

ENQUÊTE DOCUMENTAIRE

LA STYLOMÉTRIE ET SES APPLICATIONS

Molière est-il l'auteur des pièces qu'il a jouées ? Homère a-t-il écrit l'*Illiade* et l'*Odyssée* ? Quel mystérieux alphabet révèle le manuscrit de Voynich ?

Questions épineuses auxquelles la stylométrie tente de répondre. Mesurer le style, l'affaire semble une chimère. L'art est étranger à la science, défend l'esthète. D'autres y voient les facettes d'une même vérité.

Soutenant que « tout est nombre », Pythagore fut sans doute précurseur en la matière. « Le monde, l'homme tout entier est dans l'alphabet » affirmera plus tard Hugo. Le croisement de ces idées est notre fil d'Ariane.

■ Histoire de la stylométrie

La véritable incursion fut moderne : en 1851, De Morgan paria sur la longueur des mots pour identifier un auteur, cependant les résultats furent peu encourageants. Plus théorique, Markov impliqua en 1913 des processus aléatoires dans les séries littérales d'un roman de Pouchkine. Un nouveau critère, la longueur des phrases, fut proposé en 1938 par Yule. Mais les statistiques « en masse » restaient insuffisantes : en 1956, Herdan promut des statistiques « en ligne » liées au temps.

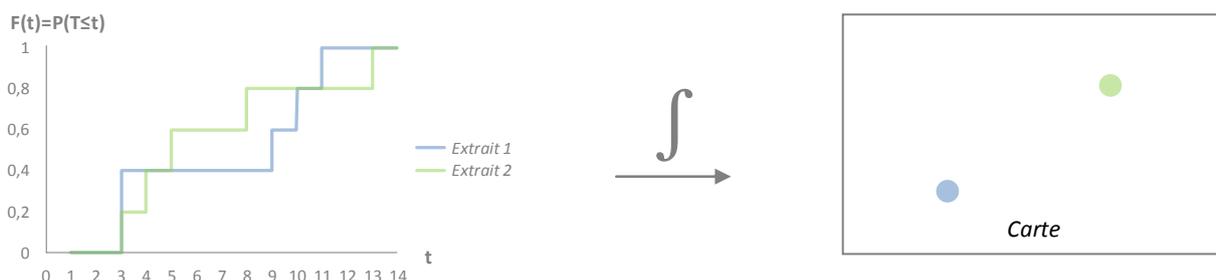
En France, le pionnier de la discipline fut Muller : à partir des textes numérisés de la base *Frantext*, il présenta en 1967 une étude sur le vocabulaire théâtral de Corneille. En 1981, Benzécri appliqua des techniques multidimensionnelles à la linguistique : avec la carte d'un corpus, les spécificités devenaient visibles. La même année, Brunet publia une histoire du vocabulaire français depuis la Révolution, avant de produire lors du bicentenaire un logiciel de statistique textuelle, *Hyperbase*.

■ Principe de mesure

La stylométrie suit une approche matérialiste, qui fait sa force et sa faiblesse. Loin de sonder l'esprit de l'auteur ou la réception du lecteur, elle recense objectivement les unités d'un texte.

Selon le point de vue linguistique, ces unités varient. Le vocabulaire porteur du sens est naturellement le plus étudié ; cependant, il implique un lemmatiseur pour ramener les formes lexicales aux entrées du dictionnaire, et sa richesse limite malencontreusement le nombre des occurrences. Les catégories grammaticales sont quant à elles plus stéréotypées, leurs emplois étant dictés par des règles ; en outre, elles réclament un étiqueteur afin d'affecter chaque terme. Ponctuation, espacements et lettres véhiculent enfin un élément plus archaïque et inconscient du langage, fait de sons et d'émotions ; objectifs et profus, les caractères sont ici privilégiés.

