ULB UNIVERSITÉ LIBRE DE BRUXELLES

Après-midi inédit du 20 février 2024

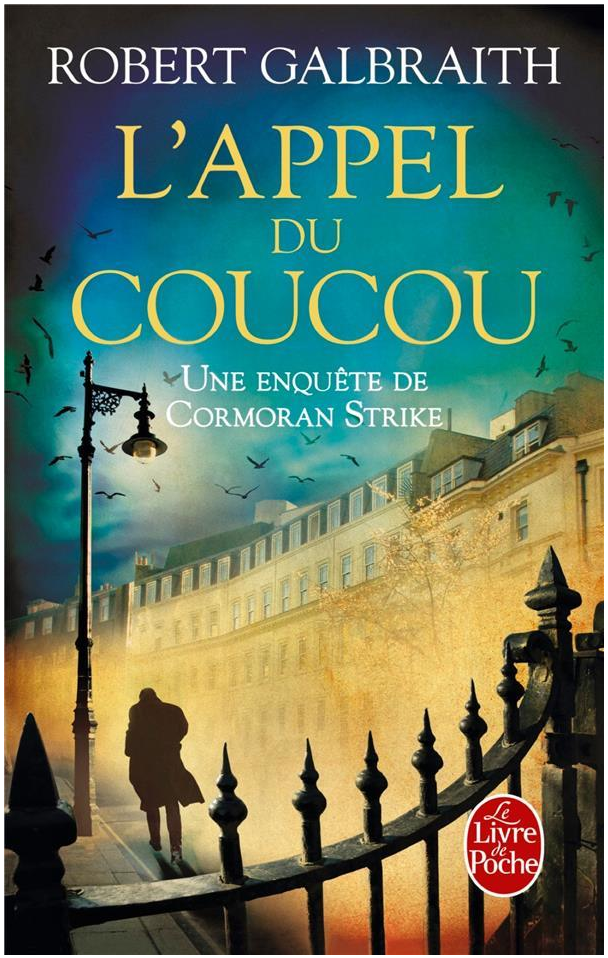
La stylométrie : quand les statistiques et l'intelligence artificielle permettent d'identifier l'auteur d'une œuvre littéraire

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

Sherlock Holmes

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

De la fiction ?



1. Définir la stylométrie

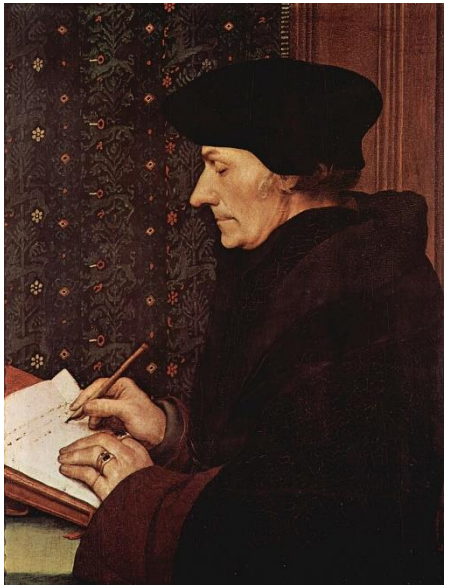
- La stylométrie est une « technique permettant de saisir le caractère souvent insaisissable du style d'un auteur, ou du moins d'une partie de ce style, en **quantifiant** certaines de ses caractéristiques » (Laan 1995, p. 271)
- Elle s'insère dans la discipline des études de paternité, dont elle se distingue par cet aspect quantificateur

1. Définir la stylométrie

- Une hypothèse centrale : chaque auteur est caractérisé par un style qui lui est personnel et qui est vérifiable : ce qu'on pourrait appeler une véritable « empreinte digitale » (Love 2002, p. 12)
  - Ce style relève de l'inconscient et ne peut être masqué ou imité
  - La littérature scientifique le nomme également « stylome »
- C'est ce « stylome » que la stylométrie cherche à quantifier

2. Un historique de la stylométrie

- Depuis l'Antiquité, les érudits s'intéressent aux études de paternité
- La stylométrie à proprement parler remonte au XIX^{e} s., mais on peut voir en Érasme (XV^{e} – XVI^{e} s.) un précurseur
- Il avance l'hypothèse que Quintilien, auteur latin du I^{er} s. ap. J.-C., utilise plutôt l'adverbe *interim* alors que les autres auteurs préfèrent *interdum* (trad. « parfois », « quelques fois »)
- Toutefois, on ne peut pas parler de stylométrie comme il n'y a pas de comptage systématique



Hans Holbein le Jeune, *Portrait d'Érasme écrivant*, 1528

2. Un historique de la stylométrie

- Pour la première trace concrète de stylométrie, il faut se tourner vers le XIX^{e} s. et Auguste de Morgan (mathématicien anglais)
- Dans une lettre de 1851, il conçoit une méthode pour étudier la paternité des épîtres attribuées à saint Paul

Photographie d'Auguste van Morgan
S. E. de Morgan, *Memoir of Augustus De Morgan*, 1882

La méthode d'A. de Morgan

Je voudrais que vous fassiez ceci : passez votre regard sur n'importe quelle partie des épîtres de saint Paul qui commence par *Paulos* – en grec, je veux dire – et sans prêter attention au sens. Faites ensuite la même chose avec l'épître aux Hébreux, et essayez d'équilibrer dans votre esprit la question de savoir si cette dernière n'utilise pas des mots plus longs que la première. J'ai toujours pensé que les questions d'attribution pourraient être réglées à peu de frais de cette manière. [...]

Comptez un grand nombre de mots dans Hérodote – disons tout le premier livre – et comptez toutes les lettres ; divisez le deuxième nombre par le premier, ce qui donne le nombre moyen de lettres pour un mot dans ce livre. Faites la même chose avec le deuxième livre. Je m'attends à ce que ces deux moyennes soient très proche, [par exemple 5,624 et 5,619]. Je ne m'étonnerais pas si le même comptage appliqué à deux livres de Thucydide donne [les moyennes 5,713 et 5,728]. C'est à dire, je m'attends à observer de légères différences entre des auteurs différents, et des différences infimes [entre les textes d'un même auteur].

Si ce fait-là était établi, et si les épîtres de saint Paul qui commencent par *Paulos* donnaient 5,428 et les les Hébreux donnaient 5,516, par exemple, je serais convaincu que le texte grec des *Hébreux* [...] n'est pas de la plume de Paul.

2. Un historique de la stylométrie

- Dans la deuxième moitié du XIX^{e} s., plusieurs tentatives pour classer les œuvres de Platon par ordre chronologique en comptant :
  - l'utilisation de mots rares (Lewis Campbell)
    → aboutit à un rapprochement thématique
  - l'utilisation de prépositions (Charles Baron)
    → aboutit à un rapprochement stylistique

2. Un historique de la stylométrie

Première période.

Charmide Hippias Major Criton	absence complète de πέρι.
Protagoras Alcibiade I Apologie	πέρι employé avec les pronoms seulement.
Ion Cratyle Phédon Euthyphron (Lachès)	πέρι, tournure encore exceptionnelle.

