

LECTAUREP

Lecture Automatique de Répertoires

799	13	Innovation	Chaillot	14	3.75
800	13	14	3.75
801	13	Decharge	Le Déan	14	3.75
802	13	...	Blanc 13	14	3.75
803	13	14	910
			...	14	3.75

Archives nationales (DMC/DMOASI)

Équipe ALMAAnACH

Atelier Culture - Inria, 22 novembre 2018

Partenariat

Archives nationales

Département du Minutier Central (DMC)

Département de la maîtrise d'ouvrage du système d'information (DMOASI)

- Marie-Françoise Limon-Bonnet (DMC)
- Aurélia Rostaing (CC DMC)
- Danis Habib (CHED DMC)
- Frédéric Zamarreno (resp. DMOASI)
- Gaetano Piraino (DMOASI)

ALMAnaCH (*Automatic Language Modelling and Analysis and Computational Humanities*)

Équipe-projet Inria-EPHE

- Marie-Laurence Bonhomme, stagiaire Master TNAH de l'ENC
- Marc Bui, EPHE
- Éric de la Clergerie, Inria
- Benjamin Kiessling, EPHE
- Marie Puren, Inria
- Laurent Romary, Inria
- Charles Riondet, Inria
- Daniel Stökl Ben Ezra, EPHE

Le corpus d'archives



Le notaire, un officier public ministériel

- Détenteur d'une part de l'*auctoritas publica* ou de la foi publique,
- Producteur d'actes authentiques en forme de minutes,
- Lesquelles sont listées dans des répertoires

Le corpus d'archives

Les répertoires des études notariales sont des registres dans lesquels un cleric consigne, dans l'ordre chronologique et pour chaque jour, les actes qui ont été passés et enregistrés dans son étude.

Ils doivent nécessairement contenir la date de l'acte, sa nature et son espèce (en minute ou en brevet*) ainsi que le nom des parties et la relation de l'enregistrement.

Un formulaire pré-imprimé à 7 colonnes.

*La minute d'un acte notarié est l'original d'un acte authentique dont le notaire ne peut se défaire ; un acte en brevet est un acte notarié dont l'original est dépourvu de formule exécutoire, et est remis aux parties.

Un exemple de
minute, les statuts de
la société civile du
« Bon Marché »

Arch. nat. MC/ET/XVI/1416



La mention de
cette même minute

- Statuts de la
société civile du
Bon Marché -
au répertoire du
notaire Gatine

https://www.siv.archives-nationales.culture.gouv.fr/siv/media/FAN_IR_051822/c1ug18c3sei4--w32j38d5447t/D_AFANCH96_016MIC06736_L

