

StilometrIA alla prova delle scritture collettive Da “Quaderni piacentini” e “Quindici” a ChatGPT

Marco De Cristofaro

Université de Mons
Université de Namur - C2W
marco.decrisofaro@umons.ac.be

Mariangela Giglio

Università di Bologna, Italia
mariangela.giglio2@unibo.it

Abstract*

La ricerca si propone di esplorare la capacità di GPT-4 di emulare lo stile di due riviste culturali italiane attive negli anni '60, «Quaderni Piacentini» e «Quindici». Basandosi su un corpus derivato dai primi numeri dei periodici, lo studio intende valutare se GPT-4 sia in grado di eludere l'analisi stilometrica generando un testo che rispecchi la strategia editoriale di una determinata rivista. Dopo la generazione di testi tramite GPT-4, è stato condotto un confronto stilometrico con il corpus originale. L'analisi con metodologie stilometriche tradizionali ha permesso di individuare i punti in cui le due riviste si distanziano e, di conseguenza, i nodi cruciali su cui il modello si concentra per la differenziazione stilistica e tematica. La ricerca intende aprire nuove prospettive sull'uso della stilometria per l'analisi computazionale di testi relativi a specifici contesti culturali; se fino ad ora, infatti, la comunità scientifica si è concentrata sulla capacità degli LLM di riprodurre fedelmente stili di diversi autori, l'esempio qui riportato dimostra la possibilità di indagini più ampie, che nel caso delle riviste implica un attraversamento delle strategie editoriali, delle tendenze di lettura e dei processi di circolazione delle idee.

611

Copyright © 2025 The Author(s)

The text in this work is licensed under the Creative Commons BY License.

<https://creativecommons.org/licenses/by/4.0/>

* Il presente contributo è stato suddiviso tra i due autori nel modo seguente: Marco De Cristofaro è responsabile dei paragrafi "Creazione del corpus di riferimento", "Strutture identitarie: cluster analysis di «Quaderni Piacentini» e «Quindici»", "Principi di riconoscibilità: Principal Component Analysis", "Lessici stranianti: un tentativo di zeta analysis sui testi delle riviste", "Scelte strutturali: una cluster analysis fondata sulle parts-of-speech"; Mariangela Giglio è responsabile dei paragrafi "OCR, accuratezza e rumore", "LLM e competenza emulativa", "Definizione del test-set tramite GPT-4", "Stilometria su testi AI generated". L'analisi complessiva, l'introduzione, la conclusione e la revisione finale del testo sono state realizzate congiuntamente da entrambi gli autori.

Parole chiave: Stilometria, Attribuzione di autorialità, Large Language Models, GPT-4, Riviste Culturali, AIUCD2024

The research aims to explore the ability of GPT-4 to emulate the style of two Italian cultural magazines active in the 1960s, «Quaderni Piacentini» and «Quindici». Using a corpus derived from the early issues of these magazines, the study assesses whether GPT-4 can bypass stylometric analysis by producing text that reflects the editorial strategy of a specific magazine. After using GPT-4 for generating emulative texts, a stylometric analysis was conducted to compare the AI generated texts with the original corpus. Comparison with traditional stylometric methodologies has allowed the identification of aspects where the two journals diverge, and consequently, the nodes on which the model focuses for stylistic and thematic differentiation. The research intends to open new applications on the use of stylometry for the computational analysis of texts related to specific cultural contexts; indeed, until now, the scientific community has focused on the ability of LLMs to faithfully reproduce the styles of different authors. Applying these methodologies to the field of magazines would allow for broader considerations on editorial strategies, reading trends, and the processes of idea circulation.

Keywords: Stylometry, Authorship attribution, Large Language Models, GPT-4, Magazines, AIUCD2024.

Introduzione

Le riviste letterarie, e più genericamente quelle culturali, hanno da sempre rappresentato per autori e intellettuali un terreno di confronto nel dibattito pubblico. La rivista si è progressivamente rivelata uno strumento fondamentale per ricostruire «le prese di posizione dei letterati, i loro trattati di alleanza e dichiarazioni di guerra» ([4]:23). La sua centralità nelle vesti di prisma interpretativo necessario a comprendere i rapporti di forza all'interno dello spazio culturale è confermata da un fortunato filone di studi che ottiene un primo riconoscimento a partire dagli anni sessanta del Novecento, quando Einaudi pubblica *La cultura italiana del '900 attraverso le riviste* [52], e prosegue nel corso dei decenni successivi grazie a una prolifica produzione scientifica ([40]; [1]; [17]; [39]). Questo campo di indagine ha avuto numerose applicazioni in ambito digitale con progetti di catalogazione e digitalizzazione, come la fondazione del Catalogo Italiano dei Periodici (ACNP); il Programma di digitalizzazione delle pubblicazioni periodiche¹ di ICCU e l'iniziativa CIRCE,² il Catalogo Informatico delle Riviste Culturali Europee ([4]:22). In un simile quadro, caratterizzato da una progressiva crescita di interesse, gli studi si sono soffermati sul ruolo delle riviste quali piattaforme di confronto tra intellettuali e sulla loro funzione di punto di incontro tra un gruppo culturale e una specifica comunità di lettori. Nel secondo Novecento, d'altro canto, i periodici assumono i connotati di vere e proprie "istituzioni", assumendo forme e caratteri diversi [26] capaci di offrire non solo un canale di accesso al campo di produzione culturale [12], ma anche uno spazio di partecipazione in costante movimento. Pur radicate in un luogo e in un gruppo ben definito, le riviste cercano di allargare il proprio spazio di influenza nel tentativo di ottenere rilevanza nazionale e, in alcuni casi, internazionale [26].

Negli anni Sessanta in Italia un ruolo di primo piano è stato ricoperto da due riviste che coniugano la dimensione culturale con precise ambizioni politiche: «Quaderni Piacentini», fondata da Piergiorgio

¹ <https://www.iccu.sbn.it/it/internet-culturale/storia-della-biblioteca-digitale-italiana-bdi/programma-di-digitalizzazione-delle-pubblicazioni-periodiche/> (ultima consultazione 28/01/2025)

² <https://r.unitn.it/it/lett/circe/circe> (ultima consultazione 01/01/2025)

Bellocchio e Grazia Cherchi nel 1962,³ e «Quindici», fondata a Roma nel 1967 dal Gruppo 63.⁴ Se diverse indagini sono state condotte sulla loro attività, sull'organizzazione redazionale, sulla loro posizione politico-culturale, ad oggi non risultano analisi computazionali dei testi pubblicati sulle pagine dei due periodici. Uno studio ad ampio raggio sullo stile degli articoli e sulle tematiche affrontate fornirebbe non solo un quadro più completo sui punti di vista dei due giornali, ma anche una più profonda comprensione delle rispettive strategie editoriali. Si tratta, in altre parole, di coniugare un'analisi dei fenomeni linguistici ricorrenti negli articoli con la traiettoria sociale e le ambizioni culturali dei loro contributori.

Il ricorso a metodologie stilometriche e a Large Language Models (LLM), come GPT-4, potrebbe così offrire un importante contributo alle indagini sulla storia culturale italiana degli anni Sessanta e alle ricerche più aggiornate sulla storia della lettura e della stampa.

A tal fine, nella prima parte dello studio viene dettagliato il procedimento di costruzione di uno specifico corpus di articoli estratti dalle due riviste prese in esame, attraverso modalità OCR e una normalizzazione e lemmatizzazione dei testi. Ad esso è stato affiancato successivamente, in un'ottica comparativa, un corpus di articoli *AI-generated* che funge da test set dell'indagine complessiva.

Una volta ottenuta la banca dati testuale completa di scritti estratti dalle riviste originali e scritti elaborati da GPT-4, è stata condotta un'analisi stilometrica ricorrendo a più approcci metodologici. In un primo momento, lo studio si sofferma sul confronto tra i testi di «Quaderni piacentini» e quelli di «Quindici». In questo caso, la Delta Analysis consente di verificare l'effettiva riconoscibilità delle riviste o di specifici articoli pubblicati su di esse. Il successivo approfondimento comparativo tra le due sedi editoriali si è rivolto a un'indagine lessicale e semantica condotta attraverso la Principal Component Analysis e la Zeta Analysis. Il confronto ha permesso di osservare la presenza di eventuali scelte tematiche distintive.

Nella terza fase la comparazione ha coinvolto anche i testi *AI-generated*. Qui, all'iniziale delta analysis, volta a inquadrare a livello complessivo il rapporto tra testi prodotti da umani e da AI, ha fatto seguito una nuova indagine lessicale, necessaria ad approfondire il grado di separazione e i diversi fattori di differenziazione tra riviste e GPT.

Il lavoro si propone di esplorare e valutare le capacità della stilometria nell'individuazione di specificità stilistiche e tematiche di ciascuna rivista. Allo stesso tempo, lo studio riflette sulle performance di GPT-4 applicate – più che sullo stile di singoli autori – su un intero corpo editoriale. L'indagine parte dal quadro teorico dello studio delle riviste culturali italiane [11] e dalla traccia metodologica offerta dal lavoro di Simone Reborà [47], che ha esaminato la capacità

³ Sui «Quaderni Piacentini» si vedano almeno le due antologie: «*Quaderni piacentini*» 1962-1968, a cura di L. Baranelli e G. Cherchi, Gulliver, Milano, 1977, e *Prima e dopo il '68*, a cura di G. Fofi e V. Giacobini, Minimum fax, Roma, 1998. Per una ricostruzione storica delle vicende della rivista si rimanda a: G. Pontremoli, *I “Piacentini”. Storia di una rivista (1962-1980)*, Edizioni dell'asino, Roma, 2017; a G. Muraca, *Cronistoria dei “Quaderni piacentini”*, in *Da Il Politecnico a “Linea D'ombra”*, Lalli, Poggibonsi, 1990, 51-95 e a Id., *Piergiorgio Bellocchio e i suoi amici. Intellettuali e riviste della sinistra eterodossa*, Ombre corte, Verona, 2018.

⁴ Per le coordinate storiche di «Quindici» si rimanda all'antologia *Quindici: una rivista e il Sessantotto*, a cura di N. Balestrini, Feltrinelli, Milano, 2008 e a F. Bortolotto e D. Paone, *Una crepa nel sistema: dalla crisi di «Quindici» alla ricostruzione di «Alfabeta»*, in F. Bortolotto, E. Fuochi, D. A. Paone, F. Parodi, *Sistema periodico. Il secolo interminabile delle riviste*, Pendragon, Bologna, 2018.

dei modelli linguistici di grande scala, come GPT-3, di eludere gli approcci stilometrici nell'attribuzione dell'autore. Qui si intende, al contrario, testare le capacità stilometriche su testi generati da LLM nel tentativo di replicare la somma stilistica dei periodici culturali.

Creazione del corpus di riferimento

Una prima complessità dello studio riguarda la scelta del corpus di riferimento. Le riviste italiane del Novecento, infatti, sono tanto numerose quanto variegate e condividono spesso scopi politico-culturali, forme espressive (racconti, reportage, articoli di critica ecc.) e approcci (commistione tra testi e immagini, traduzioni, coinvolgimento dei lettori). Occorre verificare, dunque, se sia possibile delimitare linee di demarcazione tra le singole riviste. Data la diversità degli autori, inoltre, bisogna indagare l'esistenza di elementi stilistici ricorrenti influenzati dalle sedi editoriali di pubblicazione. È chiaro, infatti, che l'ambizione delle riviste nel periodo storico individuato consiste nel proporsi quale spazio di discussione vivo e attento soprattutto alla realtà circostante. Una simile impostazione influenza l'organizzazione e la struttura testuale degli articoli in particolar modo dei numeri iniziali delle riviste, su cui, in genere, è possibile riscontrare un maggior livello di guardia da parte delle redazioni. Il desiderio di intervenire su questioni attuali e al centro del dibattito pubblico si evidenzia in una prosa lineare, in cui l'ambizione alla chiarezza espositiva è sottolineata dal ricorso più frequente alla paratassi e dalla tendenza a esplicitare terminologie, metodi e concetti che richiedono una conoscenza specifica del contesto culturale. Non è raro, d'altro canto, che i testi seguano un'alternanza di domande e risposte, rendendo il tono spesso assertivo e la sintassi del ragionamento più frammentata. Gli espedienti stilistici più complessi coinvolgono specifici ambiti di indagine, dalla critica letteraria alle scienze umane che, tuttavia, professano in quegli anni la ricerca di una prosa lineare e di un lessico ampiamente accessibile a un pubblico ampio e variegato per formazione e competenze.

