

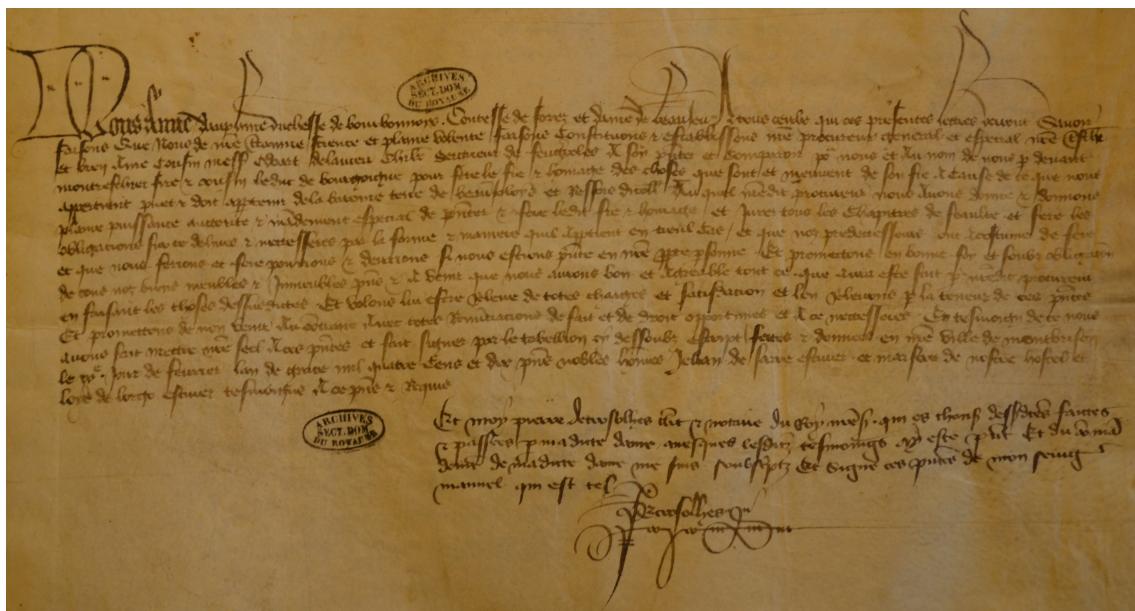
Justine Chainiau

licenciée ès lettres

diplômée de master recherche en histoire médiévale

VERS LA MISE EN LIGNE DES ACTES DES DUCS DE BOURBON (1356 - 1475)

Édition, encodage et chaîne de traitement
au sein du projet « Actes Princiers »



Mémoire pour le diplôme de master
« Technologies numériques appliquées à l'histoire »

2023

Résumé

Le projet « Actes Princiers au royaume de France (XIV^e - XVI^e siècle) » entend procéder à l'édition numérique de plusieurs corpus d'actes inédits de princes français de la fin du Moyen Âge. Parmi ces corpus se trouvent ceux composés des actes émis par les ducs de Bourbon : Louis II, Anne Dauphine, Charles I^{er} et Agnès de Bourgogne.

Ces quatre corpus ont fait l'objet d'éditions numériques via un logiciel de traitement de texte. Le présent mémoire s'attache à présenter la chaîne de traitement permettant la transformation des documents Word contenant les éditions diplomatiques des actes en des fichiers XML (TEI) en vue de leur mise en ligne.

Mots-clés : Actes princiers ; ducs de Bourbon ; chaîne de traitement ; diplomatique ; édition ; encodage ; expressions régulières ; LaTeX ; ODT ; Python ; transcription ; XML-TEI.

Informations bibliographiques : Justine Chainiau, *Vers la mise en ligne des actes des ducs de Bourbon (1356 - 1475). Édition, encodage et chaîne de traitement au sein du projet « Actes Princiers au royaume de France (XIV^e - XVI^e siècle) »*, mémoire de master « Technologies numériques appliquées à l'histoire », dir. Jean-Damien Généro et Lucence Ing, École nationale des chartes, 2023.

Soutenance : mémoire présenté et soutenu publiquement le 25 septembre 2023 à l'École nationale des chartes, devant un jury composé d'Olivier Canteaut, président, archiviste paléographe et maître de conférences, Jean-Damien Généro, ingénieur d'études du CNRS et Lucence Ing, doctorante et attachée temporaire d'enseignement et de recherche (ATER).

Illustration de couverture : Lettres patentes d'Anne Dauphine, AN, P 1389³, c. 314 (9 février 1410).

Remerciements

Mes remerciements vont tout d'abord à mon encadrant de stage Jean-Damien Généro pour son accompagnement. Son écoute, sa disponibilité et ses conseils ont été essentiels et précieux dans la conduite des missions du stage et pour la rédaction du mémoire.

Je tiens aussi à remercier particulièrement Lucence Ing, pour ses relectures et ses recommandations concernant la rédaction du présent travail de recherche, et Olivier Can teaut, pour l'intérêt porté à mon travail et les informations transmises quant à ma poursuite d'étude.

Je remercie également l'équipe pédagogique du master TNAH pour ces deux années de formation riches d'enseignements, d'acquisition de compétences et de méthodes.

Enfin, je remercie mes proches de m'avoir épaulée et soutenue pendant la rédaction.

Liste des sigles et abréviations

Institutions

AD : Archives départementales

AM : Archives municipales

AN : Archives nationales

BM : Bibliothèque municipale

BnF : Bibliothèque nationale de France

CNRS : Centre National de Recherche Scientifique

CRH : Centre de recherches historiques (CNRS - EHESS)

EHESS : École des Hautes Études en Sciences Sociales

ENC : École Nationale des Chartes

Labex : Laboratoire d'excellence

LAMOP : Laboratoire de médiévistique occidentale de Paris (CNRS - Paris 1)

UMR : Unité mixte de recherche

Informatique et technologies numériques

CSV : *Comma-Separated Values*

HTML : *Hypertext Markup Language*

ODF : *Open Document Format*

ODT : *Open Document Text*

PDF : *Portable Document Format*

TEI : *Text Encoding Initiative*

XML : *eXtensible Markup Language*

XSL : *eXtensible Stylesheet Language*

Bibliographie

Diplomatique

C.I.D., « Normes Internationales Pour l'édition Des Textes Médiévaux », dans *Travaux Préliminaires de La Commission Internationale de Diplomatique*, Saragosse, 1984 (*Folia Caesaraugustana*, 1), p. 21-93.

GUYOTJEANNIN (Olivier), *Conseils pour l'édition des textes médiévaux, Fascicule II : actes et documents d'archives*, t. 2, 3 t., CTHS - École Nationale des Chartes, Paris, 2009 (Orientations et méthodes, 14).

GUYOTJEANNIN (Olivier) et MATTÉONI (Olivier), *Jean de Berry et l'écrit : Les pratiques documentaires d'un fils de roi de France*, Éditions de la Sorbonne, École Nationale des Chartes, Paris, 2019 (Histoire ancienne et médiévale, 159), DOI : 10.4000/books.psorbonne.54103.

GUYOTJEANNIN (Olivier), PYCKE (Jacques) et TOCK (Benoît-Michel), *Diplomatique médiévale*, Brepols, Paris, 2006 (L'Atelier du médiéviste, 2).

GUYOTJEANNIN (Olivier) et VIELLIARD (Françoise), *Conseils pour l'édition des textes médiévaux, Fascicule I : conseils généraux*, t. 1, 3 t., CTHS - École Nationale des Chartes, Paris, 2014 (Orientations et méthodes, 27).

Techniques Pour l'Historien En Ligne (Theleme), École Nationale des Chartes, URL : <http://theleme.enc.sorbonne.fr/> (visité le 10/04/2023).

Historiographie

BENTON (John F.), BUR (Michel) et COURTOIS (Michèle), *Recueil des actes d'Henri le Libéral Comte de Champagne (1152-1181)*, Académie des inscriptions et belles-lettres, Paris, 2009 (Chartes et diplômes relatifs à l'histoire de France).

BÖHMER (Johann Friedrich), *Regesta Imperii inde ab anno MCCCXIII usque ad annum MCCXLVII*. S. Schmerber, Francfort-sur-le-Main, 1839.

CANTEAUT (Olivier), *Le discret langage du pouvoir, Les mentions de chancellerie du Moyen Âge au XVIIe siècle*, École Nationale des Chartes, Paris, 2019 (Etudes et rencontres de l'Ecole des Chartes, 55).

- COCKSHAW (Pierre), *Le Personnel de La Chancellerie de Bourgogne-Flandre Sous Les Ducs de Bourgogne de La Maison Des Valois (1384-1477)*, U.G.A., Kortrijk-Heule, 1982.
- COURTEL (Anne-Lise), « La chancellerie et les actes d'Eudes IV, due de Bourgogne (1315-1349) », *Bibliothèque de l'École des Chartes*, 135-1 (1977), p. 23-71, DOI : 10.3406/bec.1977.450090.
- DELISLE (Léopold) et BERGER (Élie), *Recueil des actes de Henri II, roi d'Angleterre et duc de Normandie, concernant les provinces françaises et les affaires de France*, 4 t., Imprimerie Nationale, Paris, 1909 (Chartes et diplômes relatifs à l'histoire de France).
- FAVIER (Jean), *La Guerre de Cent Ans*, Fayard, Paris, 1980.
- HÜBER (Alfons), *Die Regesten des Kaiserreichs, unter Kaiser Karl IV, 1346-1378*, G. Olms, Innsbruck, 1877.
- JONES (Michael), *Recueil Des Actes de Jean IV, Duc de Bretagne*, Université de Haute Bretagne - Rennes 2, Rennes, 1980.
- *Recueil des Actes de Charles de Blois et Jeanne de Penthievre : Duc et duchesse de Bretagne (1341-1384). Suivi des Actes de Jeanne de Penthievre (1364-1384)*, Presses universitaires de Rennes, Rennes, 2016 (Histoire), DOI : 10.4000/books.pur.28420.
- KERHERVÉ (Jean), « La Chancellerie de Bretagne Sous Louis XII et Anne de Bretagne (1499-1514) », dans *Powerbrokers in the Late Middle Ages/Les Courtiers Du Pouvoir Au Bas Moyen Âge*, Brepols, Turnhout, 2001 (Burgundica, IV), p.199-233.
- LÉMEILLAT (Marjolaine), *Actes de Pierre de Dreux, Duc de Bretagne (1213-1237) et Actes de Jean Ier, Duc de Bretagne (1237-1286)*, PUR/Société d'histoire et d'archéologie de Bretagne, Rennes, 2013 (Sources Médiévales de l'histoire de Bretagne, 1-2).
- REYDELET (Chantal), « La Chancellerie d'Humbert II Dauphin de Viennois (1333-1349) », *Archiv für Diplomatik*, 20 (1974), p. 241-383.
- STEIN (Henri), « Catalogue des Actes de Charles le Téméraire (1467-1477) », dans *Urkunden und Mandate Karls von Burgund, Grafen von Charolais, 1433-1467, bearb. von Sonja Dünnebeil*, Jan Thorbecke, Sigmaringen, 1999 (Instrumenta, 3).

Les ducs de Bourbon

- DE Boos Emmanuel, « L'armorial d'Auvergne », dans *Bourbonnais et Forez, Créer, Nonette*, 1998 (édition et étude).
- GÉNÉRO (Jean-Damien), *La chancellerie de Charles Ier de Bourbon, suivi de l'édition de ses actes et ceux de son épouse, Agnès de Bourgogne (1421-1476)*, Mémoire de Master recherche, Paris, Paris I Panthéon-Sorbonne, 2018.
- « Chanceliers et secrétaires de la Chancellerie de Charles Ier de Bourbon et d'Agnès de Bourgogne (1425-1456) », *Bulletin trimestriel de la Société bourbonnaise des*

- études locales*, Bulletin trimestriel de la Société bourbonnaise des études locales 365–23 (2021), p. 1-17.
- « Sceaux et armes de Charles Ier de Bourbon et d’Agnès de Bourgogne », *Revue française d’héraldique et de sigillographie*, 92 (2022), p. 31-51, URL : <https://hal.science/hal-03908270> (visité le 03/04/2023).
- HUILLARD-BREHOLLES (Jean Louis Alphonse) et LECOY DE LA MARCHE (Albert), *Titres de la maison ducale de Bourbon*, Plon, 1867, URL : http://archive.org/details/bub_gb_POUbAQAAQAAJ (visité le 24/06/2023).
- LEGUAI (André), « De La Seigneurie à l’Etat. Le Bourbonnais Pendant La Guerre de Cent Ans », *Annales. Histoire, Sciences Sociales*, Les Imprimeries Réunies 30–2-3 (juin 1975), p. 365-366, DOI : [10.1017/S0395264900169057](https://doi.org/10.1017/S0395264900169057).
- « Agnès de Bourgogne, Duchesse de Bourbon », *Etudes bourbonnaises* (, 1996), p. 405-417.
- MATTÉONI (Olivier), *Servir le prince : les officiers des ducs de Bourbon à la fin du Moyen âge, 1356-1523*, Paris, Paris I Panthéon-Sorbonne, 1994.
- « Écriture et pouvoir princier. La chancellerie du duc Louis II de Bourbon (1356-1410) », dans *”De part et d’autre des Alpes” (II) : chancelleries et chanceliers des princes à la fin du Moyen Âge ; actes de la table ronde de Chambéry, 5 et 6 octobre 2006*, Université de Savoie, Chambéry, 2011 (Collection Sociétés, religions, politiques, 19), t. 2, p. 137-178.
 - « À propos d’une ”ordonnance” de finance attribuée à Louis II de Bourbon : une relecture », dans *Cultures fiscales en Occident du Xe au XVIIe siècle. Études offertes à Denis Menjot*, Presses universitaires du Midi, Toulouse, 2019 (Études médiévales ibériques, 17), p. 153-163, URL : <https://shs.hal.science/halshs-02972959> (visité le 06/07/2023).
 - *Les Bourbons en leur bibliothèque (XIIIe-XVIe siècle)*, Éditions de la Sorbonne, Paris, 2022 (Histoire ancienne et médiévale).
- MATTÉONI (Olivier) et GÉNÉRO (Jean-Damien), « Les actes d’Anne Dauphine : étude diplomatique », dans *Anne Dauphine, dernière comtesse de Forez. Actes de la journée d’étude de Montbrison, 29 septembre 2017*, La Diana - Société historique et archéologique du Forez, Montbrison, 2019, p. 35-52, URL : <https://shs.hal.science/halshs-02972893> (visité le 31/03/2023).
- NORTIER (Michel), « Le sort des archives dispersées de la Chambre des Comptes de Paris », *Bibliothèque de l’École des chartes*, 123–2 (1965), p. 460-537, DOI : [10.3406/bec.1965.449701](https://doi.org/10.3406/bec.1965.449701).
- VAYSSIÈRE (Augustin), « Fragment d’un Compte de Gilles Le Tailleur, Argentier de Charles Ier, Duc de Bourbonnais, En 1448, Communication de M. Vayssière », *Bulletin archéologique du Comité des travaux historiques et scientifiques*–1 (1891), p. 54-76.

Technologies numériques

- APOLLON (Daniel), RÉGNIER (Philippe) et BÉLISLE (Claire), *L'édition critique à l'ère du numérique*, L'Harmattan, Paris, 2017 (Socio-économie de la chaîne du livre, 9).
- Beautiful Soup 4.12.0 Documentation*, Beautiful Soup Documentation, URL : <https://www.crummy.com/software/BeautifulSoup/bs4/doc/#id15> (visité le 19/04/2023).
- BURNARD (Lou), *Qu'est-ce que la Text Encoding Initiative ?*, trad. par Marjorie Burghart, OpenEdition Press, Marseille, 2015 (Encyclopédie numérique), DOI : 10.4000/books.oep.1237.
- Contenu de la documentation Python*, Python documentation, URL : <https://docs.python.org/3/contents.html> (visité le 05/05/2023).
- Créer Ou Modifier Des Fichiers .Csv à Importer Dans Outlook*, Microsoft, URL : <https://support.microsoft.com/fr-fr/office/créer-ou-modifier-des-fichiers-csv-%C3%A0-importer-dans-outlook-4518d70d-8fe9-46ad-94fa-1494247193c7> (visité le 28/06/2023).
- DE VALERIOLA (Sébastien), « L'ordinateur au service du dépouillement de sources historiques. Éléments d'analyse semi-automatique d'un corpus diplomatique homogène », *Histoire & mesure*, XXXV-2 (2020), p. 171-196, DOI : 10.4000/histoiremesure.13534.
- DIB (Firas), *Regex101 : Build, Test, and Debug Regex*, regex101, URL : <https://regex101.com/> (visité le 05/05/2023).
- Différences Entre Le Format Texte OpenDocument (.Odt) et Le Format Word (.Docx)*, Microsoft, URL : <https://support.microsoft.com/fr-fr/office/diff%C3%A9rences-entre-le-format-texte-opendocument-odt-et-le-format-word-docx-d9d51a92-56d1-4794-8b68-5efb57aebfdc> (visité le 28/06/2023).
- En Apprendre plus Sur Les Formats de Fichiers - Support Microsoft*, Microsoft, URL : <https://support.microsoft.com/fr-fr/office/en-apprendre-plus-sur-les-formats-de-fichiers-56dc3b55-7681-402e-a727-c59fa0884b30> (visité le 28/06/2023).
- GLORIEUX (Frédéric), *Le traitement de textes (odt) pour produire des documents structurés (XML/TEI) — Odette*, 8 avr. 2015, URL : <https://resultats.hypotheses.org/267> (visité le 17/04/2023).
- LaTeX*, Wikilivres, URL : <https://fr.wikibooks.org/wiki/LaTeX> (visité le 22/07/2023).
- LaTeX — DMS*, WikiDMS - Université de Montréal, URL : <https://dms.umontreal.ca/wiki/index.php/LaTeX> (visité le 22/07/2023).
- List of Regular Expressions*, LibreOffice, URL : <https://help.libreoffice.org/7.5/en-US/text/shared/01/02100001.html> (visité le 17/04/2023).
- OpenDocument Text Document Format (ODT)*, Version 1.2, ISO 26300-1 :2015, Sustainability of Digital Formats : Planning for Library of Congress Collections, 22 déc.

2021, URL : <https://www.loc.gov/preservation/digital/formats/fdd/fdd000428.shtml> (visité le 17/04/2023).

Programmation orientée objet, IBM, 22 nov. 2021, URL : <https://www.ibm.com/docs/fr/spss-modeler/saas?topic=language-object-oriented-programming> (visité le 04/08/2023).

REIGNIER (Virgile), *Vers l'indexation automatique du Trésor des chartes. Constitution, alignement et utilisation d'un référentiel d'entités nommées au sein du projet Hymenis*. Mémoire de Master TNAH, Paris, École Nationale des Chartes, Université Paris Sciences et Lettres, 2022.

TEI : Text Encoding Initiative, TEI, URL : <https://tei-c.org/> (visité le 05/04/2023).

Using Regular Expressions in Text Searches, LibreOffice, URL : https://help.libreoffice.org/7.5/en-US/text/swriter/guide/search_regexp.html (visité le 17/04/2023).

Introduction

Les études sur les actes, les documents émanant d'une autorité et agissant en vertu de cette autorité¹, ont dans un premier temps priorisé les actes pontificaux et royaux, « laissant leurs homologues princiers au rang d'actes privés »². Les actes princiers désignent les actes émis par les princes. Le terme prince se doit d'être désambiguïsé dans ce travail. Bien que se rapportant d'ordinaire à l'héritier présomptif du trône royal au sens contemporain du terme, il ne doit ici pas être entendu pour nommer celui qui est premier par le sang ou par le rang, mais bien celui qui appartient à une maison souveraine et ne règne pas³. Les premiers projets consacrés aux actes princiers voient le jour au XIX^e siècle, en Allemagne, offrant ainsi une analyse des pratiques entre princes, empereurs et évêques du Saint-Empire⁴. À l'instar de leurs voisins, les diplomatistes français se concentrent d'abord sur l'imitation des formes et des formules royales et c'est par ce biais que l'on voit émerger les prémisses d'une diplomatique princière. Les premières recherches sont impulsées par le lancement, en 1894, par l'Académie des inscriptions et belles-lettres, d'une nouvelle série, en plus de celles dédiées aux actes royaux et pontificaux, consacrée aux actes des « grands feudataires »⁵. Cette série s'ouvre avec la publication des actes d'Henri II Plantagenêt (1151-1189) relatifs à la France⁶. Malheureusement, cette entreprise n'est pas poursuivie et il faut attendre la fin du XX^e siècle pour assister au départ de nouveaux projets. La plupart de ces derniers se focalisent sur le début du bas Moyen Âge. On observe une raréfaction des éditions de documents des XII^e - XV^e siècle, à l'exception des cas bourguignon, breton et gascon⁷.

1. Olivier Guyotjeannin, Jacques Pycke et Benoît-Michel Tock, *Diplomatique médiévale*, Brepols, Paris, 2006 (L'Atelier du médiéviste, 2).

2. Canteaut, Olivier, Moufflet, Jean-François, « Les éditions d'actes princiers (XII^e - XV^e siècle) : bilan à l'ère du numérique », in : O. Guyotjeannin et Olivier Mattéoni, *Jean de Berry et l'écrit : Les pratiques documentaires d'un fils de roi de France*, Éditions de la Sorbonne, Ecole Nationale des Chartes, Paris, 2019 (Histoire ancienne et médiévale, 159), DOI : 10.4000/books.psorbonne.54103.

3. Définition CNRTL : <https://www.cnrtl.fr/definition/prince>.

4. Johann Friedrich Böhmer, *Regesta Imperii inde ab anno MCCCCXIII usque ad annum MCCCCXLVII*. S. Schmerber, Francfort-sur-le-Main, 1839., Alfons Hüber, *Die Regesten des Kaiserreichs, unter Kaiser Karl IV, 1346-1378*, G. Olms, Innsbruck, 1877.

5. O. Guyotjeannin et O. Mattéoni, *Jean de Berry et l'écrit...*

6. Léopold Delisle et Élie Berger, *Recueil des actes de Henri II, roi d'Angleterre et duc de Normandie, concernant les provinces françaises et les affaires de France*, 4 t., Imprimerie Nationale, Paris, 1909 (Chartes et diplômes relatifs à l'histoire de France).

7. Annexe Répartition chronologique des éditions et catalogues d'actes princiers produits (espace français actuel), in Canteaut, Olivier, Moufflet, Jean-François, « Les éditions d'actes princiers (XII^e - XV^e

Force est de constater, en plus des disparités chronologiques, des disparités géographiques. Concernant les actes des ducs de Bretagne, nous pouvons citer les travaux de Michael Jones, pour les actes de Charles de Blois (1341 - 1364), Jeanne de Penthièvre (1341 - 1365, † 1384) et Jean IV (1365 - 1399), de Marjolaine Lémeillat pour les actes de Pierre de Dreux (1213 - 1221) et de Jean I^{er} (1221 - 1286), ainsi que ceux de Jean Kerhervé sur la chancellerie de Bretagne au temps du duc François II (1458 - 1488), de sa fille la duchesse Anne (1488 - 1514) et du premier mari de celle-ci, le roi de France Louis XII (1498 - 1515)⁸. Le duché de Bourgogne est également l'objet de travaux, parmi lesquels ceux d'Anne-Lise Courtel sur la Bourgogne pour les ducs capétiens, puis, d'Henri Stein pour les ducs Valois, Philippe le Hardi (1363 - 1404), Jean sans Peur (1404 - 1419) et Philippe le Bon (1419 - 1467) et de Sonja Dünnebeil pour les actes de Charles le Téméraire (1467 - 1477)⁹. Pierre Cockshaw a également étudié le personnel de la chancellerie de Bourgogne-Flandre¹⁰. Même si elles sont majoritaires, les entreprises d'édition ne concernent pas que les deux maisons précédemment citées. John Benton et Michel Bur ont étudié les actes du comte de Champagne, Henri le Libéral (1152 - 1181)¹¹. La chancellerie dauphinoise sous Humbert II (1333 - 1349) a aussi engendré le dépouillement d'actes par Chantal Reydelet¹². Ces disparités géographiques et chronologiques s'expliquent par le fait que pour la période étudiée, les corpus sont importants et très dispersés. Nombre des travaux sur les actes princiers se sont limités à des analyses, à des catalogues, plutôt qu'à une édition complète face à l'abondance de la documentation.

siècles) : bilan à l'ère du numérique », in : O. Guyotjeannin et O. Mattéoni, *Jean de Berry et l'écrit...*

8. Michael Jones, *Recueil des Actes de Charles de Blois et Jeanne de Penthièvre : Duc et duchesse de Bretagne (1341-1384). Suivi des Actes de Jeanne de Penthièvre (1364-1384)*, Presses universitaires de Rennes, Rennes, 2016 (Histoire), DOI : 10.4000/books.pur.28420.

Id., *Recueil Des Actes de Jean IV, Duc de Bretagne*, Université de Haute Bretagne - Rennes 2, Rennes, 1980.

Marjolaine Lémeillat, *Actes de Pierre de Dreux, Duc de Bretagne (1213-1237) et Actes de Jean Ier, Duc de Bretagne (1237-1286)*, PUR/Société d'histoire et d'archéologie de Bretagne, Rennes, 2013 (Sources Médiévales de l'histoire de Bretagne, 1-2).

Jean Kerhervé, « La Chancellerie de Bretagne Sous Louis XII et Anne de Bretagne (1499-1514) », dans *Powerbrokers in the Late Middle Ages/Les Courtiers Du Pouvoir Au Bas Moyen Âge*, Brepols, Turnhout, 2001 (Burgundica, IV), p.199-233.

9. Anne-Lise Courtel, « La chancellerie et les actes d'Eudes IV, due de Bourgogne (1315-1349) », *Bibliothèque de l'École des Chartes*, 135-1 (1977), p. 23-71, DOI : 10.3406/bec.1977.450090

Henri Stein, « Catalogue des Actes de Charles le Téméraire (1467-1477) », dans *Urkunden und Mandate Karls von Burgund, Grafen von Charolais, 1433-1467, bearb. von Sonja Dünnebeil*, Jan Thorbecke, Sigmaringen, 1999 (Instrumenta, 3).

10. Pierre Cockshaw, *Le Personnel de La Chancellerie de Bourgogne-Flandre Sous Les Ducs de Bourgogne de La Maison Des Valois (1384-1477)*, U.G.A., Kortrijk-Heule, 1982.

11. John F. Benton, Michel Bur et Michèle Courtois, *Recueil des actes d'Henri le Libéral Comte de Champagne (1152-1181)*, Académie des inscriptions et belles-lettres, Paris, 2009 (Chartes et diplômes relatifs à l'histoire de France).

12. Chantal Reydelet, « La Chancellerie d'Humbert II Dauphin de Viennois (1333-1349) », *Archiv für Diplomatik*, 20 (1974), p. 241-383.

Toutefois, comme l'indique Olivier Canteaut, « le développement de l'édition électronique scientifique et, plus globalement, les perspectives offertes par les humanités numériques stimulent en effet depuis plus d'une décennie la réalisation d'amples corpus d'actes et la diplomatique princière laïque bénéficie de ce regain d'intérêt »¹³. En effet, l'édition numérique offre l'opportunité d'entamer de vastes chantiers d'identification et de regroupement de sources et d'en publier un premier état. Cette opportunité n'est pas négligeable dans le cas où les données sont éclatées sur le territoire, dans différents services d'archives et nécessitent le travail conjoint de plusieurs chercheurs. Ce regain s'applique dans un premier temps à publier les analyses des actes¹⁴, à d'autres supports comme les cartulaires et les chartriers¹⁵, mais aussi aux travaux déjà imprimés. Le projet « Corpus des actes royaux de la France médiévale » (CARo) s'attache à rendre accessible sous forme numérique les actes des rois de France, de Charles le Chauve (875 - 877) à Philippe Auguste (1180 - 1223), édités par l'Institut de France¹⁶. Ces entreprises sont motrices et participent au développement d'éditions dédiées aux actes des grands. C'est dans ce contexte que s'inscrit le projet consacré aux actes princiers.

Le projet « Actes princiers au royaume de France (XIV^e - XVI^e siècle) » est porté par le professeur des universités Olivier Mattéoni et rassemble quatre institutions : deux unités mixtes de recherche (UMR) que sont le Laboratoire de Médiévistique occidentale de Paris (LaMOP, Paris 1 - CNRS¹⁷), qui porte le projet et le Centre de recherches historiques (CRH, CNRS - EHESS), le centre de recherche de l'École Nationale des Chartes, Centre Jean Mabillon, et le département du Moyen Âge et de l'Ancien régime des Archives nationales. L'ambition du projet est de procéder à l'édition numérique de plusieurs corpus d'actes inédits issus des chancelleries des princes de sang de la fin du Moyen Âge. Ces corpus comprennent les actes des ducs et duchesses de Bourbon et du duc Jean de Berry (1360 - 1416). Leur nombre est voué à s'accroître et à comporter par la suite d'autres actes princiers comme ceux de la maison d'Orléans et de la maison d'Anjou. Le projet comprend le développement d'un site web et l'encodage d'une première série d'actes : ceux de Louis II (1356 - 1410) et de Charles I^{er} (1434 - 1456) de Bourbon et de leurs épouses Anne Dauphine (1371 - 1410, † 1417) et Agnès de Bourgogne (1434 - 1456, † 1476)¹⁸.

13. Canteaut, Olivier, Moufflet, Jean-François, « Les éditions d'actes princiers (XII^e - XV^e siècle) : bilan à l'ère du numérique », in :O. Guyotjeannin et O. Mattéoni, *Jean de Berry et l'écrit...*

14. Les actes pontificaux en France, <https://qed.perspectivia.net/gallia-pontificia-online/>.

Regesta Imperii Online, http://opac.regesta-imperii.de/lang_en/.

15. Digitale Monumenta Germaniae Historica <https://www.dmgm.de/>.

Cartulaires d'Île-de-France <http://elec.enc.sorbonne.fr/cartulaires/>.

Base des actes normands médiévaux Scripta <https://www.unicaen.fr/scripta/>.

Corpus de la Bourgogne au Moyen Âge CMBA <http://www.cbma-project.eu/>.