N ^o DU RÉPERTOIRE	DATES DES ACTES.	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES. INDICATIONS, SITUATIONS ET PRIX DES BIENS.	RELATION DE L'Enregistrement.	
		EN BREVETS.	EN MINUTES.		DATES.	DROITS.
446	?	1. du 10 juin 1886	Realisation	An 1886 mois d'Avril (suite) Houmard (de prêt de prêt pour crédit foncier à Charles Ferdinand) 11 rue Colson Paris		
447	3		- Récession de	Levasseur (par Caroline Henriette Le) 89 rue Ledoual à Paris	2	27.50
448	3		- Donation	à Marie Calistin Lougeard (rue des 3 bornes pour tuteur daté à Anoyant (par Edward Lion) 1 rue Sauffoy Cour à Paris		9.75
449	3		entre époux	à Lucotte Eugénie Baillet (sa fr. uninst. exppte)		"
450	3	1. du 11 avril 1886	Dépôt de Déclat	Anoyant (par le not. à leur mari)		"
451	3		- Obligation	Courmesnil - Tournehambaert (concernant la Ste) 16 place Vendôme Paris	6	11.25
452	3		- Emargation de péleri	Dubois (par Françoise Margte Adèle Bollobt (veuve) Jules Maurice) 11 rue de la Monnaie à Joseph Foin Taquille Minin de Tollenay à Tilly (S. d. O) de 10000 rcbul. au 10 juillet 1886 et d'obl. de 10000 de la loi 1882	6	153.75
453	3 + 2		- Bail	Prot, (par Paul) 67 rue Sauffoy à Jn. Ste Coste de Montgallien de lauz à Paris faub. St Denis 87 x bis pour 11 an du 1 ^{er} Oct 86 moyt. 108% - 11 8/10, 14 8/10, 16 8/10	6	80.40
454	6		- Procuration	Dauvergne (par Jeanne Maria Chassomary 21 rue Bertrand (par Celestin Louis) & Lucien Celestin Louis) - 77 rue Perrière à Boulogne Seine à M ^{rs} Den pour Ste du Bon Marché	9	7.50
455	4		- Procuration	Bouccault (par Margte Guerin Vve Gertrude Jaquez 11 rue du Tra à Lucien Frederic Henri Labitte (rue de) Rachelle pour faire & accepter promesse de M ^{rs} Dation	9	3.75
456	4		- Donation	Bouccault (par les mêmes) à la Société Civile de 30000 francs pour fonder (rue de retraite)	9	125.000 --
457	4		- Statuts de la	Société Civile du Bon Marché 11 à Paris rue du Bac	9	505.75
458	4		- Fondation	1 ^{re} Civile par acte entre M ^{rs} Vve Bouccault & partie de ses employés pour 8 an de lauz de retraite		
459	4		- Bail	Bon Marché (par laud. Ste) au profit des employés de la maison de 20% du gros intérêt	9	3.75

Le corpus d'archives



AU NOM DU PEUPLE FRANÇAIS.

BONAPARTE, PREMIER CONSUL,

PROCLAME loi de la République le décret suivant, rendu par le Corps législatif le Vingt-cinq Ventôse, an onze, conformément à la proposition faite par le Gouvernement le quatre et dix-neuf mois, communiquée au Tribunal le Vingt-neuf Ventôse.

DÉCRET:

Titre I.^{er}

Des Notaires et des actes notariés.

Section I.^{re}

Des fonctions, responsabilités des Notaires.

Article 1.^{er}

Les Notaires sont les fonctionnaires publics établis pour recevoir tous les actes et contrats auxquels les parties doivent ou veulent faire donner le caractère d'authenticité attaché aux actes de l'autorité publique, et pour en assurer la date, en conserver le dépôt, en délivrer des copies et expéditions.

Art. 1.^{er}

Depuis 1803 (25 ventôse an XI art. 29-30), le répertoire est « normalisé » et enrichi dans ses contenus

Projet LECTAUREP ⇒ **Section du corpus**
Entre 1803 et 1940, un corpus (rien qu'à Paris) de 917 notaires différents, env. 1800 registres

Le corpus d'archives

Le Minutier central des notaires de Paris, créé en 1928 au sein des Archives nationales. Il a pour vocation de conserver les fonds des 122 études notariales historiques de la capitale :

- 20 millions de minutes
- 3 300 répertoires
- 172 000 liasses et registres

⇒ 26 000 mètres linéaires.

Rechercher

Tous les mots saisis



Toutes les archives

Archives numérisées

Producteurs d'archives

DE CRITÈRES

CRITÈRES SÉLECTIONNÉS :

Producteur : Mocquard, Constant Amédée (1815-1889)

Modifier ma recherche



Trier par Pertinence

Résultats groupés par inventaires

Voir :

15

30

45

2 résultats dans 2 inventaires.

INVENTAIRE Cote(s) : MC/ET/LXVIII/1054 - MC/ET/LXVIII/1281, MC/RE/LXVIII/16 - MC/RE/LXVIII/17

Minutes et répertoires du notaire Constant Amédée MOCQUARD, 12 juillet 1853 - 21 juillet 1875 (étude LXVIII)

Voir l'inventaire

Producteur(s) :

Mocquard, Constant Amédée (1815-1889)

Etude LXVIII

[Voir le résultat](#)

Voir tous les résultats dans l'inventaire

INVENTAIRE Cote(s) : MC/RE/LXVIII/16 - MC/RE/LXVIII/17

Images des répertoires du notaire Constant Amédée Mocquard pour l'étude LXVIII

Voir l'inventaire

Producteur(s) :

Etude LXVIII

Mocquard, Constant Amédée (1815-1889)

[Voir le résultat](#)

Salle des inventaires virtuelle des Archives nationales :

Répertoires numérisées en ligne en mode image pour l'Ancien Régime comme pour le XIX^e et le premier XX^e.

