

‘Rien n'est plus trompeur qu'un fait évident’

Affaires de style et
algorithmes

Florian Cafiero^{1,2}

Jean-Baptiste Camps^{3,4}

9 mai 2022

³Univ. Ca' Foscari, Venezia; ⁴École nationale des chartes, PSL



Arthur Conan Doyle, “The Boscombe Valley Mystery”, *The Strand Magazine*, October 1891, ill. Sidney Paget.

Sherlock Holmes, le style et la forme des oreilles

The Boscombe Valley Mystery

Dans *The Boscombe Valley Mystery* (1891), Holmes est appelé à enquêter dans un meurtre, avec un ‘coupable’ évident, que tout accuse... (témoins, cris, mobile,...) mais qu'il innocentera grâce à des empreintes de semelle.



Arthur Conan Doyle, “The Boscombe Valley Mystery”, *The Strand Magazine*, October 1891, ill. Sidney Paget.

Quels textes sont les plus proches ?

Federalist, no. 6

Carthage, quoique république commerciale, fut l'agresseur dans la guerre même qui aboutit à sa destruction.

Hannibal avait porté ses armes au cœur de l'Italie et jusqu'aux portes de Rome, avant que Scipion, à son tour, ne lui fasse subir un revers dans les territoires de Carthage, et ne fasse la conquête des colonies.

Federalist, no. 18

Même au milieu de guerres défensives et dangereuses avec la Perse et la Macédoine, les membres n'agissaient jamais de concert, et étaient, plus ou moins nombreux, éternellement les dupes ou les mercenaires de l'ennemi commun. Les intervalles de la guerre étrangère étaient remplis par

Federalist, no. 32

Quant à la supposition d'une incompatibilité entre le pouvoir de taxation des États et celui de l'Union, elle ne peut être soutenue dans le sens qui entraînerait l'exclusion des États. Il est, en effet, possible qu'une taxe soit établie sur un article particulier par un État, ce qui rendrait inexpédient l'établissement d'une autre taxe sur le même article par l'Union.

Une réponse évidente...

Federalist, no. 6

Carthage, quoique république commerciale, fut l'agresseur dans la guerre même qui aboutit à sa destruction.

Hannibal avait porté ses armes au cœur de l'Italie et jusqu'aux portes de *Rome*, avant que *Scipion*, à son tour, ne lui fasse subir un revers dans les territoires de Carthage, et ne fasse la conquête des colonies.

Federalist, no. 18

Même au milieu de guerres défensives et dangereuses avec la *Perse* et la *Macédoine*, les membres n'agissaient jamais de concert, et étaient, plus ou moins nombreux, éternellement les dupes ou les mercenaires de l'ennemi commun. Les intervalles de la guerre étrangère étaient remplis par

Federalist, no. 32

Quant à la supposition d'une incompatibilité entre le pouvoir de taxation des États et celui de l'Union, elle ne peut être soutenue dans le sens qui entraînerait l'exclusion des États. Il est, en effet, possible qu'une taxe soit établie sur un article particulier par un État, ce qui rendrait inexpédient l'établissement d'une autre taxe sur le même article par l'Union.

... mais fausse!

Federalist, no. 6
Carthage, quoique république commerciale, fut l'agresseur dans la guerre même qui aboutit à sa destruction.

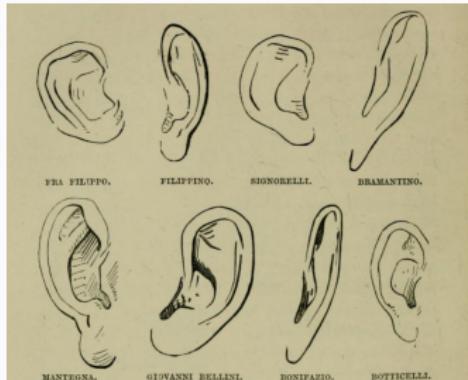
Hannibal avait porté ses armes au cœur de l'Italie et jusqu'aux portes de Rome, avant que Scipion, à son tour, ne lui fasse subir un revers dans les territoires de Carthage, et ne fasse la conquête des colonies.

A. Hamilton

Federalist, no. 18
Même au milieu de guerres défensives et dangereuses avec la Perse et la Macédoine, les membres n'agissaient jamais de concert, et étaient, plus ou moins nombreux, éternellement les dupes ou les mercenaires de l'ennemi commun. Les intervalles de la guerre étrangère étaient remplis par J. Madison

Federalist, no. 32
Quant à la supposition d'une incompatibilité entre le pouvoir de taxation des États et celui de l'Union, elle ne peut être soutenue dans le sens qui entraînerait l'exclusion des États. Il est, en effet, possible qu'une taxe soit établie sur un article particulier par un État, ce qui rendrait inexpédient l'établissement d'une autre taxe sur le même article par l'Union.
A. Hamilton

En quête d'indices de style



Giovanni Morelli, *Italian Painters*,
1892. cf. Carlo Ginzburg, 'Morelli,

Freud and Sherlock Holmes : Clues
and Scientific Method', (1980).

Richard W. Bailey (1970)
Need properties that are

- salient
- structural,
- frequent,
- easily quantifiable,
- relatively immune from conscious control

"Authorship Attribution in a Forensic Setting", 1979.

Du travail de détective à la science

Pour passer dans le domaine de la science, nous avons besoin de propriétés

- systématiques,
- fondés théoriquement ET empiriquement

Plan

Sherlock Holmes, le style et la forme des oreilles

Brève introduction à la stylométrie

Mesurer le style ?

Style, idiolecte et stylome

Propriétés inconscientes du style

Les affaires de style : un coup d'œil aux algorithmes

La vièle et le stylomètre

Corneille et Molière

QAnon

Smells like stylometry spirit

Impostors and pseudonyms (with some word of ethics)

Brève introduction à la stylométrie

Plan

Sherlock Holmes, le style et la forme des oreilles

Brève introduction à la stylométrie

Mesurer le style ?

Style, idiolecte et stylome

Propriétés inconscientes du style

Les affaires de style : un coup d'œil aux algorithmes

La vièle et le stylomètre

Corneille et Molière

QAnon

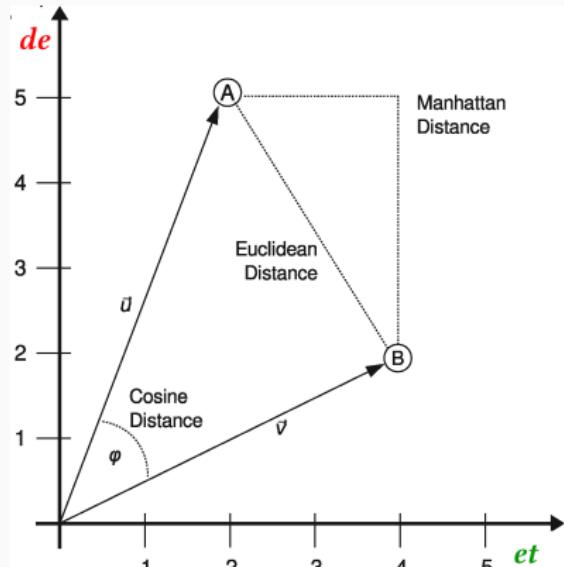
Smells like stylometry spirit

Impostors and pseudonyms (with some word of ethics)

Compter les mots, mesurer le style

Texte A
et de joie, il dit à son voisin de
classe de gauche qu'il
s'enthousiasmait de cette
conférence et de cette journée.

Texte B
et de chaque côté de lui, il vit
chiens et chats et poules et
canards.



d'après Evert et al.