16. Corpus des actes royaux, en ligne : <https://lamop.hypotheses.org/category/corpus-des-actes-royaux-caro>.

17. Laboratoire de médiévistique occidentale de Paris, <https://lamop.pantheonsorbonne.fr/>.

18. Annexe Généalogie de la famille ducale de Bourbon.

Louis II de Bourbon devient duc en 1356, dans des conditions difficiles. À la suite de la défaite de Poitiers, il est envoyé comme otage en Angleterre en échange de la libération du roi de France Jean II (1350 - 1364)¹⁹. Ses années de captivité entravent considérablement la gouvernance du duché. À son retour, en 1366, il soude la noblesse autour de l'ordre de chevalerie de l'écu d'or et mène une politique d'expansion territoriale par le biais d'alliances. Son mariage avec Anne de Forez, héritière du comté de Forez, en 1371, sanctionne le rattachement du comté à la principauté bourbonnaise. En 1400, le mariage de son fils Jean de Bourbon (1410 - 1434) avec Marie de Berry (1416 - 1434) prévoit l'intégration du duché d'Auvergne à la mort du duc Jean de Berry, intégration qui sera effective en 1425. L'Auvergne est un apanage qui doit revenir au domaine royal dans le cas où le duc, père de Marie de Berry, n'a pas d'héritier mâle. Toutefois, Jean de Berry et Louis II de Bourbon obtiennent du conseil royal et de Charles VI (1380 - 1422) que cet apanage revienne aux Bourbons du fait du mariage de leurs enfants. Cette expansion participe au développement institutionnel notable du duché impulsé par le duc, qui se traduit par d'importantes réformes avec la création d'un office de trésorier général (1372) et d'une chambre des comptes (1374)²⁰. Olivier Mattéoni parle d'une « fièvre normative » pour les domaines des finances et de la justice²¹. L'agrandissement de la principauté et les réformes qui en découlent génèrent une importante production documentaire d'actes écrits servant à créer ou confirmer des actions juridiques²². La production de ces actes s'inscrit dans un projet de réforme du gouvernement rendu nécessaire par l'agrandissement de la principauté. Les années 1370 sont marquées par une réorganisation de la chancellerie avec l'augmentation du nombre de secrétaires et une rationalisation des règles de confection des actes avec notamment la généralisation et l'obligation de la signature²³.

En 1410, Jean I^{er} succède à son père à la tête d'une principauté bien administrée. Il est duc de 1410 à 1434, mais exerce peu la réalité du pouvoir, car il est fait prisonnier par les Anglais en 1415 à Azincourt²⁴. La production documentaire du duc Jean I^{er} et de son épouse Marie de Berry est exclue de cette étude. Le travail de recensement des actes s'est

19. « Poitiers », in : Jean Favier, *La Guerre de Cent Ans*, Fayard, Paris, 1980, p.181.

20. « Les ducs de Bourbon, le roi et le royaume de France à la fin du Moyen Âge », in : O. Mattéoni, *Les Bourbons en leur bibliothèque (XIIIe-XVIe siècle)*, Éditions de la Sorbonne, Paris, 2022 (Histoire ancienne et médiévale).

21. Id., « À propos d'une "ordonnance" de finance attribuée à Louis II de Bourbon : une relecture », dans *Cultures fiscales en Occident du Xe au XVIIe siècle. Études offertes à Denis Menjot*, Presses universitaires du Midi, Toulouse, 2019 (Études médiévales ibériques, 17), p. 153-163, URL : <https://shs.hal.science/halshs-02972959> (visité le 06/07/2023).

22. O. Guyotjeannin, J. Pycke et B.M. Tock, *Diplomatique médiévale...*

23. O. Mattéoni, « Écriture et pouvoir princier. La chancellerie du duc Louis II de Bourbon (1356-1410) », dans *"De part et d'autre des Alpes" (II) : chancelleries et chanceliers des princes à la fin du Moyen Âge ; actes de la table ronde de Chambéry, 5 et 6 octobre 2006*, Université de Savoie, Chambéry, 2011 (Collection Sociétés, religions, politiques, 19), t. 2, p. 137-178.

24. « Les ducs de Bourbon, le roi et le royaume de France à la fin du Moyen Âge », in : Id., *Les Bourbons en leur bibliothèque (XIIIe-XVIe siècle)...*

concentré sur ceux de Louis II et Charles I^{er} et de leurs épouses du fait de leurs réformes dans l'administration bourbonnaise et des acquisitions territoriales importantes sous leurs règnes. De plus, les actes dénombrés pour le couple Jean I^{er} et Marie de Berry sont très peu nombreux : seuls cinquante-neuf actes ont pu être recensés (et quarante-quatre d'entre eux sont établis avant 1415)²⁵. Pendant la captivité du duc, les femmes administrent le duché : d'abord Anne Dauphine, comme duchesse douairière jusqu'en 1417, puis Marie de Berry. La période est caractérisée par la réconciliation avec la maison de Bourgogne qui se concrétise par l'alliance matrimoniale entre Charles de Bourbon et Agnès de Bourgogne, la fille du duc de Bourgogne Jean Sans Peur, en 1425.

Charles I^{er} devient duc en 1434, mais en raison de la captivité de son père, il assure le gouvernement du duché avant. Il bénéficie d'une proximité importante avec le roi de France Charles VII (1422 - 1461) qui le nomme lieutenant en Languedoc et dans les marches est du royaume et dont il est le conseiller dès les années 1420²⁶. Il est également choisi par ce dernier pour diriger la délégation royale à la paix d'Arras, en 1435. Son mariage avec Agnès de Bourgogne le met en possession du duché d'Auvergne, ce qui contribue à affirmer son influence. Il noue des alliances avec les autres princes du royaume et entre en rébellion contre le roi lors de la Praguerie en 1440²⁷. Après l'échec de cette dernière, il se concentre sur l'administration et les finances avec la création de l'office de contrôleur général. Son principat est caractérisé par le renforcement des cadres institutionnels de la principauté (finances, fiscalité...)²⁸. Il cherche également à s'assurer le soutien de la noblesse, comme le montre la création de l'Armorial de Bourbonnais et de Forez²⁹. La cour des ducs de Bourbon, implantée à Moulin, devient alors un centre artistique accueillant sculpteurs, musiciens, écrivains³⁰. L'épouse de Charles I^{er}, Agnès de Bourgogne, n'est pas sans jouer un rôle diplomatique. Elle fait régulièrement l'intermédiaire entre son mari et son frère, et s'attache à mener une politique de pacification entre les deux familles³¹. Sept de ses onze enfants sont notamment élevés à la cour de Bourgogne. La fin du principat de Charles I^{er} est marquée par sa maladie. Il meurt le 4 décembre 1456 et son épouse lui survit jusqu'en 1476.

25. Jean-Damien Généro, *La chancellerie de Charles Ier de Bourbon, suivi de l'édition de ses actes et de ceux de son épouse, Agnès de Bourgogne (1421-1476)*, Mémoire de Master recherche, Paris, Paris I Panthéon-Sorbonne, 2018.

26. André Leguai, « De La Seigneurie à l'Etat. Le Bourbonnais Pendant La Guerre de Cent Ans », *Annales. Histoire, Sciences Sociales*, Les Imprimeries Réunies 30–2–3 (juin 1975), p. 365-366, doi : 10.1017/S0395264900169057.

27. En 1440, révolte des grands contre Charles VII (à laquelle se joignit le Dauphin).

28. « Les ducs de Bourbon, le roi et le royaume de France à la fin du Moyen Âge », in : O. Mattéoni, *Les Bourbons en leur bibliothèque (XIII^e-XVI^e siècle)*...

29. Armorial : Registre ou catalogue contenant les armes ou armoiries de la noblesse d'un royaume, d'une province, d'une ville, d'une famille, dessinées, peintes ou décrites. Définition CNRTL. Emmanuel de Boos, « L'armorial d'Auvergne », dans *Bourbonnais et Forez*, Créal, Nonette, 1998 (édition et étude).

30. A. Leguai, « De La Seigneurie à l'Etat. Le Bourbonnais Pendant La Guerre de Cent Ans »...

31. Id., « Agnès de Bourgogne, Duchesse de Bourbon », *Etudes bourbonnaises* (, 1996), p. 405-417.

L'objectif du stage dans le projet « Actes princers » est de participer à la conception d'une chaîne de traitement afin de transformer des fichiers ODT contenant les éditions diplomatiques d'actes bourbonnais en des fichiers XML (TEI), pour assurer leur mise en ligne. L'un des enjeux est de tester les différentes solutions de transformation pour les documents afin de déterminer la plus adaptée au corpus, et pour des besoins futurs. Ensuite, les métadonnées associées aux corpus sont extraites via des scripts Python afin de les incorporer dans les fichiers XML/TEI produits. Ces derniers sont intégrés à une application Flask prototypée pour leur mise en ligne.

Ce travail est accompagné d'un livrable technique situé dans un repository GitLab³². Il comprend les fichiers contenant les actes édités sous différents formats (ODT, PDF, XML) ainsi que les fichiers CSV contenant les métadonnées des actes. Il comprend également une chaîne de traitement de données (scripts Python) codée par les ingénieurs du projet, fonctionnelle et documentée. Il est complété par le descriptif des étapes du travail et du code nécessaire à la réflexion sur l'édition numérique et la mise en ligne d'actes princiers.

L'enjeu de ce travail de recherche autour du projet « Actes Princiers » est de contribuer à une meilleure connaissance de la diplomatique princière qui est encore obscure pour la fin du Moyen Âge en France.

Ce projet d'édition s'articule d'abord autour de l'édition des actes dispersés des princes de Bourbon, Louis II, Anne Dauphine, Charles I^{er} et Agnès de Bourgogne.

Il s'accompagne ensuite d'une importante réflexion sur la transformation des éditions, du traitement de texte vers l'encodage, afin de choisir les opérations de préparation, le langage d'encodage et le convertisseur les plus adaptés.

Enfin, l'aboutissement du projet réside dans la conception d'une chaîne de traitement des actes via la technologie Python, afin de les pourvoir d'une structuration optimale nécessaire à la mise en ligne des données.

32. Justine Chainiau, Chaîne de traitement des actes princiers, en ligne : https://gitlab.huma-num.fr/medieval-acts/principely-acts/stage_actes_princiers.git.

Première partie

L'édition numérique des corpus de
Louis II, d'Anne Dauphine et de
Charles I^{er} et d'Agnès de Bourgogne

Chapitre 1

Présentation du corpus

Le corpus dénombre huit cent vingt-trois actes, avec presque autant d'actes pour chaque couple de prince. Quatre cent douze actes sont recensés pour Louis II (333) et Anne Dauphine (79) et quatre cent six actes pour Charles I^{er} (379) et Agnès de Bourgogne (32). L'étude de ce corpus est permise par la conservation des différents actes étudiés. Lorsqu'on y regarde de plus près, on est vite confronté à la dispersion de la documentation. Aucune série documentaire ne se rapporte à la chancellerie ni ne contient l'enregistrement des actes des ducs de Bourbon par cette dernière¹.

1. O. Mattéoni, « Écriture et pouvoir princier. La chancellerie du duc Louis II de Bourbon (1356-1410) »...

1.1 Dispersion des actes

Les actes des Bourbons sont conservés depuis le XII^e siècle². Sous la première maison de Bourbon, les seigneurs de Bourbon-Dampierre, les pièces originales sont déposées dans des coffres ou des layettes et les lettres sont petit à petit transcrrites dans des cartulaires. L'accroissement de la production documentaire donne lieu à la réalisation d'inventaires dès le XIV^e siècle. Après la création de la chambre des comptes de Moulins, en 1374, le duc Louis II ordonne que tous les actes, terriers, lettres et écritures touchant au domaine (nominations, lettres de dons et de grâce, ordonnances) y soient remis³. La chambre des comptes se sédentarise à Moulins et fonctionne avec une régularité croissante. Le « Trésor des chartes » du Bourbonnais s'agrandit et est complété par les inventaires des actes conservés dans les territoires nouvellement acquis. En effet, Louis II maintient les Chambres des comptes du comté de Forez à Montbrison et de la seigneurie de Beaujolais à Villefranche. En 1527, cinq ans après la fuite du dernier duc de Bourbon, le connétable Charles III (1505 - 1523), François I^{er} (1515 - 1547) confisque tous les biens féodaux des Bourbons tenus en apanage⁴. L'administration des domaines revient alors à la mère du roi, Louise de Savoie (1476 - 1531), petite fille du duc de Bourbon Charles I^{er}, qui en revendique alors la possession. À sa mort, cet héritage contesté retourne à la couronne. Les chambres des comptes du duché de Bourbon sont supprimées, leur juridiction est transférée à la chambre des comptes de Paris et le Trésor des chartes du Bourbonnais y est rapatrié après la réalisation d'un inventaire par Jacques Luillier⁵. Au début du XVIII^e siècle, pour éviter les pertes, les liasses sont reliées et converties en volumes et les sceaux sont coupés. Le tout est réuni dans le dépôt de la chambre d'Anjou. La série de titres et de registre qui était restée à Moulins est détruite sous le Directoire comme entaché de féodalité, celle de Villefranche subit de nombreuses pertes et les archives du Forez à Montbrison sont transférées à Lyon au bureau des trésoriers de France. L'inventaire des *Titres des Bourbons*⁶, édité par les archivistes Alphonse Huillard-Breholles et Albert Lecoy de la Marche au XIX^e siècle, dresse, de manière chronologique, la liste des actes des trois chambres des comptes du duché qui sont conservés dans la série P des Archives nationales, dans les fonds des anciennes Chambre des comptes du duché de Bourbon (P 1355 à P 1402)⁷. La majorité des actes du corpus (318) provient de cette série P, mais quelques actes ont également été identifiés dans les séries J (43) et K (7)⁸.

2. Jean Louis Alphonse Huillard-Breholles et Albert Lecoy de La Marche, *Titres de la maison ducal de Bourbon*, Plon, 1867, URL : http://archive.org/details/bub_gb_POUbAQAAQAAJ (visité le 24/06/2023).

3. *Ibid.*

4. *Ibid.*

5. *Ibid.*

6. *Ibid.*

7. Archives nationales, Série P, Chambre des comptes et comptabilité, en ligne : <http://www.archivesnationales.culture.gouv.fr/chancery/fonds/EGF/SA/SAPDF/Egfn-p.pdf>.

8. Série J, Trésor des Chartes - Série K, Monuments historiques.

Une autre partie importante du corpus est issue des services d'archives départementales (221). La localisation de ces institutions de conservation révèle la géographie du duché de Bourbon. La majorité des actes identifiés aux Archives départementales provient du service de la Loire (150). Le territoire de ce département correspond à celui de l'ancien comté de Forez, qui est rattaché à la principauté par héritage lors du mariage, en 1371, entre Louis II et Anne Dauphine. D'autres actes sont conservés dans les services de l'Allier, sur le territoire du duché de Bourbonnais, du Puy-de-Dôme, sur le territoire du duché d'Auvergne et du Cantal, dont une partie était autrefois rattaché au duché d'Auvergne. Les actes conservés dans les services d'archives municipales nous éclairent également sur la géographie de la production documentaire des ducs de Bourbon. Des actes ont été identifiés dans les services d'archives de Moulins, capitale du duché. Louis II y fait construire un château en 1375 et s'y installe définitivement. Ses successeurs font de même et la ville devient alors le siège de la cour ducale et un centre culturel important⁹. D'autres actes ont été retrouvés aux Archives municipales de Riom, la capitale du duché d'Auvergne, et de Montluçon, une autre ville d'envergure. Ils sont, pour la grande majorité, relatifs à l'administration de la principauté bourbonnaise.

La localisation des institutions de conservation témoigne également des relations entretenues par les Bourbons avec les autres princes. Des actes proviennent des services d'archives départementales de la Côte-d'Or (24), et du Nord (ancien duché de Flandres) pour le corpus de Charles I^{er}, et témoignent des alliances et accords passés avec les ducs de Bourgogne. D'abord, son mariage, en 1425, avec Agnès, la fille du duc Jean Sans Peur¹⁰, puis des traités et des abstinences de guerre conclues avec Philippe le Bon¹¹. En moindres proportions, des actes ont été recensés dans les services d'archives départementales de la Loire-Atlantique et du Tarn. Ils témoignent des liens entre les différents princes et des différentes alliances et trêves conclues pendant la guerre de Cent Ans¹². Le reste des actes du corpus provient de la Bibliothèque nationale de France (172), notamment de la collection Gaignière¹³, ou de collections d'érudits des XVII^e - XVIII^e siècles (La Mure, Hozier, Tripperet).

Le corpus étudié est donc dispersé géographiquement, institutionnellement et chronologiquement dans la mesure où il se déploie sur plus d'un siècle.

9. « Les ducs de Bourbon, le roi et le royaume de France à la fin du Moyen Âge », in : O. Mattéoni, *Les Bourbons en leur bibliothèque (XIII^e-XVI^e siècle)*...

10. AN, P 1370², c. 1919. (4 février 1425).

11. AD Côte-d'Or, B 11918, c. 118. (4 décembre 1434).

12. AD Nord, B 304, c. 15.660 (7 septembre 1435) — AD Loire-Atlantique, E 121-1, c. 42 (4 avril 1441).

13. BnF, ms. fr. 22388 et 22390.

1.2 Répartition chronologique des actes

Le corpus étudié couvre la période allant de 1357 à 1475. Nous pouvons établir une première période allant de 1357 à 1410 pour Louis II et jusqu'à 1417 pour Anne Dauphine. Puis, une seconde période de 1421 à 1456 pour Charles I^{er} et jusqu'à 1475 pour Agnès de Bourgogne. Les veuves des deux ducs leur survivent et continuent à produire des actes en raison des circonstances politiques. Anne Dauphine continue à émettre des actes pour le comté de Forez, dont elle est l'héritière, et au début de la captivité de Jean I^{er}.

À partir du graphique ci-après, nous pouvons établir une moyenne d'environ dix actes par pour notre corpus. Le nombre d'actes est en deçà au début du principat de Louis II. En revanche, on observe une augmentation des écrits à partir des années 1380, qui fait suite à ses réformes administratives. Ensuite, le nombre d'actes recensés est plutôt stable jusqu'à sa mort.

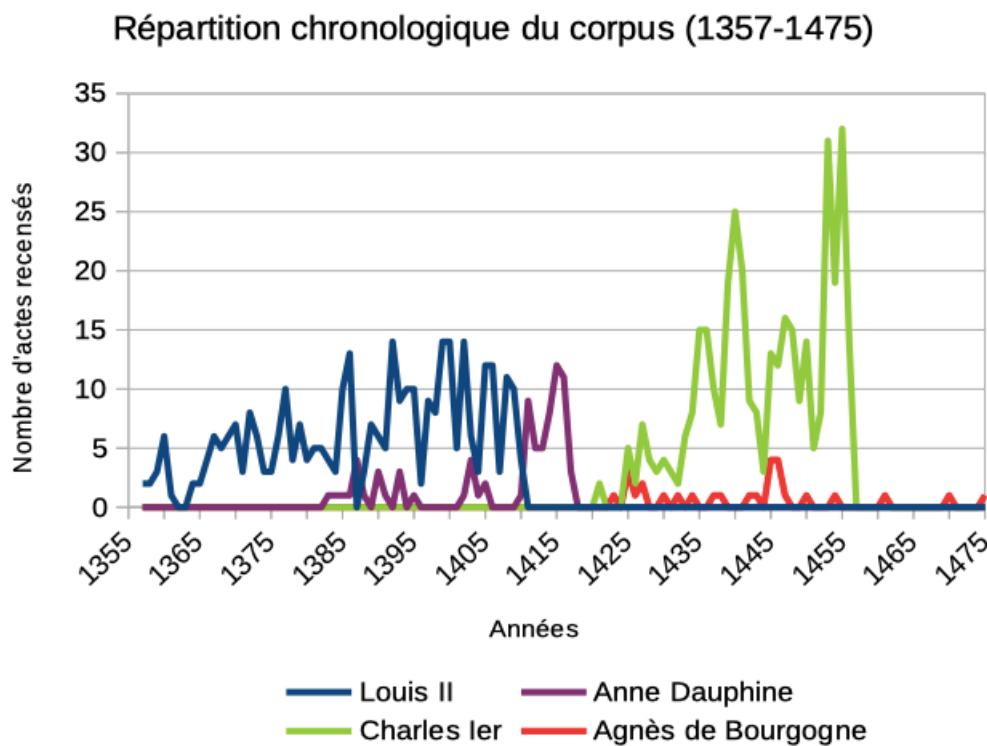


FIGURE 1.1 – Répartition chronologique des actes du corpus.

La diminution du nombre d'actes entre le principat de Louis II et celui de Charles I^{er} s'explique par le fait que Charles I^{er} n'est pas le successeur direct de Louis II d'une part. D'autre part, les actes du duc Jean I^{er} et de son épouse Marie de Berry ne sont pour l'instant pas traités par le projet *Actes princiers*. Néanmoins, des actes sont quand même le fait de la duchesse Anne Dauphine, pour laquelle on observe une nette augmentation à la mort de son époux, et de Charles de Bourbon, qui commence à administrer le duché pendant la captivité de son père. Le nombre d'actes recensés pendant le principat de Charles I^{er}, à partir de 1434, est caractérisé par une augmentation par rapport à la période de Louis II, ce qui peut faire penser à une émission de plus d'actes sur une plus courte période. Cette augmentation peut s'expliquer par la consolidation d'une administration déjà stable, mais surtout par le biais documentaire qu'induit le corpus. Nous ne pouvons en effet pas chiffrer la part de la production totale de la chancellerie bourbonnaise qui nous est parvenue par rapport à celle qui a été perdue. L'un des pics de production peut être documenté plus assurément. Il s'agit de l'augmentation au début des années 1440, pendant la Praguerie¹⁴, avec de nombreuses nominations de capitaines-châtelain et de capitaines de guerre¹⁵. Cependant, le second pic en 1455, peu avant la mort du duc, avec de nombreuses nominations et fondations pieuses afférentes au duché¹⁶, ne peut réellement être explicité. Il est certainement plutôt dû à un effet de source qu'à un regain administratif à la fin de la vie de Charles I^{er}. En effet, nous conservons un registre d'enregistrement d'actes de la Chambre des comptes du comté de Forez, à Montbrison, qui contient des copies de lettres de nomination d'officiers des années 1450 aux années 1470¹⁷.

L'émission des actes est plutôt constante malgré quelques variations dues au contexte politique du duché pour la période étudiée. Néanmoins, il ne faut pas perdre de vue que ce graphique omet les actes qui ont été perdus, ou qui n'ont pas encore été retrouvés, et comptabilise le corpus étudié, soit les actes qui ont été identifiés. Ces derniers représentent une masse conséquente aux natures et aux typologies variées.

14. J.D. Généro, « Chanceliers et secrétaires de la Chancellerie de Charles Ier de Bourbon et d'Agnès de Bourgogne (1425-1456) », *Bulletin trimestriel de la Société bourbonnaise des études locales*, Bulletin trimestriel de la Société bourbonnaise des études locales 365-23 (2021), p. 1-17.

15. BnF, ms. fr. 22299, folio 9 (20 juillet 1440) — BnF, ms. fr. 22299, folio 9 (30 juillet 1440).

16. O. Mattéoni, *Servir le prince : les officiers des ducs de Bourbon à la fin du Moyen Âge, 1356-1523*, Paris, Paris I Panthéon-Sorbonne, 1994.

17. AD Loire, B 1844.

1.3 Typologie des actes

L'étude du corpus révèle les différentes typologies documentaires. Parmi les actes à forte valeur juridique, on trouve des chartes, dont la valeur est perpétuelle, et des lettres patentées qui accordent une faveur au destinataire. Dans une moindre mesure, on trouve aussi des lettres closes, fermées d'un sceau, et des lettres missives, ces dernières ayant plus un caractère politique puisqu'elles sont destinées à une correspondance officielle. Des écrits sont dédiés à l'administration du duché, en témoignent les nombreux mandements, des ordres du prince, concernant la gestion du domaine et des quittances qui attestent du remboursement de sommes dues. Certains actes sont plus singuliers, comme les accords passés avec d'autres princes, dans la mesure où ils comportent les titulatures des deux individus¹⁸, et les différentes versions des testaments des ducs de Bourbon¹⁹. Pour chaque duc et duchesse, on recense ces différents types d'actes. Les corpus sont diversifiés, hormis celui d'Agnès de Bourgogne (essentiellement des quittances et des missives).

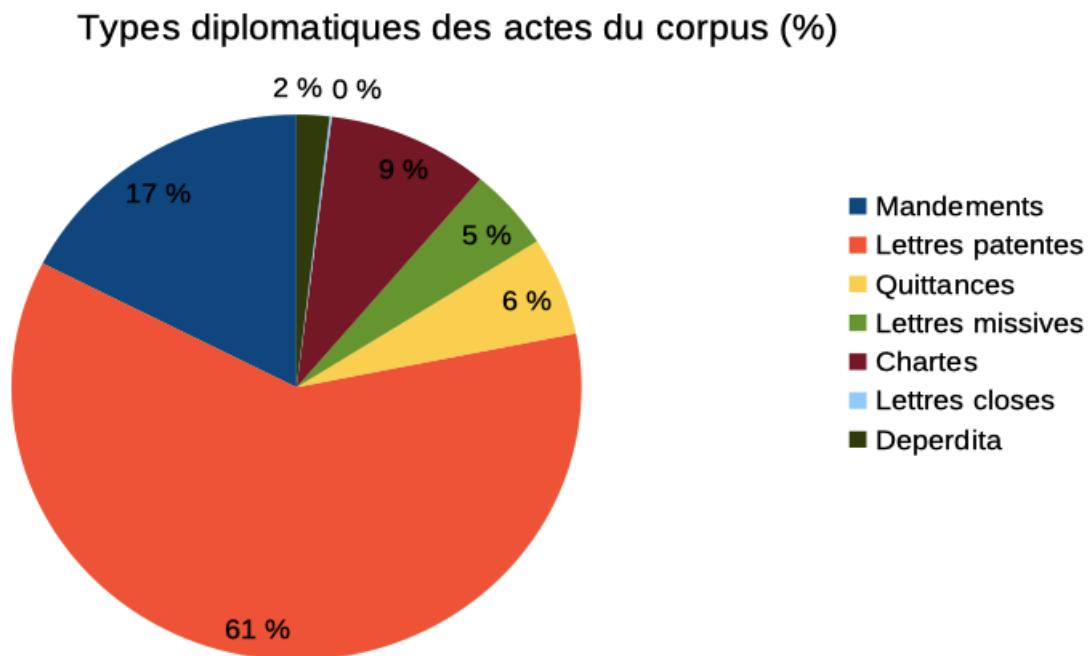


FIGURE 1.2 – Typologie des actes étudiés.

18. AN, P 1357², n° 428 (avril 1372).

19. AN, P 1370¹, c. 18791 (12 décembre 1428) — AN, P 1370¹, c. 1880 (4 décembre 1456).

Comme le montre le graphique ci-avant, les lettres patentes sont majoritaires. Ce sont des actes à caractère juridique exprimant la volonté du duc. Ces lettres sont généralement destinées à des particuliers et beaucoup d'entre elles entérinent des nominations d'officiers, des souffrances d'hommages ou encore des donations ...²⁰. Le deuxième type d'acte le plus représenté est le mandement. Ils concernent en majorité Louis II, et ce tout au long de la période²¹, et dans une moindre mesure Charles Ier²². Ces actes ont trait à l'administration instantanée du duché : paiement de taxes, impositions extraordinaires, saisies, réparations diverses et fortifications, enquêtes, requêtes ...²³. Viennent ensuite les quittances et les chartes. Cette production a trait principalement aux finances (aides et paiements), à l'administration, ces deux aspects étant intimement liés, et enfin aux relations entretenues avec le roi et les autres princes. Les documents relatifs à la gestion du domaine font preuve d'une grande variété : aides, documents relatifs à l'organisation des marchés et des foires dans les localités relevant de la juridiction²⁴, et dans une moindre mesure des fondations pieuses²⁵. Quelques mentions de trêves, d'organisation d'États ont également été relevées. Une partie de la production documentaire du duché est liée au contexte de guerre de Cent Ans : des actes mentionnent des autorisations de fortification²⁶, des nominations de capitaines²⁷ et des garnisons. Les documents afférents aux relations du duc avec les grands du royaume enregistrent des alliances²⁸, des trêves²⁹, des promesses³⁰, mais aussi des ambassades³¹, des ratifications de paix³² ou des tractations relatives aux otages³³. L'autre partie consiste en des ratifications de conventions de mariage³⁴ et en paiements de dotes³⁵.

20. AN, P 1374², n° 2420 (19 février 1403) — AD Loire, B 1844, folio n°4 recto (21 mai 1453).

21. AN, P 1376², n° 2749 (27 février 1457) — AN, P 1389³, n° 370 (30 juin 1410).

22. BnF, ms. fr. 20389, c. 78 (20 juin 1444).

23. AN, K 48, n° 6 (5 mai 1360) — AM Montluçon, AA 8 (1) (20 novembre 1360) — AM Montluçon, AA 2 (1) (1368) — AM Montluçon, n° 240, boîte 3 (7 novembre 1373) — AN, P 1376², n° 2716 (25 août 1376) — AN, P 1379², c. 3136/16 (18 août 1437) — AN, P 13911, c. 562 (23 mai 1447).

24. AN, P 1376², n° 2752 (10 septembre 1366).

25. AN, P 1355¹, n° 46 (18 août 1392).

26. AD de la Loire, B 1952 (1^{er} mai 1435).

27. BnF, ms. fr. 22299, folio 7 (24 octobre 1439).

28. AN, P 1372², c. 2113, (4 août 1427).

29. AD de la Côte-d'Or, B 11918, c. 117 (24 octobre 1433).

30. AD de la Côte-d'Or, B 11904, c. 67 (6 février 1436).

31. AN, P 1360², c. 881 (28 mai 1431).

32. AD du Nord, B 304, c. 15.660 (1^{er} octobre 1435).

33. Mention : Aubret L., Mémoires pour servir à l'histoire de Dombes, II, Guigue M.-C. (éd.), Trévoux, 1868, p. 617. Indication de provenance : « Titres de Trévoux ».

