

Hériser une cursive du 17e siècle

Élodie Paupe, Université de Neuchâtel – elodie.paupe@unine.ch

Chargée de projet auprès des AAEB (Jura, Suisse)

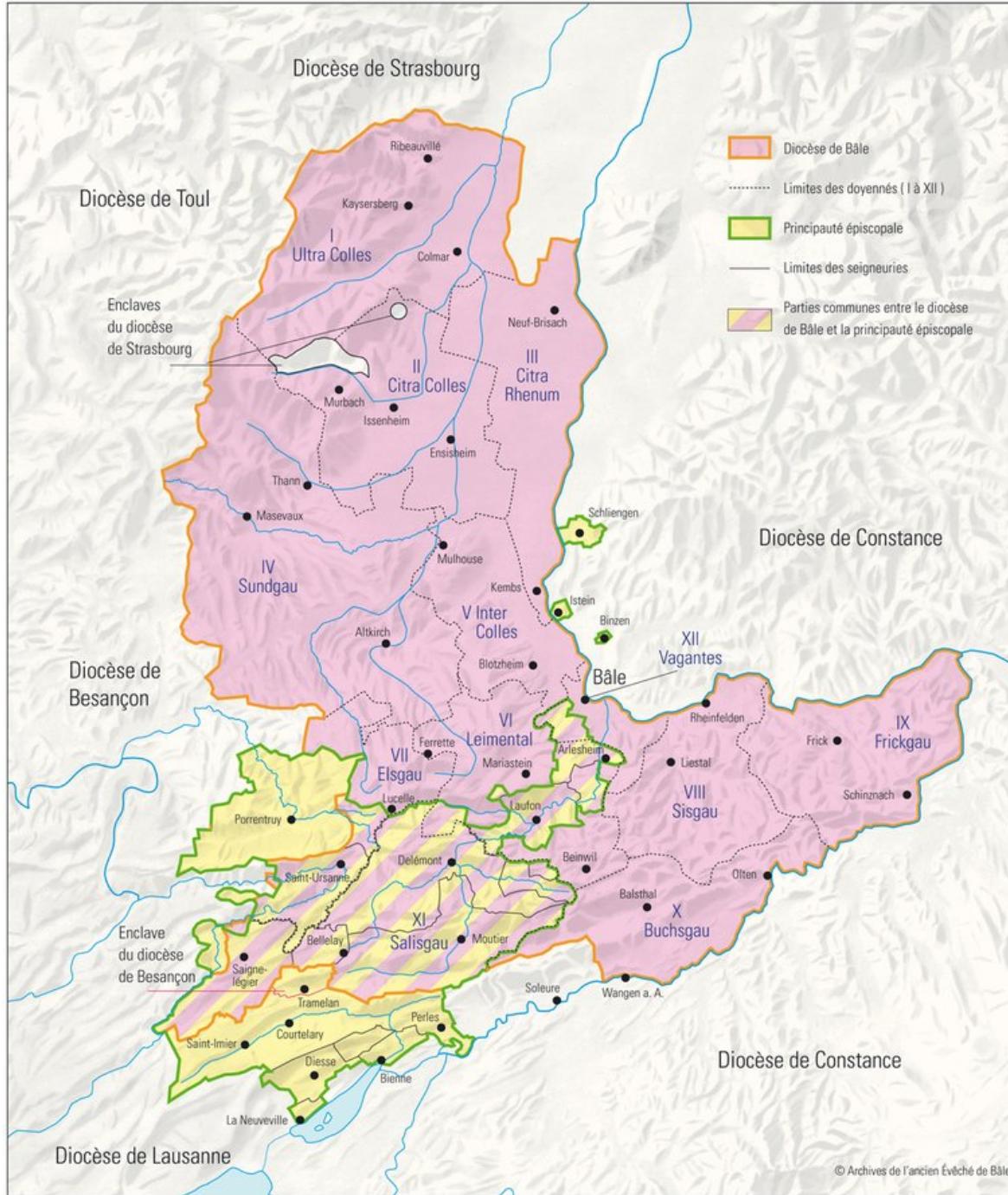
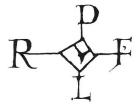
Plan de la présentation

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

1603, avril 16

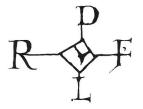
21

ame
devalay

Information / faite & par le Doyen.
Contre Anneles, fille faine de Boulay
de Conffordant de Laffos & son pere
aujor d'aujour du 16 d'april 1603.

