

# Hériser une cursive du 17e siècle

Élodie Paupe, Université de Neuchâtel – [elodie.paupe@unine.ch](mailto:elodie.paupe@unine.ch)

Chargée de projet auprès des AAEB (Jura, Suisse)



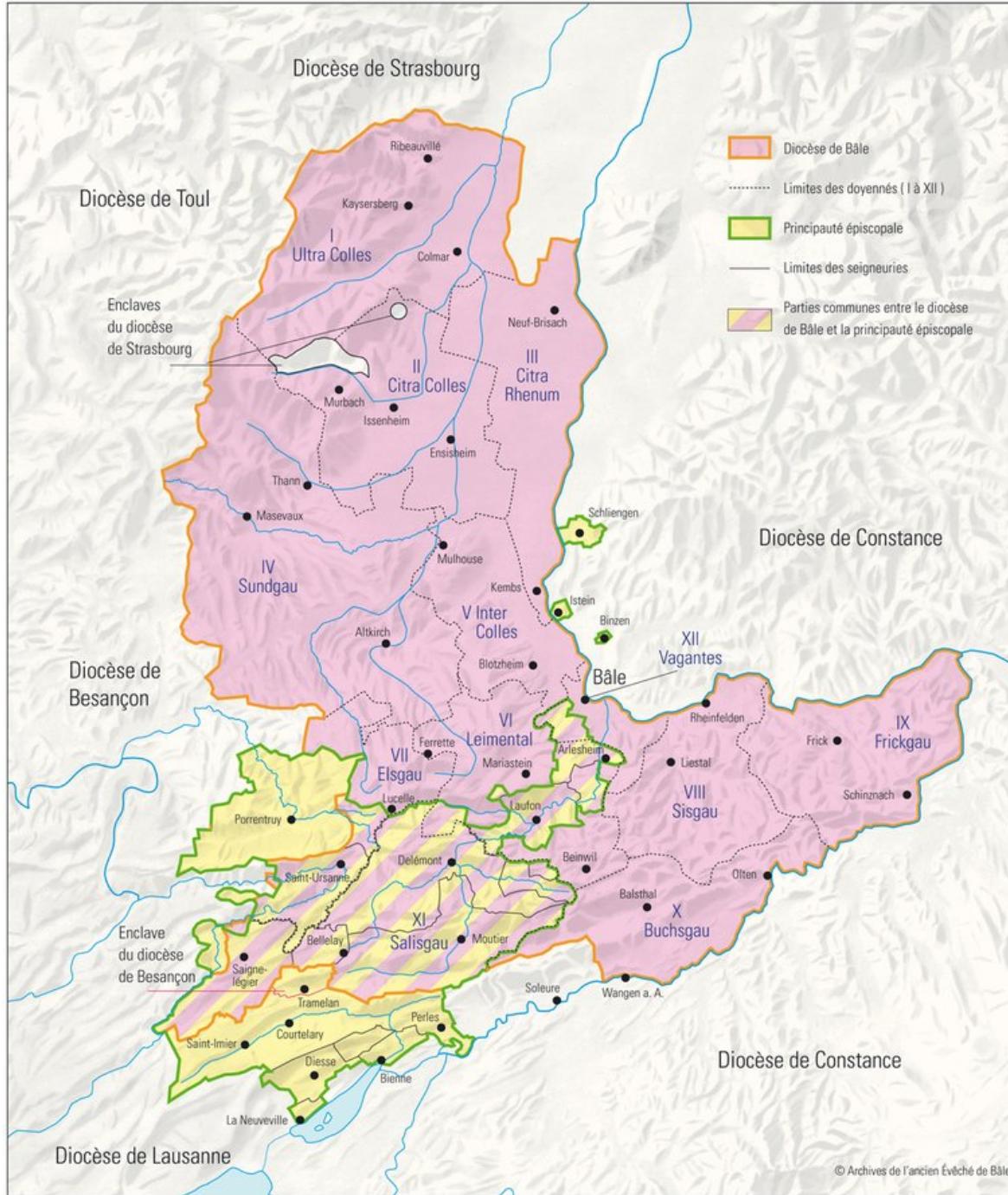
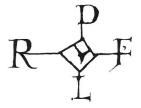
# Plan de la présentation