34. AN, P 1370², c. 1919 (4 février 1425).

35. AD de la Côte-d'Or, B 299, pièce scellée 338 (8 décembre 1425).

Malgré quelques tensions avec le roi lors de la Praguerie, Charles I^{er} mène une politique d'alliances par les mariages de ses enfants (Anjou, France, Bourgogne)³⁶ et affirme sa place au rang de prince via l'écrit et la représentation. Cette production, riche et abondante, participe à la consolidation administrative qui caractérise la période. En témoigne la création des offices de gouverneur général des finances (1437) et de contrôleur général (1445)³⁷. En fin de vie, on observe une augmentation des fondations pieuses ainsi que la présence de lettres personnelles adressées à des membres de la famille, c'est notamment le cas pour le corpus d'Agnès de Bourgogne³⁸. Quelques *deperdita* ont été recensés, des actes perdus ou détruits dont l'existence est attestée par une mention. Néanmoins, ils ne sont pas très représentatifs de l'ensemble du corpus puisqu'ils font en grande majorité partie du corpus de Charles I^{er}. Les archives des ducs de Bourbon sont progressivement versées à la chambre des comptes de Paris lors du rattachement du duché à la couronne. Une partie de ces actes a disparu dans l'incendie de la chambre des comptes de 1737, et nous est connu par des inventaires³⁹. La collection de l'historiographe et collectionneur François Roger de Gaignières, donne les analyses de nombreux actes de notre corpus⁴⁰.

Les actes recensés révèlent des typologies variées et offrent, malgré les pertes documentaires, un riche extrait de ce à quoi pouvait ressembler la production documentaire pour la période étudiée. Nonobstant la diversité typologique qui les caractérise, les actes émis par les ducs de Bourbon ne sont pas dépourvus de ressemblance avec les actes royaux.

36. AN, P 1370², c. 1915 (14 février 1437).

37. *Ibid.*

38. AN, K 188, c. 304 (18 novembre 1470) — BnF, ms. fr. 15538, folio 138 (11 novembre, après 1473).

39. Michel Nortier, « Le sort des archives dispersées de la Chambre des Comptes de Paris », *Bibliothèque de l'École des chartes*, 123–2 (1965), p. 460–537, DOI : 10.3406/bec.1965.449701.

40. BnF, ms. fr. 22388 et 22390.

1.4 L'*Imitatio regis* au cœur des pratiques documentaires des princes de Bourbon

Les princes du royaume de France adoptent le modèle royal en imitant les usages diplomatiques de la chancellerie royale. Même si ce modèle ne s'impose pas unanimement pour des raisons politiques (Bretagne, Bourgogne⁴¹) ou à cause de l'usage de pratiques locales, la majorité des actes émis par les chancelleries ducales présentent de nombreuses similitudes avec les actes royaux. Olivier Canteaut relève que « l'imitation ou le rejet des modèles sont autant d'instruments qui permettent aux autorités d'affirmer leur domination sur leurs sujets et leur identité politique face à leurs correspondants étrangers »⁴². En tant que branche cadette de la famille royale depuis le XIII^e siècle (descendants de Saint-Louis (1226 - 1270)), les Bourbons bénéficient de l'office héréditaire de grand chambrier de France⁴³ et d'une proximité particulière avec les rois de France⁴⁴. La participation de Louis II au gouvernement du royaume dès la minorité de son neveu Charles VI, ainsi que ses nombreux déplacements accompagnés de sa chancellerie, ont favorisé la circulation des conseils, des savoir-faire et des modèles⁴⁵. La chancellerie est l'organe producteur de l'écrit princier, mais aussi de l'image et du discours du prince⁴⁶. Celle des ducs de Bourbon, localisée à Moulins, est semblable en de nombreux points à la chancellerie royale dans son organisation, mais aussi dans la production des écrits. De plus, les XIV^e - XV^e siècles voient l'émergence et la réapparition de la figure du chancelier, l'officier chargé de l'écrit et de la garde du sceau. Il est en ce sens le responsable de la mise par écrit de la volonté politique du prince par la vérification des actes et leur signature une fois ceux-ci conformes. Olivier Guyotjeannin indique que « se doter d'un chancelier, homme de culture juridique, rompu aux jeux du pouvoir, attiré vers la justice en ce qu'elle est le condensé du gouvernement des hommes, apparaît ainsi bien souvent comme le premier acte de l'identité diplomatique princière »⁴⁷. Les officiers en charge de l'écrit princier sont les secrétaires, des professionnels de l'écrit public. Après délibérations du conseil ducal, et en fonction des directives de ce dernier, les secrétaires rédigent des minutes, des premières rédactions d'un acte. Les minutes sont ensuite grossoyées au propre par des clercs. La

41. Originalité qui va disparaître lorsque les Valois mettront la main sur le duché, A.L. Courtel, « La chancellerie et les actes d'Eudes IV, due de Bourgogne (1315-1349) »...

42. « Introduction », in : Olivier Canteaut, *Le discret langage du pouvoir, Les mentions de chancellerie du Moyen Âge au XVIIe siècle*, École Nationale des Chartes, Paris, 2019 (Etudes et rencontres de l'Ecole des Chartes, 55).

43. Grand chambrier : grand officier de la couronne chargé de l'intendance de la chambre du roi et de la garde du trésor royal. Définition CNRTL : <https://www.cnrtl.fr/definition/chambrier>.

44. « Les ducs de Bourbon, le roi et le royaume de France à la fin du Moyen Âge », in : O. Mattéoni, *Les Bourbons en leur bibliothèque (XIIIe-XVIe siècle)*...

45. Id., « Écriture et pouvoir princier. La chancellerie du duc Louis II de Bourbon (1356-1410) »...

46. J.D. Généro, *La chancellerie de Charles Ier de Bourbon, suivi de l'édition de ses actes et de ceux de son épouse, Agnès de Bourgogne (1421-1476)*...

47. Conclusion par Olivier Guyotjeannin O. Mattéoni, « Écriture et pouvoir princier. La chancellerie du duc Louis II de Bourbon (1356-1410) »...

production documentaire relative à l'administration du duché de Bourbon est exécutée par des hommes qualifiés qui irriguent le milieu politico-administratif et financier bourbonnais. Ce sont majoritairement des érudits laïcs (licenciés en droit ou bacheliers en décret), dont les qualités mentionnées dans les lettres de nomination font référence à la science, à la probité, à la diligence et à la vertu⁴⁸. On dénombre quinze officiers chargés de l'écrit ducal sous Louis II⁴⁹ et trois chanceliers et vingt-huit secrétaires entre 1425 et 1456 (quatorze pour Charles I^{er} et neuf pour Agnès de Bourgogne)⁵⁰. Ils sont généralement issus de la bourgeoisie locale, sauf quelques exceptions. Parmi elles, nous pouvons citer le cas d'Etienne de Bar, talentueux secrétaire venu de Paris, auparavant secrétaire du roi⁵¹. Même s'il reste isolé, ce cas de figure a pu jouer dans l'adoption des pratiques d'écriture en vigueur à la chancellerie royale.

Cette pratique d'*imitatio regis*⁵² est visible dans les écrits, au sein des caractères externes des actes. La réalisation est soignée, l'écriture conforme au modèle royal, avec des abréviations canoniques et peu nombreuses. L'organisation des différentes parties du discours suit celle des actes royaux sur un format horizontal : exposé, mentions hors teneurs, signatures organisées autour d'un paraphe cadre. Des emprunts très caractéristiques ont trait à l'ornementation avec l'utilisation de fleurs de lys, motifs typiques de l'iconographie des chartes royales contemporaines des Valois⁵³. Ce discours dynastique et sacré est également présent sous le principat de Charles I^{er}⁵⁴. On remarque des anges sur le sceau ducal et un usage plus affirmé du sceau du secret à partir des années 1440. Cette *imitatio* est également manifeste dans les caractères internes des actes. Le vocabulaire de l'exposé, le contenu des clauses, la forme de la datation ainsi que la notification « a tous présens et a venir » ou « a tous ceux qui ces presentes verront » témoignent d'une proximité importante avec la chancellerie royale. L'organisation des parties du discours et le vocabulaire suivent ceux en usage à la cour du roi de France. Néanmoins, quelques dissemblances peuvent être relevées. Les différences majeures se rapportent aux verbes usités et à la langue des actes. En Bourbonnais, la langue vernaculaire domine. Les quelques actes rédigés en latin sont généralement des actes très solennels destinés aux clercs et aux institutions religieuses.

48. *Ibid.*

49. *Ibid.*

50. J.D. Généro, *La chancellerie de Charles Ier de Bourbon, suivi de l'édition de ses actes et de ceux de son épouse, Agnès de Bourgogne (1421-1476)*...

51. O. Mattéoni, « Écriture et pouvoir princier. La chancellerie du duc Louis II de Bourbon (1356-1410) »...

52. Méliissa Barry, Cléo Rager, Élisabeth Schmit, Marie-Émeline Sterlin et Clémence Lescuyer, « « Imitatio regis » ? Pour une diplomatique des actes de Jean de Berry », in : O. Guyotjeannin et O. Mattéoni, *Jean de Berry et l'écrit...*

53. Annexe Initiales fleurdelisées.

AN, P 1363¹, c. 1151¹ (18 décembre 1448) – AN, P 1368², c. 1628 (19 février 1447).

54. J.D. Généro, « Sceaux et armes de Charles Ier de Bourbon et d'Agnès de Bourgogne », *Revue française d'héraldique et de sigillographie*, 92 (2022), p. 31-51, URL : <https://hal.science/hal-03908270> (visité le 03/04/2023).

La structuration de la chancellerie sous Louis II et sa prolongation sous Charles I^{er} marque un approfondissement de l'*imitatio regis* qui traduit l'attachement familial des Bourbons au lignage des rois de France et une volonté d'asseoir leur pouvoir par l'imitation du modèle royal. Le rôle des individus, notamment des officiers de la chancellerie, est également à souligner dans la diffusion des pratiques d'écriture pour la période étudiée.

Le corpus rassemble des actes princiers aux typologies variées, dispersés au gré des différentes institutions de conservation et s'échelonnant sur une période de plus d'un siècle. La production recensée qui a permis de former le corpus étudié est le fruit d'un travail de longue haleine de regroupement puis d'édition des actes.

Chapitre 2

Des éditions numériques à partir d'un logiciel de traitement de texte

Les premiers projets d'édition sont impulsés par des érudits au XIX^e siècle, avec comme objectif principal de rassembler des corpus et d'en fournir les éléments d'analyse¹. Les actes des princes laïcs sont quelque peu délaissés par ces entreprises au profit des actes pontificaux et royaux. Les perspectives qu'offre aujourd'hui le numérique permettent l'étude de corpus importants et dispersés sur un temps indéterminé, autant de possibilités que n'offre pas la publication d'un ouvrage, comprenant ainsi la reproduction, la publication et la diffusion de documents d'archives. C'est dans ce contexte que l'on voit se développer des entreprises d'édition consacrées aux actes princiers et dans lequel s'inscrivent les éditions des actes émis par les Bourbons.

1. « L'évolution des pratiques éditoriales », in : O. Guyotjeannin, *Conseils pour l'édition des textes médiévaux, Fascicule II : actes et documents d'archives*, t. 2, 3 t., CTHS - École Nationale des Chartes, Paris, 2009 (Orientations et méthodes, 14).

2.1 Trois projets d'édition

Les quatre corpus rassemblant les actes de Louis II, d'Anne Dauphine, de Charles I^{er} et d'Agnès de Bourgogne sont stockés, au commencement du stage sur lequel ce mémoire s'appuie, dans des fichiers de traitement de texte. Ces fichiers ont été constitués à l'issue du travail de recensement des actes dans les différentes institutions citées précédemment. Il s'agit d'éditions critiques : des éditions spécifiques qui collationnent les variantes d'un texte tout en pourvoyant ce dernier d'un apparat critique².

Ces éditions ont été menées par plusieurs chercheurs, sur des temporalités différentes et avec des objectifs différents : recherches personnelles, mémoire académique et travail préparatoire d'une communication scientifique. Les actes de Louis II ont été édités par Olivier Mattéoni, enseignant chercheur à l'Université Paris 1 Panthéon-Sorbonne, dans le cadre de ses recherches sur les Bourbons³. Il a notamment soutenu une thèse de doctorat sur les officiers bourbonnais en 1994 et publié de nombreux travaux sur le duché et les ducs de Bourbon⁴. Les actes de Charles I^{er} et d'Agnès de Bourgogne ont été édités par Jean-Damien Généro pour son mémoire de Master recherche soutenu à l'Université Panthéon-Sorbonne en 2018, sous la direction d'O. Mattéoni⁵. Les actes d'Anne Dauphine ont enfin été édités conjointement par O. Mattéoni et J.-D. Généro en vue de présenter une communication sur la diplomatique de la duchesse au colloque organisé en 2017 pour les cinq cents ans de sa mort⁶. O. Mattéoni a par la suite retravaillé les fichiers de Louis II et d'Anne Dauphine en systématisant l'ajout des notes critiques qui concernent les individus et les lieux. Il a aussi rédigé de succinctes biographies des individus qui apparaissent régulièrement dans les actes (les grands officiers, les grands prélats) et les a ajoutées comme note critique à chaque fois que ces individus apparaissent. Pour les lieux, il a précisé leurs localisations (toponyme actuel, canton, département). Dans la mesure où il n'est pas possible de prévoir comment les utilisateurs vont aborder le corpus sur le site internet ni quel sera soit leur point d'entrée, l'existence de notes concernant les individus et les lieux demeure essentielle.

2. L'apparat critique désigne l'ensemble des annotations d'érudition qu'un éditeur ajoute à un texte original. « Le destin de l'appareil critique dans l'édition scientifique numérique », in : Daniel Apollon, Philippe Régnier et Claire Bélisle, *L'édition critique à l'ère du numérique*, L'Harmattan, Paris, 2017 (Socio-économie de la chaîne du livre, 9).

3. O. Mattéoni, « Écriture et pouvoir princier. La chancellerie du duc Louis II de Bourbon (1356-1410) »...

4. Id., *Servir le prince...*

5. J.D. Généro, *La chancellerie de Charles Ier de Bourbon, suivi de l'édition de ses actes et de ceux de son épouse, Agnès de Bourgogne (1421-1476)*...

6. O. Mattéoni et J.D. Généro, « Les actes d'Anne Dauphine : étude diplomatique », dans *Anne Dauphine, dernière comtesse de Forez. Actes de la journée d'étude de Montbrison, 29 septembre 2017*, La Diana - Société historique et archéologique du Forez, Montbrison, 2019, p. 35-52, URL : <https://shs.hal.science/halshs-02972893> (visité le 31/03/2023).

Les éditions des actes de Louis II et d'Anne Dauphine n'étaient pas destinées à être immédiatement éditées sous forme papier ou électronique, tandis que celles des actes de Charles I^{er} et d'Agnès de Bourgogne ont été imprimées dans le cadre d'un travail universitaire. Ces finalités différentes supposent des choix de mise en forme et de mise en page différents, qui ont une influence sur les traitements postérieurs. En effet, concevoir un fichier de traitement de texte dans le but de l'imprimer implique de privilégier les aspects visuels de l'édition et sa mise en page.

Les actes ont été édités grâce à un logiciel de traitement de texte. Le traitement de texte qualifie les logiciels qui permettent à un utilisateur de rédiger du contenu informatisé. Ces logiciels sont multiples et utilisent différents formats (ODT, docx) plus ou moins compatibles selon les logiciels et les systèmes d'exploitation. Un format de fichier spécifie la manière dont les données sont stockées pour une application particulière⁷. La lecture des différents formats et de leur mise en forme diffère selon les applications. J.-D. Généro a édité les actes de Charles I^{er} et d'Agnès de Bourgogne sur le logiciel de traitement de texte *Pages* de MacOS avant de les convertir en ODT. Les actes d'Anne Dauphine ont également été édités sur un fichier ODT tandis que les actes de Louis II ont été édités sur un fichier docx. Le fichier docx est un document Microsoft Word au format Open XML et le fichier ODT est un document LibreOffice au format OpenDocument. Même si OpenDocument Format (ODF) prend en charge les fonctionnalités OpenOffice et Open XML, et que Microsoft Office prend en charge ODF, toutes les fonctionnalités ne sont pas prises en charge, ce qui peut occasionner des changements en termes de modification d'un document et des pertes de contenu.

Les travaux de recherche, même à leurs prémisses, sont rarement neutres, puisque les données ne le sont pas non plus, et qu'ils consistent en la reprise de données qui ont souvent déjà été traitées. Les fichiers de traitement de texte peuvent être qualifiés de *legacy data*, puisqu'ils sont le résultat de traitements antérieurs opérés par les premiers chercheurs à travailler sur le corpus⁸. Les ingénieurs du projet « Actes princiers » ont conscience qu'ils vont devoir composer avec cela tout au long de leurs traitements et intégrer ce paramètre à leur réflexion.

Ces différences de finalités et de formats posent des problèmes d'harmonisation des données lorsqu'il s'agit de regrouper les actes des ducs de Bourbon pour en proposer une édition cumulative. Malgré ces disparités dans la mise en page, les éditions comportent des similitudes, notamment dans leurs présentations que l'on peut qualifier de diplomatiques.

7. Microsoft, « En apprendre plus sur les formats de fichiers », <https://support.microsoft.com/fr-fr/office/en-apprendre-plus-sur-les-formats-de-fichiers-56dc3b55-7681-402e-a727-c59fa0884b30>.

8. Sur le concept de *legacy data*, voir : Virgile Reignier, *Vers l'indexation automatique du Trésor des chartes. Constitution, alignement et utilisation d'un référentiel d'entités nommées au sein du projet Himanis*. Mémoire de Master TNAH, Paris, École Nationale des Chartes, Université Paris Sciences et Lettres, 2022.

2.2 Des éditions diplomatiques

Les différentes éditions suivent les normes en vigueur développées dans les *Conseils pour l'édition des textes médiévaux* de l'École Nationale des Chartes⁹, inspirés des règles établies en 1984¹⁰. Les actes médiévaux étant généralement écrits d'un seul bloc, il est d'usage de restituer cette mise en forme en éditant le contenu textuel de l'acte en un seul paragraphe. Le texte édité est accompagné d'éléments de présentation. Comme l'explique Olivier Guyotjeannin, « une fois le texte établi et son apparat mis en forme, il reste à le munir d'un "habillage", composé des éléments de présentation qui guideront la lecture. L'éditeur cherche à concilier le respect d'éléments qui peuvent être signifiants et les commodités fournies au lecteur contemporain »¹¹.

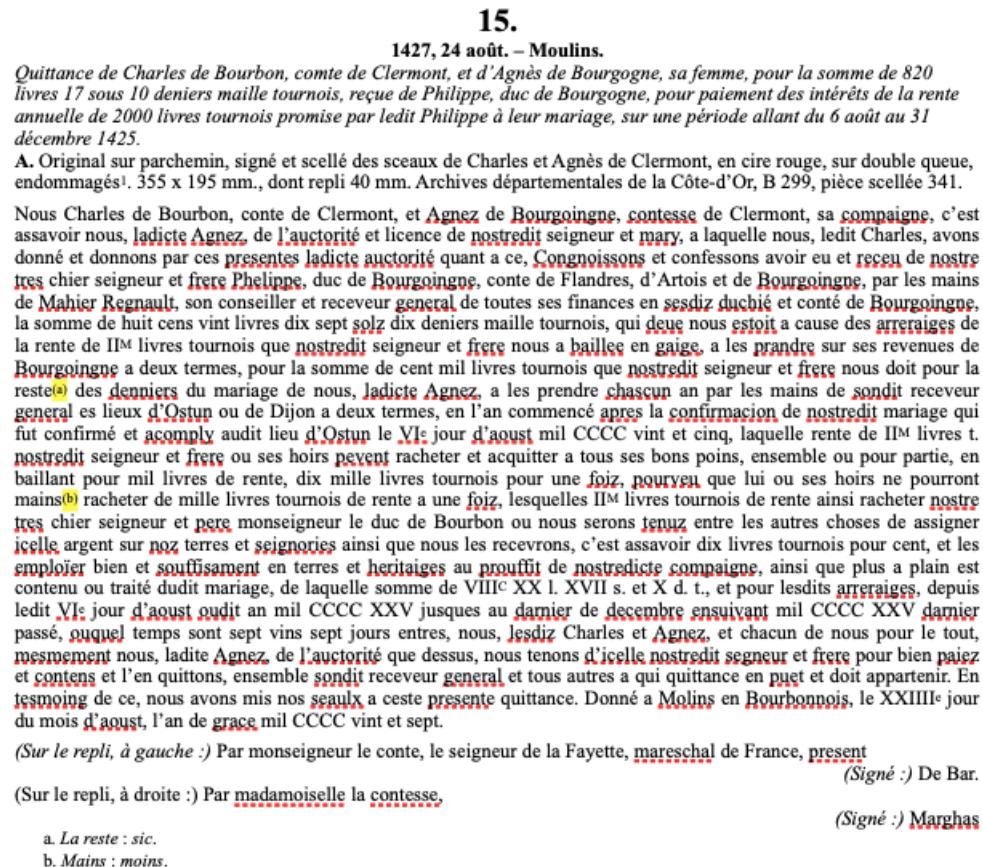


FIGURE 2.1 – Édition diplomatique de l'acte daté du 27 août 1427.

9. O. Guyotjeannin et Françoise Vielliard, *Conseils pour l'édition des textes médiévaux, Fascicule I : conseils généraux*, t. 1, 3 t., CTHS - École Nationale des Chartes, Paris, 2014 (Orientations et méthodes, 27). - O. Guyotjeannin, *Conseils pour l'édition des textes médiévaux, Fascicule II : actes et documents d'archives...*

10. C.I.D., « Normes Internationales Pour l'édition Des Textes Médiévaux », dans *Travaux Préliminaires de La Commission Internationale de Diplomatique*, Saragosse, 1984 (Folia Caesaraugustana, 1), p. 21-93.

11. « La présentation de l'édition », in : O. Guyotjeannin, *Conseils pour l'édition des textes médiévaux, Fascicule II : actes et documents d'archives...*

Comme nous pouvons le voir dans l'exemple ci-avant, chaque acte est précédé d'un numéro d'ordre permettant de le situer dans le corpus. Viennent ensuite les dates de temps et de lieu.

La date de temps est restituée sous une forme contemporaine et débute par l'inscription du millésime, suivi du quartier et du mois. Dans le cas où le millésime convertit par l'éditeur viendrait à différer de celui indiqué dans l'acte, en raison du style en vigueur chez les contemporains, ce dernier est suivi de la mention (n.st.). En effet, la chancellerie des ducs de Bourbon suivant le style de Pâques, il est nécessaire de convertir les dates ayant lieu avant cette fête en style du premier janvier. Si la date a été reconstituée, les adverbes de temps « avant » et « après » sont précisés en amont.

La date de lieu est séparée de la date de temps par un point, un blanc et un tiret et les toponymes sont restitués selon leur forme actuelle s'ils ont pu être identifiés.

Une analyse (regeste) précède le texte édité afin d'en proposer un résumé et d'en préciser la typologie. Elle est indiquée en italique et se doit être la plus efficace possible dans la compréhension de l'acte qu'elle va offrir au lecteur.

L'analyse est suivie d'un tableau de la tradition qui indique l'état de la tradition et mentionne les différents témoins ayant permis l'établissement du texte. Ce tableau commence par une description de l'original, dans le cas où il est existant, qui comporte les mentions de son support (parchemin, papier), de sa forme (rouleau, cahier ...), de ses dimensions (en millimètres), du mode de scellage, de son état de conservation ainsi que l'institution où il est conservé. Cette mention est suivie de celle des autres témoins, généralement des copies pour lesquelles on précise la nature juridique (copie d'après original, vidimus ...), la date, l'auteur, le recueil dont elle est tirée ou la source. Les éditions sont également mentionnées, notamment dans le cas des *deperdita*.

Le corps du texte suit ces éléments d'identification et de précision. Il est annoté via un appareil critique fournissant des renseignements liés à l'établissement du texte (choix d'édition : variantes, remarques ...) et une annotation historique apportant des précisions utiles à la compréhension du texte¹². Lorsque l'on édite des originaux, la localisation des signatures et des mentions hors teneurs dans l'espace graphique du parchemin est précisée, généralement en dessous du texte. Le nom de signataires est signalé en petites capitales et précédées de la mention « (Signé :) ». Les mentions hors teneurs sont également précédées de mentions « (Sur le repli :) » ou « (Sous le repli :) », selon les cas. Dans le cas où elles sont nombreuses, des précisions sur leur disposition peuvent être indiquées : « (À gauche :) » ...

12. « La mise au point du texte », in *Ibid.*

En dépit des recommandations canoniques pour la production d'éditions diplomatiques, les pratiques d'édition des deux chercheurs sont différentes. C'est le cas pour les notes de bas de page, qui ont été placées à la suite de l'acte auquel elles faisaient référence pour les éditions des actes de Louis II, et insérées à partir d'une fonction disponible sur ODT pour les autres éditions. Le corpus de Louis II est également le seul dont les actes ne comportent pas de numérotation. Même si cette dernière peut être questionnée lors de la découverte de nouveaux actes, elle participe à la structuration du corpus.

2.3 Le traitement de texte : un bon compromis ?

L'édition à partir d'un logiciel de traitement de texte n'est pas sans avantages. Contrairement à une édition papier imprimée, elle n'est pas figée. L'intégration progressive des données au fur et à mesure de l'avancée des travaux de recherche est possible, là où une édition imprimée ne permet pas ces mises à jour, ou alors via des publications additionnelles¹³. Ces fonctionnalités sont particulièrement utiles dans le cadre d'éditions diplomatiques, dans la mesure où elles permettent, lors de la découverte de nouveau états du texte, de les ajouter au corpus. Néanmoins, le traitement de texte ne permet pas le remplissage automatique des données contrairement à d'autres technologies comme Python. Même s'il permet l'intégration de nouveaux actes, celle-ci doit se faire manuellement. Cela nécessite de réfléchir en amont à la structuration la plus adaptée possible de son édition, pour éviter d'avoir des problèmes lors de l'ajout de nouvelles données.

Prenons le cas du corpus de Charles I^{er} dans lequel les actes sont numérotés un à un. Quid de cette numérotation lors de l'intégration d'actes inédits au corpus ? Lors de notre visite aux Archives nationales en avril 2023, nous avons photographié cinq actes inédits repérés par O. Mattéoni dans la série P¹⁴. Nous avons décidé de les intégrer à un CSV (comma-separated values)¹⁵ répertoriant les informations des actes. Néanmoins, cela a ensuite posé des problèmes de correspondance des données, car la numérotation ne correspondait plus une fois ces actes intégrés au corpus. La question de leur intégration immédiate au corpus s'est donc posée et il a été décidé de les intégrer postérieurement dans le projet. C'est pourquoi réfléchir à un schéma, à des pratiques d'édition communes, facilite le travail conjoint de plusieurs chercheurs sur un même corpus.

L'édition numérique ou électronique désigne une édition mise en forme et diffusée dans un environnement numérique. L'édition des actes sur un logiciel de traitement de texte peut en ce sens être vue comme une édition numérique dans la mesure où elle peut être lue sur un écran. De plus, à l'instar de l'édition numérique, elle laisse la possibilité aux ajouts, aux corrections ainsi qu'à l'élargissement du corpus. Ces mêmes ajouts et corrections peuvent être le fait différents contributeurs. Le travail peut également être fractionné en fonction des spécialités de chacun (transcription, mise en forme ...) ¹⁶. Néanmoins, l'ambition principale d'une édition numérique n'est pas sa lecture seule, mais son

13. Canteaut, Olivier, Moufflet, Jean-François, « Les éditions d'actes princiers (XII^e - XV^e siècle) : bilan à l'ère du numérique », in : O. Guyotjeannin et O. Mattéoni, *Jean de Berry et l'écrit...*

14. AN, P 452², c. 190 (24 octobre 1443) - AN, P 452², c. 307 (24 octobre 1443) - AN, P 452², c. 313 (24 octobre 1443) - AN, P 452², c. 229 (25 octobre 1443) - AN, P 452², c. 169 (22 octobre 1454).

15. Type de fichier spécial dont les valeurs sont séparées par des virgules et qu'il est possible de créer ou de modifier dans Excel, définition in : Microsoft, « Créez ou modifiez des fichiers .csv à importer dans Outlook », <https://support.microsoft.com/fr-fr/office/cr>.

16. D. Apollon, P. Régnier et C. Bélisle, *L'édition critique à l'ère du numérique...*

accessibilité. Celle-ci est permise par la mise en ligne des données, qui offre des fonctionnalités supplémentaires contribuant à renforcer la souplesse et la qualité de l'édition.

La mise en ligne des données, soit leur disponibilité et leur accessibilité sur un réseau, permet un certain nombre de possibilités qui ne sont pas à négliger dans le cadre d'une édition de textes médiévaux. Elle propose des interfaces de consultation et de visualisation qui mettent considérablement en valeur le projet. La mise en page est améliorée et offre une meilleure lisibilité au lecteur. C'est notamment le cas avec l'apparat critique, qui peut vite entraver la mise en page dans le cas d'une édition papier ou à partir d'un logiciel de traitement de texte¹⁷. L'hébergement en ligne permet des notes interactives et un certain nombre de renvois aux références sans avoir recours à des notes de bas de page qui alourdissent l'édition et qui nécessitent parfois de faire des choix sur les informations communiquées au lecteur. La présentation est dans l'ensemble plus complexe et offre un affichage multiple avec plusieurs vues du texte : affichage des autres états et des variantes (possibilité d'afficher un fac-similé via un lien). Les traductions et les transcriptions peuvent être affichées de manière interactive¹⁸. La mise en ligne permet également une publication progressive des recherches, et offre une disponibilité de ce qui a déjà été traité pour la communauté scientifique, mais aussi pour un public plus large (généalogistes...). Une édition dynamique interactive rend le travail d'édition pilotable par l'utilisateur, permettant ainsi plus de possibilités et une réflexion sur le travail d'édition¹⁹. Le lecteur a aussi la possibilité de filtrer les informations et les métadonnées essentielles à ses travaux de recherche.

Malgré des choix d'édition proches, le logiciel de traitement de texte permet de trop nombreuses variations et ambiguïtés selon le format utilisé. Ces variations posent des problèmes d'harmonisation qui ne seront pas sans conséquence lorsque se posera la question de l'accessibilité des données. En effet, l'accès aux données à la communauté scientifique est l'une des fins, pour ne pas dire la fin, du travail d'édition. Cette accessibilité est aujourd'hui largement facilitée et augmentée par la mise en ligne des travaux. L'ODT offre un compromis relativement simple à prendre en main, mais dont les possibilités d'interopérabilité²⁰ et de structuration des données nécessaires aux travaux d'édition de corpus conséquents sont vite limitées. L'harmonisation des éditions est donc primordiale et passe par le choix d'outils (automatiques ou manuels) plus adaptés et intéressants pour la recherche, offrant des possibilités de manipulation des données plus élevées.

