



CLASSIQUES
GARNIER

GLAUDES (Pierre), KHUTORETSKAYA (Olga), CERVONI (Aurélia), GUGLIELMI (Francesca), MAYAUX (Catherine), MARUSENKO (Mikhail), KURALESINA (Yekaterina), MIRETINA (Mariya), NIKITINA (Yekaterina), SOLOVYEVA (Mariya), « Jules Barbey d'Aureville et corpus journalistique. Problèmes d'attribution », in ALEXANDRE (Didier), ROE (Glenn) (dir.), *Observer la vie littéraire. Études littéraires et numériques*, p. 261-301

DOI : [10.48611/isbn.978-2-406-13347-6.p.0261](https://doi.org/10.48611/isbn.978-2-406-13347-6.p.0261)

La diffusion ou la divulgation de ce document et de son contenu via Internet ou tout autre moyen de communication ne sont pas autorisées hormis dans un cadre privé.

© 2022. Classiques Garnier, Paris.
Reproduction et traduction, même partielles, interdites.
Tous droits réservés pour tous les pays.

JULES BARBEY D'AUREVILLY ET CORPUS JOURNALISTIQUE

Problèmes d'attribution

L'attribution d'auteur est une question qui intéresse les chercheurs en littérature depuis de nombreuses années. Les médias ont été maintes fois occupés par les discussions sur la paternité de l'œuvre de Shakespeare¹ ou encore par les querelles autour de certaines pièces de Molière, que l'on attribue parfois à Corneille². De manière générale, la question de l'attribution d'auteur est souvent rencontrée par les spécialistes de littérature, lorsqu'ils doivent résoudre le problème posé par certaines œuvres publiées anonymement ou sous un pseudonyme (sans parler des textes composés par des « nègres » que tel auteur connu s'attribue).

L'œuvre critique de Jules Barbey d'Aurevilly (1808-1889), dans ce domaine, représente un cas particulièrement intéressant, qui soulève des questions diverses. Barbey d'Aurevilly mena, entre les années 1830 et 1880, une carrière de journaliste, conjointement à son activité littéraire, et fut l'un des critiques les plus redoutés des lettres françaises. Tout au long de sa vie, dans les pages de divers journaux, parmi lesquels

-
- 1 La controverse date du milieu du XIX^e siècle. Depuis, les œuvres de Shakespeare ont été attribuées à de nombreux écrivains, dont Francis Bacon, Christopher Marlowe ou John Florio. Parmi les études récentes, voir notamment : Lamberto Tassinari, *Shakespeare ? È il nome d'arte di John Florio*, Montréal, Giano Books, 2008 (*John Florio alias Shakespeare* [traduit par Michel Vais], Lormont, Le Bord de l'eau, 2016) ; Daniel Bougnoux, *Shakespeare, Le choix du spectre*, Bruxelles, Les Impressions nouvelles, 2016.
 - 2 Sur la question de la paternité des œuvres de Molière, voir notamment : Dominique Labbé, *Corneille dans l'ombre de Molière. Histoire d'une découverte*, Paris-Bruxelles, Les Impressions nouvelles, 2003 et *Si deux et deux sont quatre, Molière n'a pas écrit « Dom Juan »*, Paris, Max Milo, 2009 ; Eléna Rodionova, *Méthodes linguistiques d'attribution et de datation des œuvres littéraires (contribution à l'étude du problème « Molière – Corneille »)*, thèse en linguistique mathématique et appliquée soutenue à l'université d'État de Saint-Pétersbourg, 2008 ; Mikhaïl Marusenko et Eléna Rodionova, « Mathematical Methods for Attributing Literary Works when Solving the "Molière-Corneille" Problem », *Journal of Quantitative Linguistics*, vol. 17, n° 1, 2010, p. 30-54. Voir aussi le site de monitoring de la querelle dirigé par Georges Forestier : <www.moliere-corneille.paris-sorbonne.fr>.

la *Revue de Caen*, *La Mode*, *Le Pays*, *Le Réveil*, *Le Nain Jaune*, *L'Éclair*, *Le Constitutionnel*, il fut l'observateur attentif de l'actualité littéraire, des nouvelles productions dramatiques, mais aussi de la vie intellectuelle et politique de son temps et des phénomènes de société. Cet aspect de son activité lui importa assez pour qu'il souhaitât recueillir en volumes un grand nombre de ses articles de critique. Cependant, cette entreprise gigantesque de publication en volumes n'a jamais été achevée par l'auteur de son vivant. Une partie de sa production n'a donc paru que dans la presse et, surtout, la partie publiée anonymement ou sous un pseudonyme n'a pas encore été totalement exhumée. En effet, l'écrivain-journaliste, à certains moments de sa carrière, a collaboré à des périodiques sans signer ses articles de son nom, ce qui pose encore aux spécialistes des problèmes d'identification de ces textes et d'attribution de leur paternité.

L'objectif principal de ce projet, associant une équipe de recherche russe et une équipe de recherche française est de conjuguer les ressources de la linguistique, de l'informatique, de la philologie et de l'histoire littéraire pour tenter d'identifier les articles que Barbey d'Aurevilly n'a pas signés de son nom. Il vient en appui de l'entreprise éditoriale de grande ampleur lancée en 2005, sous la direction de Pierre Glaudes et de Catherine Mayaux, aux Belles Lettres, avec le soutien du Centre national du Livre (CNL), et visant à publier l'intégrale de l'œuvre critique de Barbey d'Aurevilly (sept volumes parus à ce jour, les deux derniers en préparation). Cette œuvre critique se compose de trois ensembles.

Un premier ensemble³ est constitué des vingt-six volumes publiés sous le titre *Les Œuvres et les Hommes*, dont les séries successives – quatre au total – s'efforcent de réunir selon une logique thématique – les écrivains religieux et politiques, les historiens, les poètes, les romanciers, etc. – la part la plus importante de la critique journalistique de Barbey d'Aurevilly. Dès les années 1850, celui-ci conçoit en effet le projet de soustraire ses articles à l'éparpillement imposé par des publications au jour le jour, assujetties à une actualité éphémère. Lorsqu'en 1860, le premier volume est enfin publié, l'ambition de l'écrivain est de faire pour l'histoire intellectuelle de la France du XIX^e siècle, l'équivalent de ce que Balzac avait réalisé dans *La Comédie humaine* pour l'histoire des mœurs. C'est ce monument qu'il tentera d'édifier jusqu'à son mort : mais lorsque celle-ci survient, seuls onze volumes, formant la première

3 Publié aux Belles Lettres, t. I-VI, 2004-2014.

série et une partie de la seconde dans *Les Œuvres et les Hommes*, ont vu le jour. Les volumes suivants – soit quinze sur vingt-six – feront l'objet d'une publication posthume dont se chargera Louise Read, l'amie de l'écrivain qui lui servait de gouvernante et de secrétaire. Avec patience, celle-ci continuera, jusqu'à sa mort en 1928, de rassembler les articles disséminés dans les journaux pour pouvoir les publier en recueil.

Un deuxième ensemble comprend les autres recueils publiés par Barbey d'Aurevilly ou par Louise Read en dehors de l'ensemble précédent. C'est le cas de ses chroniques dramatiques⁴ qui, du fait de leur caractère propre, ont fait l'objet d'une publication séparée (*Théâtre contemporain*, cinq volumes, 1887-1889 et 1892-1896) ; le cas aussi d'ouvrages comme *Les Prophètes du passé* (1851) ou *Du dandysme et de George Brummell* (1845), dont la publication à part s'explique par l'antériorité de ces ouvrages par rapport à l'ensemble éditorial formé par *Les Œuvres et les Hommes* ; le cas enfin de textes caractérisés par une forte unité thématique et par la netteté de leur accent, polémique ou satirique. Ainsi de *Goethe et Diderot* (1880) des *Ridicules du temps* (1883) et de leur prolongement, *Les Vieilles Actrices* et *Le Musée des antiques* (1884)⁵.

Enfin, le troisième ensemble⁶ rassemble la masse d'articles que Barbey d'Aurevilly et Louise Read, par la suite, ont oublié de reprendre, à moins qu'ils ne leur aient refusé le changement de statut impliqué par la mise en recueil. Ces textes seraient tombés dans l'oubli si Andrée Hirschi et Jacques Petit, il y a quelques années, ne les avaient exhumés pour les réunir en deux nouveaux volumes, qui restent cependant incomplets⁷ : il y manque des articles inédits, dont ceux qui sont conservés dans le fonds Chastain aux Archives départementales de la Manche, des articles perdus, que l'équipe éditoriale préparant actuellement l'édition de l'œuvre critique aux Belles Lettres a retrouvés, et la plupart des articles anonymes ou publiés par Barbey d'Aurevilly sous un pseudonyme, qui n'ont pas encore été identifiés.

L'objectif de l'édition en cours étant d'offrir un recensement sinon exhaustif du moins le plus complet possible des articles composés par

4 Publiées aux Belles Lettres, t. VII, 2017.

5 Ces textes seront publiés aux Belles Lettres dans le tome VIII, à paraître.

6 Il sera publié aux Belles Lettres dans le tome IX, à paraître.

7 Barbey d'Aurevilly, *Articles inédits. 1852-1884*, publiés par Andrée Hirschi et Jacques Petit, Paris, Les Belles Lettres, 1972 ; Barbey d'Aurevilly, *Premiers articles. 1834-1852*, publiés par Andrée Hirschi et Jacques Petit, Paris, Les Belles Lettres, 1973.

Barbey d'Aureville tout au long de sa carrière de journaliste, ce projet a pour objectif de progresser dans l'identification des articles que le critique n'a pas signés de son nom.

PRÉSENTATION DU CORPUS

Le corpus étudié dans le cadre de cette recherche est formé par les articles de Barbey d'Aureville non identifiés à ce jour. Il a pu être en partie constitué grâce aux informations fournies non seulement par l'écrivain lui-même dans ses *Memoranda* et dans sa correspondance, mais aussi par les études qui se sont intéressées à ses activités journalistiques ou à la presse du XIX^e siècle⁸.

Dans ce corpus, on peut ainsi distinguer cinq ensembles d'articles donnés par Barbey d'Aureville : au *Nouvelliste* entre juillet 1838 et janvier 1839⁹ ; au *Globe* entre avril 1842 et avril 1843 environ¹⁰ ; au *Phare de Dieppe* à l'automne de 1842¹¹ ; à *L'Époque* pendant quelques mois, à

8 Notons cependant que l'outil relevant de la linguistique informatique mis au point dans ce projet permettra aussi de vérifier la véracité des affirmations de Barbey, car il n'est pas exclu par principe qu'il puisse révéler que telle signature ne correspond pas à la paternité déclarée.