2. Un historique de la stylométrie

Deuxième période.

Banquet
Gorgias
Menon
Euthydème
} πέρι dans la proportion de 1/8 à 1/6.

Troisième période.

(Théétète)
République
Phèdre
} πέρι dans la proportion de 1/5 à 1/3.

Quatrième et dernière période.

(Timée)
(Critias)
Sophiste
Politique
Philèbe
Lois
} πέρι dans la proportion de 1/3 à 1/2 et au-delà.

2. Un historique de la stylométrie

- Dès les années 1960, le développement de l'ordinateur permet d'automatiser des calculs qui devaient être réalisés jusqu'alors à la main
- En 1963, les statisticiens Frederick Mosteller et David Wallace publient un article considéré comme fondateur
- C'est un article avec une portée avant tout statistique, et qui convainct un large public

2. Un historique de la stylométrie

- Celui-ci vise à identifier les auteurs qui se cachent derrière le pseudonyme « Publius » utilisé dans les *Federalist papers*
  - Essais politiques publiés en 1787-1788 dans les tout nouveaux États-Unis d'Amérique
- Sans aucun doute possible, les *papers* sont attribués à trois pères fondateurs du nouvel État : James Madison, Alexander Hamilton, John Jay

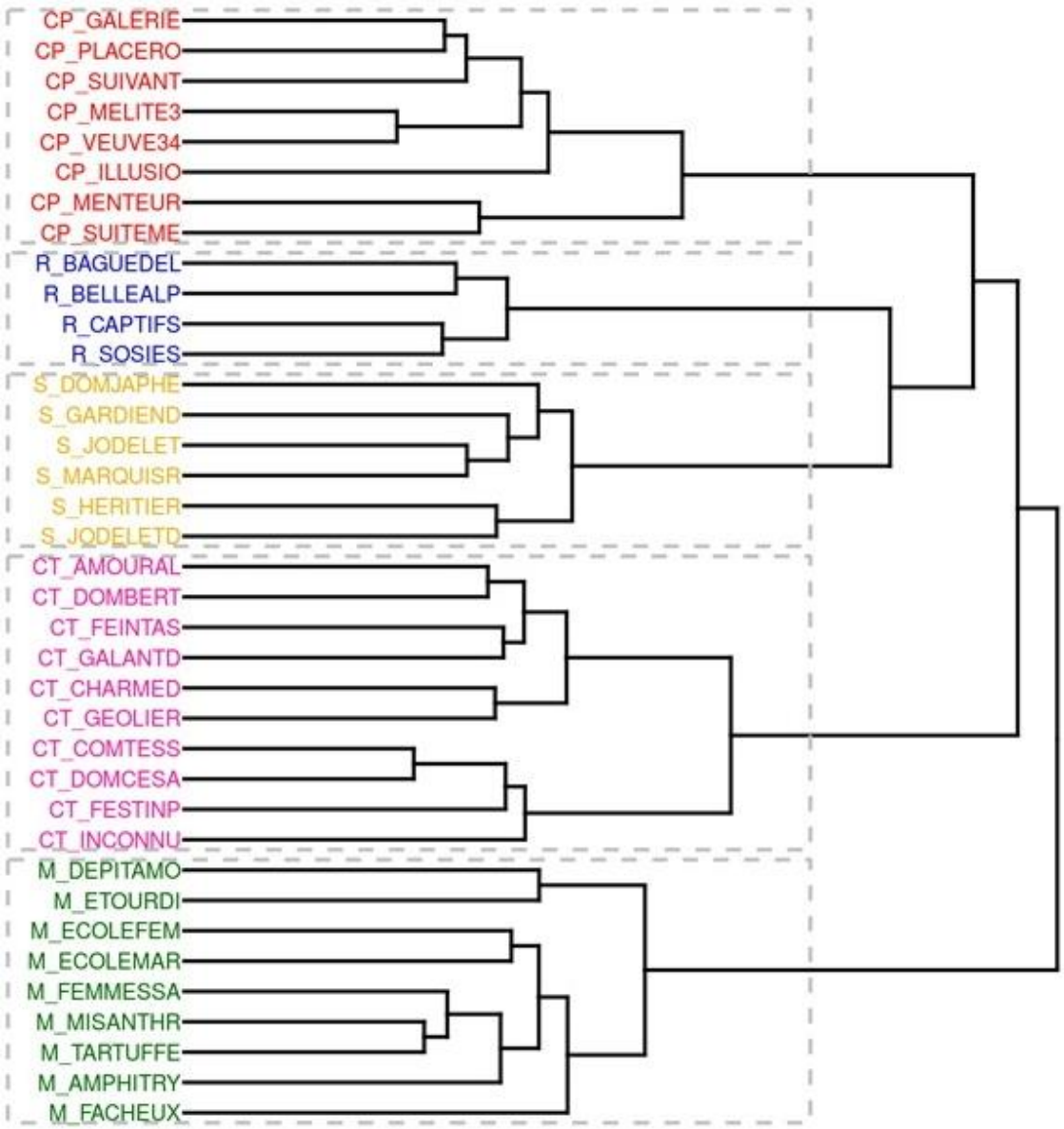
2. Un historique de la stylométrie

- Plus récemment, en 2019, Florian Cafiero et Jean-Baptiste Camps publient un article qui s'ancre dans un débat sur la paternité des œuvres de Molière
- Ses œuvres sont-elles écrites par lui ou par un porte-plume ?
- Chaque pièce devient un vecteur : on compte la fréquence des mots dans celle-ci, qui correspond à :

  nombre d'occurrences d'un mot / nombre total de mots

2. Un historique de la stylométrie

- Chaque pièce étant un vecteur, on calcule la distance entre les vecteurs et on regroupe les pièces selon leur proximité
- Pour ce faire, les auteurs utilisent un algorithme de partitionnement hiérarchique → les pièces sont rassemblées en regroupements sous la forme d'un arbre inversé


CP_GALERIE
CP_PLACERO
CP_SUIVANT
CP_MELITE3
CP_VEUVE34
CP_ILLUSIO
CP_MENTEUR
CP_SUITEME
R_BAGUEDEL
R_BELLEALP
R_CAPTIFS
R_SOSIES
S_DOMJAPHE
S_GARDIEND
S_JODELET
S_MARQUISR
S_HERITIER
S_JODELETD
CT_AMOURAL
CT_DOMBERT
CT_FEINTAS
CT_GALANTD
CT_CHARMED
CT_GEOLIER
CT_COMTESS
CT_DOMCESA
CT_FESTINP
CT_INCONNU
M_DEPITAMO
M_ETOURDI
M_ECOLEFEM
M_ECOLEMAR
M_FEMMESSA
M_MISANTHR
M_TARTUFFE
M_AMPHITRY
M_FACHEUX

3. Des stylométristes en herbe

- Maintenant, il est temps de vous-mêmes essayer d'identifier les auteurs des trois textes suivants
- Comment faire ? Il faut identifier ce qui nous permettrait d'extraire le « stylome »
  - Quels éléments des textes, à votre avis ?