Per questa indagine, sono state selezionate le riviste «Quaderni Piacentini» (QP) e «Quindici» (QU): entrambe hanno condiviso l'ambizione a considerarsi quale spazio sia culturale che politico, come emerge fin dalle dichiarazioni programmatiche dei primi numeri. La seconda ragione alla base della scelta è stata l'attenzione rivolta alla cura stilistica dei testi, posta in stretta correlazione con i rispettivi lettori di riferimento. Ora, definire uno specifico orizzonte d'attesa delle riviste risulta problematico a causa di un insieme di fattori di carattere ideologico ma anche, e soprattutto, organizzativo che ha determinato la loro distribuzione, spesso discontinua in termini di copie vendute. Risulta altresì complicato, come evidenziano i limiti della teoria della ricezione ([30]; [15]; [16]), ricostruire il rapporto tra gli autori della redazione di una singola rivista e il lettore ideale a cui ciascuno di essi intende rivolgersi. Nel tentativo di individuare un lettore probabile occorre dunque affidarsi allo studio della «storia interna» ([29]:98) dei singoli testi, facendo riferimento al contempo alle dichiarazioni esplicite degli autori e all'effettivo risultato formale. La «lettura battagliera e militante» ([64]:177) di «Quaderni piacentini» si concretizza con una visione ad ampio raggio capace di accogliere contributi interdisciplinari che spaziano dalle inchieste ai testi contenenti le rivendicazioni di scioperi nazionali e locali, da interventi poetici a scritti di critica letteraria, dalle memorie di critici militanti a intermezzi teatrali. La varietà strutturale degli articoli trova un minimo comune denominatore nella vicinanza tematica confermata dalla «larghissima diffusione» della rivista presso un pubblico specifico: «i militanti del nascente movimento studentesco» ([64]:178). Uno sviluppo analogo si riscontra nella traiettoria di «Quindici» che, nata su istanze sovversive di revisione della letteratura, trova a sua volta nel movimento studentesco uno spazio ideale di diffusione ([8]:336).

Un ulteriore aspetto di interesse è la presunta libertà espressiva affidata a ciascun collaboratore, che pone tuttavia in essere il rischio principale dell'analisi. La criticità maggiore dell'ipotesi di partenza, infatti, è la presenza di articoli scritti da autori diversi che potrebbe impedire il riconoscimento dei testi come raccolte rappresentative di un unico giornale. «Lo strumento-rivista» non può essere considerato «come il mero risultato della somma dei suoi partecipanti», dato che «le redazioni si organizzano come un nucleo più o meno ristretto di intellettuali, autori e critici che non rinunciano alla loro identità intellettuale di singoli per sposarne *in toto* una collettiva» ([10]:189).

Ancora una volta, in entrambi i casi di studio, è stata individuata una specifica dichiarazione programmatica. L'editoriale di apertura del primo numero di «Quaderni Piacentini» presenta la rivista come «un foglio di battaglia, portata non solo all'esterno ma anche all'interno» (QP 1962), lanciando un appello alla collaborazione. Pur connotata in senso ideologico, «Quaderni Piacentini» dichiara il proprio sforzo volto ad allargare lo spettro del dibattito e le modalità di espressione. Ciononostante, la radicalizzazione del dibattito politico e culturale nel decennio successivo e l'assimilazione del giornale a uno specifico fronte militante saranno le principali cause di una perdita di quella varietà espressiva e vivacità interdisciplinare degli esordi ([64]:178-179).

In «Quindici» la stessa impostazione emerge con maggiore insistenza. La rivista professa la libertà espressiva come connotato distintivo e la difesa dell'autore, nelle vesti di individuo singolo ben riconoscibile nella struttura dei suoi testi, è l'elemento che appare maggiormente in contrasto con l'ipotesi alla base dell'analisi stilometrica.

L'ultimo criterio adottato è la prossimità temporale: è vero, infatti, che il primo numero di «Quaderni Piacentini» esce nel 1962 e quello di «Quindici» nel 1967, ma entrambe le riviste affrontano tematiche riconosciute come rappresentative del contesto sociale italiano degli anni Sessanta prima del 1968. Nel corso degli anni Cinquanta la dominante culturale era stata la riflessione sulla stilistica e l'analisi testuale con la pubblicazione in Italia di due testi fondamentali come *Critica stilistica e storia della lingua* di Spitzer (1954) e *Mimesis* di Auerbach. Nel 1960 Leric pubblica in italiano *Il grado zero della scrittura* di Roland Barthes che sancisce l'ascesa dello strutturalismo e, successivamente, delle scienze umane in Italia ([63]:96). I principi fondativi del Gruppo '63 confermano la volontà di rinnovamento della critica e trovano supporto nell'evoluzione del mercato editoriale. L'editoria libraria italiana inizia a guardare con favore alle emergenti scienze umane ([44]:208-36). Allo stesso tempo, il progressivo successo planetario e mediatico degli autori che ne erano stati i principali promotori [62] allarga il perimetro del dibattito intellettuale. In un contesto di confronto tra metodi di indagine diversi, la dinamicità delle riviste si rivela uno strumento privilegiato, suscitando non solo l'interesse degli intellettuali ma anche l'attenzione degli studiosi che vedono nella dimensione di immediatezza dei fogli culturali un'occasione fondamentale per decifrare la realtà circostante ([4]:22-23).

La fertile stagione della critica letteraria – e in particolare dell'analisi testuale – e la somiglianza delle tematiche avrebbe garantito una maggiore centralità dello studio prettamente stilistico quale tratto dirimente della polarizzazione tra le due riviste: l'indagine stilometrica, in altre parole, sarebbe stata maggiormente valida, perché basata su aspetti formali e non influenzata da un'eccessiva distanza degli argomenti trattati.

OCR, accuratezza e rumore

Una volta identificati i testi di riferimento si è provveduto alla trascrizione: per «Quaderni Piacentini» la selezione degli articoli è basata su uno spoglio delle riviste reperibili integralmente sul sito della Biblioteca Gino Bianco;⁵ per «Quindici» la scelta è ricaduta sul volume curato da Nanni Balestrini [5]. Sono stati dunque trascritti nell'interezza i primi tre numeri di ciascuna rivista per un totale di circa 25.000 parole per «Quindici» e 22.000 circa per «Quaderni Piacentini».⁶ Per garantire livelli di accuratezza quanto più possibile elevati è stato scelto di combinare tecniche OCR e forme di controllo manuale, procedendo alla trascrizione dai fascicoli originari. Per la fase di riconoscimento OCR sono stati utilizzati congiuntamente due strumenti diversi: la piattaforma *Transkribus* e lo *Snipping Tool* sviluppato per Windows 11. Il risultato della trascrizione OCR *raw* (senza revisione umana) è stato di un tasso di accuratezza di circa il 99% (un errore ogni 101 caratteri). A seguito della correzione il corpus di riferimento non presenta errori immediatamente rilevabili.

Come ricordato da Holley (2009) l'accuratezza media dell'OCR senza correzione può variare fra il 71% e il 98,02%. Al 71% si troverebbero 145 caratteri erronei in un paragrafo medio di 500 caratteri. Visto da un'altra prospettiva, ciò significa che il 29% del paragrafo sarebbe scorretto: ne risultano difficoltà per la corretta lettura e interrogazione. La soglia minima di accuratezza consigliata dalla DFG (Deutsche Forschungsgemeinschaft)⁷ per le applicazioni di ricerca è di almeno il 99,95%. La trascrizione a quattro mani garantisce un'accuratezza del 99,997% (Haaf et al. 2013); in questo caso si è optato per una combinazione di testo OCR e di revisione manuale che permette di raggiungere percentuali di accuratezza paragonabili a quelle della trascrizione a quattro mani in quanto rappresenta un processo di doppia correzione del risultato della scansione OCR [33].

Sebbene la presenza di un corpus accurato sia sempre auspicabile, ricerche mostrano come la preparazione dei dati possa prendere fino all'80% del tempo di preparazione di una ricerca ([46]; [60]). È lecito dunque domandarsi quanto la presenza di rumore all'interno del proprio dataset possa effettivamente impattare sui risultati definitivi. Nel tentativo di ridurre il pre-processamento dei dati, i ricercatori hanno testato la tolleranza algoritmica al rumore [23]. Sorprendentemente, nel campo della stilometria, parrebbe che la presenza di testi rumorosi non impatti significativamente le performance di attribuzione autoriale.

Eder ([18]:612) osserva che, sebbene la tolleranza al rumore vari significativamente tra diverse lingue, anche un 20% di danno nel corpus non impatta in modo significativo il processo di corretta attribuzione. D'altro canto, numerosi approcci contemporanei all'attribuzione computazionale dell'autore si focalizzano su caratteristiche locali, come le frequenze di n-grammi di parole o n-grammi di caratteri ([34]; [55]; [32]). Tali caratteristiche tendono ad essere ampiamente diffuse e distribuite in maniera omogenea all'interno dei testi, risultando meno sparse rispetto, ad esempio, alle parole di contenuto. Questa distribuzione contribuisce a spiegare la robustezza di tali categorie

⁵ Il progetto di digitalizzazione della Biblioteca Gino Bianco è stato portato avanti dalla Fondazione Alfred Lewin in collaborazione con la Fondazione Anna Kuliscioff: <https://www.bibliotecaginobianco.it/?m=1&p=0&t=home>.

⁶ Trattandosi di un lavoro che rientra in una ricerca più ampia il corpus revisionato, unitamente agli script utilizzati, è disponibile all'interno del seguente repository GitHub: <https://github.com/mariangiglio/GPT-4-Stilometry/tree/main>

⁷ La DFG (Deutsche Forschungsgemeinschaft) è l'organizzazione centrale autonoma per il finanziamento della ricerca scientifica in Germania. <https://www.dfg.de/en/about-us/about-the-dfg/what-is-the-dfg> (ultima consultazione 01/01/2025)

nei confronti dell'introduzione di rumore stocastico di elevata intensità. Nelle analisi stilometriche, al contrario, un fattore di affidabilità dei risultati è la dimensione del corpus scelto che, stando alle ultime ricerche di Eder [20] è stabilita in una dimensione minima accettabile di almeno 2000 parole per testi in cui l'impronta autoriale è molto forte (p. 223). Ne consegue che, sebbene i campioni di grandi dimensioni siano preferibili, anche i campioni di piccole dimensioni possono, in alcuni casi, produrre risultati accurati ([19]; [20]; [48]).

LLM e competenza emulativa

ChatGPT è un chatbot rilasciato da OpenAI il 30 novembre 2022 che usa tecniche di NLP per generare dialoghi con l'utente su un'ampia varietà di argomenti. Basato sulla tecnologia *Generative Pre-trained Transformer*, il sistema è in grado di generare testi simili a quelli umani in un'ampia varietà di stili e di lingue, con una accuratezza tale che la maggior parte delle persone non riesce a distinguere tra testi generati da AI e quelli scritti da esseri umani ([35]; [57]). Anche l'ambito accademico è coinvolto da questa *chatbot invasion* [56], che si manifesterebbe nell'uso sproporzionato di alcune *keywords* distintive. Secondo i recenti studi di Gray [25]:

For the publishing year 2023, it is found that several of those keywords show a distinctive and disproportionate increase in their prevalence, individually and in combination. It is estimated that at least 60,000 papers (slightly over 1% of all articles) were LLM-assisted.

