



**HAL**  
open science

## Espionner les espions

Yvan Jauregui, Simon Gabay

► **To cite this version:**

Yvan Jauregui, Simon Gabay. Espionner les espions: Espionner les espions: cartographier les représentants des puissances étrangères à Paris d'après les rapports de police du XVIIIe siècle. *Humanistica* 2024, Association francophone des humanités numériques, May 2024, Meknès, Maroc. hal-04559418

**HAL Id: hal-04559418**

**<https://hal.science/hal-04559418>**

Submitted on 25 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Espionner les espions : cartographier les représentants des puissances étrangères à Paris d'après les rapports de police du XVIII<sup>e</sup> siècle

Yvan Jauregui et Simon Gabay

Université de Genève

prenom.nom@unige.ch

## Résumé

Pourtant omniprésents dans les villes de l'époque moderne, les étrangers apparaissent comme un objet historique difficilement saisissable. Les archives de la bibliothèque de l'Arsenal s'illustrent alors comme une porte d'accès précieuse, notamment grâce aux rapports de surveillance de la police parisienne qui dressent le portrait des populations allochtones séjournant dans la capitale entre 1725 et 1767. Ce matériau permet, à travers l'observation des diplomates, l'étude de la perception institutionnelle des étrangers, de la sociabilité diplomatique et du rôle de la surveillance policière au XVIII<sup>e</sup> siècle. Toutefois, cette remarquable documentation soulève des difficultés de traitement (grande quantité, complexité de lecture...) et d'analyse (lecture distante...). Afin de pallier ces problèmes, cet article présente une preuve de concept, partant de l'OCRisation des manuscrits jusqu'à la production de cartes remplaçant les agents diplomatiques selon les États qu'ils représentent.

## 1 Introduction

Qualifiées par le physicien Pierre Bertholon (1741-†1800) de « gouffres habités [...] où des millions d'hommes vont se précipiter » (Bertholon, 1786, p. 3), les villes européennes connaissent entre le XVII<sup>e</sup> et le XVIII<sup>e</sup> siècle une phase d'expansion sans précédent, conséquence de la plus grande mobilité des populations provoquée entre autres par le considérable essor démographique. Les centres urbains sont ainsi marqués par l'affluence toujours grandissante d'étrangers<sup>1</sup>. Au sein de ces derniers se retrouvent des migrants, mais également une myriade d'individus de passage,

1. Ce terme recoupe dans son acception ancienne tout individu allochtone, c'est-à-dire extérieur à la cité (Milliot, 2006).

qu'ils soient étudiants, marchands, diplomates ou simples voyageurs (Dubost, 2015).

Saisir leur présence n'est cependant pas chose aisée<sup>2</sup>. Paris fait alors figure d'observatoire exceptionnel grâce à la grande richesse de ses archives<sup>3</sup>, notamment policières. En effet, la Lieutenance générale de police, fondée en 1667, génère un nombre conséquent de documents alors centralisés et conservés derrière les imposants murs de la Bastille<sup>4</sup>. Nécessaire dès lors que la Cour s'éloigne de la capitale, l'accroissement de cette production écrite atteint son apogée entre 1747 et 1757 sous la direction de Nicolas-René Berryer (1703-†1762), qui réorganise la police en établissant des départements et bureaux spécialisés. Toutefois, et malgré son « goût de l'archive », l'institution croule bientôt sous la prolifération des documents qu'entraîne son « ambition totalisante » (Denis, 2019).

Façonnant la « mémoire policière », cette documentation écrite est le témoin de la modification « des relations entre la monarchie et la société » (Denis, 2019) et le résultat de l'évolution des fonctions de la police, parmi lesquelles la préservation de la sûreté publique s'affirme progressivement jusqu'à devenir prééminente au XVIII<sup>e</sup> siècle (Milliot, 2006). Afin de maintenir cette sûreté, la police intensifie le contrôle social des populations perçues comme dangereuses, c'est-à-dire les individus marginaux et/ou non-natifs de la capitale (Milliot, 2015). Parmi eux, les diplomates font l'objet de filatures incessantes, par peur de l'espionnage ou du complot (Bély, 1990, 2007).

2. Par exemple, voir : Dubost 1997, Roche 2000, Roche 2002, Kaiser 2009, Cerutti 2012, Jahan 2016.

3. Se référer au guide des recherches consacré aux étrangers à Paris entre le XVI<sup>e</sup> et 1789 : Dubost 1993.

4. Sur l'histoire de ce fonds et les difficultés d'archivage que sa taille considérable a posées dès le XVIII<sup>e</sup> s. : Lacour et Denis 2016, Denis 2019, Jacquet 2022 et, surtout, la thèse de cette dernière (Jacquet, en cours).

Institué en 1740, le sixième bureau de la lieutenance est ainsi chargé de la sûreté publique, et donc du suivi des étrangers. Des doutes persistent toutefois quant à l'existence plus ancienne d'un département de la sûreté (Couture, 2013). En effet, les archives de ce bureau – du moins celles qui subsistent et qui concernent cette surveillance – débutent en 1725<sup>5</sup>. Elles sont aujourd'hui conservées à Paris, au sein de la bibliothèque de l'Arsenal.

L'objectif est, par l'étude de ces documents, l'étude de la perception institutionnelle des étrangers, de la sociabilité diplomatique et du rôle de la surveillance policière au XVIII<sup>e</sup> siècle. Pour ce faire, il convient dans un premier temps d'extraire le texte des manuscrits, puis d'y repérer les individus pertinents. Il s'agit ensuite de les replacer sur une carte selon les États qu'ils représentent afin de jeter un regard neuf et distant, le seul possible étant donné le volume de données à analyser.