Texte 1

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

Texte 2

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

Texte 3

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

3. Des stylométristes en herbe

- Maintenant, il est temps de vous-mêmes essayer d'identifier les auteurs des trois textes suivants
- Comment faire ? Il faut identifier ce qui nous permettrait d'extraire le « stylome » :
  - Les substantifs ?

Texte 1

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

Texte 2

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

Texte 3

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

Récapitulons (> 2 occ.)

Texte 1

élève	devoir	éducation	journée	lueur
4	2	2	2	2

Texte 2

collection	papillon	beauté	créature	histoire	merveille	préservation
4	4	2	2	2	2	2

Texte 3

élève	journée	apprentissage	connaissance	école	salle	science	sentiment
3	3	2	2	2	2	2	2

3. Des stylométristes en herbe

- Maintenant, il est temps de vous-mêmes essayer d'identifier les auteurs des trois textes suivants
- Comment faire ? Il faut identifier ce qui nous permettrait d'extraire le « stylome » :
  - ~~Les substantifs ?~~ → trop liés au thème du texte
  - Les pronoms ?

Texte 1

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

Texte 2

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

Texte 3

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

Récapitulons

Texte 1

ils	se
1	1

Texte 2

je	vous	elle
3	2	1

Texte 3

je	vous	celui	qui
4	2	1	1

3. Des stylométristes en herbe

- Maintenant, il est temps de vous-mêmes essayer d'identifier les auteurs des trois textes suivants
- Comment faire ? Il faut identifier ce qui nous permettrait d'extraire le « stylome » :
  - ~~Les substantifs ?~~ → trop liés au thème du texte
  - ~~Les pronoms ?~~ → trop liés au genre littéraire
  - Quoi d'autre ?

Texte 1

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

Texte 2

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

Texte 3

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

Récapitulons

Texte 1

et	mais
7	5

Texte 2

et	mais
6	4

Texte 3

et	mais	or
1	1	1

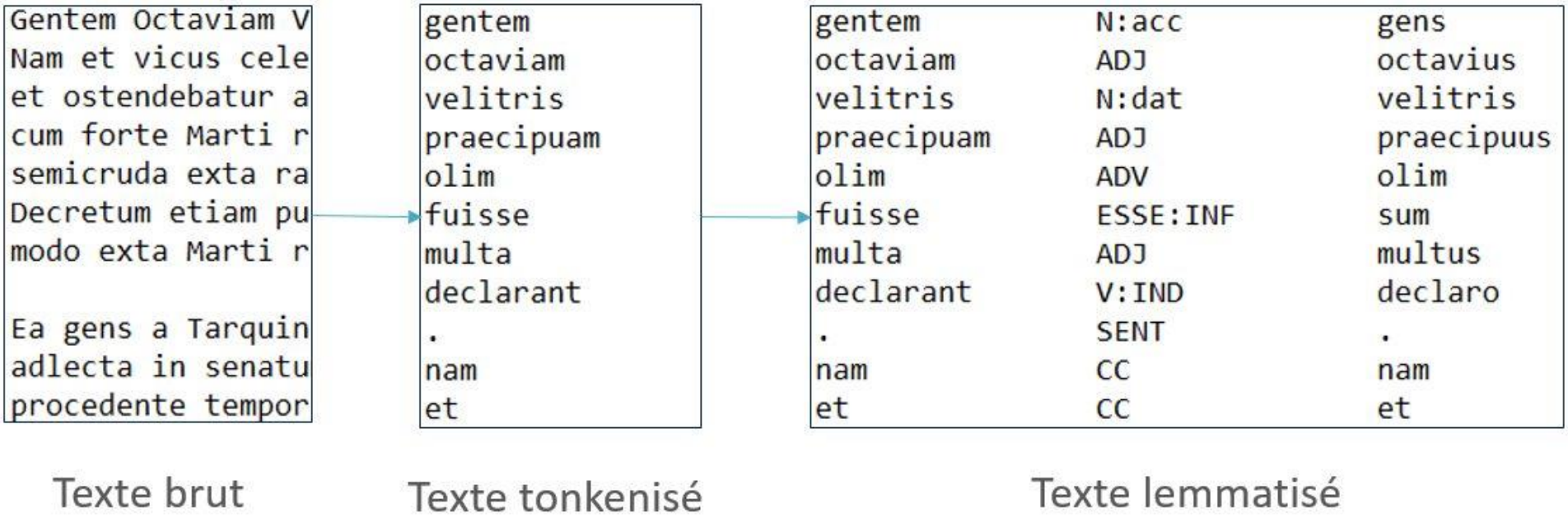
3. Des stylométristes en herbe

- Les mots vides (*stop words*) permettent d'éviter les écueils de mots trop liés au sujet du texte (substantifs, verbes, adjectifs…) ou trop liés au littéraire (par exemple les pronoms, en ce qui concerne le genre épistolaire)
- Nous nous sommes uniquement intéressés aux conjonctions de coordination, mais dans une vraie étude stylométrique, les mots vides englobent les conjonctions de subordination, les articles, certains adverbes…

4. Avec l'intelligence artificielle

- Intéressons-nous désormais à une vraie étude de cas, en latin médiéval : les écrits de trois historiens de la première croisade (tous écrits au XII^{e} s.)
  - Raoul de Caen, *Gesta Tancredi in expeditione Hierosolymitana*
  - Albert d'Aix, *Liber Christianae expeditionis pro restitutione sanctae Hierosolymitanae ecclesiae*
  - Guillaume de Tyr, *Historia rerum in partibus transmarinis gestarum*
- Les textes sont tokenisés et lemmatisés : chaque mot est renvoyé à son lemme (sa forme dans le dictionnaire)