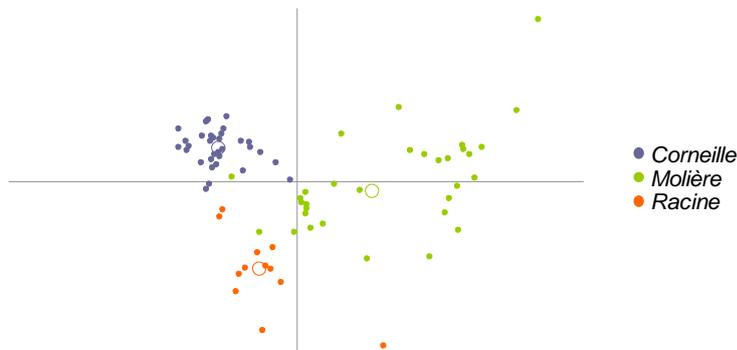
Les unités fixées, quel mètre suivre ? Pour illustrer notre propos, observons la lettre « e » dans les vers « Approuvez ma faiblesse, et souffrez ma douleur » (Corneille, *Horace*) et « Allons, Flipote, allons, que d'eux je me délivre » (Molière, *Tartuffe*). Le caractère en question, non accentué, est légèrement plus fréquent dans le premier cas (6 occurrences sur 46 signes) que dans le second (6 sur 48). Pour autant, ce comptage élémentaire ignore le rythme des apparitions. Plus finement, on enregistre les temps de retour du caractère (T vaut successivement 11, 3, 3, 9, 10 dans le premier cas ; 13, 4, 5, 3, 8 dans le second). Généralement indépendants, ces termes se synthétisent par leur fonction de répartition. Les écarts entre ces fonctions, intégrés sur l'ensemble des caractères, définissent alors une distance intertextuelle. Enfin, les distances mutuelles au sein d'un corpus sont projetées sur un plan idoine, en vue de tracer une carte.



▪ L'affaire Corneille-Molière

Derrière l'acteur, les contemporains de Molière voyaient également un auteur. D'aucuns s'interrogèrent sur une vie menée entre scène et écritoire, pressée par les demandes du roi. Baudeau de Somaize l'accusa de plagier, Donneau de Visé de rhapsoder. Néanmoins, nul ne soupçonnait une plume de l'ombre. Mais en 1919, le poète Louÿs reconnut la main de Corneille dans les vers d'*Amphitryon*, avant de généraliser à l'ensemble des pièces de Molière. Si la thèse fut reprise par des romanciers, elle trouva une nouvelle dimension par les statistiques. Analysant le vocabulaire, Labbé vit des croisements entre les deux *Menteurs* et les comédies de Molière, ainsi qu'entre *Dom Garcie* et les tragédies de Corneille. L'intuition du poète étayée par la rigueur des chiffres, une polémique opposa partisans et détracteurs des dramaturges, relayée par l'écrivain et animateur Franck Ferrand.

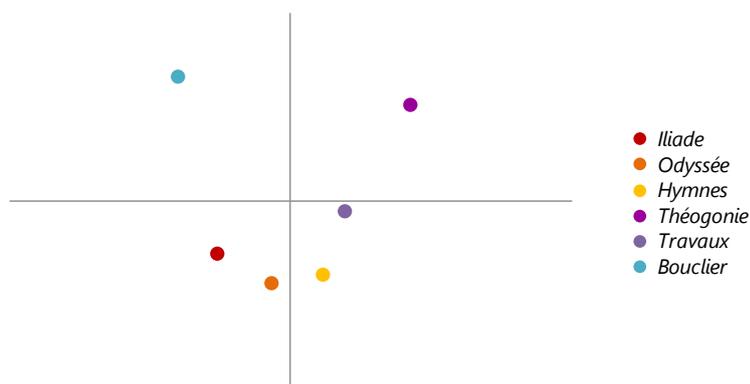
Entre les protagonistes, Racine arbitre. Sur la carte de l'ensemble des pièces, Corneille et Racine administrent minutieusement des royaumes unis, tandis qu'un Molière aventureux étend son empire à des marches lointaines, vraisemblablement inspiré par des diverses sources, françaises, italiennes ou espagnoles. Pour autant, Molière ne se saurait se rapprocher plus de Corneille que de Racine : les centres des trois nuages sont quasiment équidistants. L'axe horizontal reflète clairement les contributions du genre et de la forme : à gauche, les tragédies en vers, à droite les comédies en prose. Plus diffuse, la chronologie transparaît sur l'axe vertical, de Corneille à Racine. Si une plume de l'ombre ne peut être écartée, les mesures ne poussent guère à la dissidence, et la prudence ramène à la doctrine.



▪ La question homérique

L'Antiquité ne douta jamais de l'existence du « divin » Homère, père de la poésie grecque. Pour autant, on sait peu sur un homme situé au 8^e siècle avant notre ère, dont les traces se réduisent à des *Vies* apocryphes ou des évocations poétiques. Ces lacunes furent les ferments des conjectures de l'abbé d'Aubignac au 18^e siècle. La polémique enfla au siècle suivant, opposant un Wolf iconoclaste aux gardiens de la tradition : l'œuvre d'Homère était-elle une rhapsodie comme l'insinuaient les analystes, ou un ensemble cohérent comme l'affirmaient les unitaires ? Une autre incidence surgit au 20^e siècle, avec la thèse de Parry : la poésie d'Homère était-elle essentiellement orale, destinée aux aèdes ? Les dernières décennies s'acheminèrent vers un compromis : les odes éparses et séculaires auraient été unies et tissées par l'art d'un poète, Homère pour la postérité.