Mesurer le style : la stylométrie

- Existence d'un signature individuelle dans la langue d'une personne donnée :
stylome (génome ou empreinte digitale stylistique);
- que l'on cherche à détecter quantitativement.

Les linguistes parlent
d'**idiolecte** (« ensemble
des usages du langage
propre à un individu
donné », Wik. fr.)

Typologie

diatopique → dialecte;

diachronique →
chronolecte;

diastratique →
sociolecte;

diaphasique →



Manhunt Unabomber (TV Series)

Deux mots d'histoire...

Émergence dans la deuxième moitié du 19e s.

Contexte : Méthode d'enquête (médecine, détective, ...)

Début des approches scientifiques aux textes.

Homère, Bible, Shakespeare (c. 1850)...

Quelques dates

- 1851 : Augustus de Morgan propose une analyse quantitative du vocabulaire pour l'attribution des Épîtres de Paul.
- 1861 : Dittenberg quantifie des particularités de style non relatives au contenu philosophique pour dater les textes de Platon.
- 1897 : Lutoslawski donne à cette approche le nom de stylométrie.
- 1880-1900 : Mendenhall et sa machine à compter.

Paul Tannery, "La stylométrie : ses origines et son présent", *Revue Philosophique de la France et de l'Étranger*, 47 (1899), pp. 159-169, www.jstor.org/stable/41076911.

Plato et la naissance de la stylométrie

Problème

Comment établir la chronologie relative des dialogues de Platon ?

La méthode de Dittenberger

	1. <i>καὶ μῆν</i>	2. <i>ἄλλὰ μῆν</i>	3. <i>τι μῆν;</i>	4. <i>γε μῆν</i>	5. <i>ἄλλα- μῆν</i>	Gesamt- zahl der Beispiele von <i>μῆν</i>	Umfang nach Seiten der Hermann- schen Ausgabe
I	<i>Κρίτων</i>	—	—	—	—	—	17
	<i>Εὐθύφρων</i>	1	1	—	—	2	23
	<i>Πρωταγόρας</i>	2	3	—	—	5	63
	<i>Χαροκόπης</i>	2	2	—	—	7	29
	<i>Αάχης</i>	3	2	—	—	7	32
	<i>Τηνίας</i>	2	3	—	—	5	20
	<i>Εὐθύδημος</i>	4	5	—	—	12	45
	<i>Μένων</i>	5	5	—	—	10	46
	<i>Γοργίας</i>	9	11	—	—	24	116
	<i>Κρατύλος</i>	9	9	—	—	18	79
	<i>Φαιδῶν</i>	7	10	—	—	20	79
a	<i>Συμπόσιον</i>	9	2	—	1	2	15
	<i>Δάσις</i>	2	2	—	—	4	11
	<i>Φαιδρος</i>	3	4	11	1	1	24
	<i>Πολιτεία</i>	44	44	34	2	11	158
	<i>Θεαίητος</i>	11	6	13	1	1	38
II	<i>Παρμενίδης</i>	25	25	6	5	2	81
	<i>Φίληβος</i>	20	7	26	7	2	78
	<i>Σοφιστής</i>	24	10	12	5	2	72
	<i>Πολιτικός</i>	24	7	20	8	3	75
	<i>Νέμων</i>	26	8	48	24	2	166
b	<i>Παρμενίδης</i>	25	25	6	5	2	50
	<i>Φίληβος</i>	20	7	26	7	2	87
	<i>Σοφιστής</i>	24	10	12	5	2	82
	<i>Πολιτικός</i>	24	7	20	8	3	83
	<i>Νέμων</i>	26	8	48	24	2	117

Wincenty Lutosławski, fondateur polonais de la stylométrie

18 June 1897, discours à l'*Académie des Inscriptions et Belles-Lettres (Paris)*

« Un **stylème** est tout usage du langage qui n'est pas indispensable pour exprimer la pensée, c'est-à-dire, qui peut être remplacé par un autre. [...] la fréquence de chaque genre de mots est un stylème, même si les mots dont on calcule la fréquence sont indispensables. De même, la prévalence d'un synonyme sur un autre, la position des mots dans la phrase, la proportion des parties du discours, forment autant de stylèmes. »

« **L'affinité stylistique** de deux textes dépend du nombre de stylèmes qu'ils ont en commun et de leur importance »

« La **mesure de l'affinité stylistique** doit tenir compte de la différente importance des stylèmes. »

(publ. in *Principes de stylométrie*, 1898)

Mendenhall, un self-made-man faisant la spectrographie des textes



Mendenhall et sa machine à compter

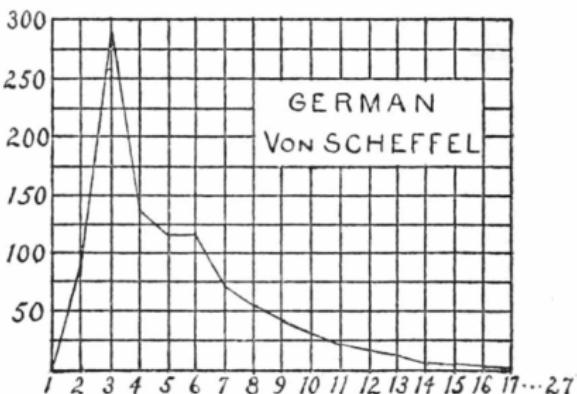
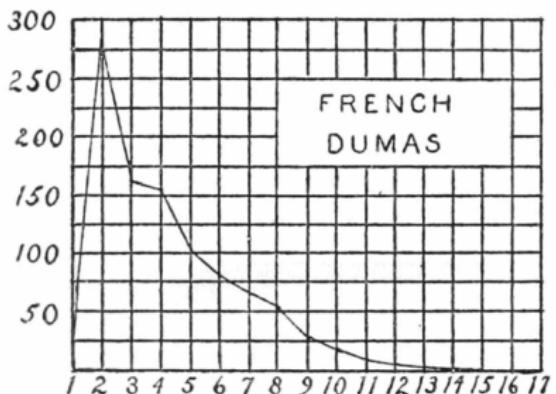
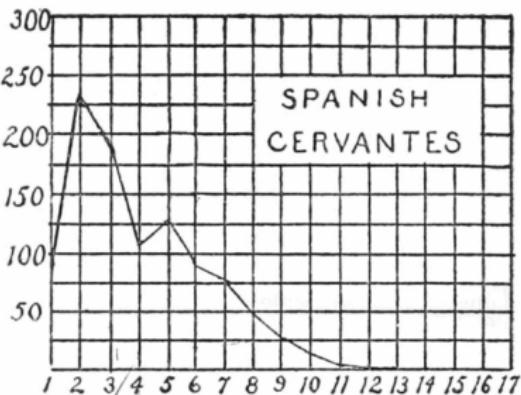
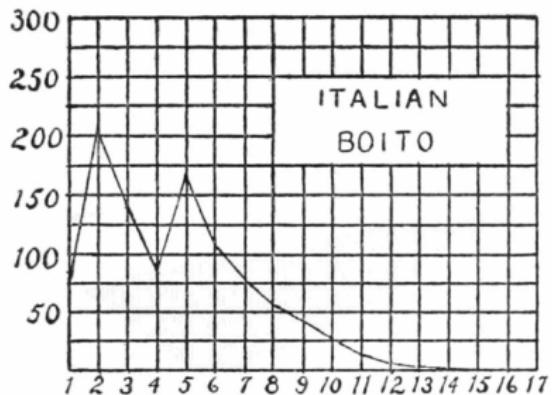
I devised a method for exhibiting graphically such peculiarities of style in composition as seemed to be almost purely mechanical and of which an author would usually be absolutely unconscious. (...) its application required no exercise of judgment, accurate enumeration being all that was necessary, and by displaying one or more phases of the mere mechanism of composition characteristics might be revealed which the author could make no attempt to conceal, being himself unaware of their existence. It was further assumed that, owing to the well-known persistence of unconscious habit, personal peculiarities in the construction of sentences, in the use of long or short words, in the number of words in a sentence, etc., will in the long run manifest themselves with such regularity that their graphic representation may become a means of identification, at least by exclusion.

