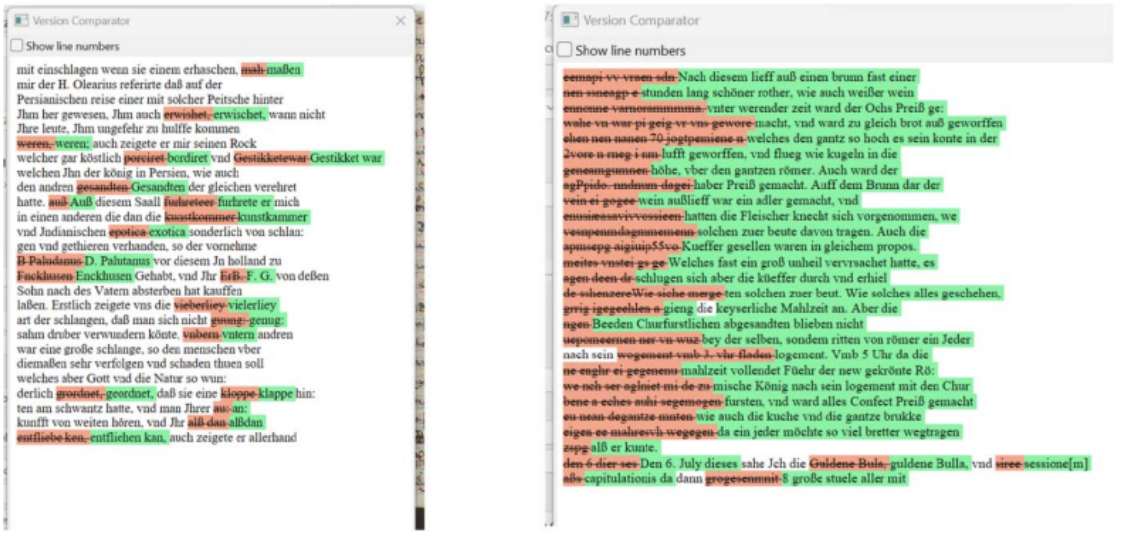
27 À partir du manuscrit de Wagener, nous avons pu examiner plus précisément l'évaluation dans Transkribus, présenter et remettre en question les possibilités d'évaluation automatisée des résultats, identifier les types d'erreurs et discuter de leurs causes possibles. Transkribus offre lui-même la possibilité d'évaluer la transcription effectuée directement dans le programme. Il convient de faire la distinction entre l'évaluation du taux d'erreur (CER) sur les pages de test et de validation et l'application du modèle entraîné sur les autres pages du manuscrit. La précision de la transcription peut être calculée à différents niveaux. Pour la comparaison, deux variantes de texte doivent être sélectionnées. Il faut d'abord une « Référence », un texte correct servant de référence pour l'identification des erreurs. Transkribus recommande « [e]n tant que Référence, choisissez une version de page qui a été correctement transcrite (*Ground Truth* :

transcription manuelle aussi proche que possible du texte original). Pour obtenir la valeur la plus significative, il serait préférable d'utiliser des pages d'un ensemble d'échantillons qui n'ont pas été utilisées dans l'entraînement et qui sont donc nouvelles pour le modèle [19] ». Pour Wagener, nous avons utilisé la version corrigée la plus récente, qui est la plus proche du manuscrit du point de vue orthographique.

28 En regard de ce texte de référence, il faut établir une « hypothèse », c'est-à-dire le texte « hypothétique » à comparer avec la variante correcte. Cela peut être la transcription automatisée par Transkribus, mais aussi des variantes textuelles ultérieures, afin d 'examiner ou de suivre certaines corrections entre collaborateurs. Transkribus recommande » en tant qu'Hypothèse, choisissez la version qui a été générée automatiquement avec un modèle de reconnaissance de texte manuscrit (HTR) et sur laquelle vous aimeriez voir à quel point le résultat est satisfaisant [20] ». Ainsi, nous avons choisi la reconnaissance de texte automatisée (décodage PyLaia). La possibilité offerte par Transkribus de comparer différentes versions de traitement de texte permet une confrontation directe entre la transcription partiellement automatisée et la correction manuelle effectuée. Cependant, il est difficile de comprendre exactement comment fonctionnent la comparaison et l'évaluation qui en découle. Contrairement à d'autres programmes de reconnaissance de texte manuscrit, cela n'est pas transparent dans le cas de Transkribus.

29 Les problèmes rencontrés lors de l'évaluation automatique concernent principalement la détection des lignes, surtout lorsqu'elle est corrigée ultérieurement ou lorsqu'il y a des améliorations dans l'ordre des lignes. Par exemple, les lignes peuvent être difficile à identifier si l'une d'entre elles est divisée en deux et que les deux éléments sont fusionnés, si une note est ajoutée au-dessus de la ligne, ou si une note en marge est séparée du texte principal après coup. Dans ces cas, l'ensemble du texte se déplace, ce qui entraîne également des parties considérées comme incorrectes ou affichées comme telles, bien qu'elles soient au moins partiellement correctes. En effet, Transkribus prétend attribuer une erreur à chaque caractère mal reconnu.

Fig. 4. Les deux captures d'écran montrent le résultat de la transcription automatique le plus réussi (à gauche) et le plus mauvais (à droite) [Des corrections ultérieures n'ont pas pu être prises en compte ici]



30 Dans cette perspective, un problème supplémentaire d'évaluation avec Transkribus est lié au fait que le système se base sur les mots et non sur les caractères individuels. Pour un observateur extérieur, les résultats semblent donc plutôt moyens, même s'ils sont en réalité nettement meilleurs. On le comprend lorsque l'on compare l'évaluation de l'écriture manuscrite de Wagener au niveau des mots et des caractères : au niveau des mots, le taux d'erreur est généralement plus élevé qu'au niveau des caractères.

31 Étant donné que Transkribus fournit l'évaluation automatique de la transcription au niveau des caractères et alors que pour les mots il le fait uniquement en pourcentages, nous avons effectué une évaluation manuelle de la transcription automatisée et nous avons compté les pages. Lors de l'évaluation manuelle de la transcription, nous avons veillé à compter toutes les reconnaissances divergentes. L'objectif était de découvrir et de comprendre la fiabilité de l'évaluation automatisée de Transkribus. Ce qui est étonnant, c'est qu'apparaissent des écarts significatifs dans les pourcentages. Ainsi, la somme des valeurs, que ce soit pour le taux d'erreur de mots (WER) et la précision des mots, ou pour le taux d'erreur de caractères (CER) et la précision des caractères, n'a pas atteint les 100 %. Il est

également frappant de noter que les valeurs lors de l'évaluation du taux d'erreur de mots (WER) dans l'évaluation automatisée correspondent approximativement à mes propres calculs ; cependant, au niveau des caractères, Transkribus présente un taux d'erreur de caractères de 41,66 %, alors qu'après le calcul manuel, seulement près de 17 % des caractères ont été mal lus. De plus, le résultat global est meilleur que ce que Transkribus avait prévu. Ainsi, pour l'écriture manuscrite de Wagener, nous avons atteint une précision d'à peine 58 % au niveau des mots mais à plus de 80 % au niveau des caractères. En tenant compte du faible nombre de pages de l'écriture manuscrite de Wagener sur lesquelles Transkribus a pu apprendre, le résultat obtenu ici est déjà bon, bien qu'il reste possible et souhaitable de l'améliorer davantage.

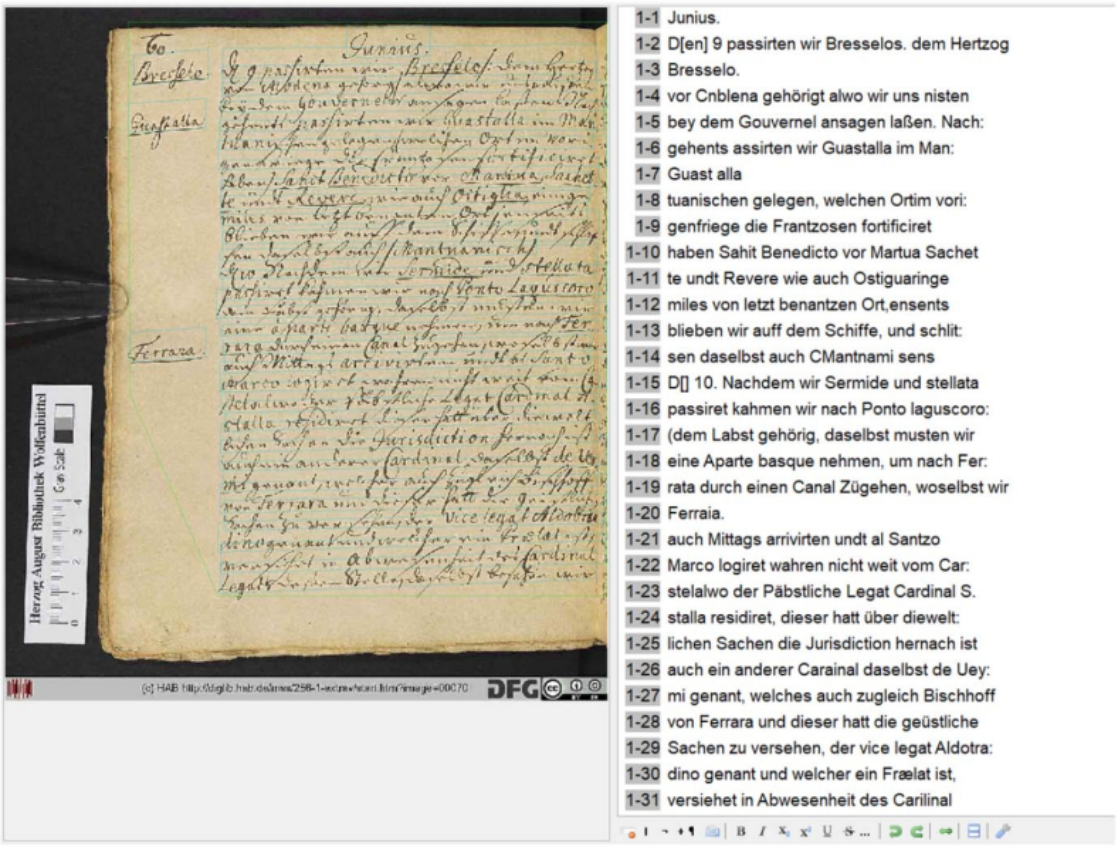
32 En plus de l'affichage de la fréquence des erreurs, une évaluation des pages manuscrites individuelles au niveau des mots et des caractères offre un aperçu plus précis du déroulement et des résultats des différentes itérations d'entraînement du modèle. Le nouveau cycle d'entraînement du modèle a-t-il été bénéfique ? Comment le résultat de la transcription se compare-t-il avec le taux d'erreur prédéfini dans l'ensemble d'entraînement et de validation ?

33 Lorsqu'on examine les résultats de la transcription partiellement automatisée, une analyse plus approfondie des erreurs survenues (types d'erreurs) est recommandée. Cela inclut notamment des lectures complètement erronées de mots et parfois même de lignes entières, qui déformaient le contenu ou le rendaient méconnaissable, rendant ainsi la lecture du manuscrit plus difficile. Comme mentionné précédemment, ces erreurs dépendaient fortement des pages et étaient causées par l'apparence de l'écriture, l'encre détrempée, des ajouts de mots isolés ou des ratures qui n'étaient généralement pas reconnus par le programme. Sur certaines pages, les erreurs étaient également causées par des petits morceaux de papier mal reliés, qui ne posaient généralement pas de problème à la reconnaissance des champs de texte par le logiciel, mais qui compliquaient la reconnaissance des caractères [21]. Dans l'ensemble, des lignes entières étaient souvent mal lues, transcrivant une séquence de lettres incompréhensible, tandis que des lignes entièrement lues sans erreur étaient plutôt l'exception.

34 En général, les erreurs concernaient souvent des lettres individuelles (notamment la confusion entre u/v, s/f, m/n, c/e, b/l, etc.), des signes de ponctuation (virgules manquantes ou mal lues, ponctuations, parenthèses) ou des espaces mal lus, trop peu ou trop nombreux. Certaines lettres ont été lues en double, d'autres ont été ignorées ; la quantité de caractères dans la version corrigée ne correspondait pas toujours à celle de la version lue automatiquement. Parfois, des signes de ponctuation ont été identifiés comme des lettres et vice versa, parfois l'ordre a été inversé ou mélangé, par exemple, « *wire* » au lieu de « *wier* ». Certaines erreurs étaient malheureusement dues uniquement à des espaces manquants, supplémentaires ou inutiles. Moins graves sont également les erreurs liées aux accents et aux trémas. Il est à noter que le logiciel n'a presque pas reconnu les soulignements, les mises en évidence, les ratures, les élévations et les indices de caractères individuels. Cela peut cependant être dû au fait que les zones de texte précédemment marquées n'ont pas nécessairement été incluses dans l'entraînement du modèle et la tâche de reconnaissance manuscrite partiellement automatisée en tant que telle. En ce qui concerne les caractères spéciaux, le programme a appris au fil de l'ajustement du modèle et s'est amélioré. Les caractères étaient généralement bien lus (par exemple, dans les indications de quantités) et seulement quelques erreurs sont survenues. Dans l'ensemble, la reconnaissance des dates et des chiffres généraux était défectueuse.

35 Le modèle a pu être optimisé au fil de son utilisation ultérieure sur des manuscrits subséquents, atteignant jusqu'à présent un taux d'erreur de 0,70 % dans l'ensemble d'entraînement et 2,60 % dans l'ensemble de validation.

Fig. 5. La capture d'écran montre comme exemple le résultat de la transcription automatique du manuscrit HAB Cod. Guelf. 256.1 Extrait à l'aide de l'application du modèle de Wagener.



36 Cependant, Transkribus atteint rapidement ses limites. Des ajustements échouent lorsque le programme ne peut plus améliorer la reconnaissance d'une écriture manuscrite. Les entraînements n'aboutissent alors pas : dans les ensembles d'entraînement et de validation, apparaissent des taux d'erreur inexplicables et élevés, dépassant parfois les 80 %, voire les 90 %. Lorsque de tels échecs surviennent, aucune amélioration supplémentaire n'est possible et il faut revenir au modèle précédemment entraîné ayant obtenu les meilleurs résultats.

À propos des possibilités et limites de Transkribus

37 Comme le montre l'exemple du manuscrit de Wagener, le succès d'une reconnaissance automatisée de l'écriture manuscrite dépend fortement de chaque page et de la ligne transcrite. De plus, la lisibilité et la qualité du scan du manuscrit, ainsi que la présence de ratures et d'ajouts (notamment entre deux lignes), sont d'une importance particulière pour le succès de la reconnaissance automatisée de l'écriture manuscrite et peuvent l'influencer négativement. Transkribus rencontre peu de problèmes de césure de mots, de distinction entre majuscules et minuscules, de ponctuation, ainsi que de reconnaissance de caractères spéciaux et d'abréviations. Cependant, des difficultés surviennent lors de la détection d'espaces et de l'apprentissage de différentes orthographes pour les mêmes mots (*vnd/und*, *aus/auß*…). Les noms propres (notamment les lieux et les personnes), les chiffres et les indications de date sont particulièrement sujets aux erreurs. La question des règles de transcription prédéfinies est alors cruciale. Pour la reconnaissance de l'écriture manuscrite, l'idéal est de rester aussi proche que possible de l'original. Dans le cas de Wagener, cela concerne, par exemple, la distinction entre les lettres I et J, la majuscule et la minuscule, ainsi que la ponctuation. Si des corrections doivent être apportées dans le cadre d'une édition critique, il est recommandé de les ajouter à une étape ultérieure pour éviter de risquer d'entraîner des erreurs dans Transkribus.

38 L'application antérieure du modèle Wagener sur d'autres manuscrits a permis une amélioration du modèle. Cependant, l'IA atteint rapidement ses limites, surentraînant le modèle et entraînant des taux d'erreur inexplicablement élevés dans les ensembles d'entraînement et de validation. Une amélioration supplémentaire du modèle précédent n'est donc plus possible. La question se pose alors de savoir dans quelle mesure ce surentraînement est spécifique au programme Transkribus ou s'il peut être attribué de manière générale à tous les systèmes de transcription manuscrite automatisée (HTR) pilotés par une IA.

39 Transkribus semble avoir joué un rôle de pionnier au cours de ces dernières années, mais il ne constitue pas la seule possibilité d'exploiter des documents historiques grâce à l'IA. Comme l'a déjà souligné Elpida Perdiki, il existe plusieurs alternatives au programme de reconnaissance manuscrite Transkribus. Outre l'application open-source eScriptorium, elle en mentionne d'autres : « python systems implémentés avec la bibliothèque TensorFlow, tels que Kraken, un système OCR pour les documents historiques, et Tesseract, le moteur OCR développé par Google et principalement utilisé dans de nombreux projets [22] ». Il serait autant plus intéressant dans cette perspective, de faire une étude comparative entre ces différents outils d'HTR à partir d'une source manuscrite exemplaire.

NOTES

1 Le présent article contient des observations déjà partagées partiellement en langue allemande sur le blog accompagnant le projet, https://grandtourdig.hypotheses.org qui ont été présentement rassemblées, traduites et complétées par des recherches supplémentaires entreprises depuis. Pour lire les rapports et témoignages périodiquement partagés sur le blog, voir notamment les deux billets suivants : Angela Göbel, « Erste Schritte in Transkribus (1) – Modelltraining und Transkription der Handschrift Wagener », in : *Grand Tour digital. Ein Forschungsblog zum Editionsprojekt frühneuzeitlicher Selbstzeugnisse an der Herzog August Bibliothek Wolfenbüttel*, mise en ligne le 6 juillet 2023 (https://grandtourdig.hypotheses.org/290), et Angela Göbel, « Erste Schritte in Transkribus (2) – Auswertung der Transkription der Handschrift Wagener », in : *Grand Tour digital. Ein Forschungsblog zum Editionsprojekt frühneuzeitlicher Selbstzeugnisse an der Herzog August Bibliothek Wolfenbüttel*, mise en ligne le 20 décembre 2023 (https://grandtourdig.hypotheses.org/1022). Il s'agit ici de deux billets de blog qui discutent de plus près de l'application et de l'évaluation de l'outil Transkribus à partir de l'exemple du premier manuscrit traité. Le présent article propose un aperçu de nos recherches courantes et vise à mettre à disposition du public francophone les premiers résultats de recherche.

2 Grand Tour digital. Numérisation, exploration et visualisation de témoignages personnels de voyages éducatifs de l'époque moderne en utilisant des procédés d'édition partiellement automatisés

3 Voir Mathis Leibetseder, « Kavalierstour – Bildungsreise – Grand Tour: Reisen, Bildung und Wissenserwerb in der Frühen Neuzeit », dans Leibniz-Institut für Europäische Geschichte (IEG) Hrsg., *Europäische Geschichte Online* (EGO), mis en ligne le 14 août 2013, URL : http://www.ieg-ego.eu/leibetsederm-2013-de (consulté le 15 janvier 2024).

4 Voir le site du portail intitulé « Selbstzeugnisse der Frühen Neuzeit in der Herzog August Bibliothek », en ligne : https://selbstzeugnisse.hab.de. Le projet « Grand Tour digital » fait suite à deux projets d'édition précédents, déjà publiés sur le portail des témoignages personnels et disponibles en libre réutilisation. Voir Jacqueline Krone, David Maus et Inga Hanna Ralle (dir.), « Herzog August der Jüngere von Braunschweig-Wolfenbüttel (1579–1666): Ephemerides. Sive Diarium (1594–1635) », dans *Selbstzeugnisse der Frühen Neuzeit in der Herzog August Bibliothek*, Wolfenbüttel, HAB, 2017, URL : http://selbstzeugnisse.hab.de/edition_august, et Andreas Herz, Jan-Hendrik Hütten, Alexander Zirr et al. (dir.), « Digitale Edition der Tagebücher von Herzog Ludwig Rudolf und Herzogin Christine Luise von Braunschweig-Wolfenbüttel », dans*Selbstzeugnisse der Frühen Neuzeit in der Herzog August Bibliothek*, Wolfenbüttel, HAB, 2019–2022, URL : http://selbstzeugnisse.hab.de/edition_sz2.

5 https://grandtourdig.hypotheses.org/

6 https://git.hab.de/forschungsdaten

7 À côté de l'original, trois autres copies de ce récit de voyage sont conservées dans les collections de la Bibliothèque Herzog August (HAB) et des Archives d'État de Basse-Saxe à Wolfenbüttel. Ces textes font également l'objet de travaux de recherche dans le cadre du projet, visant d'une part à regrouper les collections dispersées, et d'autre part, à disposer d'un corpus de sources identique ou très similaire en tant que « groupe de contrôle » pour l'indexation semi-automatisée des textes avec Transkribus. Ces copies sont conservées dans HAB Cod. Guelf. 256.1 Extrav. ; Nds. LA WF, 1 Alt 20 Nr. 94 ; et Nds. LA WF, 95 ALT Nr. 44.

8 Une grande partie est déjà accessible en format numérisé via la base de données des manuscrits de la HAB, les entrées respectives dans le portail des manuscrits et le portail des témoignages personnels sont actuellement en cours de traitement et seront préparées d'ici la fin du projet. Voir « Auswahl nach Projekt: Grand Tour digital Digitalisierung, Erschließung und Visualisierung frühneuzeitlicher Selbstzeugnisse von Bildungsreisen unter Anwendung teilautomatisierter Editionsverfahren (21 Hss.) », dans

Handschriftendatenbank, URL : https://diglib.hab.de/?db=mss. Pour y accéder cliquez « Auswahl nach … », puis « Projekt », puis sélectionnez le présent projet. Une présentation plus détaillée du corpus est à trouver sur le blog accompagnant le projet, voir https://grandtourdig.hypotheses.org/uber.

9 Dans sa thèse de doctorat, Joe Nockels, par exemple, a étudié à l'Université d'Édimbourg l'utilisation du logiciel de transcription manuscrite Transkribus dans la recherche entre 2015 et 2020. Voir Joe Nockels et al., « Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research », *Archival Science* 22 (2022), p. 367–392.

10 Si l'on souhaite transcrire un manuscrit complet avec un modèle existant sans effectuer un nouvel entraînement, et que l'on a du mal à choisir parmi la multitude de modèles possibles, Transkribus offre la possibilité de la comparaison de motifs.

11 Le « centre de ressources » de Transkribus propose des informations détaillées, fournissant plusieurs guides pour l'utilisation du logiciel, voir URL : https://readcoop.eu/transkribus/resources/.

12 Voir par exemple le billet de blog présentant l'outil par Jan Horstmann et la littérature complémentaire répertoriée dans celui-ci : Jan Horstmann, « Transkribus », in : forText. *Literatur digital erforschen*, mis en ligne en 2018, URL: https://fortext.net/tools/tools/transkribus (consulté le 15 janvier 2024) ; Marc Rothballer, « Transkribus. Der Einsatz von maschinellem Lernen und Handwritten Text Recognition in der Erschließung historischer Dokumente », in : *FORUM. Zeitschrift des Bundes der öffentlichbestellten Vermessungsingenieure e.V.* 4 (2020), p. 29–35, URL : https://www.bdvi.de/application/files/9516/1459/2541/Transkribus_FORUM_4-2020.pdf (consulté le 15 janvier 2024), ainsi que le rapport d'expérience et d'atelier de Jacob Möhrke « Training eines Transkribus Sprachmodells », in : Begleitblog zum Regensburger Projekt DEHisRe – Digitale Editionen Historischer Reiseberichte, mis en ligne le 20 juillet 2020, https://dehisre.ios-regensburg.de/training-eines-transkribus-sprachmodells/ (consulté le 15 janvier 2024).

13 Le taux d'erreur de caractères (CER) compare, selon le glossaire de Transkribus, « pour une page donnée, le nombre total de caractères (n), espaces compris, au nombre minimum d'insertions (i), substitutions (s) et suppressions (d) de caractères) » (« for a given page, the total number of characters (n), including spaces, to the minimum number of insertions (i),

substitutions (s) and deletions (d) of characters »), nécessaires pour obtenir une correspondance à cent pour cent entre la reconnaissance manuscrite et la transcription correcte du document en question. La meilleure façon de former un nouveau modèle dans Transkribus est expliquée ici étape par étape. La page est régulièrement mise à jour et offre une bonne introduction à la formation de modèles personnalisés, qui ne sera pas répétée à cet endroit. Voir URL : https://readcoop.eu/glossary/character-error-rate-cer/.

14 Pour d'autres informations sur ce travail : Angela Göbel, Maximilian Görmar, « Zu Besuch in der Gottorfischen Kunstkammer. Berichte eines Apothekergesellen auf Reisen », *HABlog*, mis en ligne le 28 mars 2023. URL : https://www.hab.de/zu-besuch-in-der-gottorfischen-kunstkammer/ (consulté le 15 janvier 2024), ainsi que dans la recherche plus ancienne les articles d'Otto Hahne, « Erlebnisse des Apothekers Wagener auf seinen Wanderjahren und in Begleitung des Herzogs Ferdinand Albrechts I. zu Braunschweig und Lüneburg (1652–1659) », in : Braunschweigisches Magazin 33 (1927), p. 81–94, et ibid., « Die Reisen des Apothekers Wagener aus Itzehoe 1652–1659 », in : Pharmazeutische Zeitung 72 (1927), p. 951–952.

15 Il est important de noter que cet entraînement peut prendre un certain temps. En général, vous vous retrouverez dans une file d'attente et selon l'heure de la journée, cela peut prendre un peu plus de temps. Il est donc recommandé de prévoir toujours une demi-journée à une journée entière.

16 Voir à ce sujet Jakob Möhrke, *op. cit.*

17 Voir par exemple Elpida Perdiki, « Transkribus: Reviewing HTR training on (Greek) manuscripts », in : RIDE – A review journal for digital editions and resources 15 (2022), URL : https://ride.i-d-e.de/issues/issue-15/transkribus/ (consulté le 15 janvier 2024), ainsi que Ivan N. Petrov, Achim Rabus, « Linguistic Analysis of Church Slavonic Documents: A Mixed-Methods Approach », in : Scando-Slavica 69 (2023), N° 1, p. 25-38, URL : https://doi.org/10.1080/00806765.2023.2189617.

18 Gabriel Viehhauser, « Digitalisierung von Handschriften und frühen Drucken, OCR (Bericht über Kurzvorstellungen und Diskussionen Sektion 1) », dans Joachim Hamm, Albrecht Hausmann, Elisabeth Lienert et Gabriel Viehhauser (dir.) : *Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik*, Oldenbourg 2022 (= BmE Themenheft 12), URL : https://doi.org/10.25619/BmE20223192 : « *Automatische Handschriftenerkennung wird selbst bei besttrainierten Modellen zu einer gewissen Fehlerrate in der Erkennung führen. Ohne nachträgliche manuelle*

Korrektur (post-processing) wird also kein philologisch einwandfreier Text entstehen ».

19 La façon de calculer la précision des modèles de reconnaissance manuscrite (HTR) est expliquée par les développeurs de Transkribus sur leur site : « En tant que “Référence”, choisissez une version de page qui a été correctement transcrite (vérité de terrain : transcription manuelle aussi fidèle que possible au texte original). Pour obtenir la valeur la plus significative, il serait préférable d'utiliser des pages d'un ensemble d'échantillons qui n'ont pas été utilisées dans l'entraînement et qui sont donc nouvelles pour le modèle. » (« As '*Reference*', *choose a page version, which was correctly transcribed (Ground Truth: manual transcription as close to the original text as possible). To get out the most significant value it would be best to use pages from a sample set which have not been used in the training and therefore are new to the model* »). URL : https://readcoop.eu/glossary/compute-accuracy/.

20 Voir *ibid.* : « En tant que “Hypothèse”, choisissez la version qui a été générée automatiquement avec un modèle HTR et sur laquelle vous aimeriez voir à quel point le résultat est satisfaisant » (« As '*Hypothesis*', *choose the version, which was automatically generated with an HTR-model and on which you would like to see, how good the result is* »).

21 Étant donné que cela ne concernait qu'un petit nombre de pages et de lignes de texte, l'effort n'a pas été fait pour extraire les pages concernées du fichier PDF généré, corriger leur orientation, et les réintégrer dans Transkribus. Cela aurait pris plus de temps que la simple correction des pages en question. Malheureusement, à notre connaissance, Transkribus ne propose pas actuellement de solution directe à ce problème dans le programme. La possibilité de rotation partielle de la vue du manuscrit ne semble pas avoir d'impact sur la reconnaissance de l'écriture manuscrite en elle-même et ne concerne malheureusement pas seulement une page. En redressant les pages scannées de travers, toutes les pages sont automatiquement redressées.

22 Elpida Perdiki, *op cit.* : « *python sytems implented with TensorFlow library, [..] Kraken, an OCR system for historical documents, and [...] Tesseract, the OCR engine developed by Google and mostly used in many projects* ». On peut utiliser Kraken, mentionné par Perdiki, soit via eScriptorium, où il est intégré, soit individuellement.