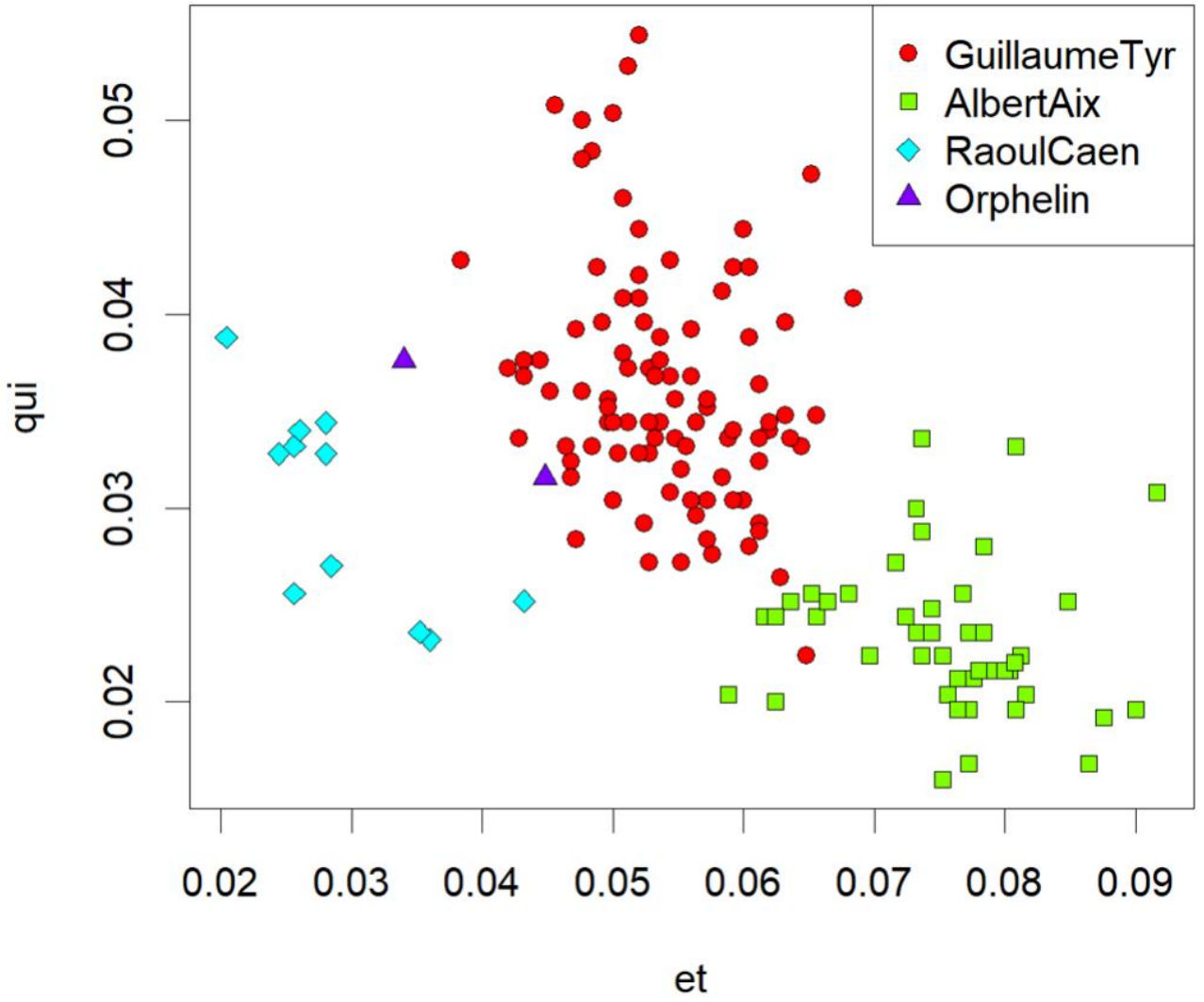
4. Avec l'intelligence artificielle



4. Avec l'intelligence artificielle

- Chaque texte est coupé en « paquets » de 2500 mots, afin de mieux refléter la variabilité des textes
- On peut décider de représenter la fréquence des deux lemmes les plus fréquents (la conjonction *et* et le pronom relatif *qui*) sur un nuage de points bidimensionnels :
  - Un point = un paquet de mots
  - Sa coordonnée en x = la fréquence relative de *et*
  - Sa cordonnée en y = la fréquence relative de *qui*
  - Sa couleur et sa forme = l'auteur du texte (ou orphelin)

4. Avec l'intelligence artificielle



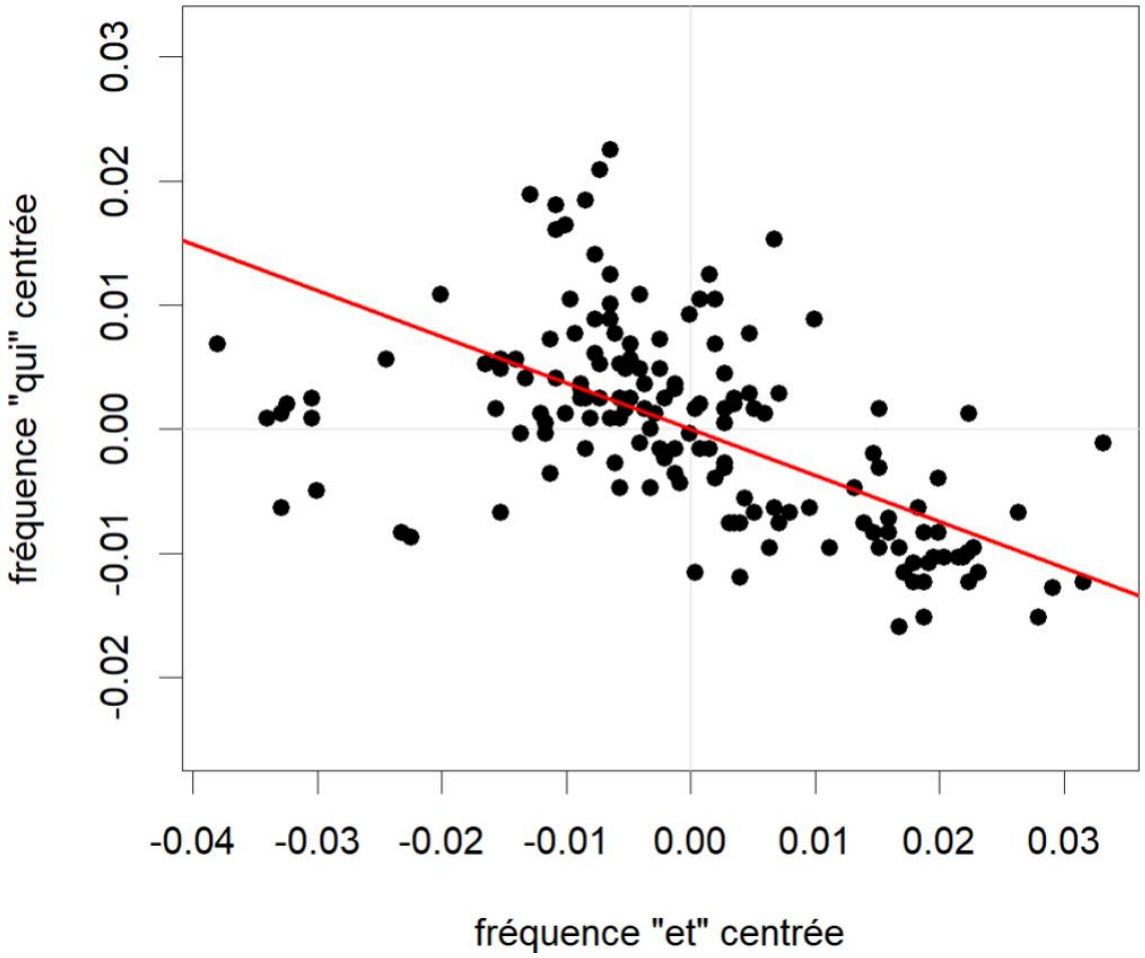
4. Avec l'intelligence artificielle

- Les points mauves (l'orphelin) se trouvent loin des verts (Albert d'Aix), mais à égale distance entre les rouges (Guillaume de Tyr) et les turquoises (Raoul de Caen)
- Ce n'est donc pas concluant… mais on n'a considéré que deux lemmes !
- Pour comparer plus de fréquences, il faut pouvoir visualiser efficacement les résultats, car il y aurait autant de dimensions que de fréquences comparées