Mendenhall (T. C.), « A Mechanical Solution of a Literary Problem », *The Popular Science Monthly*, 60 (1901), p. 97-105,

The « counting machine » (=2 female coders and a keyboard)

the heavy task of counting was performed by the ladies who undertook it, Mrs. Richard Mitchell and Miss Amy C. Whitman, of Worcester, Massachusetts. Their intelligent interest in the problem itself, together with their excellent knowledge of the various authors under examination and familiarity with the literature of the Shakespearean period, contributed greatly to the easy accomplishment of the work. The operation of counting was greatly facilitated by the construction of a simple counting machine by which a registration of a word of any given number of letters was made by touching a button marked with that number. One of the counters, with book in hand, called off 'five,' 'two,' 'three,' etc., as rapidly as possible (...) the other registering, as called, by pressing the proper buttons. Practice enabled the counters to do the work with remarkable rapidity (...) they were occupied for several months (...) The work was very exhausting, however, and could not be kept up satisfactorily more than three to five hours each day.

Mendenhall profiles



Et après

1850-1900

- spectrographie de Mendenhall : écho important (publié dans *Science*), mais ne fonctionne pas. Mesure le genre.
- stylométrie de Dittenberger-Lutosławski sur les mots-outils : fonctionne, a donné des résultats sur Platon qui sont toujours étudiés, mais postérité limitée durant les 50 années suivantes.

1900-1950

Les mathématiciens lisent des romans

- 1913 : A. Markov sur la propriété des lettres dans Pushkin, *Eugene Onegin*;
- Sherman (1893) et Yule (1939) sur la longueur des phrases (toujours générique, et pas idiolectal);
- Estoup, et Zipf (1949) sur les fréquences de mots (Joyce, *Ulysses*) et la richesse lexicale.

Tout change avec l'arrivée de l'ordinateur

Quelques dates

- 1962 : Alvar Ellegård sur les Lettres de Junius (textes anonymes critiquant Georges III);
- 1963 : Andrew Morton, sur les Épîtres de Paul;
- 1963-1964 : Mosteller et Wallace sur les *Federalist papers*
- 1966 : J. Leed fait paraître *The Computer & Literary style* (1^{er} manuel ?).
- 1980-1990's : Burrows (English) and Benzécri (français) popularisent les approches multivariées fondées sur des calculs de distance et l'analyse factorielle.

Plan

Sherlock Holmes, le style et la forme des oreilles

Brève introduction à la stylométrie

Mesurer le style ?

Style, idiolecte et stylome

Propriétés inconscientes du style

Les affaires de style : un coup d'œil aux algorithmes

La vièle et le stylomètre

Corneille et Molière

QAnon

Smells like stylometry spirit

Impostors and pseudonyms (with some word of ethics)

Style

Si la stylométrie est la mesure du style, en général dans une perspective d'attribution, **qu'est-ce que le style ?**

Est-ce vraiment unique ?

la définition d'un stylisticien (Molinier, 1996)

Le stylème est un caractérisème de littérarité

c'est-à-dire quelque chose qui n'est pas lié ni à la nécessité syntaxique, ni à la complétude sémantico-informative de l'énoncé. caractérise un discours littéraire comme littéraire (et pas seulement informatif).

Cf. Gabay, *Beyond Idiolectometry...*

Style et genre littéraire.

figure de style.

Commentaire stylistique.

Style comme un signal individuel ?



Le style, c'est l'homme même.

Georges-Louis Leclerc, comte de Buffon,
Discours sur le style (1753).

En littérature, la ‘naissance du style individuel’ en tant que représentation est une évolution historique : passage de l’imitation à l’expression de l’individualité.

Style = idiolectes ?

Stylome = génome stylistique individuel ?

Postulat de base de la stylométrie

Postulat

Chaque individu (et au-delà, chaque catégorie d'individus) emploie une langue démontrant des *propriétés particulières* et **mesurables**, appelées **stylome**.

Ce postulat est aujourd'hui généralement admis comme valide empiriquement. Même au delà de textes littéraires, on peut discerner un ensemble de traits linguistiques caractéristiques d'un individu.

L'expérience de Nimègue

van Halteren et al., "New Machine Learning Methods Demonstrate the Existence of a Human Stylome", *Journal of Quantitative Linguistics*, 2005

- Demande à des étudiants d'écrire une série de textes sur une série de 9 sujets imposés : 3 sujets de fiction, 6 sujets de non fiction.
- Rédigent ensemble dans le même amphithéâtre dans un temps imparti.
- Textes obtenus assez courts : entre 628 et 1342 mots par texte.
- Tentent de faire repérer les différents auteurs par une intelligence artificielle.
- Besoin de donner beaucoup d'information à l'intelligence artificielle (pas que les mots outils, mais le lexique, les POS), mais fini par identifier les étudiants.

Des performances de plus en plus surprenantes

Soroush Vosoughi, Helen Zhou, and Deb Roy. (2015). "Digital Stylometry : Linking Profiles Across Social Networks", *Proceedings of the 7th International Conference on Social Informatics* (SocInfo 2015).

- Etude du "Laboratory for social machines" du MIT Media Lab
- Intéressés par l'étude des contenus postés sur les médias sociaux.
- Problème : identités fragmentées : une même personne peut poster sur différentes plateformes à la fois (Facebook, Twitter etc.)
- Ont tenté de retrouver le propriétaire de comptes Facebook et Twitter, grâce à des méthodes stylométriques, sur une base de données d'environ 5500 comptes.
- Réussissent la tâche d'identifier correctement les auteurs dans un tiers des cas.
- Parait peu, mais déjà énorme, vu la brièveté des textes et le nombre d'auteurs candidats... Et progrès en cours.

Plan

Sherlock Holmes, le style et la forme des oreilles

Brève introduction à la stylométrie

Mesurer le style ?

Style, idiolecte et stylome

Propriétés inconscientes du style

Les affaires de style : un coup d'œil aux algorithmes

La vièle et le stylomètre

Corneille et Molière

QAnon

Smells like stylometry spirit

Impostors and pseudonyms (with some word of ethics)

Propriétés inconscientes du style

Mots-outils.

Calcul de leurs fréquences, et des différences entre les fréquences.

Depuis le XIXe siècle : propriétés inconscientes du style, moins dépendantes du contenu (Dittenberg, Mendenhall,...).

Mots-outils

Petite catégorie, mais très fréquente. Fonctions grammaticales, syntaxiques (structurantes), mais peu de sens en soi. E.g. : *of, the, by,*

...

Mosteller et Wallace (1963), *Federalist papers*

TABLE 2.2. FREQUENCY DISTRIBUTION FOR *war*

Rate/1000	H	M
0 (exactly)	23	15
0+-2	16	13
2-4	4	5
4-6	2	4
6-8	1	3
8-10	1	3
10-12	—	3
12-14	—	2
14-16	1	2
Totals	48	50

TABLE 2.3. FREQUENCY DISTRIBUTION FOR *upon*

Rate/1000	H	M
0 (exactly)	—	41
0+-1	1	7
1 -2	10	2
2 -3	11	
3 -4	11	
4 -5	10	
5 -6	3	
6 -7	1	
7 -8	1	
Totals	48	50

Préparez vous à compter les *f*
sur la diapo suivante...