ABSTRACTS

Français

Le projet « Grand Tour digital » vise à numériser, explorer et visualiser des témoignages personnels de voyages éducatifs de l'époque moderne, mettant l'accent sur cinq textes principaux. Le projet vise à développer expérimentalement un processus éditorial novateur en utilisant Transkribus. Financé par la Deutsche Forschungsgemeinschaft (DFG) de 2022 à 2025, le projet utilise la plateforme Transkribus pour la transcription automatique de manuscrits, combinée à la reconnaissance d'entités nommées (REN) pour l'exploration et la visualisation des témoignages. Le corpus comprend 21 journaux de voyage, rédigés entre 1550 et 1770, conservés à la Bibliothèque Herzog August à Wolfenbüttel (HAB), avec une variété de voyages couvrant l'Europe, l'Empire ottoman et le Proche-Orient. La reconnaissance automatique des récits de voyage réalisée avec Transkribus fait face à des défis tels que la variété de l'écriture, l'individualité de chaque écriture, la mise en page complexe, les polices variées, l'utilisation de différentes langues et les exigences de contenu. Au cours du projet sera développé un modèle adapté à ces exigences. Le logiciel fournit des instructions utiles pour l'édition de texte, et la recherche génère des rapports d'expérience, des discussions scientifiques et des conseils pratiques sur l'outil. La transcription manuelle des pages du journal de voyage de Wagener dans Transkribus a permis d'entraîner un nouveau modèle basé sur le modèle existant « Transkribus German handwriting M1 » de l'Université de Greifswald. Ce modèle a été ajusté progressivement en transcrivant partiellement automatiquement, en corrigeant, et en réentraînant. L'objectif était d'adapter le programme aux particularités de l'écriture de Wagener, réduisant ainsi le taux d'erreur dans la reconnaissance manuscrite. Des ajustements manuels ont été apportés au marquage des champs de texte pour minimiser les erreurs de détection. L'évaluation du modèle a montré des taux d'erreur fluctuants, atteignant 2,41 % dans l'ensemble d'entraînement et 11 % dans l'ensemble de validation lors de la première itération. L'optimisation du modèle a continué avec des ajustements itératifs, élargissant le jeu de caractères entraîné. Les résultats montrent des améliorations, bien que des questions subsistent sur la meilleure façon de traiter les pages « non utilisables ». L'évaluation automatique dans Transkribus s'est concentrée sur le taux d'erreur (CER) et la précision des mots et des caractères, révélant des écarts significatifs entre les évaluations automatiques et manuelles. Malgré des améliorations dans le modèle, Transkribus atteint ses limites, avec des échecs d'ajustement conduisant à des taux d'erreur élevés. L'étude soulève des questions sur la spécificité de ces limites à Transkribus par rapport à d'autres systèmes de transcription automatisée. L'auteure suggère également d'explorer d'autres outils d'HTR tels que Kraken et Tesseract

pour une comparaison approfondie. En conclusion, le succès de la reconnaissance automatisée dépend fortement de la qualité des pages, de la lisibilité du manuscrit, et de la présence de ratures. Bien que Transkribus ait joué un rôle pionnier, d'autres alternatives méritent une étude comparative pour évaluer les performances des différents outils d'HTR.

English
The "Grand Tour digital" project aims to digitize, explore, and visualize personal accounts of educational journeys from the modern era, focusing on five main texts. The project aims to experimentally develop an innovative editorial process using Transkribus. Funded by the Deutsche Forschungsgemeinschaft (DFG) from 2022 to 2025, the project utilizes the Transkribus platform for the automatic transcription of manuscripts, combined with Named Entity Recognition (NER) for the exploration and visualization of testimonies. The corpus consists of 21 travel journals written between 1550 and 1770, held at the Herzog August Library in Wolfenbüttel (HAB), covering various journeys across Europe, the Ottoman Empire, and the Middle East. Automatic recognition of travel narratives, especially with Transkribus, faces challenges such as writing variety, individuality of each script, complex layout, diverse fonts, use of different languages, and content requirements. A model adapted to these requirements will be developed during the project. The software provides useful instructions for text editing, and research generates experience reports, scholarly discussions, and practical advice on the tool. Manual transcription of Wagener's travel journal pages in Transkribus was used to train a new model based on the existing "Transkribus German handwriting M1" model from the University of Greifswald. This model was gradually adjusted by partially automatic transcription, correction, and retraining. The goal was to adapt the program to Wagener's writing characteristics, thus reducing the error rate in handwriting recognition. Manual adjustments were made to text field labeling to minimize detection errors. Model evaluation showed fluctuating error rates, reaching 2.41% in the training set and 11% in the validation set during the first iteration. Model optimization continued with iterative adjustments, expanding the trained character set. Results show improvements, although questions remain about the best way to handle "unusable" pages. Automatic evaluation in Transkribus focused on Character Error Rate (CER) and word and character accuracy, revealing significant discrepancies between automatic and manual evaluations. Despite improvements in the model, Transkribus reaches its limits, with adjustment failures leading to high error rates. The study raises questions about the specificity of these limits in Transkribus compared to other automated transcription systems. The author also suggests exploring other HTR tools such as Kraken and Tesseract for a comprehensive comparison. In conclusion, the success of automated recognition heavily depends on page quality, manuscript readability, and the presence of erasures. Although Transkribus has played a pioneering role, other alternatives deserve a comparative study to assess the performance of different HTR tools.

INDEX

Mots-clés
Transkribus, reconnaissance d'écriture automatique, HTR, Grand Tour, récits de voyage, édition en ligne, intelligence artificielle

Keywords
Transkribus, handwritten text recognition, HTR, Grand Tour, travelogues, online editions, artificial intelligence

AUTHOR

Angela Göbel
Université Jean Moulin Lyon 3, LARHRA UMR 5190 Herzog August Bibliothek Wolfenbüttel
IDREF : https://www.idref.fr/28172489X

La reconnaissance d'entités nommées dans les éditions numériques à l'exemple du récit de voyage du pharmacien Wagener

Recognition of named entities in digital editions: the example of pharmacist Wagener's travel account

Maximilian Görmar

Translated by Angela Göbel

DOI : 10.35562/theia.144

Copyright
CC BY-NC-SA 4.0

OUTLINE

Introduction
S
ituation de départ
Workflow et problèmes actuels
Perspectives
Conclusion

AUTHOR'S NOTES

Le présent texte est issu d'une conférence donnée lors de l'atelier « Ouverture des récits de voyage historiques : mise en forme du texte, modélisation des données et visualisation » (20-21 juillet 2023), organisé par le projet « Éditions numériques des récits de voyage historiques » (Digitale Editionen Historischer Reiseberichte, DEHisRe) à l'Institut Leibniz de recherche sur l'Europe de l'Est et du Sud-Est (IOS) à Ratisbonne. Mes remerciements les plus chaleureux vont aux organisatrices Anna Ananieva et Sandra Balck, ainsi qu'à toute l'équipe du projet, pour leur invitation à intervenir, ainsi qu'à tous les participants pour les discussions stimulantes et les échanges. À l'origine rédigé en allemand, le texte a été publié en premier lieu sous la forme d'un billet de blog dans le blog associé au projet « Grand Tour digital. Numérisation, exploitation et visualisation des témoignages autobiographiques des voyages éducatifs de l'époque moderne grâce à l'application de procédés éditoriaux partiellement automatisés » à la Bibliothèque Herzog August de Wolfenbüttel, (https://grandtourdig.hypotheses.org/949). Pour cette publication, le texte a été légèrement étendu. Je tiens à exprimer ma gratitude à Angela Göbel pour avoir accueilli et traduit l'article en français.

TEXT

Introduction

1 Les récits de voyage, en particulier ceux relatant des voyages éducatifs, constituent des témoignages importants qui peuvent fournir des informations sur les expériences et la vie de leurs auteurs. Ils documentent les rencontres avec de nouveaux lieux et personnes, les coutumes et les traditions, offrant ainsi un aperçu de l'appropriation mentale et matérielle ainsi que du traitement de l'étranger. Ils éclairent ainsi les processus de formation de soi et de la personnalité, souvent négligés jusqu'à présent par la recherche sur les témoignages personnels [2]. Particulièrement pour l'époque moderne, de vastes collections de cette catégorie de sources n'ont pas encore été suffisamment explorées ou analysées, telles que celles conservées à la Bibliothèque Herzog August de Wolfenbüttel [3]. Le projet d'édition « Grand Tour digital. Digitalisierung, Erschließung und Visualisierung frühneuzeitlicher Selbstzeugnisse von Bildungsreisen unter Anwendung teilautomatisierter Editionsverfahren » (Grand Tour digital. Numérisation, exploration et visualisation de témoignages personnels de voyages éducatifs de l'époque moderne en utilisant des procédés d'édition partiellement automatisés) part de ce problème fondamental et tente, comme le suggère déjà le titre complet, d'explorer également de nouvelles voies méthodologiques dans l'édition numérique en appliquant des procédures éditoriales partiellement automatisées [4]. Cela englobe d'une part l'utilisation de la reconnaissance automatique de l'écriture manuscrite (*Handwritten Text Recognition*, HTR) avec le programme Transkribus [5], et d'autre part, le traitement et la mise en évidence des textes identifiés avec des outils logiciels de reconnaissance des entités nommées (*Named Entity Recognition*, REN). L'étape suivante est ensuite de créer une liaison des entités nommées identifiées avec des données normalisées et d'autres sources de données. Pour ce deuxième aspect, la REN, les récits de voyage se prêtent particulièrement bien à l'étude de cas, car leur

densité en termes de noms de personnes et surtout de noms de lieux est particulièrement élevée pour alimenter une base de données.

Fig. 1 : Les voyages du pharmacien en formation Wagener, 1652–1659.



2 Ce type de travail sera montré à travers l'exemple d'un texte du projet « Grand Tour digital », rédigé par le jeune pharmacien Johannes Wagener entre 1652 et 1659 [6]. Wagener entreprit un voyage de compagnon (1652–1655), alors courant pour les futurs apothicaires à l'époque moderne [7], ainsi qu'un voyage de formation (1658/59) en compagnie du duc Ferdinand Albrecht I[er] de Brunswick-Wolfenbüttel-Bevern (Fig. 1) [8]. Le premier voyage le conduisit à travers le nord de l'Allemagne jusqu'à Königsberg en Prusse (aujourd'hui Kaliningrad), tandis que le second le mena à travers l'ouest et le sud de l'Allemagne, ainsi que la Suisse, jusqu'en France, où il décéda en 1659 à Angers des suites d'une fièvre ardente (« hitzigen Fieber ») [9]. Parallèlement, il entreprit également plusieurs voyages plus courts vers 1656/57, notamment au château de Gottorf, où il visita son célèbre cabinet de curiosités et put voir ainsi le globe géant qui s'y trouvait [10]. L'ampleur des informations géographiques, sociales et linguistiques du texte de Wagener ouvre de vastes possibilités, voire des défis, dans l'utilisation des méthodes de reconnaissance des entités nommées (REN) et dans le traitement des textes de voyage de

l'époque moderne. Celles-ci seront retracées ci-après en se basant sur les expériences et le flux de travail du projet « Grand Tour digital ».

Situation de départ

3 Tout d'abord, partons d'une définition générale : « La reconnaissance d'entités nommées (REN) désigne la détection des noms propres (entités nommées) dans les textes, ainsi que leur classification en différents types d'entités. […] par défaut, les personnes, les lieux et les organisations font partie de ces types d'entités »[11]. Il existe fondamentalement deux types de méthodes qui sont utilisées à cet effet : des approches basées sur des règles ou des connaissances, reposant par exemple sur des vocabulaires contrôlés ou des *gazetteers* (index géographique), et des approches basées sur l'apprentissage automatique. Ainsi, la REN peut être associée de manière similaire à la reconnaissance d'écriture manuscrite dans le vaste domaine de l'apprentissage automatique et de l'intelligence artificielle, auquel appartiennent également des outils tels que Chat-GPT, actuellement très populaires[12].

4 Contrairement aux grands modèles de langage (*Large Language Models*, LLM) avec des millions à des milliards de tokens et de paramètres, sur lesquels reposent Chat-GPT et d'autres outils de traitement du langage naturel pour les langues actuelles, les modèles correspondants et les ensembles de données disponibles pour les textes anciens sont beaucoup plus rares et plus petits.
Ainsi, Ehrmann *et al.*, dans un état des lieux récemment publié sur la REN dans de tels documents ont identifié 22 corpus annotés avec des données d'entraînement par rapport à 121 corpus pour les langues contemporaines[13]. Les données d'entraînement correspondent au corpus choisi pour étalonner la recherche et permettre ensuite au modèle de fonctionner. De plus, les modèles linguistiques pour les textes anciens, en raison de leur taille plus réduite et des défis spécifiques des stades linguistiques antérieurs sur lesquels nous reviendrons plus en détail, présentent généralement une fiabilité moindre (score F1 : 60-70 %) par rapport à ceux pour les langues actuelles (score F1 : généralement supérieur à

90 %)[14]. Bien entendu, même parmi eux, d'importantes différences subsistent en fonction des domaines spécifiques[15].

5 La situation en ce qui concerne la REN dans les textes anciens est donc certainement perfectible, mais il y a eu ces dernières années plusieurs initiatives et projets qui utilisent notamment des éditions numériques pour générer des ensembles de données d'entraînement correspondants. On peut citer, par exemple, le projet NERDPool en cours depuis 2020, porté par l'Académie autrichienne des sciences ainsi que les universités d'Innsbruck et de Graz, qui a rassemblé des ensembles de données d'entraînement pour la REN en allemand ancien à partir de six collections de textes ou éditions allant du XVI^e^ au début du XX^e^ siècle[16]. En outre, il existe des projets plus modestes, tels que ceux réalisés dans le cadre de mémoires de master ou de séminaires pratiques, où ont été élaborées des données d'entraînement grâce à de tels modèles et des tutoriels pour la REN dans les textes anciens[17]. Il s'agit donc d'un domaine de recherche extrêmement dynamique, particulièrement pour les sources du début de l'allemand contemporain et grâce aux récits de voyage[18].

Workflow et problèmes actuels

6 Dans ce contexte globalement positif, nous avons commencé à élaborer un travail pour la REN qui fait suite à l'établissement du texte intégral avec Transkribus. Tout a commencé par le choix d'un outil approprié pour la REN, avec l'évaluation de trois candidats au total : Stanford Named Entity Recognizer[19], WebLicht[20] et la bibliothèque Python appelé spaCy[21]. Les trois outils sont basés sur des systèmes techniques spécifiques, offrant chacun des avantages et des inconvénients ainsi que des approches différentes pour relever les défis des textes de cette époque et de leur édition numérique. Avec WebLicht, une sorte de boîte à outils en ligne pour le traitement du langage naturel (NLP), il est par exemple possible de compenser la variation linguistique des textes anciens par rapport à l'allemand moderne en normalisant les textes avant la REN à l'aide d'un outil spécialisé[22]. Une autre approche pour adapter le logiciel au matériel textuel ancien et à ses particularités consiste à former ses propres modèles de REN à l'aide de l'apprentissage automatique. Cette méthode est possible à la fois avec le Stanford Named Entity

Recognizer et avec spaCy. Cependant, avec le Stanford Named Entity Recognizer, ainsi qu'avec WebLicht, il est relativement complexe et fastidieux de traiter les données de transcription disponibles en TEI-XML [23], notamment pour que les annotations déjà effectuées restent telles quelles et que le résultat en sortie soit à nouveau en XML et conforme à la norme TEI de manière à inclure les entités reconnues et annotées par le programme. C'est pourquoi le choix s'est finalement porté sur spaCy, car grâce à cet outil il est non seulement relativement facile et rapide de mettre en œuvre un flux de travail pour la REN, mais aussi parce que la plupart des projets ayant travaillé sur l'application de la REN aux éditions numériques ont utilisé spaCy [24]. Les solutions aux problèmes déjà développées et documentées, sur lesquelles nous reviendrons plus tard, peuvent être réutilisées, ce qui non seulement représente un gain de temps considérable, mais est également conforme à une pratique de recherche durable dans le domaine de l'édition numérique.

7 SpaCy est une bibliothèque Python open source dédiée au traitement du langage naturel (NLP). En plus d'une gamme complète de scénarios d'utilisation en linguistique informatique (comme la tokenisation, la segmentation de phrases, le marquage grammatical), la bibliothèque comprend également des composants pour la REN et offre la possibilité de former des modèles personnalisés à cet effet. Parallèlement, ces modèles basés sur l'apprentissage automatique peuvent être combinés avec des listes de mots et des motifs de reconnaissance d'entités, intégrant ainsi une approche basée sur la connaissance ou sur des règles. Ses nombreuses fonctionnalités et sa facilité d'utilisation relative ont conduit, comme mentionné précédemment, à son utilisation dans plusieurs projets liés aux éditions numériques pour la NER. De là sont également nés quelques packages Python pour l'extraction de données à partir de textes annotés en TEI et pour l'intégration des résultats d'analyse, notamment l'acdh-spacytei [25] et le Standoff Converter [26].

8 Avant même de choisir un outil de REN, la décision avait été prise de sélectionner un texte pour les essais et l'entraînement. Par souci de simplicité, les premières pages du compte rendu de voyage de Wagener ont été choisies, environ 20 pages qui devaient de toute façon être transcrites manuellement pour l'entraînement de Transkribus [27]. Ainsi, sans trop d'effort supplémentaire, il a été

possible d'annoter les noms propres et les entités, formant ainsi la base de l'ensemble d'entraînement. Au total, 248 entités ont été étiquetées, dont 92 personnes, 148 lieux et 8 organismes. Il s'agit donc d'un ensemble d'entraînement relativement petit (en particulier pour les organismes), mais il semblait suffisant pour une première tentative et pour définir un modèle. Comme l'objectif initial était d'annoter le reste du compte rendu de voyage de Wagener avec cet ensemble, nous pouvions supposer que le matériel d'entraînement et le matériel cible présentaient une similarité suffisante pour obtenir des résultats acceptables.

9 Avant l'entraînement proprement dit, le texte devait être préparé d'une manière à pouvoir être traité. Ce fut le premier défi, car il s'est avéré initialement moins évident de convertir le texte TEI en un format JSON adapté à spaCy. En particulier, l'extraction de la position exacte des entités nommées s'est révélée initialement sujette à des erreurs en utilisant un script xslt, nécessitant ainsi un contrôle minutieux et chronophage [28]. Cependant, ce problème a pu être largement résolu en réutilisant des solutions et des packages de programmes existants provenant d'autres projets, comme mentionné précédemment. Après l'élaboration du flux de travail à partir d'un tutoriel d'Isabel Hansen issu de son mémoire de master rédigé à l'Université de Trèves, il a été assez simple de travailler avec le package Python acdh-spacytei [29], pour mettre le texte TEI dans la forme nécessaire pour l'ensemble d'entraînement. Ensuite, les données ont dû être divisées en ensembles d'entraînement, de validation et de test [30].

10 Ensuite, l'entraînement réel a eu lieu, ce qui peut parfois prendre beaucoup de temps. Au cours de plusieurs itérations, un réseau neuronal a été formé de manière à pouvoir prédire, en quelque sorte, quel mot ou quelle chaîne de caractères pouvait devenir une entité nommée. Cela se fait pratiquement par essais et erreurs, en attribuant d'abord aléatoirement des étiquettes aux entités nommées et en les comparant avec les annotations dans les données d'entraînement. Les erreurs qui surviennent sont mesurées et leur poids dans le réseau neuronal est ajusté en conséquence. Cela se poursuit jusqu'à ce qu'il n'y ait plus d'amélioration significative du modèle [31]. Les principales métriques utilisées pour évaluer ce processus sont la précision, le *recall* (rappel) et le score F1 calculé à

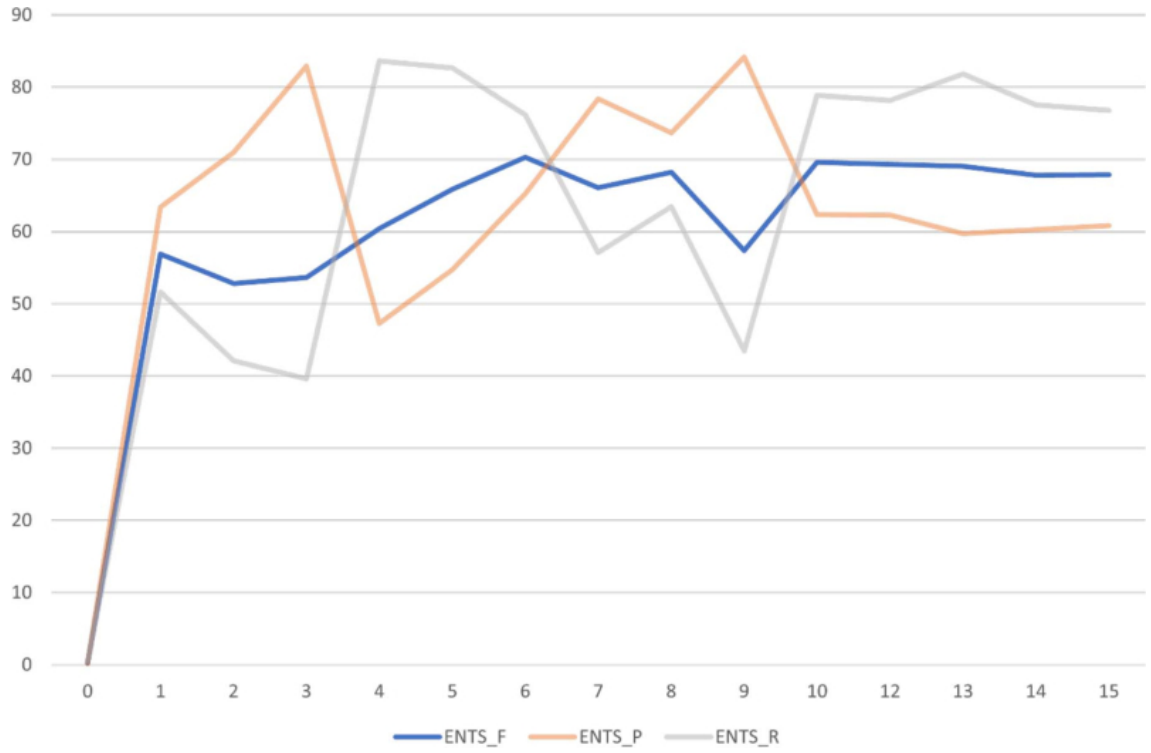
partir de ceux-ci [32]. Un aperçu de l'évolution de ces métriques au cours de l'entraînement est facilement compréhensible dans spaCy grâce à une présentation sur un tableau simple. Pour une meilleure visualisation, le déroulement de l'entraînement peut également être représenté sous forme de diagramme, montrant comment le score F1 se stabilise progressivement autour de 70 (Fig. 2 et 3). Bien que cette valeur semble relativement basse à première vue, on ne peut pas attendre trop du modèle formé. Cependant, il se situe dans la fourchette de valeurs couramment rencontrées dans la littérature pour des ensembles de textes anciens. En tenant compte de la petite taille de l'ensemble d'entraînement, la performance est en elle-même relativement bonne.

Fig. 2 : Vue d'ensemble de spaCy dans l'entraînement d'un modèle REN. Représentation simplifiée dans un tableau.

```
✓ Initialized pipeline

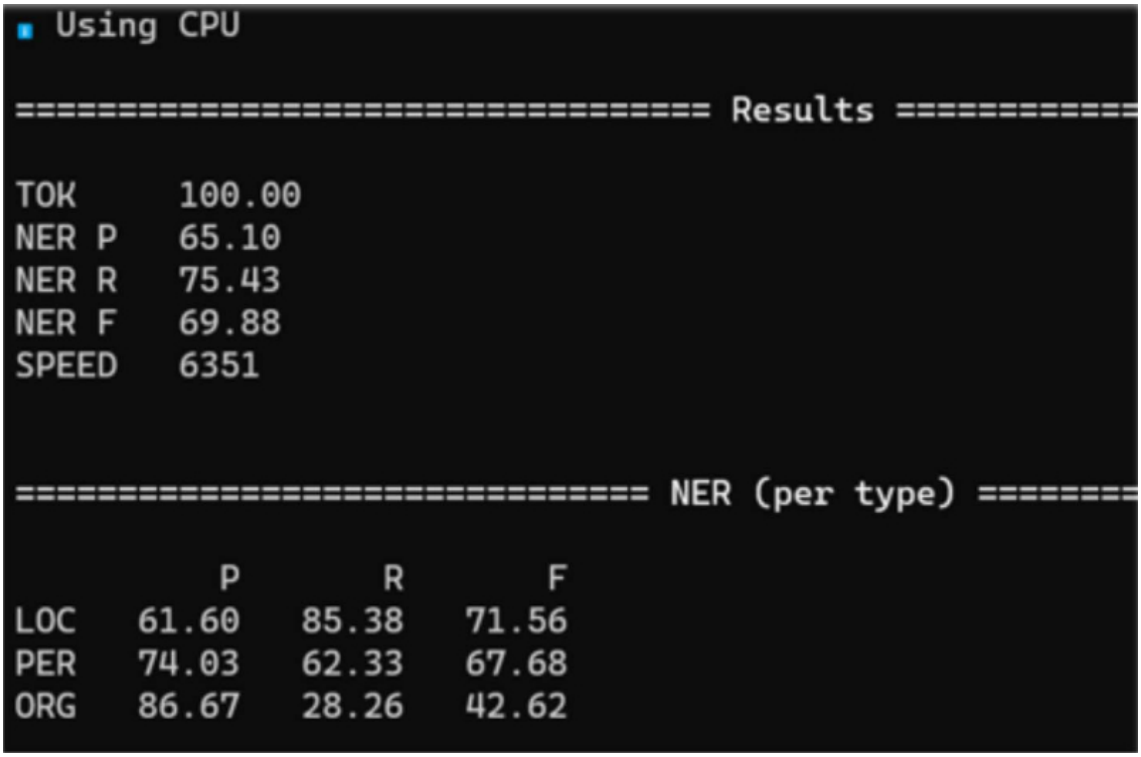
============================= Training pipeline =============================
ℹ Pipeline: ['tok2vec', 'ner']
ℹ Initial learn rate: 0.001
E    #       LOSS TOK2VEC  LOSS NER  ENTS_F  ENTS_P  ENTS_R  SCORE
---  ------  ------------  --------  ------  ------  ------  ------
  0       0          0.00    109.71    0.25    0.16    0.53    0.00
  1     200       1061.15   4155.34   56.90   63.41   51.61    0.57
  2     400         35.32   1680.59   52.84   70.96   42.09    0.53
  3     600         26.82   1498.22   53.63   82.91   39.64    0.54
  4     800        449.09   1410.32   60.40   47.28   83.60    0.60
  5    1000         34.84   1410.74   65.86   54.74   82.66    0.66
  6    1200         38.09   1273.08   70.27   65.22   76.18    0.70
  7    1400         44.37   1362.77   66.06   78.37   57.09    0.66
  8    1600         47.18   1360.55   68.19   73.69   63.46    0.68
  9    1800        200.29   1371.29   57.35   84.18   43.49    0.57
 10    2000         55.71   1265.83   69.62   62.32   78.87    0.70
 12    2200         63.78   1284.95   69.34   62.31   78.17    0.69
 13    2400         88.32   1200.16   69.06   59.74   81.84    0.69
 14    2600         88.65   1210.34   67.82   60.28   77.52    0.68
 15    2800         88.96   1217.98   67.89   60.85   76.77    0.68
✓ Saved pipeline to output directory
```

Fig. 3 : Visualisation des critères de qualité recall (ENTS_R), précision (ENTS_P) et f-score (ENTS_F) à partir du tableau ci-dessus au cours des itérations d'entraînement.



11 Une évaluation plus différenciée de la qualité du modèle peut se faire à l'aide des données du test. Il s'agit d'une partie des données annotées qui n'ont pas fait partie de l'ensemble d'entraînement. Avec elles, la précision, le *recall* et le score *F1* peuvent être calculés pour un texte inconnu, permettant ainsi d'évaluer le modèle. Là encore, spaCy fournit les valeurs correspondantes, cette fois-ci détaillées par type d'entité. Elles peuvent ainsi être facilement comparées entre elles (Fig. 4). Pour les personnes et les lieux, les valeurs sont d'environ 70, donc dans la fourchette du modèle global. On observe que la reconnaissance des lieux fonctionne légèrement mieux. En revanche, les organismes obtiennent des résultats nettement moins bons. Cela est lié à leur nombre nettement inférieur dans les données d'entraînement, et même pour les lieux et les personnes, les différences peuvent probablement être attribuées à leur fréquence respective dans l'ensemble d'entraînement.

Fig. 4 : Vue d'ensemble de spaCy des résultats de validation du modèle REN entraîné.



12 Après l'entraînement et l'évaluation, le modèle peut être appliqué au texte non annoté. Un extrait du résultat est visible dans la Fig. 5, qui montre la vue HTML générée avec displaCy, un module correspondant de spaCy [33]. Nous pouvons utiliser un simple fichier texte comme base pour cela, et le fichier HTML pourrait également être utilisé pour créer une version TEI rudimentaire du texte. Cependant, comme une partie de la balise d'origine serait perdue, le Standoff Converter de David Lassner a été utilisé [34]. Avec le convertisseur, le texte brut peut être extrait du XML, traité avec spaCy, et les nouvelles annotations (par exemple, les entités nommées) peuvent être insérées dans le document source sans que le balisage d'origine soit perdu. Le résultat est visible dans la Fig. 6.

Fig. 5 : Sortie HTML du résultat des REN (conçu avec displaCy).