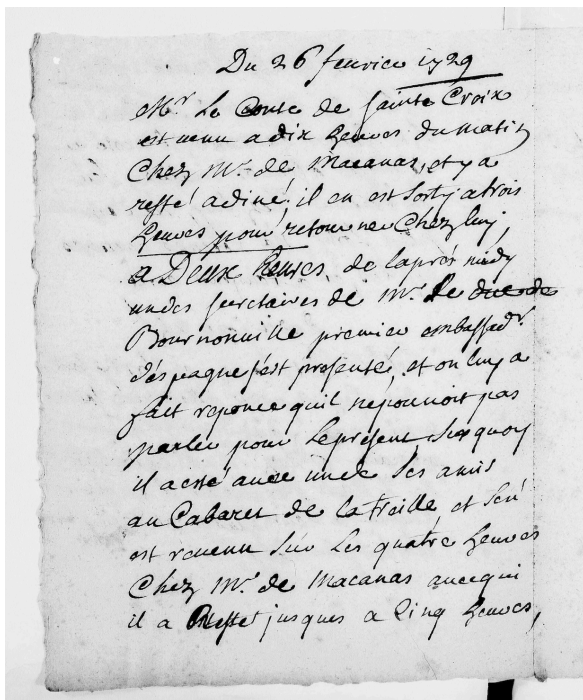


FIGURE 1 – Exemple de numérisation disponible sur Gallica. Paris, Arsenal, Ms-10283, f. 295<sup>v</sup>.

## 2 État de l'art

Plusieurs projets se sont intéressés à la numérisation des archives parisiennes, mais l'Ancien Régime n'a pas encore reçu l'attention

5. « Surveillance des étrangers : rapports envoyés au Lieutenant de police par ses officiers et ses agents » (Paris, BnF, Arsenal, Ms-10283-10292, 1725-1767).

qu'il mérite. En effet, pour la partie basse de la chronologie, Himanis s'est consacré aux documents de la chancellerie royale française (1302-1483, cf. Stutzmann et al. 2017). De l'autre côté du spectre, LECTAUREP s'est attelé aux archives du Minutier central des notaires de Paris entre 1803 et 1944 (Chagué et Rostaing, 2021).

Étant donnée l'importante quantité de données à traiter, ces deux projets se sont distingués par la conception de modèles HTR qui, malheureusement, ne couvrent pas les besoins de cette étude. Le logiciel Transkribus (Kahle et al., 2017) propose un modèle pour les mains modernes<sup>6</sup> offrant un taux d'erreur caractère de 7,8%, mais son utilisation est payante. Le moteur OCR Kraken offre, quant à lui, un premier modèle pour les cursives françaises modernes et contemporaines (Chagué et Clérice, 2022). Dénommé Manu McFrench<sup>7</sup>, son efficacité n'a pas encore été mesurée, mais un essai a confirmé son inefficacité sur le corpus de sources de l'Arsenal, vraisemblablement à cause du faible nombre d'archives modernes contenues dans son jeu d'entraînement.

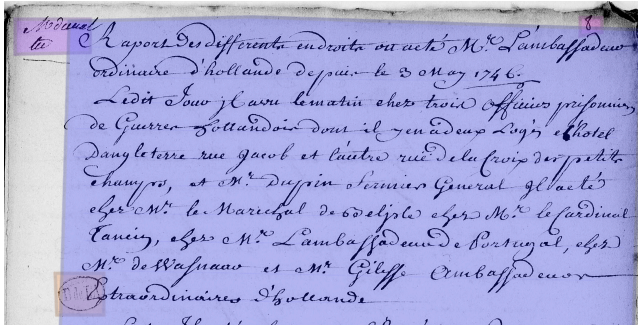
La solution la plus efficace est donc de produire un modèle, en utilisant les jeux de données publiés en ligne comme ceux répertoriés par le projet HTR-United<sup>8</sup> (Chagué et Clérice, 2023). Les données élaborées par les projets CREMMA (Clérice et al., 2022b; Chagué et al., 2022; Clérice et al., 2022a) et HTR-Romance (Norindr et Chagué, 2023) sont certainement les plus importants, avec ceux de LECTAUREP (Limon-Bonnet et Chagué, 2022), des étudiant·e·s de l'École nationale des chartes (Ceard et al., 2021), de GenAuto (Boutet et Merx, 2021) ou encore celui de l'édition de la correspondance de Constance de Salm (1767-†1845, Biay et al. 2022).

Concernant la cartographie numérique, de nombreux projets se sont penchés sur la reconnaissance des entités nommées du français de la période moderne. Si de bons résultats sont obtenus avec des mécanismes *rule based* (Katherine McDonough et van de Camp, 2019; Moncla et Gaio, 2023), ces outils ont délaissé les anthroponymes en s'intéressant principalement aux toponymes. De récentes recherches

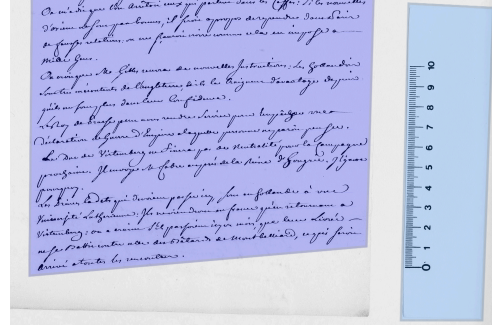
6. Modèle N°37758.

7. <https://zenodo.org/records/6657809>

8. <https://htr-united.github.io>



(a) NumberingZone (rose), MarginTextZone (mauve), MainZone (violet), StampZone (orange). Paris, Arsenal, Ms-10292, f. 8<sup>r</sup>.