1. Introduction: le projet "Crimes et châtiments"
2. Le corpus
  - i. Choix du corpus
  - ii. Présentation de la main
  - iii. Choix de transcription
  - iv. Intérêt du corpus
3. Retour d'expérience sur l'entraînement de modèles HTR
  - i. Transkribus
  - ii. Kraken (FoNDUE)
4. Conclusion et développements à venir

# "Crimes et châtiments. Procès de sorcellerie et procédures criminelles: numérisation, mise en ligne et médiation"

## Cadre institutionnel

- Fondation intercantonale (BE, JU, BL, BS) créée en 1985
- Fonds entre le IXe s. et 1815, parmi lesquels le plus ancien document en langue française de Suisse (1244)
- Dossiers
  - qui concernent le diocèse de Bâle et la principauté épiscopale
  - qui concernent l'administration française (dép. du Mont-Terrible et du Haut-Rhin 1793-1814)
  - en français, en allemand et en latin



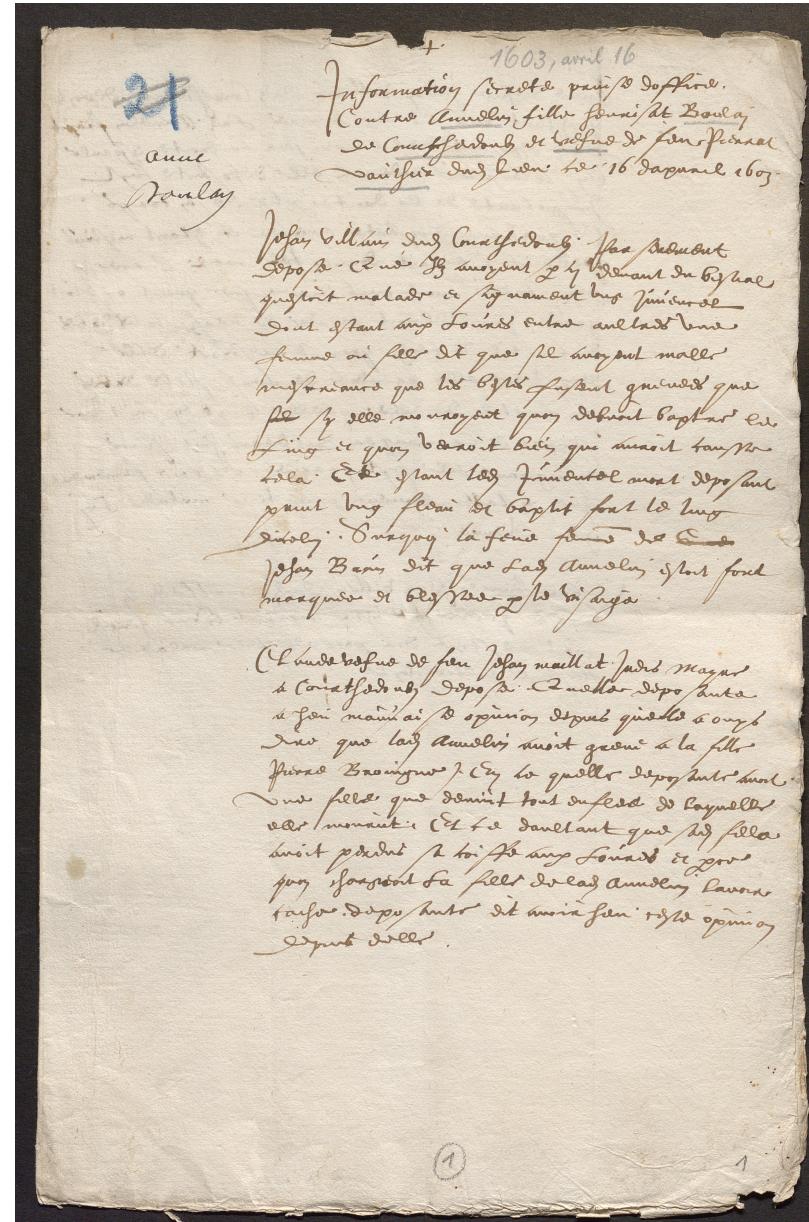
# Projet de numérisation "Crimes et châtiments" (2022-2025)