Thanks to Mike Kestemont for the idea!

Finished files are the result of
years of scientific study
combined with the experience of
many years.

How many f ?

Finished files are the result of
years of scientific study
combined with the experience of
many years.

L'expérience de Schindler

- Expérience proposée par Schindler, 1978.
- Il y a **6 f**, mais beaucoup de personnes en ratent un ou plusieurs, et répondent 4 ou 5, car ils oublient ceux qui sont dans les o.
- Illustre le peu d'attention que nous prêtions aux mots outils.

SCHINDLER, Robert M., The effect of prose context on visual search for letters. *Memory & Cognition*, 1978, vol. 6, no 2, p. 124-130.

Déjà noté par : **Healy, Alice F. (1976)**. Detection errors on the word the : Evidence for reading units larger than letters. *Journal of Experimental Psychology : Human Perception and Performance*, 2(2), 235.

An Internet meme in 2003

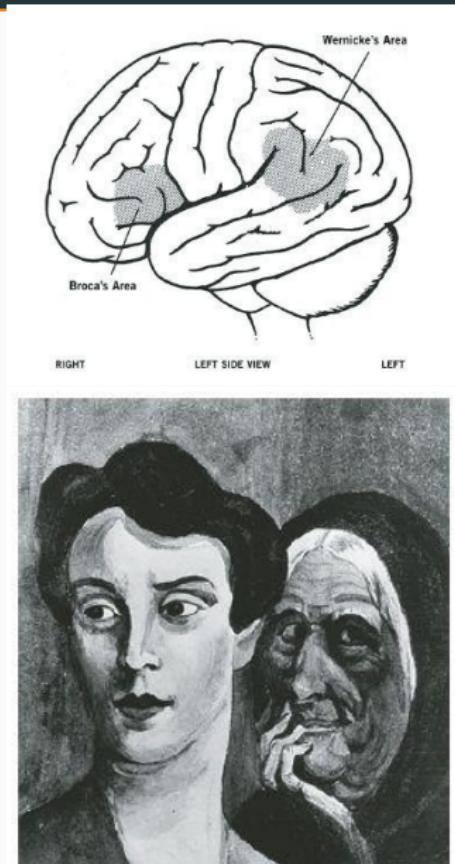
Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.

An Internet meme in 2003

Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it
deosn't mttaer in waht oredr the ltteers in a wrod are,
the olny iprmoetnt tihng is taht the frist and lsat ltteer
be at the rghit pclae. The rset can be a toatl mses and
you can sitll raed it wouthit porbelm. Tihs is bcuseae the
huamn mnid deos not raed ervey lteter by istlef, but the
wrod as a wlohe.

- toutes les lettres mélangées sauf la première et la dernière;
- beaucoup de mots préservés (longueur ≤ 3);
- surtout des mots-outils.
- Personne ne sait vraiment d'où sort ce meme;
- Cf. <https://www.mrc-cbu.cam.ac.uk/people/matt.davis/cmabridge/>;
- Graham Rawlinson, *The Significance of Letter Position in Word Recognition*, PhD Thesis (unpubl.), 1976, Nottingham University.

Perspective psycholinguistique : James Pennebaker, The Secret Life of Pronouns, 2011



Brocca's aphasia

Girl ... ummm ... woman ... ahh ... picture,
uhhh ... old. OK. Old woman.

Wernicke's aphasia

Well, right here is one of them and I think
she's next to that one. So if I see over
there you'll see her too. Let's see. I'm
thinking of her now. She's over there.

Cerveaux sociaux et profils psycholinguistiques

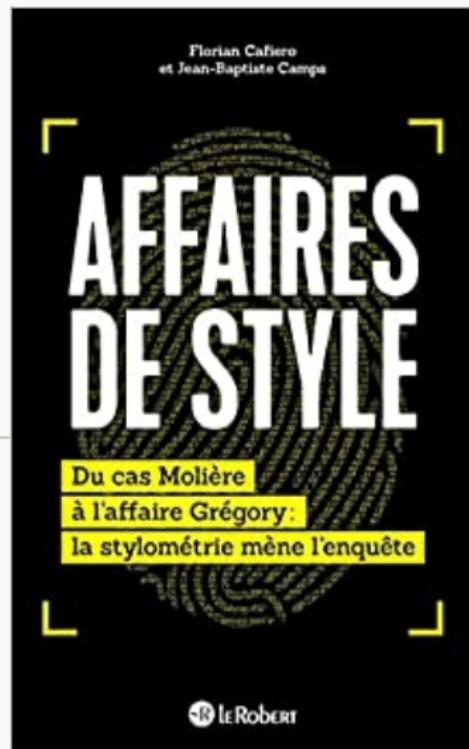
Table 2. Most important Style and Content features (by information gain) for each class of texts in each profiling problem.

Class	Style Features	Content Features
Female	personal pronoun, <i>I, me, him, my</i>	cute, love, boyfriend, mom, feel
Male	determiner, <i>the, of, preposition-matter, as</i>	system, software, game, based, site
Teens	<i>im, so, thats, dont, cant</i>	haha, school, lol, wanna, bored
Twenties	preposition, determiner, <i>of, the, in</i>	apartment, office, work, job, bar
Thirties+	preposition, <i>the, determiner, of, in</i>	years, wife, husband, daughter, children
Bulgarian	conjunction-extension, pronoun-interactant, however, <i>pronoun-conscious, and</i>	bulgaria, university, imagination, bulgarian, theoretical
Czech	personal pronoun, usually, <i>did, not, very</i>	czech, republic, able, care, started
French	indeed, conjunction-elaboration, <i>will, auxverb-future, auxverb-probability</i>	identity, europe, european, nation, gap
Russian	<i>can't, i, can, over, every</i>	russia, russian, crimes, moscow, crime
Spanish	determiner-specific, <i>this, going_to, because, although</i>	spain, restoration, comedy, related, hardcastle
Neurotic	<i>myself, subject pronoun, reflexive pronoun, preposition-behalf, pronoun-speaker</i>	put, feel, worry, says, hurt
Non-neurotic	<i>little, auxverbs-obligation, nonspecific determiner, up, preposition-agent</i>	reading, next, cool, tired, bed

Shlomo Argamon, Moshe Koppel, James W. Pennebaker, and Jonathan Schler,

« Automatically Profiling the Author of an Anonymous Text », Comm ACM 52-2 (2009),

Les affaires de style : un coup d'œil aux algorithmes



Plan

Sherlock Holmes, le style et la forme des oreilles

Brève introduction à la stylométrie

Mesurer le style ?

Style, idiolecte et stylome

Propriétés inconscientes du style

Les affaires de style : un coup d'œil aux algorithmes

La vièle et le stylomètre

Corneille et Molière

QAnon

Smells like stylometry spirit

Impostors and pseudonyms (with some word of ethics)

Stylometry at the École des chartes

Fundamental questions about the documents

Who wrote this document? When was it written? Where? Is it original? False? Altered? Co-written?

Which multiply for ancient periods

- prevalence of anonymity in medieval texts;
- quarrels of attribution.

Ex : Who wrote the lives of the troubadours ?

