

La stylométrie : quand les statistiques et l'intelligence artificielle permettent d'identifier l'auteur d'une œuvre littéraire

Sébastien de Valeriola et Guillaume Quintin
(QuaDiHum Lab, ReSIC, LTC)

Après-midi inédit du 20 février 2024

« Et l'auteur de la lettre est un Allemand. Avez-vous remarqué la construction particulière de la phrase : “Les renseignements sur vous nous sont de différentes sources venus.” ? Ni un Français, ni un Russe ne l'aurait écrite ainsi. Il n'y a qu'un Allemand pour être aussi discourtois avec ses verbes. Il reste toutefois à découvrir ce que me veut cet Allemand qui m'écrit sur papier de Bohême et préfère porter un masque plutôt que me laisser voir son visage. »

Tels sont les mots qu'Arthur Conan Doyle met dans la bouche de son célèbre protagoniste, Sherlock Holmes¹. De la fiction ? Assurément, mais notre exposé vise à montrer que des méthodes bien réelles s'en approchent.

1 Résumé

L'écrivaine JK Rowling, connue avant tout pour la saga Harry Potter, est-elle aussi l'auteure de « L'Appel du Coucou », un roman signé du nom de Robert Galbraith ? Molière a-t-il réellement écrit toutes les pièces qui lui sont attribuées ? Quels discours du président américain Ronald Reagan ont-ils vraiment été rédigés de sa main ? Dans quel ordre les œuvres de Platon ont-elles été écrites ? Qui se cache derrière QAnon, le compte Reddit anonyme connu pour ses théories complotistes ravageuses ? À ces questions, qui concernent des textes originaires d'époques et de contextes très différents, on peut apporter des réponses en effectuant des analyses du style littéraire à

¹Arthur Conan DOYLE. *Un scandale en Bohême*.

l'aide de méthodes issues de la statistique et de l'intelligence artificielle. En effet, la façon d'écrire est dans une certaine mesure révélatrice de l'identité : notre style est une empreinte indélébile sur tout texte que nous produisons.

Dans cet exposé, nous découvrirons avec les élèves la stylométrie, une boîte à outils à cheval sur les sciences exactes et les sciences humaines, qui a des applications dans des domaines très variés, depuis les études littéraires et l'histoire jusqu'à la criminologie et la politologie, en passant par la philosophie et la musicologie. Après une présentation des débuts de celle-ci et de ses principes fondamentaux, nous disséquons un exemple concret (un texte rédigé en latin au Moyen Âge). Pour ce faire, nous ferons appel à des notions élémentaires de statistique, et introduirons une technique d'intelligence artificielle de réduction de dimensionnalité, de façon accessible pour un public non initié. Nous terminerons par un panorama des domaines d'application de la stylométrie, tant dans le monde de la recherche que dans un contexte sociétal.

2 Structure

La présentation articulera des explications théoriques avec une mise en pratique lors de laquelle les élèves comprendront comment fonctionne concrètement la stylométrie :

1. Accroche avec des exemples fictifs et réels du procédé stylométrique ;
2. Définition de la stylométrie ;
3. Bref historique de la discipline ;
4. Mise en pratique : repérer les critères pertinents pour identifier le style d'un auteur ;
5. L'apport de l'intelligence artificielle à la stylométrie exemplifié par un exemple concret en latin médiéval ;
6. Ouverture vers des utilisations au-delà de la littérature.

L'exposé théorique sera interactif, grâce à des questions posées aux élèves par l'outil Wooclap. Les élèves auront besoin d'un smartphone pour y participer.

3 Contact et ressources

Les deux organisateurs de l'atelier sont joignables par email :

- Sébastien de Valeriola : sebastien.de.valeriola@ulb.be ;
- Guillaume Quintin : guillaume.quintin@ulb.be.

De plus amples informations sur les activités scientifiques du QuaDiHum Lab, laboratoire dédié aux humanités numériques à l'ULB, sont disponibles sur son site web : <https://quadihum.ulb.be/> (en anglais).

Les deux organisateurs sont membres de l'équipe enseignante du Master en Sciences et Technologies de l'Information et de la Communication (STIC), qui dispose de son propre site web : <https://mastic.ulb.be/>. Vous pourrez y retrouver une version électronique de ce dossier pédagogique ainsi que les diapositives projetées lors de l'atelier.

4 Références

4.1 Sitographie

Quelques ressources en format vidéo :

- Brève séquence de presse sur une radio française ([lien](#)), revenant sur des usages contemporains de la stylométrie (notamment l'affaire QAnon) ;
- Conférence de Jean-Baptiste CAMPS ([lien](#)) traitant en détail de l'histoire et des méthodes de la stylométrie ;
- En anglais : Vulgarisation d'un article scientifique de Mike KESTEMONT ([lien](#)) consacré à la figure d'Hildegarde de Bingen.

En outre, l'exemple d'analyse stylométrique sur les textes en latin médiéval est développé dans un document hébergé sur le site du QuaDiHum Lab ([lien](#)).