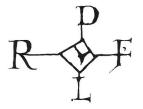
- Corpus relatif à la chasse aux sorcières et aux procès criminels (documents de 1461-1797; 25 m. linéaires/~110'000 pages)
- Axes du projet :
  - inventarisation très détaillée (en ligne : <https://archives-aaeb.jura.ch/suchinfo.aspx>)
  - numérisation des sources
  - développement de modèles HTR
  - mise en ligne des numérisations et des transcriptions dans un nouveau *front-end*
  - médiation
- Recherche de financement en cours

# Le corpus

## i. Choix du corpus

- 57 pages, soit environ 8'800 mots
- bifeuillets manuscrits
- une seule main, celle du prévôt Farine actif entre 1580-1618 (plus de 3'500 documents tous fonds confondus)
- autres mains d'archivistes pour la numérotation





En 24 de Novembre 1608

Festinuant fait au chasteau de Bâle  
Présent à la mort de l'archevêque

1. A quelle heure sera que l'heure est passée.

2. Si point elle a fasse au sommeil le plus  
Noyer pour l'assouvissement.

3. Si point elle a fait fasse au enfant sans temps.

4. Si point elle a fait jasent à la femme de dieu  
Noyer.

5. Quel point de temps se rétablit le plus  
Dame la matin de l'assassinat le moins de temps  
Noyer.

6. Qui gaigne le moins sommeil le plus de rétablissemens  
en temps de l'assassinat le moins de temps  
Noyer.

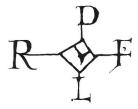
7. Qui gaigne le moins sommeil le plus de temps  
Noyer.

8. Qui a suivi la maladie de l'assassinat le moins de temps  
Noyer.

9. Qui a suivi la maladie de l'assassinat le moins de temps  
Noyer.

(3)

1



Adelbertus Bischof Bozengy  
Bottay.

187

La robe et le fumage de  
Mireb grise le 10 juillet 1616  
d'enduit malade.

Frauertord fandel la appelle pisse  
prostites plus que blanche

La fille Cappon vallat malade  
frapper malade et guérir de  
guérir.

Le 18 juillet del Cappon vallat  
guérir de 18 febb.

Roman. Mme.

Le 18 juillet quelle fait tel  
mouvement le Moys. 1616.  
Le Moys mouvement de l'espous  
et de les foetus pour le 18 febb.

Des maladie de vallat donne  
au general malade de 18 febb  
moys vallat que durent guérir.

Le 18 juillet presentat C. heide  
que l'ay mouvement de 18  
mouvement fait le 18 febb.

Cette que le voulent de 18  
J'auant a del de 18.

Le 18 juillet quelle bon une de  
l'espous Moys 1616. de l'espous  
toujours

Le 18 juillet mort a Cuillande  
de peste pour 18 juillet.

La robe enduit malade  
a Cuillande du juillet.

La malade donne au peste  
de 1616 presentat  
mort de la robe quelle  
est le 18 juillet 1616.

## ii. Présentation de la main

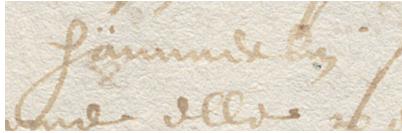
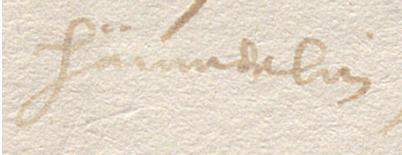
François des Arches Camis de  
Sion +  
+ obit. Commis de reportage.  
A son atté venu du feu Jeanne  
fondat. de Bouffay. Y a été  
établi au n° 18 20 - 26  
et dédié à son daugt. de la  
1609.