9 Voir : Jacques Petit et Philipp John Yarrow, *Barbey d'Aureville journaliste et critique. Bibliographie*, Annales littéraires de l'Université de Besançon, vol. 28, Paris, Les Belles Lettres, 1959, p. 70-74 ; Alain Néry, « Barbey d'Aureville au *Nouvelliste* (1838-1839) », dans *Barbey d'Aureville, ombre et lumière*, textes réunis par Joseph-Marc Bailbé, actes du colloque organisé à Rouen les 17 et 18 mai 1989, Mont-Saint-Aignan, Centre d'art esthétique et littéraire, 1990, p. 37-44.

10 Voir les lettres de Barbey à Trebutien des 16 avril 1842, [16 ou 23] avril [1843], 18 mai [1843], 29 mars 1844 et 17 juin 1850 ; Jules Barbey d'Aureville, *Correspondance générale*, Paris, Les Belles Lettres, t. I, 1980, p. 97, p. 106, p. 111 et p. 162 ; t. II, 1982, p. 173.

11 Voir les lettres de Barbey à Trebutien des 25 mars et 18 mai 1843, janvier et 4 septembre 1844 (*ibid.*, t. I, p. 103, 111, 154 et 182) et les articles suivants : Georges Lebas, « Barbey d'Aureville, polémiste à Dieppe. Une élection législative en 1842 », Dieppe, Imprimerie 7, rue Victor-Hugo, 1909, rééd. « Barbey d'Aureville, polémiste en province », *Mercur de France*, t. CIII, n° 384, 16 juin 1913, p. 767-776 ; André Boudier, À travers les siècles. Dieppe, dieppois, dieppoiseries..., 4^e série, Dieppe, Les Informations dieppoises, 1951, « Dieppe et Barbey d'Aureville », p. 10-12 ; Jacques Giffard, « Barbey d'Aureville journaliste à Dieppe, une campagne électorale en 1842 », *Précis analytique des travaux*, Académie des sciences, belles lettres et arts de Rouen, 1978, p. 133-176.

partir de septembre 1845¹² ; au *Figaro* entre le 7 novembre 1861 et le 12 janvier 1862, dans la série des « Lettres de Junius¹³ ».

Le degré de difficulté de l'attribution d'auteur varie selon les ensembles. À l'exception de dix-neuf chroniques dramatiques qui portent sa signature, la plupart des textes donnés par Barbey d'Aurevilly au *Nouvelliste* sont anonymes. À la même époque, cependant, l'écrivain-journaliste tient un journal intime (le premier de ses *Memoranda*) qui donne parfois des renseignements précis sur les articles en préparation (sujet, date de parution, taille de l'article, etc.), ce qui a permis de repérer certains textes parus dans ce journal. Ainsi, le 27 décembre 1838, Barbey d'Aurevilly révèle qu'il vient de composer « un *premier-Paris* sur le Mexique¹⁴ ». Cet article a été publié, en effet, le lendemain, 28 décembre 1838. Le 3 janvier 1839, il note qu'il a « griffonné un article de trois colonnes¹⁵ ». Les articles de cette taille étant rares dans le journal, cette indication permet d'identifier le texte de Barbey d'Aurevilly : l'article, qui porte sur la politique étrangère du gouvernement, a paru dans *Le Nouvelliste* le 4 janvier 1839.

Lorsque les indications des *Memoranda* sont trop vagues, on dispose du recueil des numéros du *Nouvelliste*, conservé à Paris à la bibliothèque de l'Arsenal. Un annotateur anonyme a marqué les articles de l'initiale du rédacteur : presque tous les articles de Barbey d'Aurevilly dont l'identification est certaine du fait des informations fournies par les *Memoranda* sont marqués de la lettre B. L'annotateur inconnu qui a procédé à l'identification s'est montré bien renseigné, mais il arrive que, dans quelques cas assez rares, il ait omis de signaler par la lettre B un texte que les sources philologiques permettent d'attribuer à Barbey d'Aurevilly. De nouveaux articles sont donc à attribuer et l'on a besoin d'étayer sur d'autres éléments de preuve l'identification de ceux qui sont marqués d'un B sans que les *Memoranda* corroborent cette identification. À ce jour, les articles du *Nouvelliste* dont l'attribution à Barbey d'Aurevilly est incertaine sont environ au nombre de cent vingt.

12 Voir les lettres de Barbey à Trebutien des 19 septembre 1845 et 2 janvier 1846 (*Correspondance générale*, éd. citée, t. II, p. 47 et p. 53) et à Yzarn-Freissinet du 16 août 1845 (*ibid.*, p. 44).

13 Voir Marie-Ève Thérenty, *La Littérature au quotidien. Poétiques journalistiques au XIX^e siècle*, Paris, Éditions du Seuil, 2007, p. 73-74.

14 Barbey d'Aurevilly, *Œuvres romanesques complètes*, Paris, Gallimard, coll. « Bibliothèque de la Pléiade », t. II, 1966, p. 1007.

15 *Ibid.*, p. 1011.

Les articles donnés par l'écrivain au *Phare de Dieppe*¹⁶ à l'automne de 1842, lors de la campagne électorale visant à élire le nouveau député de Dieppe à la suite au décès de Charles Bérigny, ne sont pas signés. Dans sa correspondance avec Trebutien, Barbey d'Aurevilly a évoqué lui-même sa participation au « journalisme en province » et la part active prise lors de l'élection à Dieppe :

Depuis que je ne vous ai écrit, mon cher Trebutien, j'ai fait du journalisme en province. J'ai été envoyé à Dieppe pour brasser une élection, et cette élection, je l'ai enlevée contre vent et marée. J'ai battu les journaux de l'administration et rallié des légitimistes à un candidat qui ne l'était pas. Ça été un coup de partie bien menée et qui m'a fait honneur¹⁷.

Il fait au moins deux fois allusion au soutien apporté à l'un des candidats, Charles-Louis Levavasseur : « Moi qui ai fait un député à Dieppe du plus gros commerçant de Rouen » ; « M^{me} Levavasseur (née de Vaudeul) est une de ces adorables mésalliées comme on a l'angoisse et quelquefois le bonheur d'en rencontrer dans notre aimable société sans divorce. C'est la perle jetée à vous savez bien qui. Son mari est le député que j'ai fait, mais pardieu ! pas à mon image¹⁸. »

Dans *Le Phare de Dieppe*, l'écrivain-journaliste mena en effet une véritable lutte politique contre un rédacteur inconnu de *La Vigie*, qui soutenait le candidat adversaire, Gustave Rouland. Il joua un rôle décisif dans l'élection de Levavasseur, qui eut lieu le 13 novembre 1842. Les circonstances qui conduisirent Barbey d'Aurevilly, alors collaborateur au *Moniteur de la Mode* et rédacteur au *Globe*, à s'installer quelque temps à Dieppe pour prêter son concours, en tant qu'agent électoral, à Levavasseur, sont encore peu connues. Les textes qu'il donna à partir du 19 octobre 1842 au *Phare de Dieppe*, en omettant volontairement sa signature, seraient au moins au nombre de dix. En 1978, Jacques Giffard a publié dans le cadre de travaux conduits à l'Académie des sciences, belles lettres et arts de Rouen tous les textes réunis entre le 19 octobre et le 19 novembre 1842 dans *Le Phare de Dieppe*, à la rubrique « Élections¹⁹ ». Dans cet ensemble

16 La bibliothèque Jean-Renoir de Dieppe conserve la collection complète du *Phare de Dieppe*, journal bi-hebdomadaire publié du 24 août 1836 au 1^{er} février 1843, fondé par Antoine Corsange.

17 Lettre de Barbey à Trebutien, 25 mars 1843 (*Correspondance générale*, éd. citée, t. I, p. 103).

18 Lettres de Barbey à Trebutien, janvier et 4 septembre 1844 (*ibid.*, p. 154 et p. 182).

19 Jacques Giffard, « Barbey d'Aurevilly journaliste à Dieppe, une campagne électorale en 1842 », art. cité.

il faudrait néanmoins distinguer les textes rédigés par Barbey de ceux qu'on doit probablement à Antoine Corsange (1786-1860), imprimeur et fondateur du journal.

Les articles publiés dans *Le Figaro*, dans la série des « Lettres de Junius » posent aussi de délicats problèmes d'attribution. Le journal ressuscité par Hippolyte de Villemessant en 1854 propose volontiers, dans les années soixante, des *steeple-chases*, « courses à la plume » arbitrées par les lecteurs, dans lesquelles rivalisent des chroniqueurs en vue, sur le modèle des courses de chevaux. Ainsi, à l'automne de 1861, un *steeple-chase* est organisé entre cinq « Junius » journalistes et pamphlétaires, qui se lancent dans la compétition et, devant écrire à tour de rôle un article, reçoivent pour l'occasion des casques de différentes couleurs. De l'aveu du directeur du *Figaro* lui-même, la casaque blanche est endossée par Jules Vallès, la casaque verte par Alphonse Duchêne²⁰. Sous les casques orange et bleue se cacheraient respectivement, selon toute probabilité, Charles Monselet²¹ et Aurélien Scholl. De fortes présomptions invitent, enfin, à identifier Barbey d'Aureville à la casaque écarlate (sa couleur fétiche), ce que confirme Alfred Delvau dans l'édition qu'il fait paraître chez Dentu en 1862 de ces *Lettres à Junius* :

La casaque orange et la veste écarlate, la première lettre et la dernière, sont signées et désignées suffisamment pour moi [...].

[...] si vous ne l'aviez su d'avance, auriez-vous reconnu dans l'un l'auteur humoristique des *Oubliés et dédaignés du treizième siècle*, et, dans l'autre, l'auteur délicat du *Dandysme*? C'est eux, pourtant; ou, si ce n'est pas eux, ce sont donc leurs frères! Je n'en connaissais pas, jusqu'ici, à M. Charles Monselet et à M. J. Barbey d'Aureville²².

Mais les dix-sept lettres signées par Junius, qui paraissent du 31 octobre 1861 au 12 janvier 1862, ne précisent la couleur de la casaque qu'à la

20 « En effet, la casaque blanche avait été endossée par notre collaborateur, M. Jules Vallès, l'auteur de l'article très remarqué : *Les Morts* » (*Le Figaro*, 14 novembre 1861, n. 2, signée H[ippolyte] de V[illemessant]). « À la fin d'un bal travesti, les gens de bonne compagnie ôtent leur masque. – Allons, mon cher Duchesne, exécutez-vous! » (*ibid.*, n. 3.)