Si tratta di percentuali difficilmente considerabili come realmente significative, ma che potrebbero essere riviste al rialzo negli anni a venire, soprattutto in considerazione della crescente popolarità che ChatGPT e altri chatbot stanno godendo. Ancor più preoccupante sembrerebbe l'infiltrazione dell'AI nell'ambito delle *peer review* che, stando agli studi di Liang et al. [37], potrebbero essere interessate da modifiche tramite AI in una percentuale che varia dal 6.5% al 16.9%.

Al di là di queste considerazioni – e delle ripercussioni sull'ambito accademico – è stata ampiamente discussa la difficoltà di identificare con certezza testi scritti da AI, tanto che OpenAI il 31 gennaio 2023 ha rilasciato un classificatore per distinguere i testi *AI generated*; è stata la stessa OpenAI, tuttavia, ad ammetterne i limiti: lo strumento presenta un elevato tasso di falsi positivi, prestazioni inadeguate per testi brevi e, soprattutto, una bassa capacità di rilevamento per le lingue diverse dall'inglese [38]. Sembra particolarmente interessante, dunque, valutare l'effettiva capacità emulativa di GPT, non del linguaggio umano in senso ampio⁸ né in relazione al singolo autore (come già esplorato da Rebora [47]) ma anche in attività più complesse come la riproduzione di articoli di una rivista.

La scelta del modello GPT-4 è stata guidata da una combinazione di fattori qualitativi e pratici. In primo luogo, la diffusa popolarità del modello e la sua pervasività nell'ecosistema della generazione automatica del linguaggio ne fanno un punto di riferimento imprescindibile per qualunque indagine comparativa sull'output stilistico prodotto da LLM. A rafforzare questa scelta concorrono le valutazioni ottenute nei benchmark indipendenti, come quelli curati dalla community di Hugging Face e le performance registrate in contesti di testing comparativo come

⁸ Riguardo la valutazione della capacità degli LLM di emulare caratteristiche proprie del linguaggio umano si veda almeno Mitchell e Krakauer [41] e il recentissimo studio di Jones e Bergen [31].

LiveBench, che evidenziano le sue potenzialità soprattutto nei compiti legati alla coerenza stilistica, alla comprensione contestuale e alla continuità discorsiva.⁹

Infine, un ulteriore elemento decisivo è stato di natura funzionale: l'interfaccia conversazionale di GPT-4, così come implementata attraverso piattaforme accessibili al pubblico, consente l'integrazione diretta di blocchi testuali e materiali di input, facilitando in modo significativo la produzione di testi su misura e contestualmente mirati. Questo aspetto è particolarmente rilevante nell'ambito della presente analisi, dove la possibilità di condizionare il modello con prompt estesi e documenti di riferimento ha permesso di generare sezioni testuali più aderenti allo stile desiderato, pur mantenendo la spontaneità tipica della generazione automatica.

Definizione del test-set tramite GPT-4

La definizione del test set ha rappresentato una fase fondamentale per valutare l'abilità emulativa di GPT-4. Il corpus delle riviste è stato suddiviso in blocchi di testo, ciascuno contenente un minimo di 5.000 parole. Si presenta qui di seguito una ricostruzione dei dati complessivi relativi al corpus di riferimento.

Il corpus è composto da 10 documenti, per un totale di 56.301 token (parole o unità lessicali) e 11.767 types (forme uniche), con un Type/Token Ratio (TTR) globale pari a 0,209.

Questo valore, già in sé moderatamente basso, è il risultato non solo della naturale tendenza del TTR aggregato a ridursi per effetto delle ripetizioni inter-documentali, ma anche del contributo diretto di testi con TTR individuali molto bassi, in particolare quelli generati dall'intelligenza artificiale.

I documenti AIQP e AIQU, infatti, presentano TTR rispettivamente di 0,237 e 0,294, le proporzioni più basse del corpus, suggerendo un lessico più ripetitivo e contenuto. Questi valori contribuiscono in modo diretto all'abbassamento della varietà lessicale complessiva e spiegano la discrepanza tra il TTR globale (0,209) e la media dei TTR individuali, che si attesta a 0,367.

Questa media più elevata è coerente con un comportamento atteso: calcolando il TTR per ciascun testo si limita l'effetto delle ripetizioni inter-documentali, permettendo di rilevare più chiaramente le variazioni stilistiche. Si osserva infatti una certa eterogeneità interna: documenti come QU2 (0,435) o QP3 (0,417) evidenziano una maggiore varietà lessicale, in netto contrasto con la ridondanza dei testi generati da AI.

Tali differenze suggeriscono che i testi prodotti da modelli generativi, pur risultando coerenti e strutturalmente comparabili in termini di lunghezza, tendano a impiegare un vocabolario più ristretto. Al contrario, la scrittura umana appare più articolata e diversificata dal punto di vista lessicale. Questi dati offrono un primo indizio quantitativo di differenze stilometriche sostanziali tra le due modalità di produzione testuale, ponendo le basi per analisi più approfondite.

⁹ Classificazione fornita dalla LLM Leaderboard di HuggingFace, raggiungibile a <https://huggingface.co/docs/leaderboards/index> (ultima consultazione (01/11/2024))

Nome file	Tokens	Types	TTR	Media Token	Minimo Token	Massimo Token	Token SD	Media TTR
AIQP	5712	1356	0.237	NA	NA	NA	NA	NA
AIQU	5493	1615	0.294	NA	NA	NA	NA	NA
QP1	6299	2283	0.362	NA	NA	NA	NA	NA
QP2	5042	2011	0.399	NA	NA	NA	NA	NA
QP3	4983	2079	0.417	NA	NA	NA	NA	NA
QP4	5276	1941	0.368	NA	NA	NA	NA	NA
QU1	6067	2148	0.354	NA	NA	NA	NA	NA
QU2	4849	2109	0.435	NA	NA	NA	NA	NA
QU3	6732	2695	0.4	NA	NA	NA	NA	NA
QU4	5848	2380	0.407	NA	NA	NA	NA	NA
CORP US	56301	11767	0.209	5630.1	4849	6732	616.9	0.367

Anche se i recenti studi di Eder [20] suggeriscono che 2.000 parole possano essere sufficienti per analisi stilometriche efficaci, in questo caso abbiamo optato per una soglia superiore. Questa scelta è giustificata dal fatto che gli scritti analizzati, essendo composti, presentano un'impronta autoriale meno marcata, rendendo così necessario un corpus più ampio per garantire analisi significative e affidabili. I blocchi testuali così generati sono stati allegati nell'interfaccia di chat GPT-4 con l'obiettivo di generare articoli che non solo rispecchiassero lo stile, ma anche le tematiche intrinseche dei materiali originali. Questo processo si è tuttavia rivelato non privo di sfide. Le limitazioni intrinseche di GPT-4, in particolare per quanto riguarda il limite di lunghezza dei testi generabili, hanno imposto restrizioni significative; è inoltre emersa una notevole variabilità nei risultati ottenuti utilizzando lo stesso prompt in sessioni di chat differenti. Per quanto si sia cercato di non influenzare eccessivamente l'elaborazione da parte di GPT, è stato necessario modificare leggermente i parametri per impedire sia la ripetizione esatta dei risultati, sia la generazione di testi poco coerenti.

In una prima fase si è testato GPT tramite API, operando principalmente e in maniera alternata sulla temperatura¹⁰ e top p,¹¹ regolando così il bilanciamento tra creatività e coerenza nei testi generati.

Per il primo test-set è stata impostata una temperatura di 0.8. Questo valore, tuttavia, ha generato testi eccessivamente simili tra loro, poco adatti ad analisi stilometriche. Dopo una serie di tentativi, si è deciso di adottare la temperatura consigliata di 1.0, la stessa utilizzata nell'interfaccia online di GPT 4 sulla quale si è poi proseguito il lavoro. Per quanto riguarda il top p, invece, si è scelto di rimanere nei valori consigliati tra 0.7 e 0.9. Valori più bassi (ad esempio, intorno a 0.5) hanno prodotto risposte apparentemente più interessanti ma che, secondo il nostro giudizio, risultavano eccessivamente vicine alla mera riformulazione del testo proposto in partenza. Sarebbe stato utile poter verificare come il modificarsi dei parametri potesse influenzare le analisi stilometriche ma, a causa della corposità dei testi richiesti, questo non è stato possibile. In generale si è tentato comunque di minimizzare l'influenza dell'aspetto umano nell'analisi stilometrica, garantendo così una maggiore oggettività nei risultati ottenuti. Lo stesso criterio è stato mantenuto nella stesura del prompt che bisognava rendere il più generico e imparziale possibile.

Il prompt scelto per la generazione di articoli è stato il seguente:

I submitted to you a word file with a few italian texts. Write a text in italian of 1000 words in the style of the texts I submitted. Try to use the same stylistic features, to address similar topics, take the same political and cultural positions taken in the articles you analyzed. Try to emulate the articles and mirror the style and stylistic features as much as possible. language of the results "Italian" length of the results: "1000 words" minimum.

Sebbene l'incapacità di GPT di contare (non soltanto i token) sia già stata ampiamente appurata,¹² si è scelto di mantenere un'indicazione di massima all'interno del prompt. La generazione testuale è stata ripetuta fino al raggiungimento della lunghezza desiderata. I testi così ottenuti sono stati poi uniti in un singolo file di circa 5000 parole l'uno.

La prima considerazione riguarda la scelta del prompt: in questo task specifico si è optato per un approccio *Zero-Shot Prompting*, in quanto a seguito di approcci *Few-Shots* è emersa una tendenza a contaminare i risultati della generazione con i prompt utilizzati. Si è già accennato, inoltre, alla difficoltà di ottenere output consistenti con lo stesso prompt. Si è deciso comunque di mantenere lo stesso prompt per la generazione di tutti i testi al fine di ottenere uno standard di riproducibilità quanto più possibile elevato, nei limiti del determinismo imposto dalla macchina.

¹⁰ La temperatura è un parametro utilizzato per regolare la distribuzione delle probabilità dei possibili token generati da un modello di linguaggio, modificando i punteggi associati alle parole, chiamati "logit". Più alta è la temperatura, più uniforme diventa la distribuzione delle probabilità delle parole, favorendo scelte meno probabili e risposte più creative. Al contrario, una temperatura più bassa accentua le differenze tra i logit, favorendo le scelte più probabili e producendo risposte più coerenti ma potenzialmente meno creative.

¹¹ Il parametro "top-p", conosciuto come "nucleus sampling", regola la distribuzione delle probabilità durante la generazione del testo, limitando le scelte del modello ai token con probabilità cumulativa più alta. Riducendo il valore di "top-p", si concentra sui token più probabili, garantendo risposte coerenti ma potenzialmente meno creative. Aumentando "top-p", si amplia la varietà di token considerati, ma a discapito della coerenza.

¹² Si veda almeno la pagina sulla tokenizzazione da Open-AI (<https://platform.openai.com/tokenizer> ultimo accesso 01/01/2025) e Asprovsk & Hunter [2].

Nonostante le difficoltà evidenziate, il processo ha portato alla redazione di un totale di otto articoli, quattro per ciascuna delle due riviste. Questi testi formano il test set che rappresenta il nucleo centrale dell'analisi stilometrica condotta in questo studio.