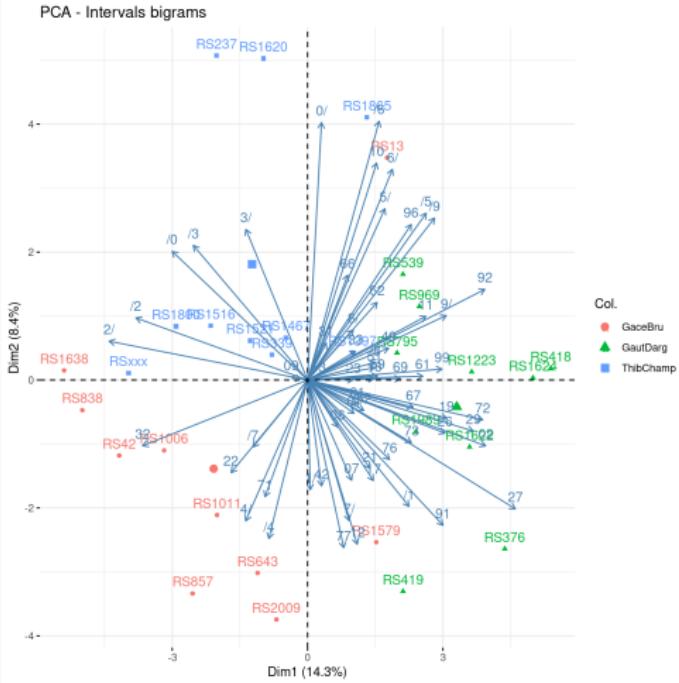
How were the great compilations of the lives of saints constituted ?

Did the trouvères compose the melodies of their songs ?



Uc de Saint-Circ, chansonnier; cf. Camps, 2009.

Trouvères texts and melodies



Camps, Chaillou-Amadieu et al., 2021.

Plan

Sherlock Holmes, le style et la forme des oreilles

Brève introduction à la stylométrie

Mesurer le style ?

Style, idiolecte et stylome

Propriétés inconscientes du style

Les affaires de style : un coup d'œil aux algorithmes

La vièle et le stylomètre

Corneille et Molière

QAnon

Smells like stylometry spirit

Impostors and pseudonyms (with some word of ethics)

The Corneille / Molière Case

Context : twilight of literary idols

In the 19th century, the emerging science of philology called into question major figures of Western culture : Homer, Shakespeare, the authors of the Bible...

Shakespeare and Molière

From the middle of the 19th century onwards, Shakespeare's authority over all or part of his plays was questioned.

1918 : Abel Lefranc, a famous chartist, publishes *Sous le masque de William Shakespeare : William Stanley*.

1919 : Pierre Louÿs, a poet in decline and literary forger, transposes it to Corneille and Molière.

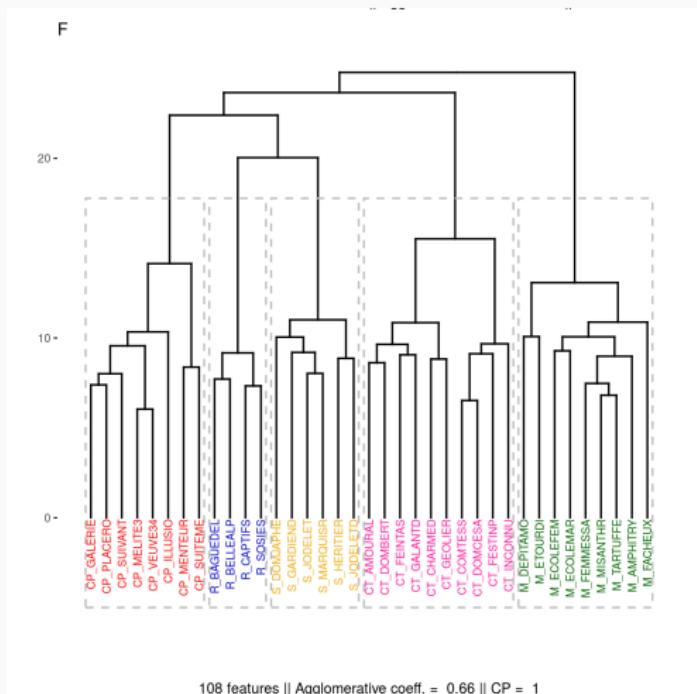
Resurfaces occasionally (revived by statistical work in the early 2000s).

The Corneille / Molière Case

Unsupervised analysis of :

1. lemmas,
2. rhymes,
3. words,
4. affixes,
5. grammatical categories,
6. function words.

The results are consistent :
pieces attributed to
Molière and Corneille
belong to two authors with
clearly different styles.