21 C'est Alfred Delvau qui attribue à Charles Monselet la casaque orange. Voir la citation qui suit dans le texte.

22 [Alfred Delvau et Alphonse Duchesne], *Lettres de Junius*, Paris, Dentu, 1862, p. 33 et p. 35.

date du 7 novembre 1861, et l'on manque d'éléments pour identifier les auteurs des autres lettres.

Les articles donnés par Barbey d'Aurevilly au *Globe* et à *L'Époque* posent, enfin, les plus sérieux problèmes d'identification, car ils sont anonymes et l'on ne dispose à ce jour que de quelques indices (en particulier dans l'œuvre critique, les *Memoranda* et la correspondance) pour les repérer dans ces journaux et pour les attribuer à l'écrivain-journaliste²³.

En l'état actuel des études aurevilliennes, les chercheurs qui se sont intéressés à ce corpus sont fort rares. Les principaux travaux qui ont porté sur la biographie de Barbey d'Aurevilly et sur son œuvre journalistique et critique se contentent de signaler le problème, sans avancer dans son élucidation²⁴. C'est la raison pour laquelle ce projet a été conçu : il vise à réparer une sérieuse lacune dans l'identification de l'ensemble des contributions de Barbey d'Aurevilly à la presse de son temps.

LA PREMIÈRE PHASE

Le projet comprend deux phases : la première porte sur le traitement d'un corpus limité de textes, sur l'attribution de dix articles de presse dont Barbey d'Aurevilly pourrait être un des auteurs (tableau 1) et sur le choix d'une méthode adéquate d'attribution ; la seconde porte sur la totalité du corpus d'articles anonymes ou pseudonymiques. Lors de la première étape visant à attribuer les textes, la méthode de reconnaissance des formes a été utilisée²⁵.

23 Voir ci-dessus, notes 7-8.

24 Hormis ceux qui ont déjà été cités plus haut, ces travaux sont les suivants : Jean Canu, *Barbey d'Aurevilly*, Paris, Robert Laffont, 1945 ; Gisèle Corbière-Gille, *Barbey d'Aurevilly, critique littéraire*, Genève-Paris, Droz-Minard, 1962 ; Michel Lécureur, *Barbey d'Aurevilly, le Sagittaire*, Paris, Fayard, 2008 ; Marie-Françoise Melmoux-Montaubin, *L'Écrivain-journaliste au XIX^e siècle : un mutant des lettres*, Saint-Étienne, Éditions des Cahiers intempestifs, 2003 [notamment p. 19-77] ; Jacques Petit, *Barbey d'Aurevilly critique*, Annales littéraires de l'université de Besançon, vol. 53, Paris, Les Belles Lettres, 1963.

25 Voir Pierre Glaudes, Francesca Guglielmi, Catherine Mayaux, Mikhail Marusenko, Yekaterina Kuralesina, Mariya Miretina, Yekaterina Nikitina, Mariya Solovyeva, Olga Khutoretskaya, « Zhyul' Barbe d'Orevil'i i korpus gazetnykh statey : probleme atributsii (chast' I) » [« Jules Barbey d'Aurevilly et le corpus journalistique :

Code	Article, références	Nombre de phrases du texte d'auteur
Texte 1	« Paris, 21 juillet » [la Gazette de Leipsick et Louis Bonaparte], Le Nouvelliste, journal politique du soir et du matin, 22 juillet 1838, p. 1.	15
Texte 2	« Du droit des gens en Europe », Le Nouvelliste, journal politique du soir et du matin, 29 juillet 1838, p. 1.	31
Texte 3	« Paris, 2 août » [Lord Palmerston et la question d'Orient], Le Nouvelliste, journal politique du soir et du matin, 3 août 1838, p. 1.	9
Texte 4	« Paris, 21 août » [La question hollando-belge], Le Nouvelliste, journal politique du soir et du matin, 22 août 1838, p. 1.	30
Texte 5	[L'archevêque de Paris], Le Nouvelliste, journal politique du soir et du matin, 1 ^{er} septembre 1838, p. 1.	21
Texte 6	« Candidature de M. Charles Levassieur », Le Phare de Dieppe, 19 octobre 1842, p. 1.	10
Texte 7	« Dieppe. Élections. Candidature de M. Charles Levassieur », Le Phare de Dieppe, 22 octobre 1842, p. 1-2.	66
Texte 8	« Candidature de M. Charles Levassieur », Le Phare de Dieppe, 26 octobre 1842, p. 1.	50
Texte 9	« Lettres de Junius. Casaque écarlate. II », Le Figaro, 7 novembre 1861, p. 6 (p. 237-242 du volume).	54
Texte 10	« Lettres de Junius. VII », Le Figaro, 12 décembre 1861, p. 3-4 (p. 103-122 du volume).	108

TABLEAU 1 – Textes à attribuer (anonymes et publiés sous pseudonyme).

Pour la première phase on a constitué deux classes *a priori* (tableaux 2 et 3).

problèmes d'attribution (première partie) », *Drevnyaya i Novaya Romaniya*, vol. 21, 2018, p. 17-40. En russe = Глод Пьер, Гульельми Франческа, Майо Катрин, Марусенко Михаил, Куралесина Екатерина, Миретина Мария, Никитина Екатерина, Соловьева Мария, Хуторецкая Ольга, « Жюль Барбе д'Оревилли и корпус газетных статей : проблемы атрибуции (часть I) », *Древняя и Новая Романия*, вып. 21, 2018, с. 17-40.

N°	Article, références
1	M. de M. ..., « De la démocratie nouvelle ou des mœurs et de la puissance des classes moyennes en France, par Édouard Alletz », <i>Revue française et étrangère ou nouvelle revue encyclopédique</i> , [deuxième semestre] 1837, p. 258-277.
2	« J.-B. d'A... », « Théâtre de la Renaissance. <i>Ruy Blas</i> et sa préface », <i>Le Nouvelliste, journal politique du soir et du matin</i> , 3 décembre 1838, p. 1-3.
3	« J. B. d'A. », « <i>Histoire de la papauté pendant les XVI^e et XVII^e siècles</i> , par M. Léopold Ranke, professeur à l'université de Berlin, traduite de l'allemand par M. Haiber, publiée et précédée d'une introduction par M. Alexandre de Saint-Chéron », <i>Journal général de l'instruction publique et des cours scientifiques et littéraires</i> , n° 85, 12 mai 1838, p. 519-520, dans la rubrique « Histoire », dans la « Partie non officielle ».
4	« Jules Barbey d'Aurevilly », « <i>Histoire de la papauté pendant les XVI^e et XVII^e siècles</i> , par Léopold Ranke, professeur à Berlin, traduite par M. Haiber et précédée d'une préface par M. de Saint-Chéron (2 ^e article) », <i>Journal général de l'instruction publique et des cours scientifiques et littéraires</i> , 15 septembre 1838, dans la rubrique « Histoire », dans la « Partie non officielle », p. 817-819.
5	J. B. d'A., « <i>Théâtre-Français. Il faut que jeunesse se passe</i> , par M. de Rougemont », <i>Le Nouvelliste, propagateur des doctrines constitutionnelles</i> , 6 juillet 1839, p. 1.
6	« J. B. d'A... », « <i>Encyclopédie du XIX^e siècle</i> », <i>Journal général de l'instruction publique et des cours scientifiques et littéraires</i> , n° 33, 24 avril 1839, p. 234-235, dans la rubrique « Littérature », dans la « Partie non officielle » ; <i>Le Nouvelliste, propagateur des doctrines constitutionnelles</i> , 27 avril 1839, p. 13.
7	J. B. D'A., « De l'Angleterre et de lord Palmerston », <i>La Mode</i> , 6 juillet 1850, p. 36.
8	J. Barbey d'Aurevilly, « Un Machiavel inédit. L'esprit d'un homme d'État, par Nicolas Machiavel, traduit par M. Joseph Delarosa », <i>Le Pays</i> , 14 novembre 1861, p. 3, dans la « Bibliographie ».
9	« Les chroniqueurs », <i>Le Nain jaune</i> , 24 février 1866, dans les « Ridicules du temps », p. 2-3.
10	« Les chats de la critique », <i>Le Nain jaune</i> , 19 mai 1866, dans les « Ridicules du temps », p. 2-3.

TABLEAU 2 – Composition de la classe Ω (Barbey d'Aurevilly).

N°	Article, références
1	Adolphe Granier de Cassagnac, « Théâtre de la Renaissance. <i>Ruy Blas</i> , drame en 5 actes et en vers, par M. Victor Hugo », <i>La Presse</i> , 11 novembre 1838, p. 1-3.
2	Théophile Gautier, « Benvenuto Cellini », <i>La Presse</i> , 17 septembre 1838, p. 1-3.
3	Théophile Gautier, « <i>La Gypsy</i> , ballet pantomime en trois actes », <i>La Presse</i> , 4 février 1839, p. 1-2.
4	<i>La Vigie</i> , « Dieppe. Élections », 23 octobre 1842, p. 1.
5	<i>La Vigie</i> , « Dieppe. Élections. Candidature de M. Rouland », 27 octobre 1842, p. 1-2.
6	Jules Vallès, « Lettres de Junius. Casaque blanche. II », <i>Le Figaro</i> , 7 novembre 1861, p. 5-6 (p. 228-237 du volume).
7	Alphonse Duchesne, « Lettres de Junius. Casaque verte. II », <i>Le Figaro</i> , 7 novembre 1861, p. 3-4 (p. 217-227 du volume).
8	Aurélien Scholl, « Causerie théâtrale », <i>L'Artiste</i> , 25 avril 1858, p. 282-284.
9	Charles Monselet, <i>Le Monde illustré</i> , 23 novembre 1861, p. 746-747.
10	Louis Veuillot, « La Question romaine », <i>L'Univers</i> , 8 juillet 1859, p. 1.

TABLEAU 3 – Composition et puissance de la classe Ω (échantillon de fond).

Conformément à la méthode que l'on a utilisée, la première étape du traitement des données textuelles consiste en la description des classes *a priori* avec le langage des paramètres qui constituent le vocabulaire des paramètres *a priori*. Pour y parvenir, on a constitué deux échantillons aléatoires de 100 phrases chacun, tirés des textes composant les classes *a priori* (tableaux 2 et 3) et l'on a évalué les valeurs de tous les paramètres pertinents pour la langue française.