Analisi stilometrica

Una prima indagine è stata effettuata su un insieme di otto testi, suddivisi in quattro gruppi (di 5000 parole circa) per ciascuna delle due riviste. Occorre precisare che gli scritti non corrispondono a uno specifico articolo ma rappresentano una somma di contributi differenti. Considerando, infatti, che il numero minimo di parole per un'analisi stilometrica affidabile è di almeno 2000,¹³ è stato necessario raccogliere gli articoli dei primi due numeri di ciascuna rivista in quattro differenti insiemi: QP1, QP2, QP3 e QP4 per quanto riguarda «Quaderni piacentini»; QU1, QU2, QU3 e QU4 per quanto riguarda «Quindici». I testi generati da AI sono classificati in due gruppi: AIQP, prodotto utilizzando testi estratti da «Quaderni piacentini» come riferimento; AIQU, realizzato fornendo i testi di «Quindici» come esempio.

Tutti i gruppi di testo sono sufficientemente lunghi per l'analisi stilometrica e dovrebbero rispecchiare una visione non frammentata della rivista. Le analisi stilometriche sono state condotte utilizzando R, sulla base della metodologia sviluppata da Maciej Eder, Jan Rybicki e Kestemont [21]. La metrica di riferimento iniziale è stata “Delta” [13]. Attraverso l'utilizzo di una serie di *features* differenti, si possono individuare traiettorie distintive tra le due riviste. Per la visualizzazione dei primi risultati la creazione di dendrogrammi [59] permette di raccogliere i materiali in gruppi separati sulla base del grado di somiglianza: il raggruppamento consente di osservare anche i diversi livelli a cui i vari contenuti sembrano avvicinarsi.

Dopo questa prima fase, per il confronto tra le riviste sono state applicate diverse metodologie. La Principal Component Analysis (PCA) ha permesso di approfondire gli elementi stilistici che differenziano maggiormente la scrittura delle due riviste. Una specifica indagine è stata condotta attraverso la Zeta Analysis [14], con cui si verifica la probabilità che determinate scelte lessicali possano rivelarsi rappresentative di uno dei due fogli culturali.

Particolarmente significativa per confermare l'affidabilità dell'analisi stilometrica è stata la Rolling Stylometry, che consente di tracciare le variazioni stilistiche progressive all'interno di un singolo testo o tra testi differenti. La Rolling Stylometry ha permesso di attribuire correttamente specifici articoli alla loro rivista di appartenenza, partendo dall'analisi di un pastiche contenente sia testi tratti da una rivista sia contenuti generati nelle fasi precedenti da GPT.

Infine, da un punto di vista stilistico, è apparso rilevante tentare un'indagine sulle *Parts-of-speech*, considerandole come elementi linguistici rappresentativi delle scelte formali degli autori. In questo caso, la cluster analysis fondata sulle parti del discorso e sulla loro frequenza di utilizzo consente anche una gerarchizzazione dei testi sulla base della complessità delle scelte sintattiche.

¹³ Come già ricordato nel paragrafo “OCR, accuratezza e rumore” e in “Definizione del test-set tramite GPT-4”.

Strutture identitarie: cluster analysis di «Quaderni Piacentini» e «Quindici»

Una prima analisi ha rilevato il numero di most frequent words (MFW) capace di offrire i risultati più stabili. Il parametro di incremento progressivo “mfw.incr=50” ha indicato tra le 100 e le 300 mfw il valore di maggiore stabilità. In nessuno dei casi si osserva una perfetta distribuzione dei gruppi di testi, dal momento che alcuni di essi presentano sempre collocazioni problematiche o attribuzioni non corrette. Ciononostante, con 300 mfw il numero di criticità si riduce ma due testi su dieci complessivi, QU2 e QU4, continuano ad avere difficoltà di collocazione.

I risultati della Cluster Analysis condotta, considerando come parametro analitico le parole, su 300 mfw e con l'utilizzo di un solo n-gram (Fig. 1) mostrano come le riviste si siano disposte sostanzialmente su tre rami differenti.

Due ramificazioni sono ben delineate raccogliendo, da un lato, QP4 e QP3, e, dall'altro, QU1 e QU3. La vicinanza è comprensibile in QP3 e QP4 che raccolgono tutti i testi pubblicati sul secondo numero di «Quaderni piacentini». Più rilevante è il legame tra QU1 e QU3 che, pur appartenendo entrambi a «Quindici», si riferiscono a due numeri diversi del giornale: QU1 al primo numero di giugno 1967 e QU3 al secondo numero uscito a luglio 1967. È ipotizzabile, dunque, una maggiore coerenza nella scelta tematica e nella costruzione dei testi in «Quindici» che si estende su numeri diversi.

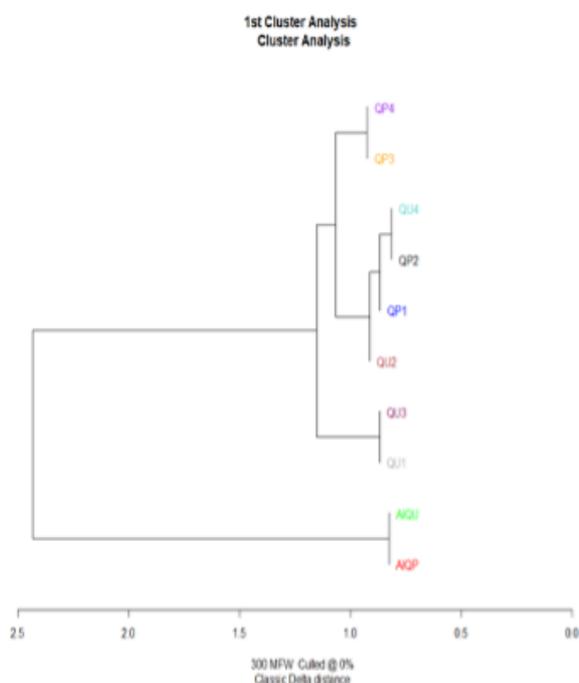


Figura 1 Cluster Analysis con 300 MFW

Il terzo raggruppamento è più stratificato e raccoglie: QP1, QP2, QU2, QU4. L'analisi stilometrica riesce a individuare la vicinanza tra due gruppi di testi, QP1 e QP2, afferenti a «Quaderni Piacentini».

Il legame tra QP1 e QP2 è rafforzato dal fatto che raccolgono tutti i testi pubblicati sul primo numero della rivista fondata da Bellocchio e Cherchi, uscito nell'aprile del 1962.

Più problematica appare la collocazione degli scritti di «Quindici». QU4, che raccoglie testi usciti nel secondo numero della rivista animata dal Gruppo '63 e apparso nel luglio del 1967, è in stretta relazione con QP1 e QP2, che raccolgono l'insieme degli scritti usciti sul primo numero di «Quaderni Piacentini» nell'aprile del 1962. QU4, dunque, presenta tratti stilistici assimilabili alla scrittura dei testi pubblicati da «Quaderni Piacentini».

Andando a verificare le motivazioni dietro a una vicinanza così stretta tra QU4 e il primo numero dei «Quaderni Piacentini», sembra che la corrispondenza avvenga in relazione ad articoli di carattere argomentativo e informativo, trattandosi nello specifico di recensioni. Non potendo dilungarci per ragioni di spazio su tutte scelte relative alle differenti tipologie testuali, si presentano qui di seguito alcuni casi paradigmatici. QP1 e QP2, infatti, comprendono i testi usciti su «Quaderni Piacentini» nell'aprile del 1962, tra cui ci sono: l'appello di Antonio Bellocchio, *Rinnoviamo la biblioteca comunale* e l'intervento di Alberto Bellocchio, *I sette di Genova vogliono il processo*, in cui si instaura una sovrapposizione tra l'operato dei nazisti nel 1943 e la morte di cinque operai a Reggio Emilia negli scontri con la polizia del 1960. In QU4, che raccoglie i testi apparsi su «Quindici» nel giugno del 1967, troviamo l'intervento di Giovanni Di Maio, *Le lotte degli studenti d'architettura* e il giudizio di Carmelo Bene sul *Nuovo teatro: un laboratorio collettivo*, in riferimento al Primo Convegno Internazionale sul Nuovo Teatro del 1967. La distribuzione dei testi su questo ramo da parte della cluster analysis fa emergere, dunque, un caso significativo di contaminazione tra gli articoli pubblicati sul primo numero di «Quaderni Piacentini», apparso nell'aprile del 1962, e i testi estratti dal secondo numero di «Quindici», pubblicato nel luglio del 1967. Nel connubio tra scritti informativi e argomentativi si riscontra il maggior grado di vicinanza tra le due riviste. Laddove «Quindici» lascia spazio alla critica letteraria, con una più marcata tendenza all'argomentazione, la rivista appare editorialmente più connotata e riconoscibile dall'analisi stilometrica. Quando, invece, un articolo affronta eventi d'attualità, la necessità di presentare un più ampio impianto di fonti e informazioni causa una perdita della riconoscibilità originaria della rivista stessa. È dunque su testi esclusivamente argomentativi che si gioca principalmente la distinzione tra le due sedi editoriali a un livello stilometrico complessivo.

Dati i risultati della cluster analysis con un unico n-gram, è stata verificata la tenuta della distribuzione a livello di proposizione. La lunghezza media delle frasi all'interno dei testi pubblicati sulle riviste è di circa 24 token. Nei materiali prodotti da AI si attesta su 21 token. Lo scarto minimo ha fatto ipotizzare una maggiore vicinanza in termini sintattici. Il parametro di 23 n-grams, ovvero la lunghezza media delle proposizioni tra tutti i testi, impostato sul livello “words”, avrebbe dovuto contribuire in linea teorica – dal momento che non avrebbe avuto un riscontro reale sul fatto testuale – all'analisi sintattica.

La distribuzione dei testi ne esce chiaramente rimodulata (Fig. 2).

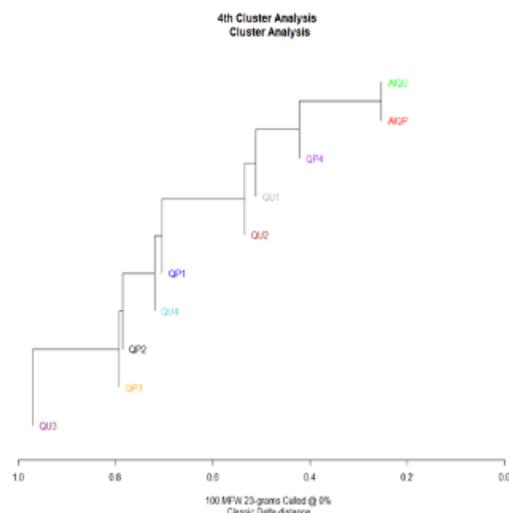


Figura 2 Cluster Analysis con 23 n-grams

Persiste l'isolamento dei materiali *AI-generated*. Sono pochi, tuttavia, anche i legami consolidati tra i testi delle riviste. Ad essere significativa è la collocazione del gruppo QU4 che raccoglie i testi finali del secondo numero di «Quindici», apparso nel luglio 1967. QU4 conserva la sua difficile collocazione e conferma la vicinanza agli scritti estratti da «Quaderni Piacentini». Anche a livello della proposizione – sebbene con un necessario e ampio grado di approssimazione per le ragioni sopra esposte – QU4 si configura come un insieme di testi di primaria importanza. Analizzando questo gruppo, emergono infatti strutture ricorrenti: una sintassi frammentaria con frasi brevi e spesso nominali; a livello lessicale si osserva una predilezione per il linguaggio colloquiale e quotidiano, che in alcuni casi imita l'oralità e in altri sfocia nel gergale (si segnalano, ad esempio, diverse trascrizioni volutamente scorrette, finalizzate a una mimesi del parlato); e si ritrovano espressioni straniere di uso comune (*Very important people, American way of life*).¹⁴ L'analisi stilometrica sembra dimostrare che queste strategie espressive sono decisamente meno frequenti negli altri numeri o sezioni di «Quindici», mentre ricorrono in «Quaderni Piacentini», probabilmente più fedeli all'intenzione originaria proclamata nel primo numero: «esser seri senza essere noiosi» [50]. Si tratta, inoltre, dell'insieme di scritti più lontani da quelli *AI-generated*. L'ipotesi potrebbe mostrare la difficoltà dell'AI di imitare strutture formali che traggono il loro principale portato semantico da uno spazio extratestuale, basandosi invece sul lessico interno maggiormente significativo. Questa prospettiva sembra trovare conferma nell'analisi dei testi raccolti in QP4, l'insieme più vicino alla produzione *AI-generated* con 23n-grams. QP4 raccoglie due contributi di Cesare Rossi e Renuccio Tirelli pubblicati su «Quaderni Piacentini» nell'aprile del 1962. I testi si presentano più estesi rispetto alla lunghezza media degli altri articoli apparsi sulla rivista. I due interventi si soffermano sul dibattito, allora in corso, intorno all'alleanza politica che avrebbe portato alla nascita, nel 1963, di quello che viene definito il primo governo di centro-sinistra della Repubblica italiana. A livello stilistico, la sintassi non è più discontinua come negli altri scritti dei primi due numeri di «Quaderni Piacentini» e le proposizioni si allungano in un più insistente ricorso all'ipotassi:

¹⁴ [1] Per una definizione più dettagliata dei fenomeni si faccia riferimento a Berruto [9], in particolare il cap. 5: «Tratti dell'italiano popolare e tratti del parlato colloquiale», pp. 81-101.