(b) DigitizationArtefactZone (bleu), MainZone (violet). Paris, Arsenal, Ms-10292, f. 2<sup>r</sup>.

FIGURE 2 – Exemple d’analyse de mise en page.

proposent néanmoins des modèles couvrant un spectre plus large d’entités et emploient des méthodes neuronales (Ortiz Suarez et Gabay, 2022) qui répondent davantage aux besoins présents.

Un cas similaire regarde la question de la désambiguïsation : l’essentiel du travail a été effectué sur les noms de lieux et non ceux de personnes. Ainsi, le projet DicoTopo<sup>9</sup> offre de belles perspectives pour les localités d’Ancien Régime, mais demeure incomplet, notamment concernant la ville de Paris qui est importante pour cette étude. D’autres projets paraissent également intéressants, mais ne couvrent que la période révolutionnaire (Keller et al., 2018).

### 3 Préparation des données d’entraînement

En l’absence de tout travail de fond sur les données françaises de l’époque, la décision a été prise de préparer des données pour la création de modèles.

#### 3.1 Corpus

Le corpus de travail pour l’entraînement du modèle d’OCR se compose de six manuscrits appartenant aux « rapports envoyés au lieutenant de police par ses officiers et agents », conservés à la bibliothèque de l’Arsenal et dont le master des microfilms est disponible sur Gallica (cf. fig. 1 pour un exemple). Un échantillonnage a été effectué selon, d’une part, le nombre de folios et, d’autre part, la représentativité des mains<sup>10</sup> (cf. tab. 1).

9. <https://dicotopo.cths.fr>.

10. Sur les 116 folios retenus, il est possible de dénombrer plus d’une vingtaine de scripteurs dont Germain-

Cote	Nombre de folios	Échantillon
Ms-10283	324	23
Ms-10285	361	10
Ms-10287	490	25
Ms-10289	458	25
Ms-10290	268	20
Ms-10292	269	13
Total	2 170	116

TABEAU 1 – Corpus de travail et échantillons retenus pour l’entraînement du modèle d’OCR.

#### 3.2 Extraction : analyse de mise en page

L’analyse de la mise en page a été exécutée grâce à l’application *open source* e-Scriptorium (Kiessling et al., 2019), dont l’Université de Genève possède une instance (Gabay et al., 2021-), en suivant les préconisations du projet SegmOnto<sup>11</sup> afin de permettre l’interopérabilité des données (Gabay et al., 2023b).

Les folios ne possédant pas de mise en page complexe, cinq types de zones ont été retenues (cf. fig. 2) :

- DigitizationArtefactZone pour les mires, règles, etc. ;
- MainZone pour le corps du texte ;
- MarginTextZone pour les intertitres et annotations ;
- NumberingZone pour la foliotation ;
- StampZone pour les tampons de la BnF.

Louis Chauvelin, Nicolas Odile de Pommereuil, Jean de Chantepie, Jean Poussot, Claude-Henry Feydeau de Marville...

11. <https://segmonto.github.io>.

Catégorie	Cas	Traitement	exemple	image
Agglutination	Segmentation imprécise	Un seul mot	<i>dela</i>	
Agglutination	Pas d'apostrophe	Respect du système	<i>dhollande</i>	
Abréviation	Nasale	Tilde	<i>nom̃e</i>	
Abréviation	Suscrite	Précédé de ^	<i>1000<sup>^</sup>lt</i>	
Alphabet	s long (⟨f⟩)	s rond (⟨s⟩)	<i>mareschal</i>	
Alphabet	⟨u⟩ consonne	Respect du système	<i>uenu</i>	
Alphabet	⟨j⟩ voyelle	Respect du système	<i>Ljncendie</i>	
Correction	Rature lisible	Entre doubles crochets	<i>[[on croit que]]</i>	
Correction	Rature illisible	Entre doubles crochets (avec un point par lettre)	<i>[[s..]]</i>	
Fin de ligne	Trait de conduite	U+23AF (↵)	<i>cheuaux —</i>	
Fin de ligne	Tiret(s) de fin	U+00AC (↵)	<i>domes↵</i>	

TABLEAU 2 – Principales règles de transcription

Les lignes ont, quant à elles, été annotées selon trois catégories (cf. fig. 3) :

- `HeadingLine` pour les lignes mises en valeur (centrées, alignées à droite, etc.) ;
- `DefaultLine` pour la ligne normale ;
- `InterlinearLine` pour les ajouts, corrections, etc.

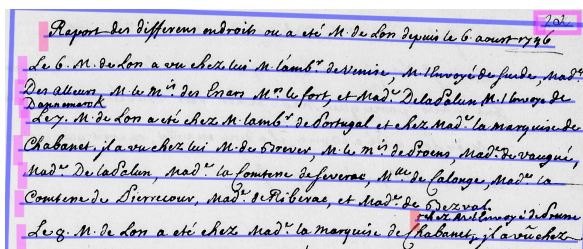


FIGURE 3 – Catégories retenues pour les lignes. `HeadingLine`, `DefaultLine`, `InterlinearLine`. Paris, Arsenal, Ms-10292, f. 202<sup>r</sup>.

### 3.3 Extraction : Reconnaissance optique de caractères

Différents guides pour la transcription de manuscrits ont été rédigés à ce jour, mais se focalisent essentiellement sur les documents médiévaux (Pinche, 2022; Clérice et al., 2023) ou de la Renaissance (Solfrini et al., 2023). Concernant la période moderne, il n'existe que des recommandations pour les imprimés (Gabay et al., 2023a). Les lignes suivantes pro-

posent ainsi d'émettre quelques règles simples qui se veulent, dans la limite du possible, interopérables avec les autres guides et états du français (cf. tab. 2).