À cette étape, le vocabulaire *a priori* standard comptant 54 paramètres qui caractérisent la structure et la composition d'une phrase et de ses composantes a été allégé de deux paramètres (X38, X39) qui n'existent pas en français (tableau 4)²⁶. Dans la phase de dépouillement des textes, il s'est avéré que les paramètres X46 et X47 n'ont pas de valeurs différentes de zéro, raison pour laquelle ces paramètres ont été éliminés de l'ensemble.

26 *Ibid.*, p. 31.

Codes	Nominations
X01	Nombre de mots dans une phrase simple indépendante
X02	Nombre de propositions simples dans une phrase
X03	Nombre de propositions principales
X04	Nombre de propositions coordonnées
X05	Nombre de propositions coordonnées sans verbe au mode personnel
X06	Nombre de propositions subordonnées
X07	Nombre de propositions subordonnées du 1 ^{er} niveau
X08	Nombre de propositions subordonnées du 2 ^e niveau
X09	Nombre de propositions subordonnées du 3 ^e niveau
X10	Nombre de propositions subordonnées du 4 ^e niveau et des niveaux postérieurs
X11	Nombre de propositions simples sans substantif sujet
X12	Nombre de propositions subordonnées sans verbe au mode personnel
X13	Nombre de propositions insérées
X14	Nombre de propositions embrassantes
X15	Nombre de mots de la première classe (mots lexicaux)
X16	Nombre de mots de la seconde classe (mots grammaticaux, mots accessoires)
X17	Nombre de substantifs
X18	Nombre d'adjectifs
X19	Nombre de pronoms
X20	Nombre de verbes au mode personnel
X21	Nombre de verbes au mode impersonnel
X22	Nombre d'adverbes
X23	Nombre de prépositions
X24	Nombre de conjonctions
X25	Nombre de conjonctions de subordination
X26	Nombre de conjonctions de coordination
X27	Nombre d'attributs
X28	Nombre de compléments directs
X29	Nombre de compléments indirects
X30	Nombre de sujets
X31	Nombre de pronoms-sujets

X32	Nombre de groupes de termes multiples
X33	Nombre de membres des groupes de termes multiples
X34	Nombre de prédicats multiples
X35	Nombre de groupes multiples de compléments du verbe
X36	Nombre de tours participiaux
X37	Nombre de membres des tours participiaux
X40	Nombre d'adjectifs qualificatifs et d'adjectifs déterminatifs
X41	Nombre d'adjectifs verbaux
X42	Nombre d'appositions
X43	Nombre d'appositions nominales
X44	Nombre de termes isolés
X45	Nombre de membres des termes isolés
X46	Nombre de propositions absolues
X47	Nombre de termes des propositions absolues
X48	Nombre de propositions infinitives
X49	Nombre de termes des propositions infinitives
X50	Nombre de groupes nominaux minimaux
X51	Nombre de syntagmes nominaux
X52	Nombre de termes des syntagmes nominaux
X53	Nombre de mots de la première classe dans le syntagme nominal
X54	Nombre de mots de la seconde classe dans le syntagme nominal

TABLEAU 4 – Jeu de paramètres *a priori*.

Conformément au schéma de Bonnegarde, les paramètres informatifs sont sélectionnés en deux étapes : à la première, ils sont sélectionnés parmi ceux qui sont utiles pour classifier les objets dont l'appartenance est bien connue. Pour cela, ces objets sont soumis à la classification automatique dans un espace à une dimension en utilisant une valeur de seuil qui est comparée à la valeur du critère de classification. Par la suite, vu les valeurs du critère t de Student :

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{1/2}} \quad (1)$$

le vocabulaire de paramètres *a priori* est divisé en deux sous-ensembles (tableau 5).

Paramètres	Ω (Barbey d'Aureville)			Ω (échantillon de fond)	t
	\bar{X}_{1j}	σ_{1j}	\bar{X}_{2j}	σ_{2j}	
X02	3,180	2,276	3,030	2,022	0,49
X03	1,210	1,192	1,170	1,101	0,25
X04	0,380	0,814	1,010	1,494	3,70
X05	0,010	0,101	0,050	0,261	1,43
X06	1,730	1,808	1,330	1,477	1,71
X07	1,230	1,179	0,980	0,985	1,63
X08	0,400	0,696	0,300	0,611	1,08
X09	0,080	0,273	0,040	0,197	1,19
X10	0,020	0,200	0,020	0,141	–
X11	2,060	1,841	1,400	1,664	2,66
X12	0,040	0,197	0,030	0,223	0,34
X13	0,485	0,705	0,180	0,411	3,73
X14	0,430	0,607	0,190	0,465	3,14
X15	20,580	14,757	17,170	11,377	1,83
X16	10,060	8,770	14,340	11,088	3,03
X17	6,240	5,253	7,510	5,482	1,67
X18	3,120	2,595	2,620	2,510	1,39
X19	2,830	2,421	2,440	2,311	1,17
X20	3,110	2,283	3,090	2,175	0,06
X21	1,620	1,763	0,890	1,569	3,09
X22	2,260	2,168	1,370	1,522	3,36

X23	4,090	3,975	4,160	3,287	0,14
X24	2,760	2,682	2,720	2,483	0,11
X25	1,730	1,847	1,450	1,527	1,17
X26	1,030	1,123	1,300	1,460	1,47
X27	0,750	0,978	0,790	0,977	0,29
X28	1,740	1,873	1,750	1,940	0,04
X29	1,000	1,073	1,040	1,442	0,22
X30	2,820	2,213	3,020	2,183	0,64
X31	1,630	1,643	1,330	1,505	1,35
X32	1,480	2,368	1,320	2,093	0,51
X33	4,380	8,980	4,040	7,705	0,29
X34	0,160	0,598	0,480	1,168	2,44
X35	0,320	0,803	0,370	0,991	0,39
X36	0,050	0,219	0,050	0,261	–
X37	0,300	1,508	0,150	0,744	0,89
X40	2,700	2,456	2,140	2,074	1,74
X41	0,170	0,451	0,280	0,552	1,54
X42	1,660	2,036	0,710	1,274	3,96
X43	1,410	1,730	0,500	1,059	4,49
X44	0,860	1,231	0,440	0,756	2,91
X45	4,390	9,321	1,760	3,644	2,63
X48	0,040	0,197	0,060	0,239	0,65
X49	0,150	0,783	0,300	1,259	1,01
X50	2,450	2,694	2,510	2,320	0,17
X51	3,450	3,319	2,710	2,587	1,76
X52	12,270	11,472	8,070	7,917	3,01
X53	7,980	7,005	7,380	6,788	0,62
X54	4,313	4,778	5,120	5,890	1,06

TABLEAU 5 – Paramètres pertinents pour la séparation des classes *a priori*
(la valeur critique de t est de 1,96).

Ainsi, à partir d'un vocabulaire *a priori* de 49 paramètres, on a obtenu un sous-ensemble de 12 paramètres (X04, X11, X13, X14, X16, X21, X22, X42, X43, X44, X45, X52) pertinents pour la séparation des classes *a priori*.

La compression ultérieure de l'espace paramétrique consiste à traiter la matrice corrélationnelle des relations entre paramètres faisant partie de ce sous-ensemble, avec les paramètres de leur groupe (corrélation moyenne intragroupe) (équation 2) et avec les paramètres qui ne font pas partie de ce groupe (corrélation moyenne extragroupe) (équation 3)²⁷ :

$$\bar{r}^m = \left[\left(\sum_{i=1}^m |r_{ij}| - 1 \right) / (m-1) \right] \quad (2)$$

$$\bar{r}^{n-m} = \frac{\left[\left(\sum_{i=1}^n |r_{ij}| - 1 \right) - \left(\sum_{i=1}^m \bar{r}_{ij} \right) - 1 \right]}{n-m-1} \quad (3)$$

Les résultats des calculs du coefficient d'efficacité de chaque paramètre pertinent pour la séparation des classes *a priori* :

$$E_j = \bar{r}_j^{n-m} / \bar{r}_j^m \quad (4)$$

figurent dans le tableau 6.

Paramètres	\bar{r}_j^{n-m}	\bar{r}_j^m	E_j
X04	0,197	0,110	1,783
X11	0,430	0,359	1,196
X13	0,209	0,254	0,824
X14	0,223	0,268	0,832
X16	0,509	0,379	1,342
X21	0,324	0,331	0,976
X22	0,335	0,346	0,968
X34	0,158	0,112	1,406

27 *Ibid.*, p. 31-32.

X42	0,337	0,420	0,803
X43	0,344	0,426	0,807
X44	0,288	0,363	0,793
X45	0,307	0,381	0,804
X52	0,444	0,462	0,961

TABLEAU 6 – Efficacité des paramètres.

Le tableau 6 montre que 4 paramètres qui ont une valeur de critère d'efficacité supérieure à 1 (X04, X11, X16, X34) répondent à la condition d'efficacité que l'on a formulée.

Dans ce qui suit, on a décrit chaque objet comme un vecteur à 4 dimensions et classifié les objets dans un espace à 4 dimensions dont ces paramètres sont les axes. Les résultats du dépouillement statistique des classes *a priori* et des objets à attribuer sont donnés dans les tableaux 7-9.

Paramètres	\bar{x}_j	σ_j
X04	0,38	0,813
X11	0,04	0,196
X16	10,06	8,769
X34	0,16	0,598

TABLEAU 7 – Coordonnées de l'étalon de la classe Ω
(Barbey d'Aureville).

Paramètres	\bar{x}_j	σ_j
X04	1,01	1,494
X11	1,4	1,663
X16	14,34	11,087
X34	0,48	1,167

TABLEAU 8 – Coordonnées de l'étalon de la classe Ω
(échantillon de fond).

Textes	Paramètres	\bar{X}_j	σ_j
Texte 1	X04	0,267	0,704
	X11	2,067	1,486
	X16	8,800	4,960
	X34	0,267	0,704
Texte 2	X04	0,742	1,237
	X11	1,871	1,648
	X16	13,355	7,369
	X34	0,194	0,601
Texte 3	X04	0,900	1,197
	X11	2,200	2,348
	X16	15,000	11,576
	X34	0,000	0,000
Texte 4	X04	1,063	1,318
	X11	2,188	1,533
	X16	15,313	11,026
	X34	0,250	0,672
Texte 5	X04	0,435	0,843
	X11	1,696	1,105
	X16	12,391	6,330
	X34	0,087	0,417
Texte 6	X04	0,444	0,882
	X11	2,333	1,732
	X16	9,556	8,398
	X34	0,667	1,000
Texte 7	X04	0,444	0,857
	X11	1,397	1,225
	X16	12,460	9,652
	X34	0,111	0,512
Texte 8	X04	0,540	1,014
	X11	1,420	1,247

	X16	13,580	11,468
	X34	0,300	0,863
Texte 9	X04	0,737	1,357
	X11	2,333	1,949
	X16	7,737	9,224
	X34	0,175	0,685
Texte 10	X04	1,566	2,931
	X11	2,470	2,716
	X16	11,651	21,483
	X34	0,422	2,090

TABLEAU 9 – Coordonnées des objets à attribuer.