Nos DU RÉPERTOIRE	DATES DES ACTES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.	
		EN BREVETS	EN MINUTES		DATES	DROITS
804	13	Procurat.		An 1920, mois d'Avril Brugnière (par Rose Suzanne Gauthier, ve de François Louis) à Paris rue Michel Ange 84, en blanc, pour toucher	15	3.75
805	13	Donation		de Vauguignouse, (par Josselin Charles Edmond) St. Thiers, avenue Hoche 13, à Marie Anne Esther Tassel sa femme (toute pp ^l)	"	"
806	13	Donation		de Vauguignouse (par Made) ni Tassel, sur veuve au d ^l son mari (toute pp ^l)	"	"
807	13	Et de mariage		Lenormand (entre Henri Théophile Louis) à Paris rue Beauregard 26, et Marie Anne Bierrat, d ^l à Paris rue Montparnasse 32 (C ^t d'acquit)	15	177.93
808	14	Procurat.		Halberci (par Louis Eugène Albert), d ^l à Paris rue Rambuteau 116, en blanc. pour radier inscript.	15	3.75
809	14	Col de pp ^l		Haas Cecile Louise Camille L'épargne 75-3234319 au nom de "Conventual Lina" femme de Max, en son vivant, d ^l à Vincennes rue du Bois 9 - de 250 fr	16	"
810	14	Maintenue		Rodes (par Antoine) à Paris rue Gaillon 10 pres et actuel ^l rue St Lazare 13 d'inscript. au 2 ^d de commune		

Ressource riche mais difficile d'accès

Depuis le XVI^e siècle, le répertoire est donc un registre qui liste, pour chaque jour, les actes passés dans une étude donnée...

... Pour exploiter ces répertoires, il faut toujours en passer par un dépouillement systématique parfois long

... Il n'en existe pas de transcription, pas de possibilité de recherche en plein texte

Objectifs du projet

- Offrir un meilleur accès à cette ressource
- Possibilité pour les chercheurs de réaliser des exploitations statistiques et quantitatives, en considérant ces répertoires non plus comme des inventaires mais comme des sources primaires.

Approche

- Reconnaissance automatique de structures et d'écriture manuscrite
- Mise à disposition des images numérisées de ces répertoires et de leur transcriptions sur une plateforme en ligne permettant d'y effectuer des recherches avancées.

Éléments de méthode

Segmentation des tableaux

- Distinction des classes d'information
- Analyse des répétitions de structures

HTR (Handwritten Text Recognition)

- Scribes multiples
- Utilisation de ressources sémantiques (référentiels des Archives Nationales)

Vers la conception d'une chaîne de traitement

1. Layout analysis : analyse de la structure physique et sémantique des répertoires
2. Reconnaissance d'écriture manuscrite
3. Extraction d'informations (Personnes, Lieux)
4. Brique participative (Correction des transcriptions, annotation)
5. Intégration dans le Système d'information des Archives Nationales
6. Mise à disposition de la communauté des chercheurs

Phase 1 du projet

Accent mis sur les premières étapes de la chaîne de traitement

1. **Layout analysis : analyse de la structure physique et sémantique des répertoires**
2. **Reconnaissance d'écriture manuscrite**
3. *Extraction d'informations (Personnes, Lieux)*
4. *Brique participative (Correction des transcriptions, annotation)*
5. *Intégration dans le Système d'information des Archives Nationales*
6. *Mise à disposition de la communauté des chercheurs*

Layout analysis
(structure de la
page)