L'essentiel de la transcription est graphématique (Stutzmann, 2011) : le nombre d'interventions demeure le plus limité possible, tout en réduisant chaque forme à son sens dans le système alphabétique du français. Les abréviations sont conservées (*nom̃e* pour *nomme*), et celles suscrites sont indiquées comme telles grâce au signe  $\wedge$ . Les majuscules sont également préservées, y compris lorsqu'elles ne seraient plus utilisées en français contemporain : aucune normalisation n'est introduite (cf. fig. 4). À propos de la segmentation des mots, elle est retouchée le moins possible : les apostrophes qui marquent l'élision ne sont pas réintroduites (*dhollande*), mais les levés de plume en milieu de mots ne sont pas conservés

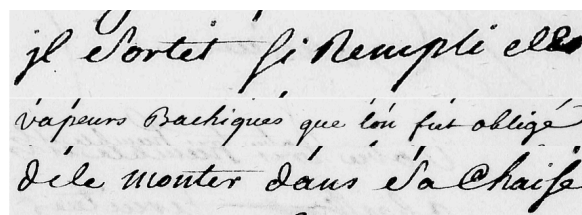


FIGURE 4 – Exemple de préservation des majuscules. « il sortit si Rempli des vapeurs Bachiques que l'on fut obligé de le monter dans sa Chaise », Paris, BnF, Arsenal, Ms-10283, f. 203<sup>r</sup>.

(*d ela* → *dela*). Les variantes allographiques, en revanche, ne sont pas sauvegardées (*s* long) *a contrario* des emplois anciens comme le ⟨u⟩ pour le son consonne et le ⟨j⟩ pour le son voyelle.

## 4 Extraction et analyse

### 4.1 Ajustement du modèle OCR et inférence

La vérité de terrain produite a permis de procéder à l'« ajustement » (angl. *fine tune*) du modèle HTR-United - ManuMc Fondue (Gabay et al., 2024a)<sup>12</sup>. Alors que de précédents tests (Gabay et al., 2024b) ont déterminé une précision de 85.27%, le nouveau modèle offre un score de plus de sept point de pourcentage supérieur (92.66%), établissant l'état de l'art. Il est distribué sous le nom d'*Ivanhoé* (Gabay et Jauregui, 2024).

Le modèle est alors appliqué sur le jeu de données, pour un total d'environ 4 000 pages (cf. tab. 3).

Ms.	Date	Vues
Ms-10283	1725, 1729	346
Ms-10284	1730	339
Ms-10285	1731	247
Ms-10286	1732-1740	338
Ms-10287	1741	512
Ms-10289	1744	545
Total		3 960

TABLEAU 3 – Liste et détail des manuscrits OCRisés.

### 4.2 Extraction de l'information

À partir de ce corpus, la trace de nombreux diplomates a été retrouvée comme celle d'Annibale Carlo (1666-†1735), comte de Maffei et serviteur des États de Savoie, d'Álvaro José Navia-Osorio y Vigil de Quiñones (1684-†1732), marquis de Santa Cruz (nommé « Comte de Sainte Croix » dans le manuscrit) et missionné par le Royaume d'Espagne, du duc Léopold-Philippe d'Arenberg et d'Arshot (1690-†1754), natif de Bruxelles alors dans les Pays-Bas espagnols et employé par le Saint-Empire romain germanique, ou encore de Giu-

12. Ce modèle a été créé en parallèle de cet article, d'où son absence dans le présent état de l'art.

lio Franchini Taviani (1694-†1759) qui œuvre pour le Grand-duché de Toscane.

Pour ce faire, une recherche d'entités nommées a été effectuée d'après le fichier exporté en .txt des documents OCRisés. Le taux d'erreur mot demeurant élevé, les outils de TAL mentionnés dans l'état de l'art (cf. § 2) ont démontré de trop faibles performances et ont été écartés : l'essentiel du travail a donc dû être mené avec des expressions régulières.

Les mots clefs privilégiés ont été les charges et fonctions – « ambassadeur », « envoyé », « ministre », « chargé d'affaire », etc. – et les toponymes accolés aux individus – « Angleterre », « Provinces-Unies, Hollande », « Espagne », « (Saint-)Empire (romain germanique) », « Wittenberg », « Venise », « Suède », « Pologne », etc. Davantage que sur les vocables complets, la recherche s'est concentrée sur des séquences de quelques caractères afin de contourner le problème des abréviations comme de pallier les erreurs de la transcription automatique et ainsi avoir une meilleure efficacité. Par exemple, les séquences « bas|f » et « f|sad » – « f » correspondant à un possible ⟨f⟩ – pour « ambassadeur », ou « voy|i » et « uoy|i » pour « envoyé ».

### 4.3 Cartographie

Sur cette base débute le relevé. Celui-ci concerne les individus et non les occurrences afin d'éviter les répétitions qui fausseraient le décompte : chaque acteur est comptabilisé une fois par manuscrit.

