

Présentation de l'outil *Transkribus* dans le cadre d'un projet d'édition de chartes médiévales

NICOLE BERGK PINTO

IRHT-CNRS, SECTION DE DIPLOMATIQUE

IRHT

Institut de recherche
et d'histoire des textes



Cadre de travail

- ❑ Institut de recherches et d'histoire des textes – CNRS, Section de diplomatique
- ❑ Projet ANR 2020-2024 : « BeCoRe – Between Composition and Reception : The Authority of Medieval Charters »
- ❑ Objet d'étude : actes médiévaux = documents juridiques dont la forme est relativement standardisée
- ❑ Objectif : étude comparative des signes textuels et graphiques de l'autorité et de la validité des actes médiévaux
- ❑ Étude menée sur base des données fournies par le portail de diffusion et d'édition d'actes [Monasterium.net](https://monasterium.net)
- ❑ Plus de 600 000 chartes, dont environ 270 000 disposant d'images des originaux, provenant de toute l'Europe

Exemplum

Et de ultimo... Domini... Benedicti...

monacho expressis professo. Hater... in firaade... humiliter... Salutem In Domino Religione zelus
 litarum scientia vite et morum sinceritas rerum providentia et negotiorum experientia aliaque probitate et virtutum merita
 quibus personam vestram fiduciosam relati accente nominis insignitiam nos merito iudicavit ut vos congrue honoribus antel
 lamus volentes vobis premissorum et aliorum meritorum et virtutum intuitu gratiam foree specialem vos in domum
 nostram seu prioratum nostrum conveniensem sancti panthani lethen in Regno anglie Evesham diocesi Qui de quibus filiabus nostri
 monasterii evasit Et qui a predicto nostro monasterio charen immediate dependet Et per monachos pnie monasterii de que
 numero et ordinem predictum expresse professus esse etiam consuevit Datum ad presentem in manibus nostris per obitum seu
 decessum sui vobis annuente quondam vltimū prioris et pssoris Cuiusque collatio promissio presentano et alia omnimoda dispositio
 quoniam locus vancatome occurrit ad nos et abbates de monasterio predicto et successores pro tempore existente pleno iure
 ab antiquo dignoscitur pertinere manemus in preceam omnimodam prioratus nostri predicti sancti panthani Et ipsam
 prioratum de presentem sic vancatem seu ale per quatuor aliam personam daret vnam pnie iuribus redditibus pncentibus
 fueribus et emolumentis ac pertinentiis suis vniuersis vobis vltimus conferimus et donamus et de pto prouidimus et prouidenda
 Et per traditionem presentium litarum iurisdictionis emestimus Ac in realem et actualem possessionem inducimus et pominu
 iurisdictionem vobis alienacionem rerum et iurium nobilium immobilium dicit nostri prioratus pter seu contra iuris formam et
 iuriam ac successorum nostrorum consensum patet et sensum Mandante propria fueribus nostre monachie et officariis
 predicti prioratus nostri ac etiam rebus monachie Reges Ecclesiarum vnciores pbes capellanos notarios tabelliones et alias
 personas publicas in iuris subsidiis regnum et regni quousque vos fuerim nicolam vel proxiem vestrum pro vobis in realem corpale
 et actualem predicti prioratus nostri possessionem iungere et pertinere eiusdem auctoritate nostra realiter ponant et iudicant
 nosq; induriam descendant facientes q; vos aut priorum vestrum pro vobis ad dictum prioratum recipi et admitti Amore ex
 nide quolibet illius detentore quem nos tenore presentium de iure possimus amouemus Ipse fuerit officarii et subditi ni
 dicit prioratus sancti panthani vobis tanq; vero et iudicato prior dicit prioratus pntant obdiant et officari iudicant Ac
 vobis auxilium et consilium presentem Atq; de omnibus rebus et pertinentiis eiusdem vobis vel priorum vestro pro vobis iureque
 respondeant et responderi faciant ut more est. Hanc rei confirmacionem sigillum nostrum literis purpureis duximus
 apponendum Datum in castro nostro de ludmo propolmaticum die Quarta mensis martii Anno domini a mille simo Quia
 Simeone sime Quadragesimo Quinque. more gallico sumpto. Nuntius venerabilibus et carissimis fratribus nostris domus
 vnderino de Campie prioris prioratus nostri de caruarvurione vncaten de rano de rano nostris de raru ordinis nostri amiaten
 et iohanne mathia puzpivo burgen clm vncaten de ranc matisten diocesi et fratribus ad premissa vocatis



Et de... Johannes... de Elmato... matistone... diocesi... publicus... imperialis et regna...
 Notum est... abbas... sigilla... in...
 Signis et subscripce... in...
 Sigillum circularis officii publici.

De... Anno...

Handwritten signature.

Cadre de travail

- IRHT : apport d'un corpus complémentaire, notamment le corpus des chartes originales de l'abbaye de Cluny (xii^e – xv^e siècles)
 - + ou - : 1200 documents
- Flow :
 - Originaux (BnF)
 - > reproductions numériques (Gallica)
 - > traitement philologique (transcriptions sur *Transkribus*)
 - > « édition » électronique (Monasterium.net)
 - > analyse graphique
 - > résultats de recherche

L'outil *Transkribus*

- Outil de transcription et de reconnaissance automatique de texte
 - Un parmi d'autres (eScriptorium)
- Importation d'images à travers le « protocole IIIf » (International Image Interoperability Framework) : ensemble de standards qui définissent un cadre d'interopérabilité pour les bibliothèques numériques
- Segmentation des images et définition de zones de texte, lignes et base d'écriture
- Alignement, ligne par ligne, de la transcription du texte
- Coopératif entre plusieurs utilisateurs

L'outil *Transkribus*

- ❑ Reconnaissance automatique de texte (HTR)
 - ❑ < intelligence artificielle
 - ❑ Apprentissage par la machine sur un corpus déjà transcrit
 - ❑ entraînement puis validation
 - ❑ > modèle présentant un certain pourcentage d'erreur
 - ❑ Ex. : Modèle « Cluniac charters » > 15%
 - Modèle « Himanis Chancery M1+ » > 6 %

Les données produites

- Une transcription du texte alignée sur l'image accompagnée de métadonnées
 - Exportable sous un grand nombre de formats
 - Exploitable par d'autres plates-formes : ex. *Monasterium.net*
- Un modèle de reconnaissance automatique de texte : mis à disposition de tous les utilisateurs de la plate-forme
 - Non réutilisable en dehors de Transkribus

Reproductibilité et interopérabilité

- ❑ Cadre d'interopérabilité en amont et en aval :
 - ❑ IIF,
 - ❑ récupération de transcriptions (word),
 - ❑ formats d'exportation de données (PDF, docx, txt, xml, ...)
- ❑ Reproductibilité :
 - ❑ Modèles sont fonctions des données d'entraînement
 - ❑ Plusieurs reconnaissances à partir d'un même modèle > plusieurs états textuels
 - ❑ Boîte noire du logiciel et de l'IA vs transparence
 - ❑ Gravité dépend des objectifs de recherche

Reproductibilité et interopérabilité

Faut-il se poser la question de la reproductibilité en philologie ?