La segmentation des actes

N ^{os} DU REPERTOIRE	DATES DES ACTES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.		
		EN BREVETS	EN MINUTES		DATES	DROITS	
1654	27	Pragmation		<p>An 1919, mois de Novembre</p> <p>Noël, (par l'ami) 5^e à Saint-Jacques le Montagne St Genevieve 47, sublane, pour remplir 2^{es} Bas, Coust. vint, comus au lieu de 2^{es} de Guillaumes) de la son domicile à Saint-Jacques le 27 août 1919 en restant pour l'année n^{os} 1325777 et 1636220 le 10^e chaux Grimbert, de Marie Sylve Timot épouse de Marie Joseph (par l'ami) à Saint-Basile 19 Doire (par l'ami) Elise Louise et l'ami Mathieu à Saint-Basile 19 Marie Valère Saint pour remplir 2^{es} à se sachet, à Saint-Basile 19 St Antoine (par l'ami) Julien, St-Basile, com De Vincennes, 112 (sublane) de pouvoirs qui lui ont été donnés par son père Gouin, fils, (par l'ami) André-Henri Gouin à Croix (2^{es}) et Francis Augustin Gouin à Saint-Jacques Dammes n^{os} 26^{es} comme membres de la société sur à Saint-Basile, n^{os} de l'acte chaton n^{os} 10, de son captif en greffe n^{os} de l'acte de la son le 16 juin 1911 n^{os} 30242 n^{os} 88 ? fe 173, com</p>	29	3.75	
1655	27	Est de pp ^{te}			"	"	
1656	27	Est de vie			"	29	1.88
1657	27	Pragmation			"	"	"
1658	27	substitution			"	15	3.75
1659	27	manuscrit			"	2	3.75

Structure du tableau dans le registre :

- Imprimé
- Lignes verticales
- Cohérent sur tout le corpus

Structure de l'acte :

- Manuscrit
- Plusieurs mains
- Pas de lignes horizontales
- Alignement aléatoire

Segmentation automatique de la page, de l'entête et des colonnes

N°	LIVRE	NATURE DE L'ÉPREUVE		NOMS, PRÉNOMS ET DOMICILES DES PARTIES	DÉLÉGUÉS	
		DE DÉCRET	DE JURY		NOM	PRÉNOM
An 10 ^{ème} mois de Junc.					5	7.1
101	1	5	7.1
102	1	5	7.1
103	1	7	10.1
104	1	7	10.1
105	1	7	10.1
106	1	7	10.1
107	1	7	10.1
108	1	7	10.1
109	1	7	10.1
110	1	7	10.1
111	1	7	10.1
112	1	7	10.1
113	1	7	10.1
114	1	7	10.1
115	1	7	10.1
116	1	7	10.1
117	1	7	10.1
118	1	7	10.1
119	1	7	10.1
120	1	7	10.1
121	1	7	10.1
122	1	7	10.1
123	1	7	10.1
124	1	7	10.1
125	1	7	10.1
126	1	7	10.1
127	1	7	10.1
128	1	7	10.1
129	1	7	10.1
130	1	7	10.1
131	1	7	10.1
132	1	7	10.1
133	1	7	10.1
134	1	7	10.1
135	1	7	10.1
136	1	7	10.1
137	1	7	10.1
138	1	7	10.1
139	1	7	10.1
140	1	7	10.1
141	1	7	10.1
142	1	7	10.1
143	1	7	10.1
144	1	7	10.1
145	1	7	10.1
146	1	7	10.1
147	1	7	10.1
148	1	7	10.1
149	1	7	10.1
150	1	7	10.1
151	1	7	10.1
152	1	7	10.1
153	1	7	10.1
154	1	7	10.1
155	1	7	10.1
156	1	7	10.1
157	1	7	10.1
158	1	7	10.1
159	1	7	10.1
160	1	7	10.1
161	1	7	10.1
162	1	7	10.1
163	1	7	10.1
164	1	7	10.1
165	1	7	10.1
166	1	7	10.1
167	1	7	10.1
168	1	7	10.1
169	1	7	10.1
170	1	7	10.1
171	1	7	10.1
172	1	7	10.1
173	1	7	10.1
174	1	7	10.1
175	1	7	10.1
176	1	7	10.1
177	1	7	10.1
178	1	7	10.1
179	1	7	10.1
180	1	7	10.1
181	1	7	10.1
182	1	7	10.1
183	1	7	10.1
184	1	7	10.1
185	1	7	10.1
186	1	7	10.1
187	1	7	10.1
188	1	7	10.1
189	1	7	10.1
190	1	7	10.1
191	1	7	10.1
192	1	7	10.1
193	1	7	10.1
194	1	7	10.1
195	1	7	10.1
196	1	7	10.1
197	1	7	10.1
198	1	7	10.1
199	1	7	10.1
200	1	7	10.1