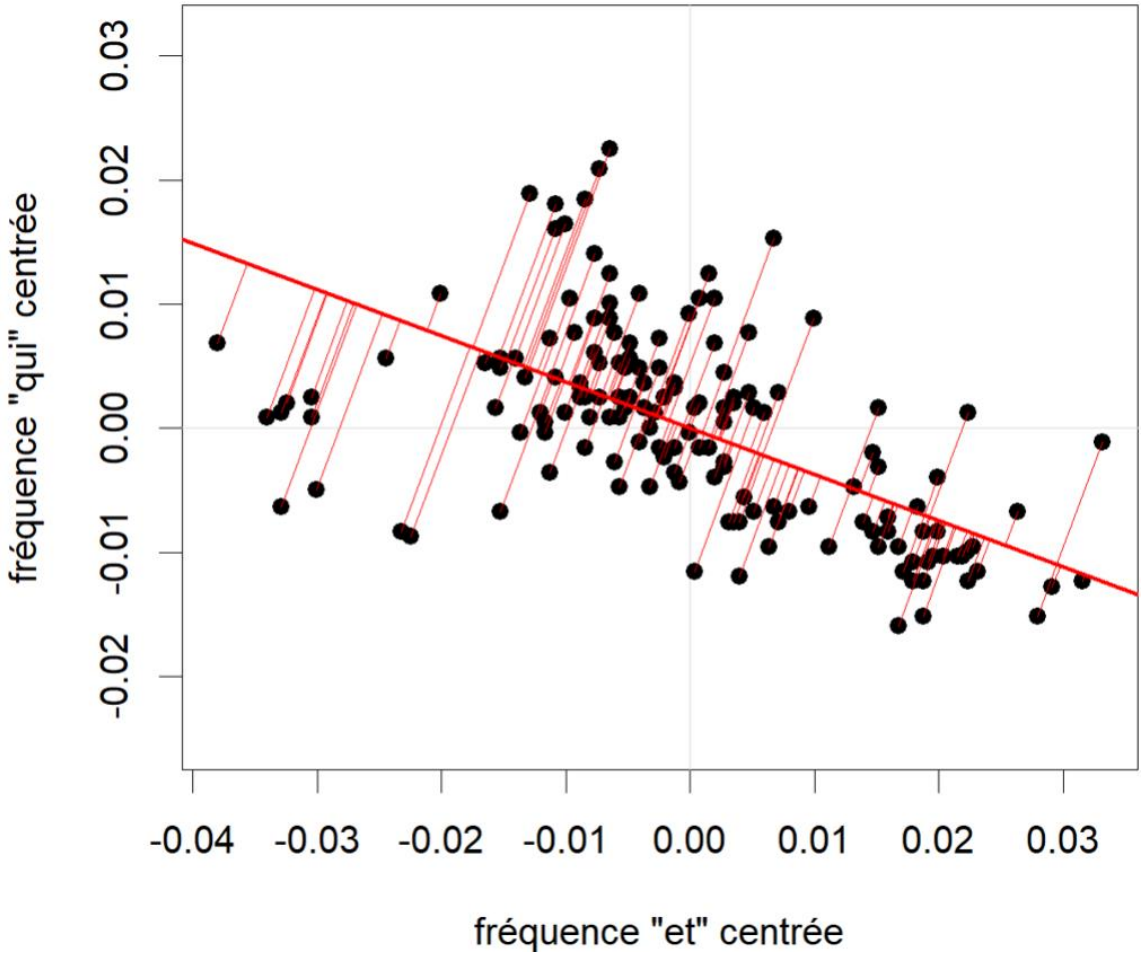
4. Avec l'intelligence artificielle

- L'intelligence artificielle nous vient en aide, avec l'analyse en composantes principales (ACP) : elle permet de synthétiser l'information en un certain nombre de composantes principales
- Prenons un exemple : le résumé de deux variables en une variable, c'est-à-dire en termes géométriques le passage d'une dimension 2 en une dimension 1

