

> Les arguments laissant à penser que Molière n'aurait pas écrit ses pièces sont d'abord historiques. Il ne nous est ainsi parvenu aucun manuscrit, pas même une lettre, une note ou un brouillon, qu'on puisse lui attribuer. Il s'agit là d'un cas unique pour un auteur aussi prolifique et célèbre. Le moliériste Georges Forestier, de l'université Paris-Sorbonne, note toutefois que les manuscrits d'auteurs du XVII<sup>e</sup> siècle, alors jugés de peu d'importance, étaient très rarement archivés, mais bien plus souvent détruits. Aucun brouillon ou manuscrit de pièces de Corneille n'a ainsi été conservé. De Racine, il ne nous est parvenu qu'un acte d'un projet de pièce finalement abandonné. Et si ses descendants l'ont précieusement gardé, c'est probablement parce que cette pièce n'a jamais été imprimée, et que ce brouillon en était la seule trace restante. Rien de si alarmant, donc, à ne rien retrouver de la main de Molière.

Autre élément ayant prêté au soupçon : Molière aurait commencé à jouer les pièces de Corneille après un voyage à Rouen en 1643, et se serait mis à écrire ses chefs-d'œuvre après un séjour en 1658 dans la même ville. Or Rouen était le fief de Corneille. Ces voyages en terre normande auraient-ils été l'occasion d'ententes secrètes entre les deux hommes ? Les spécialistes de Molière en doutent, et ne sont guère impressionnés par la coïncidence. Il est possible, mais nulle part attesté, que Molière et Corneille se soient rencontrés sur place. Et l'on ne trouve aucune trace d'un quelconque accord conclu entre les deux hommes. Dans sa correspondance, Corneille semble d'ailleurs indifférent au passage de la troupe en 1658. Troisième ville du pays, Rouen était régulièrement visitée par les troupes de théâtre, qui n'y allaient pas nécessairement voir Corneille.

Dernier argument majeur : l'éclosion tardive du talent de Molière. Celui-ci écrivit tous ses chefs-d'œuvre passé l'âge de 37 ans, ce qui n'a



Molière, de son vrai nom Jean-Baptiste Poquelin (1622-1673), gravure de James Poppelwhite publiée dans *The Gallery of Portraits: with Memoirs*, d'Arthur Thomas Malkin (Londres, 1833).

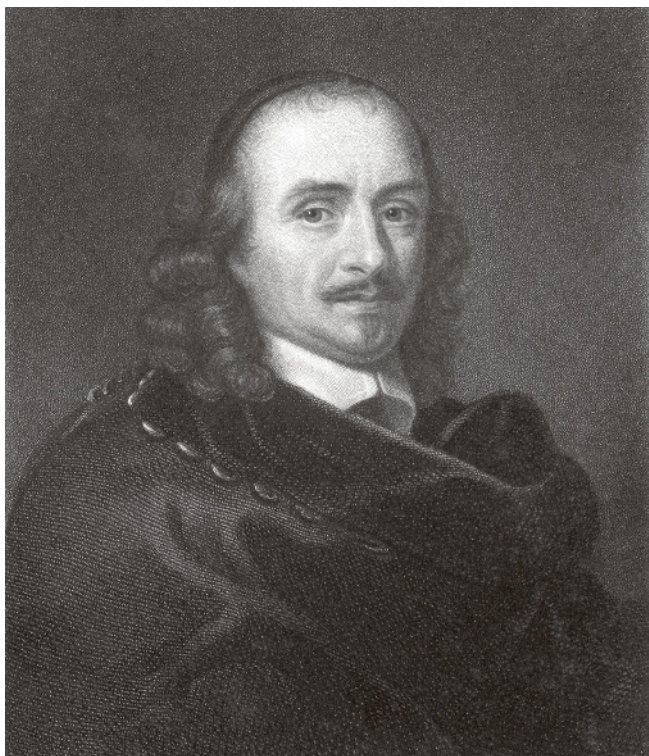
pas manqué d'étonner. Selon Hippolyte Wouters, avocat belge et auteur de théâtre, ce serait le seul cas « où un auteur médiocre jusqu'à 40 ans devient non seulement profond, mais surtout une des plus belles plumes de son temps ». Si cette situation est en effet assez rare dans l'histoire littéraire, elle n'est cependant pas inédite – Marcel Proust, James Conrad ou Umberto Eco ont débuté des carrières littéraires tout aussi tard, et ont été couronnés d'un succès certain.

## RETOUR AUX TEXTES

Faute d'archives pouvant prouver l'imposture supposée, la controverse historique semblait mener à une impasse. Pour apporter de nouveaux éléments aux débats, des chercheurs ont décidé de s'intéresser aux textes eux-mêmes. Si les textes signés de la main de Molière étaient stylistiquement proches de ceux de Pierre Corneille, ce pourrait être le signe que ce dernier en est l'auteur.