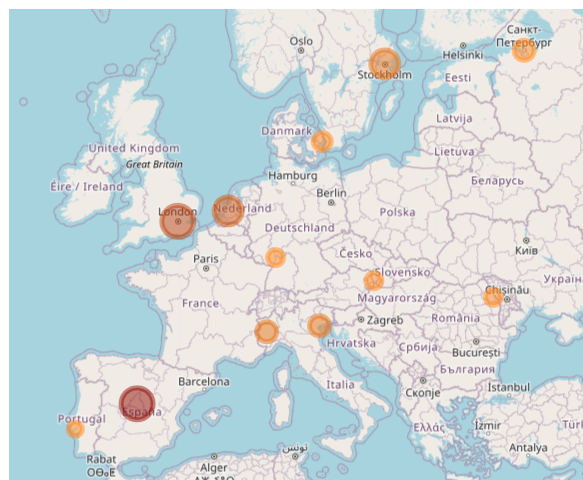


FIGURE 5 – États représentés par les diplomates filés (1725, 1729-1741, 1744)

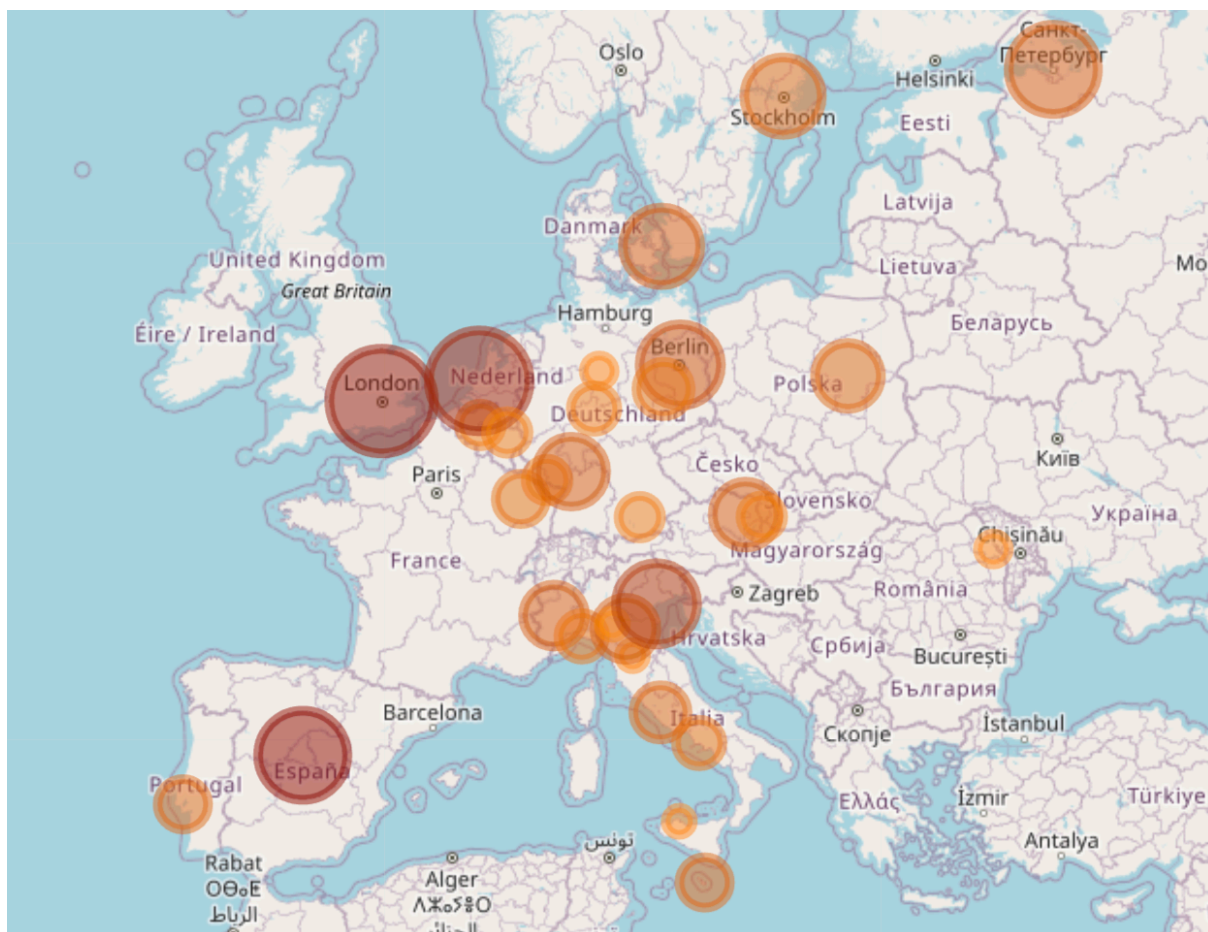


FIGURE 6 – États représentés par les diplomates rencontrés (1725, 1729-1741, 1744)

Après avoir retrouvé les États qu'ils représentent, des coordonnées géographiques – celles des capitales – leur ont été associées en vue de former des cartes. La première correspond aux diplomates filés (cf. fig. 5) et la seconde, aux agents diplomatiques rencontrés au détour du suivi d'un acteur tiers, comme d'un autre diplomate ou d'un noble parisien/ne/allochtonne (cf. fig. 6). Le rayon des cercles qui s'élargit, comme les couleurs qui se foncent, coïncident avec la quantité des individus associés à chaque point de la carte.

## 5 Discussion

Il peut paraître étonnant, de prime abord, de retrouver autant d'individus de la haute société dans des archives policières, mais il était courant que le souverain ou l'un de ses ministres exigent la filature de diplomates étrangers (cf. fig. 7), et ce pour de multiples raisons, dont la crainte de l'espionnage, d'alliances se préparant en secret ou simplement afin de se

renseigner sur les avancées ou les retombées politiques de conférences internationales. Ce dernier cas de figure explique probablement le suivi renforcé d'Espagnols en 1729, notamment celui du marquis de Santa Cruz et de Michel-Joseph duc de Bournonville et Saint-Algonde (1670-†1752), tous deux plénipotentiaires lors du Congrès de Soissons (1728-1730, Dhondt 2022).

