

La stylométrie, technique qui bouleverse les affaires criminelles

Science et justice La méthode s'immisce dans les grandes affaires pénales pour dévoiler l'identité des auteurs de textes. Une start-up suisse se cache derrière des découvertes majeures.

Luca Di Stefano

Corneille n'a pas écrit les pièces de Molière, contrairement à ce que laissait croire une vieille rumeur tenace. En revanche, certains actes d'Henri VIII sont l'œuvre de John Fletcher et non de Shakespeare, son maître. On doit ces découvertes à une même technique. En proposant d'analyser la manière dont chaque individu rédige, la stylométrie est une recherche de la signature stylistique de l'auteur d'un texte. Elle éclaire l'histoire d'un écrit et répond à une question simple : qui a écrit ces mots ?

Depuis son irrruption dans le champ judiciaire, cette technique provoque des déflagrations à la mesure des intrigues qu'on lui demande d'élucider. On peut citer l'affaire du petit Gregory. Depuis quarante et un ans, l'identité de l'assassin du petit garçon de la famille Willemain est un mystère, alors que la réponse – ou du moins des informations fondamentales – pourrait se trouver dans les lettres de menaces reçues par la famille.

«Voilà ma vengeance»

Quatre décennies plus tard, le mystère du corbeau est peut-être percé. Fin octobre, Jacqueline Jacob, 81 ans, grand-tante de l'enfant, est mise en examen. Le procureur de Dijon a reçu une expertise réalisée par la start-up suisse OrphAnalytics. Celle-ci soutient «très fortement l'hypothèse» que l'octogénaire a rédigé certains courriers de menace. «J'espère que tu mourras de chagrin le chef [...] Voilà ma vengeance. Pauvre con», disait l'une des lettres adressées au père de l'enfant. Jacqueline Jacob est-elle aujourd'hui trahie par son style ?

Ici, la notion ne renvoie pas au choix des mots, à l'élaboration des phrases ou au respect des règles de la langue.

La stylométrie, telle que pratiquée par le chercheur suisse, consiste à découper les mots employés en séquences de deux ou trois lettres (*lire ci-dessous*). Leur utilisation et la manière dont elles s'enchaînent marquent une empreinte stylistique dont l'auteur ignore tout. Mais pour aboutir, il a fallu comparer un corpus de textes – 35 au total – entre eux avant de s'appuyer sur un algorithme. La signature de l'auteur des menaces prend au final la forme d'une probabilité statistique hautement significative.

Si Claude-Alain Roten n'est pas autorisé à parler du dossier sur lequel il travaille, les conclusions de son expertise, révélées par la presse française, laissent peu de doute quant à l'identité de l'auteur des lettres.

Commentaires anonymes

Aujourd'hui, un autre dossier pourrait rebondir. Ce n'est rien de moins qu'une demande de révision qui a été transmise au président de la Chambre pénale d'appel et de révision de Genève qui a jugé l'affaire Ramadan.

Dans son volet suisse, la condamnation de l'islamologue a été confirmée par le Tribunal



La défense de Tariq Ramadan demande à la justice genevoise de rouvrir le dossier. Selon une expertise stylométrique, la plaignante aurait annoncé «une torpille [...] en prépa» avant sa rencontre avec l'islamologue (ici en mai 2023). Laurent Guiraud

fédéral, mais l'équipe de défense a adressé, il y a moins d'un mois, une demande de réouverture du dossier. À entendre les avocats de Tariq Ramadan, les recherches menées par la linguiste forensique américaine Carole E. Chaski mettent au jour «un mensonge original» de celle qui l'accuse de viol à Genève. Un nouvel élément à la lumière duquel «une nouvelle lecture doit s'appliquer sur l'ensemble du dossier».

Si le cas Ramadan se prête à la stylométrie, c'est parce qu'il repose sur des traces écrites, en l'absence de témoins directs. En l'occurrence, il s'agit de commentaires déposés sur des blogs par une rédactrice anonyme avant

la rencontre entre le Genevois et son accusatrice. Parmi ces réactions, une dénommée «Leila» écrit au bas d'un article hostile à Tariq Ramadan : «A enfoiré, enfoiré et demi.. Une vraie torpille en prépa.»

Examinant 29 documents pour un total de 4228 mots et 328 phrases, l'experte a procédé à une analyse syntaxique pour identifier l'auteur de ce commentaire ainsi que d'autres, tous dans le même esprit. Ici, la méthode est moins complexe que celle employée pour l'affaire Gregory. En l'occurrence, il s'agit de repérer les structures syntaxiques employées par la personne dissimulée derrière ces commentaires.

Au-delà des résultats statistiques, l'emploi de la locution «A enfoiré, enfoiré et demi...» a été retenu comme une signature stylistique, tout comme l'emploi des deux points (au lieu d'un point final ou des points de suspensions) qui revient avec régularité dans d'autres textes. En définitive, la linguiste est parvenue à la conclusion que l'accusatrice suisse de Tariq Ramadan est – à 99,09% et 100% selon les modèles statistiques employés – l'auteure de ces commentaires.