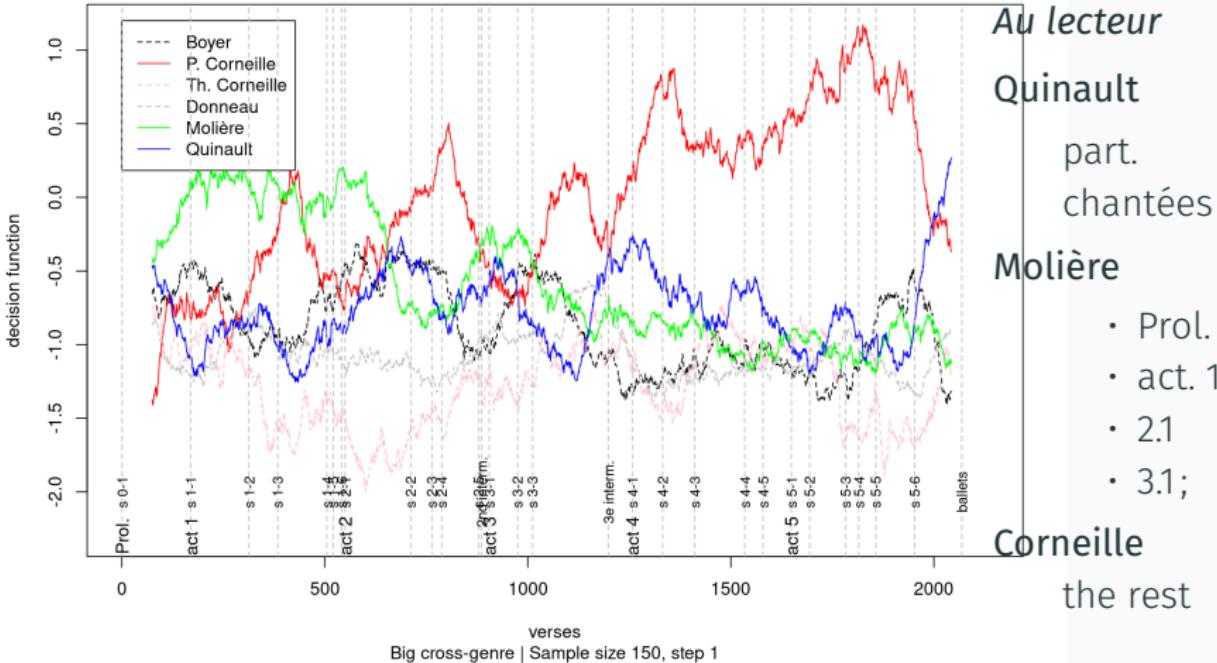


Psyche (1671), co-writing a blockbuster

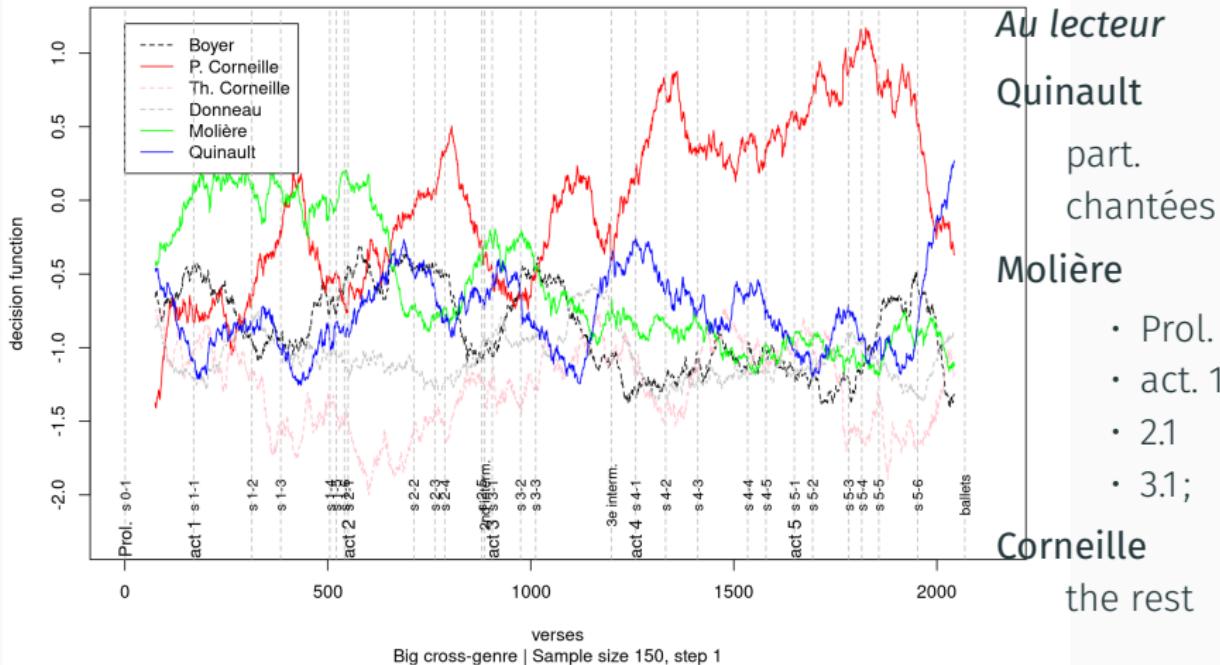
- competition opened by Louis XIV,
- for the “Salle des Machines” at the Tuileries (4000 spectators);
- subject and “pièce à machines’ (i.e., with special effects) to be produced in a very limited time (a fashionable genre in the years 1650-1670,
- commissioned by Louis XIV,
- ballets, music by Lully; Première at the Tuileries theatre (machinery, 4000 spectators.
- Molière (*Psyché*) beats Racine (*Orphée*), Quinault (*Proserpine*), ...

Cet ouvrage n'est pas tout d'une même main. M. Quinault a fait les paroles qui s'y chantent en musique, à la Réserve de la plainte italienne [de Lully]. M. de Molière a dressé le plan de la pièce, et réglé sa disposition, où il s'est plus attaché aux beautés et à la pompe du spectacle, qu'à l'exacte régularité. Quant à la versification, il n'a pas eu le loisir de la faire entière. Le carnaval approchait, et les ordres pressants du Roi, qui se voulait donner ce magnifique Divertissement plusieurs fois avant carême, l'ont mis dans la nécessité de souffrir un peu de secours. Ainsi il n'y a que le Prologue, le Premier

Rolling classification (with machine learning)



Rolling classification (with machine learning)



Conform to the declaration of the authors, except a single undeclared intervention from Corneille at the very end of the first act?

Plan

Sherlock Holmes, le style et la forme des oreilles

Brève introduction à la stylométrie

Mesurer le style ?

Style, idiolecte et stylome

Propriétés inconscientes du style

Les affaires de style : un coup d'œil aux algorithmes

La vièle et le stylomètre

Corneille et Molière

QAnon

Smells like stylometry spirit

Impostors and pseudonyms (with some word of ethics)

Who is Q?

The story in short

In October 2017, on a forum (4chan) in one of the dark and extreme places of the web

an anonymous account started posting cryptic “drops”

that ended up aggregating most conspiracy theories

and warned about a cabal of child-eating satanic elites inside the world and US governments (Obama, H. Clinton...)

and played role in the Capitol coup attempt.

Anonymous ID: B07V3bcM Sat 28 Oct 2017 16:44:28 No.147012719  Report

Quoted By: >>147017117 >>147017275 >>147036772 >>147039926.1 >>147039926.3

>>147005381

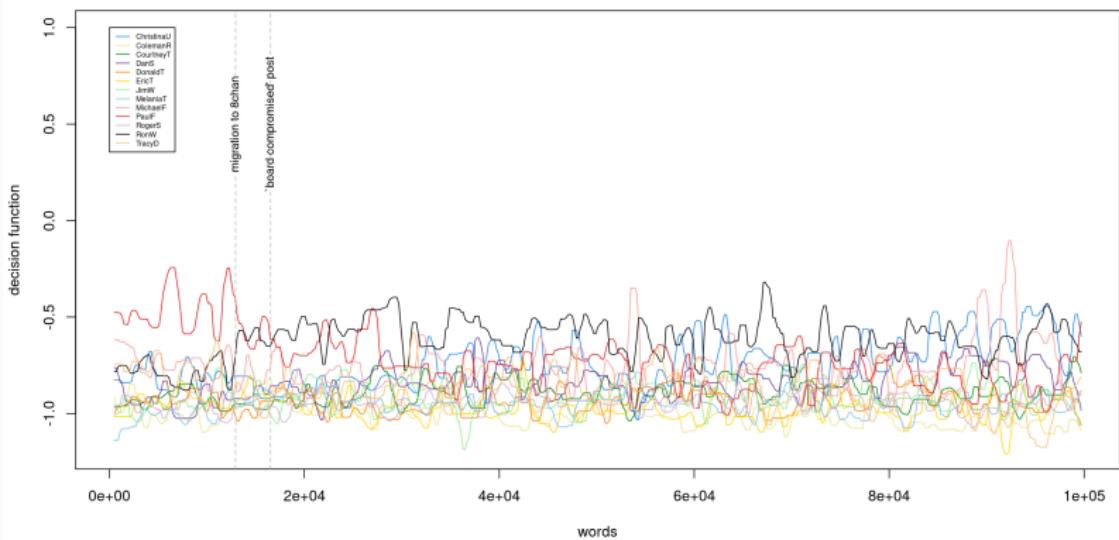
HRC extradition already in motion effective yesterday with several countries in case of cross border run. Passport approved to be flagged effective 10/30 @ 12:01am. Expect massive riots organized in defiance and others fleeing to US to occur. US M's will conduct the operation while NG activated. Proof check: Locate a NG member and ask if activated for duty 10/30 across most major cities.

Q is

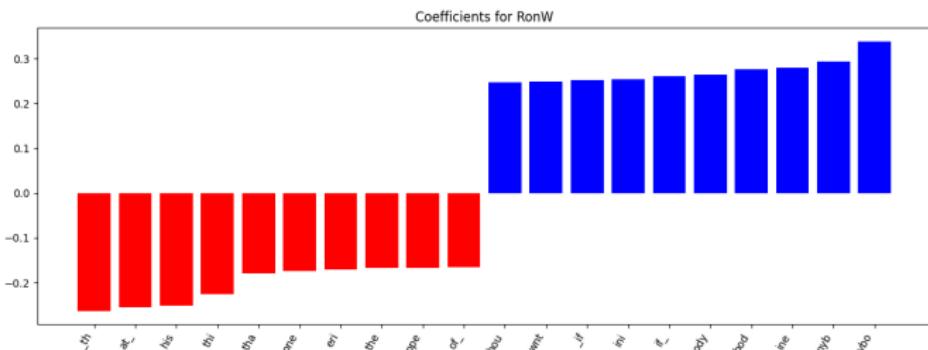
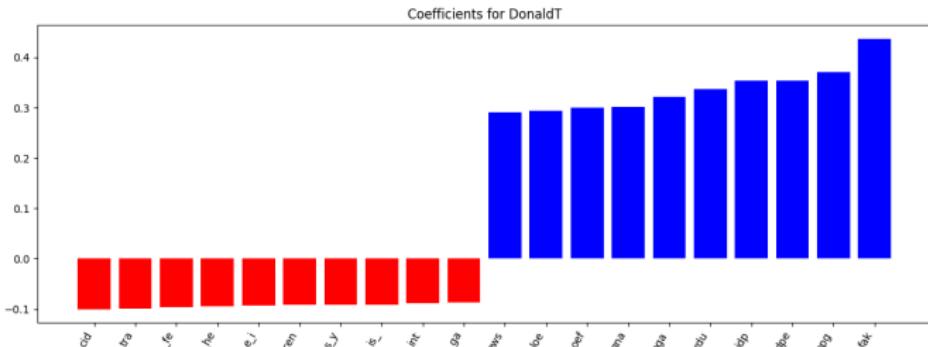
- posing as an insider exposing conspiracies of the deep state;
- thought by “true believers” to be (close to) Trump;
- thought by investigative journalists to be someone from the globalised alt-right online activism sphere (i.e., a geeky fascist in his garage).