Avec ce premier relevé, bien qu'encore très parcellaire, il est possible d'observer les effets d'une proximité socio-économique et culturelle : issus de milieux semblables, les diplomates sont enclins à établir des liens avec l'élite parisienne. Ceux-ci s'inscrivent dans des rapports de travail, d'amitié, voire d'intimité, au point de tisser des relations familiales (Lilti, 2005). Se déroulant autant dans les promenades qu'au sein de leurs hôtels, leurs fréquentes rencontres indiquent l'existence d'une sociabilité internationale, qui s'ajoute aux réseaux traditionnels. L'exercice de la sociabilité

et de la mondanité se mue en un des devoirs des diplomates (Bély, 1990 ; Lilti, 2005). Toutefois, ce sont ces « liaisons dangereuses » qui inquiètent les autorités, justifiant à leurs yeux le déploiement d'un suivi policier.

Les cartes produites témoignent alors de l'intensité de l'activité diplomatique européenne dans laquelle Paris détient une place centrale. En tant que pôle rayonnant de la mondanité, elle apparaît comme une des destinations privilégiées des diplomates<sup>13</sup> (Belissa et Schnakenbourg, 2010). Paradoxalement, alors que leur regroupement facilite les échanges et concertations entre les différentes puissances, il permet, en contrepartie, à la police et à ses mouches d'exercer une surveillance des actions et décisions politiques internationales. À travers le nombre d'agents filés, il est possible de remarquer le degré de préoccupation de la monarchie française envers les autres États européens. Sans surprise, ce sont les grandes puissances continentales qui inquiètent davantage : le Royaume d'Espagne, l'Empire britannique, la République des Provinces-Unies ou le Royaume de Suède (cf. fig. 5).

Constat étonnant : le personnel du Saint-Empire romain germanique semble presque absent de cette surveillance. Pourtant, dans le cadre de la guerre de succession d'Autriche (1740-1748), le Royaume de France est en conflit ouvert avec plusieurs des princes-électeurs de l'Empire ainsi qu'avec la monarchie des Habsbourg d'Autriche<sup>14</sup>. Or, curieusement – ou significativement –, les Espagnols, alors alliés de Louis XV (1710-†1774), figurent parmi les diplomates les plus surveillés. Est-ce un biais inhérent aux sources ou une réalité ? Seule une analyse approfondie permettra de l'éclaircir.

## 6 Pistes

Ce prototype de chaîne de traitement semble démontrer une efficacité certaine. Il convient désormais de perfectionner le modèle d'OCR<sup>15</sup>, d'agrandir le relevé et d'automati-

13. Sur la remise en question de l'hégémonie parisienne et française sur l'Europe, voir : [Beaurepaire 2008](#).

14. Sur l'action des diplomates français dans le Saint-Empire romain germanique, voir : [Bély 2007](#).

15. D'autant plus qu'une meilleure détection des graphies contenues dans les archives de la police parisienne aurait de nombreux mérites. En effet, ces fonds « ont

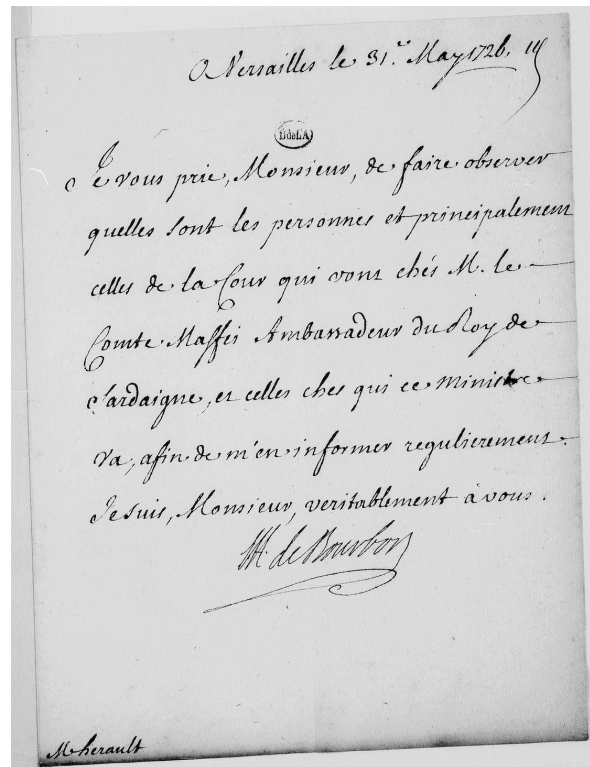


FIGURE 7 – Ordre donné au lieutenant général de police René Hérault par Louis Henri de Bourbon, prince de Condé et principal ministre d'État de Louis XV entre 1723 et 1726. « Je vous prie, Monsieur, de faire observer quelles sont les personnes et principalement celles de la Cour qui vont chez M. le Comte Maffei Ambassadeur du Roy de Sardaigne, et celles chez qui ce ministre va ; afin de m'en informer régulièrement. », Paris, BnF, Arsenal Ms-10283, f. 115<sup>r</sup>.

ser autant que possible son exécution avant d'analyser finement les données qui seront produites en les replaçant dans leur(s) contexte(s) respectif(s).

Concernant les données, les recueils de l' Arsenal sont loin d'être les seuls exploitables qui subsistent aujourd'hui. Les Archives diplomatiques de la Courneuve conservent également 95 manuscrits contenant les rapports du contrôle des étrangers que faisait parvenir le lieutenant général de police au secrétaire d'État aux Affaires étrangères<sup>16</sup>. Cependant, ces documents ne sont toujours pas disponibles au format numérique : peut-être le seront-ils prochainement ?

vocation à intéresser le gouvernement de l'ensemble du royaume, dès la fin du XVII<sup>e</sup> siècle, ce qui les place en pratique au rang des archives centrales de la monarchie » (Denis, 2019, p. 183).