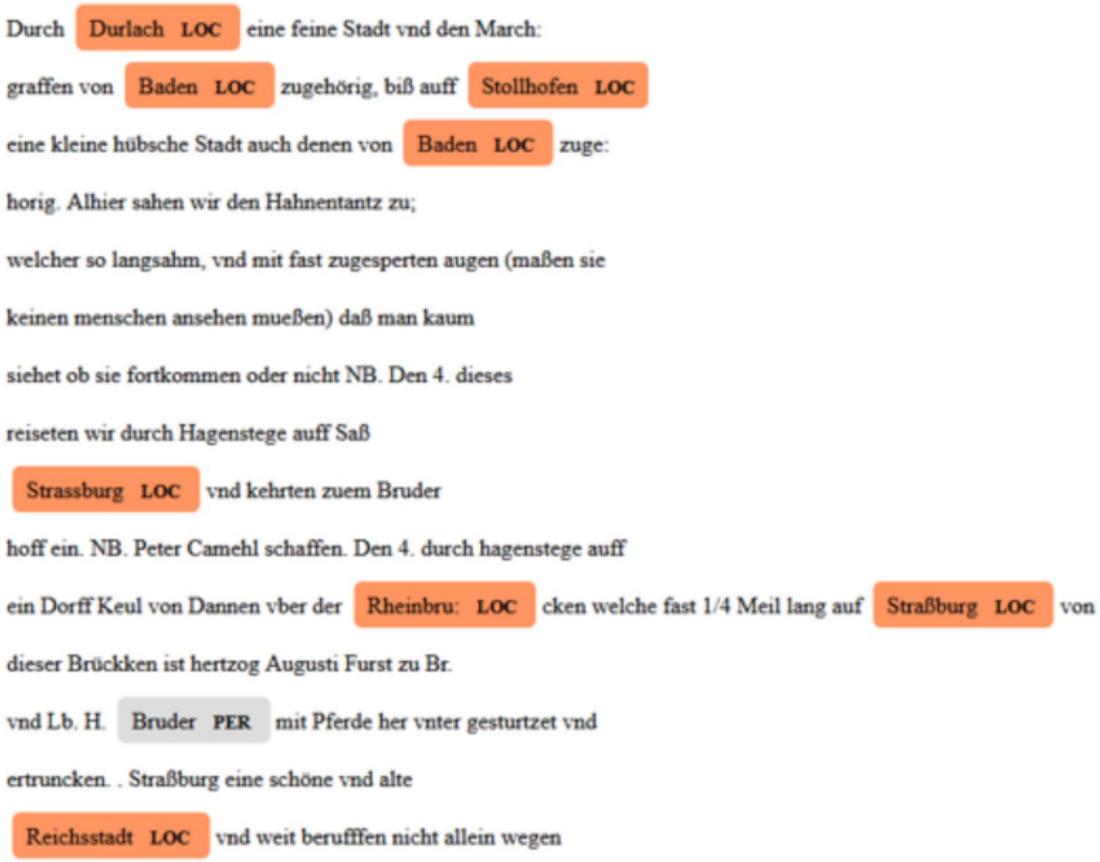


Fig. 6 : Sortie TEI du résultat des REN (conçu avec le Standoff-Converter).

```
</ab><pb facs="#facs_60" n="60" xml:id="img_0060"/><ab facs="#facs_60_r1">
   <lb facs="#facs_60_r111" n="N001"/>Die stadt <rs type="place">Franckfurt</rs> Jst eine sch&#246;ne handel stadt
   <lb facs="#facs_60_r112" n="N002"/>Ma&#223;en alle Jahr 2 gro&#223;e Me&#223;en allda gehalten werd[en]
   <lb facs="#facs_60_r113" n="N003"/>da dan viele 1000 Menschen zu handeln dahin komen
   <lb facs="#facs_60_r114" n="N004"/>der Mayn flie&#223;et mitten durch die stadt. Jst auch
   <lb facs="#facs_60_r115" n="N005"/>sonderlich beruemhbt wegen der Kayserlichen Wahl
   <lb facs="#facs_60_r116" n="N006"/>dadan sonderliche hauser vnd h&#246;ffe vor Jeden
   <lb facs="#facs_60_r117" n="N007"/>Churf&#252;rsten allezeit sein, darauff in Jhrer ankunfft
   <lb facs="#facs_60_r118" n="N008"/>zu wohnen. E&#223; hat 5 feine apothecken. Der elteste
   <lb facs="#facs_60_r119" n="N009"/>Jm hirsch H. Jacob Holtzapfel im Einhorn H. Sparr
   <lb facs="#facs_60_r1110" n="N010"/>im Kopff H. Banse im Schwan H. Saltzwedel
   <lb facs="#facs_60_r1111" n="N011"/>vnd im Engel H. Persebecher.
   <lb facs="#facs_60_r1112" n="N012"/>Oben an der Maynbrukke am thor siehet man
   <lb facs="#facs_60_r1113" n="N013"/>3 K&#246;pffe, derer, so vor diesen die stadt haben
   <lb facs="#facs_60_r1114" n="N014"/>verrathen wollen.
   <lb facs="#facs_60_r1115" n="N015"/>Den 29 Julij styl vet. haben Jhr F. G. <hi>He&#223;en</hi>
   <lb facs="#facs_60_r1116" n="N016"/>Ferdinand Albrecht hertzog zu Br. vnd <rs type="place">Luneburg</rs> mir
   <lb facs="#facs_60_r1117" n="N017"/>die gro&#223;e Gnade erwiesen, vnd von <rs type="place">Franckfurt</rs>
   <lb facs="#facs_60_r1118" n="N018"/>ab mit auf Jhre reise genommen, da wir dan
   <lb facs="#facs_60_r1119" n="N019"/>erwehnten dito auff H&#246;gst, ein klein stedlin
   <lb facs="#facs_60_r1120" n="N020"/>am Mayn lieget gereiset alhier ist ein altes zer
   <lb facs="#facs_60_r1121" n="N021"/>storetes schlo&#223;, den 30 dito haben wir vns da:
   <lb facs="#facs_60_r1122" n="N022"/>selbsten vber setzen la&#223;en vnd seind gefahren auf
   <lb facs="#facs_60_r1123" n="N023"/>Darmstadt worinnen ein schlo&#223; vnd F&#252;rstl.
   <lb facs="#facs_60_r1124" n="N024"/>Regierung. Alhier seind 3 apothecken, von hier
   <lb facs="#facs_60_r1125" n="N025"/>fuhren wir auf Everstadt ein Marckflecken
```

13 Ainsi, on distingue ici un procès rudimentaire pour la REN dans les documents TEI, et le résultat est réutilisable et peut être appliqué aux textes que nous traiterons encore dans le projet. Cependant, il reste quelques problèmes ouverts et du travail à faire. Le plus grand problème est certainement que le taux de reconnaissance avec le modèle développé jusqu'à présent peut encore être amélioré. Cela est dû à la fois à la petite taille de l'ensemble d'entraînement, mais aussi aux particularités linguistiques des sources. Ils posent un défi aux outils de traitement automatique du langage naturel (NLP) qui sont conçus pour des langues contemporaines relativement normalisées. En effet, la langue au XVII^{e} et XVIII^{e} siècle présente de grandes variations orthographiques, des différences de vocabulaire, de morphologie et de structure syntaxique par rapport à l'allemand actuel. On trouve également dans les textes anciens un multilinguisme récurrent [35]. Quelques mots en latin, français, italien ou d'autres langues étaient généralement intégrés sans distinction typographique. Souvent, des passages entiers en une ou plusieurs langues étrangères étaient également intégrés dans un texte principalement allemand. Il s'agissait fréquemment de citations étrangères utilisées par les voyageurs qui se formaient à l'apprentissage des langues étrangères [36]. Surtout, les noms propres étaient régulièrement traduits dans d'autres langues, sans parler du fait qu'il n'y avait pas de conventions uniformes pour leur écriture, y compris chez un même auteur. Par ailleurs, l'utilisation de majuscules n'était pas appliquée de manière cohérente à cette époque [37].

14 Cette incertitude sur les noms propres entraîne alors les deux autres problèmes : l'identification et la désambiguïsation des entités, ainsi que leur liaison avec des données normatives telles que Geonames [38], GND [39], VIAF [40] ou encore Wikidata [41]. Pour ces raisons, on recueille souvent, en plus d'une entrée principale, des formes de noms historiques alternatives ou des formes dans d'autres langues. Une source de ces formes de noms peut être des éditions (numériques), ce qui nous amène à notre troisième point, « Perspectives ».

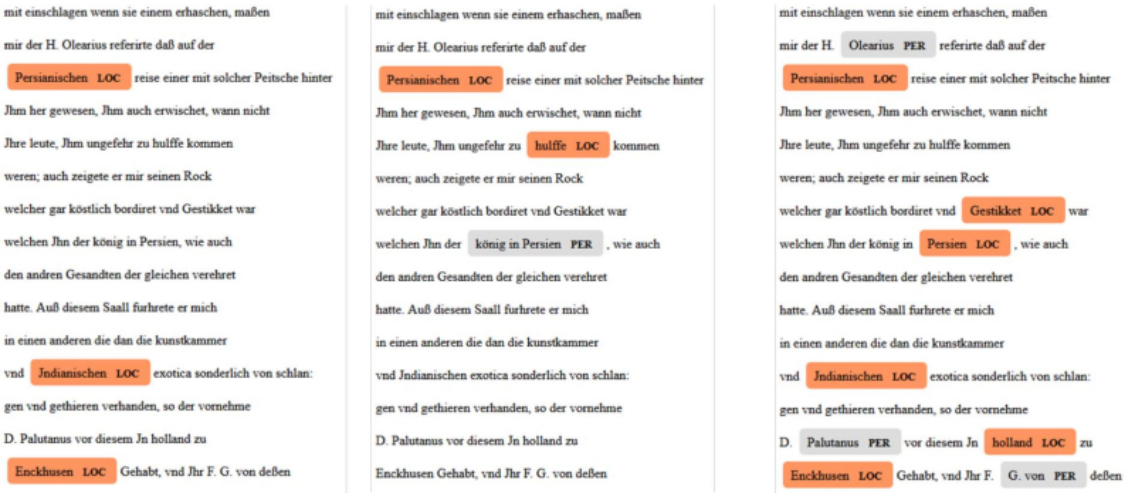
Perspectives

15 Quelles sont les solutions possibles pour les problèmes mentionnés ? Nous voyons surtout deux approches : l'élargissement de la base de

données pour l'entraînement de modèles linguistiques historiques et la combinaison de ces modèles basés sur des algorithmes d'apprentissage automatique avec des approches basées sur des règles ou des connaissances. Pour cette dernière approche, il est nécessaire de développer des bases de données ou des *gazetteers*, comme c'est déjà partiellement le cas. Ils peuvent et doivent certainement être étendus et complétés dans certains domaines, ce qui justifie des éditions numériques annotées. Dans les lignes qui suivent, nous allons brièvement présenter nos expériences faites jusqu'à présent dans ce domaine du projet et esquisser les perspectives qui en découlent.

16 L'entraînement de la REN peut être adapté à l'élargissement de la base de données d'autant plus qu'il existe déjà des collections de données similaires qui peuvent potentiellement être réutilisées. Un exemple concret serait le projet NERDPool mentionné précédemment [42]. Cependant, les tentatives menées jusqu'à présent pour entraîner des modèles plus importants avec des éditions de la Herzog August Bibliothek n'ont conduit qu'à des résultats légèrement meilleurs pour la REN. En plus des problèmes déjà formulés inhérents aux sources, des défis supplémentaires sont apparus en raison d'approches éditoriales différentes en fonction des sources écrites [43]. Ces défis concernent les différentes pratiques de balisage par rapport aux entités nommées, les différents niveaux de normalisation du langage source et les différences en ce qui concerne la segmentation des textes, ce qui peut être intégré en tant qu'information contextuelle dans le modèle entraîné [44]. Cela rend nécessaire une adaptation partiellement manuelle des textes et de leur codage, ce qui représente une charge de travail non négligeable. Par conséquent, il semble judicieux de ne pas simplement intégrer de nombreux textes dans un ensemble d'entraînement, mais plutôt de créer des échantillons représentatifs et spécifiques au domaine (par exemple, des récits de voyage) codés selon des normes uniformes. De plus, il peut être utile de soumettre les données textuelles à d'autres opérations de traitement automatique du langage naturel avant l'entraînement, telles que la division en phrases individuelles. Cela peut également réduire le temps de calcul nécessaire pour créer un modèle.

Fig. 7 : Sortie HTML pour différents modèles NER avec et sans découpage de phrases.

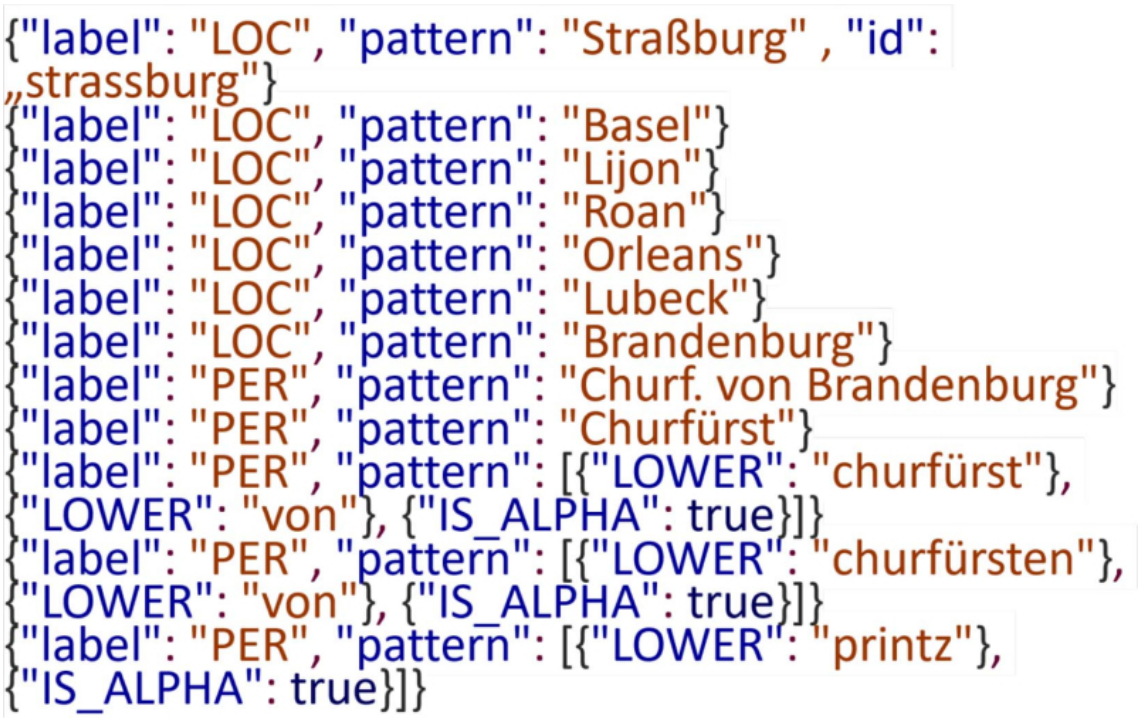


17 Un contrepoint comparatif des résultats obtenus jusqu'à présent, qui intègre déjà en partie ces considérations, est visualisé dans la Fig. 7 avec displaCy. À gauche on trouve le modèle de base. Au milieu, c'est le modèle qui entraîne le texte de Wagener combiné avec des journaux du duc Auguste le Jeune de Brunswick-Wolfenbüttel [45] et du prince Christian II d'Anhalt-Bernburg [46]. À droite, c'est un modèle basé sur Wagener et où le texte a de nouveau été divisé en phrases avec spaCy avant l'entraînement (ce qui repose bien sûr sur des modèles linguistiques pour des langues actuelles et est donc sujet aux erreurs correspondantes) [47].

18 L'intégration de *gazetteers* et de modèles de reconnaissance pour la REN est relativement facile à faire dans spaCy. Il suffit de mettre les données correspondantes, par exemple les formes de noms de lieux, dans un format JSON simplifié basé sur des lignes et de les intégrer dans le *pipeline* de traitement. Théoriquement, il est possible de construire un flux de travail REN uniquement à l'aide d'une telle liste [48]. Il serait alors possible de comprendre et de contrôler ce qui est reconnu comme entité et ce qui ne l'est pas. Ce n'est pas le cas avec les réseaux neuronaux auto-apprenants. Ils s'apparentent à une boîte noire à l'intérieur de laquelle même les développeurs ne peuvent pas regarder ou de manière très limitée. Il est aussi possible de réutiliser les ressources existantes telles que les éditions numériques. Et il est possible d'attribuer des identifiants, ce qui

permet la liaison avec des registres et des données normatives déjà existants (Fig. 8). Cela permettrait également d'automatiser dans une certaine mesure le lien avec les entités nommées, qui doit suivre la REN lors de la création de l'édition numérique [49].

Fig. 8 : extrait d'un fichier JSONL avec les modèles et notions pour la REN.



19 Les différentes formes de noms induisent l'essentiel des inconvénients. Dans la mesure où les personnes n'apparaissent souvent que dans des contextes spécifiques et où la variabilité des noms est encore plus élevée, il semble judicieux d'interroger des termes fixes surtout pour les lieux et les organismes, qui sont plus ou moins des entités de longue durée. Mises à part quelques personnalités importantes (Jésus-Christ, Luther, Calvin, Charlemagne etc.), il est plus judicieux pour les personnes de définir des modèles plus généraux, comme avec « Kurfürst von… » (prince-électeur de…). Cependant, pour les membres de la haute noblesse ou les souverains, le problème des entités dites « imbriquées » se pose souvent, où une partie du nom peut être simultanément un toponyme [50]. De plus, les indications de lieu peuvent être relatives à l'emplacement de l'écrivain et donc être fortement dépendantes du contexte, de sorte que des mots tels que « ici », « en cet endroit », « là-bas » peuvent déformer

la REN s'ils ne sont pas éliminés de la base de données. Il en va de même pour les pronoms personnels, qui peuvent également faire référence de manière contextuelle aux personnes. Malgré ces inconvénients, l'utilisation de *gazetteers* ou de listes de données peut améliorer de manière décisive la REN, en particulier lorsque peu de données d'entraînement sont disponibles.

Conclusion

20 Il est devenu évident que la REN est une composante exigeante mais judicieuse de la boîte à outils offerte par le traitement automatique du langage naturel pour l'élaboration d'éditions numériques. Cependant, il est important d'éviter la fausse allégation répandue selon laquelle « Big Data supposedly lets you get away with dirty data »[51]. Bien que nous ayons besoin de plus de données pour l'entraînement des modèles linguistiques anciens, notamment ceux de la REN, cela ne signifie pas nécessairement l'usage de *big data*. Des échantillons judicieusement choisis et soigneusement organisés devraient conduire à de bons résultats. Cela ne rend pas l'éditeur et le chercheur humains superflus, mais cela leur apportera un soutien significatif dans un avenir proche.

21 En outre, les problèmes liés à la REN ne doivent pas être considérés de manière isolée. Ils sont notamment liés à la modélisation des données et à l'enrichissement sémantique des éditions numériques, en particulier dans le domaine des récits de voyages anciens. Les défis mentionnés précédemment lors de la conversion des données textuelles et d'annotations vers et depuis le TEI-XML illustrent cette complexité. Ces problématiques prennent une place encore plus importante dans le cadre du développement d'ontologies visant à modéliser les relations entre les entités présentes dans les récits de voyages, ainsi que dans la création de jeux de données *Linked Open Data* réutilisables à partir d'éditions numériques. Ces efforts sont actuellement au cœur du projet de Ratisbonne, « Digitale Editionen Historischer Reiseberichte[52] » (DEHisRe, Édition numérique des récits de voyage historiques), et démontrent que des travaux de recherche sur la REN dans les récits de voyages sont essentiels pour la création d'éditions numériques fiables. Ces efforts soulignent la pertinence de la reconnaissance des entités nommées

(REN) dans le domaine de l'édition numérique. Il est évident que la question de la REN dans les récits de voyages historiques constitue actuellement un domaine de recherche très dynamique.

NOTES

1 Terme informatique désignant une unité de valeur numérique, un objet identifié numériquement.

2 Voir en introduction, et avec une bibliographie complémentaire Mathis Leibetseder, « Kavalierstour – Bildungsreise – Grand Tour : Reisen, Bildung und Wissenserwerb in der Frühen Neuzeit », dans Leibniz-Institut für Europäische Geschichte (IEG) (dir.), *Europäische Geschichte Online* (*EGO*), mis en ligne le 14 août 2013, consulté le 15 janvier 2024, (http://www.ieg-ego.eu/leibetsederm-2013-de).

3 Voir le projet d'édition en ligne : Inga Hanna Ralle (dir.), *Selbstzeugnisse der Frühen Neuzeit in der Herzog August Bibliothek. Digitales Selbstzeugnis-Repertorium*, Wolfenbüttel, Herzog August Bibliothek, mis en ligne en 2014-2017, consulté le 15 janvier 2024, (http://selbstzeugnisse.hab.de/repertorium/). Pour une présentation du projet, voir Inga Hanna Ralle (dir.) *Selbstzeugnisse der Frühen Neuzeit in der Herzog August Bibliothek : Digitale Edition des Diariums von Herzog August dem Jüngeren. Technische Konzeption und Begleitung durch David Maus, unter Mitarbeit von Jacqueline Krone*, Wolfenbüttel, Wolfenbütteler Digitale Editionen 1, 2015–2017, (https://hainhofer.hab.de/informationen-zur-edition/bibliographie/ralle_diarium_2015-2017).

4 Voir la présentation en cours de construction des résultats du projet, notamment des éditions en voie d'élaboration, sur le portail des témoignages autobiographiques de la Bibliothèque Herzog August à l'adresse suivante : http://selbstzeugnisse.hab.de/edition_grand_tour ; ainsi que le blog accompagnant le projet à l'adresse suivante : https://grandtourdig.hypotheses.org/.

5 Voir la contribution d'Angela Göbel dans ce dossier, de même que Tobias Hodel, « Konsequenzen der Handschriftenerkennung und des maschinellen Lernens für die Geschichtswissenschaft. Anwendung, Einordnung und Methodenkritik », *Historische Zeitschrift*, 2023, n° 316, p. 151–180.

6 HAB Wolfenbüttel, Cod. Guelf. 267.1. Voir Wolf-Dieter Otte, *Die neueren Handschriften der Gruppe Extravagantes*, vol. 3, Francfort-sur-le-Main, Klostermann, 1993, p. 169 ; I. H. Ralle, *op. cit.*, voir ici l'entrée à propos du journal de voyage de Wagener, http://selbstzeugnisse.hab.de/repertorium/eintrag/267-1-extrav.

7 Voir Werner Gaude, *Die alte Apotheke. Eine tausendjährige Kulturgeschichte*, 3e éd., Leipzig, Koehler & Amelang, 1985, p. 25.

8 Voir Jill Bepler, *Ferdinand Albrecht, Duke of Braunschweig-Lüneburg (1636-1687). A Traveller and his Travelogue*, Wiesbaden, Harassowitz, 1988.

9 *Ibid.*, p. 293.

10 Voir Angela Göbel, Maximilian Görmar, « Zu Besuch in der Gottorfischen Kunstkammer. Berichte eines Apothekergesellen auf Reisen », *HABlog*, mis en ligne le 28 mars 2023, consulté le 15 janvier 2024, (https://www.hab.de/zu-besuch-in-der-gottorfischen-kunstkammer/).

11 Elisabeth Eder, « Named Entity Recognition (NER) », dans Helmut W. Klug (dir.), *KONDE Weißbuch*, mis en ligne en 2021, (https://gams.uni-graz.at/o :konde.141) : « Named Entity Recognition (NER) bezeichnet die Erkennung von Eigennamen (named entities) in Texten sowie auch deren Klassifizierung in verschiedene Entitätstypen. […] Standardmäßig zählen Personen, Orte und Organisationen zu diesen Entitätstypen ».

12 Voir de manière introductive Manuela Lenzen, *Künstliche Intelligenz. Fakten, Chancen, Risiken*, München, Beck, 2020.

13 Maud Ehrmann, Ahmed Hamdi, Elvys Linhares Pontes, Matteo Romanello, Antoine Doucet, « Named Entity Recognition and Classification on Historical Documents: A Survey », ACM *Computing Surveys*, n° 56/2, mis en ligne le 14 septembre 2023, p. 1–43, URL : https://doi.org/10.1145/3604931, ici p. 17.

14 *Ibid.*, p. 22.

15 Voir Mareike Schumacher, *Orte und Räume im Roman. Ein Beitrag zur digitalen Literaturwissenschaft*, Berlin, J.B. Metzler, 2023, p. 86–90. Le score F1 est calculé comme la moyenne harmonique des valeurs de la *precision* et du *recall*. La *precision* indique combien de *tokens* marqués comme noms propres ont été correctement reconnus. Le *recall* est la valeur représentant la proportion de *tokens* correctement identifiés par rapport à

l'ensemble des *tokens* appartenant à la catégorie d'entité respective, *ibid.*, p. 86.

16 *NERDPool : Data Pool for Named Entity Recognition*, consulté le 15 janvier 2024, (https://nerdpool-api.acdh-dev.oeaw.ac.at/). Voir également Peter Andorfer, Roman Bleier, Matthias Schlögl, « NERDPool. Datenpool für Named Entity Recognition », dans Michaela Geierhos (dir.), *DHd2022: Kulturen des digitalen Gedächtnisses. Konferenzabstracts*, Potsdam, 2022, p. 333, URL: https://doi.org/10.5281/zenodo.630459.

17 Voir Isabel Hansen, *Named Entity Recognition. Einsatz und Verwendung der Technologie in den Digital Humanities am Beispiel historischer Texte*, mémoire de Master (M. Sc.), Université de Trèves, 2023 (un grand merci à Madame Hansen et Joëlle Weis pour la mise à disposition du travail !) ainsi que le tutoriel associé sous URL : https://easyh.github.io/NerDH/, et le modèle sous URL : https://huggingface.co/easyh/de_fnhd_nerdh ; Jacqueline More, *Theorie und Anwendung von Named Entity Recognition in den Digital Humanities mit Fokus auf historische Texte des 17. Jahrhunderts*, mémoire de Master, Université de Graz, 2021, URL : https://resolver.obvsg.at/urn :nbn :at :at-ubg:1-166482 ainsi que le dépôt GitHub correspondant sur URL : https://github.com/jackymore/NER_historical_texts ; voir de plus Tatiana Bessonova, Lisa Braune, Sarah Ondraszek *et al.*, *Reiseberichte karTRIERt*, mis en ligne le 24 avril 2023, (https://kartriert.github.io/index.html) ; voir encore d'autres tutoriels : Susan Grunewald, Andrew Janco, « Finding Places in Text with the World Historical Gazetteer », Programming Historian, 2022, n° 11, (https://doi.org/10.46430/phen0096 ; Mareike Schumacher, « Named Entity Recognition (NER) », *forTEXT. Literatur digital erforschen*, mis en ligne le 18 mai 2018, (https ://fortext.net/routinen/methoden/named-entity-recognition-ner).

18 Parallèlement, des journaux de l'époque moderne et des dossiers criminels ont également été examinés dans des études de cas à l'aide de la REN. Voir Thomas Kirchmair, Nina C. Rastinger, Claudia Resch, « Die historische ‚Wiener Zeitung' und ihre Sterbelisten als Fundament einer Vienna Time Machine. Digitale Ansätze zur automatischen Identifikation von Toponymen », *Wiener Digitale Revue*, 2022, n° 4, (https://doi.org/10.25365/wdr-04-03-04) ; Tobias Hodel, Ismail Prada Ziegler, Christa Schneider, « Pre-Modern Data: Applying Language Modeling and Named Entity Recognition on Criminal Records in the City of Bern », dans Anne Baillot, Toma Tasovac, Walter Scholger, Georg Vogeler (dir.), *Digital*

Humanities 2023: Book of Abstracts, Graz, 2023, p. 384.,URL : https://doi.org/10.5281/zenodo.7961822 ; voir également le modèle NER correspondant, consulté le 15 janvier 2024, (https://huggingface.co/dh-unibe/turmbuecher-lm-v1).

19 *Stanford Named Entity Recognizer*, Version 4.2.0, mis en ligne le 17 novembre 2020, consulté le 15 janvier 2024, (https://nlp.stanford.edu/software/CRF-NER.shtml). Voir aussi Mareike Schumacher, « Stanford Named Entity Recognizer », *forTEXT. Literatur digital erforschen*, mis en ligne le 20 septembre 2018, (https://fortext.net/tools/tools/stanford-named-entity-recognizer) ; *Id.*, « Named Entity Recognition mit dem Stanford Named Entity Recognizer », *forTEXT. Literatur digital erforschen*, mis en ligne le 26 août 2019, (https://fortext.net/routinen/lerneinheiten/named-entity-recognition-mit-dem-stanford-named-entity-recognizer).

20 CLARIN-D/Seminar für Sprachwissenschaft, Université de Tübingen : WebLicht : *Web-Based Linguistic Chaining Tool*, mis en ligne en 2012, (https://weblicht.sfs.uni-tuebingen.de). Pour une documentation et une introduction à l'utilisation, voir le wiki correspondant, consulté le 15 janvier 2024, (https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page).

21 ExplosionAI GmbH, *spaCy*, mis en ligne en 2016-2024 (https://spacy.io/). Voir aussi Peter Andorfer, Matthias Schlögl, , Helmut W. Klug (dir.), *KONDE Weißbuch*, « spaCy », mis en ligne en 2021, (https://www.digitale-edition.at/o :konde.170).