Al fondo di questo complesso di inferiorità, che a volte assume anche la figura della vergine che non vuole andare per il mondo per paura di perdere ciò che infine non ha, perché lo ha solo in quanto non ha mai subito nessun tentativo di seduzione (perché sa che i tentativi di seduzione fatti dal neocapitalismo sono tutti riusciti, almeno nella prima fase), vi è la mediocrità della sinistra e la non volontà della stessa di impegnarsi ([49]:25).

All'accresciuta complessità sintattica si affianca la ricerca di un lessico specializzato di ambito politico con espressioni cristallizzate in una retorica fortemente stereotipata. I riferimenti alla teoria marxista suggeriscono sfere semantiche più ampie rispetto al significato letterale delle singole parole. A risentirne è soprattutto la linearità del ragionamento che attraverso l'anastrofe inverte il nesso logico e la causa “non aver subito nessun tentativo di seduzione” è posticipata rispetto alla conseguenza “rinunciare ad andare per il mondo per paura di perdere ciò che infine non ha”. Il chiarimento dell'inversione strutturale avviene attraverso il riferimento al neocapitalismo come forma di seduzione a cui, secondo Rossi, si arrendono gli esponenti della sinistra ormai disposti ad aprirsi a forme più accettabili di ascendenza liberale. Sulle espressioni desunte dalla teoria marxista o dal lessico politico più semanticamente connotato si fondano i testi *AI-generated*, che fanno particolare affidamento a topoi della retorica di sinistra. Nei testi di GPT vediamo ad esempio ricorrere l'espressione “vecchio ordine”, slogan quali “Avanti, compagni!” ma anche riferimenti al ruolo delle “aule universitarie come palcoscenico politico”. L'insieme di articoli contenuti in QP4, apparsi nell'aprile del 1962 su «Quaderni Piacentini» sembra, dunque, distanziarsi dai principi ispiratori del giornale fondato da Bellocchio e Cherchi. La struttura dei suoi testi, tuttavia, si è rivelata molto più fertile nel fornire riferimenti formali, soprattutto in ambito lessicale, a ChatGPT per la generazione dei testi.

Principi di riconoscibilità: Principal Component Analysis

Le variabili nella distribuzione della cluster analysis hanno indicato le direttive lungo cui ricercare i principali elementi distintivi tra i gruppi di testi. A questo scopo, è stata condotta una Principal Component Analysis (PCA) prima sugli articoli estratti dalle riviste e successivamente sull'insieme del corpus.

L'Analisi delle Componenti Principali mostra la quantità di varianza nei dati che è spiegata da ciascuna componente principale, dove con varianza si intende una misura della variabilità nei dati. Pertanto, nei grafici successivi (vedi Fig.3), una percentuale di varianza più alta per una componente principale indica che quella componente rappresenta una porzione maggiore della variabilità totale dei dati. In altre parole, le percentuali forniscono un'idea di quanto ogni componente principale contribuisca a spiegare le differenze stilometriche nei testi.

Nella PCA ogni punto rappresenta un testo diverso nel corpus, e la distanza tra i punti riflette la somiglianza stilometrica tra i testi. I testi stilisticamente simili sono dunque vicini tra loro nel grafico. Sono stati eseguiti due diversi tipi di PCA: con matrice di covarianza (cov) e con matrice di correlazione (corr.). La differenza principale tra le due riguarda il tipo di matrice utilizzata nell'analisi delle componenti principali. PCA cov utilizza una matrice di covarianza, che tiene conto delle relazioni di variazione tra le parole nei testi, ossia attribuisce maggiore importanza alle parole che variano molto da un testo all'altro. Al contrario PCA corr. utilizza una matrice di correlazione, che tiene conto delle relazioni di associazione tra le parole nei testi anziché della loro variazione, risultando dunque meno influenzato dalle differenze di variazione tra le parole.

In sostanza, PCA con matrice di covarianza considera quanto le parole cambiano insieme tra i testi, mentre PCA con matrice di correlazione considera quanto sono associate le parole tra i testi.

Anche in questo caso la distribuzione è solo parzialmente omogenea (Fig. 3). QU2, QU3 e QU4, che raccolgono i testi usciti nella seconda parte del primo numero e tutto il secondo numero di «Quindici», si allineano lungo lo stesso asse. QU1, dove si trovano le intenzioni programmatiche di «Quindici» uscite nel primo numero del giugno 1967, appare decisamente distanziato. L'elemento principale che determina la distanza è l'uso del verbo essere alla terza persona singolare del presente: «è». Lo specifico token denota un tentativo classificatorio e autodeterminante che si ritrova nel manifesto pubblicato sul primo numero della rivista: «è un periodico redatto da un gruppo», «ciascuno dei suoi appartenenti è perfettamente libero», «la libertà è garantita dal basso costo d'esercizio», «lo sciopero è in atto da anni, il silenzio è completo», «lo sciopero è compatto e unanime», «è un giornale continuo». ([51]:1)

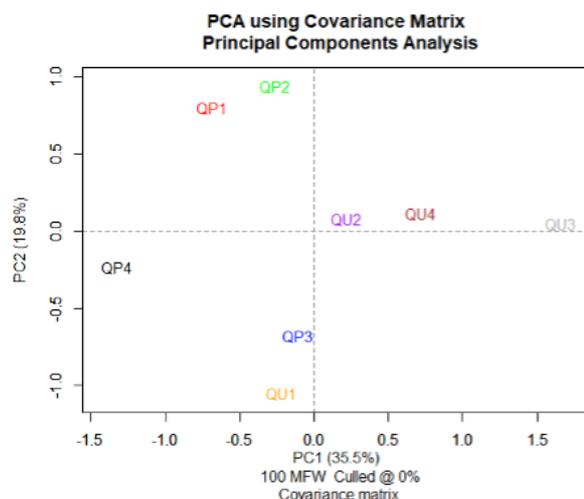


Figura 3 Principal Component Analysis con matrice di covarianza

Lo stesso tentativo di definire uno spazio di indagine si trova nell'articolo di Alfredo Giuliani, *Le cerimonie sadiche della critica*, pubblicato a sua volta sul primo numero: «lo Spazio Letterario è scomodo», «L'atteggiamento tipico dello scrittore [...] è una specie di inconsapevolezza attiva nei confronti del linguaggio», «la sua intenzionalità[...] è, al contrario, una specie di consapevolezza passiva del linguaggio», «ingenua metafora è il credere che l'opera sia un "linguaggio primo"», «La letteratura come menzogna è un libro altamente provocatorio». ([24]:1)

Ad ogni proposizione corrisponde una definizione: si delineano gli obiettivi della rivista, chi sono i suoi collaboratori, quali sono i punti di partenza e le prospettive di arrivo, qual è l'atteggiamento verso la critica e il linguaggio. Le prese di posizione nette e provocatorie diventano tanto caratterizzanti in QU1 da essere in contrasto con lo stile adottato nel corso del tempo da «Quindici», come denuncia la distanza dagli altri gruppi di scritti, QU2, QU3, QU4.

Più omogenea è la distribuzione quando viene considerata la matrice di correlazione che porta a una disposizione distesa in tutto il quadro cartesiano. Le *function words* conducono a questo risultato e in particolare la congiunzione "e" che trova una frequenza quasi identica in tutti i testi. Il linguaggio delle due riviste, per quanto ambisca a trattare aspetti di carattere culturale, politico e sociale, conserva la sua aderenza a uno stile comunicativo che appare necessario per affrontare con immediatezza l'attualità, scopo ultimo di entrambi i fogli. Due sono le considerazioni che

hanno condotto alla distribuzione della cluster e della principal component analysis e, di conseguenza, al confronto stilometrico tra gli stili editoriali.

L’uso di un linguaggio generalmente riconducibile all’italiano standard costituisce una base di riferimento stabile. L’inserimento raro e talvolta inaspettato di forme che deviano dalla norma generale (ad es. formule mutuare dal parlato, adozione di prestiti da lingue straniere, introduzione di periodi più marcatamente ipotattici) risulta determinante nell’individuazione di stili specifici e distintivi, utili per la comprensione dell’afferenza alla rivista.

Lessici stranianti: un tentativo di zeta analysis sui testi delle riviste

Le ultime fasi del confronto tra le riviste mirano a esplicitare il più possibile fenomeni lessicali e sintattici capaci di rappresentare l’una o l’altra strategia editoriale, pur consapevoli, come dimostrato finora, che non esiste una distinzione netta tra i due approcci a livello di scrittura.

La principale metodologia stilometrica di comparazione lessicale tra corpora è il metodo Zeta, con cui è stato effettuato il confronto tra i testi delle riviste. I risultati segnano un primo sostanziale scarto rispetto alle considerazioni emerse nelle fasi di indagine precedenti. Nella Zeta Analysis, infatti, le scelte lessicali tracciano con maggior nettezza la posizione di ciascuna rivista (Fig. 4).