4.2 Bibliographie

Les articles et ouvrages consacrés à la stylométrie sont nombreux, et majoritairement rédigés en anglais. Citons d'abord quelques références en français :

- Bref article consacré à Louise Labé, figure littéraire de la Renaissance : Florian CAFIERO, Jean-Baptiste CAMPS et Simon GABAY. « Louise Labé : une créature de papier ? » In : *Humanistica 2023*. Moyen Âge - Renaissance. Association francophone des humanités numériques. Genève, Suisse, juin 2023. URL : <https://hal.science/hal-04090284> ;
- Ouvrage de référence : Florian CAFIERO et Jean-Baptiste CAMPS. *Affaires de style*. Le Robert, avr. 2022. URL : <https://hal.science/hal-03936009>.

Citons également quelques exemples anglophones :

- Article cité dans la sitographie, sur Hildegarde de Bingen : Mike KESTEMONT, Sara MOENS et Jeroen DEPLOIGE. « Collaborative authorship in the twelfth century : A stylometric study of Hildegard of Bingen and Guibert of Gembloux ». In : *Digital Scholarship in the Humanities* 30.2 (2015), p. 199-224. DOI : [10.1093/LLC/FQT063](https://doi.org/10.1093/LLC/FQT063) ;
- Article consacré aux œuvres de Molière : Florian CAFIERO et Jean-Baptiste CAMPS. « Why Molière most likely did write his plays ». In : *Science Advances* 5.11 (nov. 2019), p. 1-14. DOI : [10.1126/sciadv.aax5489](https://doi.org/10.1126/sciadv.aax5489) ;
- Article fondateur de la stylométrie moderne, dans lequel des statisticiens se sont intéressés à des essais politiques publiés sous pseudonyme au lendemain de l'indépendance des États-Unis d'Amérique : F. MOSTELLER et D. L. WALLACE. « Inference in an authorship problem. A comparative study of discrimination methods applied to the authorship of the disputed Federalist papers ». English. In : *Journal of the American Statistical Association* 58 (1963), p. 275-309. ISSN : 0162-1459. DOI : [10.2307/2283270](https://doi.org/10.2307/2283270).

5 Annexes

5.1 Texte 1

Au XIXe siècle, la journée d'un élève était marquée par la rigueur des études et la simplicité de la vie quotidienne. Dès l'aube, l'écolier se levait avec les premières lueurs du jour et entamait le chemin escarpé vers l'école. Les salles de classe modestes, mais propices à l'apprentissage, accueillaient les élèves avec l'odeur caractéristique du bois et de l'encre. Le maître, exigeant mais bienveillant, guidait les élèves dans leur quête de connaissance. Les devoirs étaient nombreux, mais ils forgeaient la discipline et la persévérance des écoliers. La pause déjeuner offrait un répit bienvenu, mais l'après-midi était dédié à de nouvelles leçons exigeantes. Le soir tombant, les devoirs remplissaient les foyers modestes, mais les familles encourageaient l'éducation comme une voie vers un avenir meilleur. La faible lueur des bougies éclairait les cahiers et livres, créant une ambiance studieuse. Malgré les défis économiques et les conditions parfois difficiles, la journée d'un élève au XIXe siècle était une marche vers l'éducation, marquée par la détermination, les rires d'enfants et l'espoir d'un savoir toujours plus grand.

5.2 Texte 2

Je prends la plume pour partager avec vous l'histoire de ma passionnante et parfois surprenante collection de papillons. Depuis des années, j'ai consacré mon temps à chasser ces créatures délicates, et ma collection est devenue le témoin de la diversité infinie de ces merveilles ailées. Chaque papillon que j'y ajoute apporte une nouvelle dimension de beauté avec ses motifs uniques et ses couleurs éclatantes. Chasser ces papillons n'est pas facile, mais la patience est la clé pour saisir la perfection de chaque instant. Les triomphes sont nombreux, mais parfois, la rareté d'une espèce est un défi qui rend la chasse plus passionnante. La préservation de leur habitat naturel est cruciale, et c'est là que réside le paradoxe de ma collection. Elle représente la beauté de la nature, mais aussi l'impact des défis environnementaux et sociétaux auxquels ces créatures fragiles sont confrontées. Ainsi, ma collection devient une fenêtre ouverte sur la biodiversité et une invitation à la préservation. Chaque papillon représente une histoire unique, mais aussi un rappel de la fragilité de notre écosystème. En espérant que ma passion pour ces merveilles ailées puisse susciter en vous le même enthousiasme.

5.3 Texte 3

Je vous adresse mes salutations respectueuses dans cette lettre où je vais tenter de vous narrer ma journée à l'école. Dès le matin, j'ai entamé le chemin menant à l'établissement scolaire. Les salles de classe, propices à l'apprentissage, accueillait les élèves avec une atmosphère studieuse. Les matières enseignées étaient variées, allant des classiques aux sciences émergentes. Comme à mon habitude, j'ai particulièrement apprécié les leçons qui concernaient les sciences exactes, comme la géométrie. Le maître, à la fois exigeant et bienveillant, guidait nos esprits curieux dans la quête du savoir. La discipline régnait dans la salle, or une forte camaraderie entre les élèves créait un environnement propice à l'apprentissage. La pause déjeuner était un moment de répit. L'après-midi apportait son lot de défis à surmonter. Les devoirs, bien que souvent décriés par les élèves, représentaient une opportunité de renforcer nos connaissances. En fin de journée, la fatigue était palpable, mais le sentiment d'avoir acquis de nouvelles connaissances faisait naître un sentiment de satisfaction. Ma journée à l'école était ainsi une expérience inestimable pour mon développement intellectuel, ainsi que pour celui des autres élèves.