22 Il s'agit de l'outil de Bryan Jurish, *DTA::CAB – "Cascaded Analysis Broker" for error-tolerant linguistic analysis*, (https://kaskade.dwds.de/~moocow/software/dta-cab/). Voir aussi *Id.*, *Finite-state Canonicalization Techniques for Historical German*, thèse de doctorat, Université de Potsdam, 2012, consulté le 15 janvier 2024, URL : https://publishup.uni-potsdam.de/files/5562/jurish_diss.pdf. Le *Cascaded Analysis Broker* est utilisé, entre autres, dans le German Text Archive. Voir « Software im Deutschen Textarchiv », dans Berlin-Brandenburgische Akademie der Wissenschaften (dir.), *Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache*, mis en ligne le 4 mars 2020, consulté le 15 janvier 2024, (https://www.deutschestextarchiv.de/doku/software).

23 TEI est le standard *de facto* pour la codification des textes dans le domaine des éditions numériques. Voir Text Encoding Initiative (dir.), *TEI: Guidelines for Electronic Text Encoding and Interchange. P5*, Version 4.7.0,

mis en ligne le 16 novembre 2023, consulté le 15 janvier 2024, (https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html).

24 Voir ci-dessus les notes 17 et 18.

25 Peter Andorfer, Matthias Schlögl, Saranya Balasubramanian, *acdh-spacytei*, Version 0.1.2, mis en ligne le 19 juillet 2019, (https://pypi.org/project/acdh-spacytei/). Voir aussi Peter Andorfer, Matthias Schlögl, Helmut W. Klug (dir.), *KONDE Weißbuch*, « acdh-spacytei », mis en ligne en 2021, consulté le 15 janvier 2024, (https://www.digitale-edition.at/o :konde.2).

26 David Lassner, *standoffconverter*, Version 0.8.11, mis en ligne le 28 avril 2022,(https://pypi.org/project/standoffconverter). Voir *Id.*, *Analysis of textual variants with robust machine learning methods: Towards novel insights for the digital humanities*, thèse de doctorat, TU Berlin, 2023. URL : https://doi.org/10.14279/depositonce-17717, p. 10–16 et 123–125.

27 Voir Angela Göbel, « Erste Schritte in Transkribus (1) – Modelltraining und Transkription der Handschrift Wagener », dans *Grand Tour digital. Ein Forschungsblog zum Editionsprojekt frühneuzeitlicher Selbstzeugnisse an der Herzog August Bibliothek Wolfenbüttel*, mis en ligne le 6 juillet 2023, (https://grandtourdig.hypotheses.org/290).

28 Ont été utilisés dans ce cadre Arunmozhi, *NER Text Annotator*, Version 1.3.0, mis en ligne le 12 mai 2023,(https://tecoholic.github.io/ner-annotator/).

29 P. Andorfer, M. Schlögl, S. Balasubramanian, *op. cit.*

30 Cela a été réalisé en préparant 70 % de l'ensemble de données comme données d'entraînement, 20 % comme données de validation et 10 % comme données d'essai.

31 Voir *Training Pipelines & Models*, (https://spacy.io/usage/training).

32 Pour expliquer ces indicateurs, voir la note 15 ci-dessus.

33 ExplosionAI GmbH, *displaCy Named Entity Visualizer*, (https://demos.explosion.ai/displacy-ent).

34 D. Lassner, *op. cit.*

35 Voir par exemple Joachim Peters, Sabrina Freund, « *Vormittags rien fait quj vaille.* Codeswitching im Fürstentagebuch Christians von Anhalt-Bernburg (1599–1656) », dans Elvira Glaser, Michael Prinz, Stefaniya Ptashnyk (dir.), *Historisches Codeswitching mit Deutsch.*

Multilinguale Sprachpraktiken in der Sprachgeschichte, Berlin/Boston, de Gruyter, 2021, p. 331–366.

36 Voir Arturo Tosi, *Language and the Grand Tour. Linguistic experiences of travelling in Early modern Europe*, Cambridge, Cambridge University Press, 2020.

37 Voir Johannes Volmert, « Geschichte der deutschen Sprache », dans *Id.* (dir.), *Grundkurs Sprachwissenschaft. Eine Einführung in die Sprachwissenschaft für Lehramtsstudiengänge*, 4e éd., Munich, W. Fink, 2000, p. 29–46, ici p. 42–48 ; de plus voir comme exemple Andreas Herz, « Zu den Schreibweisen Fürst Christians II. von Anhalt-Bernburg », dans Ronald G. Asch, Peter Burschel, Arndt Schreiber *et al.* (dir.), *Digitale Edition und Kommentierung der Tagebücher des Fürsten Christian II. von Anhalt-Bernburg (1599–1656). Christian II. von Anhalt-Bernburg*, Wolfenbüttel, Herzog August Bibliothek, 2013–2024, consulté le 15 janvier 2024, (http://diglib.hab.de/edoc/ed000228/id/edoc_ed000228_introduction_spelling_sm/start.htm).

38 https://www.geonames.org/.

39 https://gnd.network.

40 https://viaf.org/.

41 https://www.wikidata.org.

42 Voir la note 16 ci-dessus.

43 Outre le cadre fourni par la TEI, qui est déjà assez avancé et offre généralement plusieurs possibilités pour annoter le même phénomène textuel, de nombreuses éditions de la période de l'époque moderne s'orientent également d'après : Arbeitskreis Editionsprobleme der Frühen Neuzeit, « Empfehlungen zur Edition von frühneuzeitlichen Texten », *HEIMATFORSCHUNG-REGENSBURG.DE*, mis en ligne le 11 février 2016, (http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:355-rbh-280-9). Cependant, il existe généralement différentes déviations, de sorte que chaque édition établit fondamentalement ses propres directives de transcription et d'édition, qui ne sont pas nécessairement compatibles avec d'autres. Cette diversité, fondamentalement justifiée compte tenu de l'individualité des sources traitées et des axes de recherche possibles des projets d'édition respectifs, entraîne un niveau élevé de confusion et limite l'interopérabilité des données d'édition. Cependant, un degré élevé de différenciation et de spécialisation dans les directives d'édition et la confusion qui en résulte ne sont pas spécifiques aux éditions numériques,

mais concernent également depuis longtemps les éditions imprimées. Voir Bodo Plachta, *Editionswissenschaft. Eine Einführung in Methode und Praxis der Edition neuerer Texte*, édition complétée et actualisée, Stuttgart, Reclam, 2013 ; Hans-Gert Roloff, « Epochenprofilierung durch Editionen », dans *Id.* (dir.), *Editionsdesiderate der Frühen Neuzeit. Beiträge einer Tagung der Kommission für die Edition von Texten der Frühen Neuzeit*, vol. 1, Amsterdam/Atlanta, Rodopi, 1997, p. 1–13.

44 Pour la relation entre la segmentation du texte et l'analyse (numérique), voir Sabine Bartsch, Evelyn Gius, Marcus Müller, Andrea Rapp, Thomas Weitin, « Sinn und Segment. Wie die digitale Analysepraxis unsere Begriffe schärft », *Zeitschrift für digitale Geisteswissenschaften*, 2023, n° 8, (https://doi.org/10.17175/2023_003).

45 Inga Hanna Ralle, Jacqueline Krone, David Maus (dir.), *Herzog August der Jüngere von Braunschweig-Wolfenbüttel (1579–1666) : Ephemerides. Sive Diarium (1594–1635)*, Wolfenbüttel, Herzog August Bibliothek, 2017, (https://doi.org/10.15499/edoc/ed000225).

46 Ronald G. Asch, Peter Burschel, Arndt Schreiber *et al.* (dir.), *Digitale Edition und Kommentierung der Tagebücher des Fürsten Christian II. von Anhalt-Bernburg (1599-1656). Christian II. von Anhalt-Bernburg*, Wolfenbüttel, Herzog August Bibliothek, 2013–2024, (http://diglib.hab.de/edoc/ed000228/start.htm).

47 Pour différentes possibilités techniques de segmentation de phrases dans spaCy voir *Linguistic Features*, consulté le 15 janvier 2024, (https://spacy.io/usage/linguistic-features).

48 Voir *Rule-based matching*, (https://spacy.io/usage/rule-based-matching) ; ainsi que plus loin S. Grunewald, A. Janco, *op. cit.*

49 Voir Sina Menzel, Hannes Schnaitter, Josefine Zinck *et al.* « Named Entity Linking mit Wikidata und GND – Das Potenzial handkuratierter und strukturierter Datenquellen für die semantische Anreicherung von Volltexten », dans Michael Franke-Maier, Anna Kasprzik, Andreas Ledl, Hans Schürmann (dir.), *Qualität in der Inhaltserschließung*, Berlin/Boston, de Gruyter Saur, 2021, p. 229–257.

50 Voir Jenny Rose Finkel, Christopher D. Manning, « Nested Named Entity Recognition », dans Philipp Koehn, Rada Mihalcea (dir.), *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Singapour, Association for Computational Linguistics, 2009, URL: https://aclanthology.org/D09-1015, p. 141–150.

51 Tim Hitchcock, « Big Data, Small Data and Meaning », *Historyonics*, mis en ligne le 9 novembre 2014 (http://historyonics.blogspot.com/2014/11/big-data-small-data-and-meaning_9.html).

52 Voir Sandra Balck, Ingo Frank, Hermann Beyer-Thoma, Anna Ananieva, « Interlinking Text and Data with Semantic Annotation and Ontology Design Patterns to Analyse Historical Travelogues », *digital humanities quaterly*, n° 17/3, 2023 (http://www.digitalhumanities.org/dhq/vol/17/3/000726/000726.html).

ABSTRACTS

Français

Dans le domaine des éditions numériques savantes, les applications de l'intelligence artificielle (IA) et de l'apprentissage automatique gagnent de plus en plus d'influence. De nos jours, la transcription ainsi que l'annotation de textes peuvent être facilitées par des outils puissants tels que Transkribus, avec lesquels des modèles d'IA peuvent être entraînés pour effectuer une reconnaissance de texte manuscrit (HTR). Dans une deuxième étape, les données textuelles peuvent être traitées par des logiciels conçus pour le Traitement du langage naturel (NLP) afin d'extraire et d'annoter des caractéristiques syntaxiques, morphologiques ainsi que des informations sémantiques. Par exemple, il est possible de baliser des personnes, des lieux et des organisations via la Reconnaissance d'entités nommées (NER), ce qui est particulièrement pertinent pour les éditions numériques de sources historiques.

Ces considérations sont l'une des principales préoccupations du projet de recherche et d'édition « Grand Tour digital » à la Bibliothèque Herzog August de Wolfenbüttel. Il vise à établir la faisabilité des méthodes d'IA pour l'édition savante de sources historiques et à adapter de telles méthodes dans un flux de travail complet et durable qui pourrait être applicable à d'autres projets également. Ce faisant, nous devons identifier les possibilités et, plus important encore, les défis potentiels de la technologie en ce qui concerne son adaptabilité aux sources historiques. Cet article présente un travail réalisable grâce à la NER en utilisant l'exemple d'un récit de voyage du milieu du xvii^{e} siècle écrit par un jeune apothicaire itinérant qui a voyagé dans la région baltique, à travers le nord de l'Allemagne, faisant fonction de laquais dans la suite du jeune duc Ferdinand Albrecht I^{er} de Brunswick-Wolfenbüttel-Bevern, en Suisse et en France.

En observant ce processus, apparaîtront certains problèmes et difficultés liés aux particularités des textes de l'époque moderne par rapport aux textes contemporains pour lesquels pratiquement tous les outils de NLP et de NER ont été initialement conçus. Tout d'abord, les textes actuels sont orthographiquement et grammaticalement beaucoup plus normalisés que

de nombreux textes anciens. À l'époque moderne, par exemple, le même scribe pouvait utiliser différentes orthographes du même mot sur la même page. De plus, de nombreux scribes et leurs textes étaient multilingues plutôt que monolingues, et même les noms de personnes ou de lieux pouvaient parfois être donnés dans différentes langues et versions, par exemple le prénom allemand Johannes ou Hans pouvait parfois apparaître dans sa forme française Jean même si la même personne était visée.
Il existe plusieurs approches pour atténuer ces difficultés et celles qui y sont liées, dont certaines seront évaluées dans cet article. La première consiste à utiliser des techniques d'apprentissage automatique pour former des modèles NER spécifiquement sur des textes de la période et de la langue qui nous intéressent. Le problème est qu'il existe, en général, très peu de jeux de données d'entraînement disponibles à partir de textes historiques qui peuvent être utilisés pour former des modèles spécialisés. Une autre approche est l'utilisation de gazetiers ou de dictionnaires de noms avec lesquels le programme peut reconnaître certains tokens [1] comme des noms. Encore une fois, il existe relativement peu de ressources pour les textes de l'époque moderne par rapport aux textes contemporains et surtout les noms de personnes montrent une grande variété et sont souvent très spécifiques à des textes individuels. Ainsi, l'approche basée sur le dictionnaire ou les règles ne peut être utilisée de manière significative que pour des entités qui existent sur une période relativement longue et peuvent apparaître, par conséquent, dans un plus grand nombre de textes, comme des lieux ou des organisations. Une troisième manière de faire face aux problèmes posés par les sources historiques pour la NER consiste en la combinaison des deux approches décrites ci-dessus. C'est possible avec certaines applications de NLP, par exemple spaCy, qui a été utilisé dans l'étude entre autre pour cette raison.
Dans l'ensemble, cet article présente une étude de cas pour l'application de méthodes NER aux éditions numériques savantes de textes de l'époque moderne. Il analyse les possibilités et les défis de cette entreprise et propose des solutions en cas de difficultés. Si ces réflexions peuvent être utiles à d'autres projets, elles sont encore à un stade préliminaire et nécessitent des tests et des améliorations supplémentaires.

English
In the field of scholarly digital editions applications of Artificial Intelligence (AI) and machine learning gain more and more influence. Nowadays, the transcription as well as the annotation of texts can be facilitated by powerful tools such as Transkribus with which AI-models can be trained to perform Handwritten Text Recognition (HTR). In a second step, the textual data can be processed by software designed for Natural Language Processing (NLP) to extract and annotate syntactical and morphological features as well as semantic information. For example, it is possible to markup persons, places and organisations via Named Entity Recognition (NER), which is especially relevant for digital editions of historical sources. These considerations are one main concern of the research and edition

project “Grand Tour digital” at the Herzog August Library Wolfenbüttel. It aims to establish the feasibility of AI-methods for the scholarly editing of historical sources and to adapt such methods into a comprehensive and sustainable workflow which might be applicable for other projects as well. In doing so, we need to identify the possibilities and, more importantly, potential challenges for the technology regarding its adaptability to historical sources. The present article outlines a possible workflow for NER using the example of a mid-17th century travelogue written by a young apothecary journeyman who travelled to the Baltic region, through northern Germany and, as some sort of lackey in the party of the young Duke Ferdinand Albrecht I. of Brunswick-Wolfenbüttel-Bevern, to Switzerland and France.

While discussing the workflow, some problems and difficulties will become apparent which are a result of the peculiarities of early modern in contrast to modern texts for which virtually all NLP and NER tools were originally designed. First and foremost, modern texts and languages are orthographically and grammatically far more normalized than many historical texts. In the early modern era, for example, the same scribe could employ different spellings of the same word on the same page. Additionally, many scribes and their texts were multilingual rather than monolingual, and even the names of persons or places could occasionally be given in different languages and different versions, e. g. the German given name Johannes or Hans would sometimes appear in its French form Jean even if the same person was meant by it.

There are several approaches to mitigate these and related difficulties, some of which will be evaluated in the paper. The first one is to use machine learning techniques to train NER models specifically on texts of the time period and language one is interested in. One problem regarding this solution is that there are, generally speaking, very few and rather small available training sets of suitable data from historical texts which can be used to train specialised models. Another approach is the use of gazetteers or dictionaries of names with which the program can recognize certain tokens as names. Again, there are comparatively few resources for pre-modern than modern texts to build and especially person names show a great variety and are often very specific for single texts. Thus, the dictionary- or rule-based approach can only be used in a meaningful way for entities that are existent over a relatively long time period and may appear, therefore, in a greater number of texts, such as places or organisations. A third way of addressing the problems posed by historical source for NER consists of the combination of the two approaches outlined above. This is possible with some NLP-applications, e. g. spaCy which was used in the present study because of that reason among others.

All in all, this article presents a case study for the application of NER methods to scholarly digital editions of early modern texts. It analyses the possibilities and challenges of this venture, and proposes some solutions for potential problems and difficulties. While these will be hopefully useful for

other projects with similar concerns, they are still in a preliminary state and need further testing and improvement.

INDEX

Mots-clés
éditions numériques savantes, reconnaissance d'entités nommées, récit de voyage, époque moderne, Grand Tour, apprentissage automatique, intelligence artificielle

Keywords
scholarly digital editions, named entity recognition, travelogue, early modern era, Grand Tour, machine learning, artificial intelligence

AUTHOR

Maximilian Görmar
IDREF : https://www.idref.fr/281724733

TRANSLATOR

Angela Göbel
Herzog August Bibliothek Wolfenbüttel
IDREF : https://www.idref.fr/28172489X

Éditer la recherche et les données de la recherche : conjuguer plateforme de livres et entrepôt de données

Publishing research and research data: combining a book platform and a data warehouse

Christine Chadier

DOI : 10.35562/theia.161

Copyright
CC BY-NC-SA 4.0

OUTLINE

Pourquoi compléter des sites de publications en ligne par un entrepôt de données ?
Pourquoi choisir Nakala ?
  Une plateforme portée par une infrastructure de recherche
  Nakala ? Qu'est que c'est ?
Construire ses données dans Nakala
  Les métadonnées dans Nakala
    Les métadonnées générées par Nakala
    Les métadonnées à ajouter par les déposants
    Où peuvent se retrouver les métadonnées en dehors de Nakala ?
  Les identifiants dans Nakala
  Les images dans Nakala
    IIIF : International Image Interoperability Framework
  Générer un site web personnalisé avec « Nakala_Press »
La « collection » Chrétiens et Sociétés sur Nakala
  Le rôle de l'éditrice ou de l'éditeur

AUTHOR'S NOTES

Cette communication se veut « un retour d'expérience » dans le cadre des choix éditoriaux faits au sein du LARHRA pour les éditions « Chrétiens et Sociétés ». Il n'a pas pour ambition de valoriser tel ou tel procès, mais d'exposer les raisons qui ont conduit un laboratoire-éditeur à faire de tels choix en matière de stockage de données. Le dispositif est évolutif et continue d'être développé actuellement.

TEXT

Pourquoi compléter des sites de publications en ligne par un entrepôt de données ?

1 En France, l'ouverture des données de recherche a été prévue dans le premier volet de la loi pour une République numérique (loi dite « Axelle Lemaire » n° 2016-1321 du 7 octobre 2016) puis inscrite depuis dans le code de la recherche :

> Article 533-4-II du code de la recherche
> I.- Lorsqu'un écrit scientifique issu d'une activité de recherche financée au moins pour moitié par des dotations de l'État, des collectivités territoriales ou des établissements publics, par des subventions d'agences de financement nationales ou par des fonds de l'Union européenne est publié dans un périodique paraissant au moins une fois par an, son auteur dispose, même après avoir accordé des droits exclusifs à un éditeur, du droit de mettre à disposition gratuitement dans un format ouvert, par voie numérique, sous réserve de l'accord des éventuels coauteurs […]
> II.- Dès lors que les données issues d'une activité de recherche […] ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre.
> III.- L'éditeur d'un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication.

2 Les jeux de données rassemblés au cours d'une recherche deviennent ainsi aussi importants que la publication des résultats scientifiques. Leur mise à disposition à l'issue d'un programme de recherche peut même être exigée par les partenaires financiers (demandes de financement de projets européens H2020, programmes ANR [1]). L'éditeur, particulièrement l'éditeur scientifique public, se doit donc

d'accompagner ce mouvement afin que la publication des données n'entre pas en concurrence avec la publication des écrits scientifiques (articles ou ouvrages) et afin de ne pas déconnecter les données de la recherche des résultats de la recherche.

3 Si on veut que les données soient publiées selon les critères FAIR (Facile à trouver, Accessible, Interopérable et Réutilisable) qui favorisent, outre l'ouverture des données, leur mise à disposition avec un triple objectif de qualité des données et des métadonnées, d'inscription dans un cycle de vie maitrisé par les scientifiques et enfin de pérennité des données sur le long terme (accès, intégrité, contextualisation de la production des données), il convient de proposer, en appui des revues et des collections d'ouvrages une solution pour l'hébergement des données de la recherche [2] et de déterminer la solution la mieux adaptée à la discipline et aux choix éditoriaux.

Pourquoi choisir Nakala ?

4 Avec l'obligation d'ouvrir les données de la recherche, différents entrepôts de données se mettent en place dont plusieurs spécialisés en SHS, mais certains sont institutionnels (Collection Sciences Po - data.sciencespo réservé à la communauté académique de Sciences Po et partenaires de Sciences Po), d'autres réservés à un type de données comme Quetelet-Progedo (données quantitatives – données d'enquêtes, données quantitatives tabulées) ou le Conservatoire national des données 3D qui utilise un Schéma de métadonnées « maison » vu la spécificité des données hébergées [3].

Une plateforme portée par une infra-structure de recherche

5 Nakala, portée par l'infrastructure de recherche Huma-Num mise en œuvre par le CNRS avec le Campus Condorcet et Aix-Marseille Université a pour principale mission de construire, avec les communautés et à partir d'un pilotage scientifique, une infrastructure numérique de niveau international inscrits dans les principes FAIR. Engagée dans l'European Open Science Cloud, elle porte la participation de la France dans l'European Research

Infrastructure Consortium [ERIC DARIAH (Digital Research Infrastructure for the Arts and Humanities)].

6 Depuis janvier 2021, Huma-Num porte avec OpenEdition et Métopes le Consortium de moyens mutualisés pour des services et données ouvertes en SHS (COMMONS). COMMONS, projet lauréat de l'appel à manifestations d'intérêt « Équipements structurants pour la recherche : EquipEx+ » du Programme d'investissements d'avenir (PIA 3), associe trois infrastructures françaises qui jouent un rôle essentiel dans les Sciences humaines et sociales et en particulier dans l'édition en SHS. COMMONS permet d'aborder l'ensemble de la chaîne de production des connaissances : de la gestion des données à la diffusion et l'édition. Publiant en accès ouvert nos ouvrages et revue sur OpenBooks et OpenJournals, les plateformes d'OpenEdition et utilisant la chaine Métopes [4] pour le travail éditorial, il était donc logique que nous nous tournions vers Huma-Num et Nakala pour choisir le lieu d'hébergement pour publier les données ayant été utilisées pour la rédaction des publications éditées par nos soins ou les données complémentaires qui pourront ainsi plus facilement être réutilisées que des banque d'images, des tableaux statistiques, des reproduction de documents, etc.

7 Nous avons donc choisi pour accompagner la revue *Chrétiens et Sociétés* XVI^e^-XXI^e^ *siècles* (https://journals.openedition.org/chretienssocietes/) et la collection « Chrétiens et Sociétés. Documents et mémoires » (https://books.openedition.org/larhra/634) de créer une collection dans Nakala.

Nakala ? Qu'est que c'est ?

8 Nakala est l'un des centres de référence de l'écosystème national Recherche Data Gouv.fr, Un écosystème au service du partage et de l'ouverture des données de recherche. Se rapprocher de réseaux structurés comme ceux sur lesquels s'appuie Nakala permet de bénéficier de services offrant stabilité et visibilité… Contrairement à d'autres entrepôts de données comme Zenodo porté par le CERN (Conseil européen pour la recherche nucléaire), il s'agit d'un entrepôt de données spécialisé en SHS. Lancé dès 2014, il répond aux besoins d'accès persistant et interopérable aux données numériques. Il se compose de deux ensembles :

- un visible de tous permettant l'accès aux données
- un qui sera utilisé par les robots pour pouvoir moissonner les données qui contient la présentation des métadonnées

9 La qualité et le soin apporté au traitement des métadonnées sont la clef de voûte pour une meilleure exposition des données de la recherche. Les métadonnées, c'est-à-dire des informations de description ou « données à propos des données », permettent de décrire les caractéristiques d'une donnée ou d'un jeu de données, par exemple :

- auteurs/autrices
- description du contenu
- date de création
- lieu de capture/de production
- raison pour laquelle les données ont été générées
- comment les données ont été créées
- etc.

10 Les métadonnées doivent être aussi riches, précises et exactes que possible.

11 Nakala propose des services interopérables de présentation des métadonnées

- exposition dans un triplestore RDF [5]
- accès via un entrepôt OAI-PMH [6].

12 Outre cette structuration des données, Nakala assure l'accès aux données :

- accès et citation par un identifiant pérenne (Digital Object Identifier, DOI)
- stockage sécurisé
- accès permanent.

Construire ses données dans Nakala

13 Un dépôt, au sens où Nakala l'entend, est un regroupement d'informations contenant

des données : des fichiers numériques (au minimum 1)
- des métadonnées
- un identifiant pérenne de type DOI attribué automatiquement à chaque donnée publique et déclaré auprès de DataCite.

14 Nakala contient aussi d'autres informations

- des collections (regroupements de références à des données)
- des référentiels (vocabulaires utilisés dans les descriptions).

15 C'est cet ensemble qui doit être appréhendé pour comprendre et pouvoir réutiliser les données.

Les métadonnées dans Nakala

16 Les données déposées, mais aussi les collections et les référentiels, doivent être accompagnés de métadonnées :

- certaines sont générées par Nakala
- les autres sont ajoutées par le déposant.

Les métadonnées générées par Nakala

17 Les métadonnées générées automatiquement par Nakala se composent d'informations de gestion :

- dates de dépôt et de dernière modification, version, statut, identification du déposant, identifiant du dépôt

18 et d'informations techniques sur les fichiers :

- taille (en octets), type de média (mime), empreinte digitale (sha-1).

Les métadonnées à ajouter par les déposants

19 Ce sont les informations documentaires qui vont permettre d'identifier la donnée à deux niveaux.

- Au niveau du dépôt :

20 avec cinq champs obligatoires : Titre(s), Auteur(s), date de création, licence, type
et des champs facultatifs : le déposant peut ajouter l'ensemble des champs du vocabulaire Dublin-Core qualifié [7].

- Au niveau des fichiers :

21 la date de visibilité (embargo) : par défaut c'est la date du dépôt
une description facultative mais fortement conseillée afin de valoriser les données.

Où peuvent se retrouver les métadonnées en dehors de Nakala ?

22 Excepté la page d'administration qui n'est visible que par le seul déposant, tous les autres accès sont susceptibles d'être consommés par des services extérieurs à Nakala. Toutes les personnes et services qui moissonnent ces métadonnées sont susceptibles de les réexposer et leurs réexpositions d'être moissonnées et réutilisées à nouveau par d'autres personnes ou services.

23 Cela peut se faire :

- automatiquement par la découverte par des robots
- par les moteurs de recherche type Google, Google data search, Google scholar
- dans leurs articles, ouvrages, blogs, etc.

Les identifiants dans Nakala

24 Chaque donnée déposée dans Nakala se voit attribuer un identifiant pérenne : un Digital Object Identifier (DOI) [8]. Les DOI sont mis à disposition par Datacite [9]. L'identifiant (DOI) sert à construire des URLs pour obtenir l'accès aux différents éléments d'un dépôt

25 Pour accéder à la page d'accueil (*landing page*) d'un dépôt les URLs sont les suivants :

- https://nakala.fr/[DOI] ou https://doi.org/[DOI]
- https://nakala.fr/10.34847/nkl.4b33r2h4
- https://doi.org/10.34847/nkl.4b33r2h4.

26 Par extension cet URL est souvent considéré comme l'identifiant du dépôt.

Les images dans Nakala

27 Nakala intègre le protocole IIIF (International Image Interoperability Framework). Intégrer la communauté permet de mieux valoriser les documents numérisés (reproduction de documents, photographies d'œuvres ou de lieux, etc.) en donnant aux utilisateurs la possibilité d'accéder facilement aux images et au contexte de production de celles-ci. Mais surtout il s'agit d'un format interactif qui apporte aux images des fonctionnalités avancées de visualisation et de manipulation des documents (zoom profond, visualisation de documents en haute résolution, gains d'accessibilité, facilités d'accès et de partage des ressources grâce à des liens en principe pérennes, etc.).

IIIF : International Image Interoperability Framework

28 IIIF désigne à la fois une communauté et un cadre d'interopérabilité pour diffuser, présenter et annoter des images et documents audio/vidéo sur le Web, qui s'est imposé en quelques années comme un standard et une brique technologique essentielle pour décloisonner les collections numérisées des institutions patrimoniales à l'échelle mondiale. Cela recouvre un modèle pour présenter et annoter des contenus numériques (images, documents audio et vidéo) et un cadre de réflexion sur les mécanismes d'échange des images entre entrepôts numériques : Nakala, BnF, BL, Cornell, Los Alamos National Laboratory, NL of Norway, Oxford, Stanford…

29 IIIF propose un cadre technique commun grâce auquel les fournisseurs et créateurs d'images numériques peuvent :

- délivrer leurs images de manière standardisée sur le Web
- les rendre consultables, manipulables et annotables
- rendre cela possible pour n'importe quelle application ou logiciel compatible.

30 Utiliser les standards du IIIF c'est offrir aux images utilisées pour la rédaction d'ouvrages et d'articles scientifiques une visibilité et une interopérabilité qu'elles ne pourraient avoir en restant seulement

insérées dans un texte même publié en accès ouvert sur une plateforme [10].