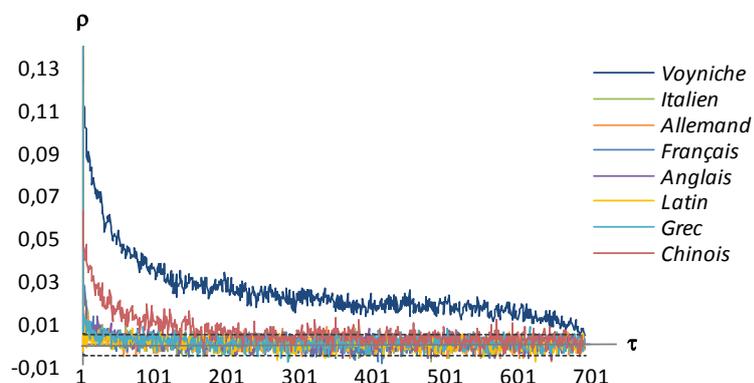
Le corpus comprend l'*Illiade* et l'*Odyssée*, attribuées à Homère, la *Théogonie* et les *Travaux* attestés à Hésiode, ainsi que deux satellites, les *Hymnes* dits homériques et le *Bouclier* d'un pseudo Hésiode. L'analyse des caractères permet de travailler sur le grec ancien, sans impliquer la connaissance de la langue. Avec des distances réduites sur la carte, le noyau d'Homère apparaît soudé face au groupe d'Hésiode ; les *Hymnes*, proches de l'*Odyssée*, sont dans le foyer homérique, tandis que le *Bouclier*, imitation du chant 18 de l'*Illiade* d'après Aristophane de Byzance, tombe vers Homère. D'une vue générale, la guerre semble grever la gauche du plan, tandis que la paix règne à droite. In fine, les résultats confortent la tradition et remettent en cause les analyses modernes. Sans trancher, nous parions que « l'aveugle de Chios » a vécu et composé ses poèmes.



▪ Le manuscrit de Voynich

La dernière pièce de notre trilogie est assurément la plus étrange. Actuellement détenu par l'université de Yale, le manuscrit doit son nom à un bibliophile polonais, acquéreur en 1912 lors d'un voyage en Italie. Dans un fonds ancien tenu par des jésuites, il découvrit un codex singulier, tracé de glyphes et de figures énigmatiques. Glissée entre les folios, une lettre datée de 1666 indiquait qu'il avait appartenu à Rodolphe II. Fêré d'ésotérisme et de science, l'empereur germanique l'aurait acheté en 1586, espérant mettre la main sur une œuvre de Bacon, alchimiste anglais du 13^e siècle. Cependant, le style du manuscrit évoque l'Europe centrale, et le carbone 14 situe le parchemin au début du 15^e siècle. Certains rapprochent par ailleurs le « voynichien » des langues asiatiques. D'autres enfin y voient un artifice, si ce n'est un canular.

À l'aune des caractères, la composition du manuscrit est normale, cependant la chronologie présente des corrélations inédites parmi les langues européennes, partiellement retrouvées dans le chinois. La piste asiatique de Stolfi, repérée sur les longueurs de mots, se voit quelque peu rouverte. L'ampleur du phénomène pourrait naître d'un processus aléatoire. Mais ces résonnances, constatées à d'autres niveaux linguistiques, feraient du manuscrit une construction complexe, manifestement hors de portée d'un savant du Moyen Âge, a fortiori d'un faussaire pressé par l'argent. L'hypothèse la plus vraisemblable reste celle d'un alchimiste, s'appuyant sur un langage naturellement rythmé pour transmettre une incantation ou une initiation. Néanmoins, le mystère demeure.



▪ Champ de la stylométrie

L'autorité des chiffres ne confère pour autant aucun imperium à la stylométrie. Le texte est résumé à quelques caractéristiques, dépendant des unités comme des statistiques – sous couvert d'objectivité, la subjectivité subsiste. De surcroît, un écrit ne saurait cerner un auteur, pensons à la falsification ou à l'imitation, et une enquête historique assemble nécessairement les documents et les témoignages avant de conclure. En somme, la stylométrie n'arrête rien ; néanmoins, elle comble les voies traditionnelles pour éclairer des points en suspens.

▪ Liens

Graphoscopie (détail de la méthode et des études)

Lexicométrica (revue de statistique textuelle)