16. Paris, Archives diplomatiques, Courneuve, 1QO/1-94 (1771-1791). Pour une analyse qualitative de ce corpus, du moins concernant les diplomates et les salons, se référer à [Lilti 2005](#).



## Remerciements

Merci à Marie Houllémare, Fabrice Brandli et Mathieu Nicati pour leurs conseils avisés.

## Bibliographie

- Pierre-Yves Beaurepaire. 2008. *Le mythe de l'Europe française au XVIIIe siècle. Diplomatie, culture et sociabilités au temps des Lumières*. Autrement, Paris.
- Marc Belissa et Eric Schnakenbourg. 2010. **Les circulations diplomatiques en Europe au XVIIIe siècle : représentation, information, diffusion des modèles culturels**. In Pierre-Yves Beaurepaire et Pierrick Pourchasse, éditeurs, *Les circulations internationales en Europe, années 1680-années 1780*, pages 33–49. Presses universitaires de Rennes, Rennes.
- Pierre Bertholon. 1786. *De la salubrité de l'air des villes et en particulier des moyens de la procurer ; [...]*. impr. de J. Martel aîné, Montpellier.
- Sébastien Biay, Alix Chagué, et Thibault Clérice. 2022. **Éditer la correspondance de Constance de Salm (1767-1845)**.
- Jean-François Boutet et Jean-Pierre Merx. 2021. **Genauto td corpus**.
- Lucien Bély. 1990. *Espions et ambassadeurs au temps de Louis XIV*. Fayard, Paris.
- Lucien Bély. 2007. *L'Art de la paix en Europe. Naissance de la diplomatie moderne, XVIe-XVIIIe siècle*. Presses universitaires de Rennes, Rennes.
- Lien Ceard, Cécile Sajdak, et Fanny Lebreton. 2021. **Projet correspondance Berlioz**.
- Simona Cerutti. 2012. *Etrangers. Etude d'une condition d'incertitude dans une société d'Ancien Régime*. Bayard, Paris.
- Alix Chagué et Thibault Clérice. 2023. **"I'm here to fight for ground truth" : HTR-United, a solution towards a common for HTR training data**. In *Digital Humanities 2023 : Collaboration as Opportunity*, Graz, Austria. Alliance of Digital Humanities Organizations and University of Graz.
- Alix Chagué et Aurélia Rostaing. 2021. **LECTAUREP : Lecture Automatique des Répertoires de Notaires Parisiens**. In *Fantastic Futures 2021 / Futures Fantastiques 2021*, Paris, France. AI4LAM and BnF and Université Paris Saclay.
- Alix Chagué et Thibault Clérice. 2022. **Htr-United - manu mcfrench v1 (manuscripts of modern and contemporaneous french)**.
- Alix Chagué, Thibault Clérice, Jade Norindr, Elsa Falcoz, et Margaux Faure. 2022. **CREMMA - a repository of 18th century manuscripts**.
- Thibault Clérice, Malamatenia Vlachou-Efstathiou, et Alix Chagué. 2023. **CREMMA Medii Aevi : Literary manuscript text recognition in Latin**. *Journal of Open Humanities Data*, 9 :4.
- Thibault Clérice, Alix Chagué, Baudouin Davoury, Soline Doat, Margaux Faure, et Maxime Humeau. 2022a. **CREMMA - a repository of 19th century manuscripts**.
- Thibault Clérice, Alix Chagué, Margaux Faure, Jade Norindr, Anais Mazoue, et Baudouin Davoury. 2022b. **CREMMA - a repository of 17th century manuscripts**.
- Rachel Couture. 2013. *"Inspirer la crainte, le respect et l'amour du public" : les inspecteurs de police parisiens, 1740-1789*. Thèse de doctorat, Université du Québec à Montréal et Université Caen Basse-Normandie, Montréal et Caen. Vol. 1.
- Vincent Denis. 2019. **Quand la police a le goût de l'archive : réflexions sur les archives de la police parisienne au XVIIIe siècle**. In Maria Donato et Anne Saada, éditeurs, *Pratiques d'archives à l'époque moderne. Europe, mondes coloniaux*, pages 183–203. Garnier, Paris.
- Frederik Dhondt. 2022. **"Bringing the divided powers of Europe nearer one another" the congress of Soissons, 1728-1730**. *Nuova Antologia Militare*, III :535–642.
- Jean-François Dubost. 1993. *Les étrangers en France XVIe siècle-1789. Guide de recherches aux archives nationales*. Archives nationales, Paris.
- Jean-François Dubost. 1997. *La France italienne, XVIe-XVIIe siècle*. Aubier, Paris.
- Jean-François Dubost. 2015. **Naissance de la police des étrangers dans le royaume de France (XVIe-XVIIIe siècle)**. In Marie-Claude Blanc-Chaléard, Caroline Douki, Nicole Dyonet, et Vincent Milliot, éditeurs, *Police et migrants. France 1667-1939*, pages 33–49. Presses universitaires de Rennes, Rennes.
- Simon Gabay, Alix Chagué, et Thibault Clérice. 2024a. **Htr-United - manu mc fondue (manuscripts of modern and contemporaneous french - manu mcfrench v4)**.
- Simon Gabay, Robin Champenois, Pierre Kuenzli, Jean-Luc Falcone, et Christophe Charpillot. 2021-. **Formes Numérisées et Détection Unifiée des Écritures (FoNDUE)**.
- Simon Gabay, Thibault Clérice, et Christian Reul. 2023a. **OCR17 : Ground Truth and Models for 17th c. French Prints (and hopefully more)**. *Journal of Data Mining and Digital Humanities*, 2023.
- Simon Gabay et Yvan Jauregui. 2024. **Ivanhoé - ocr model for Parisian archives of the 18th c.**
- Simon Gabay, Ariane Pinche, Kelly Christensen, et Jean-Baptiste Camps. 2023b. **SegmOnto : A Controlled Vocabulary to Describe and Process Digital Facsimiles**. Document de travail.