Isay Villain des Conffordants. Ses adoubent
Doyen &c mesme moyent le deau de Boulay
qu'oit malade & le auant les viii ans
doul grant temps souffrir au lait veue
fame ou fles & que le auoyent malade
mesme auant que les bys fayent gencel que
le / yee auoyent quoy debout cognoit en
l'iso le quoy venoit blyz qui auoit tenu la
tela. Et le grant temps furent auant Doyen
auant long temps de baptis fait le long
distret. Or auoyant la feme fane de Isay
Isay Bais dit que lad Anneles s'est fonde
auoyante de led Doyen & le laigera.

Ch'auant auant de fene Isay meillat Paris Major
& Conffordant. Doyen &c mesme Doyen auant
la fene auant le opiniot dequelques temps
davantage que lad Anneles auoit gencel a la fene
Pere Boulay & le quello dequel temps auant
une felle que deuant tout auant dequelles
est auant. Et le auant que led felle
auoit pereut a toz les temps souffrir et por
pour auant la felle de lad Anneles auant
touz depon auant de auant fene & le opiniot
de quelles.



En 24 de Novembre 1608

Festinuant fait au chasteau de Bâle
Présent à la Cour de Bâle

1. A quelle heure sera que le bœuf est déboulé.

2. Si point être à faire au sommeil le plus
Noyer pour le sommeil.

3. Si point être à faire au enfant sans temps.

4. Si point être à faire au grand et la femme de la
Noyer.

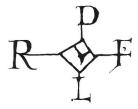
5. Quel poisson de bœuf sera le meilleur
de la main de M. Noyer, lequel sera temps
qu'il sera, nullement de la main de M. Noyer
que il sera temps.

6. Qui sera le plus de la main de M. Noyer
pour lequel sera le meilleur.

7. Pourquoi sera le bœuf le plus que l'autre.

8. Qui sera le meilleur saumon qui sera au bœuf
de la main de M. Noyer.

9. Qui a suivi la maladie de la femme de la
main de M. Noyer et lequel est le plus
pour la panette.



Adelbertus Bischof Bozengy
Bottay.

187

La robe et le fumage de
Mireb grise le 10 juillet 1600
d'Edouard malade.

Frauertord fand le 10 juillet 1600
morte le 10 juillet 1600

La fille Cappon vallat malade
frapper malade et guérir le 10
juillet.

Le 10 juillet 1600 Cappon vallat
guérir le 10 juillet.

Roman de Mars.

Le 10 juillet quelle fait sa
morte le 10 juillet 1600
de nos mors de la grippe
et le 10 juillet faire la mort le 10 juillet.

Des malades vallat domine
du general malade le 10 juillet
moy vallat que j'aurai guérir

Le 10 juillet présentat Céleste
que l'ay mort auquel le 10 juillet
morte et faire le 10 juillet.

Cette que je veux le 10
juillet a elle le 10 juillet.

Le 10 juillet quelle bonne de
l'abbé Mme Böly à la mort
faire

Le 10 juillet mort à Cenklam
de pitié pour le 10 juillet.

Le 10 juillet domine malade
à Cenklam du pitié.

La malade domine auquel
le 10 juillet presentat
mort de la grippe quelle
faire le 10 juillet

Domine pitié

ii. Présentation de la main

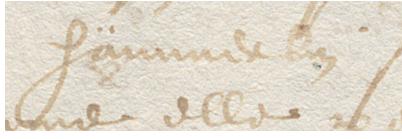
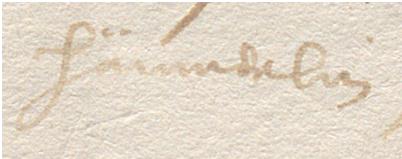
François des Arches Camis de
Sion +
+ obit. Commis de reportage.
A son atté venu du feu Jeanne
fondat. de Bouffay. Y a été
établi au n° 18 20 - 26
et dédié à son daugt. de la
1609.

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.