AAEB B 168/15-10.3 p. 1

### iii. Choix de transcription

- La transcription n'est pas allographétique.
- Les lettres aux formes variables selon leur position dans le mot sont transcrites par une forme unique.
- Les majuscules se divisent en deux groupes:
  - celles qui ont une morphologie distincte de la lettre minuscule ne posent pas de problème (B, E, P, Q, R, S, T...).
  - celles qui ne se distinguent des minuscules que par la taille (A, C, D, G, V...) et sont donc difficiles à différencier.
- Les <i/j> sont identiques et systématiquement sous forme majuscule au début d'un mot.

- On peut également noter qu'il manque souvent un jambage au groupe "mm", comme dans le nom de famille Hämmerlin.

"Hämmerlin" avec six jambages	"Hämmerlin" avec cinq jambages
	
AAEB B 168/15-23.3 p. 2	AAEB B 168/15-23.3 p. 1

- Les abréviations sont toutes développées.
- Les césures et apostrophes sont rendus d'après l'original.

## iv. Intérêt du corpus

- Histoire, sociohistoire, *gender studies*: grand intérêt pour la question de la chasse au sorcière
- Histoire du droit: procédures criminelles
- Philologie: documents en langue d'oïl dans une région frontière (au sud, le franco-provençal; à l'est, les langues germaniques)

# Retour d'expérience sur l'entraînement de modèles HTR

## Deux outils testés

- Transkribus
- FoNDUE (basé sur le moteur HTR Kraken)  
⇒ vendredi, 9h30, *FoNDUE – Mise en place d'une infrastructure eScriptorium à Genève par Simon Gabay*

# Transkribus

## Retour d'expérience avec Transkribus

1. Entraîner un modèle à partir des données du seul corpus :

Modèle	Nombre de mots	CER (set de validation)
AAEB_v2	8'800 mots	11%

## 2. Entraîner un modèle sur un modèle pré-existant (*fine tuning*) :

Modèle	Nombre de mots	Modèle de base	Epochs	CER (set de validation)
AAEB_v4	8'800 mots	Charter Scripts XIII-XV_M1	100	5.69%

## Résultat avec le modèle AAEB\_v4

de mesme quicelui lui donnat **ung** du  
poulsat quest dedans du pappier aupres  
de la bouete de graisse, duquel, elle  
**deluoit** faire a mourir gens et bestes  
**Iontesffois** quelle navoit encor fait  
a mourir **p drsonne** Mais quelle avoit  
fait a mourir les bestes suigvantes

B168/15-10.2, p. 3.

### 3. Entraîner un modèle sur un modèle préexistant + le corpus avec les abréviations développées

Modèle	Nombre de mots	Modèle de base	Epochs	CER (set de validation)
AAEB_ab_v1	8'800 mots	Charter Scripts XIII-XV_M1	100	4,68%

## Résultat avec le modèle AAEB\_ab\_v1

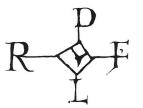
de mesme quicelui lui donnat **ung** du  
poulsat quest dedans du pappier aupres  
de la bouete de graisse, duquel, elle  
**de boit faire a mourir gens et bestes**  
**Contesfois** quelle navoit encor fait  
a mourir personne Mais quelle avoit  
fait a mourir les bestes suigvantes

B168/15-10.2, p. 3.

Du 19 Juillet 1609.

Riguelat le 19 juillet 1609 fonda le  
Bonfog et l'autre partie de la Commune  
fame. et suprême de la plus grande  
partie de la Commune et Brayles  
La que trace de partez. Le quadrille  
de formable partez. A que l'obligation  
de l'autre partie de la Commune  
ne vol pas longer pas.