Les résultats de la mise en œuvre de l'algorithme déterministe de l'attribution figurent dans le tableau 10.

Objets	Paramètres	t (Barbey d'Aureville)	t (échantillon de fond)
Texte 1	X04	0,569	3,160
	X11	5,274	1,594
	X16	0,812	3,271
	X34	0,558	0,988
Texte 2	X04	1,529	1,001
	X11	6,172	1,387
	X16	2,075	0,571
	X34	0,272	1,801
Texte 3	X04	1,343	0,270
	X11	2,909	1,052
	X16	1,312	0,173
	X34	2,674	4,111
Texte 4	X04	2,765	0,190
	X11	7,902	2,476

	X16	2,457	0,434
	X34	0,677	1,381
Texte 5	X04	0,283	2,493
	X11	7,159	1,040
	X16	1,471	1,130
	X34	0,692	2,700
Texte 6	X04	0,211	1,715
	X11	3,970	1,553
	X16	0,172	1,589
	X34	1,496	0,529
Texte 7	X04	0,477	3,068
	X11	8,719	0,014
	X16	1,601	1,142
	X34	0,556	2,766
Texte 8	X04	0,970	2,269
	X11	7,778	0,082
	X16	1,909	0,387
	X34	1,030	1,066
Texte 9	X04	1,809	1,169
	X11	8,859	3,039
	X16	1,545	4,002
	X34	0,142	2,060
Texte 10	X04	3,575	1,568
	X11	8,134	3,134
	X16	0,632	1,032
	X34	1,104	0,227

TABLEAU 10 – Résultats de la mise en œuvre de l'algorithme déterministe.

L'analyse du tableau 10 montre que d'après les résultats de la mise en œuvre de l'algorithme déterministe, il n'y a que deux objets – Texte 2 et Texte 6 (les valeurs du t-critère pour tous les paramètres inférieurs à la valeur critique 1,96 pour $\alpha = 0,05$) qui peuvent être attribués à

la classe Ω (échantillon de fond). Il faut souligner que les différences maximales par rapport à la classe Ω (Barbey d'Aurevilly) sont dues au paramètre X11 – nombre de propositions simples sans substantif sujet. Il semble que Barbey d'Aurevilly n'a pratiquement jamais utilisé cette construction syntaxique.

Après l'algorithme déterministe, c'est l'algorithme probabiliste évaluant la probabilité de l'appartenance des objets à attribuer aux classes *a priori* qui est mise en œuvre :

$$P_{ij} = \frac{1}{d_{ji}} \left(\sum_k \frac{1}{d_{jk}} \right)^{-1} \tag{5}$$

S'agissant de deux classes *a priori*, il est utile de fixer la valeur du seuil à laquelle un objet est attribué à une classe égale à 0,66. Les résultats de la mise en œuvre de l'algorithme probabiliste figurent dans le tableau 11.

Objets	D (Barbey d'Aurevilly)	D (échantillon de fond)	P (Barbey d'Aurevilly)	P (échantillon de fond)
Texte 1	2,391	5,633	0,70	0,30
Texte 3	5,437	1,148	0,17	0,83
Texte 4	5,716	1,273	0,18	0,82
Texte 5	2,860	2,090	0,42	0,58
Texte 7	2,758	1,997	0,42	0,58
Texte 8	3,786	0,911	0,19	0,81
Texte 9	3,283	6,681	0,67	0,33
Texte 10	3,148	2,947	0,48	0,52

TABLEAU 11 – Résultats de la mise en œuvre de l'algorithme probabiliste.

Les colonnes 2-3 du tableau 11 donnent les valeurs des distances euclidiennes entre les objets à attribuer et les colonnes 4-5, les probabilités de l'appartenance des objets à ces classes calculées à la base de ces distances²⁸.

28 *Ibid.*, p. 33.

Ainsi, par l'algorithme probabiliste, les Texte 1 et Texte 9 sont attribués à la classe Ω (Barbey d'Aureville) et les objets Texte 3, Texte 4 et Texte 8 – à la classe Ω (échantillon de fond). Trois objets (Texte 5, Texte 7 et Texte 10) restent non attribués.

Finalement, la matrice des décisions du système de reconnaissance aboutit à ceci (tableau 12) :

Objets	Ω (Barbey d'Aureville)	Ω (échantillon de fond)
Texte 1	1*	0
Texte 2	0	1
Texte 3	0	1*
Texte 4	0	1*
Texte 5	-	-
Texte 6	0	1
Texte 7	-	-
Texte 8	0	1*
Texte 9	1*	0
Texte 10	-	-

TABLEAU 12 – Matrice des décisions. (1 : décision prise par l'algorithme déterministe, 1* : décision prise par l'algorithme probabiliste).

Il reste à noter que dans le cas de deux classes *a priori*, si l'on s'était tenu à la valeur standard du seuil de probabilité ($P = 0,5$) les objets restés non attribués auraient été attribués à la classe Ω (échantillon de fond).

RÉSULTATS DE LA PREMIÈRE PHASE

Les résultats du traitement du corpus limité de textes de journaux (10 textes à attribuer) ont permis de cerner les problèmes qui vont se poser pendant le traitement du corpus complet de textes susceptible d'être attribués à Barbey d'Aureville qui peut compter plusieurs dizaines d'œuvres.

De nombreux textes sont de faible volume, ce qui ne permet pas d'estimer d'une façon fiable leurs caractéristiques statistiques. Il s'ensuit que pendant la seconde étape du projet on sera amené à utiliser une méthode d'attribution élaborée à propos de textes courts, bien que leur efficacité et leur fiabilité puissent être moindres.

L'utilisation de la classe Ω (échantillon de fond), qui représente en quelque sorte « la température moyenne » est une mesure forcée due à la structure de l'hypothèse formulée à la première étape du projet. Par la suite, il sera utile de décomposer la classe Ω (échantillon de fond) en plusieurs classes *a priori* personnalisées appartenant aux autres auteurs présumés (Jules Vallès, Alphonse Duchêne, Charles Monselet, Aurélien Scholl, etc.).

La formulation d'une nouvelle hypothèse d'attribution qui comprendrait un plus grand nombre de classes *a priori* permettrait certainement d'améliorer la qualité et la fiabilité de l'attribution.

L'HYPOTHÈSE D'ATTRIBUTION À LA SECONDE PHASE

À la seconde étape, afin d'attribuer les textes anonymes, pseudonymiques et litigieux dont Barbey d'Aureville pourrait être l'auteur, on a utilisé le sous-corpus du corpus journalistique, constitué par la partie française du collectif scientifique.

Dans le corpus principal, les textes sont divisés en cinq groupes : des journaux *Le Nouvelliste* (1838–1839), *Le Globe* (1842–1843), *Le Phare de Dieppe* (1842), *L'Époque* (1845) et *Le Figaro* (1861–1862). Le corpus à traiter comprend 88 textes. Certains documents sont signés du nom de Barbey d'Aureville ou de ses initiales.

Pour le traitement, les titres des articles ont été remplacés par les codes T01 – T88 afin d'éviter l'adaptation des résultats à l'hypothèse traditionnelle ou intuitive. Le tableau 6 contient aussi des informations sur les volumes de ces textes (nombre de mots).

Les volumes des textes figurant dans le tableau 13 varient : le texte le plus court est de 76 mots (T43), le texte le plus long est de 3270 mots (T88).

Codes	Titres, dates	Volume (nombre de mots)
T01	De la comédie. 1838.07.03	1876
T02	Du sacre de la reine Victoria. 1838.07.05	935
T03	Une lettre écrite au Morning-Chronicle. 1838.07.06	1058
T04	Quels que soient les bruits contradictoires. 1838.07.09	999
T05	Au moment où la question belge. 1838.07.11	551
T06	L'enthousiasme que le maréchal Soult inspire. 1838.07.14	929
T07	S'il fallait en croire La Gazette d'Augsbourg. 1838.07.16	566
T08	Théâtre français. Première représentation de <i>Philippe III</i> , tragédie en 5 actes, de M. Andraud. 1838.07.16	1444
T09	<i>La Quotidienne</i> nous répond ce matin. 1838.07.17	310
T10	Banquet de Guildhall. 1838.07.18	193
T11	<i>La Quotidienne</i> publie ce matin un article. 1838.07.19	977
T12	Jusqu'ici on avait pu remarquer. 1838.07.22	388
T13	Le Docteur Francia. 1838.07.22	1849
T14	<i>L'Europe</i> prétend ce matin. 1838.07.24	178
T15	En rendant compte des actes les plus insignifiants. 1838.07.25	222
T16	<i>L'Europe</i> nous apprend ce matin. 1838.07.25	388
T17	Ce que nous disions, il y a quelques jours. 1838.07.26	403
T18	Lorsque <i>La Quotidienne</i> demande. 1838.07.28	521
T19	Du droit des gens en Europe. 1838.07.29	1038
T20	Au moment où le ridicule projet. 1838.07.30-31	459
T21	Plusieurs journaux ont reproché à lord Palmerston. 1838.08.03	366
T22	Nous avons déjà fait remarquer. 1838.08.04	821