17. « Le destin de l'appareil critique dans l'édition numérique scientifique », in : *Ibid.*

18. C'est par exemple le cas des dossiers documentaires du THELEME. *Techniques Pour l'Historien En Ligne (Theleme)*, École Nationale des Chartes, URL : <http://theleme.ENC.sorbonne.fr/> (visité le 10/04/2023).

19. Id., *L'édition critique à l'ère du numérique...*

20. L'interopérabilité désigne la capacité de systèmes, unités, matériels à opérer ensemble.

Deuxième partie

Du traitement de texte à l'encodage

Chapitre 3

Préparer le traitement de texte à l'encodage en XML

Les actes du corpus étant édités aux formats ODT et docx, il s'agit dans un premier temps de les nettoyer, c'est-à-dire de les préparer à un encodage, afin de permettre aux chercheurs de manipuler le corpus. XML (Extensible Markup Language) est un langage de balisage qui révèle la structure et la sémantique des données, conçu pour la description des documents textuels, et en ce sens adapté pour l'édition critique. Le recours au langage de balisage XML permet une structuration des éléments textuels à l'aide de balises, ce que le traitement de texte ne permet pas. Depuis 1998, XML est un langage libre et documenté qui facilite la lisibilité des données, leur échange et leur migration vers d'autres plateformes, logiciels et formats. C'est aujourd'hui un standard du W3C¹ qui comporte différents formats comme la TEI (Text Encoding Initiative) pour l'encodage de documents patrimoniaux.

1. World Wide Web Consortium : communauté internationale qui développe des normes ouvertes pour assurer la croissance à long terme du Web.

3.1 Le recours à un langage de balisage

Le chercheur en philologie Claus Huitfeld définit le balisage comme « l'usage de chaînes de caractères réservés, insérées dans des séquences de caractères dans des fichiers de documents numériques, afin de dénoter ou de signaler des caractéristiques d'un document qui ne peuvent pas être transmises par des caractères représentant directement leur contenu verbal »². Le terme balisage renvoie à des marqueurs spéciaux permettant de signaler certaines propriétés du document numérique (ponctuation, notes...) via un ensemble de balises³. L'édition scientifique actuelle fait appel à des méthodes de publication informatisées qui induisent du balisage. Les systèmes de traitement de texte ou de publication sur le web les plus utilisées actuellement disposent d'un système de balisage en « arrière-plan » que l'usager n'a pas à gérer : il intervient directement sur le document via une mise en page visuelle (WYSIWYG : « what you see is what you get »⁴). Ces outils sont utiles pour l'affichage à l'écran ou l'impression papier, mais dès lors que l'on souhaite proposer des outils de recherche performants, de comptage, ou encore d'analyse, il devient nécessaire d'effectuer un balisage généralisé du texte. Différents langages de balisage existent et comprennent une syntaxe, un vocabulaire contrôlé, des méthodes, des pratiques et des outils associés⁵.

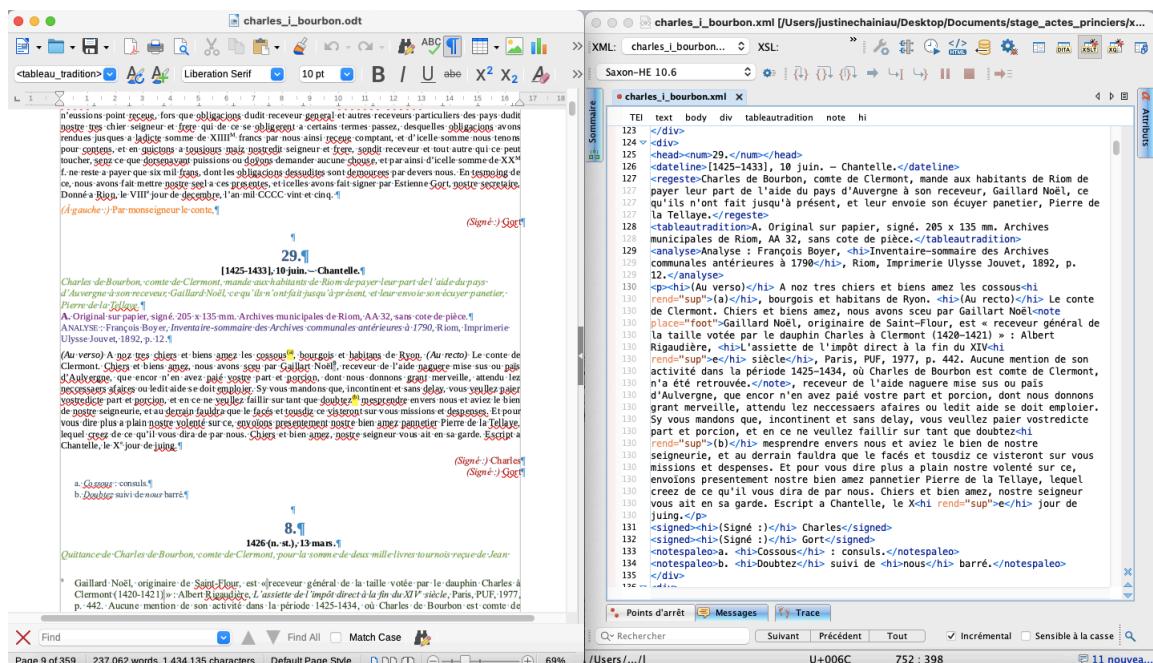


FIGURE 3.1 – Traitement de texte et encodage de l'acte n°29 daté du 10 juin 1425-1433.

2. « Systèmes de balisage de textes et édition critiques », in : *Ibid.*
3. *Ibid.*
4. « Ce que vous voyez est ce que vous obtenez » : l'utilisateur voit directement le résultat de sa saisie.
5. « Systèmes de balisage de textes et édition critiques », in : *Ibid.*

Le langage XML est le plus approprié pour l'édition numérique. Contrairement au langage HTML (HyperText Markup Language) dont les balises définissent l'apparence des données (titres, paragraphes), les balises du langage XML définissent la structure et la signification sémantique des données, ce qui permet leur réutilisation. La particularité du XML est que son affichage nécessite une interprétation, soit une transformation du document. Le document XML est alors transformé via des feuilles de style XSL (Extensible stylesheet language)⁶ en document HTML pour une disponibilité sur le web ou en PDF pour une impression de qualité. Cette possibilité de présentation dans une variété de formats sans qu'il soit nécessaire de modifier le code assure l'intégrité des données. En effet, la conversion d'un document produit sous traitement de texte peut être difficile, manquer de fiabilité. De plus, les formats propriétaires ne sont pas documentés publiquement (usage commercial) et comportent un risque d'obsolescence qu'il est important de prendre en compte lorsque l'on souhaite produire un travail pérenne. Ainsi, l'un des principaux avantages du balisage généralisé est son interopérabilité, qui garantit une utilisabilité future. Dans le cadre de la recherche en sciences humaines et sociales, de l'édition en particulier, il est primordial de faciliter l'échange et la réutilisation des documents numériques et d'assurer leur conservation sous une forme qui les rend accessibles à la recherche. Un autre avantage du balisage généralisé est son adaptabilité à différents projets d'édition. Pour cela, différents langages XML ont été développés.

En 1994 sont publiées les Recommandations méthodologiques pour le balisage et l'échange de textes électroniques de la TEI⁷, soit un ensemble de balises qui peuvent être combinées et adaptées à des besoins spécifiques⁸. La TEI est un langage descriptif de balisage adapté à la description fine des différents aspects d'une source, permettant ainsi au chercheur de caractériser les informations utiles à la production d'une édition. Il s'agit d'un ensemble de recommandations pour les chercheurs sur la création de ressources textuelles informatisées et lisibles par des machines, adaptées aux besoins de la recherche et extensibles, puisque ces besoins changent et évoluent⁹. Les Guidelines de la TEI définissent plusieurs centaines de concepts différents permettant une description scientifique et sémantique d'un texte pour des projets d'édition variés (théâtre, poésie, documents d'archives...).

6. Extensible stylesheet language : langage des feuilles de style associées à XML.

7. Text Encoding Initiative : consortium qui élabore et maintient collectivement une norme pour la représentation des textes sous forme numérique : <https://tei-c.org/>.

8. « Systèmes de balisage de textes et édition critiques », in : *Ibid.*

9. Lou Burnard, *Qu'est-ce que la Text Encoding Initiative ?*, trad. par Marjorie Burghart, OpenEdition Press, Marseille, 2015 (Encyclopédie numérique), DOI : 10.4000/books.oep.1237.

Les technologies de balisage adaptées à l'édition numérique rendent possible la séparation entre la description et la présentation. Cette différenciation entre la transcription des données textuelles et leur présentation offre à l'éditeur la possibilité de développer de manière distincte l'enregistrement du document et l'affichage des données et des métadonnées le caractérisant (notes, index...) ¹⁰. La technologie remet donc en question les pratiques traditionnelles de l'édition. Les standards de balisage du texte assurent l'interopérabilité des données, contribuent à améliorer la qualité de l'édition et de ses éventuelles révisions, et sont en ce sens novateurs. Toutefois, les langages de balisage nécessitent de se familiariser avec leur syntaxe, ce qui peut parfois motiver le choix d'une édition via un logiciel de traitement de texte, comme dans notre cas.

Deux possibilités d'encodage s'offrent désormais à nous pour transformer le corpus : manuelle ou automatique. L'encodage manuel étant chronophage, la conversion automatique est plus appropriée. Toutefois, la conversion des ODT en XML sans opérations préparatoires risque de générer une « dette technique » (maintenances qui auront une incidence importante sur l'ensemble du projet). En effet, parmi les opérations de traitement des XML obtenus, certaines auraient pu être évitées si la mise en forme du document ODT d'origine avait été améliorée. Pour pallier cette « dette », il est nécessaire d'harmoniser au maximum les documents en amont.

10. « Éditions critiques et séparation de la description et de la présentation, in : D. Apollon, P. Régnier et C. Bélisle, *L'édition critique à l'ère du numérique...*

3.2 Utilisation de regex pour une première structuration du texte

Le recours à des expressions régulières (regex)¹¹ s'est révélé opportun pour automatiser le nettoyage des documents¹². Les regex sont des chaînes de caractères utilisées pour décrire, dans une syntaxe précise, un ensemble de chaînes de caractères possibles. Leur emploi peut être exécuté à partir de la fonction rechercher/remplacer du traitement de texte. Elles peuvent également être utilisées pour rechercher des caractères non spécifiés ou même invisibles (espaces, sauts de lignes...). Le corpus des actes émis par Agnès de Bourgogne étant moins volumineux, il a tenu lieu d'expérimentation afin de tester la pertinence des regex et leur prochaine application au corpus plus important que représentent ceux de Louis II et de Charles I^{er}.

L'utilisation de regex simples a permis dans un premier temps de parfaire la mise en page du document en retirant les doubles espaces, les sauts de lignes ou encore les pages inutiles¹³. Cette étape a été exécutée avec facilité pour le premier corpus relatif aux actes d'Agnès de Bourgogne. Néanmoins, le traitement d'un fichier plus volumineux, comme celui de Charles I^{er}, nous confronte à des problèmes de mise en page. Nous remarquons la présence d'espaces similaires à des tabulations devant les mentions hors teneurs. L'affichage de « la mise en page » sur Word a permis de repérer ces anomalies, à savoir l'apparition de mentions « sur le repli » par la suite masquées en blanc afin d'aligner les mentions hors teneur en vue de l'impression du mémoire de Master. Le recours à une regex plus ciblée a permis leur suppression et a réglé les problèmes d'indentation. Dans un second temps, les regex ont contribué à améliorer la lisibilité du document. Les notes paléographiques ont bénéficié d'un retour à la ligne et les astérisques faisant office de séparateur des actes ont été supprimées. Seule la numérotation des actes a été retenue, car contrairement aux astérisques, elle est présente dès le premier acte, ce qui évitera un encodage manuel de ce dernier, permettant ainsi d'automatiser totalement le processus.

11. *List of Regular Expressions*, LibreOffice, URL : <https://help.libreoffice.org/7.5/en-US/text/shared/01/02100001.html> (visité le 17/04/2023).

12. *Using Regular Expressions in Text Searches*, LibreOffice, URL : https://help.libreoffice.org/7.5/en-US/text/swriter/guide/search_regex.html (visité le 17/04/2023).

13. Annexe Liste des regex utilisées.

L’automatisation du nettoyage des données via des regex s’est révélée être une méthode puissante pour uniformiser rapidement les éditions. Néanmoins, certaines étapes du nettoyage ont dû être effectuées manuellement, afin d’éviter une perte de données. C’est le cas tout au long du travail avec les fautes orthographiques qui peuvent être repérées progressivement. Les notes de bas de page présentes dans les dates ont été déplacées à la fin de l’analyse afin d’éviter de surcharger l’encodage. Certaines erreurs trop volumineuses n’ont pas encore été résolues et nécessitent un questionnement. C’est le cas de la présence aléatoire de points à la suite des dates. Il y a aussi des problèmes d’harmonisation des mentions hors teneurs et des signataires qui ne sont parfois pas signalés en amont, ce qui ne permet pas de les identifier et d’intervenir dessus via des regex.

Ces deux premières étapes n’ayant pas permis de résoudre tous les problèmes de mise en page et d’uniformisation, les dernières phases du processus auront lieu plus tardivement dans le projet lors de l’utilisation d’autres technologies. La mention « (Signé :) » devant le nom de chaque signataire n’étant pas continue, il sera plus simple de l’ajouter ou de la retirer plus tardivement lors de l’encodage en XML à partir d’un script Python. Le nettoyage des données découle donc des différentes phases du projet et est aussi dépendant des formats (ODT, XML) et des technologies utilisées (Python...), même si des étapes peuvent être pensées afin de préparer les données à une conversion.

3.3 Proposition de stylage des actes en vue d'une conversion

Les logiciels de traitement de texte permettent d'appliquer des styles différents à des parties du texte. Le recours aux regex s'est avéré efficace pour structurer le texte via le stylage des différentes parties des éditions. Un style est un « ensemble de règles de mise en forme, regroupées, nommées, hiérarchisées et enregistrées, afin de pouvoir être facilement réutilisées en bloc par la suite »¹⁴. Ces derniers offrent à l'utilisateur un contrôle de la mise en forme des documents. Le format OpenDocument est une archive ZIP contenant des fichiers XML qui renferment le contenu du document (content.xml), les styles utilisés (styles.xml) et les métadonnées associées au fichier (meta.xml)¹⁵. Ainsi, lors de la transformation des documents ODT vers XML, les différents styles associés aux parties des éditions seront interprétés par l'éditeur de code, ce qui permettra de les distinguer dans des balises structurantes.

Toutes les différentes parties des éditions : datation, analyse, tableau de la tradition, mentions hors teneurs, signataires, notes paléographiques et notes de bas de page ont été stylées via des regex¹⁶. Dans la mesure où la mise en forme de toutes ces parties était relativement normalisée, les expressions régulières ont permis d'identifier les chaînes de caractères ciblées et de les styler. Les styles ont été créés à l'aide de la fenêtre flottante « Styles et formatage » qui permet de visualiser les différents styles, de les utiliser, de les modifier, d'en créer ou d'en supprimer. Ces styles ont ensuite été réutilisés et appliqués aux autres documents ODT pour les actes de chaque prince. Nous avons opté pour un stylage coloré afin d'avoir un meilleur aperçu des différentes parties des éditions visuellement. Comme le montre les différents exemples ci-après, le stylage n'a pas été appliqué aux mêmes parties de l'édition selon l'état des actes. Les originaux sont les actes qui ont demandé le plus de stylage dans la mesure où ils comportent toutes les parties de l'édition.

14. Open Office, « Utiliser Styles et Modèles avec OpenOffice.org ».

15. Microsoft, « Différences entre le format Texte OpenDocument (.ODT) et le format Word (.docx) », <https://support.microsoft.com/fr-fr/office/diff>.

16. Annexe Liste des regex utilisées.

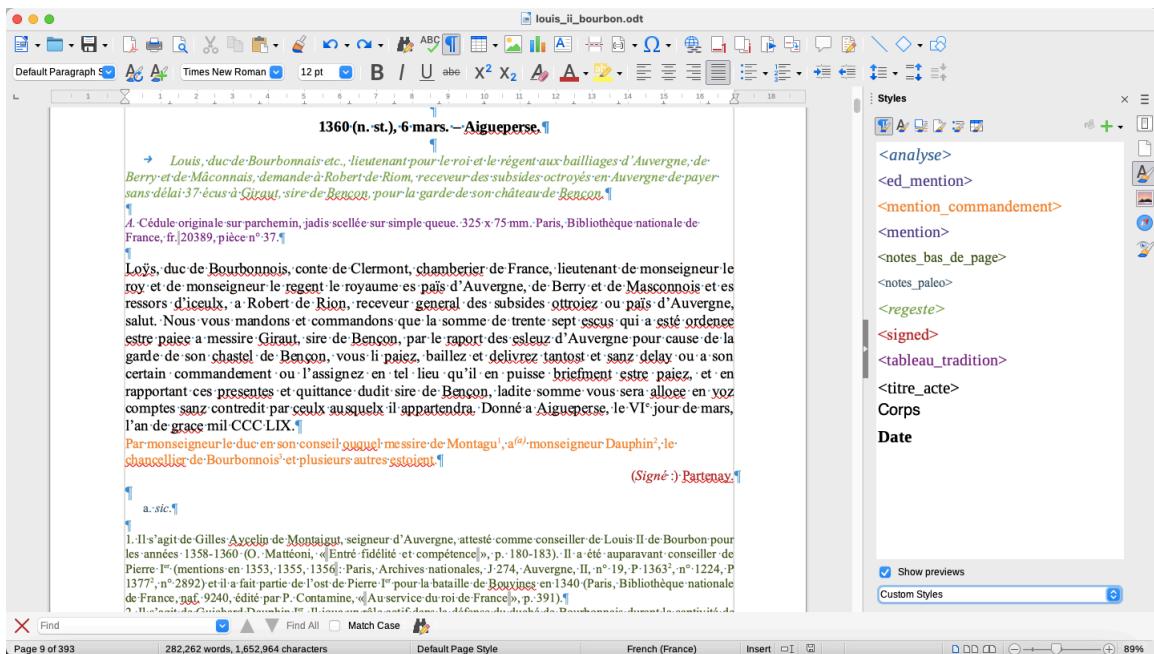


FIGURE 3.2 – Original daté du 6 mars 1360.

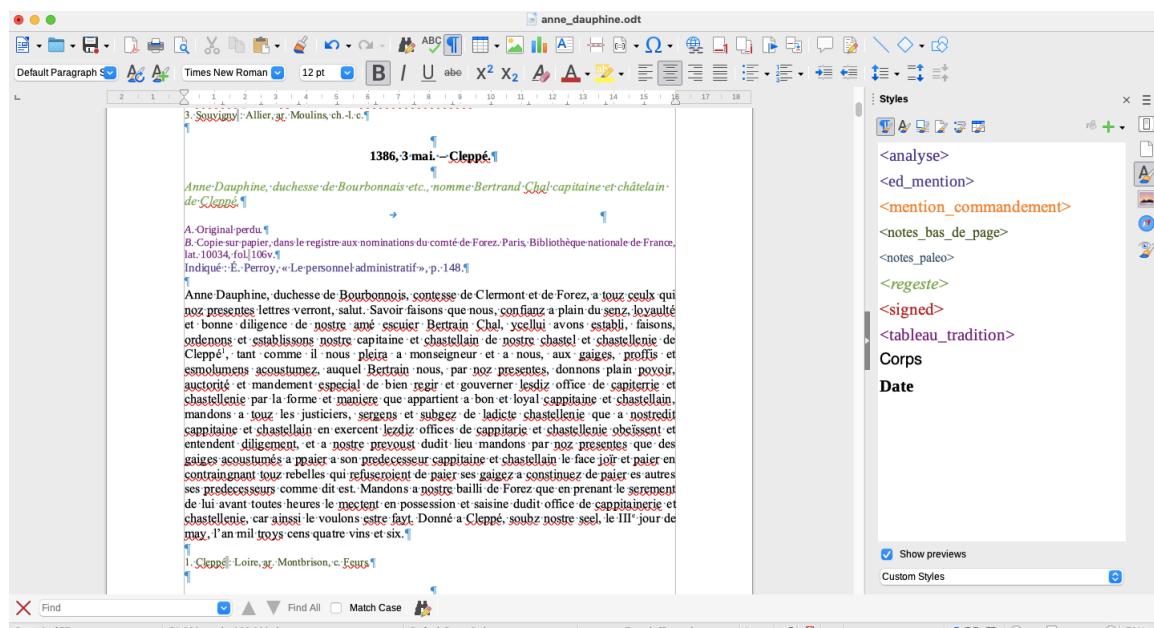


FIGURE 3.3 – Copie datée du 3 mai 1386.

Les copies et vidimus ne comportent pas de stylage pour les mentions hors teneurs et les signataires puisque l'acte copié ou vidimé est présenté d'un bloc.

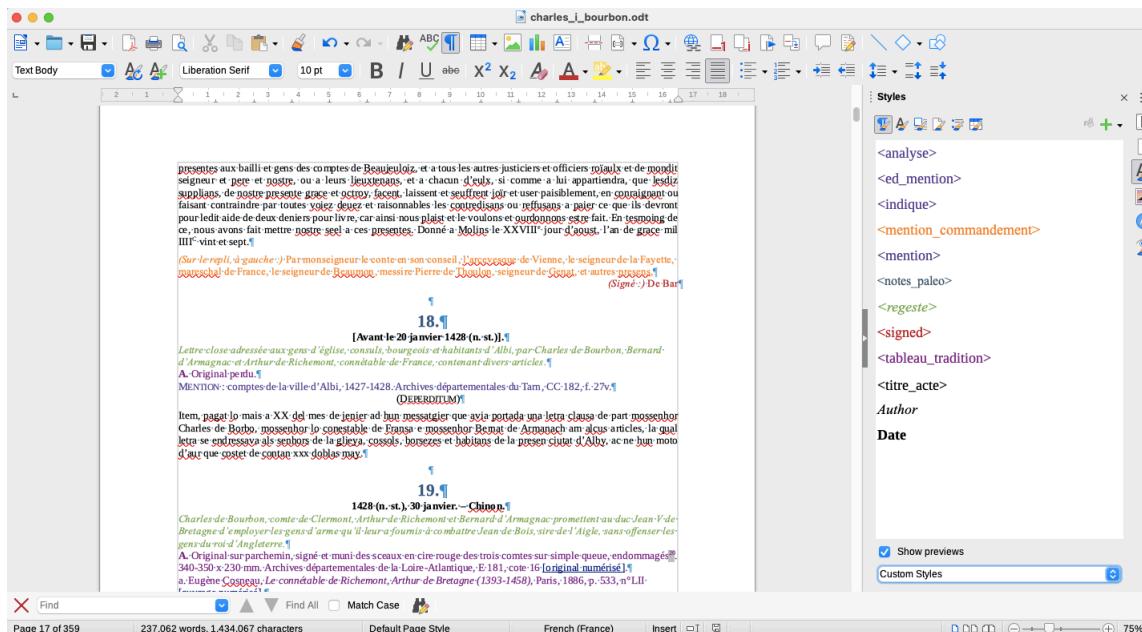


FIGURE 3.4 – **Deperditum** numéroté daté d'avant le 20 janvier 1428.

Les **deperditum** sont les actes pour lesquels le stylage est le plus bref et comporte des variations. À l'instar des autres, les éléments de date et d'analyse sont stylés, le tableau de la tradition également, même s'il est très sommaire : il comporte seulement l'indication que l'original est perdu. La mention du **deperditum** à la suite du tableau de la tradition a cependant nécessité la création d'un style dédié. Le contenu de cette mention est parfois présent, même si ce n'est pas toujours le cas, ce qui peut impliquer des notes.

Une autre variation est attribuable aux différents choix d'édition. En effet, les actes transcrits par Jean-Damien Généro, ceux de Charles I^{er} et d'Agnès de Bourgogne, sont numérotés, ce qui a impliqué la création d'un style titre structurant pour les numéros. Ces étapes ont nécessité de penser des regex plus ciblées et adaptées aux différents corpus. C'est le cas pour les dates dans le corpus de Charles I^{er}. En effet, quelques actes de ce corpus ne sont pas datés de manière précise et obligent à recourir à des adverbes de temps (avant, après et entre). Là où il s'est avéré relativement simples pour les autres corpus, le stylage des analyses des actes de Charles I^{er} a nécessité une harmonisation de ces dernières (normalisation de leur préambule) afin d'automatiser cette étape. Les analyses des premiers actes reflètent la mise en place progressive de la chancellerie de Charles de Bourbon, qui n'est pas encore duc de Bourbon, mais comte de Clermont, et l'évolution de sa titulature. Les corrections se font plus rares après sa nomination, la titulature de "Charles, duc de Bourbonnais et d'Auvergne" étant définitive. Néanmoins, la plupart des corrections apportées ont permis de pallier un problème normatif. Pour beaucoup d'actes, la mention "de Bourbon" après Charles a été oubliée dans les analyses alors qu'elles étaient présentes dans les actes, elle a donc été rajoutée.

L'autre modification concerne les substantifs placés en début d'analyse, que l'on a souvent transformé en verbes, placés à la suite de la titulature du duc. Par exemple, le “consentement donné par Charles de Bourbon” devient “Charles de Bourbon [...] donne son consentement”. Ensuite, le stylage de ces analyses s'est fait par le recours à de nouvelles regex permettant d'attraper les analyses normalisées. Quelques opérations manuelles ont été réalisées pour les analyses uniques, comme pour les testaments ou les ratifications de contrats de mariage. En tout, trente-deux éditions ont fait l'objet d'une modification manuelle. Les mentions hors teneurs situées à la suite du texte ont bénéficié d'une indentation, étape préalable au stylage de toutes les mentions.

Néanmoins, le stylage de certains titres ou sections qui n'ont pas été identifiés par les regex a parfois été réalisé manuellement. Une partie du tableau de la tradition mentionnant les éditions au sein du corpus de Louis II n'a pas pu être stylée automatiquement en raison d'une trop grande proximité stylistique avec les notes paléographiques. En effet, la numérotation des éditions est la même que celle des notes paléographiques (a., b., c., etc). De plus, une partie de ces dernières ont, pour des raisons inconnues, échappé au stylage via des regex, ce qui montre leurs limites d'application à des éditions complexes et non normalisées. Le recours à des logiciels spécialisés dans l'édition critique comme *Classical Text Editor*¹⁷ aurait permis un gain de temps sur toutes ces étapes et la génération d'une édition standard. Enfin, certaines parties comme le contenu des actes n'ont pu être stylées et structurées grâce à des regex en raison de leur complexité (intitulations) ou de leur typographie et pourront être manipulées plus facilement après la conversion.

17. *Classical Text Editor*, en ligne : <https://cte.oeaw.ac.at/>

Chapitre 4

Conversion des fichiers ODT vers XML/TEI

Le format de document texte OpenDocument (ODT) est un format pour les documents textuels modifiables et un des sous-types de la famille ODF pour certaines catégories de contenu, notamment pour les applications de traitement de texte¹. La structure et le texte d'un fichier ODT sont tous représentés en XML : les paragraphes et les sections sont facilement reconnaissables, tout comme les en-têtes et les pieds de page, ainsi que les constructions de niveau supérieur grâce à l'utilisation cohérente des styles (par exemple, pour les titres), des tables des matières et des index générés automatiquement, et des modèles structurés. Tout le formatage est représenté dans le document ODT et stocké dans des fichiers XML afin de faciliter la séparation du texte et de la structure sémantique des caractéristiques de mise en page. Les formats de traitement de texte tels que l'ODT stockent de nombreuses informations associées au processus de création et d'examen des documents, y compris les modifications. Toutes ces informations seront interprétées lors de la conversion. L'enjeu est de trouver un convertisseur adapté, dont le résultat reflètera le mieux possible l'édition des actes. Plusieurs convertisseurs ont été testés afin de mesurer leur convenance, celle-ci étant essentiellement basée sur la reproduction de l'organisation des documents.

1. OpenDocument Text Document Format (ODT), Version 1.2, ISO 26300-1 :2015, en ligne : <https://borntocode.fr/latex-comment-insérer-et-customiser-du-code-source/>.

4.1 Les convertisseurs libres

D'abord, nous avons testé, par curiosité, deux convertisseurs libres. Premièrement, le niveau de structuration après conversion s'est avéré relativement mauvais. *AnyConv* sépare les actes dans des sections, avec un niveau de titre correspondant à leur numéro². Hormis cela, tous les éléments sont ensuite encodés dans des balises `<para>` sans distinction aucune ni précision de la typographie ou des styles.



```

 7 <sect1>
 8   <title>1.</title>
 9   <para>[Avant le 15 janvier 1421 (n. st.)].</para>
10   <para>Lettre close aux consuls d'Albi, signée de Charles de Bourbon, lieutenant en Languedoc, du comte d'Armagnac et de Charles d'Albret, par laquelle ils engagent les habitants à être bons, fidèles et obéissants au roi et au dauphin-régent, et à eux-mêmes comme représentants dudit régent, et non à d'autres.</para>
11   <para>A. Original perdu, jadis signé par les princes et scellés de leurs sceaux (sagelada de lors sagels e senhada de lors propias mas).</para>
12   <para>Mention : comptes de la ville d'Albi, 1420-1421. Archives départementales du Tarn, CC 176, f. 27r.</para>
13   <para>(Deperditum)</para>
14   <para>Item pagat a XV del dich mes de janier a Gilet de Lacort, cavalgador de mossenhor Charles de Borbo, per mossenhor lo dalfi regen en lo pays de lengua d'oc loctenen, lo qual de part del dich mossenhor Charles de Borbo, mossenhor lo compte d'Armanhac et de mossenhor Charles de Lebret, avia portada una letra clausa als senhors cossolls e manans en la presen ciutat d'Albi, sagelada de lors sagels e senhada de lors propias mas, per las quals los mandavo(a) els borgueses esser bos e fizels et hoberdiens al rey nostre señor et a mossenhor lo dalfi regen lo rialme et als comeses per luy e non a d'autres (...).</para>
15   <para>a. Mandavo suivi de que rayé.</para>
16   <para/>
17 </sect1>
18 <sect1>
19   <title>2.</title>
20   <para>[Avant le 18 janvier 1421 (n. st.)].</para>
21   <para>Charles de Bourbon, Jean IV, comte d'Armagnac, et Charles d'Albret, informent les consuls, bourgeois et habitants de Saint-Flour de bonnes nouvelles, et leur demandant de rester fidèles au roi et au régent Charles. Le porteur des lettres est chargé d'en apporter de semblables à toutes les bonnes villes d'Auvergne, du Gévaudan et du Velay.</para>
22   <para>A. Original perdu, jadis signé par les princes et scellés de leurs sceaux (sagelada de lors sagels e senhada de lors propias mas).</para>
23   <para>Mention : compte de la ville de Saint-Flour, année 1420. Archives municipales de Saint-Flour, ch. XI, art. 2., n°32, f. 43v.</para>
24   <para>Indiqué : Albert Rigaudière, Saint-Flour, ville d'Auvergne au bas Moyen Âge, II, Paris, PUF, 1982, p. 620.</para>

```

FIGURE 4.1 – Document XML converti avec *AnyConv*.