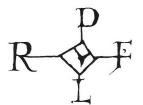
AAEB_ab_v1	AAEB_v4
Du 19 daougst 1609	D 19 daougst 1609
Aignelatte vefve de The fouvent de	Aignelatte vefve de The fouvent de
Bonfolz estant chargee <b>par la Comune</b>	Bonfolz estant chargee <b>p la Comme</b>
<b>caive et reputation</b> de La plus grande	<b>came et repuentx</b> de La plus grande
<b>partie de la Comunaulte dudit Baisol</b>	<b>ptie de la Comunaulte dud Baise</b>
A estee tiree en prison Sur precedentes	A estee tiree en prison. Sur precedentes
<b>informations</b> prises, A estee interrogué	<b>u formapt</b> prises, A estee interrogué
<b>ce jourdhi</b> matin. Laquelle na	<b>ceourdhoi</b> matin. Laquelle na
rien volsus cougnoistre	rien volsus congnoistre.



# Modèle testé sur des mains proches

- Main de Nicolas Farine (1612)

WER	CER	Word Acc	Char Acc
68.02%	29.81%	20.87%	60.31%



- Main du Dr Faibvre (1612)

WER	CER	Word Acc	Char Acc
85.92%	44.29%	8.15%	42.87%

Document 6. La forme de ce poème ob-  
elle est tout à made que j'ose et.  
reformé si je n'est pas dans un village où  
elle je ont son opinion qu'elle est tout  
vraiment. Du 12 de Mars 1613

Pierre Magni cage de 40 ans aussi  
de Dampierre et assomante répose  
qu'elle est infâme de sort lege et  
comme il croit de bons amis de leur  
village servis de 3 des parents dielle  
Jeanne. Et de la femme l'origine  
bien entendue de Beaumouys du village  
quelle leur porté des dommages b'importe  
elles. Si elles n'indit.

Jeanette Mandry femme dud'  
Pierre cage de 40 ans environ et  
ses parents n'espèce dépose le b'importe  
que son mari de faire.

Du 16 d'août.

Marquerite femme de Mauris O'Gill  
duz Dampierre cage de 31 ans  
environ apres l'arrant pastre dépose  
qui il ya vingt ans qu'elle entant repon-  
der à la Beaumouys à Tonnerre.  
Et qu'illes croit que la pluralité  
de leur commandement est de cette opion  
Et qu'il ya que ans qui l'auve de varie de  
deposition se remettre une cause quel que  
vingt pas. Cinq de la main de la

# FoNDUE

## Retour d'expérience

1. Entraînement à partir des données du corpus :

Modèle	Nombre de mots	Accuracy
AAEB_v1	8'800 mots	83,8%

## i. Tentative de binarisation (méthode Otsu)

Modèle	Nombre de mots	Accuracy
AAEB_bin_v1	8'800 mots	75,1%

Information / acte paroissial d'orthographe.  
Contre Annely, fille fondaire de Baudry  
de Conflans, veuve de Jean Pichot  
d'aujux dans l'encre le 16 d'april 1603.

## ii. Modification de l'architecture d'apprentissage

Un second modèle a été entraîné avec une architecture transformée:

```
-r 0.0001 -s '[1,0,0,3 Cr3,3,16 Mp3,3 Lfys64 Lbx128 Lbx256 Do]'
```

**-r** modifie le taux d'apprentissage (valeur par défaut: 0.001)

**-s** modifie l'architecture des réseaux de neurones (par défaut: [1,48,0,1 Cr3,3,32 Do0.1,2 Mp2,2 Cr3,3,64 Do0.1,2 Mp2,2 S1(1x12)1,3 Lbx100 Do])

# Conclusions

- Transkribus
  - Modèle développé sur la main du prévôt Farine satisfaisant
  - Résultats sur d'autres mains, même proches, insatisfaisants
- FoNDUE
  - ??
  - ??

## Prochaines étapes:

- créer de la vérité de terrain pour quelques mains bien représentées dans les fonds ⇒ créer des modèles spécifiques
- à partir de ces données ⇒ créer un modèle générique
- tester d'autres architectures pour améliorer le score
- publier les transcriptions et les numérisations

**Merci de votre attention !**