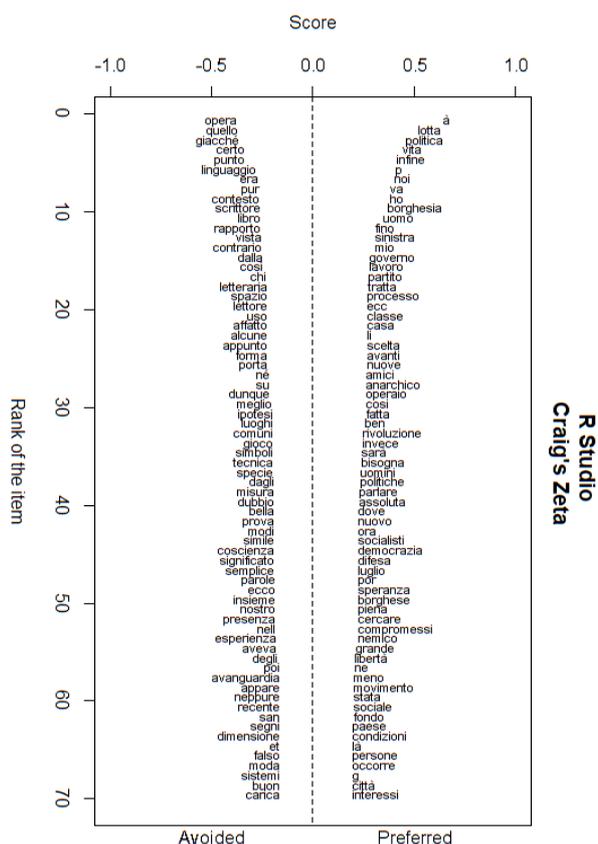


Figura 4 Zeta Analysis. Mostra i termini preferiti/evitati dalle riviste

Di particolare rilievo in questo caso è la distanza tra i campi semantici adottati. «Quaderni piacentini» manifesta un maggior ricorso a termini di ambito politico come “lotta”, “comunisti”, “democrazia”, “compagni”, “socialista”, “fascista”. La ricerca di un intervento pratico sia a livello locale che nazionale, dichiarato in più di un’occasione non solo dai direttori della rivista, ma anche dai vari contributori, ha prodotto il consolidarsi di una terminologia che ben definisce il campo d’azione. La Zeta analysis conferma l’interesse per una dimensione sociopolitica di «Quaderni piacentini», sottolineando il rilievo di specifiche scelte lessicali, da “borghesia” a “classe”, da “anarchico” a “rivoluzione”, ma permette anche un’indagine più profonda sulle strategie che la rivista persegue: un impegno prima di tutto locale dove il rilievo di parole come “casa”, “città”, “persone” fa da eco all’articolo di Antonio Bellocchio sulla necessità di rinnovare la biblioteca Passerini-Landi di Piacenza. Come osservato da Zinato [64], «Quaderni Piacentini» ha dato spazio a un importante filone della critica letteraria e culturale italiana. La rivista conserva una riconoscibile impronta terminologica radicata nella discussione politica. Un confronto a livello diacronico tra il lessico utilizzato negli articoli dei primi anni e quello adottato negli articoli dei decenni successivi potrebbe confutare queste considerazioni, aprendo ulteriori riflessioni sullo sviluppo storico del giornale.

In «Quindici» gli interessi dei fondatori fanno propendere verso una terminologia tipica della critica letteraria con parole quali “opera”, “linguaggio”, “contesto”, “scrittore”, “lettore”, “ipotesi”, “simboli”, “tecnica”, “avanguardia”, “segni”. Anche in questo caso, la Zeta Analysis non si limita a definire un campo semantico complessivo, ma suggerisce i temi di un dibattito che rimanda a una specifica stagione culturale. Il rilievo accordato alla triade opera-linguaggio-contesto non può che riferirsi alla controversia intorno allo strutturalismo a cui il Gruppo ’63 era particolarmente sensibile: collocandoli come termini distintivi si vengono a configurare quali identità lessicali della rivista, un elemento che sarà significativo nel confronto con i testi *AI-generated*. A completare la prima costellazione del lessico di «Quindici» è il binomio scrittore-lettore che sarà decisivo nel passaggio a una nuova fase della critica letteraria degli anni Settanta con la diffusione in Italia della teoria ermeneutica.

L’indagine lessicale esprime, dunque, una più evidente contrapposizione tra le due riviste, che è fondamentale tener presente nel confronto tra i testi originali e quelli generati da Large Language Models. La Zeta Analysis si rivela uno strumento significativo quando si vuole tentare uno studio diacronico del linguaggio. Nel passaggio dalla centralità del lessico sociopolitico dei primi anni di «Quaderni piacentini» a quello intriso della terminologia strutturalista o delle scienze sociali di «Quindici» alla fine degli anni ’60, si osserva lo sviluppo della critica culturale italiana in due delle sedi editoriali più attente all’attualità. Per quanto ciascuna rivista rivendicasse un approccio ideologicamente diverso dall’altra, entrambe professano un impegno ad un tempo sociale, cioè immerso nella realtà circostante, e culturale, in particolare fondato sulla critica letteraria. Questo duplice approccio è smentito dalla Zeta Analysis che evidenzia una propensione di «Quaderni piacentini» verso il primo filone e di «Quindici» verso il secondo.

Scelte strutturali: una cluster analysis fondata sulle parts-of-speech

L’ultima analisi di questa prima fase dello studio è stata un’indagine delle *parts-of-speech*.

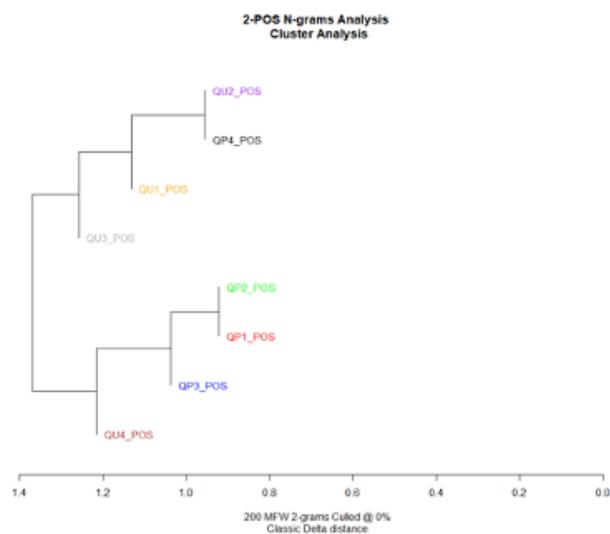


Figura 5 Indagine delle parts of speech, 200 MFW, 2 n-grammi

Per preparare il corpus la funzione `udpipe` ha annotato i testi sulla base del framework Universal Dependencies. A differenza di quanto accadeva con la prima cluster analysis condotta sui testi non annotati, ma in linea con i risultati della Zeta Analysis, anche qui la contrapposizione tra la scrittura delle due riviste appare ben più definita. La prima verifica è stata condotta su 200 MFW, che presenta la maggiore stabilità anche nell'analisi iniziale, prendendo in considerazione 2 n-grams a livello *word*. La scelta del numero degli n-gram non è stata casuale: il livello *word* qui corrisponde a una specifica *part-of-speech* e, di conseguenza, 2 n-grams possono restituire alcune considerazioni preliminari sulla costruzione sintattica delle proposizioni.

I risultati mostrano una distribuzione più stabile rispetto alla cluster analysis condotta sui materiali di partenza (Fig. 5). Quasi tutti i gruppi di testi sono collocati nel ramo corrispondente ed evidenziano una ben precisa riconoscibilità nella scrittura. Allargando leggermente i parametri, con 400 MFW e 4 n-grams la stabilità è assoluta (Fig. 6). Ogni rivista, così come i materiali *AI-generated*, occupa correttamente il proprio ramo.

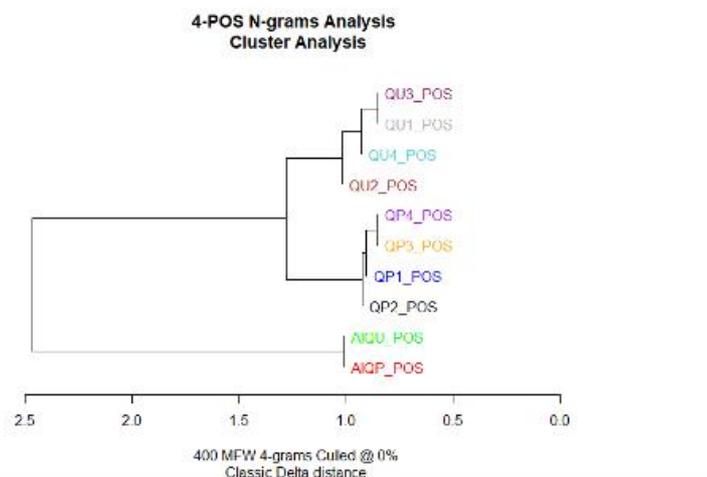


Figura 6 Cluster Analysis con 400 MFW, 4 parts of speech

Simili risultati suggeriscono una riflessione non soltanto sulla riconoscibilità delle riviste, ma anche sull'importanza della stilometria come metodologia di analisi stilistica a livello di raccolte collettive.

Il tratto distintivo di ciascun corpus, infatti, è fondato sulla costruzione sintattica della frase. La distribuzione più stabile si ottiene con la successione delle *parts-of-speech* che rafforzano i risultati lessicali e ne confermano l'affidabilità. Mentre a livello tematico è immediatamente verificabile e più comprensibile una distanza tra i testi, a livello sintattico si registrano spesso le stesse strutture, con il ricorso alla paratassi e a proposizioni brevi. Nel momento in cui si prendono in considerazione elementi di maggior dettaglio come l'alternanza di diverse parti del discorso, le stime complessive vengono confutate e la riconoscibilità stilistica emerge più marcatamente, suggerendo l'esistenza di uno stile privilegiato da ciascuna specifica rivista. Una considerazione che ha un rilievo ancora maggiore se si pensa che alcuni degli autori firmano contributi sia per l'uno che per l'altro giornale. In questo caso, i risultati suggeriscono che l'adattamento dello stile espressivo non avviene solo a livello lessicale – elemento che appare molto più ovvio – ma anche a livello sintattico, denunciando una ricezione delle modalità di scrittura presso il pubblico tanto radicata da determinare esigenze e scelte linguistiche.

La stilometria si rivela, dunque, uno strumento cruciale nello studio di grandi corpora collettivi. La sua capacità di individuare e distribuire correttamente i testi in base alla propria collocazione editoriale potrebbe aprire nuove riflessioni sulla storia della lettura e dell'editoria, e sullo studio dei movimenti culturali.

Stilometria su testi AI generated

Una volta verificata la specificità stilistica delle singole riviste di riferimento, è stato possibile includere nell'analisi comparativa anche i testi generati da GPT-4. Per ciascuna rivista è stato creato un blocco testuale contenente l'intero insieme degli articoli generati artificialmente, con l'obiettivo di valutare la qualità dell'emulazione stilistica rispetto ai materiali originali.

I risultati, rappresentati nei dendrogrammi (Fig. 1) e nell'albero di consenso, indicano chiaramente che i testi prodotti da GPT-4 presentano una cifra stilistica riconoscibile e distinta.

Le analisi di clustering mostrano una netta separazione tra i contenuti generati artificialmente e quelli autentici, provenienti da Quaderni Piacentini» e «Quindici». I testi creati dall’AI tendono a costituire cluster separati, segno di una distanza stilistica e tematica persistente rispetto alle configurazioni proprie delle riviste.

Tale esito, tuttavia, non deve sorprendere. L’utilizzo di un LLM generalista come GPT-4, progettato per affrontare una vasta gamma di compiti linguistici con impostazione neutrale e standardizzata, comporta limiti prevedibili nella capacità di riprodurre le sottigliezze stilistiche, ideologiche e retoriche tipiche di pubblicazioni storiche con una forte identità editoriale. In questo contesto, l’asimmetria osservata nei cluster non va intesa come una mancanza tecnica del modello, ma piuttosto come una conseguenza metodologica del tipo di strumento utilizzato.

Diverso potrebbe essere l’esito se si impiegassero modelli customizzati o specificamente ottimizzati per compiti di emulazione stilistica e autoriale. In letteratura recente, modelli altamente specializzati — i cosiddetti Highly Literate Language Models (HLLM) — hanno mostrato una maggiore capacità di riprodurre strutture discorsive e stili autoriali complessi, grazie a un addestramento mirato su corpora curati e omogenei [58]. In questo senso, i risultati qui discussi offrono una prima base di confronto e suggeriscono una linea di ricerca futura orientata all’utilizzo di modelli più aderenti ai domini stilistici analizzati.

