

**Atelier BnF DataLab**  
**Humanités numériques, littérature et intelligence artificielle**  
**Vendredi 18 juin 2021 de 14h à 17h30<sup>1</sup>**

---

**Inscriptions**

[https://bnf-fr.zoom.us/meeting/register/tJlIduGtpj0pHNeyLvGk39TjNcJs71\\_uk3l5](https://bnf-fr.zoom.us/meeting/register/tJlIduGtpj0pHNeyLvGk39TjNcJs71_uk3l5)

**Organisateurs :**

Emmanuelle Bermès, Marie Carlin, Antoine Silvestre de Sacy et Thaïs Bach

**Descriptif de l'atelier :**

Littérature et traitement automatique semblent de prime abord plutôt antinomiques : si l'une renvoie aux principes de sensibilité et de création, l'autre au contraire relève de la froide mécanique de l'automatisation. Et pourtant, la littérature et la linguistique comptent au premier rang des disciplines dont les pratiques de recherche ont largement été bouleversées ces dernières années par les outils du numérique. Ces outils ont permis de découper le langage en « donnée première » d'analyse, avec pour but de saisir, comprendre et analyser son fonctionnement sur de larges corpus. L'automatisation des traitements a permis une approche plus globale, beaucoup moins laborieuse que l'approche manuelle, permettant l'émergence de nouveaux champs de recherche comme la lexicométrie, la textométrie, de la stylométrie, voire de la « stylistique outillée ».

La BnF accompagne évidemment ce mouvement : de par la richesse de ses collections, qui font de l'institution un formidable « réservoir de données », mais aussi par l'ensemble des outils et l'accompagnement qu'elle met en place pour faciliter la consultation, l'extraction, l'analyse de ces vastes corpus. Les expertises et savoirs faire professionnels des différents acteurs de la BnF se retrouvent donc au service d'un objectif commun avec les acteurs de la recherche : celui de mieux connaître et valoriser les fonds qu'elle renferme.

Cet atelier du BnF DataLab propose un tour d'horizon des différentes formes d'automatisation de l'analyse littéraire, par la présentation de projets de recherche actuels.

---

<sup>1</sup> Cet atelier intervient dans le cadre de la mise en place de manifestations au BnF DataLab nouveau service est à destination des chercheurs qui souhaitent travailler sur les collections numériques de la BnF, qui doit ouvrir ses portes en octobre 2021.

## Déroulé de l'atelier

**14h** - Accueil des participants par Marie Carlin, BnF, coordinatrice du BnF DataLab

**14h05** – Emmanuelle Bermès, BnF, Adjointe chargée des questions scientifiques et techniques auprès du Directeur des services et des réseaux et Glenn Roe, Professeur de littérature française et humanités numériques à la Faculté des Lettres de Sorbonne Université, responsable de l'équipe-projet ObTIC : *Enjeux de la lecture distante : notions et corpus*.

- **1er partie : Stylométrie et analyse stylistique :**

**15h00** : Jean-Baptiste Camps (Maître de conférences en philologie computationnelle, École nationale des chartes, PSL) et Florian Cafiero (Ingénieur de recherche, GEMASS-CNRS, Paris-Sorbonne) : *Œuvres collectives et stylométrie*.

**15h20** : Antoine Silvestre de Sacy (Ingénieur d'étude, CNRS Huma-Num) : *La méthode des motifs : identification de schémas lexico-grammaticaux caractéristiques pour identifier les styles ou les genres littéraires*.

**15h40** : Questions

\_\_ **16h** : Pause \_\_

- **2e partie : Littérature et IA**

**16h15** : Carmen Brando (Ingénieure de recherche, EHESS) : *Entités nommées et textes littéraires : annotation, expériences et visualisations*.

**16h35** : Thierry Poibeau (Directeur de recherche, CNRS) : *Génération combinatoire de sonnet à partir d'un corpus de poésie du 19e siècle*.

**17h** : Questions

**17h30** : Conclusion générale par Laurence Tarin (directrice du département de l'Orientation et de la recherche bibliographique, BnF)