- + tableau et images constants
- la page peut-être en rotation

étapes

- détection de lignes horizontales et verticales du tableau (à l'aide de transformations morphologiques et de composantes connectés)
- détection et correction de la rotation de la page
- segmentation de la page de l'arrière-fond
- segmentation en colonnes et entête de la page
- suppression des séparateurs du tableau

Indices de segmentation

REPERTOIRE	ACTES	EN BREVETS	EN MINUTES	INDICATIONS. SITUATIONS ET PRIX DES BIENS	DATES	DROITS
1654	27	Prumet		An 1919, mois de Novembre J. Toël (pa. Linné) et G. Linné (pa. Linné) ont par acte en date du 27 novembre 1919, fait passer à G. Linné le bien (pa. Linné) situé au lieu dit de St. Jean n° 29 le 27 avril 1919 en vertu d'un acte de Prumet n° 1322777 et 1635220 le 10 ^o et G. Linné (pa. Linné) et G. Linné (pa. Linné) ont par acte en date du 27 novembre 1919, fait passer à G. Linné le bien (pa. Linné) situé au lieu dit de St. Jean n° 29 le 27 avril 1919 en vertu d'un acte de Prumet n° 1322777 et 1635220 le 10 ^o et G. Linné (pa. Linné) et G. Linné (pa. Linné) ont par acte en date du 27 novembre 1919, fait passer à G. Linné le bien (pa. Linné) situé au lieu dit de St. Jean n° 29 le 27 avril 1919 en vertu d'un acte de Prumet n° 1322777 et 1635220 le 10 ^o et	29	3.75
1655	27	Prumet				
1656	27	Prumet			29	1.88
1657	27	Prumet				
1658	27	Prumet			15	3.75
1659	27	Prumet			2	3.75