Générer un site web personnalisé avec « Nakala_Press »

31 À terme, il est prévu de générer un site web avec Nakala-Press, lorsque nous aurons accumulé dans notre collection Nakala suffisamment de données liées à *Chrétiens et Sociétés* pour l'alimenter de façon conséquente. Nakala-Press est un module de publication qui permet de créer un site web autour de données publiques déposées dans Nakala. Ce site web permettra de mettre en valeur les données rassemblées et d'en faciliter l'accès à un public plus large.

La « collection » Chrétiens et Sociétés sur Nakala

32 Le premier volume publié en association OpenBooks/Nakala sera le tome 4 des *Matériaux pour l'histoire religieuse du peuple français Bourgogne, Franche-Comté et Lyonnais Savoie et Dauphiné Grand Midi et Algérie* (figure 1) [11].

33 Les notices seront mises en ligne sous la forme d'un volume de la collection *Chrétiens et Sociétés. Documents et Mémoires* sur OpenBooks (https://books.openedition.org/larhra/634) [12].

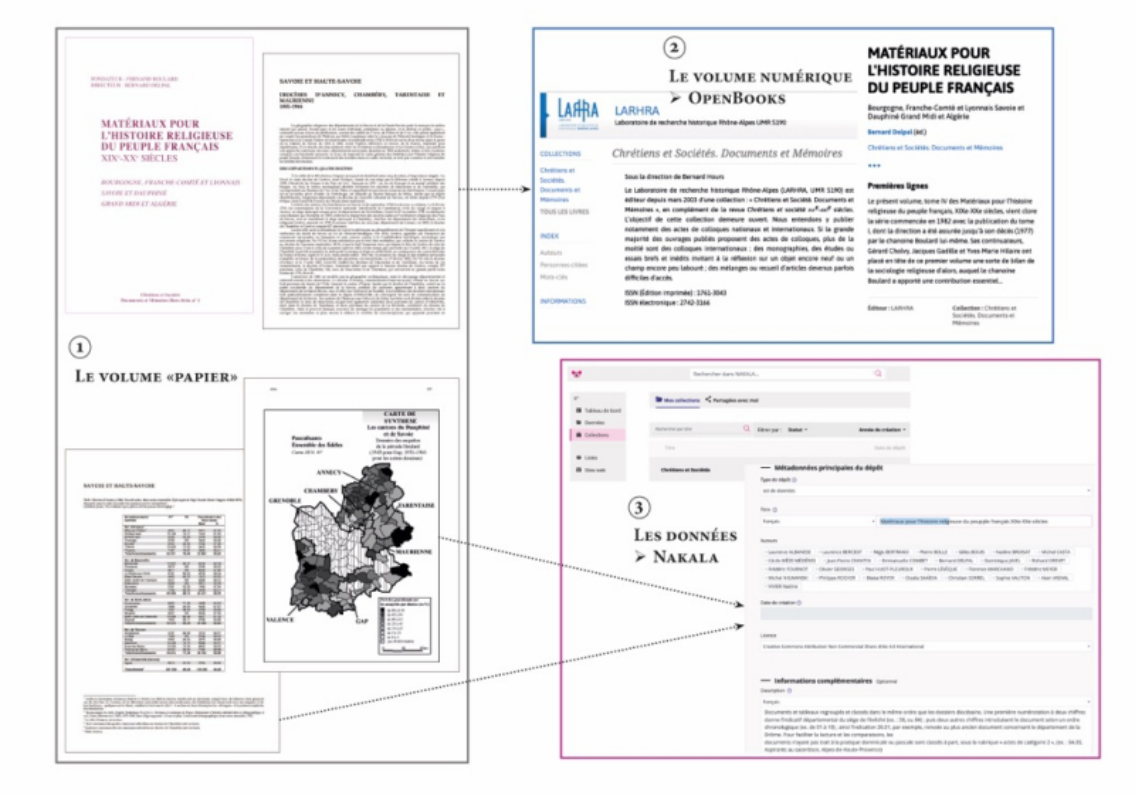
34 Il s'agira ensuite de déposer le reste du volume sur Nakala. Nous aurons deux types de données :

- une « donnée » [au sens Nakala] qui rassemblera l'ensemble des tableaux avec un fichier par département, ce qui reprend la structuration du volume papier, avec une feuille pour chaque tableau, le tout en .csv [13].
- une « donnée » qui rassemblera l'ensemble des cartes au format pdf. Le format pdf a été retenu de façon à pouvoir conserver la même configuration que pour les tableaux, soit un fichier par département contenant un ensemble de carte.

35 La convention de nommage des fichiers reprendra celle des tableaux de la version papier. Le choix de la même structure pour les fichiers des tableaux et des cartes permettra de mettre facilement en regard

les tableaux mais ne permettra pas d'appliquer le protocole IIIF. C'est un choix éditorial pour s'intégrer au mieux à la structure des textes du volume mis en ligne sur OpenBooks. Cela n'empêchera pas de créer un jeu de métadonnées développé pour donner visibilité et accessibilité tant aux cartes qu'aux tableaux.

Figure 1. Mise en ligne du tome 4 des *Matériaux pour l'histoire religieuse du peuple français Bourgogne, Franche-Comté et Lyonnais Savoie et Dauphiné Grand Midi et Algérie*



Réalisation Christine Chadier

Le rôle de l'éditrice ou de l'éditeur[14]

36 Ce nécessaire travail de relecture et d'harmonisation qui apparaît dans le cadre de la mise en ligne des données des « Matériaux Boulard » soulève la question de l'éditorialisation des données[15]. En effet, si on peut envisager, dans le cadre d'un projet de recherche ou d'un simple « *work in progress* », de déposer des données directement issues de la recherche qui constituent en soi un ensemble unique, cela me paraît plus difficile dans le cadre d'une

collection en appui de plateformes d'édition. L'édition impliquant en effet tout un travail d'harmonisation et de suivi afin de valoriser le texte publié, il convient de prévoir le même travail d'harmonisation sur les données si l'on souhaite les valoriser au-delà de la seule mise à disposition de la communauté. Le travail d'éditorialisation est essentiel pour favoriser la lisibilité des données. Le dépôt des données « brutes » suffit pour de la simple récupération de données en vue de leur réutilisation ou le moissonnage par les machines, mais si l'on veut qu'elles puissent être consultées aisément, il faut un travail de préparation.

37 Dans ce cadre, il me paraît indispensable d'avoir des données structurées de façon identique, par exemple par ouvrage ou par numéro de revue, avec des conventions de nommage qui permettent de rattacher facilement la donnée à son ouvrage ou son numéro de revue et obtenir des URL un minimum standardisées. C'est pourquoi il me semble nécessaire que la personne responsable de la collection dans Nakala soit également la·le responsable des sites de publication sur les plateformes. La publication des données viendrait donc s'ajouter aux compétences de l'éditrice ou de l'éditeur. L'édition électronique est venue s'ajouter à l'édition imprimée à la fin des années 1990 et au début des années 2000 comme la publication des données et la gestion d'un entrepôt de données vient s'ajouter aujourd'hui à l'édition électronique.

38 L'évolution des compétences doit donc suivre l'évolution des types de contenus à publier et des outils. Cela nécessite une collaboration accrue avec les auteurs pour valider l'éditorialisation des données qu'il a précieusement accumulées au cours de ses recherches et qu'il n'avait pas forcément prévu de publier. Le rôle des chercheurs est de veiller à la qualité des données publiées et le rôle de l'éditrice ou de l'éditeur, comme elle·il le fait déjà pour les revues ou les ouvrages, est de veiller au respect du travail du chercheur mais également à la bonne visibilité des données, aux conditions de réutilisation et à ce qu'elles puissent bénéficier d'un stockage « pérenne », autant que nous puissions l'envisager.

39 Il est à noter que cette montée en compétences de l'éditrice ou de l'éditeur, rendue possible en s'inscrivant, par exemple, dans des réseaux « métiers » ou des listes de diffusion [16], permet une montée

en qualité des revues ou des collections et de quitter le statut de « laboratoire-éditeur » pour rejoindre le cercle des 106 éditeurs scientifiques publics recensés par Caroline Dandurand lors de son enquête de 2022 [17] aux côtés des Presses universitaires de Lyon, des Presses universitaires de Rennes, de Sorbonne Université, des Presses UGA Éditions, des Publications de l'École française de Rome, d'ENS Éditions, de MOM Éditions - Maison de l'Orient et de la Méditerranée, etc. En choisissant de rejoindre en 2023 Nakala après avoir rejoint en 2008 OpenEdition, Chrétiens et Sociétés suit le même chemin que *GalliaArchéologies des Gaules*, revue CNRS. Cette évolution ne serait pas possible sans s'appuyer sur un contenu scientifique reconnu par la communauté et validé par ses pairs. La raison d'être du travail réalisé est d'apporter des outils de diffusion de qualité à un contenu de qualité.

Figure 2. Le rôle de l'éditrice ou de l'éditeur



Musique d'ensemble, Testu & Massin (Paris) [18], EST 5 C/43,
Estampe https://selene.bordeaux.fr/ark:/27705/330636101_EST_5_C_43/v0001.simple.highlight=musique%

NOTES

1 Cécile Arènes, Lionel Maurel, Stéphanie Rennes, *Guide d'application de la Loi pour une République numérique pour les données de la recherche*, Comité pour la science ouverte, 2022 ⟨hal-03968218⟩

2 Romain Feret, Françoise Catherine Gouzi, Sandra Guigonis, Hélène Jouguet, Nicolas Larrousse, *et al.*, Recommandations aux revues souhaitant définir une « politique de données » liées aux publications [Rapport Technique], Comité pour la science ouverte, 2020, 8 p. ⟨hal-03594383⟩

3 « Publication d'une liste d'entrepôts pour les données en SHS », *Le blog d'Huma-Num et de ses consortiums*, URL : https://humanum.hypotheses.org/9880

4 Métopes fournit un ensemble d'outils et de méthodes permettant de créer des fonds éditoriaux structurés et d'assurer la diffusion des produits éditoriaux, numériques ou imprimés, à partir de fichiers texte et facilite ainsi la diffusion des résultats et des données de la recherche.

5 Un triplestore RDF (Resource Description Framework) agit comme une base de données relationnelle : il stocke des données et les récupère via un langage de requête. Mais contrairement à une base de données relationnelle, un triplestore ne stocke qu'un seul type de données : le triplet (sujet - prédicat - objet).

6 L'Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) est un protocole informatique développé par l'Open Archives Initiative afin d'échanger des métadonnées.

7 Le Dublin Core est un format descriptif simple et générique créé en 1995 à Dublin (Ohio) par OCLC (Online Computer Library Center) et le NCSA (National Center for Supercomputing Applications). L'objectif du Dublin Core est de fournir un socle commun d'éléments descriptifs pour améliorer le signalement et la recherche de ressources au-delà des diverses communautés et des nombreux formats descriptifs propres à chaque spécialité, tout en restant suffisamment structuré. Pour en savoir plus : https://www.bnf.fr/fr/dublin-core.

8 Les DOI sont attribués pour les données scientifiques mais aussi pour d'autres types de ressources telles que des publications, des jeux et bases de données, des logiciels, des images, des cartes, etc.

9 DataCite est un consortium international de bibliothèques et services spécialisés dans les sciences de l'information, qui vise à faciliter l'archivage numérique ainsi que l'accès aux ressources numériques sur Internet, notamment par l'attribution d'un DOI à chacune d'entre elles. DataCite a été fondé le 1er décembre 2009 à Londres

10 Pour en savoir plus sur le IIIF : https://francearchives.gouv.fr/fr/article/705250527.

11 Les données ont été préparées par Anna Sébille, stagiaire du Master Sciences des religions et sociétés.

12 Suite à la fermeture de la plateforme OpenBooks du 19 février 2024 jusqu'à fin mars 2024 au plus tôt pour permettre la migration du système d'exploitation, la mise en ligne du volume est retardée.

13 Le format .csv (Comma Separated Values) est un format ouvert contrairement au format Document Microsoft Excel (.xls) ou Document Apple Numbers (.numbers) : c'est donc un format qui va maximiser la possibilité de réutilisation des données.

14 Nous entendons par « éditrice ou éditeur », l'ingénieur chargé de l'édition d'un ouvrage ou d'une revue et/ou la responsable éditoriale d'une collection.

15 Stéphane Renault, Blandine Nouvel, Micaël Allainguillaume, Astrid Aschehoug, Nicolas Coquet et Marie-Adèle Turkovics, « Harmoniser les pratiques éditoriales numériques des revues françaises d'archéologie », *Humanités numériques* [En ligne], 2 | 2020, mis en ligne le 01 juin 2020, consulté le 14 juin 2023. URL : http://journals.openedition.org/revuehn/483 ; DOI : https://doi.org/10.4000/revuehn.483

16 Par exemple le réseau Médici, réseau des métiers de l'édition scientifique ou la liste accesouvert@groupes.renater.fr, liste de discussion de la communauté du libre accès francophone.

17 Caroline Dandurand, *Préfiguration d'une structuration collective des éditeurs scientifiques publics engagés dans la science ouverte* [Rapport de recherche], Comité pour la science ouverte, 2022, 86 p. ⟨hal-03713434⟩

ABSTRACTS

Français
Il s'agit de présenter ici pourquoi et comment optimiser la mise en ligne d'un d'un article ou d'un ouvrage à l'ère de la science ouverte en présentant la démarche retenue pour les éditions « Chrétiens et Sociétés » publiée par le Laboratoire de recherche historique Rhône-Alpes.

English
The aim here is to explain why and how to optimise the online publication of an article or a book in the era of open science by presenting the approach adopted for the 'Chrétiens et Sociétés' editions published by the Rhône-Alpes Historical Research Laboratory.

INDEX

Mots-clés
science ouverte, entrepôt de données, Nakala, Huma-Num, OpenEdition, Métopes, COMMONS

Keywords
open science, data warehouse, Nakala, Huma-Num, OpenEdition, Métopes, COMMONS

AUTHOR

Christine Chadier
Université de Lyon Jean Moulin Lyon 3, LARHRA UMR 5190
IDREF : https://www.idref.fr/132141450
ISNI : http://www.isni.org/0000000357866609

Notes et documents

Saragosse lieu de siège, Saragosse pittoresque : les vues de Saragosse de Louis-François Lejeune pour le *Voyage pittoresque et historique en Espagne*

Saragossa, siege site, picturesque Saragossa: Louis-François Lejeune's views of Saragossa for the Voyage pittoresque et historique en Espagne

Béatrice Denis

DOI : 10.35562/theia.178

Copyright
CC BY-NC-SA 4.0

OUTLINE

La contribution de Lejeune à l'ouvrage de Laborde
Lejeune et l'Espagne

TEXT

1 Dans un article du *Moniteur* daté du 16 mai 1823, on fait l'annonce d'une nouvelle édition de 200 exemplaires du *Voyage historique et pittoresque de l'Espagne* [1], d'Alexandre de Laborde (1773-1842), qui avait paru en quatre volumes entre 1806 et 1820. L'annonce détaille longuement le contenu de ces exemplaires, dont une planche d'après un dessin de Louis-François Lejeune (1775-1848) :

> La manière de voyager en Espagne, d'après un tableau du général Lejeune, gravé par Bovinet. On y représente la manière de voyager des pèlerins, des moines, des militaires ; sur le devant on remarque le coche de Colleras attelé de six mules, espèce de berlingos, tels qu'ils étaient sous Louis XIV ; au fond des charrettes à roues pleines, telles qu'elles étaient du temps des Romains, etc., etc. [2]

2 Le dernier volume de l'ouvrage de Laborde, soit la seconde partie du second tome, publiée en 1820, comporte en effet cette *Manière de voyager en Espagne*, gravée par Edmé Bovinet (1767-1832) et Jean

Duplessis-Bertaux (1750-1818) d'après Lejeune (fig.1), accompagnant une description de la Castille. Cette planche semble être un point fort de l'ouvrage, le nom de Lejeune utilisé ici pour faire la publicité des nouveaux exemplaires, trois ans après la première publication. Officier lors des guerres révolutionnaires et napoléoniennes, Lejeune est surtout connu pour ses peintures de batailles, batailles auxquelles il a lui-même participé, pour la plupart. Une recension de toutes les mentions de Lejeune dans la presse de l'époque démontre qu'il était une figure médiatique relativement connue, surtout dans les années 1820.

3 La planche présente une scène pittoresque, animée de personnages de différentes classes sociales et de différentes occupations, devant se déplacer par des moyens de transport variés, et l'explication de la planche détaille trois de ces « manières de voyager en Espagne », reprises par l'annonce du *Moniteur*[3]. Cette planche est d'autant plus curieuse qu'elle va au-delà de la simple représentation d'un paysage, faisant intrusion dans la production paysagère de Lejeune. Ce dernier est en effet initié au paysage par Pierre-Henri de Valenciennes (1750-1819), formation qui imprègne le rendu des décors de ses scènes de bataille ainsi qu'une production de paysages indépendants. Le ton et le style de la composition s'avèrent être assez révélateurs de la versatilité de Lejeune, au tournant des années 1820, après la fin de l'Empire.

4 C'est justement sur cette versatilité que cet article souhaite attirer l'attention. En plus de la *Manière de voyager en Espagne*, c'est sept autres vues que Lejeune produit pour l'illustration de la seconde partie du second tome de l'ouvrage de Laborde, soit deux vues de Pampelune, accompagnant la description de la Navarre, et cinq de Saragosse, accompagnant la description de l'Aragon. Ces vues de Saragosse permettent plus particulièrement de considérer un pan très peu étudié de l'œuvre de Lejeune, ainsi que d'avancer certaines hypothèses quant à l'origine et à la singularité de la contribution de ce dernier à l'ouvrage de Laborde. La littérature sur Lejeune concerne majoritairement ses peintures de bataille, et l'étude de sa production paysagère a surtout été circonscrite à son œuvre peint. Sa production dessinée a été très peu considérée et s'est surtout limitée à ses dessins ou esquisses pour ses tableaux, en plus du contenu de ses agendas ou de ses carnets[4]. Sa contribution à certaines entreprises

illustrées, comme celle de l'ouvrage de Laborde, a été largement ignorée jusqu'à maintenant. Si l'ouvrage de Laborde a lui-même été l'objet de considération académique, l'implication de Lejeune dans l'illustration de l'ouvrage reste non documentée, et ses vues, très peu reproduites depuis le 19e siècle [5]. C'est bien malgré la numérisation par l'INHA, dès le milieu des années 2000, des volumes de l'ouvrage, mais aussi de 255 dessins originaux qui ont servi à l'illustrer et qui sont conservés dans la collection Jacques Doucet. Dans cette collection de l'INHA, sont conservés trois dessins originaux de Lejeune, non datés [6].

La contribution de Lejeune à l'ouvrage de Laborde

5 Il est possible que Lejeune ait rencontré Laborde alors que ce dernier était en ambassade avec Berthier en 1809 [7]. Lejeune venait d'assister aux sièges de Saragosse par l'armée française. Si ses mémoires, *Souvenirs d'un officier de l'Empire*, [8] ne sont parus qu'après sa mort, il publie son compte-rendu des *Sièges de Saragosse* de son vivant, en 1840 [9] : ce texte témoigne à plusieurs endroits de la familiarité qu'a Lejeune avec la géographie de la ville et de ses environs [10]. Il fait par ailleurs un tableau [11], présenté au Salon de 1827, représentant l'assaut du couvent de Santa Engracia du 27 janvier 1809. Ce même couvent fait l'objet de deux vues de Lejeune dans l'ouvrage de Laborde, une intérieure (fig.2) et une extérieure (fig.3), vues d'intérêt, considérant que le couvent a été détruit pendant les sièges.

6 Par ailleurs, avec une vue d'ensemble des planches produites pour l'ouvrage, on ne peut éviter de remarquer que ses vues ont impliqué des graveurs qui se sont démarqués dans la représentation militaire. Si la plupart des graveurs employés pour l'ouvrage ont réalisé un ensemble de planches, Duplessis-Bertaux et Bovinet (qui grave la *Bataille d'Aboukir* de Lejeune, en 1808, pour le compte du Dépôt de la Guerre), n'ont contribué qu'à *La manière de voyager en Espagne*. Leur participation à un tel ouvrage ne semblerait expliquée que par ce lien avec Lejeune. De même, contrairement aux autres dessinateurs du quatrième volume – Dominique Vivant Denon, F. Liger, Dutailly, Six, etc. – Lejeune ne contribue pas à l'illustration des autres volumes. La ponctualité de sa participation à l'ouvrage de Laborde s'expliquerait

ainsi par sa connaissance de l'Aragon et des régions traversées lors de la campagne d'Espagne, décrites dans ce volume.

Lejeune et l'Espagne

7 Rien n'indique que Lejeune ait revisité l'Espagne après les guerres napoléoniennes. Ces vues ne seraient donc le résultat que de son expérience militaire des lieux. Que Lejeune s'inscrive alors dans une représentation pittoresque et romantique [12] d'un pays qu'il a connu et traversé par la guerre mérite particulièrement d'être investigué. En effet, le couvent de Santa Engracia se retrouve ici représenté par Lejeune, alors qu'il a lui-même contribué à sa destruction en tant qu'officier de l'Empire, jetant un nouvel éclairage sur sa relation avec la capitale aragonaise.

8 Lejeune a à l'occasion dessiné des compositions destinées uniquement à la gravure, comme c'est le cas pour ses représentions du champ de bataille d'Eylau [13] et de la rencontre de Napoléon et du tsar Alexandre à Tilsit [14], deux dessins produits pour être gravés au Dépôt de la Guerre. Ces deux dessins à l'aquarelle (1807) ont probablement été réalisés à l'instigation du maréchal Alexandre Berthier (1753-1815), alors ministre de la Guerre, de qui Lejeune était l'aide de camp de 1800 à 1812. Les vues destinées à l'ouvrage de Laborde sont toutefois entièrement produites hors du contexte commanditaire napoléonien, et ne sont, de surcroit, pas des représentations de nature militaire.

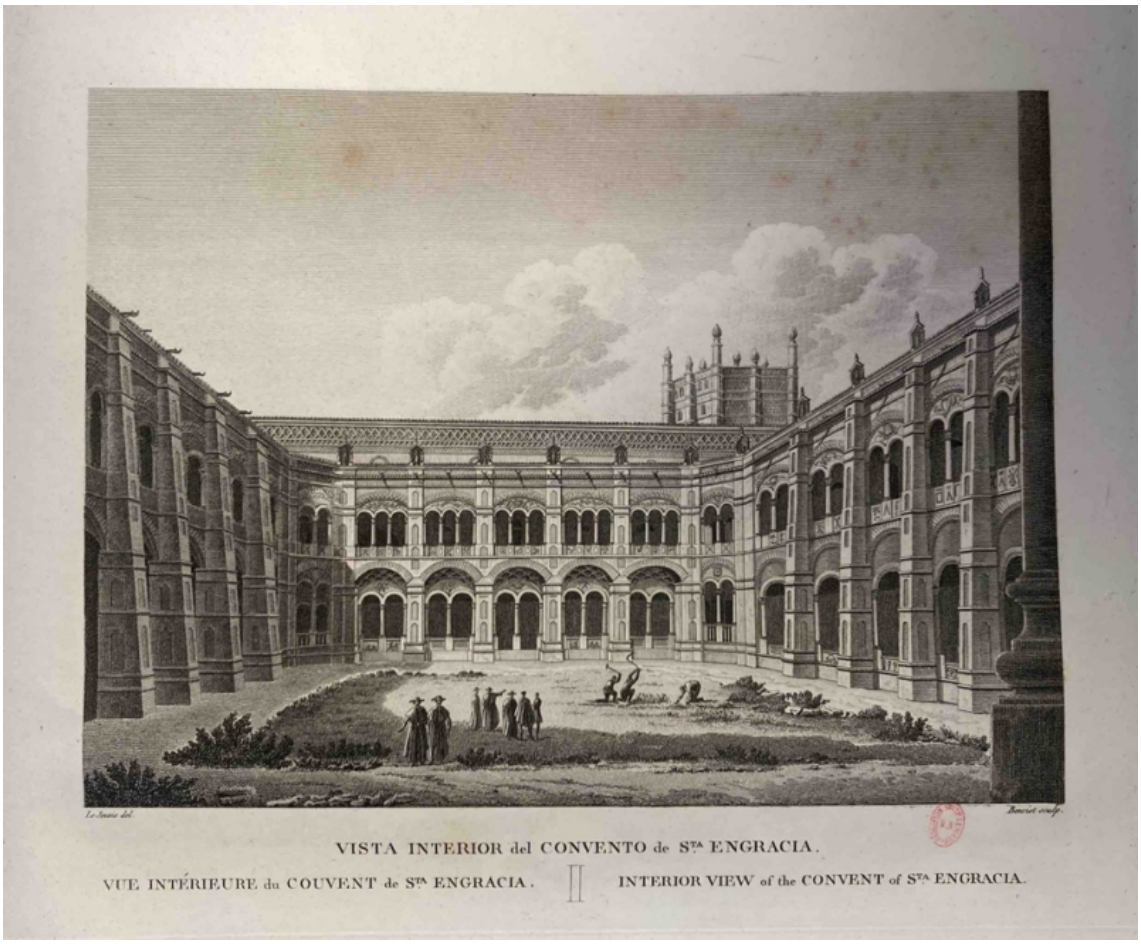
9 Cette production particulière au sein de l'œuvre de Lejeune permet ainsi de réfléchir à des aspects fascinants de sa relation avec l'Espagne, et avec Saragosse tout particulièrement : entre destruction, pendant les sièges, et conservation, avec ses vues et ses mémoires.

Fig. 1 – Louis-François Lejeune (dessiné par), Edmé Bovinet et Jean Duplessis-Bertaux (gravé par). *Manière de voyager en Espagne*. Planche de l'ouvrage *Voyage pittoresque et historique en Espagne*, Paris, P. Didot l'Aîné, 1820



Bibliothèque nationale de France. SMITH LESOUEF R-1043. Photo : Béatrice Denis

Fig. 2 – Louis-François Lejeune (dessiné par) et [M.A.] Benoist (gravé par). *Vue intérieure du couvent de Santa Engracia*. Planche de l'ouvrage *Voyage pittoresque et historique en Espagne*, Paris, P. Didot l'Aîné, 1820



SMITH LESOUEF R-1043, Bibliothèque nationale de France. Photo : Béatrice Denis

Fig. 3 – Louis-François Lejeune (dessiné par) et Robert Daudet (gravé par). *Vue extérieure du couvent de Santa Engracia*. Planche de l'ouvrage *Voyage pittoresque et historique en Espagne*, Paris : P. Didot l'Aîné, 1820



SMITH LESOUEF R-1043, Bibliothèque nationale de France. Photo : Béatrice Denis

NOTES

1 Alexandre de Laborde, *Voyage pittoresque et historique de l'Espagne* (Paris : Imprimerie de P. Didot l'Aîné, 1806-1820).

2 « Voyage historique et pittoresque de l'Espagne, par le… », *Gazette nationale ou le Moniteur universel*, 16 mai 1823.

3 Pour l'explication de la planche, voir A. Laborde, *op. cit.*, t.2 seconde partie, p. 32-33.

4 Le principal ouvrage concernant l'œuvre de Lejeune est le catalogue de l'exposition sur l'artiste qui a eu lieu à Versailles en 2012, dirigé par Valérie Bajou (dir.), *Les Guerres de Napoléon : Louis-François Lejeune, Général*

et Peintre, Paris, Hazan, 2012. Bien que des dessins paysagers de Lejeune provenant de ses carnets soient inclus dans le catalogue, on n'y trouve aucune mention des dessins réalisés pour l'ouvrage de Laborde.

5 L'ouvrage de Laborde a été étudié autant du côté francophone qu'hispanophone. Les dessins et illustrations ont cependant plus rarement fait l'objet d'études, voir par exemple Jordi Casanovas i Miró et Francesc M. Quílez i Corella, *El viatge a Espanya d'Alexandre de Laborde (1806-1820) : dibuixos preparatoris*, Barcelone, Museu Nacional d'Art de Catalunya, 2006 ; Suzanne Damiron, « Dessins originaux pour l'illustration du *Voyage pittoresque et historique de l'Espagne* d'Alexandre de Laborde », dans *Actas del XXIII Congreso internacional de historia del arte, Granada, 1973. España entre el Mediterraneo y el Atlantico*, vol. 2, Granada, Universidad de Granada, 1976, p.484-487 ; Núria Llorens, « Las escenas de paisaje en el "Viaje a España" de Alexandre de Laborde », *Anuario del Departamento de Historia y Teoría del Arte*, 2007, nº 19, p. 159-78. Les illustrations de Lejeune, si elles sont parfois montrées, sont rarement commentées. Deux des vues de Lejeune sont reproduites par exemple dans l'article de Maria Dolores Cabra, « La España prerromántica de Alexandre Laborde », *Historia 16*, 1984, nº 103, p. 23-32, mais Lejeune n'est pas nommé comme dessinateur et les images, non commentées.

6 *Autre couvent près de Saragosse*, *Seconde vue de Pampelune* et *Vue intérieure du couvent de Santa Engracia*, ca. 1820, mine de plomb, plume, lavis. NUM MS 463, Bibliothèque de l'Institut National d'Histoire de l'Art, collections Jacques Doucet, URL : https://bibliotheque-numerique.inha.fr/idurl/1/9077. D'autres dessins originaux de l'ouvrage sont aussi conservés dans la collection du Musée national d'art de Catalogne, à Barcelone.

7 Elena Fernández Herr, *Les Origines de l'Espagne romantique : les récits de voyage, 1755-1823*, Paris, Didier, 1974, p. 212.

