

Contenu de la documentation

Chapitre 1

HTR

1.1 Problématique

Quatre à cinq mains différentes ont été repérées jusqu'à présent dans la correspondance de Constance de Salm (CdS). Cette variété d'écritures peut sérieusement entraver les performances d'un modèle de reconnaissance.

Deux pistes méthodologiques se dessinent :

1. Rassembler dans un premier temps des lettres qui sont de la même main, pour voir quels sont les résultats du modèle qu'H. Souvay a commencé à entraîner ;
2. Reprendre un modèle déjà entraîné à travailler sur plusieurs mains ; c'est l'option qui a été privilégiée par le projet Lectaurep¹).

1.2 Choisir un corpus d'entraînement

Les recueils de lettres constituent la part du corpus la plus normée sur le plan paléographique. La distribution des mains y est variable selon les tomes :

1. Le premier volume² présente une grande variété de main s'enchaînant fréquemment les unes aux autres ;
2. Le deuxième volume³ présente en revanche une meilleure cohérence paléographique : la même main peut se suivre sur un bon nombre de pages consécutives, facilitant l'entraînement d'un modèle sur une écriture particulière. En partie utilisé par H. Souvay pour ses tests, nous avons repris ce volume pour constituer un premier sous-corpus paléographiquement cohérent.

1. Alix Chagué, *Création de modèles de transcription pour le projet LECTAUREP #2*, Lectaurep : l'intelligence artificielle appliquée aux archives notariales, URL : <https://lectaurep.hypotheses.org/488> (visité le 05/04/2022).

2. Constance de Salm, *Correspondance générale, seconde copie, 1er volume, 1785-1814*, URL : <https://constance-de-salm.de/archiv/#/document/11215> (visité le 11/04/2022).

3. Id., *Correspondance générale, seconde copie, 2ème volume, 1815-1821*, URL : <https://constance-de-salm.de/archiv/#/document/11216> (visité le 11/04/2022).

1.2.1 Main 1

Nous avons établi une liste de 47 images (soit 47 doubles pages) au sein du 2e volume attestant une écriture homogène que nous dénommons *Main 1*. Nous avons pour cela arbitrairement découpé les lettres afin de ne travailler que sur un seul type d'écriture, sachant que les changements de main interviennent souvent en milieu de page. Quelques corrections de la main de CdS apparaissent ponctuellement ⁴.

1.2.2 Écriture de CdS

Le site ne publie aucune lettre originale de la main de CdS, mais 52 brouillons (*Entwurf*) ⁵.

Entraîner un modèle de reconnaissance sur cette écriture supposerait un travail délicat de transcription pour une écriture particulièrement cursive (compter environ deux semaines pour disposer d'une bonne vingtaine de pages), mais l'investissement peut en valoir la peine.

1.3 Segmentation et typage des zones d'écriture

Nous

4. Des reproductions en qualité réduite de ces images ont été placées dans le dossier `./htr/img/main1bd`

5. Le dépouillement se trouve dans le fichier `./htr/mains/brouillonsCDS.md`