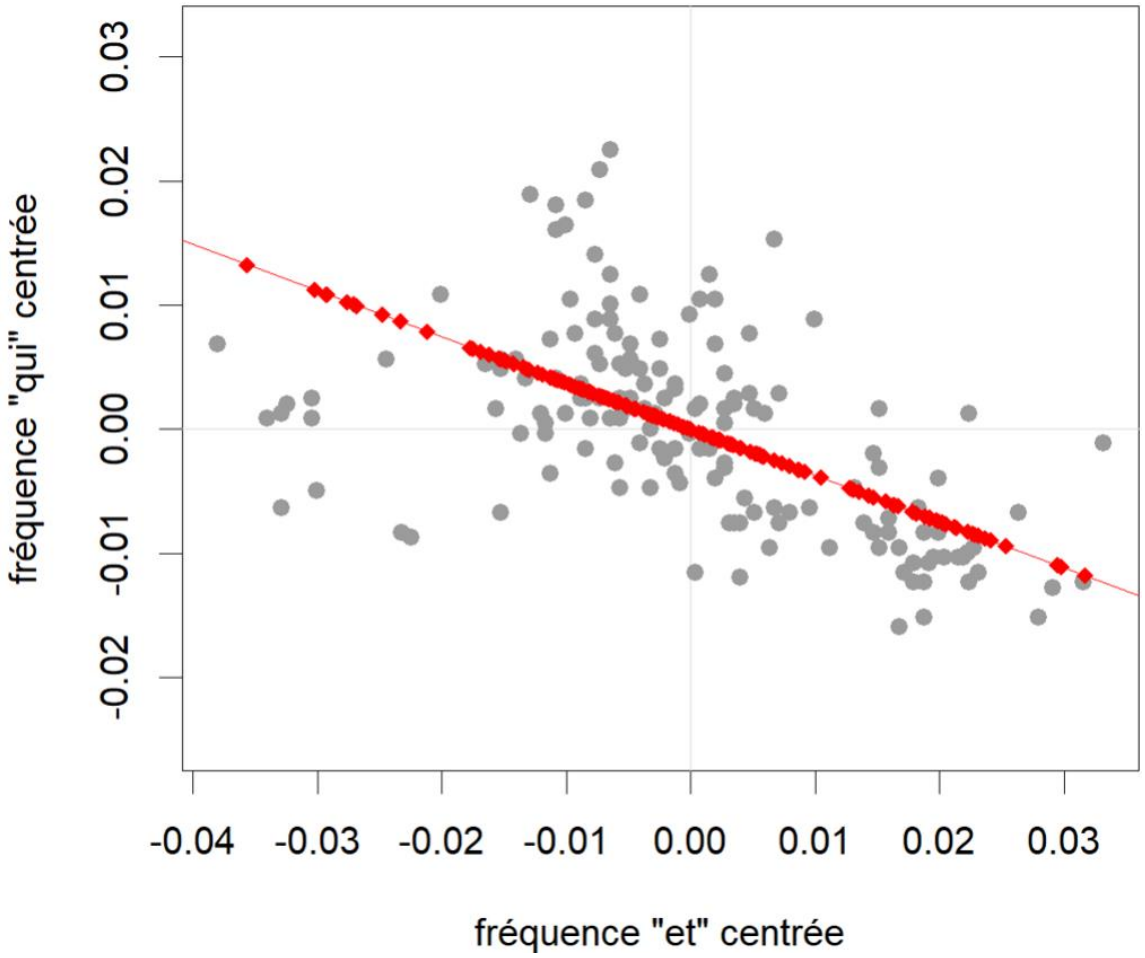
4. Avec l'intelligence artificielle



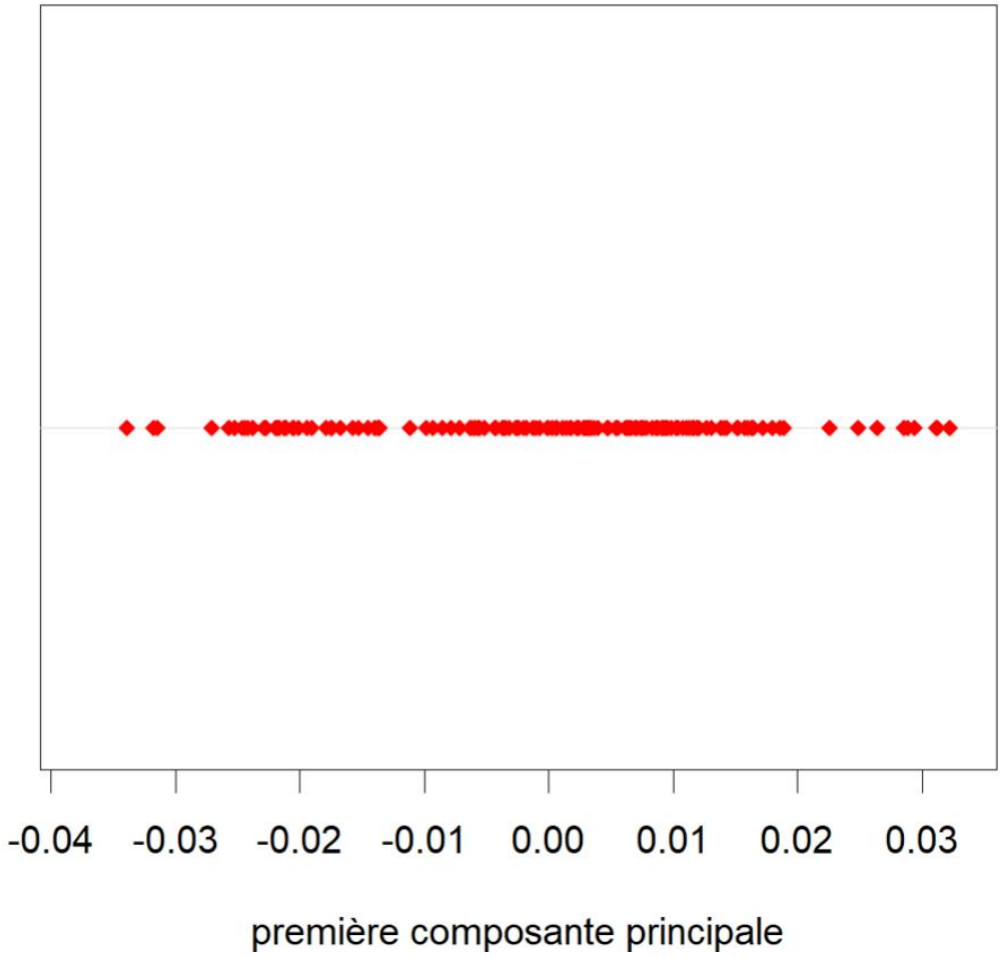
4. Avec l’intelligence artificielle



4. Avec l’intelligence artificielle

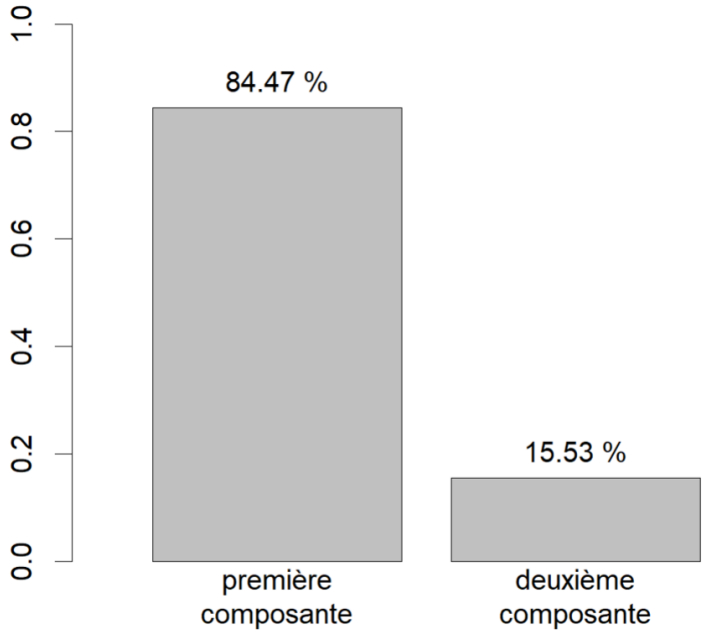


4. Avec l’intelligence artificielle

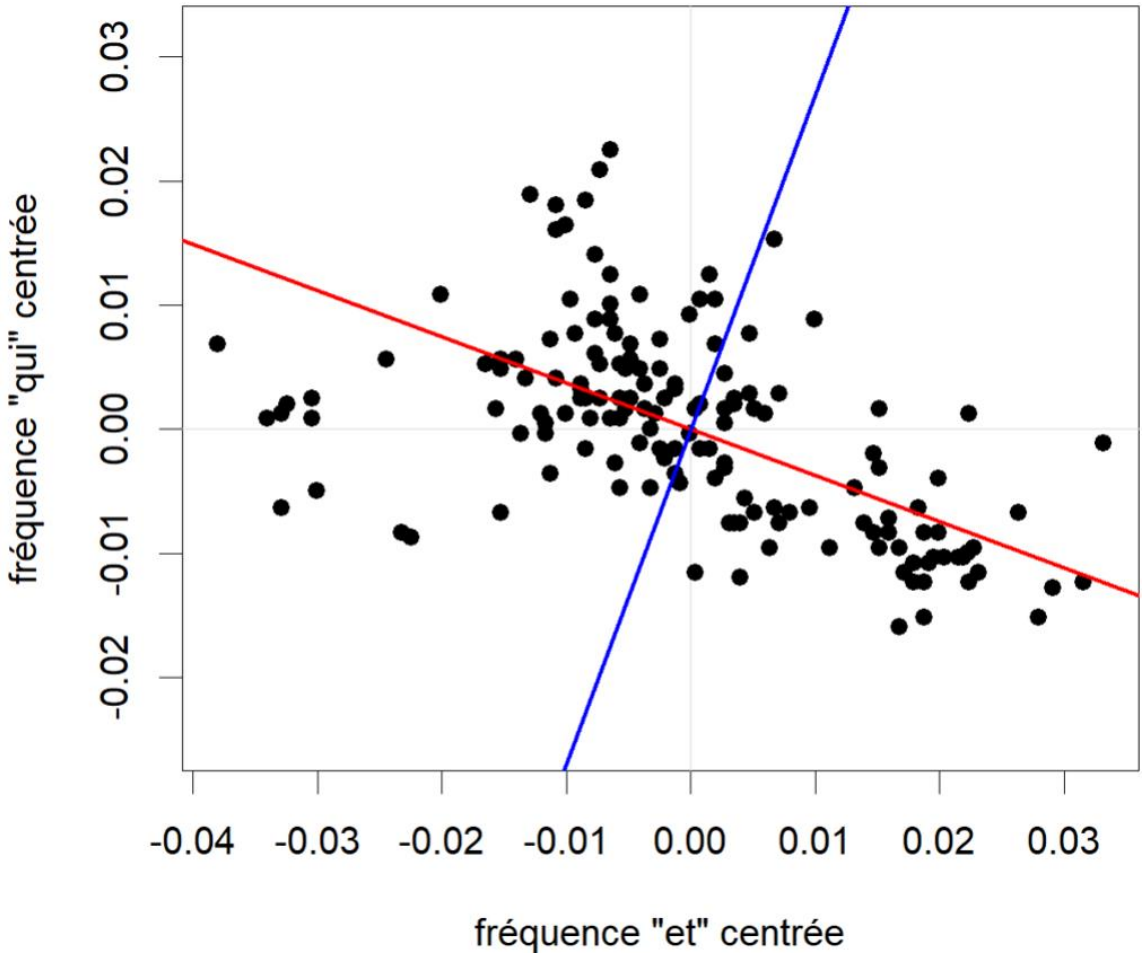


4. Avec l'intelligence artificielle

- L'ACP permet de résumer l'information, mais elle en perd forcément : on appelle première composante principale la composante qui résume au mieux l'information