But who is really Q?

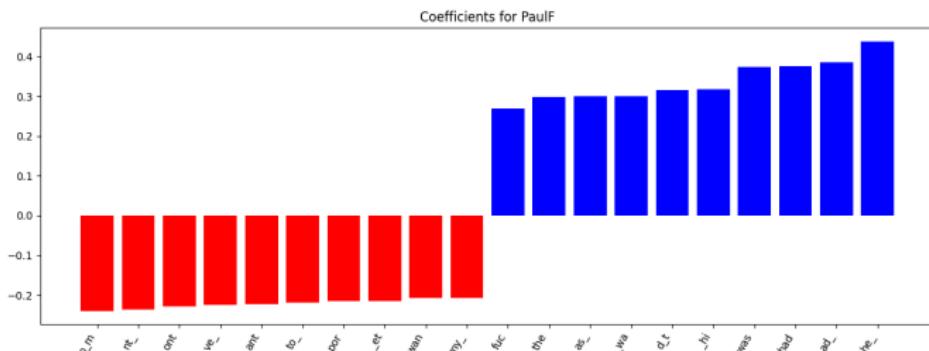
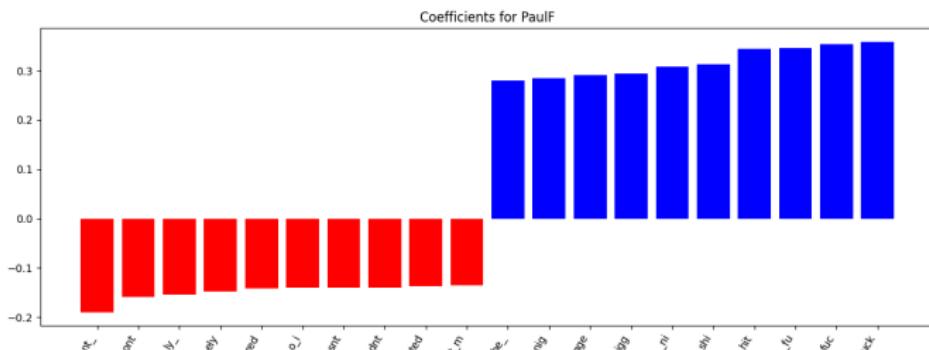
Who is Q? Rolling analysis



Who is Q? Insights on individual style and generic variation



Who is Q? Insights on individual style and generic variation



Plan

Sherlock Holmes, le style et la forme des oreilles

Brève introduction à la stylométrie

Mesurer le style ?

Style, idiolecte et stylome

Propriétés inconscientes du style

Les affaires de style : un coup d'œil aux algorithmes

La vièle et le stylomètre

Corneille et Molière

QAnon

Smells like stylometry spirit

Impostors and pseudonyms (with some word of ethics)

Detecting suicide notes

- Carole Chaski : training on 1000 sample letters of an AI detecting the authenticity of a suicide letter.
 - Good performances, even if not perfect.
 - Paradox : better performance on short notes.
 - Application : Kurt Cobain's suicide note

Detection of socio-demographic characteristics

Table 2. Most important Style and Content features (by information gain) for each class of texts in each profiling problem.

Class	Style Features	Content Features
Female	personal pronoun, I, me, him, my	cute, love, boyfriend, mom, feel
Male	determiner, the, of, preposition-matter, as	system, software, game, based, site
Teens	im, so, thats, dont, cant	haha, school, lol, wanna, bored
Twenties	preposition, determiner, of, the, in	apartment, office, work, job, bar
Thirties+	preposition, the, determiner, of, in	years, wife, husband, daughter, children
Bulgarian	conjunction-extension, pronoun-interactant, however, pronoun-conscious, and	bulgaria, university, imagination, bulgarian, theoretical
Czech	personal pronoun, usually, did, not, very	czech, republic, able, care, started
French	indeed, conjunction-elaboration, will, auxverb-future, auxverb-probability	identity, europe, european, nation, gap
Russian	can't, i, can, over, every	russia, russian, crimes, moscow, crime
Spanish	determiner-specific, this, going_to, because, although	spain, restoration, comedy, related, hardcastle
Neurotic	myself, subject pronoun, reflexive pronoun, preposition-behalf, pronoun-speaker	put, feel, worry, says, hurt
Non-neurotic	little, auxverbs-obligation, nonspecific determiner, up, preposition-agent	reading, next, cool, tired, bed

Shlomo Argamon, Moshe Koppel, James W. Pennebaker, and Jonathan Schler,
« Automatically Profiling the Author of an Anonymous Text », Comm ACM 52-2 (2009),
10.1145/1461928.1461959.

Plan

Sherlock Holmes, le style et la forme des oreilles

Brève introduction à la stylométrie

Mesurer le style ?

Style, idiolecte et stylome

Propriétés inconscientes du style

Les affaires de style : un coup d'œil aux algorithmes

La vièle et le stylomètre

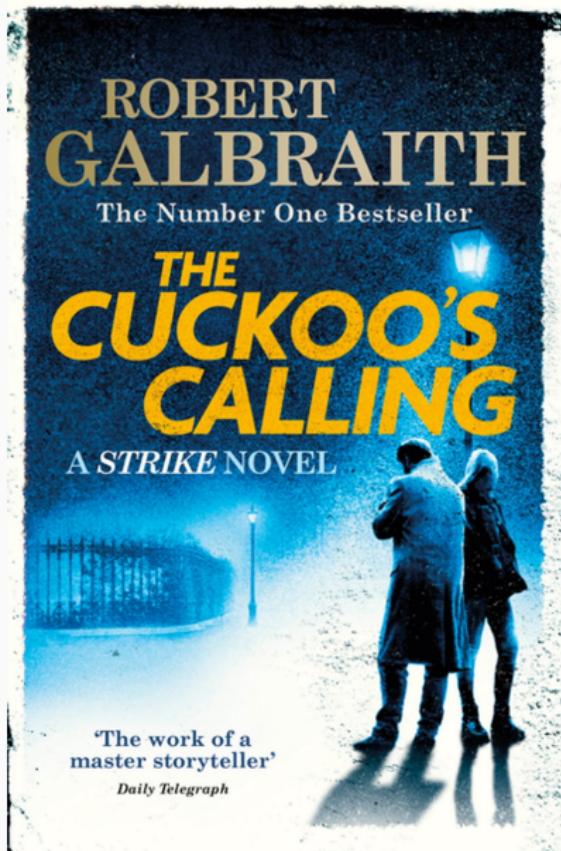
Corneille et Molière

QAnon

Smells like stylometry spirit

Impostors and pseudonyms (with some word of ethics)