T23	Affaires d'Orient. Deuxième article. 1838.08.05	1593
T24	Des troubles ont éclaté à Stockholm. 1838.08.10	567
T25	Chronique. 1838.08.11	90
T26	On a beau dire que la question hollando-belge. 1838.08.11	1024
T27	Un journal ministériel reproduit ce matin.1838.08.12	186
T28	Ceux qui savent que tout signifie quelque chose. 1838.08.13	311
T29	Théâtre de la porte St-Martin. 1838.08.13	1649
T30	<i>La Quotidienne</i> annonce en triomphe. 1838.08.14	217
T31	Tœplitz continue toujours d'alimenter les organes d'un parti. 1838.08.15	1186
T32	Théâtre-français. <i>Le Ménestrel</i> , comédie en cinq actes et en vers, par M. Camille Bernay. 1838.08.18	2807
T33	De la Suède, et de sa situation dans le nouvel équilibre de l'Europe (Premier article). 1838.08.19	1022
T34	Le discours de la reine d'Angleterre. 1838.08.20	231
T35	De la Suède, et de sa situation dans le nou- vel équilibre de l'Europe (Deuxième article). 1838.08.21	800
T36	Au milieu de tous les commérages légitimistes. 1838.08.21	370
T37	Paris, 21 août. 1838.08.22	1179
T38	Après avoir gardé, depuis quelques jours, un silence. 1838.08.25	584
T39	La journée d'hier a été heureuse pour la France. 1838.08.26	423
T40	Les paroles prononcées à la tribune anglaise. 1838.08.27	892
T41	Affaire du Mexique.1838.08.28	866

T42	Les conséquences de la faute. 1838.08.29	867
T43	<i>La Quotidienne</i> , avec ce sentiment de dépit inquiet. 1838.08.30	76
T44	<i>Le Moniteur</i> a publié hier trois ordonnances du roi. 1838.08.31	608
T45	Hier, nous avons répondu au manifeste de la police. 1838.10.01	439
T46	L'allocution que le Saint-Père vient d'adresser. 1838.10.01	417
T47	<i>Le Journal de Francfort</i> contient aujourd'hui. 1838.10.02	333
T48	En annonçant que la diète d'Allemagne. 1838.10.02	231
T49	Les Mexicains viennent de nous donner un échantillon. 1838.10.03	449
T50	Nous avons signalé, il y a deux jours. 1838.10.04	76
T51	De nos alliances. 1838.10.05	949
T52	La lettre, que publie un journal ce matin. 1838.10.06	946
T53	Le gouvernement russe vient de nous faire. 1838.10.07	407
T54	<i>L'Europe</i> , aujourd'hui, passe toutes bornes. 1838.10.07	212
T55	La nouvelle suivante, que la <i>Revue de Paris</i> donne aujourd'hui. 1838.10.08	182
T56	S'il fut jamais un thème fertile. 1838.10.09	742
T57	Louis-Napoléon a quitté la Suisse. 1838.10.10	637
T58	La réponse du directoire de Lucerne. 1838.10.11	650
T59	Le journal qui avait annoncé la déclaration d'indépendance. 1838.10.11	214
T60	Nous pensions que le dernier mot avait été prononcé. 1838.10.12	561
T61	Théâtre-Français. <i>Richard Savage</i> , drame en 4 actes. – Par MM. Desnoyer et Labat. 1838.10.15	1617

T62	Théâtre-Français. M ^{lle} Rachel. 1838.10.29	2027
T63	Théâtre-Français. <i>Maria Padilla</i> , par M. Ancelot. 1838.11.01	2187
T64	Un fait qui parle plus haut. 1838.11.08	935
T65	Affaire de Cologne. 1838.11.12	1114
T66	La <i>Revue de Paris</i> nous accuse aujourd'hui. 1838.11.12	148
T67	Le discours du roi des Belges. 1838.11.18	827
T68	Le Roi gouverne. 1838.11.24	1287
T69	La grande objection du ministère contre la coalition. 1838.11.29	943
T70	Nous avons réfuté hier l'unique objection du ministère. 1838.11.30	934
T71	Il y a parfois des révélations. 1839.01.04	1136
T72	Calme comme la force et comme la vérité. 1839.01.06	772
T73	La profonde impression produite hier. 1839.01.13	630
T74	Au milieu des débats de l'adresse. 1839.01.23	474
T75	Théâtre de l'Opéra. <i>La Gypsy</i> , par MM. Saint-Georges et Mazillier. 1839.02.05	1813
T76	Théâtre-Français. <i>Le Comité de bienfaisance</i> , par MM. Duveyrier et Delorme. 1839.02.05	258
T77	Théâtre-Français. <i>Les Serments</i> , par M. Viennet. 1839.02.25	859
T78	Théâtre de la Renaissance. <i>Diane de Chivry</i> , par M. Soulié. 1839.02.25	373
T79	Théâtre-Français. M ^{lle} de Belle-Isle, par Alexandre Dumas. 1839.04.12	2139
T80	Encyclopédie du XIX ^e siècle. 1839.04.27	2389
T81	Théâtre-Français. <i>Il faut que jeunesse se passe</i> , par M. de Rougemont. 1839.07.06	698
T82	Jusqu'ici on avait pu remarquer. 1838.07.22	388
T83	Les journaux de toutes les opinions. 1838.09.01	725
T84	Candidature de M. Charles Levavasseur. 1842.10.19	222

T85	Dieppe. Élections. Candidature de M. Charles Levavasseur. 1842.10.22	1554
T86	Candidature de M. Charles Levavasseur. 1842.10.26	1207
T87	Lettres de Junius. (Casaque écarlate) II Junius à monsieur de Villemessant. 1861.11.07	1508
T88	Septième lettre. 1861.12.12	3270

TABLEAU 13 – Textes à attribuer.

En dehors des textes à attribuer, les chercheurs français ont fourni des textes permettant de constituer les classes *a priori*, parmi lesquels 9 documents appartiennent à Barbey d'Aurevilly (tableau 14) et 12 textes, aux autres auteurs, dont 8 à Charles Monselet (tableau 15), 3 à Théophile Gautier (tableau 16) et 2 à Jules Janin (tableau 17).

Code	Auteur, titre, références	Nombre de mots
JB01	M. de M... De la démocratie nouvelle ou des mœurs et de la puissance des classes moyennes en France, par Édouard Alletz // Revue française et étrangère ou nouvelle revue encyclopédique, [deuxième semestre] 1837, p. 258-277.	7104
JB02	J. B. d'A. <i>Histoire de la papauté pendant les XVI^e et XVII^e siècles</i> , par M. Léopold Ranke, professeur à l'université de Berlin, traduite de l'allemand par M. Haiber, publiée et précédée d'une introduction par M. Alexandre de Saint-Chéron // <i>Journal général de l'instruction publique et des cours scientifiques et littéraires</i> , n° 85, 12 mai 1838, p. 519-520 (dans la rubrique « Histoire », dans la partie non officielle).	3591
JB03	Jules Barbey d'Aurevilly. <i>Histoire de la papauté pendant les XVI^e et XVII^e siècles</i> , par Léopold Ranke, professeur à Berlin, traduite par M. Haiber et précédée d'une préface par M. de Saint-Chéron (2 ^e article) // <i>Journal général de l'instruction publique et des cours scientifiques et littéraires</i> , 15 septembre 1838 (dans la rubrique « Histoire », dans la partie non officielle), p. 817-819.	2548

JB04	J.-B. d'A... Théâtre de la Renaissance. <i>Ruy Blas</i> et sa préface // <i>Le Nouvelliste, journal politique du soir et du matin</i> , 3 décembre 1838, p. 1-3.	2488
JB05	J. B. d'A. Encyclopédie du XIX ^e siècle // Journal général de l'Instruction publique et des cours scientifiques et littéraires, n° 33, 24 avril 1839, p. 234-235 (dans la rubrique « Littérature », dans la partie non officielle); <i>Le Nouvelliste</i> , propagateur des doctrines constitutionnelles, 27 avril 1839, p. 1-3.	2370
JB06	J. B. d'A. Théâtre-Français. <i>Il faut que jeunesse se passe</i> , par M. de Rougemont // <i>Le Nouvelliste, propagateur des doctrines constitutionnelles</i> , 6 juillet 1839, p. 1. + De l'Angleterre et de lord Palmerston // <i>La Mode</i> , 6 juillet 1850, p. 3-6.	688 + 1953
JB07	J. Barbey d'Aureville. <i>Un Machiavel inédit. L'esprit d'un homme d'État</i> , par Nicolas Machiavel, traduit par M. Joseph Delaroa // <i>Le Pays</i> , 14 novembre 1861, p. 3 (dans la Bibliographie).	2853
JB08	Les chroniqueurs // <i>Le Nain jaune</i> , 24 février 1866 (dans les « Ridicules du temps »), p. 2-3.	2750
JB09	Les chats de la critique // <i>Le Nain jaune</i> , 19 mai 1866 (dans les « Ridicules du temps »), p. 2-3.	2222

TABLEAU 14 – Composition et puissance de la classe *a priori* Ω
(Barbey d'Aureville – abr. JB).

Code	Auteur, titre, références	Nombre de mots
CM01	Charles Monselet. <i>Il Pleut, Il Pleut</i> , [n/a]	557
CM02	Charles Monselet, <i>Histoire anecdotique du tribunal révolutionnaire</i> (17 août-29 novembre 1792), D. Giraud et J. Dagneau (1853)	2488
CM_LES AVEUX 1-6	Charles Monselet, <i>Les Aveux</i> (6 fragments), Librairie de la Société des gens de Lettres, 1854	600 dans chaque fragment

TABLEAU 15 – Composition et puissance de la classe *a priori* Ω
(Charles Monselet – abr. CM).

Code	Auteur, titre, références	Nombre de mots
TG01	Théophile Gautier. Chronique dramatique 11 juillet // <i>La Presse</i> , 11 juillet 1837	2114
TG02	Théophile Gautier. Chronique dramatique 24 juillet // <i>La Presse</i> , 24 juillet 1837.	1475
TG03	Théophile Gautier. Chronique dramatique 31 juillet // <i>La Presse</i> , 31 juillet 1837.	1936

TABLEAU 16 – Composition et puissance de la classe *a priori* Ω
(Théophile Gautier – abr. TG).

Codes	Auteur, titre, références	Nombre de mots
JJ01, JJ02	Jules Janin. « Madame Putiphar » (2 fragments) // <i>Journal des débats politiques et littéraires</i> , 3 juin 1839, p. 1-3 (dans le Feuilleton du <i>Journal des débats</i>).	3309 + 2691

TABLEAU 17 – Composition et puissance de la classe *a priori* Ω (Jules Janin – abr. JJ).

Comme le démontrent les tableaux 14-17, 5 textes de Barbey d'Aurevilly sont datés des années 1838-1839 et les autres, des années 1850-1866. Tous les textes de Janin et Gautier sont aussi datés des années 1838-1839 et ceux de Charles Monselet, de l'année 1853.