8 Louis-François Lejeune, *Souvenirs d'un officier de l'Empire*, 2 vol. (Toulouse : Typographie Viguier, 1851).

9 Louis-François Lejeune, *Sièges de Saragosse, histoire et peinture des événements qui ont eu lieu dans cette ville ouverte pendant les deux sièges qu'elle a soutenus en 1808 et 1809* (Paris : Firmin-Didot Frères, 1840).

10 Pour une description de l'emplacement de Santa Engracia, par exemple, voir L.-F. Lejeune, *Sièges*, p. 23 ; une description de la ville, p. 38-42.

11 Louis-François Lejeune, *Une scène du siège de Saragosse, en 1809*, 1827. Huile sur toile, 162 x 129 cm. Versailles, Musée national des châteaux de

Versailles et de Trianon.

12 Sur une vision pittoresque du paysage et des coutumes espagnoles diffusées par la littérature de voyage à laquelle l'ouvrage de Laborde a contribué, voir E. F. Herr, *op. cit.*, p. 13-16 ; Gloria Mora, « *Voyage et itinéraire*, les descriptions de l'Espagne dans la seconde moitié du XVIIIème siècle et la première moitié du XIX^{e} siècle », dans J. Casanovas i Miró et F. M. Quílez i Corella, *op. cit.*, p. 272-275.

13 Louis-François Lejeune, *Épisode de la campagne de Pologne (Eylau)*, 1807. Aquarelle et lavis gris sur papier vergé crème, 35,9 x 59 cm. Paris, Musée du Louvre.

14 Louis-François Lejeune, *Tilsitt. – Entrevue de Napoléon et d'Alexandre, le 24 juin*, 1807. Aquarelle, 94 cm x 66 cm. Paris, École militaire.

ABSTRACTS

Français
Louis-François Lejeune (1775-1848), officier du génie pendant les guerres révolutionnaires et napoléoniennes, est surtout connu pour ses peintures de bataille. Sa curieuse contribution, par des dessins originaux, à l'illustration de l'ouvrage d'Alexandre de Laborde (1773-1842), *Voyage pittoresque et historique en Espagne* (1806-1820), révèle toutefois une production paysagère méconnue. Par cette contribution, Lejeune participe paradoxalement à la représentation visuelle du territoire espagnol qu'il a connu par la guerre. Ses vues de Saragosse, notamment, présentent des monuments détruits lors des sièges de la ville (1808-1809), auxquels Lejeune a assisté. Cet article revisite, malgré une documentation lacunaire, la genèse de cette contribution.

English
Louis-François Lejeune (1775-1848), an engineer officer during the Revolutionary and Napoleonic wars, is best known for his battle paintings. His curious contribution of original drawings to the illustration of Alexandre de Laborde's (1773-1842) *Voyage pittoresque et historique en Espagne* (1806-1820), however, reveals a little-known body of landscape work. Through this contribution, Lejeune paradoxically participated in the visual representation of the Spanish territory that he had come to know through war. His views of Zaragoza, in particular, show monuments destroyed during the sieges of the city (1808-1809), which Lejeune attended. Despite a lack of documentation, this article revisits the genesis of this contribution.

INDEX

Mots-clés
Lejeune (Louis-François), paysage, livre de voyage,
sièges de Saragosse (1808-1809), Laborde (Alexandre de)

Keywords
Lejeune (Louis-François), landscape, travel book, Sieges of Zaragoza (1808-1809), Laborde (Alexandre de)

Schlagwortindex
Lejeune (Louis-François), Landschaft, Reisebeschreibung,
Belagerungen von Saragossa (1808-1809), Laborde (Alexandre de)

Palabras claves
Lejeune (Louis-François), paisaje, libro de viajes, sitios de Zaragoza (1808-1809), Laborde (Alexandre de)

Parole chiave
Lejeune (Louis-François), paesaggio, libro di viaggio,
assedio di Saragozza (1808-1809), Laborde (Alexandre de)

AUTHOR

Béatrice Denis
Université de Montréal
IDREF : https://www.idref.fr/281724563

« Échos d'un coin du Chili » : l'Amérique latine à portée de main

« Échos d'un coin du Chili »: Latin America at your fingertips

Ellie Douska

DOI : 10.35562/theia.205

Copyright
CC BY-NC-SA 4.0

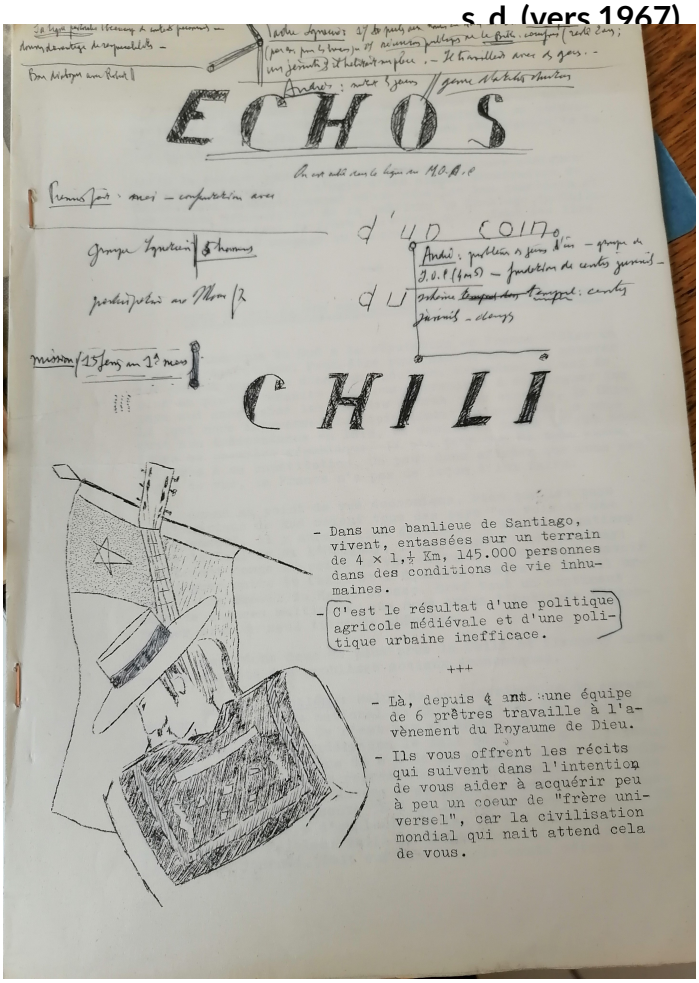
TEXT

1 Les années 1960-70 sont celles d'un véritable moment Amérique latine en Europe, en France en particulier. Les images, les nouvelles, la littérature et la musique véhiculées à travers l'Atlantique ont permis au public européen de (re)découvrir un monde quelque peu oublié [1]. Les travaux participant de l'élaboration de l'histoire transatlantique qui en découle ne manquent pas. Ils se sont même démultipliés ces dernières années, en diversifiant objets et angles d'étude [2]. Nous connaissons ainsi mieux les divers vecteurs et médiateurs qui ont animé, pendant plus d'une décennie, ces circulations transatlantiques multiformes. Rares sont pourtant les travaux qui se sont penchés sur les acteurs *religieux* participant de ces flux. De fait, dans un contexte marqué par la guerre froide, l'Amérique latine ne devient pas seulement un théâtre du duel idéologique mondial, mais aussi le centre de l'attention vaticane. Fort d'une population estimée à plus de 80 % catholique, le sous-continent constitue un enjeu majeur pour l'Église, au moment même où l'« Occident chrétien » semble perdre la foi. Or, ce « trésor catholique » serait menacé : à l'avancée du marxisme athée s'ajoute un déficit de prêtres qui pourrait mettre à mal des siècles d'évangélisation. En 1957, via l'encyclique *Fidei donum* (« le don de la foi »), Pie XII appelle ainsi à une mobilisation générale, invitant les épiscopats européens et nord-américains à prêter certains prêtres diocésains à leurs Églises sœurs en besoin de pasteurs [3]. Nombreux sont les prêtres français qui répondent à l'appel pontifical et qui partent sous l'égide du Comité Épiscopal France Amérique latine (CEFAL), dont l'histoire vient de faire l'objet d'une étude d'envergure [4].

Pour le reste, ces trajectoires restent méconnues, laissant inexploités des fonds et un terrain historiographique très riches. C'était précisément l'objet d'un mémoire soutenu en juin 2024 [5] qui proposait, à l'appui d'un corpus documentaire dormant, de retracer les parcours d'une fraction de ces missionnaires d'un genre nouveau.

2 Bien au-delà de l'histoire religieuse ou missionnaire dans laquelle elles s'inscrivent *a priori*, ces trajectoires cléricales entre deux rives offrent des perspectives analytiques plurielles pouvant contribuer au renouvellement de l'histoire des représentations, de la construction d'imaginaires, des circulations intellectuelles ou culturelles. Partis pour évangéliser, ces prêtres rencontrent un continent et des peuples avec lesquels ils solidariseront. L'émerveillement du départ se meut en désillusion au fur et à mesure qu'ils s'immergent dans les bidonvilles des mégalopoles latino-américaines et qu'ils aperçoivent l'abîme entre deux mondes que tout oppose : celui des « riches » et celui des « pauvres ». Ce qui soulève la question : et eux, à quel monde appartiennent-ils ? La frustration nourrit un besoin de comprendre, de témoigner, tandis que l'indignation née de la rencontre avec la misère les pousse à dénoncer. Ces prêtres prennent alors la plume pour verbaliser ce qu'ils vivent sur le terrain, s'improvisant en journalistes, sociologues, économistes, politologues et moralistes. En découlent des écrits d'une grande diversité dont l'originalité et l'intérêt heuristique tient à leur plasticité : entre récit missionnaire édifiant et discours tiers-mondiste porteur de contestation et moteur d'action. Il en va ainsi des « Échos d'un coin du Chili [6] » (figure 1).

Figure 1. Page de couverture des « Échos d'un coin du Chili », dossier anonyme s. d. (vers 1967)



Archives de l'Institut du Prado (site de Lyon)

3 Présenté sous forme d'une trentaine de pages dactylographiées, ce document a subi l'effet du temps. L'auteur n'a pas signé son œuvre. Or, le titre et son illustration suscitent la curiosité. Les lignes qui les accompagnent sont un appel à la lecture :

> Dans une banlieue de Santiago, vivent entassées sur un terrain de 4x1 ½ km, 145 000 personnes dans des conditions de vie inhumaines. Là, depuis 4 ans, une équipe de 6 prêtres travaille à l'avènement du Royaume de Dieu. Ils vous offrent les récits qui suivent dans l'intention de vous aider à acquérir peu à peu un cœur de « frère universel », car la civilisation mondiale qui nait attend cela de vous [7].

4 Bref et efficace, cet incipit annonce d'emblée la couleur du contenu. Sous couvert de témoignage, c'est une leçon d'humanité qui est proposée. Le propos s'organise en deux parties : il s'ouvre par ce qui

se veut une contextualisation, à travers une simple « énumération de faits ». Si l'auteur assure ne pas avoir l'intention de « faire un bilan critique du Chili », c'est bien à cela qu'il se prête. Le lecteur est juge, mais la conclusion à laquelle il est conduit est univoque : l'histoire du Chili – et de l'Amérique latine – serait celle d'une domination-exploitation perpétuelle où seuls les colonisateurs changent de visage. Il s'agirait en même temps d'une situation de dépendance structurelle aux racines si profondes qu'il est impossible de s'en « libérer ». C'est que dans l'ordre manichéen du monde dressé par l'auteur, le Chili appartient au camp des « petits », « pauvres » et « dépendants », voués à être subjugués par les « grands », « riches » et « libres ». Les chiffres avancés en fournissent la preuve mathématique, tandis que le lexique puisé dans un ensemble de théories socio-économiques transforme le propos en argument d'autorité, tout en donnant un indice sur les affinités idéologiques de l'auteur : un socialisme christianisé, un tiers-mondisme revisité. Sans les citer, c'est aux travaux de Raul Prebisch [8] et du duo Cardoso-Faletto que l'auteur semble s'appuyer, avant même qu'ils ne soient traduits en français [9].

5 Dans un second temps, le lecteur est invité à un « voyage imaginaire gratuit », au cours duquel il sera emmené à la découverte du monde populaire chilien. Comment y renoncer ? Le narrateur promet une description fidèle à la réalité et le récit est certes d'un grand intérêt historique, en ce qu'il offre une vision documentée des réalités du Chili. Mais les dates sont absentes, l'auteur demeure anonyme et les noms des personnages ne permettent aucune identification. On se dit alors qu'il faut changer de lunettes, ou plutôt d'approche. De fait, il ne s'agit pas d'un carnet de bord ou d'un rapport, mais d'un récit pédagogique porteur de leçons universelles et intemporelles. Aux côtés du narrateur, le lecteur découvre les conditions de vie des milieux populaires chiliens. Il y fait la rencontre d'Alberto, l'ouvrier qui lutte pour la survie de sa famille ; Nena, la mère qui fait des miracles culinaires au quotidien avec du thé et du pain ; de Lucho, ce travailleur victime d'exploitation salariale qui représente, nous dit-on, 15 % des Chiliens. Chaque épisode prend les traits d'une parabole biblique censée transmettre un enseignement moral et illustrer une valeur : le partage, l'hospitalité, l'amour du prochain.

6 La visée sous-jacente du document est triple : interpeler les consciences ; dénoncer un ordre social injuste ; appeler enfin à l'action, car « si les deux milliards de voix qui parlent ainsi dans leur silence ne sont pas entendues, elles parleront par la violence ». Se démarquant du récit missionnaire commun, ce document dévoile de manière inédite le conflit d'appartenances que traversent les prêtres Fidei donum et le glissement inhérent de leur terrain action : de la mission religieuse à l'engagement social, voire politique. En permettant de la sorte au public catholique français de découvrir l'Amérique latine par procuration et en le sensibilisant à des enjeux sociaux ou internationaux, les prêtres Fidei donum se font aussi des agents discrets du rapprochement entre deux continents et de l'émergence d'un tiermondisme catholique[10]. La démarche fonctionne grâce à ses artifices littéraires : un langage simple et accessible, des mots choisis pour construire des images aux contrastes frappants. Au charme du Chili répond la misère de ses milliers de pauvres ; à l'exiguïté des espaces, la concentration humaine. La conversion systématique des salaires et des prix à leur montant en francs suggère au lecteur français qu'il pourrait très bien être à la place de son frère chilien. Ce qui ressort est surtout la force d'un peuple qui ne répond pas à la pauvreté par la résignation, mais par l'espoir inébranlable en un avenir meilleur. En montrant ainsi au lecteur la résilience d'un « monde qui marche dans les ténèbres », l'auteur pose indirectement une question : Et vous, que faites-vous ? Le dépliant de collecte de fonds glissé entre les pages du dossier suggère une réponse (figure 2).

Figure 2. Feuillet de collecte de fonds pour une initiative de quartier populaire au Chili, vers 1965



REGARDEZ BIEN CETTE PHOTO...

Il s'agit d'une vue partielle d'un quartier ouvrier de la banlieue de Santiago du Chili. Là, des milliers de jeunes connaissent la misère et l'incertitude du lendemain, parce qu'ils n'ont pas de métier en mains et aucune chance d'en avoir un, un jour.

CEPENDANT

dans toutes les villes du Chili des jeunes veulent apprendre un métier et se sont organisés en « **Foyers de Jeunes** ».

Avec l'aide d'ouvriers spécialisés ou de techniciens ils ont organisé des **cours d'apprentissage.**

Mais ces cours coûtent chers et ces jeunes ont peu de moyens...

C'EST POURQUOI...

vous pouvez, VOUS, collaborer au succès de cet éveil populaire en envoyant un peu d'argent aux jeunes travailleurs et travailleuses du Chili.

Vous pouvez le faire, soit en adressant un virement postal ou un mandat au C. C. P. suivant : M. l'Abbé André Lanson, 288-38 Orléans (Loiret), soit en remettant votre offrande à la personne qui vous aura donné ce prospectus.

Les Jeunes du Chili vous remercient.

Archives de l'Institut du Prado (site de Lyon)

NOTES

1 Mona Huerta, « Le latino-américanisme français en perspective », *Caravelle*, nº 100, 2013, p. 39-62.

2 Pour un aperçu historiographique, voir entre autres : Olivier Compagnon, Anaïs Flechet, Gabriela Pellegrino Soares, « Écrire une histoire culturelle transatlantique (XVIII^{e}-XXI^{e} siècles) : enjeux, défis, méthodes », *Diogène*, vol. 258, nº 2, 2017, p. 237-250 ; les publications de l'Institut des Amériques et les travaux du groupe TRACS.

3 On appelle ainsi « prêtres Fidei donum », les prêtres diocésains qui répondent à l'appel pontifical et quittent leurs diocèses d'origine pour se mettre au service des Églises du tiers-monde.

4 Olivier Chatelan, *Des allers sans retours ? Les prêtres français en Amérique Latine (1961-1985)*, LARHRA, 2023, 2 vol.

5 Ellie Douska, *Les prêtres du Prado envoyés au Chili à titre de « prêtres Fidei donum » (1957-1990). Micro-histoire d'une rencontre et d'une conversion*, mémoire de Master préparé sous la direction d'Olivier Chatelan, Université Lyon III, 2024.

6 Archives du Prado (site de Lyon), Chili I, « Échos d'un coin du Chili », anonyme, s. d. (vers 1967).

7 *Ibid.*

8 Économiste argentin, Raul Prebisch a joué un rôle de premier plan au sein de la Commission Économique pour l'Amérique Latine (CEPAL) dès sa fondation en 1948. Son premier rapport, *El desarrollo economico y algunos de sus principales problemas*, fait date dans la critique de la division internationale du travail alors prédominante et des rapports Nord-Sud, tout en posant des éléments centraux de la théorie développementiste diffusée par la CEPAL. L'auteur y dénonce notamment les relations entre pays en introduisant les notions de centres-périphéries et de déséquilibres structurels. À ce sujet voir : Anne-Sophie SAVIGNAT, « Les premiers travaux de Raul Prebisch à la CEPAL », *Mondes en développement*, 2001, n° 113-114, p. 13-14.

9 L'ouvrage *Dépendance et développement en Amérique Latine* des auteurs Fernando Henrique Cardoso et Enzo Faletto a constitué une rupture dans la vision développementiste élaborée et diffusée par la CEPAL et ses principaux économistes depuis les années 1950, tout en posant les bases pour ce qui sera ultérieurement connu comme la « théorie de la dépendance ». Rédigé entre 1962 et 1967 sous le titre *Dependencia y desarrollo en America Latina*, l'ouvrage n'est traduit en français et diffusé par les Presses Universitaires de France qu'en 1978.

10 Denis Pelletier, « De la mission au tiers-mondisme: crise ou mutation d'un modèle d'engagement catholique », *Le Mouvement social*, 1996, n° 177, p. 3-8.

ABSTRACTS

Français
Les années 1960-70 sont celles d'un véritable « moment Amérique latine » en Europe. Les images, les nouvelles et toutes les productions culturelles qui sont véhiculées à travers l'Atlantique permettent au public européen de découvrir un continent un peu oublié, de s'éprendre de ses cultures, de suivre avec intérêt les évènements qui ébranlent ses sociétés. C'est que dans un contexte marqué par la guerre froide, le sous-continent américain devient le théâtre des tensions géopolitiques et idéologiques internationales. Alors que la crainte du communisme devient une véritable obsession, les voisins du Nord ne sont pas les seuls à se mobiliser. Le Vatican se soucie également du devenir d'un continent par excellence catholique. L'actualisation de la mission traditionnelle dans une logique d'endiguement du marxisme athée se solde, en 1957, par un appel pontifical à la mobilisation des épiscopats européens en faveur des diocèses pauvres du tiers-monde. Nombreux sont les prêtres français qui y répondront et qui partiront se mettre au service des Églises d'Amérique latine. Au-delà de la mission religieuse qui leur a été confiée, leur histoire est celle d'une rencontre avec un monde et de ses défis qu'ils voudront faire découvrir au public français. L'objectif de cet article est de présenter un document inédit qui découle de cette expérience originale, un témoignage qui se veut aussi un récit édifiant et un appel à la solidarité fraternelle avec un monde victime d'injustice. Indirectement, il offre également un regard documenté du Chili des années 1960 et permet au chercheur d'aborder une période complexe sous plusieurs angles.

English
The 1960s and 70s saw the emergence of a "Latin American moment" in Europe. The images, news and music transmitted across the Atlantic enabled European audiences to discover a somewhat forgotten continent, to grow fond of its cultures, and to follow with close interest the events that challenged its nations. In a context marked by the Cold War, the American subcontinent became the stage of international geopolitical and ideological tensions. As the fear of communism turned into a true obsession, Northern American neighbours weren't the only ones concerned. The Vatican became equally concerned about the future of a predominantly Catholic subcontinent. The actualisation of religious mission into a means of combatting Marxism led, in 1957, to a papal plea urged European episcopates to mobilize their efforts in favour of the poor dioceses of the Third World. Many French priests responded, leaving their dioceses to serve those of Latin America. Beyond the religious mission entrusted to them, their stories are those of an encounter with a world they learned to embrace, and which they wanted the French public to discover. The aim of this article is to present an original document that stems from this original missionary experience: a testimony that is also an educational essay and a call for fraternal solidarity with a world subjected to injustice. It also

indirectly offers a documented view of Chile in the 1960s, enabling to approach a complex period from a variety of perspectives.

INDEX

Mots-clés
XXe siècle, Chili, Amérique latine, circulations, Église catholique

Keywords
20th century, Latin America, Chile, Catholic Church

AUTHOR

Ellie Douska
Université Jean Moulin Lyon 3, LARHRA UMR 5190
IDREF : https://www.idref.fr/281724520

Victorine Meurent, modèle et artiste. Relecture d'un mythe de la décadence artistique

Victorine Meurent, model and artist. Rereading a myth of artistic decadence

Adeline Dollet

DOI : 10.35562/theia.216

Copyright
CC BY-NC-SA 4.0

TEXT

1 Au regard de l'histoire de l'art, Victorine Meurent (1844-1927) n'a été qu'un nom, une anecdote dans l'œuvre de l'artiste de la modernité Édouard Manet (1832-1883). Selon l'historienne de l'art britannique Frances Borzello, cette femme artiste et modèle est l'exemple parfait du traitement marginal accordé aux modèles d'atelier qui ont servi à la représentation en peinture [1]. À la croisée de l'histoire sociale de l'art, des études de genre et des modèles vivants, cette recherche a eu pour ambition de rassembler l'ensemble des connaissances consacrées à la vie et à l'œuvre de Victorine Meurent [2]. Cette étude a permis d'offrir une nouvelle compréhension aux sources archivistiques et aux œuvres pour lesquelles Meurent a posé. À cela s'ajoute l'analyse technique et iconographique de quatre tableaux de l'artiste exposés au musée municipal d'art et d'histoire de Colombes [3] et au Museum of Fine Arts de Boston [4]. Ces recherches ont surtout permis d'accéder à un manuscrit non publié du journaliste Adolphe Tabarant (1863-1950).

2 Tabarant est considéré comme l'un des premiers historiens de l'impressionnisme français [5]. Rédacteur du *Bulletin de la vie artistique* et proche des artistes néo-impressionnistes, il écrivit de nombreux articles sur la peinture de cette période. En 1931, puis en 1947, il publie deux ouvrages sur Édouard Manet, fruit d'années de recherche et d'entretiens qui constituent la source de toute étude sur l'artiste [6]. Ces ouvrages réunissent la biographie du peintre et la première tentative de catalogue raisonné de ses œuvres, proposant une

réflexion sur le processus de création, les influences, mais également la réception de ses œuvres. Tabarant est la principale source écrite consacrée à la vie de Manet, souvent citée dans les études sur l'artiste et dans les expositions qui lui sont consacrées [7].

3 Le document d'archive étudié est un manuscrit que Tabarant a rédigé entre octobre 1948 et mars 1949. Il s'agit plus précisément d'une copie de l'original tapuscrit, qui a disparu à ce jour. L'accès à ce document a été rendu possible grâce à Eunice Lipton, historienne de l'art féministe étasunienne. En 1988, Lipton obtient un fac-similé du document par le biais de l'assistante de recherche de la précédente propriétaire du fonds d'archives de Tabarant, Mina Curtiss. Lipton cède en 2023 cet exemplaire à la Morgan Library & Museum de New York, au moment où nous effectuons des recherches sur l'œuvre et la vie de Meurent [8].

4 L'analyse de ce document nous fait se confronter à une réalité complexe. D'un côté, s'impose la figure de Tabarant comme source reconnue de l'histoire de l'art, depuis la publication de la biographie *Manet, histoire catalographique* publiée en 1931. Cependant, ce document ne présente pas les qualités des précédents ouvrages de l'auteur, n'étant pas une source historique fiable de première main. Tabarant répète des éléments qu'il a déjà évoqués en 1931 et en 1947, à partir de sa réflexion personnelle, de ses hypothèses et des témoignages de l'époque. Les sources de Tabarant sont difficilement identifiables, en dehors des propos de Léon Koëlla-Leenhoff (1852-1927), fils de Suzanne Manet, la veuve du peintre, de l'artiste Suzanne Valadon (1865-1938), ainsi que des *Mémoires* du romancier et critique d'art Georges Moore (1852-1933). Ces propos rapportés sont un gage de véracité, puisque ceux de Koëlla sont associés au lien familial avec Manet et ceux de Valadon, à son statut de personnalité artistique de Montmartre et proche de Tabarant. S'agissant des propos de Moore, ils ne peuvent pas être considérés comme une source fiable, car l'auteur anglais a eu tendance à modifier les récits et les dates des évènements au gré des publications [9]. L'ensemble des propos rapportés par Tabarant s'apparentent ainsi à des « on-dit », des rumeurs montmartroises datées [10] que les deux principaux témoins confirment et alimentent :

> Je vais conter, sans en rien taire, une bouleversante histoire dont le déroulement dépasse l'horizon du romanesque, une authentique histoire qui démontre que la réalité peut quelques fois défier la fiction. [11]

5 D'un point de vue rédactionnel, le manuscrit s'oppose totalement aux précédents ouvrages. Tabarant est omniprésent en tant qu'auteur-narrateur. Les premières et dernières phrases du manuscrit le démontrent. Il se présente comme le conteur d'une histoire, celle de Meurent artiste et modèle dont la vie a longtemps été tue et qu'il révèle au grand public. Plus étonnant encore pour une démarche historique, Tabarant affirme répondre à un devoir, celui de raconter la vérité, de partager l'histoire de Meurent telle qu'il la connaît. Il se présente en homme accablé par ses informations et qui doit se soulager de ses découvertes.

> Miette après miette, je nourrissais mon dossier. Et je faisais d'affreuses découvertes. Les rendrais-je publiques. […] dans Manet et ses œuvres, j'en disais moins encore, m'arrêtant effrayé au bord du gouffre que j'avais découvert. […] J'y suis enfin descendu. M'imposant la tâche d'évoquer Victorine Meurent à toutes ses heures et toute entière, je suis allé droit à cette héroïne jusque-là si indéterminée, attirante et repoussante. [12]

6 Les recherches postérieures au récit de Tabarant démontrent qu'il fait erreur sur l'identification de l'artiste et modèle. Il présente Meurent comme une fille de Montmartre dont le « millésime de sa naissance est imprécis » [13]. En 1967, l'étude de l'historien Jacques Goedorp découvre que Meurent est finalement née le 18 février 1844 à Paris dans le quartier ouvrier de Popincourt. Son enfance semble plutôt se situer sur la rive gauche dans le quartier de la place Maubert [14]. Plus important encore, Tabarant situe le décès de Victorine Meurent en 1892, alors qu'elle réside encore à Paris jusqu'en 1906, puis à Colombes où elle décède finalement en 1927 [15].

7 Quant à la carrière de peintre de Meurent, les propos du narrateur-auteur démontre surtout sa profonde admiration pour la peinture de Manet. La vie personnelle de Meurent prend le pas sur sa création artistique et le jugement de Tabarant conduit à une interprétation biaisée et fausse, comme le démontre l'extrait suivant :

> […] Elle faisait de la peinture, disait-elle. Courant à quelques académies, elle circulait avec un carton à dessins sous le bras, des toiles, une boîte à couleurs. On sut que celle qui eût pu recevoir l'enseignement de Stevens se rendait rue Turgot à l'atelier d'Étienne Leroy, obscur peintre de portraits et de genre qui exposait irrégulièrement au Palais de L'Industrie. Quelle étrange fille ! [16]

8 À l'image de nombreuses femmes artistes, Victorine Meurent se tourne en effet vers les académies privées. Elle suit les cours du soir de l'Académie Julian, académie privée la plus renommée de Paris. Elle est présente dans les différents livrets listant les élèves féminines de l'Académie depuis 1868 [17]. Dans les catalogues des Salons de 1876, 1879, 1885 et 1904, elle est présentée comme l'élève d'Étienne Leroy (1828-1876). Peintre portraitiste, élève de François-Édouard Picot (1786-1868), exposant au Salon chaque année entre 1857 et 1873 [18], Leroy ne correspond pas à l'esthétique défendue par Tabarant. Ses œuvres sont appréciées des critiques de Salon, notamment Edmond de Laqueuille et Gonzague Privat qui vantent ses choix chromatiques, mais fustigent la position de ses œuvres au Salon [19]. Cet investissement à l'Académie Julian et auprès de Leroy témoigne d'une carrière professionnelle déjà engagée à cette période, où Meurent pose en parallèle pour Manet. Rappelons enfin que Meurent reste une peintre active jusqu'en 1913, au moins selon les registres de Salon et les annuaires artistiques que notre étude a mis à jour [20].