«Pas une science exacte»

Désormais, l'expertise privée se trouve en mains du juge, laissant les pénalistes poursuivre leur bataille procédurale. Alors que le recours à la stylométrie est également convoqué dans le dossier français (le procès aura lieu en 2026), la défense de l'islamologue pointe «un simulacre» de l'accusatrice.

À partir des commentaires publiés par la plaignante avant qu'elle ne prenne contact avec Tariq Ramadan, «on découvre en réalité un projet qui a été mûri, ficelé et discuté en amont par cette dernière», écrivent les avocats. En d'autres termes, sa véritable intention au moment de contacter l'enseignant aurait été «de lui envoyer une torpille».

Quel poids doit-on accorder à cette expertise privée ? Avocate de l'accusatrice dont le viol a été reconnu par la justice, M^e Véronique Fontana doute de sa valeur probante. «Ce type d'expertise repose sur des analyses statistiques, et n'est donc pas une science exacte ou infaillible», relativise-t-elle.

En trente-cinq ans de barreau, la pénaliste avoue être confrontée à la stylométrie pour la première fois. Elle préconise de l'utiliser «avec réserve et prudence» et de l'apprécier «en complément à d'autres preuves du dossier et non isolément». Seule certitude, tout comme dans l'affaire du petit Gregory, l'appréciation finale appartient aux juges.

«Chacun écrit avec son vécu, ses affinités, ses blessures»

Si la stylométrie fait figure de «spécialité suisse», elle le doit beaucoup à Claude-Alain Roten, fondateur de la start-up OrphAnalytics. Mandaté pour travailler sur l'affaire du petit Gregory, le chercheur vaudois n'est pas autorisé à parler du dossier. En revanche, il décrit le fonctionnement d'une technique reposant sur les statistiques et l'IA.

Comment travaillez-vous lorsqu'il vous est demandé de retrouver l'auteur d'un texte ?

Après avoir rassemblé les textes, on les homogénéise pour les rendre comparables. La question est la suivante : peut-on les regrouper ou faut-il les distinguer ? On va systématiquement découper le texte en unités stylométriques, c'est-à-dire des modules. Ce sont, par exemple, des

séquences de deux ou trois lettres qui, dans leurs récurrences et leurs enchaînements, sont utilisées par chacun de manière singulière. Prenons l'exemple de La Fontaine : notre logiciel va décomposer le vers «Maître Corbeau, sur un arbre perché» pour obtenir une combinaison de paires de lettres que l'on retrouvera dans une proportion similaire dans les autres écrits du même auteur.

En quoi la récurrence de lettres permet-elle de renvoyer à un même auteur ?

Parce que chacun écrit par registre. De manière presque inconsciente, on écrit avec son vécu, ses affinités, ses blessures qui rendent chaque style unique et que nous retrouvons dans des détails dissimulés parmi une masse d'information.

Selon certains experts, la stylométrie est une discipline ancienne. Quel est l'apport de la technologie aujourd'hui ?

Au XIX^e siècle, on cherchait déjà des critères pour identifier ce qui est profondément personnel dans un texte : longueur des mots, des phrases, etc. Cela fonctionne. Mais la grande force aujourd'hui, c'est le machine learning, l'apprentissage automatique qui rend l'algorithme de plus en plus efficace. Il permet de regrouper les textes d'un même auteur ou, au contraire, de les séparer s'ils ne sont pas d'une même personne. Cinquante caractères peuvent être suffisants.

La stylométrie intervient dans des affaires retentissantes et fait beaucoup de bruit quand

elle délivre ses résultats. Fâchés par votre travail, des avocats vous ont traité de charlatan.

On m'a même qualifié de «zin-zin helvétique !» Ce n'est jamais agréable, mais je rappelle que ces résultats sont validés par des calculs de certitude de 99%. Notre travail consiste à estimer que nos résultats ne sont pas le produit du hasard, lui-même estimé à moins de 1%. Quant aux avocats qui tentent de me discréder, je remarque qu'ils sont peut-être d'excellents pénalistes, mais que leur démarche n'a rien de scientifique.

À l'heure où beaucoup de textes sont produits avec l'aide de l'IA, les analystes de la stylométrie peuvent-ils être menés en bateau ?

Pas du tout. Si on demande à ChatGPT de faire du Shakespeare, il en est incapable. L'intelligence artificielle que nous utilisons permet de distinguer une copie d'un texte original.

Comment le microbiologiste que vous êtes en est-il arrivé à la stylométrie ?

Lorsque j'ai fait ma thèse dans les années 90, on commençait à lire les génomes sous forme de chaînes de quatre caractères nucléiques. Lorsque je voulais vulgariser mes recherches, je prenais des extraits de véritables textes pour rendre mes explications plus claires, car la complexité des séquences biologiques est comparable à celle de nos caractères. Cela m'a conduit à utiliser ces techniques pour authentifier des textes.