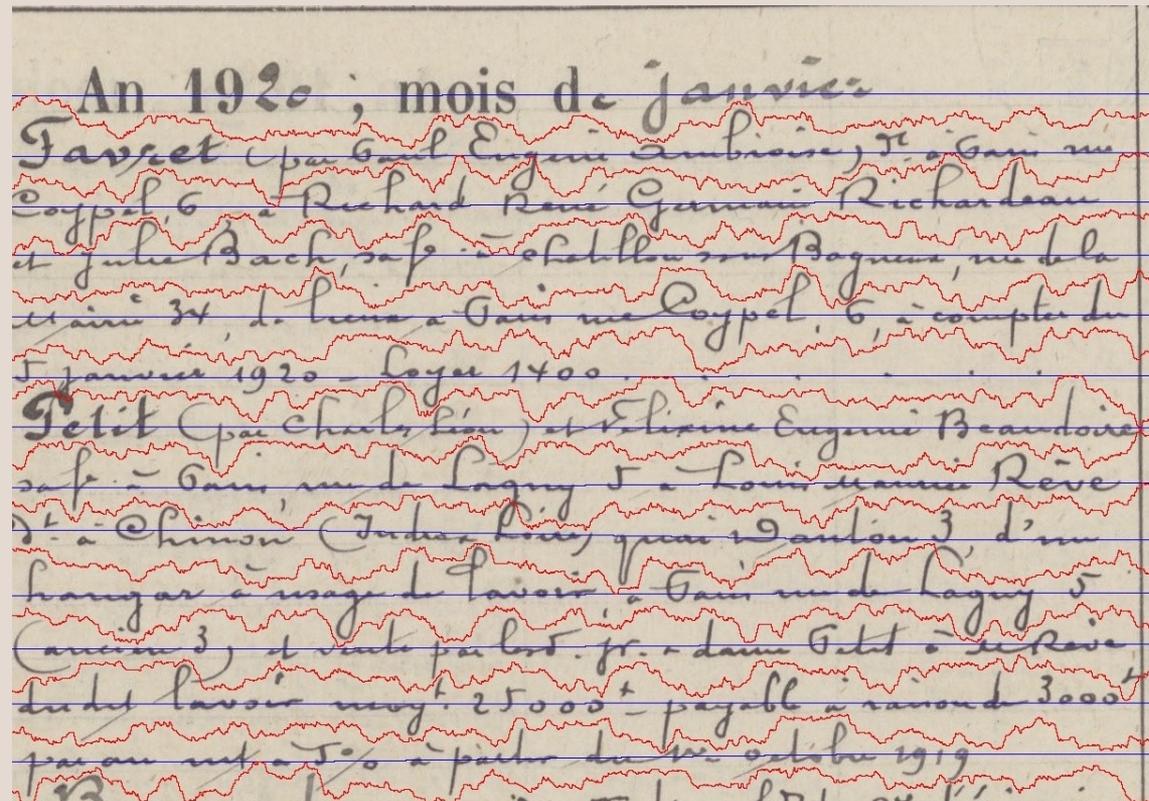
Limite supérieure de l'acte:

- N° d'ordre
- Date de l'acte
- Typologie
- Nom du signataire en gras

Limite inférieure de l'acte:

- Description de l'acte: dernière ligne parfois incomplète
- Date de l'enregistrement de l'acte
- Taxes acquittées

Segmentation automatique des lignes



An 1920 ; mois de janvier
Favret (par Gaud. Eugène Ambroise) Pl. à Gaurin rue
Coyzel, 6 à Richard René Germain Richardeau
et Julie Bach, s'af. à Echellousson Bagnères, rue de la
Mairie 34, de l'écrite à Gaurin rue Coyzel, 6, à compter du
1^{er} janvier 1920 - loyer 1400
Petit (par Charles Lion) et Elisine Eugène Beaudouin
s'af. à Gaurin rue de Lagry 5 à Louis Marie Réve
d^t à Chinson (Julien Pichet) quai d'Antoine 3, d'un
hangar à usage de lavoir à Gaurin rue de Lagry 5
(Ancien 3) et vendu par les J^r. et dame Petit à M. Réve
du dit lavoir moyt. 25000^f - payabl. à raison de 3000^f
par an net à 5% à partir du 1^{er} octobre 1919

Seamcarving (chemin d'énergie minimale) pour segmenter les lignes de la colonne de description de l'acte.

Extension de la segmentation des lignes au reste du tableau

N°	DAILLIS		NATURE ET ESPÈCE		NUMS, PRÉNOMS ET DOMICILES DES PARTIES	RELATION			
	SÉRIÉRIE	ACTES	DES ACTES :			L'Enregistrement			
			EN DAIYARD	EN PIERRES		MOIS	JOURS		
59	1				du 10 ^e mois de Janvier				
60	10							12	3.27
61	10								
62	10								
63	10								
64	10								
65	10							13	4.27
66	10								
67	10							15	6.00
68	10								

Bons résultats en général:
Alignements horizontaux
parfois décalés

Difficultés de segmentation

Présence d'un obstacle (tampon) dans la colonne 2 ⇒
Numéro de l'acte décalé dans la marge (colonne 1)

Nos	DATES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.	
		EN BREVETS	EN MINUTES		DATES	DROITS
43	8		Mantoux	An 1920, mois de Janvier Talbert (par Lucien Talbot) et sa femme Roussant, dem ^t Mendon, rue des Vignes 21, d'inscriptions au 2 ^e de la commune le 11 9 ^e 1909, n ^o 13459. contre et c. M. Gabriel. Dix années à choisir Le Roi, avenue V. Hugo, 29	9	7.38