9 Ces observations démontrent que ce manuscrit peut difficilement être étudié comme porteur d'une vérité historique, mais bien comme le témoignage de la construction d'un mythe, dont Meurent fera les frais tout au long de sa carrière artistique. Par ailleurs, l'étude de Tabarant s'inscrit dans un contexte particulier : le transfert de l'*Olympia* (1863) de Manet au musée du Jeu de Paume en 1947. Ce contexte institutionnel entraine des discussions et des interrogations sur le modèle du tableau. Le texte de Tabarant s'assimile donc bien plus à une tentative d'écrire au sujet du tableau, plutôt que de traiter de Victorine Meurent. Tabarant résume par ces mots le récit qu'il partage : « L'histoire d'une créature féminine, longtemps parées d'attraits, belle, intelligente, artiste, et qui sombra dans la pire ignominie. » [21] Assimilée à Olympia, Meurent devient la figure mythique de la fille des rues qui connaît la lumière grâce à un artiste

de génie et qui, lorsqu'elle s'en détache, ne rencontre que des échecs à travers les choix qu'elle fait. Si le manuscrit resté inédit ne peut constituer la source unique dans la construction de ce mythe décadent, les précédents écrits de Tabarant et les sous-entendus sur la vie de misère de l'artiste auront contribué à le forger. Cette image se renforce à travers les interprétations sur l'hypersexualisation et les prétendus vices cachés des femmes modèles, que l'on retrouve dans la presse de l'époque, dans les écrits de Tabarant et à travers ce corps offert au spectateur qu'est l'Olympia de Manet.

APPENDIX

« Celle qui fut l'Olympia de Manet »

« Je vais conter, sans en rien taire, une bouleversante histoire dont le déroulement dépasse l'horizon du romanesque, une authentique histoire qui démontre que la réalité peut quelques fois défier la fiction. L'histoire d'une créature féminine, longtemps parées d'attraits, belle, intelligente, artiste, et qui sombra dans la pire ignominie. L'histoire de cette Victorine Meurent que représentent plusieurs figures peintes par Édouard Manet, entre lesquelles la fameuse Olympia. » [22]

« Elle chantait dans les rues et les cabarets en s'accompagnant de la guitare, et c'était son gagne-pain, cette quête errante n'humiliant pas sa fière humeur. Elle recevait des sous sans dire merci. Bien que je connaisse rien [sic] des premiers ébats qu'elle avait pu prendre, j'ai l'impression que tout en étant de petite vertu, elle savait se défendre contre les pièges du pavé. » [23]

« La gravité de son profil vieillissait un peu cette blonde ardente, mais la face entière, où vivaient des yeux hardis, attestait une printanière jeunesse, avec cela le corps nerveux des filles de Paris, suggestif en chacun de ses détails. Le buste élancé faisant valoir le style discret de ses seins. Un œil de peintre est prompt à parcourir ces perspectives. » [24]

« Nouveau tournant ! Celle qui avait posé la jeune femme de ce tableau, celle qui avait paru dans la Partie de crocket [sic], venait encore une fois de s'évader du milieu où gravitait le cordial Stevens, non pour un retour en Amérique : elle ne s'était pas envolée, mais pour vivre à l'écart, s'y créer des

habitudes diamétralement opposées à celles qu'on lui avait connues. […] Elle faisait de la peinture, disait-elle. Courant à quelques académies, elle circulait avec un carton à dessins sous le bras, des toiles, une boîte à couleurs. On sut que celle qui eût pu recevoir l'enseignement de Stevens se rendait rue Turgot à l'atelier d'Étienne Leroy, obscur peintre de portraits et de genre qui exposait irrégulièrement au Palais de L'Industrie. Quelle étrange fille ! » [25]

« Il y eut presque unanimité pour exclure Manet. Eh bien ! Ce jury bannisseur de Manet accepta un envoi fait par … Victorine Meurent ! Victorine, oui, recommandée, croit-on, et c'est là le petit mystère, par l'inconnu qui en payait les faveurs. Un monsieur évidemment bien en place. On allait pouvoir lire sur le livret du Salon : "Meurent (Melle Victorine), née à Paris, - Boulevard de Clichy, 1, Portrait de l'auteur." Son propre portrait peint selon les recettes d'Étienne Leroy, et qu'il ne m'a pas été possible de retrouver. » [26]

« Il [Étienne Leroy] lui fit peindre sous ses yeux une toile s'adaptant aux sympathies traditionnelles d'un jury, et qu'elle soumit à celui de 1879, que présidait Bonnat. Cela passerait-il, "en douce", comme avait passé l'envoi de 1876 ? Eh bien, oui, cela passa ! Au livret du Salon de 1879, on pourrait lire, en regard du numéro 2128 : "Meurent (Melle Victorine-Louise) née à Paris. Élève de E. Leroy. Rue Bréda 21". Suivrait - ô désolation ! - le titre de l'œuvre : Bourgeoise de Nuremberg au XVIe siècle. Qu'un tel sujet eût pu être choisi par celle qui avait approché de si près la plus moderne conception de l'art, n'était-ce pas stupéfiant ? » [27]

« Elle [Marie Pellegrin, proche de Meurent] l'aperçoit, se jette dans ses bras, l'étreint, la dévore goulûment. Coup de foudre qui s'amplifie, l'ivresse dissipée. Elle jura, elle fit jurer à Victorine qu'elles ne se quitteraient plus. … Elle arrêta de faire sien le logis de Victorine, y transporta son minier, y répandit des fleurs, une crémaillère du nouveau ménage étant bruyamment plantée.

Le 1 du boulevard de Clichy, au troisième, devint le rendez-vous d'une femellerie [sic] spéciale, filles à marlous et filles pour filles, et du jour comme de nuit la maison retentit de rires et de gueulades, l'entente ne régnant pas à perpète entre ces dames, qui s'embrassaient et s'arrachait les cheveux simultanément. » [28]

« En 1881, à bout d'expédients, elle prit la résolution, qui lui coûtait, d'en appeler à Manet, ceci m'ayant été dit par Léon Koëlla. Elle ne le voyait plus,

Manet ? S'il n'avait personnellement rien contre elle, il n'évitait pas moins de la croiser sur sa route. (…) Il la reçut. Pouvait-elle n'être plus pour lui la forme vivante d'œuvres glorieuses ? Forme en dégénérescence. L'éblouissant édifice de chair du Bain, l'offertoire de luxure de l'Olympia, qu'étaient devenues en elles ces trois périssables choses ? Il eut pitié. Il lui ouvrit sa bourse. Il s'occuperait de lui avoir un petit emploi peu accaparant, celui d'ouvreuse étant indiqué par elle. Il en parlerait à des amis, gens de théâtre. En parla-t-il ? Le certain, c'est qu'il ne l'oublia pas, que plus d'une fois il lui fit passer quelques secours. » [29]

« Cette entrée de Victorine dans la galanterie ouverte, se place vers la fin de 1886, et je dois cette indication nette - je vais lui en devoir tout un lot - à ma grande amie Suzanne Valadon, qui ouvrit de bonne heure des yeux vifs, cueillant tout, sur le Paris diurne et nocturne qui s'ébrouait devant elle. Valadon, qui me fut bien chère, avec qui je fus tant bavard du passé. » [30]

« C'est à partir de 86 qu'on la vit tous les soirs à l'Élysée-Montmartre, continuait Valadon. Soi-disant comme artiste. En 87, elle y était tout à fait putain. » [31]

« Elle n'était que de seconde jeunesse, l'artiste exposant. Il ne voyait pas en elle une marcheuse. Il fallait que Victorine joua le jeu serré de l'honnêteté en proie aux cruautés de la vie pour qu'il prit quelque chaleur. Confidences à l'oreille ? Elle entrait dans le vif et le monsieur commençait à mordre. Il était le premier à faire des avances. Il pressait, elle résistait. Vaincue, elle disait oui, gagnait la sortie sans en avoir l'air, suivie à peu de distance par le partenaire impatient. La conclusion s'expédiait dans un hôtel du voisinage. » [32]

« La date extrême des échos que j'aie recueillis atteint bien juste 1892. Après cela, parier pour la vie, c'est jouer à pile ou face. N'aurait-elle pas été ramassée par la police et recluse au dépôt de mendicité, qui la guettait ? Réponse négative de la Préfecture de Nanterre. Trépassée ? Une terminaison sans cérémonie, qui ne se serait pas produite dans le trou de chiennerie où croupissait la vivante. Les avis émanés des vieilles cervelles - avis obtenus par moi - s'accordaient sur la mort vers le terme indique, environ 92. » [33]

« Olympia au Louvre, ce qui amena le nom de Victorine Meurent sur quelques lèvres, ce qui me le fit écrire dans un quotidien, trot [sic] de ma plume.

Le nom, et rien outre [sic]. Qu'aurais-je pu noter touchant la personne ? Qu'en savais-je ? Qu'en savait-on ?

Il faut qu'ici je dise une dernière fois, en mettant les points sur les i, il faut que je fasse comprendre qu'en 1907 comme en 1892, Victorine Meurent n'avait pas d'existence dans le souvenir, et que le fait d'écrire son nom ou de le prononcer ne pouvait être qu'un réflexe accidentel. » [34]

« J'y révélais [35] - et c'était bien une révélation - qu'elle avait été peintre et qu'en 1876 le salon avait reçu d'elle et présenté un portrait de femme, le sien. Je ne parlais pas de sa contribution à deux autres Salons, et pour cause. Je l'ignorais, et nul n'était à même de me le signaler.

Patience ! Le vrai venait lentement à moi par des routes souvent interrompues. Miette après miette, je nourrissais mon dossier. Et je faisais d'affreuses découvertes. Les rendrais-je publiques.

Dans mon Manet, histoire catalographique en 1931, je dénonçais la tare d'ivrognerie et je m'en affligeais. Le 10 juillet 1932, publiant dans l'Œuvre, (…) je ne dissimulais pas ce que j'appelais "la pire déchéance", tout en m'interdisant de la rapporter. » [36]

« Et dans Manet et ses œuvres, j'en disais moins encore, m'arrêtant effrayé au bord du gouffre que j'avais découvert.

J'y suis enfin descendu. M'imposant la tâche d'évoquer Victorine Meurent à toutes ses heures et tout entière, je suis allé droit à cette héroïne jusque-là si indéterminée, attirante et repoussante. J'ai décisivement déroulé le récit de cette vie de damnation que tout menait au désastre. On vient de la lire et je n'ai pas à y ajouter. » [37]

NOTES

1 Frances Borzello, *The Artist's Model*, Londres, Junction Books, 1982, p. 115.

2 Adeline Dollet, *Victorine Meurent (1844-1927), modèle et artiste*, mémoire de master 2 Histoire de l'art – Arts et cultures visuelles, sous la direction de Damien Delille, soutenu en 2023, Université Lumière Lyon 2.

3 Victorine Meurent, *Le Jour des Rameaux*, 1886, huile sur toile, 41,3 x 32,2 cm, Colombes, Musée Municipal d'Art et d'Histoire ; Victorine Meurent, *Le Briquet* ou *Le Gamin de Paris*, 1896, huile sur toile, 46 x 38 cm, Colombes,

Musée Municipal d'Art et d'Histoire ; Victorine Meurent, *Jup*, non daté, huile sur toile, 24 x 19 cm, Colombes, Musée Municipal d'Art et d'Histoire.

4 Victorine Meurent, *Autoportrait de l'artiste*, huile sur toile, 35 x 27 cm, Boston, Fine Arts Museum.

5 Jules Colmar (dir.), *Le cahier rouge des impressionnistes*, Paris, Grasset, 2019, p. 104.

6 Adolphe Tabarant, *Manet : histoire catalographique*, Paris, Éditions Montaigne, 1931 et Adolphe Tabarant, *Manet et ses œuvres*, Paris, Gallimard, 1947.

7 Tabarant est notamment l'une des sources utilisées pour l'exposition *Manet/Degas*, réalisée par le Musée d'Orsay du 28 mars au 23 juillet 2023. Voir Laurence Des Cars, Stéphane Guégan, Isolde Pludermacher (dir.), *Manet/Degas* [catalogue d'exposition, Paris, Musée d'Orsay, du 28 mars au 23 juillet 2023], Paris, Gallimard, 2023.

8 Adolphe Tabarant, « Celle qui fut l'Olympia de Manet by Adolphe Tabarant », non publié, Record ID 447552, Morgan Library & Museum, New York.

9 Margaret Armbrust-Seibert, *A biography of Victorine-Louise Meurent and her role in the of Édouard Manet*, thèse de doctorat, sous la direction de Mathew Herban III, Ohio State University, 1989, p. 306-307.

10 Emmanuel Laurent, *Mademoiselle* V, *Journal d'une insouciante*, Paris, La Différence, 2003, p. 264.

11 Adolphe Tabarant, art. cit., p. 1.

12 *Ibidem*, p. 83-83.

13 *Ibid.*, p. 1.

14 Jacques Goedorp, « La fin d'une légende : L'Olympia n'était pas une montmartroise », *Le journal de l'amateur de l'art*, 10-25 février 1967, p. 7.

15 Eunice Lipton, *Alias Olympia, a woman search for Manet's notorious model & her own desire*, New York, Meridian, 1992, p. 161-163.

16 Adolphe Tabarant, art. cit., p. 38.

17 Fonds de l'Académie Julian, Ateliers de l'Académie et divers, 63AS/13, Archives nationales de France, Pierrefitte.

18 Emmanuel Bénézit, *Dictionnaire des peintres, sculpteurs, dessinateurs et graveurs*, vol. 8, Paris, Gründ, 1999, p. 558.

19 Edmond de Laqueuille, *Les Beaux-Arts : revue nouvelle*, 1er juillet 1861, vol. 2, tome 3, p. 49 et Gonzague Privat, *Place aux jeunes, causeries critiques sur le Salon de* 1865, Paris, F. Cournot, 1865, p. 150.

20 *Annuaire de la curiosité et des Beaux-Arts : Paris, départements, étrangers*, Paris, s. n., 1913, p. 474.

21 A. Tabarant, art. cit., p. 1.

22 Adolphe Tabarant, « Celle qui fut l'Olympia de Manet by Adolphe Tabarant », non publié, Record ID 447552, Morgan Library & Museum, New York, p. 1. Extraits proposés et commentés par Adeline Dollet, pour l'ensemble des extraits.

23 *Ibid.*, p. 1-2.

24 *Ibid.*, p. 2. La gravité du visage de Meurent rappelle le portrait de Manet réalise autour de 1862, conservé au Museum of Fine Art de Boston. Considéré comme la première représentation de Meurent par Manet, il s'agit d'un petit tableau dans lequel apparaît clairement le regard qui deviendra l'attribut de Meurent. Tabarant n'ayant jamais rencontré Victorine Meurent, cette description teintée d'érotisme pointe les caractéristiques de Meurent en tant que modèle grisette, à savoir, un corps chétif faisant plus vieux que son âge et une petite poitrine. Cette description prépare le lecteur à la suite du récit.

25 *Ibid.*, p. 38.

26 *Ibid.*, p. 39, commentaire de A. Tabarant suite au Salon de 1879. Victorine Meurent est exposante avec son tableau *Bourgeoise de Nuremberg, au XVIème siècle* (localisation inconnue), en tant qu'élève d'Étienne Leroy.

27 *Ibid.*, p. 50, au sujet du Salon de 1879.

28 *Ibid.*, p. 41, au sujet du bar Le Rat mort.

29 *Ibid.*, p. 52, au sujet de la précarité de Meurent qui déménage au début des années 1880 au 21, rue Bréda à Paris, dans un appartement à 280 francs par an.

30 *Ibid.*, p. 58.

31 *Ibid.*, p. 60.

32 *Ibid.*, p. 62.

33 *Ibid.*, p. 78. A. Tabarant se trompe sur la date de décès de Meurent. Dans les années 1890, Meurent est toujours présente à Paris et change

régulièrement de logement. Elle expose au Salon de 1904, alors qu'elle a soixante ans. En 1906, elle quitte Paris pour s'installer à Colombes. Elle figure aux recensements de la commune de 1906, 1907 et 1921. Toujours en 1906, elle devient sociétaire de la Société des Artistes Français et reçoit des aides financières en 1909 et 1919. Elle est citée dans *L'Annuaire de la Curiosité et des Beaux-Arts* en 1912 et 1913, ce qui démontre qu'elle est encore une peintre active à un âge très avancé. Son acte de décès dans les archives communales de Colombes indique le 17 mars 1927, morte à l'âge de 83 ans.

34 *Ibid.*, p. 81.

35 *Bulletin de la vie Artistique*, 15 mai 1921, p. 17-19.

36 *Ibid.*, p. 82.

37 *Ibid.*, p. 83, au sujet des ouvrages Adolphe Tabarant, *Manet, histoire catalographique*, Paris, Éditions Montaigne, 1931 et *Id.*, *Manet et ses œuvres*, Paris, Gallimard, 1947.

ABSTRACTS

Français

Adolphe Tabarant, journaliste et critique d'art, a publié en 1931 et 1947 deux ouvrages qui constituent une source de premier ordre dans l'étude de l'œuvre et de la vie du peintre Édouard Manet. En 2023, la Morgan Library and Museum de New York a acquis une photocopie d'un manuscrit inédit et non publié de Tabarant. Ce manuscrit intitulé « Celle qui fut l'Olympia de Manet » se présente comme un récit biographie de Victorine Meurent, modèle préféré de Manet pour qui elle a posé entre 1862 et 1872. En rapport avec le mémoire de master consacrée à Meurent, l'objectif de cet article est d'évaluer certains éléments du manuscrit qui permettent d'établir sa valeur historique dans l'élaboration du mythe du modèle décadent, associé à la figure de Meurent.

English

Adolphe Tabarant, journalist and art critic, published two books in 1931 and 1947 that constitute a first-rate source for the study of the work and life of the painter Édouard Manet. In 2023, the Morgan Library and Museum in New York acquired a photocopy of an unpublished manuscript by Tabarant. The manuscript, entitled "Celle qui fut l'Olympia de Manet," is a biographical account of Victorine Meurent, Manet's favorite model, for whom she posed between 1862 and 1872. In connection with the master's thesis devoted to Meurent, the aim of this article is to evaluate certain elements of the manuscript that make it possible to establish its historical value in the

development of the myth of the decadent model, associated with the figure of Meurent.

INDEX

Mots-clés
Meurent (Victorine), Manet (Édouard), Tabarant (Adolphe), modèle, femme peintre, féminisme

Keywords
Meurent (Victorine), Manet (Édouard), Tabarant (Adolphe), model, woman painter, feminism

AUTHOR

Adeline Dollet
INHA
IDREF : https://www.idref.fr/281724393

Remise du diplôme de Doctorat *Honoris Causa* à Orhan Pamuk, écrivain et prix Nobel de Littérature, Université Lumière Lyon 2, 24 juin 2024

On love and museums

Orhan Pamuk

DOI : 10.35562/theia.230

Copyright
CC BY-NC-SA 4.0

TEXT

1 Four years after the publication of my novel *The Museum of Innocence* in late August 2008, I opened the actual Museum of Innocence in the same building in Çukurcuma where Füsun and her family used to live. Those wishing to visit the museum can take their copy of the novel along and enter for free by having the ticket page 485 stamped at the door; inside, they will find Füsun's earrings, hairclips, dresses, shoes, driver's license, and many more of her belongings on display. As well as Füsun's personal possessions, the museum, which opened in 2012, also features clothes and day-to-day items worn and used by Füsun's mother Nesibe Hanım and her father Tarık Bey, by Kemal and his father, and by many other characters in the novel, the little gifts they exchanged, their cigarette stubs, their lighters, and a vast array of images, photographs, video clips, and postcards of Istanbul from the time when the novel is set.

2 I started thinking of the novel *The Museum of Innocence* and of the museum itself at the same time, starting in the mid-1990s. I planned to write a love story while also creating a museum where I could exhibit the objects that appeared in that story. I have talked at length about these objects and every display cabinet in the museum in the illustrated volume *The Innocence of Objects*, in which I have described how the idea for the museum evolved over time. Now I would like to say a few words about how the novel itself developed, about objects, and about the love story in this book.

3 I will begin with the location of the museum and of Füsun's home. Between 1996 and 2000, I would drop my daughter Rüya off to school every morning. From the entrance of her school in the depths of Tophane (some 300 metres from the home of the Keskins), I would

walk through the back alleys of Beyoğlu, Çukurcuma, Firuzağa, and Cihangir towards my office, thinking all the while about what I was going to write that day (*My Name is Red*, say, or *Snow*). I loved roaming the streets in the cool air of the morning, as the shops slowly began to open their doors, the smell of bread and traditional Turkish sesame bagels wafted out of the bakeries, and schoolchildren quickened their steps to make it to class in time. Perhaps it was because I knew I had a happy day ahead of me, a page or two of a novel to write… There was also the satisfaction of noticing that so many of the things I remembered from my childhood and youth still endured, and had neither fallen into complete decay, nor been veneered into some artificial appearance of glossiness. I would tell myself that those streets and their people would never lose the air of timelessness that distinguished them. The sights I encountered in these neighbourhoods – the piles of oven-fresh loaves and bagels displayed on bakery shelves; the rather ancient-looking painkiller advertisement I spotted in a pharmacy once, showing the human body and its internal organs; the colours of the vast jars arranged in neat rows in the window of the pickle shop – would remind me of the delights of seeing and observing; I found myself wanting to possess those scenes, to put them in a frame and study them, and to make sure that I would never lose them.

4 During the time it took me to write my novel and open the museum, Çukurcuma's humble flea market continued to grow. There were stalls selling all sorts of odds and ends, from old tables to ashtrays and cutlery sets, from the locally manufactured toys I remembered from my childhood to quince graters and saltshakers. Shops specialising in old magazines, books, maps, and photographs also fuelled my growing desire to frame everything I saw and preserve it forever. It was around this time that I began to think about buying items from these shops and using them and other objects I had collected from my family and from my own life to create a 'house-museum'.

5 There was a period when I spent a long time walking around these streets in search of an old house that I could transform into a museum. When I managed in 1999 to find such a building (where the museum is now housed), the aspiring collector inside of me gained courage. Yet I knew that I lacked the spirit of a true collector.

Whenever I bought a used saltshaker, a cigarette holder, a meter stripped from an old taxi, or a bottle of eau de cologne I had seen in a shop window, I did not do so in order to create a collection out of these objects, but because I wanted to incorporate them into the novel I was writing. It made me happy to buy all these objects I planned to use and place them upon my writing desk. Sometimes I would see something in a display that I hadn't even thought of using before, and in my eagerness, I would buy it anyway and take it home with me.

6 The world was teeming with objects I could put in my novel and museum. But my excitement was not that of a collector seeking to complete a series; rather, it was the head-spinning enthusiasm of a writer-artist dreaming of turning those objects into components of a novel and a museum. Like so many other things in my life, I became attached to these objects because I knew I could potentially make them part of a book. Sometimes I would succeed, placing the object in question before me and writing about it in my story, describing what I could see just as a 'realist' novelist would. But most times I would limit myself to mentioning these objects in passing, seeking to protect my novel from the deceptive allure of reality and from the dangers of excessive descriptiveness. Every now and then I would insert more familiar items into the story: that is how I ended up giving my father's old ties to Kemal's father, and my mother's knitting needles to Füsun's mother, for the pleasure of having my protagonists use objects from my own life and family. Just as in my novel, where the wealthier branch of the family sends hand-me-downs to poorer distant relations, I was finding things that I knew from my own life and that had left some kind of mark on me, and handing them over to my characters. Or I would recall an object that had left an impression on me as a child, such as the yellow water jug which had been a fixture of my aunt's dinner table for many years, and transport it onto my characters' table without adding it to the museum collection. In 2008, after I had finished the novel and published it, I was tidying up my office when I found a box; inside were a number of items which I had bought from junk shops but then completely forgotten about. As I studied a doorbell taken from some wealthy family's historic mansion, or a rusty lantern from one of those old horse-drawn carriages that

still operated on the Princes' Islands, I was tempted to write a completely different novel that would mention these objects too.

7 But even before the publication of *The Museum of Innocence*, I had already discovered that I could dream up a story, indeed a whole novel, just by observing a series of objects – and that this might become something of a habit. The literary theorist Viktor Shklovsky, one of the principal figures of Russian Formalism, argued that what we call plot is a line which joins the points and themes a novel has set out to describe and investigate. Once we have gathered before us an instinctively selected series of items, and begun to imagine a story that might connect them all, as well as ways to introduce them into our characters' lives, we have essentially already begun to write a novel. Murder mysteries too – which, together with Dostoyevsky's *Crime and Punishment* and the short stories of Edgar Allan Poe, have also had a lasting influence on the shaping of the modern novel – are based on their hero, the Detective, coming up with a story to connect a series of separate clues.

8 But if the objects we have collected are to be capable of transporting us into a plot, and from there into the lavish, coherent, and humane universe of a novel, we need to be able to form an emotional attachment to them. The only way to imagine a novel into existence is by using a sequence of objects that have elicited an emotive, poetic response in us. The phenomenon most conducive to granting objects the power to affect our emotions and awaken our memories is, of course, that of romantic love. I wasn't writing my novel just so that I could go on to create a museum, but also as my own attempt to examine the emotion we call love.

9 My primary objective with this novel was not, therefore, to make a museum, but to undertake a level-headed exploration of the convoluted psychological, cultural, and anthropological experience that is love. I did not wish to place love on a pedestal and describe it, as so many popular songs do, as 'the most wonderful feeling in the world'. Instead, I wanted to show how it is something that – much like a car accident – can simply happen to people during the course of their lives, and often causes us far more pain than we would want. *The Museum of Innocence* is, above all else, a meditation on love.

10 Everyone's experience of love varies in accordance with their class, gender, cultural background, nationality, and even their religious beliefs. The love in *The Museum of Innocence* is the love of an upper middle-class gentleman living in Istanbul during the second half of the twentieth century. Even as I was writing the novel, I suspected that its readers' ideas about love would be far more complex and profound than Kemal's. After the novel was published, some readers complained of Kemal's selfishness and argued that there was nothing 'romantic' or 'sensitive' about him. Other readers were able to forgive Kemal because of all the suffering he endured as the story progressed, and even concluded that he was 'romantic' after all. As far as I'm concerned, both responses seem entirely appropriate.

11 Which brings us to the question I have most frequently been asked since the publication of the novel: 'Mr Pamuk, are you Kemal?' It is easy enough to say 'no, I am not Kemal; he is a character I invented.' But it is a lot harder to persuade the reader who has explored and believed every facet of Kemal's most private emotions that I might not have personally experienced those same emotions too. Indeed, the novelist's art resides in the ability to write about one's own feelings as if they belonged to someone else, and of the feelings of others as if they were one's own. Like all novelists, no matter how many times I tell people that Kemal is an imaginary character, I also want my readers to believe that the love affair I wrote about is one that I also experienced myself.

12 As for those readers who ask 'Mr Pamuk, have you ever fallen in love and started collecting objects that belonged to your beloved, like your protagonist does?', the best response I have is to show them just how much this novel owes to real life. My aunt's family used to have a 1956 model Chevrolet; their driver was called Çetin. When my father worked as a general manager at Aygaz, his office was right in front of the statue of Atatürk at the entrance of the military base in Harbiye – precisely where the Satsat offices are located. Every New Year's Eve, my paternal grandmother would gather all her children and their wives and husbands for dinner at the Pamuk Apartments, and set up a game of tombola for her grandchildren, for which she would choose and prepare the winners' prizes months in advance. Between the 1950s and the 1970s, many homes and shops in Istanbul used to have a canary in a cage or a small aquarium, but with the arrival and

growing popularity of televisions, these had all but disappeared – showing us, in the process, that our relationship to these animals had gone no deeper than the simple desire to have something to look at. In 1983, newly married and in need of some additional income, I began to turn my first novel *Cevdet Bey and His Sons* into a screenplay, encouraged by a film director who was a fan of the novel, but the film was never made, and I never completed the screenplay. Around this time, a friend of mine who was also a director would take me to the bars and clubs in Beyoğlu that filmmakers usually frequented, and when he saw how little it took to get me inebriated – just a few snatches of scandalous gossip overheard among the loud chatter of the starlets around us, and a couple of glasses of beer – he would start laughing at me, teasing me benevolently. From 1974, when I stopped painting and studying architecture, to 1995, I smoked an average of thirty cigarettes a day, and in 1995, I gave up smoking for the first time. To me, the expression 'to smoke like a Turk' – used in some western countries – is not really about consuming too much tobacco or becoming enveloped in a cloud of smoke; rather, it signifies a familiarity with certain social gestures and their individual interpretations (as well as the ability to know and recognise those interpretations) – from opening a packet of cigarettes to offering one to a stranger, still in its packet, in the spirit of friendship and peace; from rolling an unlit cigarette between your fingers until it is at precisely the right firmness to be smoked to the hundreds of different ways you can hold it in your fingertips or blow out the smoke. Towards the end of the 1960s, we used to go to an open-air cinema similar to the one I described in my novel *Silent House* in a small village and holiday resort on the coast of the Marmara Sea, where we would watch Turkish films in the company of the pungent smell of manure and the mooing of cows from a nearby barn. I also clearly remember watching films in the early 1970s with my university friends and thousands of others sunflower seed-chewing spectators in the famous park in Beşiktaş known as the Hunchback's Place. When my mother decided, at the start of the 1960s, to get herself a driver's license, she started taking my brother and I along to her classes so that we would not get too bored at home through the hot summer days. We would sit at the back and giggle – or feel apprehensive – every time the car juddered to a stall. Only when I decided ten years later, at the age of eighteen, to learn how to drive

myself, and ended up failing the practical test over and over again, did I finally understand my mother's struggle. Some of the wealthy people I have written about in the novel are inspired by my father and my uncle's friends, some by my cousins and their friends, and others by my own high school friends. As for seeking to summarise how my own experiences informed the descriptions of the 'luxury' restaurants, the *meyhane*s on the shores of the Bosphorus, the Istanbul streets, and the many different shops that appear in the novel, it would be something of an endless task, akin to trying to explain the extent to which all of my books feed upon the city of Istanbul.