The Robert Galbraith case

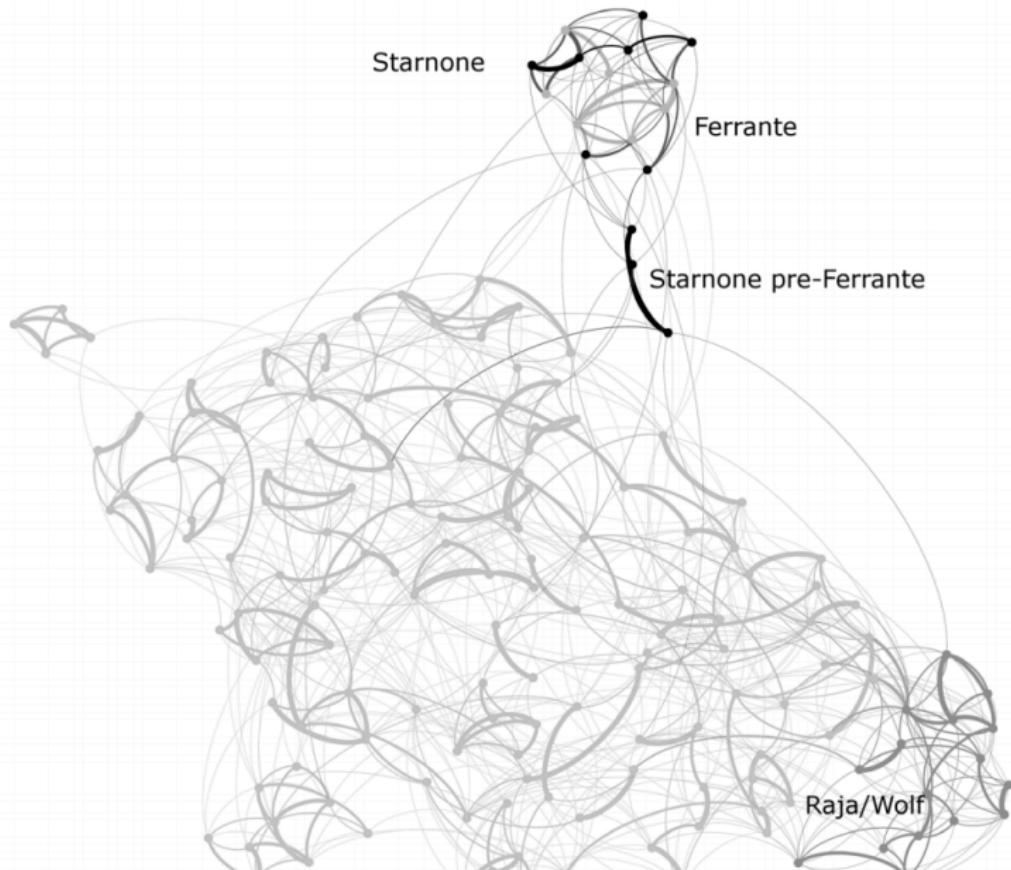


Unknown novelist
(military police veteran?)
makes a success on its first
novel...
but is it really the first?

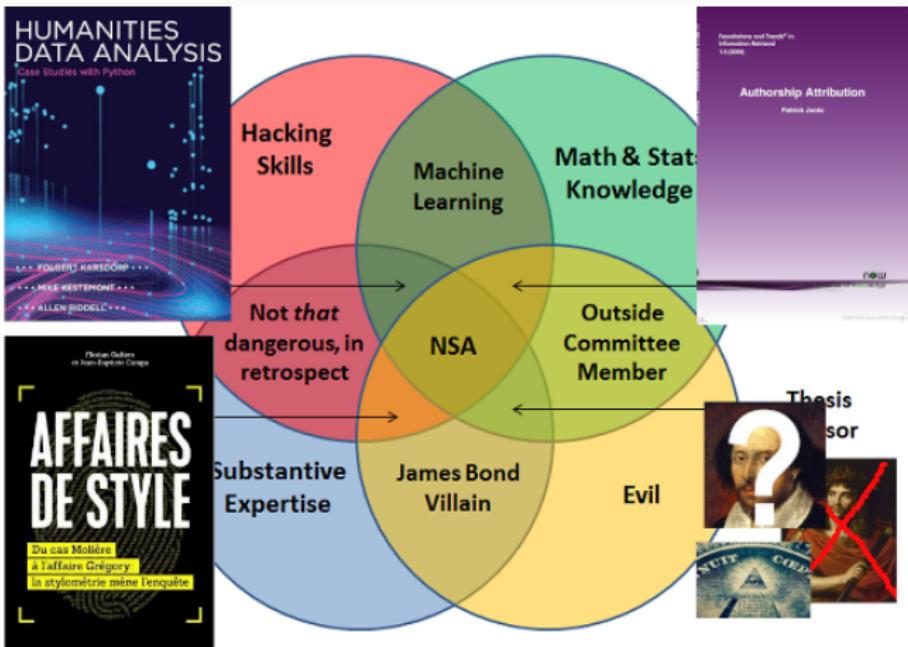
Clues

- 'experienced' writer
- women clothes
- editor

Who is Elena Ferrante?

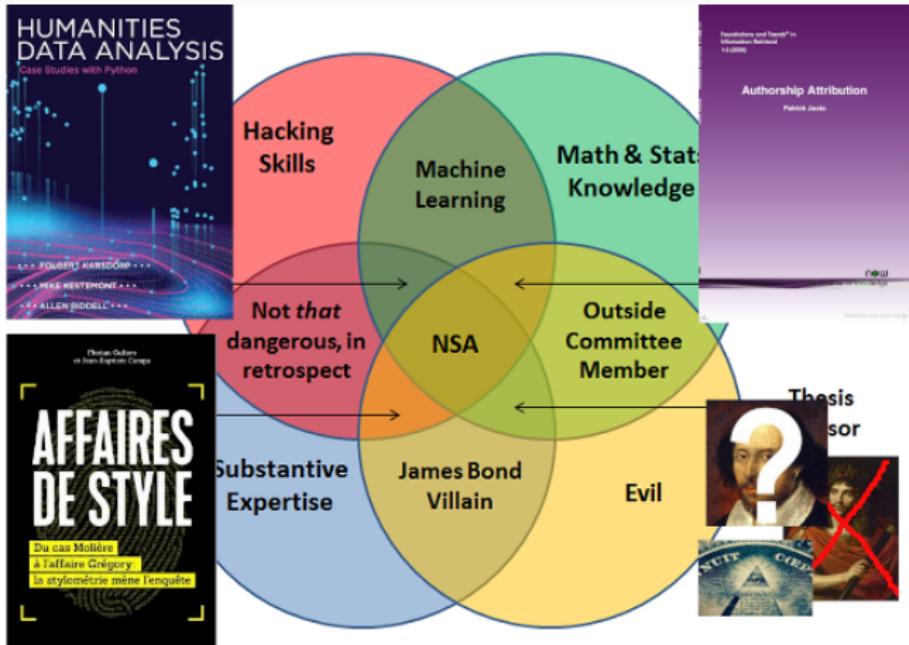


Quelques ouvrages



- F. Cafiero, J.B. Camps, *Affaires de style*, Paris, 2022.
- Patrick Juola, *Authorship Attribution*, Delft, 2008.
- F. Karsdorp, M. Kestemont, A. Riddel, *Humanities Data Analysis : Case Studies with Python*, 2021, <https://www.humanitiesdataanalysis.org/>.

Quelques ouvrages



Merci de votre attention!