- Simon Gabay, Ariane Pinche, Peter Nahon, Alix Chagué, Pauline Jacsont, Élodie Paupe, Jean-Claude Rebetz, Maxime Humeau, Christine Payot, Thibault Maillard, Yvan Jauregui, Elina Leblanc, et Loraine Chappuis. 2024b. Vers un modèle diachronique pour les mains modernes françaises. In *Humanistica 2024*, Meknès, Maroc.
- Marie-Elisabeth Jacquet. 2022. Vie et mort d'un dépôt d'archives. les archives "de la bastille" dans les années 1780. *Crime, Histoire & Sociétés*, 1(16).
- Marie-Elisabeth Jacquet. en cours. *Les Archives de la Lieutenance générale de police de Paris (1716-1789-1889)*. Thèse de doctorat, Université Paris 8 | Vincennes - Saint-Denis.
- Sébastien Jahan. 2016. *Les étrangers en Poitou au XVIIIe siècle. Traverser ou rester*. Geste éditions, La Crèche.
- Philip Kahle, Sebastian Colutto, Günter Hackl, et Günter Mühlberger. 2017. Transkribus - a service platform for transcription, recognition and retrieval of historical documents. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 04, pages 19–24.
- Wolfgang Kaiser. 2009. Extranéités urbaines à l'époque moderne. In Pilar González-Bernaldo, Manuela Martini, et Marie-Louise Pelus-Kaplan, éditeurs, *Étrangers et Sociétés. Représentations, coexistences, interactions dans la longue durée*, pages 77–86. Presses universitaires de Rennes, Rennes.
- Ludovic Moncla Katherine McDonough et Matje van de Camp. 2019. Named entity recognition goes to old regime france : geographic text analysis for early modern french corpora. *International Journal of Geographical Information Science*, 33(12) :2498–2522.
- Antoine Keller, Nathalie Abadie, Bertrand Dumenuieu, Stéphane Baciocchi, et Eric Kergosien. 2018. Vers la construction d'une base de connaissances sur la réorganisation territoriale française à la Révolution. In *Conférence Sagéo 2018 Atelier Exces*, Actes de l'atelier Exces, conférence Sageo 2018, Montpellier, France.
- Benjamin Kiessling, Robin Tissot, Peter Stokes, et Daniel Stökl Ben Ezra. 2019. eScriptorium : An open source platform for historical document analysis. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, volume 2.
- Pierre-Yves Lacour et Vincent Denis. 2016. La logistique des savoirs. surabondance d'informations et technologies de papier au XVIIIe siècle. *Genèses*, 1(102) :107–122.
- Antoine Lilti. 2005. *Le monde des salons. Sociabilité et mondanité à Paris au XVIIIe siècle*. Fayard, Paris.
- Françoise Limon-Bonnet et Alix Chagué. 2022. *No-taires de Paris – Bronod*.
- Vincent Milliot. 2006. Réformer les polices urbaines au siècle des lumières : le révélateur de la mobilité. *Crime, Histoire & Sociétés*, 10(1) :25–50.
- Vincent Milliot. 2015. Migrants et "étrangers" sous l'œil de la police : la surveillance des lieux d'accueil parisiens au siècle des lumières. In Marie-Claude Blanc-Chaléard, Caroline Douki, Nicole Dyonet, et Vincent Milliot, éditeurs, *Police et migrants. France 1667-1939*, pages 315–331. Presses universitaires de Rennes, Rennes.
- Ludovic Moncla et Mauro Gaio. 2023. *Perdido : Python library for geoparsing and geocoding French texts*. In *First International Workshop on Geographic Information Extraction from Texts (GeoExT)*, Dublin, Irlande.
- Jade Norindr et Alix Chagué. 2023. *Corpus modern roman languages*.
- Pedro Ortiz Suarez et Simon Gabay. 2022. A Data-driven Approach to Named Entity Recognition for Early Modern French. In *Proceedings of the 29th International Conference on Computational Linguistics*, Gyeongju, South Korea. International Committee on Computational Linguistics.
- Ariane Pinche. 2022. *Guide de transcription pour les manuscrits du Xe au XVe siècle*. Document de travail.
- Daniel Roche. 2000. *La ville promise. Mobilité et accueil à Paris (fin XVIIe-début XIXe siècle)*. Fayard, Paris.
- Daniel Roche. 2002. L'étranger insaisissable (xviii-xviii siècles). *Mélanges de l'école française de Rome*, 114(2) :837–874.
- Sonia Solfrini, Simon Gabay, Geneviève Gross, Pierre-Olivier Beaulnes, Aurélia Marques Oliveira, et Daniela Solfaroli Camillocci. 2023. *Guide de transcription pour les imprimés français du XVIe siècle en caractères gothiques : Version A*. Document de travail.
- Dominique Stutzmann. 2011. Paléographie statistique pour décrire, identifier, dater... Normaliser pour coopérer et aller plus loin? In Franz Fischer, Christiane Fritze, et Georg Vogeler, éditeurs, *Kodikologie und Paläographie im digitalen Zeitalter 2 = Codicology and Palaeography in the Digital Age 2*, Schriften des Instituts für Dokumentologie und Editorik, pages 247–277. BoD.
- Dominique Stutzmann, Jean-François Moufflet, et Sébastien Hamel. 2017. *La recherche en plein texte dans les sources manuscrites médiévales : enjeux et perspectives du projet humanis pour l'édition électronique*. *Médiévales*, 73 :67–96.