Le constat de départ de Dominique et Cyril Labbé est simple : deux textes écrits par un même auteur ont plus de vocabulaire en commun que deux textes écrits par des auteurs différents. En prenant comme base la fréquence d'apparition de chaque mot dans un texte, ils ont calculé une « distance intertextuelle », qui mesure la différence entre deux textes dans le choix du vocabulaire. Si la distance intertextuelle séparant deux pièces est

# L'éclosion tardive du talent d'un auteur est une situation rare, mais pas inédite



Pierre Corneille (1606-1684), gravure de Thomas Woolnoth publiée dans *The Gallery of Portraits : with Memoirs*, d'Arthur Thomas Malkin (Londres, 1833). On a supposé, à tort semble-t-il, qu'il était l'auteur des pièces de Molière.

inférieure à un certain seuil, cela signifie selon eux qu'elles ont été écrites par le même auteur.

D'après les calculs de Dominique et Cyril Labbé, les œuvres attribuées à Corneille et de nombreuses pièces de Molière sont trop proches pour avoir été écrites par deux personnes différentes. Cela porterait à croire que Corneille a écrit la plupart des chefs-d'œuvre signés de Molière (l'hypothèse selon laquelle Molière aurait écrit les pièces de Corneille étant invraisemblable et impossible pour des raisons chronologiques).

Depuis ces travaux fondateurs, les techniques permettant de découvrir l'auteur d'un texte ont évolué. Leur usage, tant par des philologues cherchant à identifier la plume derrière un poème jamais signé que par les services de renseignement pour démasquer l'auteur d'une lettre anonyme, s'est également étendu.

## ANALYSER LES HABITUDES D'ÉCRITURE ET LES TICS DE LANGAGE

Ces méthodes reposent sur l'analyse statistique des habitudes d'écriture et des tics de langage. En écrivant un texte, nous laissons involontairement des empreintes qui trahissent notre identité : chaque individu utilise avec une fréquence particulière des mots, des expressions ou des structures grammaticales. En les relevant systématiquement, on peut se faire une excellente idée de qui a écrit un texte.

Depuis quelques années, la démarche la plus couramment adoptée pour révéler l'auteur d'un texte repose sur l'intelligence artificielle. On réunit d'abord une sélection de textes dont on connaît les auteurs avec certitude. On entraîne alors l'ordinateur à reconnaître les auteurs de chacun de ces textes avec la meilleure précision possible. Puis on lui fournit des textes anonymes, pour lesquels la machine prédit l'auteur le plus vraisemblable.

Mais pour construire le profil type des textes d'un auteur, encore faut-il être d'accord sur les pièces que cet auteur a réellement rédigées. Or, dans notre débat, tout le monde n'est pas d'accord sur qui a écrit quoi. La théorie des « comédiens poètes », défendue notamment par Dominique Labbé, voudrait que, comme dans les spectacles d'humour d'aujourd'hui, dont on connaît l'interprète mais rarement le ou les auteurs, on aurait au XVII<sup>e</sup> siècle mis en avant le comédien principal et entièrement omis de citer le dramaturge. Quelque 90% des comédies de l'époque auraient alors été signées par un prête-nom. Si l'on veut prendre au sérieux cette possibilité, il faut donc adopter une autre approche.

Faute d'accorder une totale confiance à l'apprentissage automatique d'une machine, tout en nous assurant au maximum de la solidité de nos conclusions, nous avons utilisé, parallèlement, six méthodes différentes considérées comme fiables pour distinguer l'auteur d'un texte. Testées sur un corpus de trente grandes comédies en vers écrites après la mort de Corneille et Molière par six auteurs, trois de ces méthodes prédisent le bon auteur dans 100% des cas, deux dans 93% et une dans 87% des cas.

## FRÉQUENCES DES MOTS

La première méthode consiste simplement à compter le nombre d'occurrences de chaque mot qui apparaît dans les textes étudiés. On considère le texte comme un « sac de mots ». Les fréquences d'apparition de chaque mot dans ce sac varient d'un texte à l'autre. La distribution de ces fréquences constitue une signature : à chaque auteur ses préférences quant au vocabulaire qu'il utilise.

On peut parfois considérer que certaines variations autour d'un même mot ne sont pas significatives. Utiliser « galère » ou « galères », « vois » ou « voit » peut paraître indifférent, ce qui compterait étant de dire « galère » plutôt que « navire », d'utiliser le verbe « voir » plutôt que « regarder ». La conjugaison, le pluriel, etc. ne seraient qu'accidentels par rapport au terme choisi. Dans ce cas, on travaille non pas sur le mot, mais sur le *lemme* : un usage de « vois », « voit » ou « voyons » est alors considéré dans le calcul comme un usage de « voir », un usage de « galères » comme un usage de « galère », et ainsi de suite. Mais cette ➤