Le second convertisseur libre testé, *Online ODT converter*, affiche dans un premier temps l'ensemble des styles utilisés, ce qui entrave la lecture³. Ensuite, la structuration est assez chaotique : tout le texte est encodé dans des balises `<text>`. Malgré la reconnaissance des différents styles ODT et leur affichage comme attributs des balises `<text>`, il n'y a pas de structuration permettant de refléter la séparation des différents actes.

2. AnyConv : Convertisseur de ODT en XML, en ligne : <https://anyconv.com/fr/convertisseur-de-odt-en-xml/>.

3. Online ODT converter, en ligne : <https://onlineconvertfree.com/convert/odt/>.

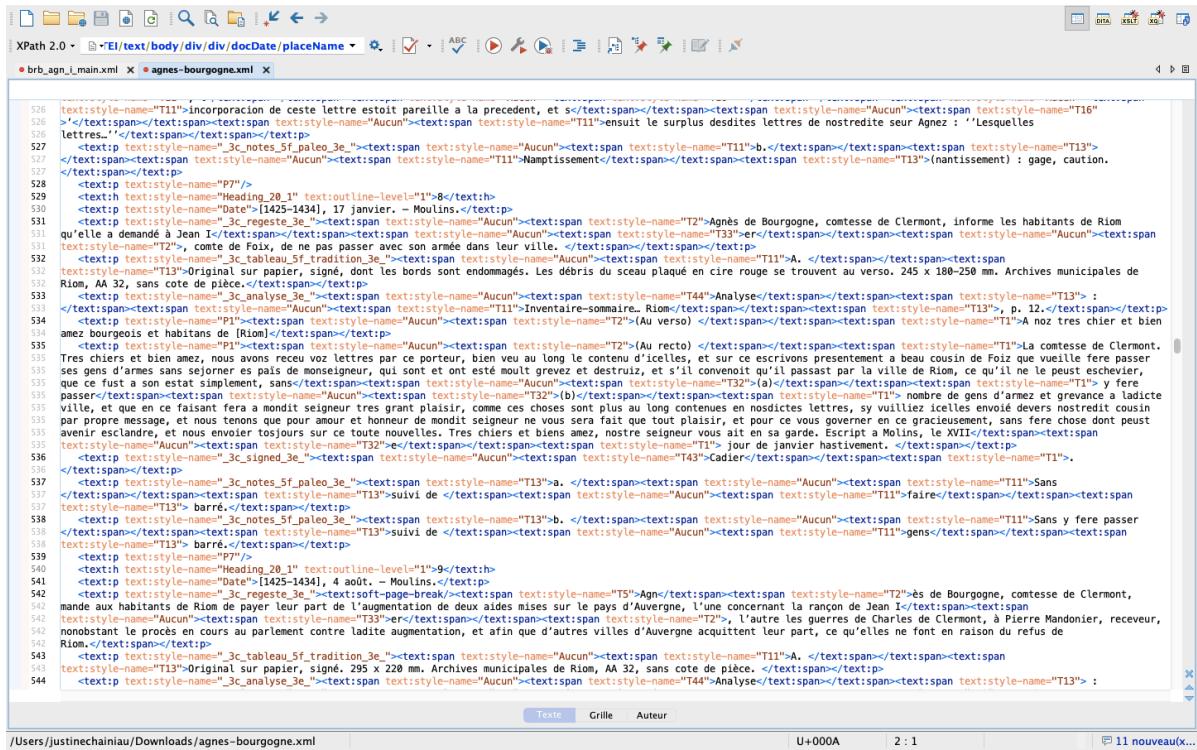


FIGURE 4.2 – Document XML converti avec *Online ODT converter*.

De plus, l'utilisation de ces convertisseurs pose un problème déontologique avec des développeurs non identifiés et des éditeurs peu fiables. Dans le cadre d'un travail scientifique, il est plus pertinent de favoriser les logiciels développés par la communauté scientifique, généralement plus adaptés aux besoins de cette dernière. D'autre part, ces deux premiers logiciels renvoient des fichiers XML, mais pas au standard TEI.

4.2 Essais de conversion avec Pandoc

Ensuite, nous avons testé une conversion avec *Pandoc* : une bibliothèque Haskell permettant de convertir un format de balisage en un autre, et un outil en ligne de commande qui utilise cette bibliothèque⁴. Ce convertisseur a été développé par John MacFarlane, un professeur de philosophie, pour la recherche en Sciences humaines et sociales. Il propose de très nombreuses options de conversion entre différents formats de balisage et de traitement de texte, centrées sur la préservation des éléments structurels d'un document. Nous avons pu tester une conversion de ODT vers XML/TEI, qui s'est montrée plus aboutie.

```

</div>
<div type="level1" xml:id="section-2">
  <head>8</head>
  <p>[1425–1434], 17 janvier. – Moulins.</p>
  <p><hi rendition="simple:italic">Agnès de Bourgogne, comtesse de
Clermont, informe les habitants de Riom qu'elle a demandé à Jean
I</hi><hi rendition="simple:superscript"><hi rendition="simple:italic">er</i>
comte de Foix, de ne pas passer avec son armée dans leur ville.
</hi></p>
<p><hi rendition="simple:italic">A. </hi>Original sur papier, signé,
dont les bords sont endommagés. Les débris du sceau plaqué en cire
rouge se trouvent au verso. 245 x 180–250 mm. Archives municipales de
Riom, AA 32, sans cote de pièce.</p>
<p>Analyse : <hi rendition="simple:italic">Inventaire-sommaire...
Riom</hi>, p. 12.</p>
<p><hi rendition="simple:italic">(Au verso) </hi>A noz tres chier et
bien amez bourgeois et habitans de [Riom]</p>
<p><hi rendition="simple:italic">(Au recto) </hi>La comtesse de
Clermont. Tres chiers et bien amez, nous avons receu voz lettres par
ce porteur, bien veu au long le contenu d'icelles, et sur ce escrivons
presentement a beau cousin de Foiz que vueille fere passer ses gens
d'armes sans sejorner es païs de monseigneur, qui sont et ont esté
moult grevez et destruiz, et s'il convenoit qu'il passast par la ville
de Riom, ce qu'il ne le peust eschevier, que ce fust a son estat
simplement, sans<hi rendition="simple:superscript">(a)</hi> y fere
passer<hi rendition="simple:superscript">(b)</hi> nombre de gens
d'armez et grevance a ladicta ville, et que en ce faisant fera a
mondit seigneur tres grant plaisir, comme ces choses sont plus au long
contenues en nosdictes lettres, sy vuilliez icelles envoié devers
nostredit cousin par propre message, et nous tenons que pour amour et
honneur de mondit seigneur ne vous sera fait que tout plaisir, et pour
ce vous governer en ce gracieusement, sans fere chose dont peust
avenir esclandre, et nous envoier tosjours sur ce toute nouvelles.
Tres chiers et biens amez, nostre seigneur vous ait en sa garde.
Escript a Molins, le XVII<hi rendition="simple:superscript">e</hi>
jour de janvier hastivement. </p>
<p>Cadier. </p>
<p>a. <hi rendition="simple:italic">Sans </hi>suivi de
<hi rendition="simple:italic">faire</hi> barré.</p>
<p>b. <hi rendition="simple:italic">Sans y fere passer </hi>suivi de
<hi rendition="simple:italic">gens</hi> barré.</p>
<p></p>
</div>
```

FIGURE 4.3 – Document XML/TEI converti avec *Pandoc*.

4. Pandoc : a universal document converter, en ligne : <https://pandoc.org/>.

En effet, cela a permis la génération d'un TEI Header, à compléter. Le niveau de structuration s'est révélé plus lisible, avec l'inscription de chaque acte dans une balise `<div>`, avec un niveau de titre pour son numéro (`<head>`). Tous les éléments sont cette fois-ci encodés dans des balises `<p>`. Cette conversion a fait ressortir l'organisation du document en différenciant les niveaux de titre (numéro, datation) du corps du texte et les notes de bas de page des notes paléographiques. Toutefois, cette reconnaissance manque de précision dans la mesure où elle ne distingue pas les différentes rubriques de l'édition diplomatique qui ont été stylées avec ODT, que sont l'analyse, le tableau de la tradition et le contenu de l'acte, malgré les différents styles qui leur avaient été appliqués. La différenciation des paragraphes ne se fait qu'à partir des graisses (gras, italique) et non des styles ODT appliqués aux différents éléments. Enfin, une conversion d'un docx vers XML/TEI nous a renvoyé une structuration similaire, mais avec un niveau de précision plus fin et une meilleure reconnaissance de la typographie appliquée aux différents éléments du document. Mais cette conversion n'a pas encore permis l'affichage des différents styles comme attributs.

Malgré le niveau de description convenable apporté par *Pandoc* dans le cadre d'une conversion de docx vers XML/TEI, nous avons poursuivi nos tests, afin de voir si la hiérarchisation des éléments et la lisibilité pouvaient être améliorées.

4.3 Choix du convertisseur Teinte

Enfin, nous avons testé le convertisseur *Odette*, pour textes bureautiques (ODT) en TEI⁵. *Odette* a été développée par Frédéric Glorieux, ingénieur de recherche en informatique linguistique et documentaire au Labex OBVIL (Observatoire de la vie littéraire) rattaché à la Sorbonne. Le convertisseur produit des documents structurés en XML/TEI via XSLT, de la qualité la plus élevée possible, à partir d'un traitement de texte⁶. La conversion permet par exemple d'éditer des pièces de théâtre classique, avec numérotation des vers, paratexte, index ou glossaire, sans perte d'information. La conversion a permis la génération d'un TEI Header à remplir plus développé. La structuration du document est également mieux respectée : chaque acte est encodé dans une `<div>`. Le numéro d'acte est encodé dans une balise `<head>`, les différents éléments sont encodés dans des balises `<p>` avec comme attributs les différents styles appliqués en amont dans le document ODT, permettant ainsi de refléter une hiérarchie et une structure plus précises. De plus, la date de l'acte est encodée dans une balise `<dateline>`, ce qui montre que le convertisseur a relevé et reconnu cet élément de datation.

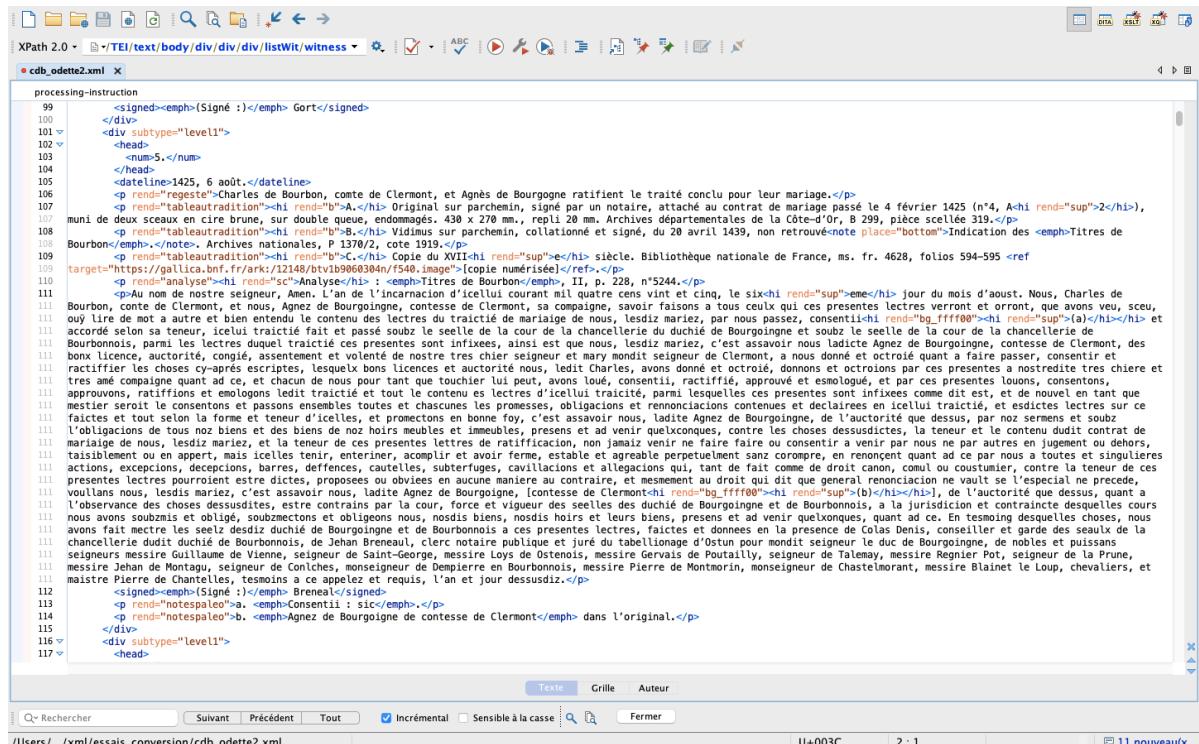


FIGURE 4.4 – Document XML converti avec *Odette*.

5. Frédéric Glorieux, Odette, convertissez vos textes bureautiques (ODT) en TEI, en ligne : <http://obvil.lip6.fr/Odette/>.

6. Frédéric Glorieux, *Le traitement de textes (odt) pour produire des documents structurés (XML/TEI) — Odette*, 8 avr. 2015, URL : <https://resultats.hypotheses.org/267> (visité le 17/04/2023).

Néanmoins, avant de trancher sur ce convertisseur, il s'agit d'être certain de sa pérennité et de sa stabilité pour une utilisation ultérieure à d'autres corpus dans le cadre du projet Actes Princiers. Afin de s'en assurer, nous avons contacté le développeur d'*Odette*, F. Glorieux. Il nous a parlé de la refonte d'*Odette* dans *Teinte*, une application en développement continu qui permet la conversion de fichiers TEI, Docx, ePub et txt en TEI, Docx, HTML et txt⁷. Il s'agit d'un logiciel libre voué à la pérennité dont l'interface a été financée par le Labex ObTIC (Observatoire des textes, des idées et des corpus) rattaché à la Sorbonne. Nous avons tenté une conversion avec *Teinte*, d'un docx vers XML/TEI. Cette dernière s'est révélée très pertinente. Le document XML/TEI retourné est pourvu d'un TEI Header à remplir. Toutefois, ce dernier reste succinct et sera à repenser dans le cadre d'une édition globale des actes princiers.

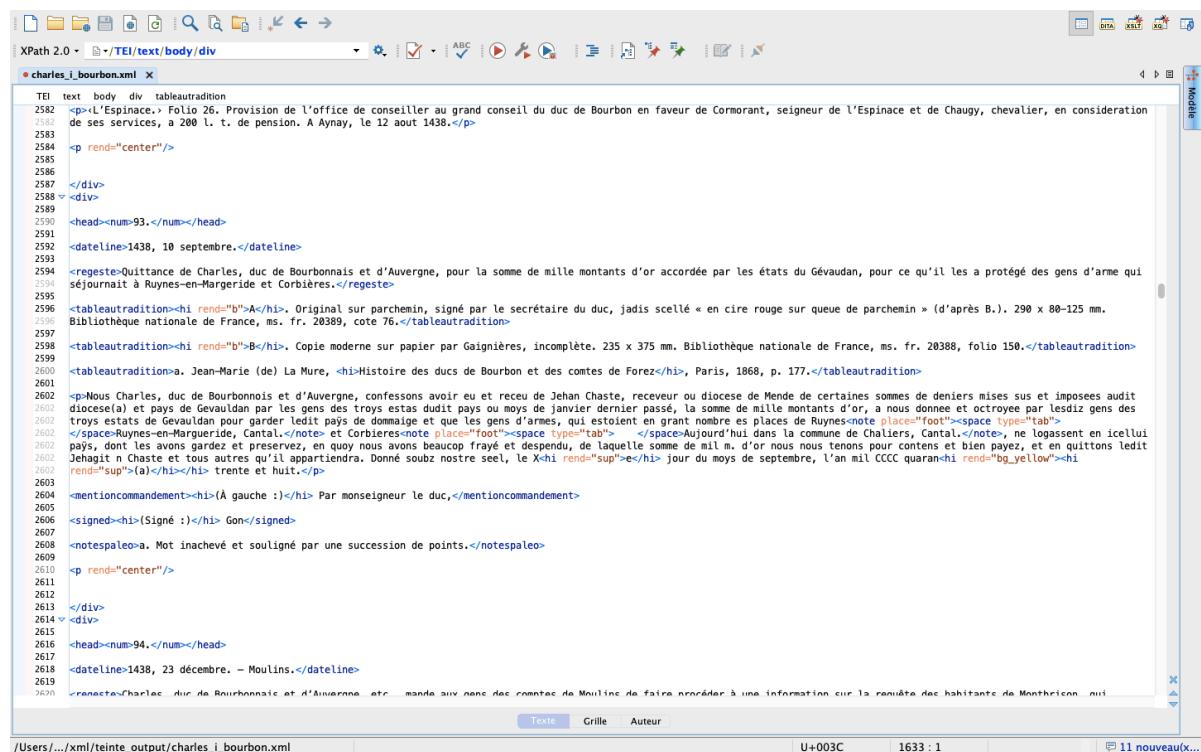


FIGURE 4.5 – Document XML/TEI converti avec *Teinte*.

Comme nous pouvons le voir ci-dessus, le recours à des styles personnalisés et normalisés par un utilisateur sur ODT permet la transformation de ces derniers en éléments (balise avec son contenu textuel). Cela implique une meilleure hiérarchisation des différents éléments diplomatiques. La compréhension est améliorée via la structuration du document et des différentes parties de l'édition dans des balises appropriées. De plus, *Teinte* est un projet développé pour la recherche en sciences humaines et sociales, voué à se pérenniser. Même si le résultat de la transformation ne renvoie pas des balises XML, ces dernières, ainsi que l'organisation du document pourront être repensés et retravaillés.

7. Obtic, Teinte, en ligne : <https://obtic.huma-num.fr/teinte/>.

Troisième partie

La chaîne de traitement des actes princiers

Chapitre 5

Structuration des XML à partir d'un script Python

La conversion des documents ODT en XML avec *Teinte* renvoie des documents TEI relativement structurés. Néanmoins, les balises ne sont pas conformes au format TEI. De plus, la structuration du document peut être encore améliorée, via la création d'attributs (caractéristiques d'une balise) par exemple. Le langage de programmation Python va permettre de lire les fichiers et d'intervenir dessus afin de parfaire l'organisation des actes. Python est un langage de programmation développé dans les années 1990 par le développeur Guido van Rossum. Il fournit une approche simple et efficace de la programmation orientée objet¹, ce qui en fait un langage idéal pour les scripts et le développement d'applications dans de nombreux domaines².

1. La programmation orientée objet, ou programmation par objet, est une manière d'approcher la programmation via l'interaction de briques logicielles appelées objets. Les objets sont des structures de données avec le comportement associé, manipulés dans un programme. In : *Programmation orientée objet*, IBM, 22 nov. 2021, URL : <https://www.ibm.com/docs/fr/spss-modeler/saas?topic=language-object-oriented-programming> (visité le 04/08/2023).

2. « Le tutoriel Python », in : *Contenu de la documentation Python*, Python documentation, URL : <https://docs.python.org/3/contents.html> (visité le 05/05/2023).

5.1 Extraction de données à partir des fichiers CSV et injections dans le XML

Une partie des données relatives aux actes de Charles I^{er}, d’Agnès de Bourgogne et d’Anne Dauphine sont présentes dans un document CSV. Le format CSV, dont les valeurs sont séparées par des virgules, est le format le plus commun dans l’importation et l’exportation de feuilles de calculs et de bases de données³. Les CSV répertorient pour chaque duc et duchesse des informations essentielles sur les actes comme leur numéro, l’institution de conservation, le fond dont ils sont issus (identifiants), leur état, leur type diplomatique, l’année et la date de rédaction, les signataires, le lieu, les mentions de commandement, la langue de l’acte, et la présence ou non d’un sceau.

	A	B	C	D	E	F	G	H	I
1	Acte_id	Num	Institution	idno1	idno2	Etat	Type diplomatique	Annee	Date
2	brb_agn_i_1423_05_12a	1	Archives départementales Côte d'Or	B 297	pièce scellée 305	Vidimus	Lettres patentes	1423	12/05/1423
3	brb_agn_i_1425_01_17a	2	Archives municipales Riom	AA 32		Original	Lettre missive	1425-1434	17 janvier
4	brb_agn_i_1425_08_04a	3	Archives municipales Riom	AA 32		Original	Lettre missive	1425-1434	4 août
5	brb_agn_i_1425_11_25a	4	Archives municipales Riom	AA 16	n° 511	Original	Lettre missive	1425-1434	25 novembre
6	brb_agn_i_1435_03_22a	5	Archives municipales Lyon	AA 25	cote 5	Original	Lettre missive	1435-1449	22 mars
7	brb_agn_i_1435_06_21a	6	Archives municipales Lyon	AA 25	cote 6	Original	Lettre missive	1435-1456	21 juin
8	brb_agn_i_1437_02_24a	7	Bibliothèque nationale de France	ms. fr. 20389	cote 89	Original	Quittance	1437	24/02/1437
9	brb_agn_i_1438_12_27a	8	Bibliothèque nationale de France	ms. fr. 20389	cote 90	Original	Quittance	1438	27/12/1438
10	brb_agn_i_1442_02_01a	9	Archives départementales Allier	A 160		Original	Quittance	1442	01/02/1442
11	brb_agn_i_1443_12_15a	10	Bibliothèque nationale de France	ms. fr. 20389	cote 91	Original	Quittance	1443	15/12/1443
12	brb_agn_i_1445_07_13a	11	Bibliothèque nationale de France	ms. fr. 20388	f. 170	Copie	Quittance	1445	13/07/1445
13	brb_agn_i_1445_09_01a	12	Bibliothèque nationale de France	ms. fr. 20389	cote 93	Original	Quittance	1445	01/09/1445
14	brb_agn_i_1445_09_24a	13	Bibliothèque nationale de France	ms. fr. 20389	cote 94	Original	Quittance	1445	24/09/1445
15	brb_agn_i_1445_11_13a	14	Bibliothèque nationale de France	ms. fr. 20389	cote 95	Original	Quittance	1445	13/11/1445
16	brb_agn_i_1446_03_08a	15	Bibliothèque nationale de France	ms. fr. 20389	cote 92	Original	Quittance	1446	08/03/1446
17	brb_agn_i_1446_08_18a	16	Bibliothèque nationale de France	ms. fr. 20389	cote 96	Original	Quittance	1446	18/08/1446
18	brb_agn_i_1446_09_18a	17	Bibliothèque nationale de France	ms. fr. 20389	cote 97	Original	Quittance	1446	18/09/1446
19	brb_agn_i_1446_12_14a	18	Bibliothèque nationale de France	ms. fr. 20389	cote 98	Original	Quittance	1446	14/12/1446
20	brb_agn_i_1447a	19	Archives nationales	P 1357/1	cote 346	Minute	Lettres patentes	1447	NS
21	brb_agn_i_1450_03_12a	20	Archives départementales Côte d'Or	B 11942	cote 247	Original	Lettre close	1450	12/03/1450
22	brb_agn_i_1454_11_11a	21	Archives nationales	P 1365/1	cote 1415/2	Copie	Lettre missive	1454	12/11/1454
23	brb_agn_i_1461_07_20a	22	Bibliothèque nationale de France	ms. fr. 20389	cote 99	Original	Quittance	1461	20/07/1461
24	brb_agn_i_1470_11_18a	23	Archives nationales	K 188	cote 30/4	Copie	Lettres patentes	1470	18/11/1470
25	brb_agn_i_1471_09_03a	24	Bibliothèque nationale de France	ms. fr. 3924	f. 20 verso	Copie	Lettre missive	NS	3 septembre
26	brb_agn_i_1471_11_11a	25	Bibliothèque nationale de France	ms. fr. 15538	folio 138	Copie	Lettre missive	NS	11 novembre

FIGURE 5.1 – CSV actes d’Agnès de Bourgogne.

L’enjeu est d’extraire une partie des données des CSV : la date, l’état et le type diplomatique de chaque acte (valeurs surlignées en rose dans le tableau ci-dessus) afin de les ajouter comme attributs de la balise `<div>` du fichier XML créé pour chaque corpus. La date permettra également la création d’un attribut prenant la valeur d’un identifiant unique (@xml :id), pour chaque acte. Afin d’illustrer au mieux les modifications réalisées via Python, les figures qui suivent (figures 5.2 à 6.2) montrent les états successifs de l’acte daté du vingt-quatre février 1437 dans la chaîne de traitement des actes princiers.

3. « Lecture et écriture de fichiers CSV », in : *Ibid.*

Pour extraire les données en question, un script Python a été écrit⁴.

D'abord, une première fonction assure la lecture des fichiers XML à l'aide de la librairie *BeautifulSoup*. Cette dernière permet l'extraction de données de fichiers HTML et XML. Afin de pouvoir l'utiliser correctement avec des fichiers XML, il est nécessaire d'importer le module « bs4 » et d'installer le parseur « lxml » conçu à cet effet. *BeautifulSoup* fournit des moyens de navigation, de recherche et de modification de l'arbre XML.

Ensuite, une deuxième fonction constitue une liste avec les données du CSV qui seront transformées en attributs de la balise <div>. Elle nécessite d'importer le module « csv », qui permet de lire et d'écrire des données au format tabulaire CSV via les objets « reader » et « writer ». La fonction applique aux différents formats de dates un format standard afin de les agréger à un identifiant unique : xmlid. Deux modules sont importés pour assurer ces opérations : le module « datetime », qui fournit des classes pour manipuler les dates qui est complété par le module « locale » qui prend en compte les spécificités de datation locales. La fonction comporte une première boucle qui itère sur chaque ligne du CSV. Si la valeur de la date est égale à « Dep. » (**deperditum**), alors la date est manquante et un identifiant est créé. Si la date n'est pas « Dep. », une série de formats de date possibles est définie et stockée dans une liste. La fonction comprend une seconde boucle qui itère sur chaque format de date, les analyse, et en assure la conversion dans un format standard (YYYYMMDD). Elle permet également la création d'un identifiant unique pour chaque acte, à partir de cette date. La fonction retourne un dictionnaire composé des données extraites du CSV : date standard, état et type diplomatique, en préparation de leur injection aux fichiers XML.

Enfin, une troisième fonction injecte les données dans le XML. Après ouverture et lecture du fichier XML, elle itère sur toutes les balises <div> du <body> et leur ajoute les données extraites et modifiées comme attributs. L'attribut « n » prend comme valeur la date au format standard, l'attribut « xml:id », l'identifiant unique créé à partir de cette date, l'attribut « type », l'état et l'attribut « subtype », le type diplomatique. Quelques opérations de mise en forme des documents ont été effectuées, comme la suppression des lignes vides dans le nouveau document XML, via des expressions régulières, ce qui a nécessité l'importation du module « re ». Les apostrophes odt sont également remplacées par des apostrophes standard et les balises <dateline> par des balises <head>. Les fichiers de sortie comportent les modifications réalisées par les trois fonctions.

4. GitLab (py/add_csv_infos.py).

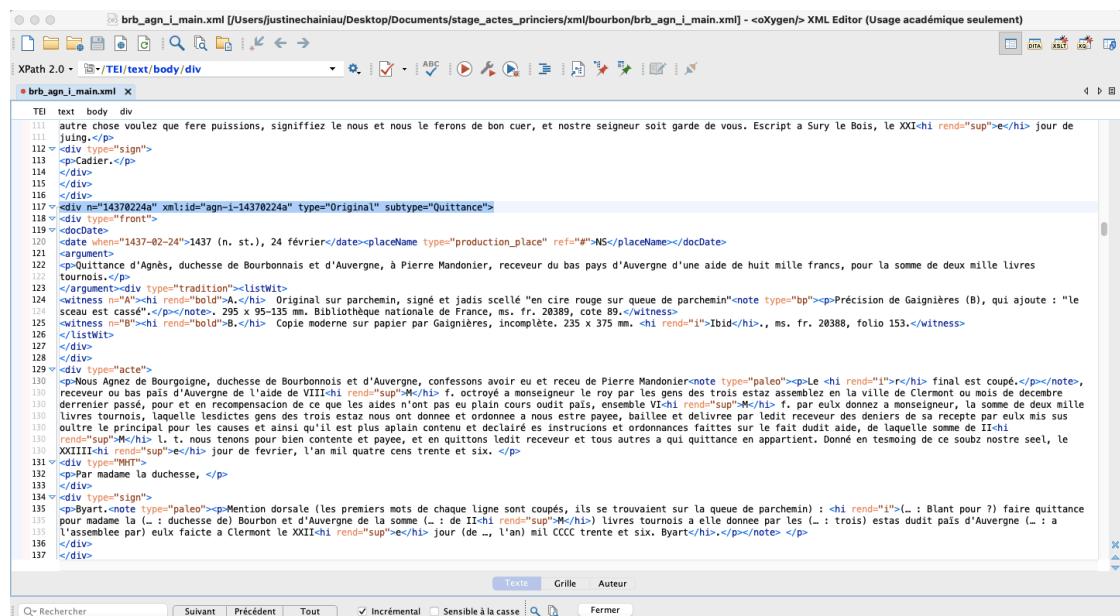


FIGURE 5.2 – XML après injection des données avec Python.