Difficultés de segmentation

- Dépassement du texte sur plusieurs colonnes
- Ruptures textuelles aussi lors des mentions introductives ou légales du juge

30	6	Contribution de requête	Le surplus payé Jossilevitch C. des 30000 francs comptant sur le prix de vente par les h ^{rs} Ledrille, répertorié n ^o 28 qui précède.	8	3.75
		Verifié le présent répertorie cent vingt huit actes, - d.c.r.	contenait depuis le dernier visa l'inscript ^{ion} de S ^r Paris le 7 janvier 1920, folio 17. Alsoacome		
31	7	Cat. devic	Tessereau, de Louis Jules dument à Gains de Bondy, 66.	8	1.88

Reconnaissance d'écriture manuscrite

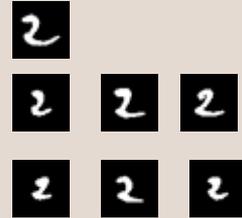
HTR et Classification

1. Classification par reconnaissance de formes
 - Colonnes de chiffres ou de nombres
 - Dates
 - N° d'actes
 - Taxes
 - Colonnes de texte « simple » des types d'actes
2. HTR : Obtention de données d'entraînement pour construire un modèle de reconnaissance
 - Transcription collaborative d'une cinquantaine de pages avec *Transkribus* d'un répertoire de l'étude Marotte (importation des images numérisées, segmentation, transcription cellule par cellule)

Classification

Classifications par reconnaissance de formes

- Chiffres (notamment numéros de jour) : 98,66 % de précision (pour les chiffres de 0 à 9) grâce à un algorithme de classification basé un réseau de neurones convolutif



Classification

Typologies d'actes : Premier travail de classification manuelle à poursuivre. Difficultés :

- Hétérogénéité des formes
- Typologies multiples pour un acte
- Typologie absente (quand un acte est la "suite" d'un premier acte)

Problèmes	Forme "normalisée"	Variation(s)	
Formes hétérogènes	Contrat conditionnel	<i>Ct conditionel</i>	<i>Contt condel</i>
	Contrat de mariage	<i>Ct de mariage</i>	<i>Ct de mge</i>
Types multiples	Liquidation	<i>Liquidation et partage</i>	<i>Liquidation et présentation de compte de tutelle</i>
Acte faisant suite à un 1 ^{er}	Suite de l'acte du 15 décembre 1919	<i>Ste 15 xbr 1919</i>	

HTR: Entraînement et premiers résultats

- Entraînement d'un modèle HTR à partir des pages transcrites manuellement "*vérité terrain*" (un seul scribe) :
 - M1 : 40 pages
 - M2 : 50 pages (1 million de mots)
- Character Error Rate (CER) sur un échantillon test du même répertoire:
 - M1 : **13,5 %**
 - M2 : **10,4 %**

⇒ ce qui reste considérable, mais peut être amélioré
- Le modèle M2 a été testé sur quelques pages d'autres registres et donc d'autres mains. Les résultats ne sont naturellement pas bons (CER autour de 40 %) : nécessité de données d'entraînement plus hétérogènes.

Défis de la phase 2

- Acquisition d'autres données d'entraînement pour d'autres mains, d'autres périodes et créer un modèle mixte
- Passage en production
 - ⇒ Envisager une plateforme de transcription collaborative
 - ⇒ Collaboration avec le projet SCRIPTA-PSL (transcription automatique de manuscrits historiques multilingues)
 - ⇒ Utilisation du logiciel d'HTR [Kraken](#) (Benjamin Kiessling, ALMAAnaCH) pour plus de contrôle et d'autonomie
- Utilisation des données sémantiques
 - ⇒ Référentiels spécifiques au corpus : typologie des actes, terminologie notariale
 - ⇒ Référentiels d'autorités pour les entités nommées (personnes et lieux : utilisation d'annuaires)
- Expérimentation sur les répétitions (word spotting)