Les textes trop courts du corpus (comme le document JB06, par exemple) ont été réunis, alors que les textes longs ont été divisés en deux ou plusieurs fragments (JJ01, CM_LES AVEUX). Cette division existait déjà dans le corpus fourni, elle s'explique par le fait que le volume égal des textes permet un plus grand nombre d'expériences avec l'échantillonnage et le choix du volume du dictionnaire des fréquences, mais aussi, selon l'avis de nombreux chercheurs, qu'il assure l'exactitude de l'attribution.

Compte tenu de la composition des classes *a priori*, l'hypothèse nulle et l'hypothèse alternative du projet peuvent être formulées de la manière suivante :

$$\begin{aligned}
 H_0 &: \text{tous les textes attribués appartiennent} \\
 &\text{à J. Barbey d'Aurevilly,} \\
 H_1^a &: \text{aucun des textes attribués n'appartient} \\
 &\text{à J. Barbey d'Aurevilly,} \\
 H_1^b &: \text{tous les textes attribués n'appartiennent pas} \\
 &\text{à J. Barbey d'Aurevilly} \\
 H_1^c &: \text{certains d'entre eux appartiennent à Ch. Monselet,} \\
 &\text{à Th. Gautier, à J. Janin ou à d'autres auteurs}
 \end{aligned}$$

MÉTHODES ET PROCÉDURES

Compte tenu du grand nombre de textes soumis à l'attribution, une méthode fondée sur l'application du progiciel d'analyse stylométrique Stylo écrit en langage de programmation R²⁹ a été utilisée dans la deuxième phase de la recherche. La détermination de l'appartenance des textes étudiés s'effectue en mesurant les différences entre les textes et en divisant ensuite le corpus des textes en catégories correspondant aux auteurs présumés.

La mesure Delta proposée par John Burrows³⁰ a été choisie comme mesure de la proximité et des différences entre les styles des auteurs : elle a la réputation d'un outil fiable pour résoudre les problèmes d'attribution. La mesure Delta a été testée sur des textes prosaïques et poétiques de différents genres, écrits dans les langues indo-européennes les plus répandues :

- étude des romans anglais du début du xx^e siècle³¹ ;
- étude de la poésie anglaise contemporaine³² ;
- tests sur des romans, poèmes et traductions en anglais, des romans polonais et français, la littérature allemande, des romans hongrois et italiens, ainsi que sur la prose et la poésie latine³³ ;
- étude des romans anglais, français et allemands des XIX^e et XX^e siècles³⁴ ;

29 Voir Maciej Eder, Jan Rybicki, Mike, « Stylometry with R : A package for computational text analysis », *The R Journal*, vol. 8 (1), p. 107-121.

30 Voir John Burrows, « Delta : A Measure of Stylistic Difference and a Guide to Likely Authorship », *Literary and Linguistic Computing*, Vol. 17 (3), 2002, p. 267-287.

31 *Ibid.*

32 Voir Maciej Eder, Jan Rybicki, Mike Kestemont, « Stylometry with R : A package for computational text analysis », *The R Journal*, Vol. 8 (1), p. 107-121.

33 Voir Jan Rybicki, Maciej Eder, « Deeper Delta across genres and languages. Do we really need the most frequent words? », *Literary and Linguistic Computing*, 2011, Vol. 26 (3), p. 315-321.

34 Voir Stefan Evert, Thomas Proisl, Fotis Jannidis, Steffen Pielström and others, « Towards a better understanding of Burrows's Delta in literary authorship attribution », *Paper presented at NAACL HLT Fourth Workshop on Computational Linguistics for Literature*, Denver, Colorado, 2015, p. 79-88 ; Fotis Jannidis, Steffen Pielström, Christof Schöch, Thorsten Vitt, « Improving Burrows' Delta. An empirical evaluation of text distance measures », Paper presented at Conference DH 2015 – Global Digital Humanities. Sydney, Australia, URL : <https://www.

- attribution des discours des présidents américains, ainsi que des articles parus dans le journal *Le Fédéraliste*³⁵.

La justification mathématique du Delta est la suivante : soit un ensemble de n mots et une paire de textes D, D' , sur lesquels la mesure Delta sera calculée. Désignons cet ensemble de mots par $\{w_i\}$, en définissant $f_i(D)$ comme la fréquence du mot w_i dans le texte D , μ_i comme la fréquence moyenne du mot dans tout l'échantillon et σ_i comme l'écart-type de cette fréquence. Ensuite, l'estimation standard (estimation z) de la fréquence d'utilisation du mot w_i dans le texte D est calculée par :

$$z(f_i(D)) = \frac{f_i(D) - \mu_i}{\sigma_i}. \quad (6)$$

Ainsi, on a l'expression mathématique suivante, appelée la mesure Delta et correspondant à la moyenne des valeurs absolues des différences d'estimations standard (estimations z) des fréquences de mots de $\{w_i\}$ entre les textes D et D' :

$$(D, D') = \frac{1}{n} \sum_{i=1}^n |z(f_i(D)) - z(f_i(D'))|. \quad (7)$$

Cette formule peut être convertie comme suit :

$$\begin{aligned} \Delta(D, D') &= \frac{1}{n} \sum_{i=1}^n |z(f_i(D)) - z(f_i(D'))| = \frac{1}{n} \sum_{i=1}^n \left| \frac{f_i(D) - \mu_i}{\sigma_i} - \frac{f_i(D') - \mu_i}{\sigma_i} \right| = \\ &= \frac{1}{n} \sum_{i=1}^n \left| \frac{f_i(D) - f_i(D')}{\sigma_i} \right|. \end{aligned} \quad (8)$$

La transformée montre que le Delta est indépendant des fréquences moyennes des mots μ_i dans l'échantillon et peut être considéré comme une mesure normalisée de la différence entre les fréquences

researchgate.net/publication/280086768_Improving_Burrows%27_Delta_-_An_empirical_evaluation_of_text_distance_measures> (consulté le 21 janvier 2022).

35 Voir F. Mosteller et D.L. Wallace, 1963, « Inference in an authorship problem », *Journal of the American Statistical Association*, 58 (302), 1963, p. 275-309.

de chacun des mots dans les textes D et D' . Puisque le calcul de la moyenne implique de diviser la somme par une constante égale au nombre de mots n considérés, cette opération peut être omise lors de la comparaison des résultats calculés. Il est donc raisonnable de simplifier la formule (8) à :

$$\Delta^{(n)}(D, D') \doteq \sum_{i=1}^n \frac{1}{\sigma_i} |f_i(D) - f_i(D')|, \tag{9}$$

c'est-à-dire que la mesure Delta par rapport à la paire de textes (D, D') est prise comme égale à la somme sur l'ensemble des mots w_i des valeurs absolues des différences de fréquence des mots entre les textes D et D' divisée par l'écart-type σ_i .

Ainsi, en utilisant Delta à des fins d'attribution, on essaie de comparer les candidats à l'attribution du texte D' en estimant la distance multidimensionnelle au texte D , chaque dimension (fréquence d'utilisation du mot) étant mise à l'échelle par le facteur $\frac{1}{\sigma_i}$ (ainsi, de petites déviations dans la fréquence du mot peuvent affecter le résultat si la fréquence de ce mot dans l'échantillon varie peu).

L'une des modifications du Delta est le Delta d'Eder³⁶, qui suppose un poids accru de mots fréquents et est recommandé pour l'étude de textes en langues synthétiques :

$$\Delta_E^{(n)}(D, D') = \frac{1}{n} \sum_{i=1}^n \left(\frac{|f_i(D) - f_i(D')|}{\sigma_i} \cdot \frac{n-i+1}{n} \right). \tag{10}$$

Une autre version de Delta est le Delta cosinus :

$$\Delta_C(D, D') = 1 - \frac{\vec{f}(D) \cdot \vec{f}(D')}{\vec{f}(D)_2 \cdot \vec{f}(D')_2} \tag{11}$$

où

36 Voir Maciej Eder, Jan Rybicki, Mike Kestemont, « Stylometry with R : A package for computational text analysis », art. cité.

$$\vec{f}(D) = (f_1(D), f_2(D), \dots, f_n(D)), \vec{f}(D') = (f_1(D'), f_2(D'), \dots, f_n(D')).$$

Avec le Delta, les algorithmes de clustering sont souvent utilisés pour obtenir le résultat sous la forme d'un dendrogramme. L'algorithme implique le choix à chaque étape de l'association de clusters qui entraînera une augmentation minimale de la fonction cible. Généralement, le rôle de la fonction cible est joué par la somme des carrés d'erreur pour chaque cluster, ou par une fonction qui la résume en utilisant une distance différente, auquel cas la mesure Delta est utilisée. Supposons que la première étape concerne l'unification des clusters A_i и A_j , $i \neq j$, $i, j \in \{0, \dots, n\}$. Pour déterminer à la fin de l'étape si une telle union est souhaitable, la fonction cible est calculée :

$$z = \sum_{k \in A_i \cup A_j} \Delta^2 \left(k, (A_i \cup A_j)^* \right), \quad (12)$$

où $(A_i \cup A_j)^*$ est le centroïde du cluster nouvellement formé. Ensuite, toutes sortes d'autres associations de clusters appariés sont considérées de la même manière. À la fin de l'étape, les valeurs de la fonction cible sont comparées ; l'union à laquelle correspond la plus petite valeur parmi celles calculées dans cette étape est sélectionnée. L'algorithme d'analyse de cluster décrit permet de construire un dendrogramme en conséquence.

Les dendrogrammes visualisent le fonctionnement de l'algorithme de clustering et se lisent de droite à gauche. Chaque désignation apparaît pour la première fois au moment où le texte correspondant est inclus dans un cluster « étranger ». L'algorithme de clustering vise à relier les objets les plus similaires en affaiblissant progressivement les exigences du degré de similarité ; ainsi, les clusters formés au début correspondent aux liens les plus étroits entre les objets. L'échelle en bas de l'image indique les valeurs de la fonction cible (voir figure 1).

Lors de l'utilisation du logiciel indiqué, l'analyse ne doit s'étendre que sur les fragments du texte créés par l'auteur lui-même. À cet effet, les colonnes-titres, les numéros de pages ainsi que les citations et les insertions d'autres auteurs devraient être supprimés. Ces dernières peuvent créer des difficultés surtout si elles ne sont pas nettement signalées.