Les fonctions créées ont permis l'extraction de données de fichiers CSV et l'ajout de ces mêmes données dans du XML pour les actes d'Anne Dauphine, de Charles I^{er} et d'Agnès de Bourgogne. Toutefois, les actes du corpus de Louis II ne sont pas répertoriés dans un CSV. Nous avons donc inversé la procédure afin d'en créer un.

La fonction `xml_to_csv` créé un fichier CSV à partir de données extraites du fichier XML répertoriant les actes de Louis II⁵. Une boucle itère sur chaque balise `<div>` et remplit, à partir d'information récupérées dans le XML, une liste. La fonction récupère le contenu de la balise `<dateline>`, qu'elle sépare en deux parties à partir d'une boucle. Si l'élément séparateur « - » est présent dans le texte de la variable date, alors ce texte est divisé en deux parties qui forment deux items, afin de séparer la date de temps de la date de lieu. Sinon, le lieu n'est pas précisé et la mention « NS »⁶ est inscrite. La chaîne de caractère « Louis II » est définie par défaut comme avant-dernier item de la liste, comme valeur de la colonne « prince ». Pour la colonne « signataire », si la balise `<signed>` est présente dans les `<div>`, alors on complète la liste avec le contenu textuel de la balise `<signed>`, en omettant les diverses mentions « (signé :) ». Autrement, la mention « NS » lui est ajoutée. La mention « fr » est répertoriée par défaut comme dernier item de la liste, comme valeur de la colonne « langue ». En effet, seuls trois actes de Louis II sont en latin et ont été repérés préalablement : la colonne langue prend alors la valeur « lat » pour les actes concernés⁷.

5. GitLab (py/div_attributes.py).

6. NS : Non signalée.

7. La Mure III, p.132 (10 juin 1370) - AD Allier, 1 G 3, fol. 1-6 (6 décembre 1386) AN, P 1356², c. 280 (18 septembre 1399).

Ensuite, le fichier CSV est ouvert via la fonction « open() » de Python, et les lignes sont remplies à partir des données stockées dans la liste. La ligne d'en-tête du fichier CSV est écrite à partir de la spécification du nom des colonnes : « Date », « Lieu », « Prince », « Signataire », « Langue », « État », « Type diplomatique ». Néanmoins, ces deux dernières colonnes n'ont pu être complétées automatiquement, mais manuellement via le relevé de l'état de chaque acte et de son type diplomatique à partir du document ODT.

La programmation en Python permet d'automatiser la manipulation des données, et ainsi de gagner du temps dans la chaîne de traitement. Toutefois, le processus se fonde sur les premiers corpus, qui disposent de fichiers CSV répertoriant les actes. En effet, la chaîne de traitement des actes princiers, telle qu'elle a été pensée par les ingénieurs du projet, nécessite qu'un certain nombre d'informations sur les actes soient présentes dans des fichiers CSV. Il faudra donc que les nouveaux corpus ajoutés au projet aient un CSV associé. Autrement, Python permet leur création via des fonctions, mais toute la procédure ne peut pas être automatisée, et toutes les colonnes remplies à partir de l'extraction de données. Des notions de diplomatique sont nécessaires afin de repérer, via la lecture de tous les actes, les états et les types diplomatiques, ce qui prend un certain temps, notamment dans le cas de corpus volumineux, comme ceux de Louis II et de Charles I^{er}. Le projet « Actes princiers » pourrait donc conseiller aux chercheurs qui élaborent des éditions diplomatiques de dresser en parallèle un tableur avec les différentes métadonnées. Les corpus disposent désormais tous d'un fichier CSV rempli, et les fichiers XML d'une balise <div> pourvue d'attributs qui contribuent à affiner l'organisation du document. Chaque acte est contenu dans une division (<div>) du document XML et identifié en tant que tel. Comme l'ensemble des balises n'est pas normalisé, les fonctions vont permettre de restructurer de manière automatique les documents XML.

5.2 Restructuration des documents XML issus de la conversion

Les documents XML obtenus en sortie, après intervention via un script Python, comportent désormais une balise `<div>` structurée. Toutefois, c'est l'ensemble du document qui nécessite d'être hiérarchisé. Il s'agit donc en utilisant Python de restructurer les documents selon le plan d'encodage ci-après, notamment en remplaçant les balises obtenues après conversion avec *Teinte* en balises TEI afin d'obtenir un document XML valide. Comme l'indique le schéma ci-après, chaque acte doit être encodé dans une balise `<div>` comportant cinq autres balises `<div>`, qui représentent des subdivisions du corps du document⁸.

La première d'entre elles comporte les éléments de datation et d'analyse. Les dates de temps et de lieu sont encodées dans deux balises distinctes. Le tableau de la tradition est ensuite structuré en une liste de témoins (`<listWit>`). Chaque témoin est encodé dans une balise dans laquelle sont imbriquées des balises pour le texte, l'institution de conservation et les identifiants (fonds et cote). Ensuite, une troisième `<div>` contient la teneur de l'acte dans une balise `<p>`, qui marque le début d'un paragraphe en prose⁹. Si l'acte comporte un titre, il est encodé dans une balise `<head>`, qui marque l'en-tête (`<head>Texte établi d'après A.</head>, <head>(Deperditum)</head>`)¹⁰. Pour finir, deux dernières balises `<div>` contiennent les mentions hors teneurs et les signataires dans des balises `<p>`. Ces opérations sont consignées dans des fichiers Python individuels pour chaque corpus : `regex_anne_dauphine`¹¹, `regex_brb_agnes`¹², `regex_charles`¹³, `regex_louis`¹⁴. Il n'a pas été possible de traiter les quatre corpus à partir d'un seul script, car les documents ODT initiaux ne sont pas exactement structurés de la même manière.

8. TEI Guidelines, `<div>`, en ligne : <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html>.

9. TEI Guidelines, `<p>`, en ligne : <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-p.html>.

10. TEI Guidelines, `<head>`, en ligne : <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-head.html>.

11. GitLab (py/regex_anne_dauphine.ipynb).

12. GitLab (py/regex_brb_agnes.ipynb).

13. Gitlab (py/regex_charles_i.ipynb).

14. GitLab (py/regex_louis_ii.ipynb).

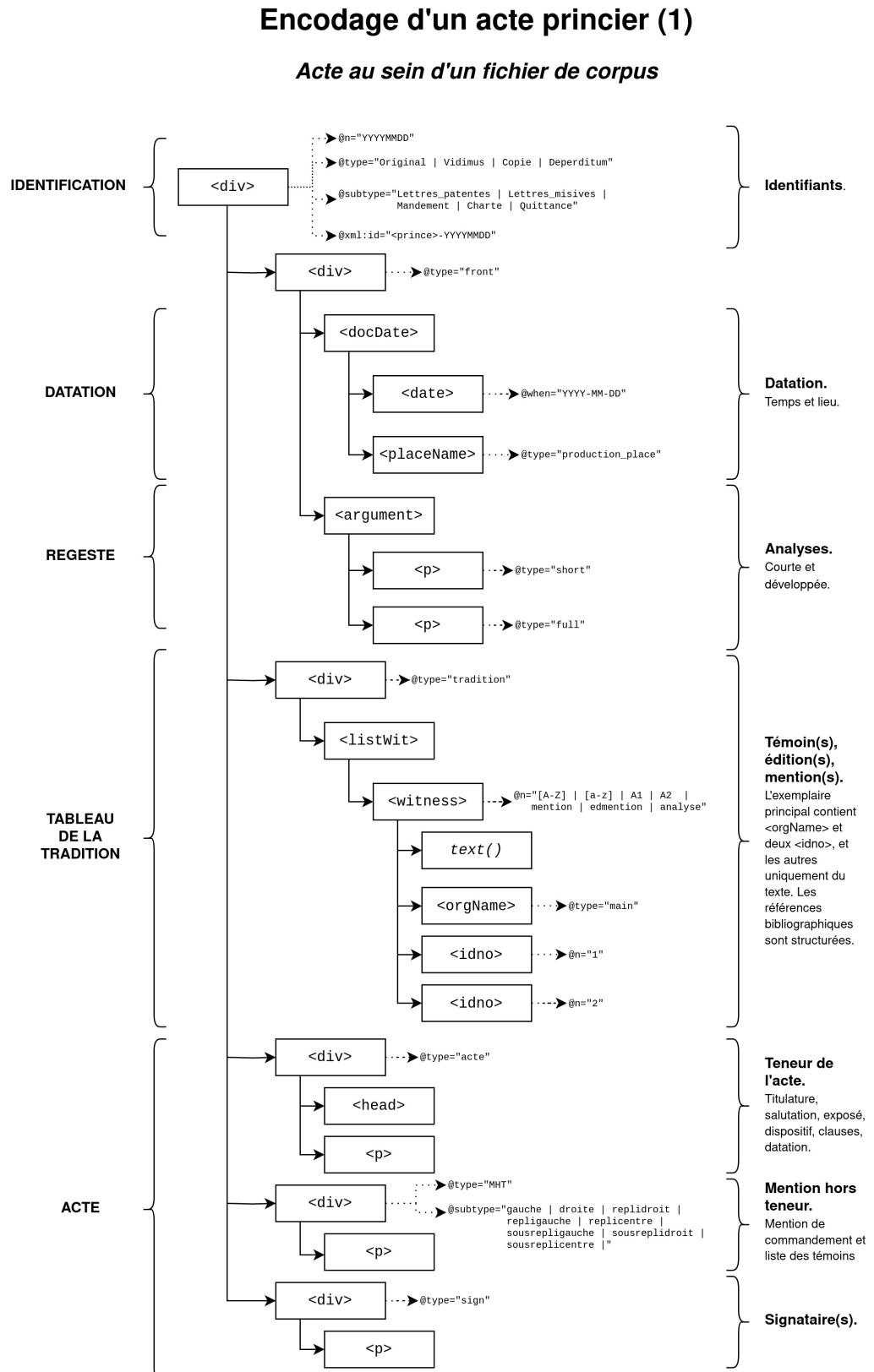


FIGURE 5.3 – Schéma d'encodage des actes en XML. Source : Jean-Damien Généro.

D'abord, le script réalise des traitements automatiques via des regex pour améliorer la mise en forme des documents. Grâce à la fonction « re.sub », le code supprime un certain nombre d'éléments inutiles (balises inutiles, lignes vides) et substitue certaines balises afin d'obtenir une structuration plus fine. Ces opérations de substitution permettent l'insertion des principales `<div>` et éléments structurants comme `<docDate>`, spécifique pour la datation de documents¹⁵, `<argument>`, cette dernière contenant une liste formelle ou une description en prose des sujets abordés dans une subdivision¹⁶ et `<listWit>`, qui énumère tous les témoins mentionnés par un appareil critique, éventuellement regroupés hiérarchiquement¹⁷. Les illustrations à la page suivante donnent un aperçu des fichiers avant et après les interventions. Les premières balises structurantes sont ajoutées : `<div type='front'>` et `<docDate>` avec ses balises enfants. La balise `<regeste>` (qui peut être suivi d'une balise `<hi>`, marquant un mot ou une phrase comme graphiquement distinct du texte environnant¹⁸) est substituée par la balise `<argument>` suivie d'une balise `<p>`. La fonction « re.sub » permet la fermeture de certaines balises comme `<p>` et `<argument>` et d'intégrer les balises ouvrantes suivantes : `<div type='tradition'>` et `<listWit>`. Cette opération est répétée afin de fermer la balise `</listWit>` et d'ajouter les suivantes : `</div>`, `</div>`, `<div type="acte">` et `<p>`. Les cas particuliers des actes commençant par un titre ou comportant la mention *deperditum* sont réglés par des regex. Puis, la balise `<div type="acte">` est fermée, ainsi que celle du dernier acte, dont le traitement n'est pas automatique. Le `<tableautradition>` est supplété par la balise `<witness>` contenant la description précise d'un seul témoin¹⁹, avec ses attributs et suivi du n° d'exemplaire. Les spécifications `<analyse>`, `<edmention>` et `<mention>` sont toutes remplacées par la balise `<witness>`, dont la balise fermante est ajoutée. Les dernières balises `<mention-commandement>` et `<signed>` sont changées en balises `<div>` disposant respectivement d'attributs « MHT » et « sign ».

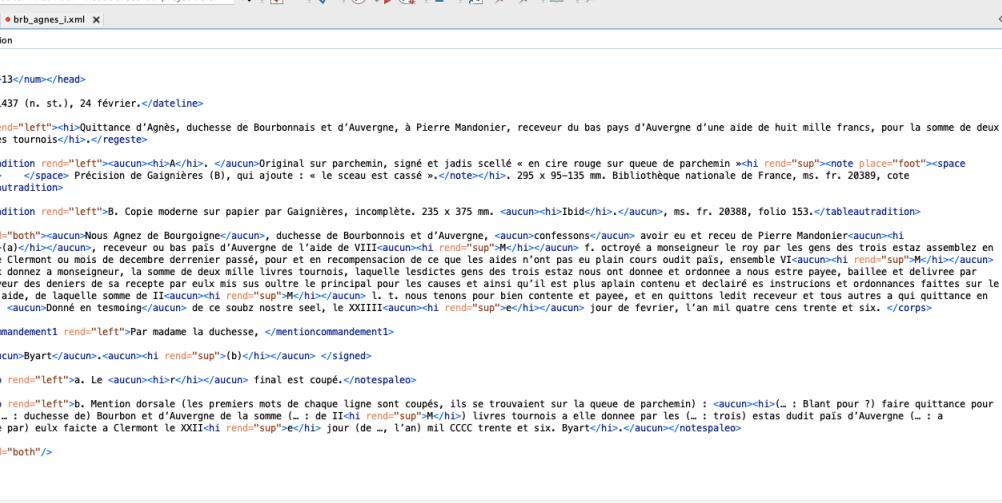
15. TEI Guidelines, `<docDate>`, en ligne : <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-docDate.html>.

16. TEI Guidelines, `<argument>`, en ligne : <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-argument.html>.

17. TEI Guidelines, `<listWit>`, en ligne : <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-listWit.html>.

18. TEI Guidelines, `<hi>`, en ligne : <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-hi.html>.

19. TEI Guidelines, `<witness>`, en ligne : <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-witness.html>.



brb_agnes_i.xml [/Users/justinechainais/Desktop/Documents/stage_actes_princiers/xml/teinte_output;brb_agnes_i.xml] - <oXygen> XML Editor (Usage académique seulement)

XPath 2.0 - Exécuter XPath sur 'Ressources du projet sélectionné'

brb_agnes_i.xml

processing-instruction

176 <div>

177 <head><num>13</num></head>

178 <dateline>1437 (n. st.), 24 février.</dateline>

179 <regeste rend="left"><hi>Quittance d'Agnes, duchesse de Bourbonnais et d'Auvergne, à Pierre Mandonier, receveur du bas pays d'Auvergne d'une aide de huit mille francs, pour la somme de deux

180 mille livres tournois</hi>.</regeste>

181 <tableauradition rend="left"><acun><hi>A</hi>.</acun>Original sur parchemin, signé et jadis scellé « en cire rouge sur queue de parchemin »<hi rend="sup"><note place="foot"><space>

182 <space> Précision de Gaignières (B), qui ajoute : « le sceau est cassé ».</note></hi>, 295 x 95-135 mm. Bibliothèque nationale de France, ms. fr. 20389, cote

183 89.</tableauradition>

184 <tableauradition rend="left">B. Copie moderne sur papier par Gaignières, incomplète. 235 x 375 mm. <acun><hi>b1b</hi>.</acun>, ms. fr. 20388, folio 153.</tableauradition>

185 <corps rend="both">aucun:Nous Agnez de Bourgoigne</acun>, duchesse de Bourbonnais et d'Auvergne, <acun>confessons</acun> avoir eu et receue de Pierre Mandonier<acun><hi>

186 rend="sup">(a)</hi>.</acun>, receveur ou bas pais d'Auvergne de l'aide de VIIIaucun><hi rend="sup">M</hi></acun> f. octroya a monseigneur le roy par les gens des trois estaz assemblez en

187 la ville de Clermont ou mois de decembre derrenier passé, pour et recompensation de ce que les aides n'ont pas eu plain cours oudit pais, ensemble VIaucun><hi rend="sup">M</hi></acun>

188 f. pour eux dous et asselelement la somme de deux mille livres tournois laquelle lessidtes gens des trois estaz nous ont deute et ordonnee a estre payee a balletee et delivree par

189 et au devant des deniers de sa recete par eux mesme sus oultre le principal pour les causes et ainsi qu'il est plus aplain contenu et declaré es instructions et ordonnances faites sur le

190 fait dudit aide, de laquelle somme de IIaucun><hi rend="sup">M</hi></acun> t. t. nous tenons pour bien contente et payee, et en quittions ledit receveur et tous autres a qui quittance en

191 appertient.<acun>Donné en temsloix</acun> de ce soub nostre seel, le XXIIIaucun><hi rend="sup">e</hi></acun> jour de fevrier, l'an mil quatre cens trente et six.</corps>

192 <mentioncommandement rend="left">Par madame la duchesse, <mentioncommandement>

193 <signed><acun>Byart</acun>.<acun><hi rend="sup">(b)</hi></acun> <signed>

194 <notespaleo rend="left">a. Le <acun><hi>r</hi></acun> final est coupé.</notespaleo>

195 <notespaleo rend="left">b. Mention dorsale (les premiers mots de chaque ligne sont coupés, ils se trouvaient sur la queue de parchemin) : <acun><hi>(... : Blant pour ?) faire quittance pour

196 madame la (...) duchesse de Bourbon et d'Auvergne de la somme (...) :<hi rend="sup">M</hi> livres tournois a elle donnee par les (...) : trois) estas dudit pais d'Auvergne (... : a

197 l'assemblée par) eux faicta a Clermont le XXII<hi rend="sup">e</hi> jour (de .., l'an) mil CCCC trente et six. Byart</hi>.</acun><notespaleo>

198 <corps rend="both">

199 <corps rend="both">

200 <corps rend="both">

201 </div>

202 </div>

Titre Grille Auteur

Suivant Précédent Tout Incrémental Sensible à la casse Fermer

Rechercher

U:/Users/justinechainais/Desktop/Documents/stage_actes_princiers/xml/teinte_output;brb_agnes_i.xml U+003C 2 : 1

11 nouveau(x)

FIGURE 5.4 – Encodage de l'acte daté du 24 février 1437 avant intervention.

brb_agnes_i.xml [Users/justinechainiau/Desktop/Documents/stage_actes_princiers/xml/corpus_prince(brb_agnes_i.xml)] - <oXygen/> XML Editor (Usage académique seulement)

XPath 2.0 • Exécuter XPath sur 'Ressources du projet sélectionné' • ADF •

brb_agnes_i.xml X brb_agnes_i.xml X

```
168 </div>
169 </div>
170 </div>
171 <div where="#14370224a" xml:id="agn-i-14370224a" type="Original" subtype="Quittance">
172 <div type="front">
173 <div>
174 <date when="1437-02-24">1437 (n. st.), 24 février.</date><placeName type="production_place" ref="#">NS</placeName></docDate>
175 <argument>
176 <p>Quittance d'Agnes, duchesse de Bourbonnais et d'Auvergne, à Pierre Mandonier, receveur du bas pays d'Auvergne d'une aide de huit mille francs, pour la somme de deux mille livres tournois.</p>
177 </argument>
178 <div type="tradition">
179 <listwit>
180 < witness >A<hi rend="bold">A.</hi> Original sur parchemin, signé et jadis scellé "en cire rouge sur queue de parchemin"<note type="bp"><p>Précision de Gaignières (B), qui ajoute : "le sceau est cassé".</p></note> 295 x 95-135 mm. Bibliothèque nationale de France, ms. fr. 20389, cote 89.</witness>
181 < witness >B<hi rend="bold">B.</hi> Copie moderne sur papier par Gaignières, incomplète. 235 x 375 mm.<hi rend="i">Ibid</hi>, ms. fr. 20388, folio 153.</witness>
182 </listwit>
183 </div>
184 </div>
185 <div type="ctc">
186 <div> Agnes de Bourgogne, duchesse de Bourbonnais et d'Auvergne, confessons avoir eu et reçu de Pierre Mandonier<note type="paled"><p>Le <hi rend="i">r</hi> final est coupé.</p>
187 <note> receveur ou bailli d'Auvergne de l'aide de VIII<hi rend="sup"><sup>M</sup></hi> f. octroyé a monseigneur le roy par les gens des trois estats assemblés en la ville de Clermont ou mois de decembre derrenier passé, pour et en recompension de ce que les aides n'ont pas eu plain cours ouidit pais, ensemble VI<hi rend="sup"><sup>M</sup></hi> f. par eux donnez a monseigneur, la somme de deux mille livres tournois, laquelle lesdites gens des trois estats nous ont donnee et ordonnee a nous estre payee, bailee et gardée par ledit receveur des deniers de sa recepe por eux mis sus outre le principal pour les causes et autres qu'il est plus plaisir contenu et payee, et en quittons ledit receveur et tous autres a qui quittance en appertient. Donné en temsion de ce soubz nostre sei, le XXXIII<hi rend="sup"><sup>M</sup></hi> l. t. nous tenons pour bien contente et payee, et en quittions ledit receveur et tous autres a qui quittance en appertient. Donné en temsion de ce soubz nostre sei, le
188 <div type="MHT">
189 <p>Par madame la duchesse, </p>
190 </div>
191 <div type="sign">
192 <p>Byart.<note type="paled"><p>Mention dorsale (les premiers mots de chaque ligne sont coupés, ils se trouvaient sur la queue de parchemin) : <hi rend="i">(... : Blant pour ?) faire quittance pour madame la (...) : duchesse de Bourbon et d'Auvergne de la somme (...) : de II<hi rend="sup"><sup>M</sup></hi> livres tournois a elle donnee par les (...) : trois) estas dudit pais d'Auvergne (...) : a l'assemblée par eux faicté a Clermont le XXXII<hi rend="sup"><sup>E</sup></hi> jour (de ..., l'an) mil CCCC trente et six. Byart</hi>.</p>
193 </note></p>
194 </div>
195 </div>
196 </div>
```

FIGURE 5.5 – Encodage de l'acte daté du 24 février 1437 après intervention.

Ensuite, le script réalise des traitements semi-automatiques. En raison des disparités entre les quatre corpus, toutes les opérations ne peuvent pas être réalisées de manière automatique. De plus, les regex de l'étape précédente n'ont pas toujours fonctionné. En effet, la chaîne de traitement bute sur des erreurs de mise en forme des éditeurs. L'option de validation d'Oxygen nous a permis de repérer les problèmes d'encodage et de les corriger un à un dans le script via une regex dédiée. Ces irrégularités sont pour une grande partie liées à la balise `<hi>` qui est parfois présente en trop ou manquante. De plus, les mentions « (Signé :) » comportent des fautes d'encodage et n'ont, par conséquent, pas toujours été supprimées par le script. Les erreurs spécifiques à chaque corpus sont traitées individuellement. Toutes les corrections sont stockées dans un dictionnaire qui se substituera aux valeurs précédentes. Les dernières erreurs concernent la mise en forme des notes de bas de page et des notes paléographiques. Le nombre d'appels de note diffère du nombre de notes. Cela est généralement dû aux manquements dans le stylage effectué précédemment, ou à leur présence en double. Après quelques corrections de mise en forme : traitement des espaces insécables, des guillemets, des balises `<hi>` sans attributs, le fichier est écrit dans un document. Le parseur `lxml` repère les fautes de syntaxe et traite les caractères non valides. Les éléments HTML, s'il y en a, sont retirés et les octets XML sont décodés en une chaîne de caractères afin d'inscrire le contenu XML dans un fichier.

Enfin, en fonction des erreurs relevées, des corrections sont effectuées avec `lxml`. L'objectif principal est la structuration des dates en un bloc. Pour cela, il s'agit d'ajouter à la nouvelle balise `<date>` un attribut « `when` » avec la date au format standard `YYYYMMDD`, ou `YYYYMM` ou `YYYY`. Puis, les notes paléographiques sont alignées sur leurs appels de note. La dernière correction concerne les attributs « `xml:id` » et « `n` ». En effet, lorsque deux actes sont datés du même jour, ils ont le même identifiant, or ces derniers doivent être uniques dans un document XML.

La suppression et la substitution de balises via une fonction Python a permis de structurer très finement les documents avec des balises XML imbriquées et dotées d'attributs. Toutefois, certaines opérations se sont montrées plus complexes, comme la hiérarchisation du tableau de la tradition, qui n'est pas aussi fine que sur le schéma d'encodage établi en amont²⁰. En effet, la balise `<witness>` ne comporte pas les balises `<text>`, `<orgName>` et `<idno>`. Même si le document obtenu en sortie ne respecte pas exactement le schéma théorique, la plupart des balises ont été transformées. Les fichiers XML bénéficient désormais d'une structure robuste permettant leur division en fichiers XML individuels pour chaque acte.

20. Figure 5.3 – Schéma d'encodage des actes en XML.

5.3 Séparation des fichiers XML de corpus en fichiers XML d'actes

L'objectif est de transformer le fichier XML contenant l'ensemble des actes d'un prince obtenu après interventions en autant de fichiers XML qu'il y a d'actes. Une fonction Python : `split_corpus`²¹ est utilisée pour réaliser cette opération. Cette fois-ci, une seule fonction est adaptable pour les quatre corpus, dans la mesure où ils ont été normalisés en amont. La première étape est la création d'un fichier XML par acte. Cela passe par la récupération des `<div>` des actes en itérant sur chaque balise. Les données sont alors stockées dans une liste, qui, couplée au canevas TEI suivant, permet la constitution du fichier.

CANEVAS TEI

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
<fileDesc>
<titleStmt>
<title level="s">Actes princiers</title>
<title level="m">+++TITLE+++</title>
<title level="a"></title>
<respStmt>
<resp>transcribed by</resp>
<name>+++TRANSCRIPTEUR+++</name>
</respStmt>
</titleStmt>
<editionStmt>
<edition>Acte édité dans le cadre du programme Actes princiers.</edition>
<respStmt>
<resp>direction scientifique</resp>
<name>Olivier Mattéoni</name>
</respStmt>
<respStmt>
<resp>direction technique</resp>
<name>Jean-Damien Généro</name>
</respStmt>
<respStmt>
<resp>direction technique</resp>
<name>Nicolas Perreaux</name>
```

21. GitLab (`py/split_corpus.ipynb`).