Le analisi con 100 MFW e 23 n-grams, in particolare, (Fig. 2) hanno confermato che la stilometria è in grado di discriminare efficacemente tra i testi delle riviste e quelli generati da GPT.¹⁵ La presenza di strutture formali e lessicali uniche nelle riviste risulta in una marcata differenziazione rispetto ai testi *AI-generated*, indicando che le sfumature stilistiche e tematiche rimangono al di fuori della portata attuale di GPT-4. La distanza lessicale e sintattica porta a interrogarsi sulla reale influenza dell’orizzonte d’attesa sulle scelte compositive. Come osservato in precedenza, il lettore ideale non coincide necessariamente con il lettore reale e risulta complesso, e poco proficuo, definire entrambe le figure sulla base delle sole caratteristiche interne dei testi. Nel caso di testi *AI generated* potrebbe invece rivelarsi utile approfondire, in un lavoro futuro e attraverso il ricorso a feedback specifici, l’esistenza di un pubblico di riferimento sulla base del quale sono stati sviluppati i testi.

La distanza tra gli articoli originari e quelli realizzati da GPT-4 appare, in maniera ancora più accentuata, anche nel Multidimensional Scaling (MDS) come evidenziato in Fig.7.

¹⁵ Per queste analisi si è seguita parzialmente la traccia metodologica proposta da Zaitsev & Jin [61].

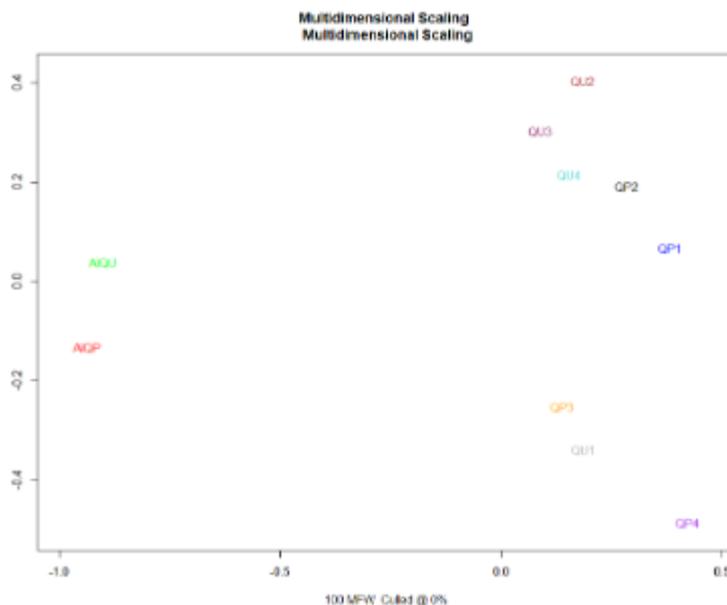


Figura 7 Multidimensional Scaling con 100 MFW

In questo caso entrambi i testi scritti da AI si pongono in netta separazione con quelli delle riviste andando ad occupare gli estremi sul piano cartesiano.

Analogamente, la PCA conferma questa separazione (Fig.8). Gli articoli delle riviste si distribuiscono lungo l'asse principale, indicando variazioni significative tra di loro che riflettono diversità stilistiche, sebbene con parziali sovrapposizioni. I testi AI, al contrario, si raggruppano vicini all'origine, suggerendo una minore varianza tra di loro rispetto ai materiali delle riviste.

Osservando le frequenze lessicali emergono invece spunti significativi per evidenziare le differenze stilistiche tra i testi generati e quelli originali. In particolare, gli n-grams riscontrati nei testi AI tendono ad avere frequenze molto basse o addirittura assenti negli articoli delle riviste, suggerendo una certa uniformità e mancanza di specificità tematica nei testi generati rispetto alla ricchezza lessicale dei testi originali. Al contrario, le frequenze derivate dalle riviste rivelano una maggiore varietà lessicale e un ricorso più frequente a sequenze di parole che riflettono tematiche e scelte stilistiche specifiche. Questa differenza sottolinea l'incapacità dell'intelligenza artificiale di riprodurre pienamente le complessità stilistiche che caratterizzano le pubblicazioni umane. Inoltre, la coerenza interna osservata in alcuni n-grams comuni tra testi di una stessa rivista indica l'esistenza di linee editoriali ricorrenti e di tematiche che attraversano più articoli, suggerendo una identità stilistica delle riviste stesse.

L'analisi delle frequenze, dunque, contribuisce a rafforzare la distinzione già emersa dalle cluster analysis e dal multidimensional scaling, mostrando come GPT-4 tenda a produrre scritti coerenti sul piano linguistico, ma meno adatti a rispecchiare le scelte lessicali e tematiche specifiche che definiscono l'identità editoriale delle due pubblicazioni.

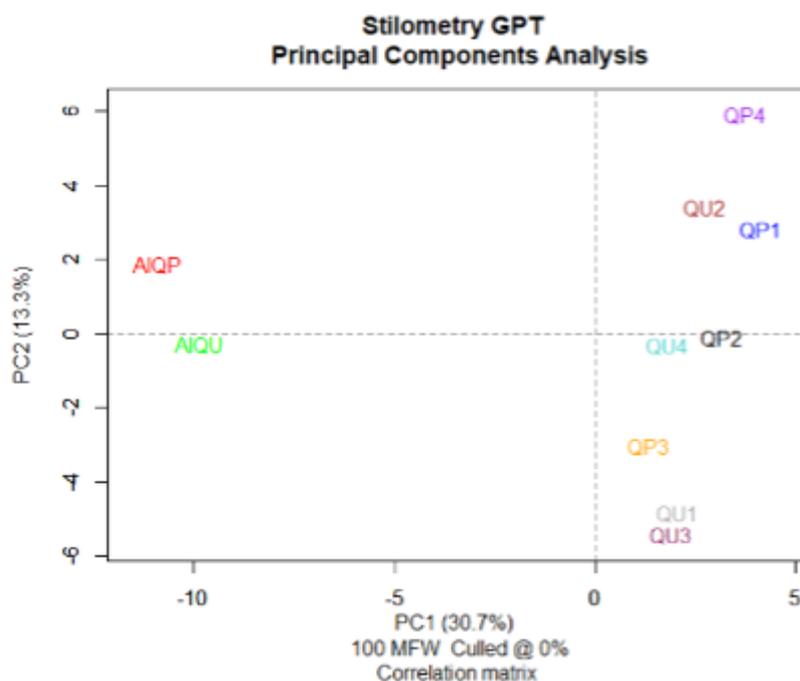


Figura 8 Principal Component Analysis 100 MFW con matrice di correlazione

Non diversi sono i risultati con l'applicazione della Zeta Analysis (Fig. 9). Quest'ultima mostra una separazione a livello lessicale tra i due insiemi di testi tanto più significativa se ci soffermiamo sui termini maggiormente utilizzati da ChatGPT in relazione alla scrittura delle riviste. Tra le parole che determinano il divario troviamo “società”, “ruolo”, “sfide”, “cambiamento”, “contesto”, “arte”, “futuro”, “culturale”, “sociale”, tutti termini portatori di significato e con un rilevante peso semantico. Sorprendentemente questi termini non risultano utilizzati da GPT, sicché il modello, nel tentativo di imitare lo stile dei giornali culturali, pare affidarsi più alle *function words* che ai nuclei tematici. A confermare questa ipotesi è la scelta delle parole *preferred*: la colonna dei termini preferiti da GPT-4 mostra un ricorso significativo a pronomi (es. “se”, “io”, “tutto”) e a parole funzionali o legate alla struttura discorsiva, come “quando”, “perché”, “ancora”. Questo suggerisce che l'AI tende a favorire un linguaggio orientato alla narrazione o all'argomentazione personale e diretta. Ancora i testi AI mostrano una preferenza netta per termini più astratti e generalizzati, come “qualche” e “uomo”, mentre evitano parole maggiormente connotate culturalmente e specifiche, come “libertà” e “identità”. Questo pattern lessicale evidenzia come GPT-4, pur riuscendo a produrre testi linguisticamente coerenti, faticò a cogliere le sfumature tematiche e stilistiche caratteristiche di pubblicazioni editoriali storiche e si limitò a produrre testi caratterizzati da un vocabolario neutro o universale. Di conseguenza, la capacità di GPT di emulare contesti culturali specifici risulta limitata, confermando la necessità di ulteriori sviluppi per modellare con maggiore precisione le scelte lessicali di contesti umani distinti.

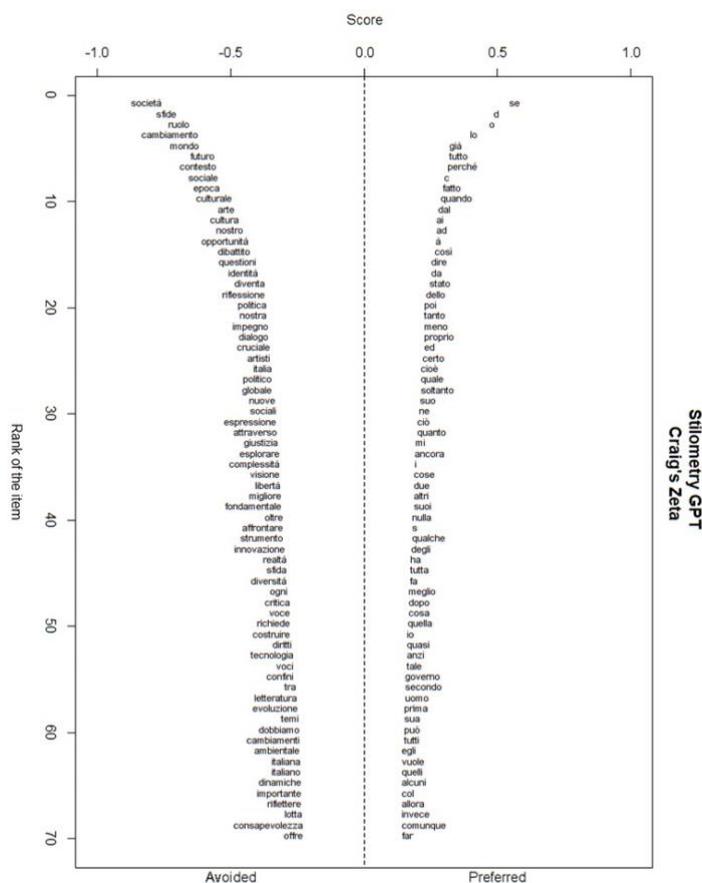


Figura 9

Infine, è stata condotta un’analisi di *Rolling Stylometry* tramite la funzione `rolling.classify` del pacchetto `Stylo`, con l’obiettivo di testare la capacità dell’algoritmo di individuare transizioni stilistiche all’interno di un testo composito. A questo scopo è stato appositamente realizzato un *pastiche* suddiviso in tre sezioni, ciascuna rappresentativa di una diversa “fonte autoriale”: la prima interamente generata da Gemini, modello linguistico di grandi dimensioni sviluppato da Google DeepMind, la seconda composta da testi prodotti da GPT-4, e la terza costituita da estratti autentici dai *Quaderni Piacentini*.

L’inclusione della sezione generata da Gemini non risponde a una semplice esigenza di confronto tra modelli, bensì alla volontà di introdurre un elemento deliberatamente perturbante nel contesto dell’esperienza stilometrica. Questo approccio si ispira alla metodologia del cosiddetto *impostors method* ([36]; ma si veda anche *impostors* nel pacchetto `idiolect`),¹⁶ secondo cui l’inserimento di testi “estranei” rispetto al ventaglio delle classi note può rivelarsi utile per affinare la capacità discriminativa degli algoritmi di attribuzione autoriale. In questo caso, il

¹⁶ <https://www.rdocumentation.org/packages/idiolect/versions/1.0.1/topics/impostors> (ultima consultazione maggio 2025)

segmento generato da Gemini funge da “falso autore”, permettendo di esplorare fino a che punto il sistema sia in grado di riconoscere e tracciare discontinuità stilistiche anche tra testi generati da modelli simili ma distinti per architettura, addestramento e output linguistico.