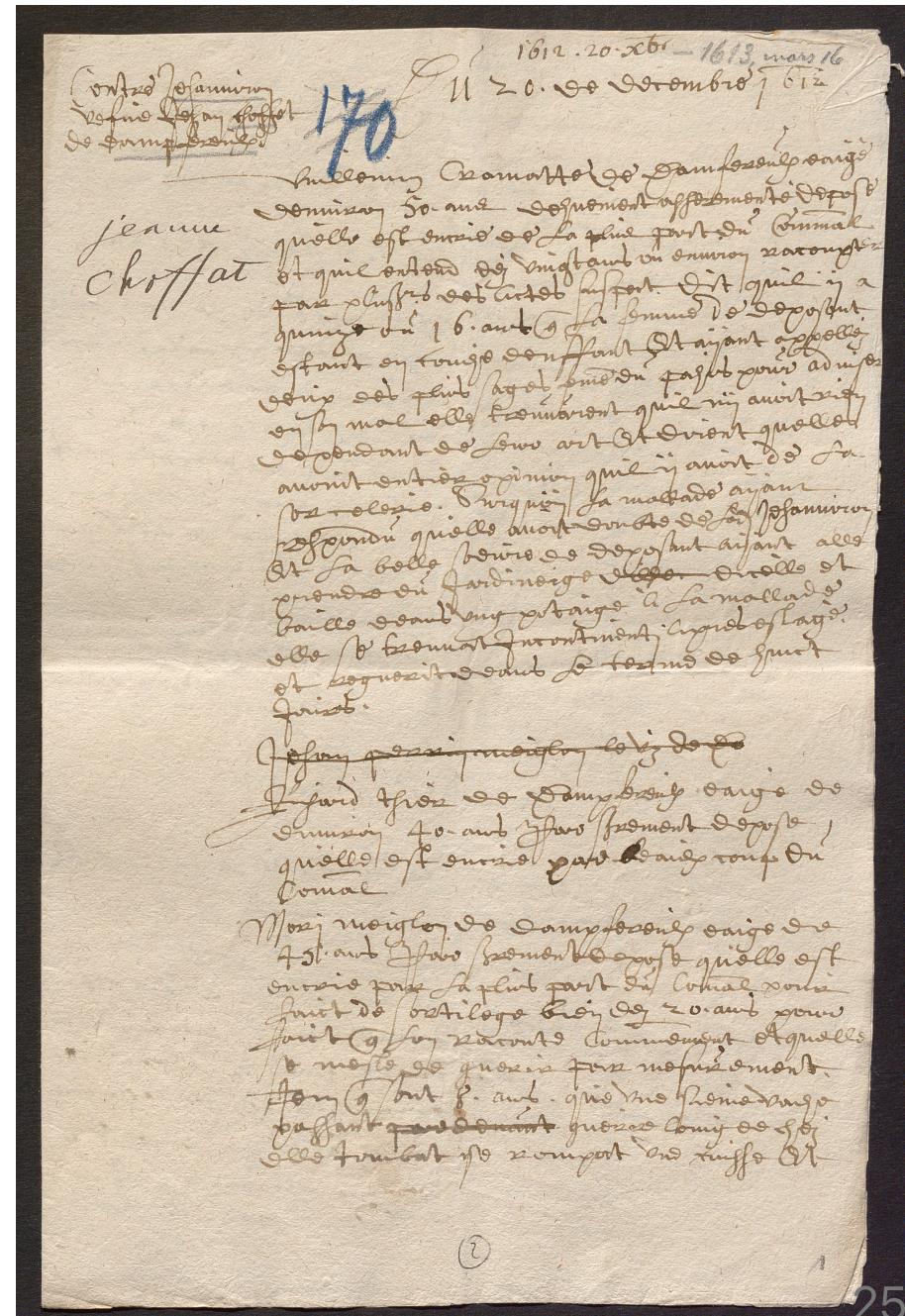
Riguelat le 19 juillet 1609 fonda le
Bonfog et l'autre partie de la Comme
famille et superstite de la plus grande
partie de la Comme autre qui est Bayloz
La quel trace de partez. Le quadrille
de partez partez. A quel Judex
et bonfog matte. Laquelle na
rrue vol les tongnes nos.

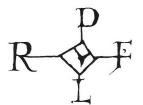
AAEB_ab_v1	AAEB_v4
Du 19 daougst 1609	D 19 daougst 1609
Aignelatte vefve de The fouvent de	Aignelatte vefve de The fouvent de
Bonfolz estant chargee par la Comune	Bonfolz estant chargee p la Comme
caive et reputation de La plus grande	came et repuentx de La plus grande
partie de la Comunaulte dudit Baisol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