- Dans notre exemple, la première composante résume près de 85 % de l'information

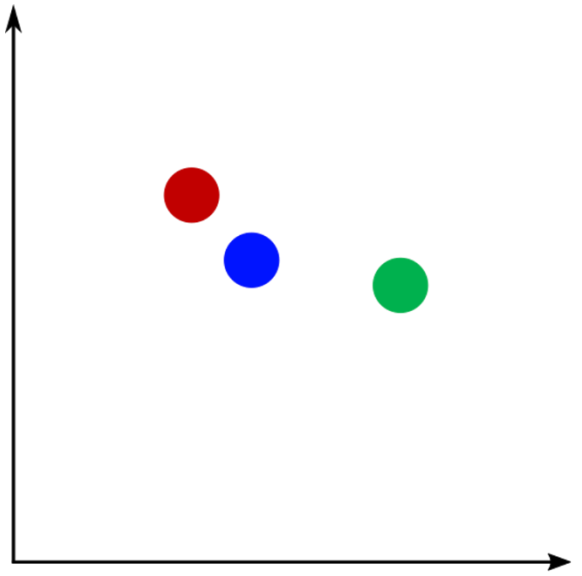


4. Avec l’intelligence artificielle



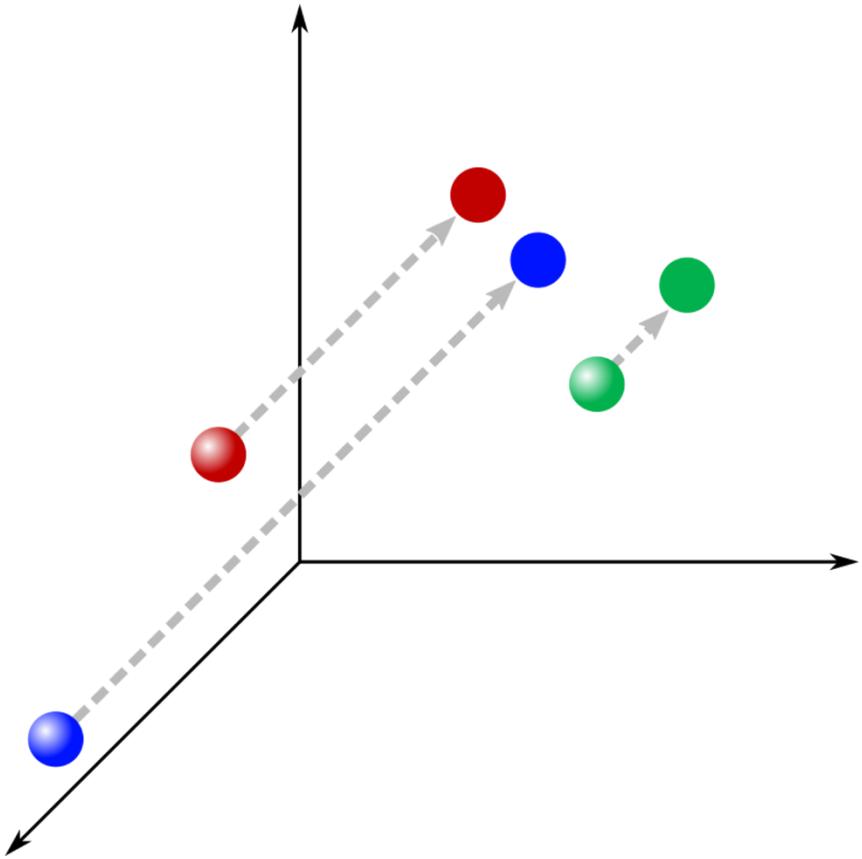
4. Avec l’intelligence artificielle

- Pour mieux comprendre la perte d’information, considérons le graphique en deux dimensions ci-contre
- Les trois ronds ont l’air d’être proches l’un de l’autre…