```

</respStmt>
<respStmt>
<resp>stagiaire de l'École nationale des chartes</resp>
<name>Justine Chainiau</name>
</respStmt>
</editionStmt>
<publicationStmt>
<publisher>Laboratoire de Médiévistique occidentale de Paris (UMR 8589), Centre de recherches historiques (UMR 8558)</publisher>
<authority>Olivier Mattéoni</authority>
<date when="2023">2023</date>
<availability><licence source="https://github.com/etalab/licence-ouverte/blob/master/open-licence.md">Distributed under an Open License 2.0</licence></availability>
</publicationStmt>
<sourceDesc>
<msDesc>
<msIdentifier>
<repository></repository>
</msIdentifier>
<msContents>
<msItem>
<docDate></docDate>
</msItem>
</msContents>
</msDesc>
<listPerson>
<listPerson type="prince"/>
<listPerson type="signatory"/>
</listPerson>
</sourceDesc>
</fileDesc>
<profileDesc>
<abstract> <p></p> </abstract>
</profileDesc>
</teiHeader>
<text>
<body>
</body>
</text> </TEI>

```

La fonction complète le TEI Header ci-dessus avec une partie des données présente dans le CSV. Le nom du prince à l'origine de l'acte est ajouté à la balise `<listPerson type="prince">`, ceux des signataires dans la balise `<listPerson type="signatory">`. L'institution de conservation, le fonds et la cote sont également extraits des colonnes du CSV. Les éléments HTML sont encore une fois retirés, et les octets XML décodés en une chaîne de caractère afin d'inscrire le contenu XML dans un fichier.

L'activation des fonctions passe par la création d'un dictionnaire contenant en clef l'identifiant de la maison et celui du prince, et en valeur une liste avec : le chemin vers le fichier XML du prince depuis le dossier /py, le nom du CSV du prince, le nom du dossier où seront enregistrés les XML des actes, le titre du corpus et le nom du transcriveur. La fonction est activée via une itération sur le dictionnaire et permet la transformation des fichiers. Le fichier ci-après montre le résultat obtenu après division du fichier contenant tous les actes du corpus.

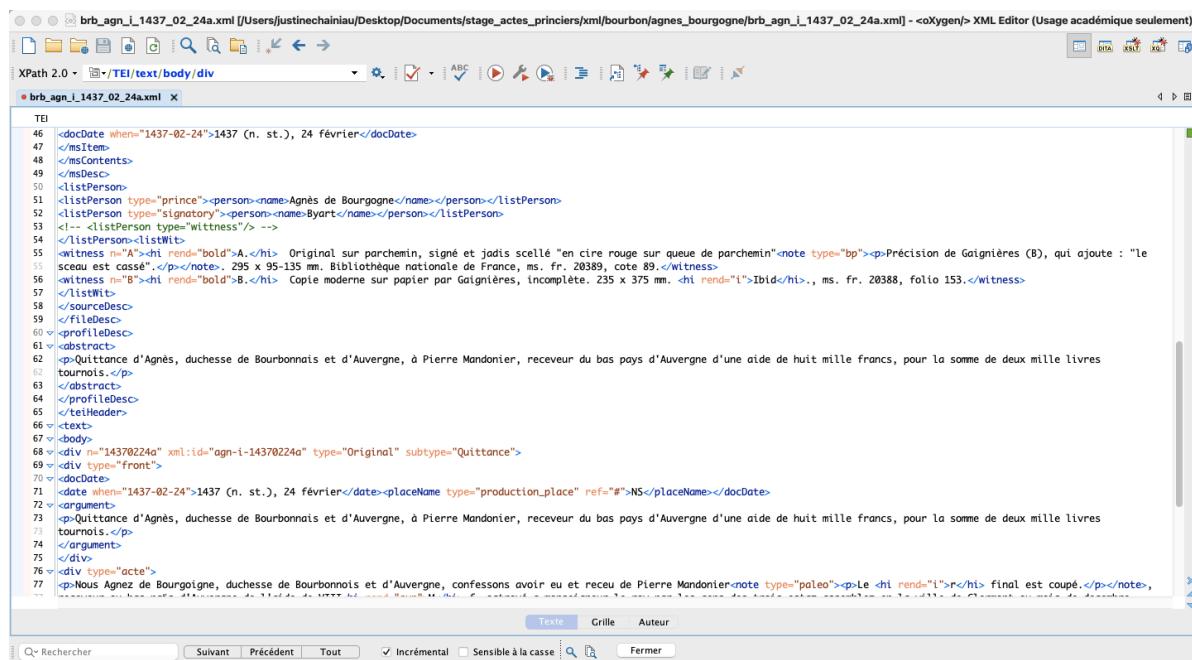


FIGURE 5.6 – Fichier XML de l'acte daté du 24 février 1437.

Les fichiers individuels sont pourvus d'un TEI Header complété d'informations générales, ajoutées à tous les actes. Le `<Filedesc>` comporte la description bibliographique du document. La balise `<titleStmt>` comprend plusieurs balises imbriquées correspondant à plusieurs niveaux de titre. Ensuite, la balise `<editionStmt>` présente le projet d'édition ainsi que les personnes en charge de celui-ci. Les indications sur la diffusion du fichier numérique et ses droits d'utilisation (licence) sont dans le `<publicationStmt>`. Les données individuelles, propres à chaque fichier et extraites du CSV sont placées dans le `<sourceDesc>`, qui décrit le document source. Cette balise contient les éléments d'identification, de conservation, de datation de l'acte ainsi que la liste des personnes concernées (signataires et prince à l'origine de l'acte) et le tableau de la tradition. Ces éléments sont suivis du `<profileDesc>`, qui fournit une description détaillée des aspects non bibliographiques, des informations contextuelles, notamment via l'analyse de l'acte en question. Le TEI Header est suivi de l'édition encodée pour chaque acte.

L'utilisation de Python via des fonctions a permis de manipuler les données et ainsi d'améliorer la structuration des documents. D'abord via l'extraction de données dans d'autres documents qui n'étaient pas au format XML (CSV), puis via la substitution des balises initiales issues de la conversion en des balises TEI standard, enfin par la séparation des fichiers XML obtenus en fichiers individuels. Ces étapes ont répondu à l'objectif fixé initialement, à savoir l'encodage des éditions des actes selon le schéma établi en amont, et sont préalables à la mise en ligne des documents.

Chapitre 6

Préparation à la mise en ligne des données

Ce chapitre présente les dernières étapes de la chaîne de traitement, préalables à la mise en ligne des données. Avant cette dernière, les fichiers XML sont uniformisés, et bénéficient de corrections individuelles, souvent manuelles, puis sont transformés en LaTeX via XSLT et compilés en PDF. L'objectif est de pouvoir relire les actes et d'y repérer les erreurs, sans avoir à ouvrir un à un les documents XML. LaTeX est un langage et un système de composition de documents largement employé par la communauté scientifique qui met l'accent sur la structure et le contenu du texte, en codant la mise en page des documents¹.

1. *LaTeX — DMS*, WikiDMS - Université de Montréal, URL : <https://dms.umontreal.ca/wiki/index.php/LaTeX> (visité le 22/07/2023).

6.1 Transformation du XML en LaTeX et compilation en PDF

Le langage LaTeX sépare le fond de la forme en définissant sémantiquement le contenu via des commandes. Il permet une gestion automatique du référencement (gestion des différentes parties d'un ouvrage : chapitres, citations, notes ...) et une typographie de qualité². Pour l'édition d'actes, les différentes parties sont mises en valeur. Chaque acte édité est inséré dans un groupe. La date fait office de titre en gras, l'analyse figure ensuite en italique, puis viennent le tableau de la tradition et le contenu textuel. Toutes ces parties sont séparées par des commandes de mise en forme (espaces, sauts de lignes ...).

```

356
357 \section{14370224a}
358
359
360
361 \begin{group}
362   \begin{center}
363     \textbf{(1437 (n. st.), 24 février)}
364     \textit{Agnès d'Agnès, duchesse de Bourgogne, à Pierre Mandonier, receveur du bas pays d'Auvergne d'une aide de huit mille francs, pour la somme de deux mille livres tournois.}
365   \end{center}
366 \textit{Quittance d'Agnès, duchesse de Bourgogne et d'Auvergne, à Pierre Mandonier, receveur du bas pays d'Auvergne d'une aide de huit mille francs, pour la somme de deux mille livres tournois.}
367 }
368 \bigskip
369 \begin{small}
370   \begin{center}
371     \textit{Original sur parchemin, signé et jadis scellé "en cire rouge sur queue de parchemin"\footnote{Précision de Gaignières (B), qui ajoute : "le sceau est cassé".} 295 x 95-136 mm. Bibliothèque nationale de France, ms. fr. 20389, cote 89.}
372     \textit{Copie moderne sur papier par Gaignières, incomplète. 235 x 375 mm. \textit{Ibid.}, ms. fr. 20388, folio 153.}
373   \end{center}
374 \end{small}
375 \end{group}
376
377 \bigskip
378
379
380
381
382 Nous Agnes de Bourgogne, duchesse de Bourgogne et d'Auvergne, confessons avoir eu et reçus de Pierre Mandonier\footnote{Le \textit{(t)} final est coupé.}, receveur ou bas pais d'Auvergne de l'aide de VIII\textsuperscript{(M)} f. octroyé a monseigneur le roy par les gens des trois estaz assembléz en la ville de Clermont ou mois de decembre derrenier passé, pour et en recompensation de ce que les aides n'ont pas eu plain cour ou dit pais, ensemble V\textsuperscript{(M)} f. par eux donnez a monseigneur, la somme de deux mille livres tournois, laquelle lessidtes gens des trois estaz nous ont donnee et ordonne a nous estre payee, baillée et delivree par ledit receveur des deniers de sa recepte par eux mis sus contre le principal pour les causes et ainsi qu'il est plus aplain contenu et declaré es instrusions et ordonnances faittes sur le fait dudit aide, de laquelle somme de II\textsuperscript{(M)} l. t. nous tenons pour bien contente et payee, et en quittons ledit receveur et tous autres a qui quittance en appartient. Donné en \textit{testimony} de ce sousz nostre sceau, le XXIII\textsuperscript{(e)} jour de fevrier, l'an mil quatre cens trente et six. Byart.}
383
384 Par madame la duchesse,
385
386 \textit{(Signé :.) Byart\footnote{Mention dorsale (les premiers mots de chaque ligne sont coupés, ils se trouvaient sur la queue de parchemin) : \textit{(dots : Blant pour ?) faire quittance pour madame la (dots : duchesse de Bourgogne et d'Auvergne de la somme (dots : de II\textsuperscript{(M)}) livres tournois a elle donnee par les (dots : trois) estas dudit pais d'Auvergne (dots : a l'assemblée par eux faite a Clermont le XXII\textsuperscript{(e)} jour (de dots , l'an) mil CCCC trente et six. Byart.)}}}
387
388
389
390
391 \medskip
392
393
394
395
396
397

```

FIGURE 6.1 – Fichier tex des actes d'Agnès de Bourgogne.

2. *Ibid.*

L’objectif est d’assembler à nouveau tous les fichiers XML individuels obtenus pour chaque acte dans un seul fichier XML. La transformation de ce fichier en LaTeX permettra sa compilation et donc sa conversion et lecture au format PDF. Cela est réalisé grâce à une fonction Python³ qui transforme la langage XML en langage tex via XSLT. XSLT est le langage conçu pour transformer des documents XML en d’autres documents selon les spécifications XSL. XSL (eXtensible Stylesheet Language) désigne les spécifications pour écrire des feuilles de style, soit de quelle manière doivent être transformés les documents XML. Une fois le document correctement codé, une étape de compilation permet d’en produire un rendu final, ici un PDF mis en page.

monseigneur que ce ne vous puise trouver a prejudice au temps a venir. Tres chiers et bien amés,
nostre seigneur vous ait en sa garde. Escript a Molins, le III^e jour d'aoust. (*Sigé :*) de Bar.

14251125a

[1425-1434], 25 novembre – Moulins

Agnès de Bourgogne, comtesse de Clermont, mande les gens de Riom d'envoyer un représentant à l'assemblée des Etats qui aura lieu le 8 décembre suivant à Issoire, en présence de Pierre de Chantelle et Pierre de Carnoy, conseillers du duc.

A. Original sur papier, signé, 285 x 170 mm. Riom, Archives municipales, AA 16, n° 511.

(Au verso) A nos tres chiers et bons amis les consulz, bourgeois et habitans de Rion (Au recto) La contessez de Clermont. Tres chiers et bons amis, pour vous dire aucunes choses touchans lez grant bien et profouill de monseigneur le roy et de monseigneur et de ses pais et seigneours, et mesmeun de la chose publicke, nous vous prions et requeres tres a certes et tant que pouvons que veilliez envoyer aucuns de vous le VIII^e jour du mois de decembre prochain venant au lieu d'Ussore, avec les autres gens des trois estats qui semblaiblement y sont mandez pour beugnages et concilure sur ledictes choses que vous seront exposees par noe amez et feaux conseillers de monseigneur et de nous, maistres Pierre de Chantelle et Pierre de Carnonne, lesquelz pour coste cause envoynys apres de la. Tres chiers et bons amis, nostre seigneur vostre ait en sa garde. Escrit a Molins, le XXV^e jour de novembre. (Signé) Gourriet.

14350322a

[1434-1449], 22 mars - Moulins

[1434-1445], 22 mars — Moudras

A. Original sur papier, signé. 220 x 270 mm. Lyon, Archives municipales, AA 25, cote 5.

(Au verso) A nos très chiers et bien aimés les bourgeois et habitants de Lyon. *(Au recto)* La duchesse de Bourbonniès et d'Auvergne. Très chiers et bien aimés, présentement nous envoyons par devers vous nos ames et faveurs chevaleres et conseillers de monseigneur et de nous, les seigneurs de l'Espinace et de la Buxière, pour vous dire et exposer certaines choses touchans le fait et la fin de monsieur seigneur. Si vous prirez les plus adertes que fare pouva nos nodis conseillere veuleables fealmente croire de ce qu'ils vous diront de par nous, et adjuster plaine fay à leur rapport comme a nous mesme, et sur ce tant que mandez seigneur et nous vous en sonons as toujours tenu. Si ne nous y voullez fallir ny escouderne sur tout le plaisir que fare nous desiriez, et sur ce nous pourrons tenre bonne volonté, et ensemble s'autence chose voulés que par vous faire puissions pour le bon de baron. Très chiers et bien aimés, le saint espirit ait en saut une bénédiction sur le bataille de BONNEMIRE.

¹ Agnès est duchesse à partir de 1434, et le signataire de la lettre, Laurent Andraut, décède en 1449 (voir l'acte du duc Charles Ier de Bourbon du 26 février 1449).

14350621a

[1434-1456], 21 juin – Sury-le-Bois

Agnès, duchesse de Bourbnois et d'Auvergne, informe les habitants de Lyon de l'existence d'une ordonnance empêchant la sortie de vivres du duché de Bourbon, au nom de laquelle a été saisis un chepTEL empêché par des bouchers et marchands de ladite ville à une foire; néanmoins, elle accepte de le libérer, tout en priant les Lyonnais de la prévenir lorsqu'ils auront besoin de faire venir des vivres du duché.

A. Original sur papier, signé, 200 x 240 mm. Lyon, Archives municipales, AA 25, cote 6.

^(a) Le recours au subjonctif *soit* nécessite l'ajout d'une conjonction (*que*), oubliée par le scribe. Sans la conjonction, la phrase aurait dû être *pour ce que nous voulions nostreidz soit et ordonneance entre entretenu*.

14370224a

1437 (n. st.) 24 février

Quittance d'Agnès, duchesse de Bourbonnais et d'Auvergne, à Pierre Mandonier, receveur du bas pays d'Auvergne d'une aide de huit mille francs, pour la somme de deux mille livres tournois.

A. Original sur parchemin, signé et jadis scellé "en cire rouge sur queue de parchemin"². 295 x 95-135 mm. Bibliothèque nationale de France, ms. fr. 20389, cote 89. B. Copie moderne sur papier par Gaignières, incomplète. 235 x 375 mm. *Ibid.*, ms. fr. 20388, folio 153.

2. Précision de Gaignières (B), qui ajoute : "le sceau est cassé".

FIGURE 6.2 – Fichier PDF des actes d’Agnès de Bourgogne issu de la compilation du fichier tex

Les éditions une fois compilées en PDF rassemblent l'intégralité des actes pour chaque duc et duchesse. Chaque acte se distingue par son identifiant unique en gras et dans une taille plus importante. Ensuite, les dates de temps et de lieux apparaissent également en gras, séparées par une virgule. L'analyse en italique précède le tableau de la tradition, dont la numérotation des mentions est indiquée en gras. Puis vient le texte de l'acte, suivi des mentions hors teneurs indentées et du nom des signataires précédé de la mention « (Signé :) ». Selon les cas, les notes paléographiques sont situées après le texte, en toutes petites capitales, alors que les notes de bas de page apparaissent en fin de page.

³. GitLab ([py/compilation_latex.ipynb](#)).

La conversion du XML en tex permet la compilation puis une lecture agréable des éditions en PDF. L'ensemble de ces langages sont privilégiés par les travaux de recherche dans la mesure où ils assurent l'intégrité des données dans le temps. LaTeX est un logiciel libre, gratuit, multiplateforme et donc compatible avec différents systèmes d'exploitation et sans risque d'obsolescence. De plus, le programme est robuste, stable, compilé puisqu'il se met à jour lors de chaque compilation⁴. L'ensemble des technologies utilisées pour la transformation des documents permet donc la génération d'édition standard et mises en forme selon les *Conseil pour l'édition de texte médiévaux* de l'École Nationale des Chartes⁵. Toutefois, certaines erreurs de mise en page ont été repérées au sein des PDF.

4. *LaTeX*, Wikilivres, URL : <https://fr.wikibooks.org/wiki/LaTeX> (visité le 22/07/2023).

5. O. Guyotjeannin, *Conseils pour l'édition des textes médiévaux, Fascicule II : actes et documents d'archives...*

O. Guyotjeannin et F. Vielliard, *Conseils pour l'édition des textes médiévaux, Fascicule I : conseils généraux...*

6.2 Uniformisation des fichiers XML

Après relecture des documents PDF générés, des anomalies figurent dans l'édition des actes⁶. Elles sont visibles et repérables au sein des PDF. Une fois identifiées, nous avons constaté qu'elles étaient souvent dues à des erreurs dans l'encodage des documents XML et à des erreurs de mise en forme. Les corrections sont effectuées au sein des fichiers individuels dans la mesure où les fichiers regroupant tous les actes pour chaque duc et duchesse, permettant par la suite la transformation en tex puis la compilation en PDF, sont générés à partir des fichiers individuels. Le mode projet d'Oxygen (l'éditeur XML) a permis de corriger ces écarts sur l'ensemble des fichiers individuels pour chaque prince, et même sur l'ensemble des quatre corpus, dans la mesure où il permet de naviguer dans l'arborescence du dossier XML du projet.

Les erreurs individuelles, propres à chaque corpus, ont été révisées en première intention. Le corpus de Louis II présente le plus d'anomalies. L'une d'entre elles a été réglée directement via le script Python. Il s'agit de la suppression de la commande « textsuper-scriber » qui apparaissait dans la date de temps des actes datés du premier jour d'un mois. Le reste a été corrigé directement dans les fichiers XML. Premièrement, des regex ont été utilisées pour pallier les problèmes d'encodage. Les balises `<hi>` ne sont pas toujours encodées correctement. Certaines ne contiennent que des points ou une parenthèse. La mention « signé » n'est pas continuellement insérée dans des balises `<hi>`, et est parfois fractionné. De plus, elle est parfois présente en double dans le document PDF, et n'est pas toujours indentée. Une centaine d'actes sont affectés par ces doubles mentions. En effet, elles sont ajoutées par XSLT, d'où la nécessité de les supprimer directement des XML via une regex. Les expressions régulières ont permis de remplacer les éléments (balises et leur contenu) incomplets par des éléments normalisés. Par ailleurs, une partie des défauts relevés dans les PDF proviennent des CSV. En effet, c'est par la ponction d'informations dans ces fichiers et leur ajout au XML que des éléments déficients se sont retrouvés par la suite dans le PDF. L'attribut de la balise `<subtype>` « Traité-Accord » qui indique le type diplomatique de l'acte et atteste d'un accord passé entre plusieurs princes correspond en fait à des lettres patentes⁷. Cet élément n'avait pas été revu dans le CSV des actes de Louis II et a ainsi été ajouté comme attribut de la première balise `<div>` dans les fichiers XML. À l'instar des XML, les fichiers CSV ont également bénéficié d'une révision, en supprimant et en remplaçant les mentions défaillantes par les mentions correctes via des regex.

6. GitLab ([medieval-acts/princely-acts/stage_actes_princiers#7](https://gitlab.com/medieval-acts/princely-acts/stage_actes_princiers#7)).

7. AN, P 1357², n°425 (13 août 1377). - AN, P 1370², c. 1920 (21 avril 1386).

Deuxièmement, il est possible de corriger certains éléments en utilisant le Xpath, le chemin qui permet de naviguer au sein de l’arbre XML. Des points d’interrogations, qui marquaient des incertitudes sur les types diplomatiques dans les CSV, ont été relevés au sein des balises <subtype> et <type>, puis éliminés. Troisièmement, les erreurs singulières ont été corrigées manuellement. C’est le cas d’une date de temps qui était doublée (acte daté du vingt-deux septembre puis signé le vingt-sept septembre) et encodée dans la date de lieu⁸. Une référence a aussi été trouvée dans une date de lieu. Le corpus d’Anne Dauphine ne comporte quant à lui qu’une date de temps encodée dans une date de lieu, qui a également été retirée manuellement.

Le corpus de Charles I^{er} comporte une erreur majeure due au remplissage du CSV, qu’il est possible de corriger via une regex. Le type de  nomatique n’est pas normalisé : on retrouve la mention « dep » au lieu de « **deperditum** ». Les expressions régulières ont permis le remplacement instantané des cent trente-trois occurrences à la fois dans le CSV et dans les fichiers XML. Le corpus d’Agnès de Bourgogne n’est pas non plus exempt d’erreurs d’encodage. En effet, la plupart des dates de lieu des actes sont encodées dans la balise <date>. Ainsi, seules des mentions « NS » sont encodées dans la balise <placeName>. En conséquence, lors de la transformation, les mentions « NS » ne sont pas interprétées et la date de temps et la date de lieux ne sont donc pas séparées par des tirets dans le PDF puisqu’elles sont encodées dans la même balise. Cela a été corrigé manuellement dans la mesure où les actes concernés n’étaient pas trop nombreux.

Les problèmes de mise en page généraux, qui concernent les quatre corpus, sont résolus à la fin grâce à l’emploi de regex. C’est le cas des lignes vides entre la balise fermante <listwist> et la <div> suivante, ainsi que des nombreux retours à la ligne après et avant les notes, qui ont été éliminés. Parfois, la numérotation des états du texte est signalée en gras : « **A.** » de manière complètement aléatoire. Or, les numéros (A, B, C) d’exemplaires doivent tous être en gras. Une regex a permis de les attraper et de les styler uniformément. Des opérations d’harmonisation des fichiers XML ont été décidées, comme l’homogénéisation des dénominations des services d’archives. Par exemple, « Archives municipales de Montluçon » devient « Montluçon, Archives municipales ». La « Bibliothèque municipale de Lyon » est remplacée par « Lyon, Bibliothèque municipale ». Les « Archives départementales de la Loire » deviennent « Archives départementales Loire ». Pour les Archives nationales, on indique « Paris, Archives nationales ». Cet alignement nécessite une réflexion conjointe des chercheurs travaillant sur le projet en amont, afin de se mettre d’accord sur des standards, et de les communiquer ensuite afin qu’ils soient respectés lors des prochains ajouts au projet.

8. AN, P 1357², n°442 (22 et 27 septembre 1399).

Les dénominations des lieux bénéficient également d'une harmonisation. Certaines dates de lieu comportent des parenthèses dont l'usage n'est pas justifié : Bourbon(-l'Archambault), Clermont (en Auvergne). Des actes comportent parfois la mention « (château de) » ou « au château » après la date de lieu. Nous avons convenu que les toponymes seraient retranscrits en entier sans parenthèses. Toutefois, lorsqu'il y a une précision en plus de la ville, celle-ci doit figurer entre parenthèses : Moulins (Château), Paris (Hôtel de Bourbon), Riom (Palais). Certaines étapes plus précises n'ont pu être effectuées automatiquement. Lorsqu'il y a au sein d'une édition une référence à un acte qui était auparavant numéroté (ex : cf. acte n°XX), cette référence est remplacée par l'identifiant du fichier XML de l'acte en question. Les actes sans mesures du corpus de Louis II ont également été repérés en vue de compléter les balises <witness> dans la mesure du possible.

L'emploi de regex et le recours au Xpath ont permis d'automatiser un certain nombre d'opérations visant à améliorer et à homogénéiser le plus possible l'encodage des actes. Toutefois, des opérations n'ont pu être automatisées et certaines choses n'ont pas pu être corrigées en raison de leur complexité. C'est le cas de la balise <witness> qui n'est en définitive pas aussi structurée que sur le schéma théorique comme balise parent des balises <text>, <orgname>, <idno1>, <idno2>⁹. Les mentions hors teneur ne sont pas dotées non plus d'un attribut « subtype » indiquant leur positionnement. Néanmoins, les étapes précédentes tendent à rendre la structuration des fichiers XML la plus uniforme possible pour leur mise en ligne.

9. Figure 5.3 – Schéma d'encodage des actes en XML.

6.3 Encodage d'actes inédits

Après avoir détaillé la chaîne de traitement des données, se pose la question des actes inédits. Au vu de la dispersion du corpus, nous ne sommes pas à l'abri de découvrir des actes passés au nom des ducs de Bourbon qui auraient échappé à notre recensement. Comment intégrer de nouveaux actes au corpus ? Une partie de la réponse à cette question a été développée en amont. En effet, le schéma d'encodage précis établi précédemment détaille la procédure à suivre pour encoder de nouvelles données afin qu'elles soient conformes au corpus¹⁰. Nous avons expérimenté ce procédé avec l'intégration des actes de Charles I^{er} repérés aux archives¹¹. Dans un premier temps, le travail a lieu sur place. Il s'agit de photographier et de mesurer les actes en question. Les prises de vues permettront par la suite d'analyser les actes et d'alimenter le site web. Cinq actes ont été repérés dans le registre P 452², pour Charles I^{er}. Les séries P 452 et 453 renferment les documents relatifs au bourbonnais, plus précisément les anciens hommages, effectués entre 1355 et 1530¹². Les actes concernés sont respectivement datés du vingt-quatre octobre 1443 pour trois d'entre eux, du 25 octobre 1443 et du 22 octobre 1454¹³. Les cinq actes attestent de l'hommage d'individus du duché pour leurs possessions dans celui-ci : Jean Chauvet, Jean de Meschatain, Pierre de Saint Aubin, Jean de Bar et Jean de la Porte.

10. CANEVAS TEI, dans Chapitre 5, 5.3. Séparation des fichiers XML de corpus en fichiers XML d'actes

11. AN, P 4522, c. 190 (24 octobre 1443) - AN, P 4522, c. 307 (24 octobre 1443) - AN, P 4522, c. 313 (24 octobre 1443) - AN, P 4522, c. 229 (25 octobre 1443) - AN, P 4522, c. 169 (22 octobre 1454), dans chapitre 2, 2.3 Le traitement de texte : un bon compromis ?

12. Archives Nationales, « Série P Chambre des comptes et comptabilité », en ligne : <http://www.archivesnationales.culture.gouv.fr/cham/cham/fonds/EGF/SA/SAPDF/Egfn-p.pdf>.

13. Annexe Actes inédits.

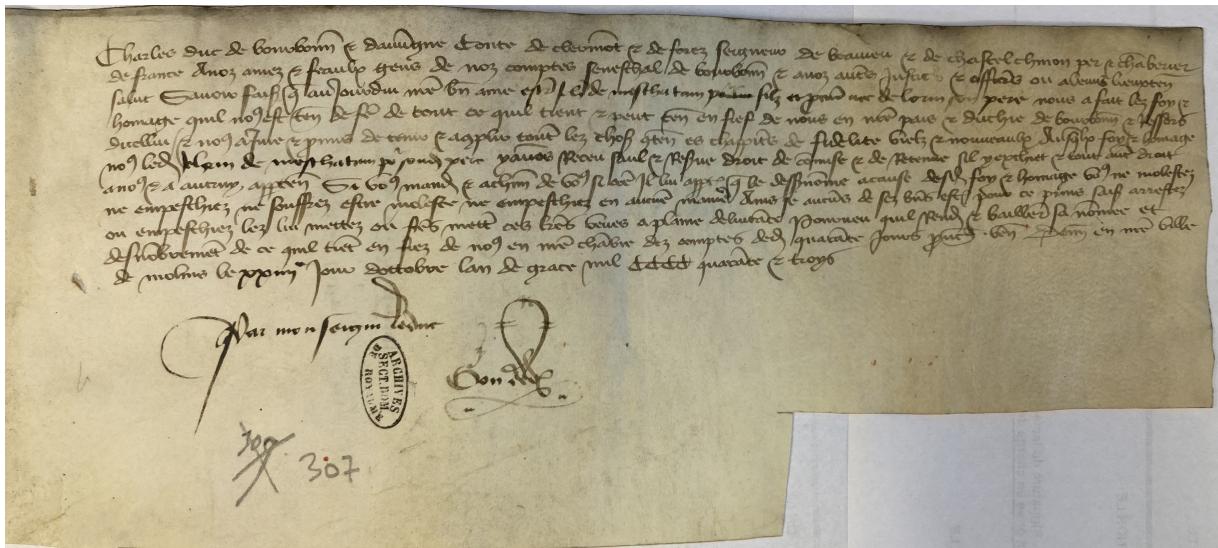


FIGURE 6.3 – Photographie de l'acte daté du 24 octobre 1443.

Charles, duc de Bourgogne et d'Auvergne, etc., notifie aux gens de ses comptes qu'il a reçu l'hommage de Jean de Meschatain pour tout ce que celui-ci tient de lui en fief dans le duché de Bourgogne.

À partir des photographies prises, les actes sont transcrits selon les normes développées dans les *Conseils pour l'édition des textes médiévaux*¹⁴. La transcription donne accès au contenu du texte et à des informations essentielles comme la datation, le lieu de rédaction, ou encore des noms d'individus, qui vont permettre de compléter le schéma XML. La deuxième étape consiste à encoder l'acte. La transcription peut avoir lieu directement à cette étape si le texte n'est pas trop difficile à lire¹⁵. Le TEI Header est complété avec des informations d'identification de l'acte. La date est nécessaire pour remplir l'attribut « xml :id » de la balise <TEI> ainsi que le titre de niveau « a ». Ensuite, c'est tout le contenu de la balise <sourceDesc>, décrivant le document, soit l'acte concerné, qui est à remplir avec les identifiants de l'acte : institution de conservation, fonds, cote ainsi que sa date de rédaction. Puis, viennent la liste des personnes mentionnées : prince à l'origine de l'acte et signataire. Enfin, le <sourceDesc> est complété par le tableau de la tradition. Le <profileDesc> contient les informations contextuelles, soit un résumé en quelques mots de l'acte en question.

14. O. Guyotjeannin, *Conseils pour l'édition des textes médiévaux, Fascicule II : actes et documents d'archives...*

15. Encodage des actes concernés : Gitlab
(xml/bourbon/charles_i_bourbon/brb_ch_i_1443_10_24a.xml),
(xml/bourbon/charles_i_bourbon/brb_ch_i_1443_10_24b.xml),
(xml/bourbon/charles_i_bourbon/brb_ch_i_1443_10_24c.xml),
(xml/bourbon/charles_i_bourbon/brb_ch_i_1443_10_25a.xml),
(xml/bourbon/charles_i_bourbon/brb_ch_i_1454_10_22a.xml).

Une fois le TEI Header complété, c'est le <text> avec les attributs de la première <div> : identifiants, état, type diplomatique. Les balises enfant de la balise <docDate> sont enrichies des dates de temps et de lieu. Ensuite, l'analyse précède le texte transcrit. Le texte de l'acte est retranscrit tel quel et les lettres et les graphies manquantes ou erronées sont déclarées au sein de l'apparat critique. Des balises signalent les personnes et les lieux du texte. Ces deux derniers éléments, ainsi que le signataire, qui clôture le texte, sont référencés.