Dans le cadre du traitement du matériel textuel, on a enlevé les noms propres, les citations, les artefacts apparents de la reconnaissance de texte,

les signatures et les numéros de pages, on a remplacé les caractères illisibles dans le codage choisi et converti tous les fichiers dans la norme de codage correcte (en l'occurrence, UTF-8), ainsi que leurs titres au format requis pour Stylo. L'étape du prétraitement comprend aussi le dépouillement (culling) du texte qui consiste en l'élimination d'un certain nombre de tokens uniques pour le texte donné. Cet aspect peut se révéler important si dans les critiques, les titres des pièces de théâtre, les noms des auteurs et des personnages se trouvent parmi les mots les plus fréquents. Pourtant, le nombre de ces mots parmi les noms communs n'est pas grand et du fait du faible volume des documents, le dépouillement appauvrit encore le dictionnaire des mots fréquents. En l'utilisant lors du test, on a pu constater la baisse de similarité des clusters formés. Par conséquent, à l'étape principale de l'attribution, la fonction de dépouillement n'a pas été mise en œuvre. La tokenisation et l'élaboration du dictionnaire de fréquences de mots auxquelles on procède habituellement à l'étape du prétraitement, sont effectuées par le logiciel.

On a retenu comme optimaux les paramètres suivants : absence du dépouillement et de l'échantillonnage, dictionnaire de fréquences de 200 mots, Delta en cosinus en tant que métrique pour estimer la proximité.

Au-delà du contrôle de la stabilité, de l'uniformité et de la représentativité des résultats de l'analyse en fonction de métriques différentes et de différents volumes du dictionnaire de fréquences de mots, avant de procéder à l'étape principale de la recherche, il a fallu tester le système avec les paramètres retenus sur les classes *a priori*, afin de pouvoir comparer les résultats de l'analyse avec les données disponibles.

Pour effectuer ce test, on a choisi une approche reposant sur la méthode de validation croisée qui assure l'utilisation la plus équilibrée des données disponibles pour estimer la fiabilité du fonctionnement du système avec les paramètres déterminés.

La méthode de validation croisée permet de vérifier sur un ensemble disponible de données la capacité d'un algorithme à traiter correctement de nouvelles données. Elle est surtout pertinente pour valider les systèmes destinés à la prédiction, avant qu'ils soient utilisés en pratique. Un cycle de validation croisée suppose la division de l'ensemble de données en parties, dont une partie appelée ensemble d'apprentissage est utilisée pour construire ou expérimenter le modèle, et l'autre appelée ensemble de test est utilisée pour valider le modèle.

Dans notre cas, les données disponibles ont déjà été réparties en classes *a priori*, il fallait donc vérifier l'efficacité de l'algorithme pour chaque classe *a priori*. Pour ce faire, tous les textes disponibles du même auteur ont été divisés en deux parties plus ou moins égales : l'une était utilisée comme ensemble test et l'autre, associée aux textes d'un autre auteur, a formé l'ensemble d'apprentissage (en fait, un nouveau corpus *a priori* de deux auteurs). Après quoi, on a mis en œuvre le système pour une attribution déterministe des textes composant l'ensemble de test.

La validation a été effectuée pour chaque classe *a priori* en combinaison avec la classe de Barbey d'Aurevilly, puis entre les autres auteurs, Barbey excepté, pris deux par deux. Une telle approche est déterminée par la tâche centrale de l'analyse qui est de séparer les textes écrits par Barbey de tous les autres textes, tandis que la tâche de différenciation des autres auteurs entre eux reste secondaire. C'est la validation deux par deux qui permet d'estimer la capacité du système à identifier correctement les textes de Barbey et à décider s'il peut être efficacement utilisé pour l'attribution d'ouvrages. Le système vise alors à ce qu'aucun texte de l'ensemble de test ne soit attribué par erreur au deuxième auteur de la paire. Le rapport entre le nombre de décisions erronées et le volume total de l'ensemble de test du corpus est établi pour estimer l'efficacité du fonctionnement de l'algorithme avec les paramètres donnés.

Selon les résultats du test, l'appartenance des textes de la classe *a priori* a été confirmée pour toutes les paires d'auteurs avec une précision de 100 % pour les paramètres choisis comme optimaux (dictionnaire des 200 mots les plus fréquents, Delta en cosinus), et avec une précision moyenne d'environ 75 % pour les paramètres moins rigoureux (100 mots de fréquence avec des métriques différentes). Notons bien qu'une telle précision a été constatée non seulement pour les paires « Barbey d'Aurevilly – un autre auteur », mais aussi entre les autres auteurs. Même avec des volumes faibles de certaines classes *a priori*, le résultat obtenu peut être considéré comme probant et suffisant pour utiliser le système dans le cadre de l'attribution des textes.

Pour améliorer la fiabilité de l'attribution, on a utilisé la méthode itérative analogue à la méthode des centres mobiles utilisée en classification automatique³⁷. Chaque itération comprend deux procédures

37 Voir Jean-Marie Bouroche, Gilbert Saporta, *L'Analyse des données*, quatrième édition, Paris, Presses universitaires de France, 1980, p. 51-52.

du logiciel Stylo : le clustering (l'analyse stylométrique anonyme qui consiste à réunir les textes *a priori* et les textes à attribuer, à analyser leurs similitudes stylistiques et à créer les corpus des textes les plus proches) et la classification (l'attribution proprement dite au cours de laquelle on attribue chaque texte traité à une des classes *a priori* dont le style est le plus proche ou bien, si l'auteur est déjà connu, dont la paternité est confirmée ou contestée sur la base d'une similitude plus ou moins grande entre les textes traités et la classe *a priori*). Le clustering permet de révéler le degré de similarité stylistique des textes traités et des textes *a priori* et l'appartenance des textes traités à un autre cluster éloigné des classes *a priori*. La classification détermine et garde les frontières du cluster dans lequel sont réunis les textes relativement proches les uns des autres et les éléments de la classe *a priori* en les attribuant au même auteur : si un texte traité se trouve à la frontière de deux clusters, il n'est pas toujours évident de déterminer de quel auteur il est le plus proche et c'est alors ce que la classification permet de révéler. Ainsi, la structure de la méthode itérative mixte peut être la suivante : à chaque itération on procède d'abord au clustering selon les résultats duquel les textes qui se sont trouvés dans les mêmes clusters que les éléments des classes *a priori*, peuvent être potentiellement ajoutés à ces classes et on ajoute les noms des auteurs auxquels les classes se rapportent à leurs métadonnées.

On passe ensuite à la classification dans le but de l'affirmation de la paternité de ces textes : si la fidélité de l'attribution est confirmée, on les ajoute aux classes *a priori* et on effectue l'itération suivante.

Lorsque la composition des classes devient stable et que l'ajout à une classe concrète de textes stylistiquement proches ne peut être confirmé à l'étape 2, c'est que l'étape principale de cette procédure est finie. Les éléments qui n'ont pas été attribués peuvent être soumis à la procédure standard de la classification, compte tenu de la mise à jour de toutes les classes.

Cette méthode itérative mixte a été appliquée au corpus des textes à attribuer (tableau 1). 18 textes parmi 88 (soit 20,4 % du nombre total de textes à attribuer) ont été attribués à la classe *a priori* Ω (Barbey d'Aureville), 4 textes à la classe Ω (Charles Monselet) et 1 texte à la classe Ω (Théophile Gautier).

Par ailleurs, la première itération a déjà révélé que le taux de similarité entre deux documents textuels dépasse toutes les différences possibles pour les textes qui se rapprochent stylistiquement et que leur Delta ou bien est égal à zéro, ou bien tend vers zéro. Il s'est avéré que le corpus des textes à attribuer contenait deux doubles (les textes dont les codes sont T12 et T82). Après l'exclusion d'un de ces doubles, le nombre total de textes s'est ramené à 87 et le pourcentage des textes de la classe Ω (Barbey d'Aurevilly) est passé à 20,7 % (tableau 18).

CONCLUSION

Codes du document	Résultats de l'attribution
T88	Ω (J. Barbey d'Aurevilly)
T87	Ω (Ch. Monselet)
T86	Ω (J. Barbey d'Aurevilly)
T81	Ω (J. Barbey d'Aurevilly)
T80	Ω (J. Barbey d'Aurevilly) double
T79	Ω (J. Barbey d'Aurevilly)
T78	Ω (J. Barbey d'Aurevilly)
T77	Ω (J. Barbey d'Aurevilly)
T75	Ω (J. Barbey d'Aurevilly)
T74	Ω (J. Barbey d'Aurevilly)
T72	Ω (J. Barbey d'Aurevilly)
T69	Ω (J. Barbey d'Aurevilly)
T65	Ω (Ch. Monselet)
T63	Ω (J. Barbey d'Aurevilly)
T61	Ω (J. Barbey d'Aurevilly)
T59	Ω (J. Barbey d'Aurevilly)
T48	Ω (Ch. Monselet)

T32	Ω (J. Barbey d'Aurevilly)
T29	Ω (J. Barbey d'Aurevilly)
T27	Ω (Ch. Monselet)
T8	Ω (J. Barbey d'Aurevilly)
T7	Ω (Th. Gautier)
T1	Ω (J. Barbey d'Aurevilly)

TABLEAU 18 – Résultats de l'attribution.

La similarité des autres textes ne s'est pas avérée statistiquement significative au cours des itérations : la divergence des valeurs de leur delta était quelquefois plus importante que pour les objets appartenant à des classes certaines. Cela signifie que dans la liste des auteurs potentiels, il manque un ou plusieurs auteurs dont les classes *a priori* seraient beaucoup plus proches de ces textes d'après la valeur de la métrique choisie. Aussi est-il nécessaire de reformuler l'hypothèse d'attribution et d'inclure une ou plusieurs nouvelles classes *a priori*. Les clusters reçus sont représentés à l'aide du dendrogramme (figure 1).

Finalement, l'hypothèse nulle de l'appartenance de tous les textes à Barbey d'Aureville a été invalidée ; l'hypothèse alternative H_a^2 a été confirmée : il n'y a qu'une partie des textes traités qui appartiennent à Barbey d'Aureville.

Ces textes se sont retrouvés sous des titres proches dans le corpus analysé (11 parmi 18 textes se trouvent dans les vingt derniers textes dans l'ordre chronologique). Leurs volumes varient (de 211 à 2700 mots) tandis que le volume moyen de ces textes qui dépasse le volume moyen du corpus de 73 %, est de 1422 mots.

Pierre GLAUDES,
Aurélia CERVONI,
Francesca GUGLIELMI
Sorbonne Université

Catherine MAYAUX
Université de Cergy-Pontoise

Mikhail MARUSENKO,
Yekaterina KURALESINA,
Mariya MIRETINA,
Yekaterina NIKITINA,
Mariya SOLOVYEVA,
Olga KHUTORETSKAYA
Université d'État
de Saint-Pétersbourg, Russie