4. Avec l'intelligence artificielle

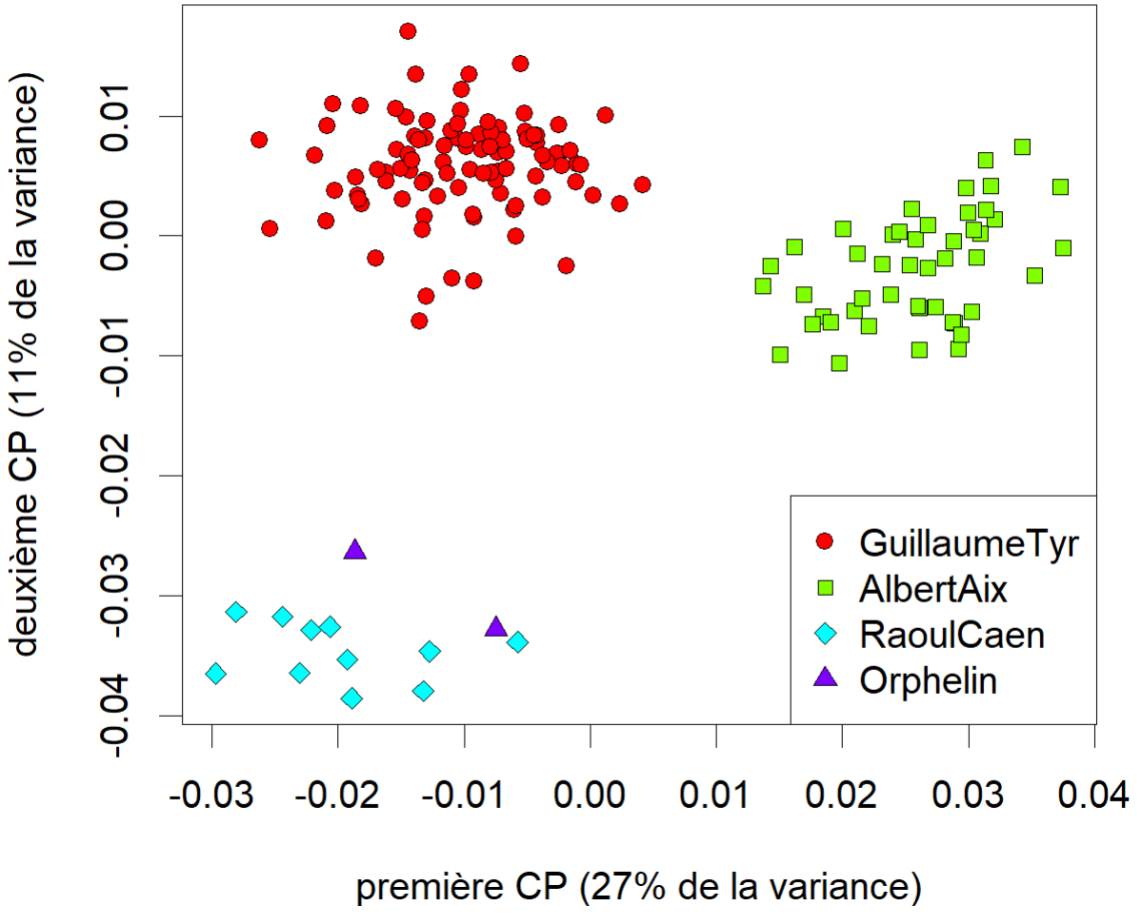
- Les trois ronds ont l'air d'être proches l'un de l'autre…
- Mais l'axe z (la profondeur) a été complétement écrasée lors du passage en dimension 2
- Les sphères ne sont pas proches les unes des autres en dimension 3 !



4. Avec l'intelligence artificielle

- Revenons sur notre exemple médiéval : on décide de prendre la fréquence de tous les 7843 lemmes, soit un résultat en autant de dimensions
- Pour le visualiser, on fait une ACP pour réduire toutes ces dimensions en 2 dimensions, les deux premières composantes principales :
  - Elles représentent 27 % et 11 % de l'information respectivement
  - On perd donc 62 % de l'information !

4. Avec l'intelligence artificielle



4. Avec l'intelligence artificielle

- On constate que le résultat, malgré la perte d'information, est sans équivoque : les deux points mauves s'éloignent considérablement des points rouges et des points verts
  - Il faut en effet considérer l'éloignement et non le rapprochement sur le graphique
  - Rappelez-vous de la projection des sphères en dimension 2
- Notre anonyme est donc Raoul de Caen, ce qui est tout à fait vrai : il est bien l'auteur du texte qui a « perdu » son étiquette d'auteur pour le bien de notre exemple

5. Au-delà de la littérature

- La stylométrie est utilisée dans un contexte judiciaire :
  - Affaire Unabomber aux États-Unis : Theodore Kaczynski est identifié grâce à la ressemblance de style entre le manifeste publié sous le nom d'Unabomber et une lettre écrite en son nom propre
  - Affaire Grégory en France : lors d'un rebondissement en 2021 dans cette vieille affaire criminelle, une analyse stylométrique désigne la grand-tante comme autrice de la lettre de revendication du meurtre

5. Au-delà de la littérature

- Q est le pseudonyme du fondateur du mouvement conspirationniste QAnon, qui diffuse ses thèses complotistes sur Internet dans l'anonymat
- Il y avait de nombreuses hypothèses sur son identité : un haut fonctionnaire de l'administration américaine, un haut gradé militaire, voire Trump lui-même
- Toutefois les analyses stylométriques pointent vers deux personnages, qui ont été successivement Q : Paul Furber puis Ron Watkins

5. Au-delà de la littérature

- Les chercheurs ont utilisé les mots-outils, difficiles à manipuler
- On peut facilement changer son vocabulaire, par exemple utiliser souvent le mot « fake » pour parler comme Trump
- Mais c'est plus difficile de contrôler les conjonctions ou les prépositions qu'on emploie → domaine de l'inconscient

5. Au-delà de la littérature

- La stylométrie pose aussi des questions éthiques : est-ce toujours justifié de vouloir identifier quelqu'un qui souhaite garder l'anonymat ?
- C'est le cas de Satoshi Nakamoto, pseudonyme du créateur du Bitcoin : plusieurs analyses stylométriques ont été menées et plusieurs identités proposées
- Même si rien n'a été concluant jusqu'ici, la question de la garantie de l'anonymat sur Internet se pose

Merci pour votre attention !

DES QUESTIONS ?

Contact :

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

Bibliographie

Baron, C. (1897), « Contributions à la chronologie des dialogues de Platon », in *Revue des Études Grecques*, 10, nº 39,

p. 264-278.

Cafiero, F. et Camps, J.-B. (2019), « Why Molière Most Likely Did Write His Plays », in *Science Advances*, 5, nº 11.

Campbell , L. (1867), *The Sophistes and Politicus of Plato, with a revised text and English notes*, 1867.

de Morgan, S. E. (1882), *Memoir of Augustus de Morgan by his Wife Sophia Elizabeth de Morgan With Selections From His Letter*, Longmans, Green, et Co, 1882.

Laan, N. M. (1995), « Stylometry and Method. The Case of Euripides », in *Literary and Linguistic Computing*, 10, nº 4, p. 271-278.

Love, H. (2002), *Attributing Authorship : An Introduction,* Cambridge University Press.

Mosteller, F. et Wallace, D. L. (1963), « Inference in an Authorship Problem », in *Journal of the American Statistical Association*, 58, nº 302, p. 275-305.

Presse

Affaire Grégory :

- https://www.20minutes.fr/justice/3021903-20210423-affaire-gregory-experts-stylometrie-principal-corbeau-jacqueline-jacob
- https://www.francetvinfo.fr/faits-divers/affaire-du-petit-gregory/affaire-gregory-en-quoi-consiste-la-stylometrie-cette-technique-d-analyse-de-l-ecriture-qui-a-permis-de-relancer-l-enquete_4221997.html

QAnon :

- https://legrandcontinent.eu/fr/2022/03/18/comment-nous-avons-trouve-qui-etait-derriere-qanon/
- https://www.sciencesetavenir.fr/high-tech/la-stylometrie-a-la-recherche-de-q-le-mysterieux-internaute-a-l-origine-du-mouvement-conspirationniste-qanon_161694
- https://www.philomag.com/articles/comment-la-stylometrie-permis-didentifier-qui-se-cachait-derriere-qanon

Satoshi Nakamoto :

- https://news.bitcoin.com/a-look-at-stylometry-can-we-uncover-satoshi-through-literary-quirks/
- https://www.developpez.com/actu/157522/La-NSA-aurait-decouvert-l-identite-reelle-de-Satoshi-Nakamoto-le-mysterieux-createur-du-Bitcoin/