13 I have always wished I could respond in the affirmative when asked 'Orhan, are you Kemal?' – a question which I have encountered with renewed regularity after the opening of the museum in 2012. Perhaps that is why I have come up with the following response: 'Yes, I too spent my childhood and youth among the 1950s to 1990s Nişantaşı bourgeoisie who are the subjects of this novel. Kemal's family and friends closely resemble my own circle, and the places he lives in and visits are similar to those I frequented. Eventually both Kemal and I were pushed out of the social class we belonged to and the milieux we moved in. In a sense, we both fell outside of our class – Kemal due to his love for Füsun; and I due to my love of literature, as well as for political reasons. Neither of us has any regrets.'

AUTHOR

Orhan Pamuk
Prix Nobel de littérature (2006)
IDREF : https://www.idref.fr/029268249
ISNI : http://www.isni.org/0000000120323809
BNF : https://data.bnf.fr/fr/12093134

Ce qu'une historienne peut dire des romans d'Orhan Pamuk

Zeynep Bursa-Millet

DOI : 10.35562/theia.232

Copyright
CC BY-NC-SA 4.0

TEXT

1 J'ai longuement réfléchi au contenu de mon discours, car ce n'est jamais une tâche facile. Il m'a semblé qu'il était intéressant de proposer quelques réflexions sur le lien entre littérature et histoire, un vieux débat qui reste néanmoins d'actualité. Quand on parle du lien entre littérature et histoire, on a peut-être tendance à penser trop vite aux romans réalistes, mais il faut le dire dès le début : Orhan Pamuk a toujours refusé l'injonction d'écrire l'histoire de son monde et de son époque. Cependant, ce refus n'a pas été facile. Une imposante tradition littéraire turque le précédait et le contexte politique en Turquie était toujours pesant.

2 Entre la construction de la République dans les années 1920 et les années 1950, le réalisme est devenu la norme dominante dans la littérature en partie soutenue par l'État turc. Dans ce contexte, les écrivains ont joué un rôle important dans la légitimation de la jeune république. L'écrivain turc avait alors pour tâche d'écrire l'histoire officielle et la vision idéale du futur de son pays. À partir des années 1950, une nouvelle génération d'écrivains (*köy romancıları*) est apparue en Turquie. Influencés par le marxisme, ils se sont intéressés à la condition des paysans en Anatolie et aux défis et conflits des peuples de l'Est de la Turquie. Ce genre littéraire a connu un succès inégalé, particulièrement avec le roman *Mehmet le Mince* de Yaşar Kemal.

3 Les années 1970 en Turquie voient une forte polarisation sociale et politisation de la société turque. L'influence continue du marxisme dans la littérature turque en témoigne. Les œuvres se situent dans les grandes villes ou les villages de province, mais se concentrent surtout

sur les conflits idéologiques, la violence politique, la pauvreté et la souffrance sociale. En 1970, Pamuk était un jeune étudiant à l'Université technique d'Istanbul en architecture. Jusqu'à la fin de cette décennie, ce qu'il a lu dans la littérature turque était en grand partie marqué par le réalisme. Dès le début de sa carrière, il s'est interrogé sur la question de savoir si l'écrivain devait écrire pour le goût et l'esthétique littéraire, ou pour être utile et transmettre un message politique à son peuple [1]. La contestation du réalisme par Pamuk était donc courageuse mais encore une fois difficile.

4 Quand son premier roman est publié en 1982, Pamuk a déjà connu trois coups d'État militaires en Turquie : en 1960, alors qu'il avait 8 ans, puis en 1971 et 1980, lorsqu'il était étudiant et puis jeune écrivain [2]. Ses deux premiers romans – *Cevdet Bey et ses fils* et *La maison du silence* – portent les traces du réalisme et ils sont marqués par leur époque. Cependant, par la suite, il a cherché à s'émanciper de ce courant et de sortir de ce cadre imposé. Il s'est cependant souvent inspiré des dynamiques sociales et politiques de la Turquie, comme on peut le voir dans *Le Livre Noir*, *Le Château Blanc* ou *La Nouvelle Vie*. Son roman *Neige*, qu'il décrit ouvertement comme un roman politique, montre son talent littéraire en intégrant les thèmes chers aux romans réalistes.

5 Orhan Pamuk est un écrivain important pour les historiens. Pourquoi ? Premièrement, les romans d'Orhan Pamuk sont très lus en Turquie. En 1998, bien avant son prix Nobel de littérature, son roman *Mon Nom est Rouge* avait atteint un chiffre record de ventes en Turquie. Pamuk est ainsi devenu non seulement un grand écrivain, mais aussi une figure très connue du grand public. Ses romans sont très discutés. C'est pourquoi ils sont un objet d'étude pour les historiens. Comme le dit l'historien Christophe Charle, la démarche historienne « n'est pas applicable à n'importe quelle œuvre (littéraire), il faut que celle-ci ait eu un impact social et collectif large et, sur ce point, les critères de l'historien ne sont pas ceux du littéraire ou de l'historien d'art » [3].

6 Je pense qu'un chercheur/une chercheuse voulant comprendre la société française ou la société russe du milieu du XIX^{e} siècle se penchent souvent sur les romans de Flaubert ou de Tourgueniev. Ce n'est pas un hasard si les historiens de l'économie empruntent

l'optique d'Émile Zola ou de Charles Dickens. Quant à la Turquie, pour comprendre ce pays, les romans d'Orhan Pamuk sont inévitables aujourd'hui. De nombreux mémoires et thèses de doctorat en sciences sociales en Turquie et à l'étranger témoignent de cet intérêt. Mais pour moi, ce besoin ne traduit pas qu'une recherche de la réalité ou de l'illustration parfaite de telle ou telle période historique dans les romans. C'est, au contraire la volonté de profiter du travail de quelqu'un qui a plus de liberté dans l'usage de l'imagination… Parce qu'on sait que les meilleures questions historiques viennent à l'esprit quand on se demande : pourquoi cela a-t-il été ainsi et pas autrement ? Avec parfois une certaine jalousie, je dois reconnaitre qu'Orhan Pamuk a la liberté d'utiliser son imagination sans être contraint par les archives et les critères scientifiques du monde académique. Il a ainsi plus de chances de montrer ce que l'on ne prend pas en considération en tant qu'historien.

7 Ce qui est encore très précieux, c'est surement sa position littéraire sur l'histoire. Orhan Pamuk disait en 2022 : « Pour moi, l'histoire diplomatique importe peu, ce qui compte, c'est l'histoire des choses de la vie quotidienne. Là, il faut être très juste, respecter les objets, alors que je peux utiliser mon imagination pour l'histoire diplomatique ou raconter une guerre »[4]. Pour les historiens, et d'ailleurs en général pour les chercheurs/chercheuses en sciences sociales et humaines, il est toujours difficile de mener une recherche sur la vie quotidienne, sur la vie des acteurs/individus ordinaires, sur la vie des couches marginalisées. Orhan Pamuk, visiblement, a toujours été prêt à relever ce défi. *Le musée de l'Innocence* est ainsi un parfait exemple de ces efforts de Pamuk pour défendre notre quotidien et notre condition humaine.

8 Une dernière chose importante chez Pamuk pour un historien, c'est la question de la mémoire. Paul Ricoeur, dans son livre *La Mémoire, l'histoire, l'oubli*, reconnaît que l'histoire a une relative dépendance à l'égard de la mémoire[5]. Cette question est l'une des plus frappantes dans les romans et la carrière d'écrivain d'Orhan Pamuk. Paru en turc en 2003, traduit et publié en français en 2007, *Istanbul, Souvenirs d'une ville* se présente non seulement comme un récit autobiographique, mais aussi comme une réflexion sur la nature complexe de la mémoire. On peut ainsi parler d'un autre point

commun entre les historiens et les romanciers, en ce qu'ils permettent de faire vivre des lieux, des personnes ou des objets.

9 Je me permets de finir mon discours avec une phrase de ce livre : « Parfois je me prends à penser : La vie ne peut pas être à ce point mauvaise. Mais, quoi qu'il en soit on peut toujours, en fin de compte, aller marcher du côté du Bosphore » [6]. Une phrase ancrée dans son quotidien depuis sa jeunesse au cœur de l'histoire d'une ville.

NOTES

1 Lors d'un entretien en français en 2004, Orhan Pamuk dit : « Lorsque j'écrivais mes premiers romans au milieu des années 1970, la culture occidentale diffusait deux images de l'écrivain : l'auteur "engagé" sur le modèle de Zola et Sartre, et celui qui écrit pour la beauté de l'écriture, comme Proust ou Nabokov. J'ai tâché d'écrire à l'image de ces derniers ». Il comprend d'ailleurs pourquoi il est souvent présenté comme un romancier engagé : « … avec la popularité, les sollicitations se sont accrues. Durant la dernière décennie, j'ai signé des pétitions, écrit des essais polémiques », Orhan Pamuk, « Il faut débattre d'une définition de l'Europe », propos recueillis par Nicolas Monceau, *Le Monde*, Supplément « Le Mondes Livres », 2 avril 2004, https://www.lemonde.fr/archives/article/2004/04/02/il-faut-debattre-d-une-definition-de-l-europe_4295727_1819218.html (accès : 29 août 2024).

2 Les coups d'État militaires ont des natures différentes. Le coup d'État de 1960 a été organisé par un groupe de jeunes officiers contre le parti conservateur, le Parti démocrate. Ce parti est jugé comme anti-kémaliste, et surtout anti-laïc, par une partie de l'armée turque. Paradoxalement, les militaires putschistes ont fourni une occasion de mettre en place une constitution parmi les plus libertaires de l'histoire de la Turquie. Avec cette nouvelle constitution, on assiste à un grand mouvement social d'ouvriers, d'étudiants et de groupes opprimés (comme les femmes, les Kurdes et les Alévis) pendant les années 1960 et 1970. La gauche marxiste devient alors un acteur important de la vie politique et sociale. Les deux derniers coups d'État ont pour objectif d'écraser ces mouvements sociaux et de reconstituer le pays sur des bases autoritaires, nationalistes et religieuses. Le coup d'État de 1980, notamment, est souvent considéré comme le plus violent de l'histoire de ce pays.

3 « Christophe Charle. Le parcours et les idées d'un historien » entretien réalisé par Chloé Maurel, *Cahiers d'Histoire*, n° 127, 2015, p. 149-166.

4 https://www.radiofrance.fr/franceculture/podcasts/le-cours-de-l-histoire/orhan-pamuk-est-fou-d-histoire-8713237

5 Paul Ricoeur, *La mémoire, l'histoire, l'oubli*, Paris, Le Seuil, 2000.

6 Orhan Pamuk, *Istanbul. Souvenirs d'une ville*, trad. Savaș Demirel, Valérie Gaye-Aksoy et Jean-François Pérouse, Istanbul, Gallimard, 2007, p.96.

AUTHOR

Zeynep Bursa-Millet
Université Lumière Lyon 2, LARHRA UMR 5190
IDREF : https://www.idref.fr/256721351

La spirale inspirée

Laurent Baridon

DOI : 10.35562/theia.234

Copyright
CC BY-NC-SA 4.0

TEXT

1 « Je pourrais passer ma vie dans une chambre ». Ces mots d'Orhan Pamuk dans *Bir odada* [*Dans une chambre*] disent l'honneur qu'il nous fait ce soir en étant dans cet amphithéâtre pour recevoir l'hommage de notre université, un hommage modeste au regard de tous ceux qu'il a déjà reçus dans le monde entier, mais un hommage absolument sincère. Évoquer en l'espace de quelques minutes son œuvre immense est vain et dérisoire. Qu'il me soit permis, en guise d'éloge, de dire quelques mots du rapport qu'elle entretient aux objets et aux images, ne serait-ce que parce qu'en tant qu'historien de l'art, je suis censé savoir en parler ; censé seulement car, simple lecteur parmi tant d'autres à travers le monde, j'éprouve comme eux une fascination qui déjoue mes compétences académiques.

2 Tout lecteur d'Orhan Pamuk aura ressenti la même impression : son écriture, avec une force singulière, fait naître ou renaître des espaces, des moments, des images. Elle fait vivre des instants ou des époques, des pays et des villes, des quartiers et des rues, des maisons, des appartements… et des chambres. Elle nous les rend aussi coutumiers que si nous y avions vécu ou, plus précisément, aussi familiers que si nous les avions observés au prisme d'une image, graphique ou photographique, qui aurait durablement imprimé notre mémoire. Dans cette chambre où il dit pouvoir passer sa vie, comme dans une chambre photographique, se projettent les images des espaces et des objets d'autres vies, celles de ses personnages. Le roman, selon Pamuk, « raconte notre propre vie comme si c'était celle d'un autre et nous offre la possibilité d'écrire la vie d'autres personnes comme s'il s'agissait de la nôtre » (*D'autres couleurs*, Paris, Gallimard, 2006, p. 299). Nous, ses lecteurs, vivons ces vies et ces espaces en nous les appropriant. Ils s'insinuent dans notre propre expérience vécue. Ce

qui nous rend ces vies si familières, au point qu'elles se confondent avec les nôtres, est sans doute lié au rapport très particulier de l'écrivain aux images et aux objets.

3 Alors qu'il était étudiant en architecture, le jeune Orhan Pamuk semble avoir été surtout intéressé par l'histoire et le dessin plus que par les aspects techniques de cette formation. Un de ses amis de l'époque a dit avec quelle jubilation ils avaient ensemble parcouru les rues du quartier de Beşiktaş pour en dessiner ses maisons anciennes, probablement parce qu'elles offraient la possibilité de plonger tant dans l'histoire de ce faubourg populaire que dans celles de ses habitants. De cette période, des dessins de jeunesse ont été retrouvés qui témoignent d'une recherche littéraire et visuelle et même de leur interaction puisqu'il s'agit parfois de calligrammes. Je tiens à mentionner également des caricatures et des dessins satiriques parce que, comme souvent dans ce genre graphique, l'effet obtenu procède de du frottement de l'image et des mots qui la légendent.

4 Si, très jeune, Orhan Pamuk a renoncé à sa vocation de peintre, il a, toute sa vie, continué à dessiner. *Souvenirs des montagnes au loin* reproduit ses *Carnets dessinés*. Pour autant il ne s'agit pas de livres illustrés, un genre dans lequel très peu d'écrivains se sont essayés, même quand ils dessinent ; mais plutôt de dessins accompagnés de texte. Les images de Pamuk capturent des moments, des odeurs et des formes. Flâneur baudelairien, c'est au cours de promenades dans le quartier de Çukurkuma, dans ses odeurs et ses ambiances, dans le fouillis de son marché aux puces, qu'il imagine le roman *Le Musée de l'Innocence*. Et le véritable musée lui-même est installé dans une maison découverte au fil de semblables pérégrinations, une maison dont les façades repeintes en rouge contribuent à nier ses propriétés architecturales pour en faire un grand objet destiné à contenir tous ceux qui ont été glanés dans les rues de ce quartier. En exposant ainsi les matériaux (au sens propre et figuré) de l'œuvre littéraire, il donne à ses lecteurs la possibilité de partager le processus d'écriture et de construction de la fiction. Ces objets sont des condensateurs d'expérience. Que l'on pense au sac à main de luxe que vend Füsun à Kemal, caractérisé par le papier de rembourrage couleur crème de la pochette centrale, dont le crissement entre les mains de Füsun, seulement imaginé par le lecteur, rend l'instant incroyablement présent.

5 Le lecteur de Marcel Proust pourra rapprocher cette quête nostalgique du moment disparu et de l'être aimé du rôle des espaces, des objets et des parfums dans l'œuvre de l'écrivain français. Même coalescence du présent et du passé pour revivre le temps d'une existence, « dans le temps » – les derniers mots de la *Recherche* de Proust –, mais aussi et peut-être davantage, chez Pamuk, hors du temps – comme l'inconscient qui selon Freud « ignore le temps ». L'objet est le médiateur de cette expérience du temps, très différente néanmoins chez Proust et Pamuk. Les « vivantes échasses grandissant sans cesse » sur lesquelles les personnages de Proust sont juchés plongent « dans le temps » caractérisant un écrivain aux prises avec l'angoisse des ruptures de la modernité du début du XX^{e} siècle. Chez Pamuk, un siècle plus tard, le temps ne présente pas cette verticalité plongeante car le récit est fondé sur le parcours de la découverte des objets, que cette découverte soit celle de l'écrivain, au fil de ses promenades, ou qu'elle soit celle de son lecteur/visiteur du musée de l'Innocence. Ainsi que Pamuk l'a souligné, en se référant au formaliste russe Victor Chklovski, les objets sont des points de passage reliés par la ligne qu'emprunte la fiction. De la même manière que dans *The Life and Opinions of Tristram Shandy, Gentleman*, de Laurence Sterne, cette ligne du temps du récit n'est pas droite – elle est dessinée comme une arabesque par l'auteur lui-même – la ligne du temps pamukien n'a pas la rectitude des échasses proustiennes qui tentent de retrouver un passé menacé d'abolition. Elle relie les temps et les lieux par de nombreux parcours, comme l'histoire polyphonique de la ville d'Istanbul qui s'écrit entre deux continents, plusieurs civilisations et de multiples cultures.

6 Cette conception plurielle et polysémique de l'histoire et du temps, que l'on peut qualifier de post-moderne, s'incarne dans la spirale peinte sur le sol du rez-de-chaussée du Musée de l'Innocence. Elle représente tant les pérégrinations du flâneur Pamuk à la recherche des objets et de la maison musée, que la construction du récit du lui-même et le parcours du visiteur d'un objet à l'autre. Le récit spirale le temps de nos vies faites de chemins plus ou moins erratiques – quels que soient nos efforts pour les organiser et nous convaincre que nous y parvenons.

7 L'architecte et critique Charles Jencks, un des fondateurs de la notion de postmodernité, a accordé une attention particulière à la forme de la spirale dans son *Garden of Cosmic Speculation*. Ainsi que ce nom l'indique, Jencks a inscrit dans le paysage des confins de l'Angleterre et de l'Ecosse, à la manière d'un pétroglyphe, la forme de l'univers, ou tout au moins celle de notre galaxie. La spirale inscrite sur le sol du Musée de l'Innocence, de la même façon, relie les objets, les personnages et les époques de l'histoire, l'écrivain et ses lecteurs, le temps révolu et le temps vécu, comme autant de corps célestes qui gravitent dans des temporalités diverses et complexes. C'est peut-être ce qui rend l'écriture de Pamuk à la fois si intime et si universelle : les objets et les images qui la fondent sont les fragments ou les vestiges d'autres mondes, d'autres univers intérieurs. Ils se sont formés dans la chambre où Pamuk écrit, devant les objets qu'il y a rapporté ; cette chambre bien semblable à toutes celles que traverse le visiteur de la maison du musée de l'innocence, toute semblable aussi à cette chambre où nous lisons le récit d'existences qui sont toutes celles de Pamuk et toutes les nôtres.

8 Je termine, comme j'ai commencé, avec une citation d'Orhan Pamuk :

> Les romans sont des structures singulières qui nous permettent de n'avoir aucun mal à faire coexister dans notre esprit des pensées contradictoires, et de comprendre simultanément des points de vue divergents. [1]

9 Les livres d'Orhan Pamuk nous aident à mieux comprendre le monde, à apprécier sa diversité en nous ouvrant aux autres, d'où qu'ils viennent, plutôt que de refermer notre histoire sur elle-même en la figeant. Lire Pamuk, c'est suivre la spirale qui part de sa chambre pour parcourir, connaître et embrasser l'universalité du monde à travers chacun des individus qui le composent.

NOTES

1 *Le romancier naïf et le romancier sentimental*, trad. de l'anglais par Stéphanie Levet, Paris, Gallimard, coll. « Arcades », 2012, p. 35.

AUTHOR

Laurent Baridon
Université Lumière Lyon 2, LARHRA UMR 5190
IDREF : https://www.idref.fr/034505083
ISNI : http://www.isni.org/0000000061546032
BNF : https://data.bnf.fr/fr/12525490

In Memoriam

De la grandeur d'une historienne

Hommage à Natalie Zemon Davis

A Historian's Greatness. Homage to Natalie Zemon Davis

Monica Martinat

DOI : 10.35562/theia.236

Copyright
CC BY-NC-SA 4.0

TEXT

1 Elle aura été l'une des plus grandes historiennes d'un XX^e^ siècle qu'elle a traversé avec passion et engagement, en transformant la manière de faire et penser l'histoire. De quoi est-elle faite, cette grandeur ?

2 On pourrait y répondre en puisant dans son œuvre les éléments qui nourrissent la réflexion de tous ceux qui étudient les différents objets sur lesquels elle a posé ses yeux : des imprimeurs et femmes lyonnaises au XVI^e^ siècle et leur adhésion à la Réforme protestante à Lazare Sainéan, objet de son dernier livre, en passant par Martin Guerre, Léon l'Africain, Maria Sibylla Merian, Gluckel von Hameln ou Marie de l'Incarnation – ces dernières rassemblées dans un livre qui tisse leurs trajectoires différentes à partir d'une interrogation portant sur les possibilités des femmes au XVII^e^ siècle. Et bien d'autres encore. Mais cela serait insuffisant à rendre compte de ce qu'elle a représenté. Parce que son importance repose surtout, à mes yeux, sur la manière dont elle a exercé notre métier, la liberté avec laquelle elle a su créer une cohérence à partir d'objets divers et la curiosité qui l'a poussée à chercher toujours l'altérité, à se laisser étonner par ses découvertes et par les gens, d'hier et d'aujourd'hui, qu'elle rencontrait.

3 Son expérience en tant que conseillère historique sur le tournage du film *Le Retour de Martin Guerre* (1983) lui inspire une recherche propre, historienne, sur ce cas étonnant de fausse identité au XVI^e^ siècle ; mais elle entame aussi une réflexion qu'elle poursuivra longtemps sur les rapports de l'histoire avec la fiction, en

revendiquant à la première une ouverture aux possibilités que la seconde fermerait par un choix interprétatif unique de l'auteur. C'est la lecture du roman d'Amin Maalouf, *Léon l'Africain*, qui relance cette réflexion : elle la proposera lorsqu'elle sera invitée à prononcer la Conférence Marc Bloch en 1995 et elle en fera un « prétexte » pour aborder à son tour, en historienne, l'histoire de celui qu'elle appellera systématique Hassan et Léon, comme à souligner la multiplicité, voire les ambiguïtés, des identités qui libèrent les individus. Son livre aura un retour intéressant à la fiction d'où il était parti lorsqu'en 2017 Wajdi Mouawad en proposera une mise en scène théâtrale avec *Tous des Oiseaux* et dont elle acceptera avec enthousiasme le déplacement total proposé par le réalisateur qui en fera une œuvre à lui, et non pas une réécriture de son livre à elle. En 2006, elle revient à l'étude de la fiction, avec *Slaves on screen* dans lequel elle interroge, encore, la manière dont la fiction s'approprie l'histoire et en donne une lecture particulière, dont elle met à l'épreuve, de manière critique et fine, le rapport à la réalité du passé, à cette réalité que l'historien se doit de saisir, ne serait-ce qu'en mode incertain.

4 Son engagement politique, qui lui a valu des poursuites par le gouvernement américain imprégné du maccarthysme anticommuniste – entre autre le retrait du passeport lui empêchant de retourner en France pour terminer le livre qu'elle aurait voulu écrire sur les imprimeurs lyonnais – a inspiré ses recherches, toujours tournées vers les plus fragiles, qu'elle a traité avec délicatesse et rigueur, en les plaçant à la fois à côté d'elle et à distance : si l'historienne invente un dialogue possible avec les « trois femmes en marges » auxquelles elle consacre un livre en 1995 qui dévoile les sentiments et les intentions qui portent cette étude, elles reprennent leur place distante, bien assises dans un XVII^{e} siècle « autre ». Ainsi, cette étude retrouve ses allures toutes scientifiques, prouvées et fiables, en dialogue non plus avec les Marie, Maria Sybilla et Gluckel, mais avec les sources et les lecteurs.

5 Sa manière d'être historienne est une leçon pour nous toutes et nous tous, particulièrement dans un temps où la recherche semble s'imbriquer à nos vies personnelles d'une manière sans doute inattendue même par ceux qui avaient revendiqué, il y a quelques décennies, l'importance de la subjectivité de l'historien dans la recherche. Passionnée de récit, elle n'a jamais fait de celui-ci la

modalité unique de l'opération historienne, ni de ses sentiments l'origine de ses interprétations – ou du moins, elle n'a jamais pensé à les mettre au centre du dispositif de la preuve, en reléguant l'empathie pour les gens qu'elle étudiait à sa juste place : un mobile de la recherche, non pas une recherche d'adhésion à ses propres interprétations, qu'elle proposait de manière à la fois décidée et réservée, basé sur un dialogue à plusieurs voix, entre passé et présent, entre historiens, entre l'autrice et ses lecteurs. Elle a su admirablement partager avec nous une idée de l'histoire profondément démocratique, ouverte, passionnée. Intelligente et respectueuse par-dessus tout.

INDEX

Mots-clés
Zemon Davis (Nathalie), historiographie, biographie

Keywords
Zemon Davis (Nathalie), historiography, biography

AUTHOR

Monica Martinat
Université Lumière Lyon 2, LARHRA UMR 5190
IDREF : https://www.idref.fr/085786748
HAL : https://cv.archives-ouvertes.fr/monica-martinat
ISNI : http://www.isni.org/000000007729030X
BNF : https://data.bnf.fr/fr/14610331

Hommage à Gaylord Brouhot (1981-2024)

Homage to Gaylord Brouhot (1981-2024)

Joana Barreto

DOI : 10.35562/theia.244

Copyright
CC BY-NC-SA 4.0

EDITOR'S NOTES

Hommage paru initialement dans https://sfdes.hypotheses.org/7919 le 17 mars 2024.

TEXT

1 La tragique disparition de Gaylord Brouhot à l'âge de 42 ans laisse un vide considérable. Je l'ai personnellement connu lorsqu'il suivait un des travaux dirigés que je donnais alors à l'université Paris 1 Panthéon-Sorbonne. Néanmoins, la rencontre amicale n'eut pas véritablement lieu à cette époque mais au moment où il choisit de s'inscrire en thèse d'histoire de l'art et de se spécialiser sur la Renaissance italienne. Loin de suivre une méthodologie tracée d'avance, Gaylord va alors profondément enrichir notre discipline car il n'arrive pas seulement avec son bagage académique très conséquent. En effet, s'il décide de consacrer sa thèse aux costumes dans les portraits des Médicis, c'est en apportant à ce sujet son expertise unique dans les industries du luxe et, en particulier, sa connaissance de l'histoire des tissus et de la mode. On lui doit ainsi la notion si stimulante de « portrait du costume ».

2 Gaylord a travaillé avec et sur les grands groupes de mode contemporains, et a longtemps enseigné dans le cadre de la formation à ces métiers, tout en assumant également des charges de cours portant sur une histoire de l'art plus classique. Attaché temporaire de recherche à l'université Paris 1 puis à l'université Rennes 2, il était très apprécié par les étudiants auxquels il savait transmettre sa passion, son savoir et la finesse d'observation qui le

caractérisait. Son expertise lui avait valu d'être consulté lors de la rénovation des tentures du *Studiolo* de Ferdinand de Médicis à la Villa Médicis – qui l'a accueilli plus tard comme pensionnaire. Personne mieux que lui ne savait parler avec précision, enthousiasme et chaleur de la trame des tissus, des techniques de teinture propres à la Renaissance, de l'origine des effets moirés ou des circulations de goût entre cours, comme des variations de terminologie qui rendent la compréhension des archives si délicates. Sa générosité faisait de lui un passeur entre deux mondes qui se parlent difficilement.

3 La même expertise lui a fait jouer un rôle cardinal dans la reconstruction de la rencontre du Camp du Drap d'Or ou dans l'investigation des contacts entre la cour de François I^{er} et l'Italie. Elle lui a encore permis de développer une histoire stimulante de la braguette à la Renaissance. En parallèle du travail colossal qui a abouti à sa thèse de doctorat méticuleuse et foisonnante, Gaylord n'a jamais hésité à sortir de son champ propre pour devenir un compagnon de pensée précieux de nombreux interlocuteurs, comme lors du colloque sur la sculpture dans le Nord de l'Europe à l'époque moderne.

4 Se situant avec une grande délicatesse au-delà des luttes intestines propres au milieu académique, Gaylord a ainsi su tracer un itinéraire fidèle à ses amitiés. Il est bien difficile de dire adieu à sa douceur et à sa joie radieuse sous le soleil provençal de Vaugines, terre de son cœur.

AUTHOR

Joana Barreto
Université Lumière Lyon 2, LARHRA UMR 5190
IDREF : https://www.idref.fr/146815823
ORCID : http://orcid.org/0000-0003-0906-3342
ISNI : http://www.isni.org/0000000426632460
BNF : https://data.bnf.fr/fr/15884695