```

68 <text>
69   <body>
70     <div n="14431024b" subtype="Lettre_close" type="Original" xml:id="ch_i_1443_10_24b">
71       <div type="front">
72         <date n="a" when="1443-10-24">1443, 24 octobre</date>
73         <placeName ref="#Moulins" type="production_place">Moulins</placeName>
74       </docDate><argument>
75         <p>Charles, duc de Bourbonnais et d'Auvergne, etc., notifie aux gens de ses comptes qu'il a reçu l'hommage de Jean de Meschatain pour tout ce que celui-ci tient de lui en
76 fief dans le duché de Bourbonnais.</p>
77       </argument></div>
78     <div type="acte">
79       <p>Charles, duc de Bourbonnais et d'Auvergne, comte de Clermont et de Forez, seigneur de Beaujeu et de Chastel Chinon, per et chamberier de France, a nos amés et feaux gens de
80 noz comptes, seneschal de Bourbonnais et a noz autres justiciers et officiers ou a leurs lieutenants, salut.</p>
81       <p>Savoir faisons que aujourd'hui notre bien amé escuyer <persName ref="#deMeschatain1">Jehan de Meschatain</persName>, filz et premier né de Lorin, son père, nous a fait les
82 foy et homage qu'il nous est tenu de faire de tout ce qu'il tient et peut tenir en fief de nous en notre pais et <placeName ref="#Bourgogne1">duché de Bourbonnais</placeName> et resort
83 d'icellui, et nous a juré et promis de tenir et accomplir toutes lez choses contenues es chapitres de fidélité vielz et nouveaux, ausquelz foy et homage nous ledit <persName
84 ref="#deMeschatain2">Jehan de Meschatain</persName> pour dit pere y avons receu saul et réservé droit de commise et de retenus s'il y exchiet et tout autre droit a nous et autrui
85 appartenent.</p>
86       <p>Si vous mandons et a chacun de vous si come il lui appartient que le dessus nommé a cause desdits foy et homage, vous ne molestez, ne empeschez, ne souffrez estre molester
87 ne empeschez en aucune maniere, ains se aucun des ses biens estoient pour ce prins, saisi, arrestez ou empeschez lez lui mectez ou faictes mettre ces lettres veues a plaine delivrance
88 pourveu qu'il rendra et baillera sa nomine et desnombrément de ce qu'il tient en fief de nous en notre chambre des comptes dedans quarante jours prochains venans.</p>
89       <p>Donné en nostre ville de <placeName ref="#Moulins">Moulins</placeName> le XXIIII<sup>e</sup> jour d'octobre l'an de grace mil CCCC quarante et troys.</p>
90     </div>
91   </body>
</text>
```

FIGURE 6.4 – Encodage TEI du texte de l'acte daté du 24 octobre 1443.

Toutes les étapes préalables à la mise en ligne des données détaillées dans les sections précédentes sont répliquables, ce qui rend possible l'élargissement du corpus lorsque de nouveaux actes de même nature sont découverts. En effet, il est possible de relancer les scripts afin d'assurer la transformation en LaTeX et la compilation en PDF du corpus, une fois les nouveaux actes encodés en XML, et ainsi les ajouter aux éléments déjà mis en ligne. Toutefois, ce procédé fonctionne et représente un gain de temps pour des corpus édités sur du traitement de texte et dont les inédits ne sont pas trop nombreux à ajouter. Afin de traiter des corpus qui n'auraient pas été édités, il serait plus pertinent de fragmenter le texte des actes et les informations le concernant, nécessaires à l'élaboration d'une édition, pour les intégrer à une base de données, qui, parsée par un script Python, pourrait permettre de générer un encodage en XML¹⁶.

16. Cette méthode est développé dans : Sébastien de Valeriola, « L'ordinateur au service du dépouillement de sources historiques. Éléments d'analyse semi-automatique d'un corpus diplomatique homogène », *Histoire & mesure*, XXXV-2 (2020), p. 171-196, DOI : 10.4000/histoiremesure.13534.

Conclusion

La mobilisation de techniques numériques a permis de faciliter l'étude des corpus conséquents et dispersés que sont ceux de Louis II, d'Anne Dauphine, de Charles I^{er} et d'Agnès de Bourgogne.

Le présent mémoire s'est attaché à présenter les documents étudiés : datation, typologie, dispersion, similitudes des actes des ducs de Bourbon. Nous sommes également revenus sur la manière dont ces derniers ont été traités initialement, car le projet d'encodage des actes pour la mise en ligne ne part pas de rien. En effet, les actes ont d'abord été édités avec un logiciel de traitement de texte. Nous avons alors développé une chaîne de traitement des documents à partir des éditions utilisées comme *legacy data*. Ensuite, nous avons travaillé sur les documents afin de les préparer à l'encodage en XML/TEI. Cette étape s'est articulée autour d'une proposition de méthode de stylage des éditions en vue d'une conversion, puis du choix d'un convertisseur approprié. Puis, après conversion des documents, plusieurs scripts Python ont facilité les interventions sur les éditions de façon à améliorer et affiner leur structuration. Ces étapes ont également permis d'extraire et d'ajouter des données. Enfin, les fichiers ont été préparés à la mise en ligne via la génération d'un fichier XML pour chaque acte, puis une compilation en PDF (via une transformation en LaTeX). Ces dernières étapes ont rendu visibles les anomalies et possibles les corrections sur le document et initié la réflexion sur le projet (intégration d'actes inédits ...).

Toutes les opérations de traitement des actes explicitées dans ce travail peuvent être qualifiées de semi-automatiques, dans la mesure où il y a une alternance entre des traitements automatiques et des traitements manuels. Sébastien de Valeriola insiste sur la complémentarité de ces deux méthodes en affirmant que « le travail de la machine ne remplace aucunement celui de l'historien, mais l'accompagne et le facilite »¹⁷. L'enjeu de ce travail sur le projet « Actes Princiers » était d'automatiser ce qui pouvait l'être. Comme nous portions d'éditions réalisées avec un logiciel de traitement de texte, de nombreuses opérations ont été semi-automatiques, comme le stylage des actes via des regex en première intention et manuel pour les cas les plus compliqués. L'ordinateur est un bon outil

17. *Ibid.*

pour réaliser les traitements les plus volumineux comme la transformation en XML et la séparation des corpus de chaque prince en fichiers individuels. Cette automatisation évite les erreurs humaines d'encodage dans notre cas. Mais, il est nécessaire de contrôler l'issue de ces manipulations, car il est fréquent que la machine bute sur des erreurs antérieures. L'automatisation permet également de rendre applicable le processus de traitement des données à l'avenir à d'autres actes et éventuels corpus. Toutefois, malgré sa réplicabilité, l'ensemble d'opérations développé est adapté à un corpus relativement homogène. Il ne faut pas perdre de vue qu'il est difficile de proposer une méthodologie cohérente, car même en suivant un modèle d'édition précis et reconnu, on ne peut pas prévenir toutes les individualités inhérentes à la recherche. L'enjeu reste donc de tendre au plus uniforme possible.

L'uniformité, lorsqu'elle est atteinte, ou du moins approchée au maximum, entraîne une automatisation accrue des procédés de traitement des données et facilite ainsi l'extraction d'un maximum d'informations des corpus. Dans ce sens, la méthode de stylage des documents pourrait être encore affinée et couplée à de la diplomatique en distinguant les différentes parties de l'acte (intitulation ...), afin d'étudier leurs caractéristiques via des méthodes quantitatives et ainsi de comparer les formules utilisées, la rédaction et l'utilisation de modèles dans certains cas ou encore l'évolution des pratiques d'écriture. Des études prosopographies pourraient également être développées à partir des personnes référencées dans chaque corpus et ainsi mettre en lumière des réseaux sociaux. En effet, les actes sont riches en renseignement sur les acteurs dans la mesure où ils mentionnent souvent une activité, un lieu d'habitation, parfois des relations. Plus largement, le site pourrait être enrichi d'autres corpus. Dans un premier temps, ceux de Jean I^{er} et de Marie de Berry pourraient bénéficier d'un travail dans la mesure où ils n'ont pas encore été très approfondis et où il serait pertinent de faire le lien entre les deux périodes de ducat de Louis II et de Charles I^{er}. Cela permettrait aussi d'étudier l'administration du duché pendant la période particulière que fut celle de la captivité du duc, en pleine guerre de Cent Ans.

Le projet « Actes Princiers » et les travaux menés dessus donnent l'impulsion aux recherches sur les actes des princes du royaume de France. Enfin, ce type de projet rend les données disponibles sur le web et accessibles à la communauté scientifique en proposant des *legacy data* homogènes et facilement exploitables et contribue ainsi à une meilleure connaissance de la diplomatique princière pour la période étudiée.

Annexes

Chronologie des éditions

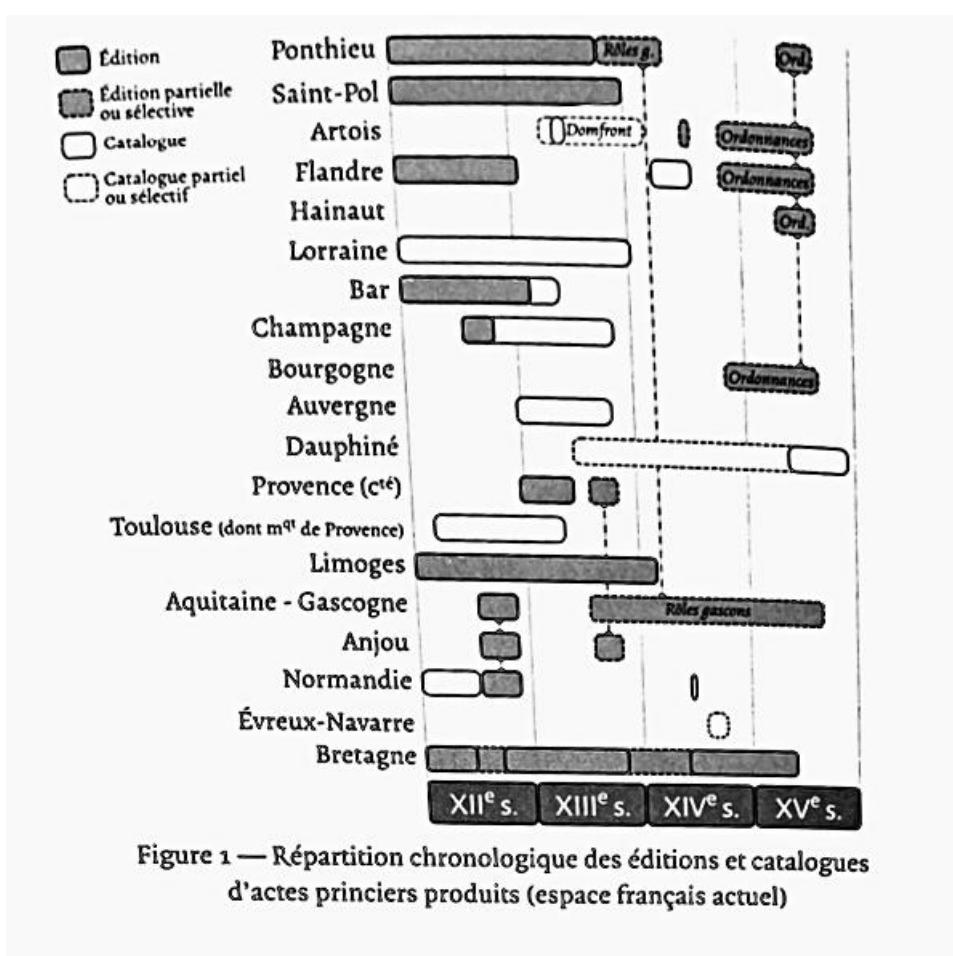


Figure 1 — Répartition chronologique des éditions et catalogues d'actes princiers produits (espace français actuel)

Canteaut, Olivier, Moufflet, Jean-François, « Les éditions d'actes princiers (XII^e - XV^e siècle) : bilan à l'ère du numérique », in : Olivier Guyotjeannin et Olivier Mattéoni (éd.), Jean de Berry et l'écrit : Les pratiques documentaires d'un fils de roi de France, Éditions de la Sorbonne, École Nationale des Chartes, Paris, 2019 (Histoire ancienne et médiévale, 159).

Généalogie de la famille ducale de Bourbon

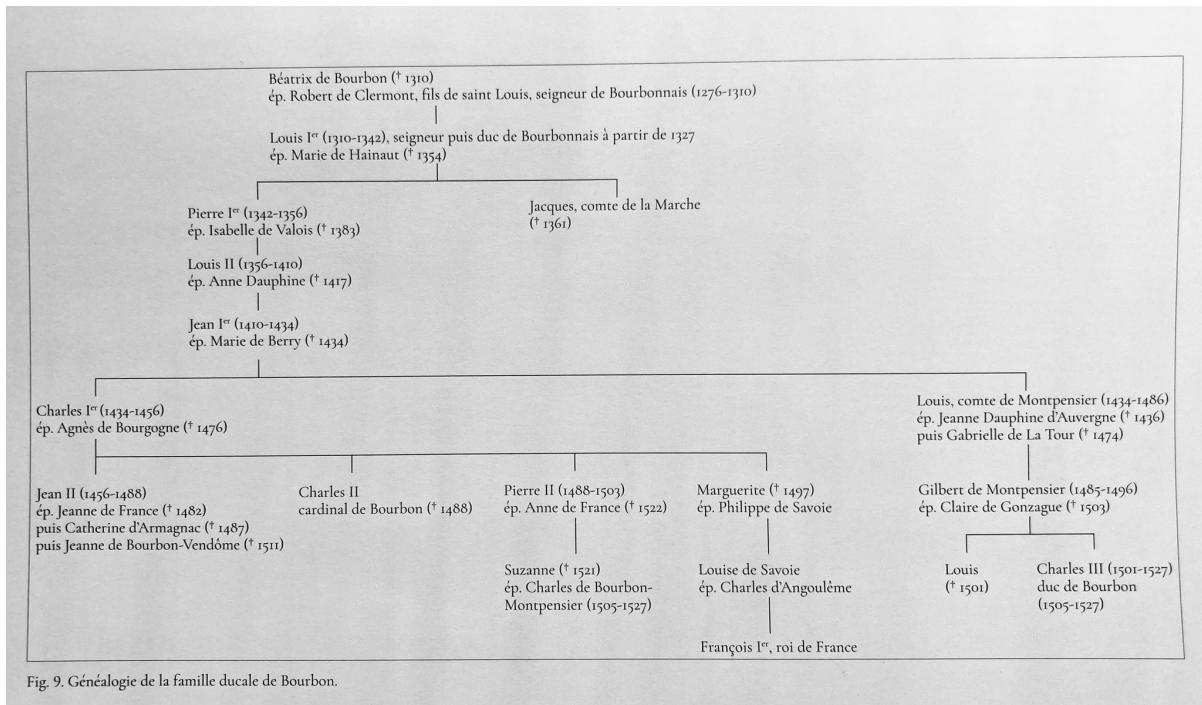
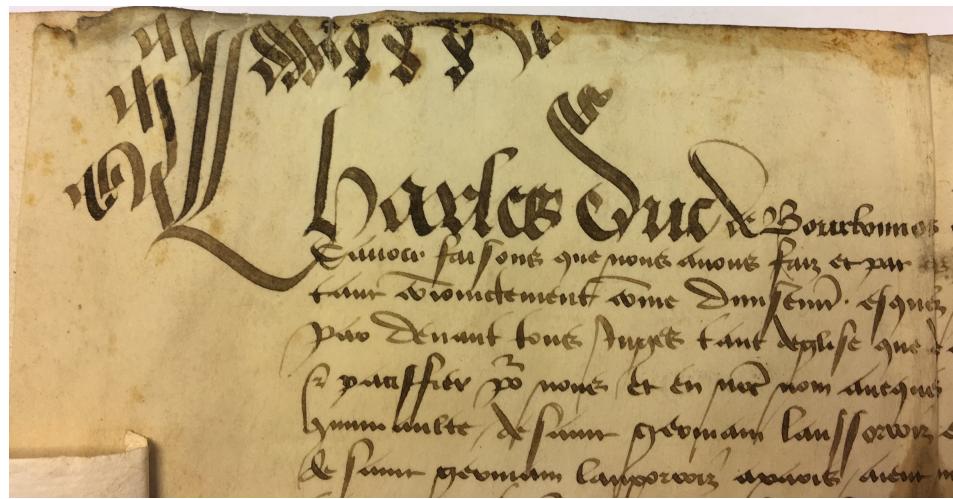


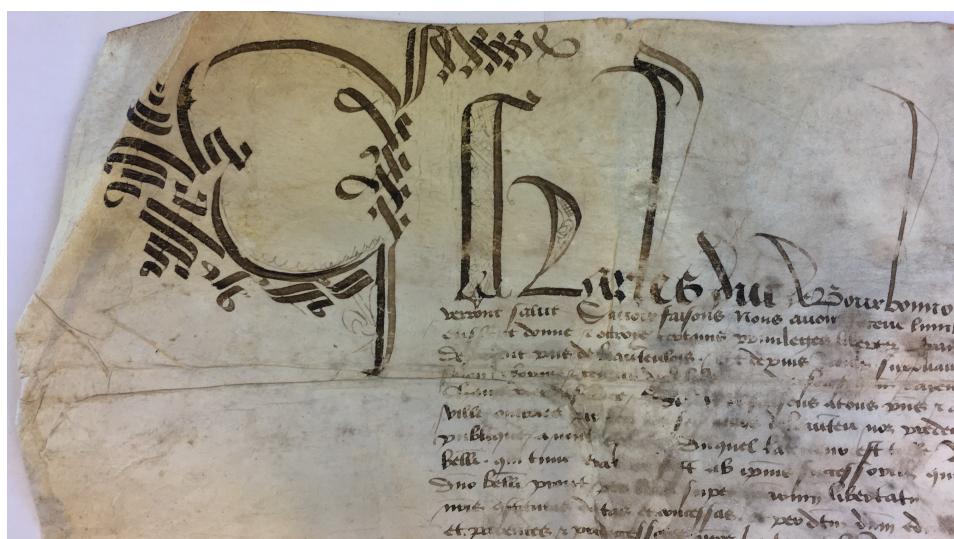
Fig. 9. Généalogie de la famille ducale de Bourbon.

Généalogie de la famille ducale de Bourbon, in MATTÉONI (Olivier), *Les Bourbons en leur bibliothèque (XIII^e-XVI^e siècle)*, Éditions de la Sorbonne, Paris, 2022 (Histoire ancienne et médiévale).

Initiales fleurdelisées



AN, P 13631, c. 11511 (18 décembre 1448).

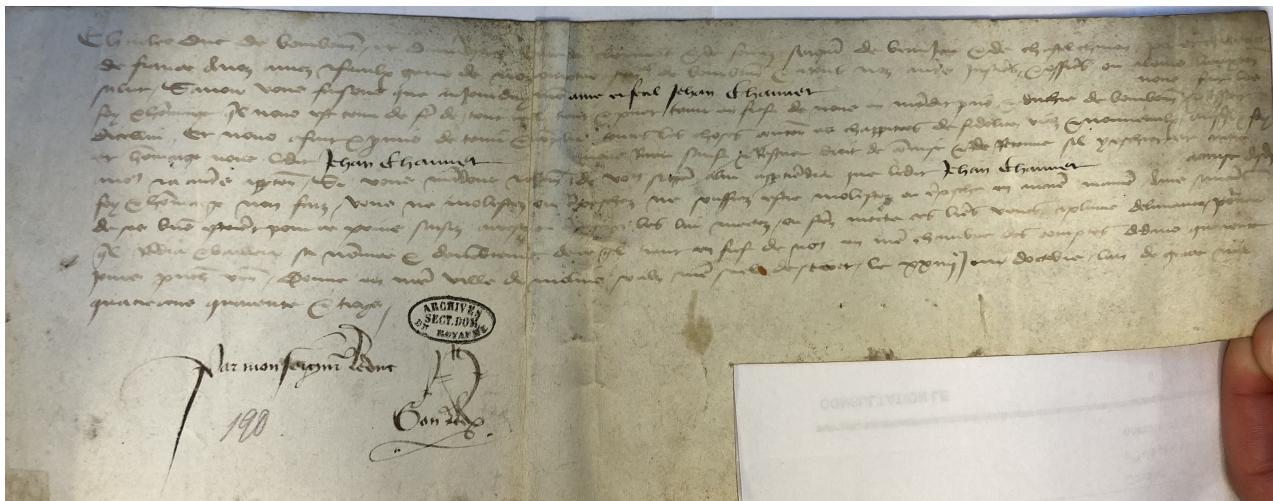


AN, P 13682, c. 1628 (19 février 1447).

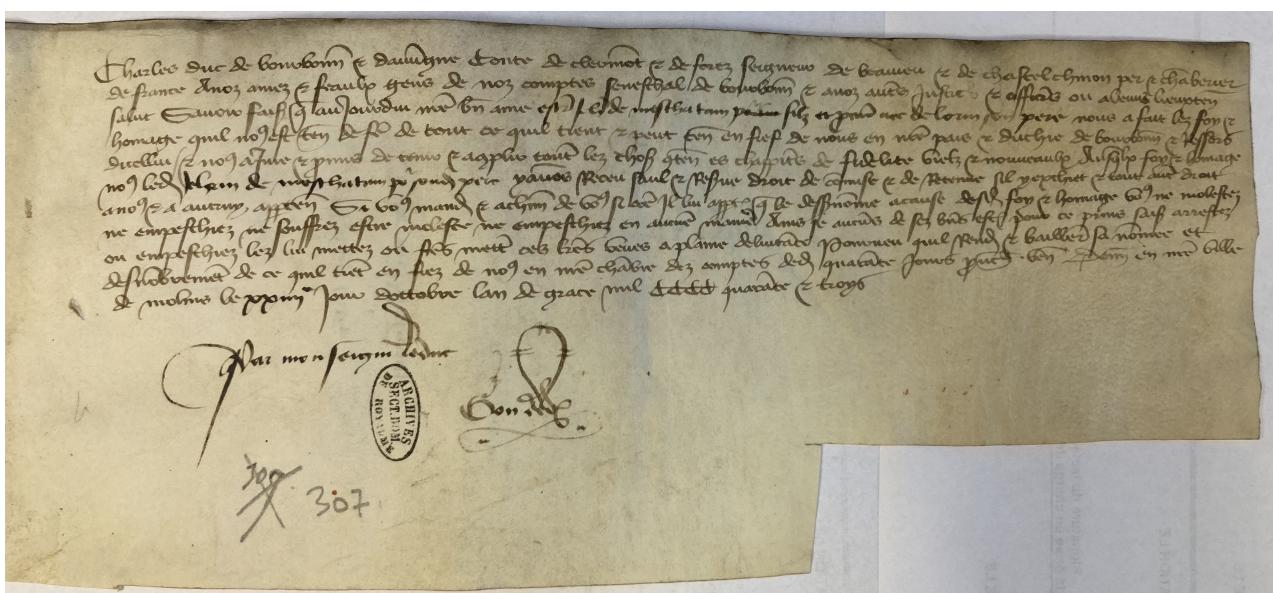
Liste des regex utilisées

Action	Regex	Remplacement
Retrait des doubles espaces	\s{2}	
Retrait des sauts de lignes	\n	
Retrait des sauts de pages	\r	
Retrait des mentions « Sur le repli »(en blanc)	^\(Sur le repli\) ([^Par].+)	\\$1
Stylage et indentation des notes paléographiques	— ([b-z]\.).	\n\$1
Suppression des astérisques	(*)+	
Stylage des dates	(^[\[\]\d\-\-] + (\((n\.\.st\.\.)?)?.+) ^(Avant)?\d{4}.+ ^(14..+) ^\\ [(14.+ Avant.+ Après.+ Entre.+) ^(14..+) ^\\ [(.+)	
Stylage des analyses	^?Louis, ? duc de Bourbonnais, ?.+ Louis, ? duc de Bourbonnais, comte de .+ ^Charles de Bourbon, comte de Clermont.+ ^Quittance de Charles.+ ^Lettre close.+ ^Marie de Berry, duchesse de Bourbonnais.+ ^Lettre de Charles.+ ^Charles, duc de Bourbonnais.+ ^Ordonnance de Charles.+ ^\\t[A-Z][a-z].+ ^Anne Dauphine, duchesse de Bourbonnais.+	
Stylage du tableau de la tradition	^A\\.\\. + ^B\\.\\.+	
Stylage des mentions	^Mention :.+ ^Analyse :.+ ^Indique :.+	
Stylage du titre des actes	^\(deperditum\) ^texte établi.+	
Stylage des mentions hors teneurs	(\Sur.+) \(Sur.+ \(Sous.+ \(A g.+ (Par madame la duchesse,)	\n\$1
Stylage des signataires	^\(Signé :\).+	

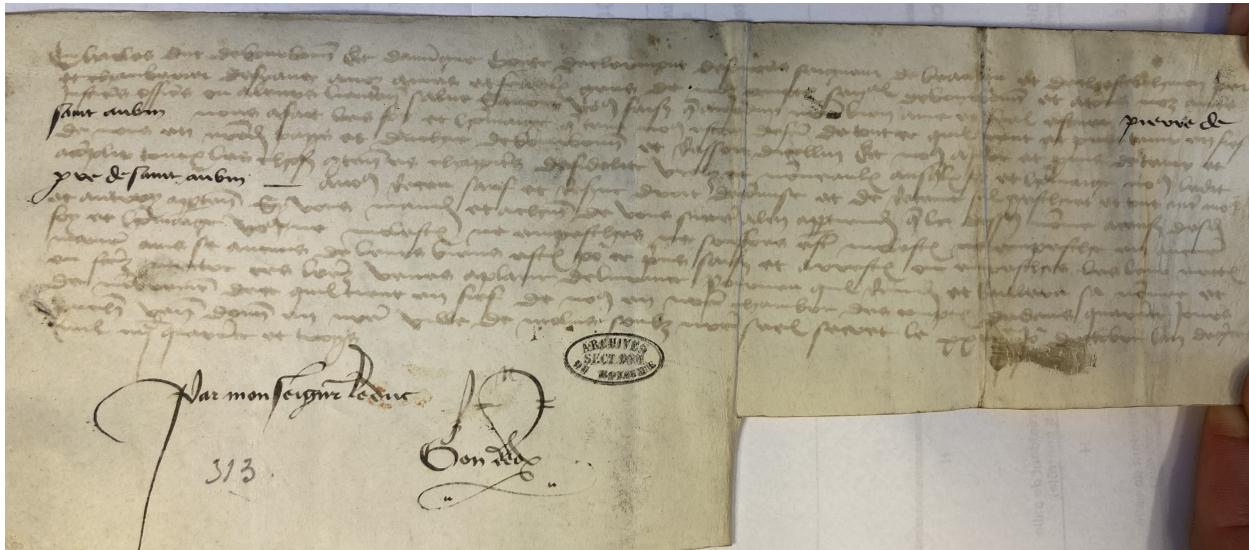
Actes inédits



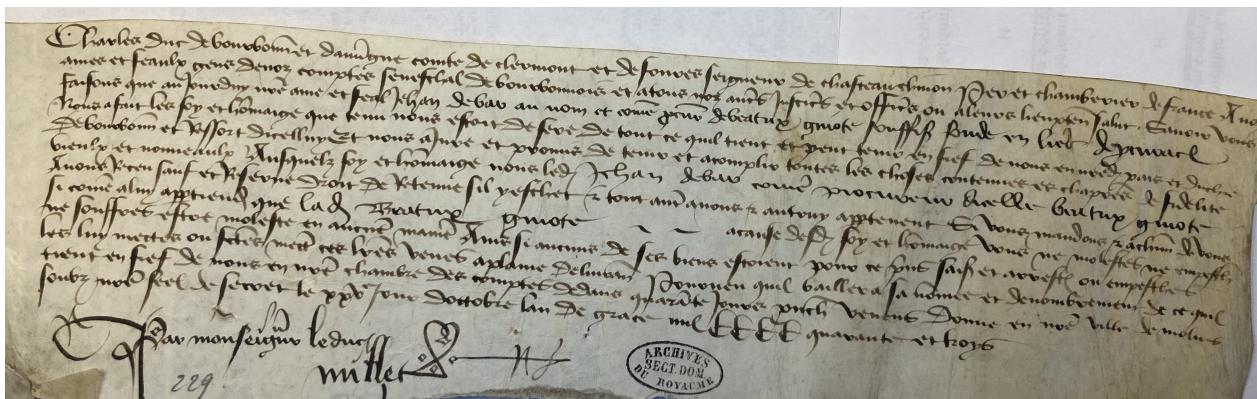
Acte daté du 24 octobre 1443.



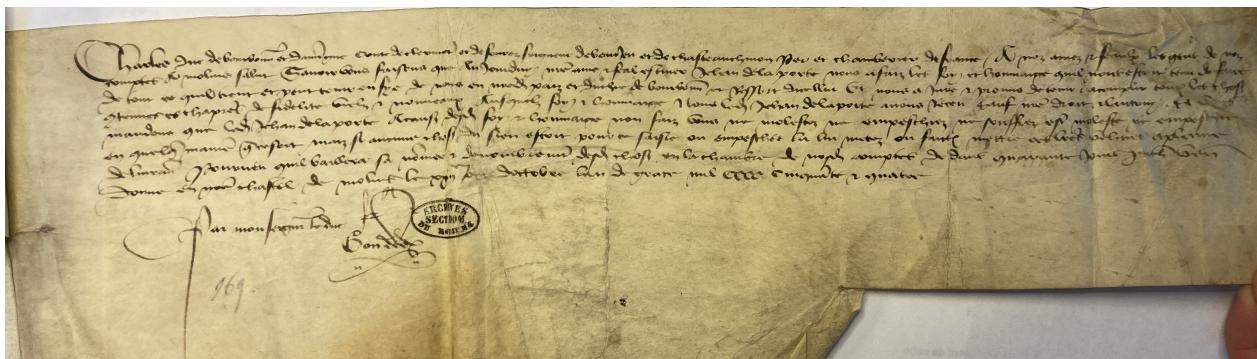
Acte daté du 24 octobre 1443.



Acte daté du 24 octobre 1443.



Acte daté du 25 octobre 1443.



Acte daté du 22 octobre 1454.

Table des matières

Résumé	i
Remerciements	iii
Liste des sigles et abréviations	v
Bibliographie	xii
Introduction	xiii
I L'édition numérique des corpus de Louis II, d'Anne Dau-	
phine et de Charles I^{er} et d'Agnès de Bourgogne	1
1 Présentation du corpus	3
1.1 Dispersion des actes	4
1.2 Répartition chronologique des actes	6
1.3 Typologie des actes	8
1.4 Pratiques d' <i>imitatio regis</i>	11
2 Éditions numériques	15
2.1 Trois projets d'édition	16
2.2 Des éditions diplomatiques	18
2.3 Le traitement de texte : un bon compromis ?	21
II Du traitement de texte à l'encodage	25
3 Préparation à l'encodage	27
3.1 Le recours à un langage de balisage	28
3.2 Utilisation de regex	31
3.3 Proposition de stylage	33

4 Conversion des fichiers ODT vers XML/TEI	39
4.1 Les convertisseurs libres	40
4.2 Essais de conversion avec Pandoc	42
4.3 Choix du convertisseur Teinte	44
III La chaîne de traitement des actes princiers	47
5 Structuration des XML	49
5.1 Extraction / Injection des données	50
5.2 Restructuration des documents XML	54
5.3 Séparation des fichiers	59
6 Préparation à la mise en ligne des données	65
6.1 Transformation et compilation des fichiers	66
6.2 Uniformisation des fichiers XML	69
6.3 Encodage d'actes inédits	72
Conclusion	77
Annexes	81
Chronologie des éditions	81
Généalogie de la famille ducale de Bourbon	82
Initiales fleurdelisées	83
Liste des regex utilisées	84
Actes inédits	85
Table des matières	87