L’analisi di *Rolling Stylometry* è stata condotta segmentando il pastiche in blocchi consecutivi di 500 parole, come da prassi metodologica del pacchetto Stylo, che consente di mappare la variazione stilistica lungo un continuum testuale. Tale approccio, che assume una logica scorrevole (*rolling*) anziché discreta, permette di osservare transizioni progressive e non necessariamente nette tra diversi stili, risultando particolarmente utile in contesti di autorialità composita o ambigua.

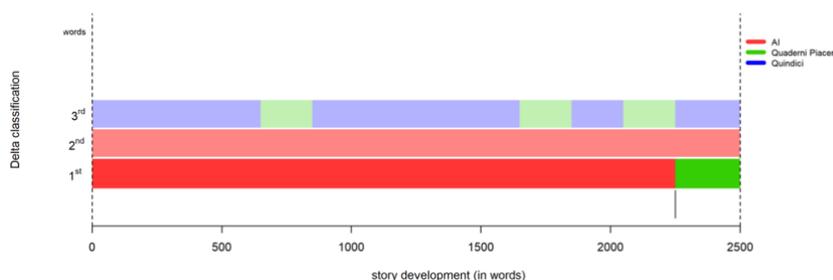


Figura 10 Rolling stylometry applicata su un Pastiche composto da AI e testi di riviste

I risultati (Fig.10) confermano la solidità dello strumento nel riconoscere discontinuità stilistiche macroscopiche: la Rolling Stylometry ha individuato con buona accuratezza i confini tra le tre sezioni del pastiche. In particolare, si osserva una distinzione netta tra i blocchi generati da modelli di AI (GPT-4 e Gemini) e quelli tratti da testi umani (in questo caso da «Quaderni Piacentini»). Tuttavia, è significativo notare che il metodo non è riuscito a discriminare con chiarezza tra GPT-4 e Gemini, suggerendo che, nonostante le differenze nei rispettivi modelli di addestramento e produzione linguistica, il loro output stilistico risulta simile dal punto di vista delle metriche adottate nella stilometria classica.

Questa mancata distinzione tra modelli di AI può essere interpretata in più modi. Da un lato, evidenzia i limiti dell’approccio basato su frequenze lessicali e misure derivate (come le frequenze di *function words* o n-grammi), che potrebbero non essere sufficientemente sensibili a variazioni stilistiche generate da architetture di deep learning simili. Dall’altro lato, sottolinea una convergenza stilistica tra modelli LLM, i cui output tendono a stabilizzarsi attorno a uno “stile mediano”, spesso ottimizzato per coerenza e leggibilità ma privo della complessità retorica e sintattica che caratterizza testi umani autentici.

Al contrario, i testi tratti dai «Quaderni Piacentini» si sono distinti chiaramente nel tracciato stilometrico, suggerendo una maggiore ricchezza stilistica, probabilmente attribuibile a una sintassi più varia, a un lessico ideologicamente marcato, e a una densità argomentativa che riflette la natura culturale e politica della rivista. La Rolling Stylometry, in questo senso, non ha solo svolto una funzione classificatoria, ma ha anche offerto una rappresentazione topografica della variazione stilistica, rivelando come l’intelligenza artificiale, pur riuscendo a imitare certi registri, non riesca ancora a eguagliare la complessità pragmatica e stilistica del discorso umano situato.

In sintesi, questa sperimentazione conferma l'utilità della Rolling Stylometry come strumento esplorativo nei contesti di testualità ibrida e, al contempo, ne mostra alcune limitazioni epistemologiche: in particolare, la difficoltà a distinguere tra stili artificiali simili suggerisce la necessità di integrare tali metodi con approcci più sofisticati, eventualmente basati su rappresentazioni distribuzionali (*embeddings*), modelli neurali o metodi multivariati non lineari.

Conclusioni

Secondo quanto emerso, risulta chiara la riconoscibilità dei testi *AI generated* rispetto a quelli scritti da umani. È altresì innegabile come l'emulazione di una specifica impostazione editoriale rappresenti una sfida ancora difficile da affrontare per GPT, dal momento che l'analisi stilometrica delinea una chiara distanza tra gruppi di testi. Ciononostante, lo studio potrebbe rivelarsi utile per futuri sviluppi, in particolare per il processo di indagine qui adottato.

La costruzione del corpus di riferimento ha portato a definire i tratti linguistici maggiormente rilevanti in un'analisi che tenga conto non di un singolo testo ma di un insieme di scritti all'interno di una stessa sede editoriale, la rivista appunto. Il lavoro su insiemi di testi, scritti da autori differenti ma inquadrati in una stessa cornice di pubblicazione, e il tentativo di stabilire una strategia ad ampio raggio potrebbero dischiudere nuovi percorsi di approfondimento sulle pratiche di scrittura e di lettura, nonché sulle specifiche modalità di fruizione dei prodotti culturali in determinati periodi storici. La successiva analisi stilometrica, condotta nella seconda fase del lavoro, ha sondato in via preliminare l'esistenza di possibili fattori identitari in grado di determinare la riconoscibilità di singole riviste. Un primo approccio generico attraverso la delta analysis ha mostrato una distanza tra le due riviste analizzate, «Quaderni piacentini» e «Quindici», confermata anche a livello lessicale dalla Principal Component Analysis e dalla Zeta Analysis. Le diverse scelte sintattiche e lessicali hanno evidenziato specifici posizionamenti dal punto di vista stilistico e tematico di ciascuno dei due giornali, con alcune sostanziali eccezioni. Nei casi in cui la distanza tra le riviste viene meno, le ragioni sono riconducibili a scelte relative alla tipologia testuale di riferimento che evidenzia una maggiore omologazione espressiva quando si ricorre a testi informativi relativi a eventi di attualità e registra, al contrario, una distanza più marcata in occasione di testi argomentativi appartenenti ad ambiti specifici come la critica letteraria.

Il successivo confronto stilometrico tra articoli realizzati da umani e da GPT-4 ha consentito ulteriori riflessioni non solo sulla capacità espressiva dell'LLM di OpenAI di produrre scritti appartenenti a varie tipologie testuali, ma anche sulle scelte strutturali più profonde delle riviste. L'ingresso nell'analisi di articoli *AI-generated* ha portato a uno squilibrio, nel Multidimensional Scaling, anche tra gli scritti originari ricavati dalle riviste, richiedendo un'indagine dettagliata delle potenzialità dell'analisi stilometrica in merito allo studio di corpora collettivi. I risultati di quest'ultima fase suggeriscono di rapportare la produzione testuale di GPT-4 a specifiche tipologie di scrittura attraverso feedback che inducano il Large Language Model a proseguire nella stesura degli articoli. A questa possibile direzione di indagine si lega strettamente l'esistenza o meno di un orizzonte d'attesa interno ai testi prodotti da GPT-4: uno spazio proficuo di riflessione ancora da sondare ed esplorare.

Considerata la crescente diffusione degli LLM nella produzione culturale e accademica, appare particolarmente promettente estendere l'analisi comparativa anche ad altri modelli generativi, sottoponendoli a prompt analoghi per valutarne le performance relative. In tal senso, la replica del protocollo qui adottato su modelli differenti – come Gemini, Claude o Mistral – permetterebbe di interrogare non solo le capacità stilistiche dei singoli LLM ma anche le scelte editoriali implicite nei loro dataset e nella loro architettura. Un'analisi comparata fornirebbe, inoltre, una mappatura preliminare delle "zone d'ombra" in cui tali modelli tendono a convergere o fallire, utile per raffinare

i prompt o progettare interventi di fine-tuning più mirati. I risultati negativi – intesi come le discrepanze più marcate rispetto al modello editoriale atteso – non andrebbero dunque considerati meri limiti operativi, bensì come preziosi indicatori di ciò che i modelli non sono ancora in grado di fare, costituendo al contempo prezioso materiale diagnostico, utile non solo al fine dell’affinamento del prompting, ma anche per l’addestramento futuro dei modelli stessi.

Il confronto tra scritti umani ed *AI-generated* idealmente finalizzati alla pubblicazione su una sede editoriale – il giornale letterario appunto – potrebbe portare a nuove riflessioni sulle modalità di addestramento degli LLM stessi, sulle loro attuali capacità generative e sull’insieme di tecniche adottate per il loro sviluppo: individuare e riprodurre una strategia testuale rappresenta, ciononostante, una sfida ancora complessa e un terreno da esplorare. In questo senso, è auspicabile che l’analisi stilometrica si rivolga in futuro, come solo in parte ha già fatto, a miscellanee, che vadano dalle riviste alle raccolte collettive di opere di diversi autori. Un simile orizzonte di studi si affianca a un approfondimento parallelo della capacità imitativa degli LLM rispetto a insiemi complessi di testi, aprendo a importanti novità nella storia della lettura e nella teoria della ricezione.

Acknowledgements

Il progetto di Marco De Cristofaro è finanziato dall’Unione Europea nell’ambito del programma Horizon 2020 – Marie Skłodowska Curie COFUND action – grant agreement No 101034383.

References

- [1] Asor Rosa, Alberto. 1975. «La cultura». In *Dall’Unità a oggi*. Vol. IV. Storia d’Italia. Torino: Einaudi.
- [2] Asprovska, Marijana, e Nathan Hunter. 2024. «The Tokenization Problem: Understanding Generative AI’s Computational Language Bias». *Ubiquity Proceedings*. <https://doi.org/10.5334/uproc.123>.
- [3] Auerbach, Erich. *Mimesis: il realismo nella letteratura occidentale*. 2 voll. Torino: Einaudi, 1956.
- [4] Baldini, Anna. 2023. *A regola d’arte: storia e geografia del campo letterario italiano (1902-1936)*. Prima edizione. Letteratura tradotta in Italia. Macerata: Quodlibet.
- [5] Balestrini, Nanni, a c. di. 2008. *Quindici: una rivista e il Sessantotto*. Saggi universale economica Feltrinelli. Milano: Feltrinelli.
- [6] Baranelli, Luca, e Grazia Cherchi, a c. di. 1977. *“Quaderni piacentini” 1962-1968*. Milano: Gulliver.
- [7] Barthes, Roland. *Il grado zero della scrittura*. Milano: Lerici, s.d.
- [8] Bazzocchi, Marco A. 2021. “Comico e marginalità”. In *Cento anni di letteratura italiana 1910-2010*, edited by M. A. Bazzocchi. Einaudi.
- [9] Berruto, Gaetano. 1987. *Sociolinguistica dell’italiano contemporaneo*. La nuova Italia Scientifica.

- [10] Bortolotto, Francesco, e Davide Paone. 2018. «Una crepa nel sistema: dalla crisi di “Quindici” alla ricostruzione di “Alfabeta”». In *Sistema periodico. Il secolo interminabile delle riviste*, 189–209. Bologna: Pendragon.
- [11] Bortolotto, Francesco, Eleonora Fuochi, Davide Antonio Paone, e Federica Parodi, a c. di. 2018. *Sistema periodico. Il secolo interminabile delle riviste*. Bologna: Pendragon.
- [12] Bourdieu, Pierre. 1992. *Les Règles de l'art. Genèse et structure du champ littéraire*. Paris: Le Seuil.
- [13] Burrows, J. 2002. «“Delta”: A Measure of Stylistic Difference and a Guide to Likely Authorship». *Literary and Linguistic Computing* 17 (3): 267–87.
<https://doi.org/10.1093/llc/17.3.267>.
- [14] Burrows, J. 2007. «All the Way Through: Testing for Authorship in Different Frequency Strata». *Literary and Linguistic Computing* 22 (1): 27–47.
<https://doi.org/10.1093/llc/fqi067>.
- [15] Cadioli, A. 1998. *La ricezione*. Laterza.
- [16] Cadioli, A., Decleva, E. and Spinazzola, V., eds. 1999. *La mediazione editoriale*. Il Saggiatore/Fondazione Arnoldo e Alberto Mondadori.
- [17] Carpi, Umberto. 1981. *L'estrema avanguardia del Novecento*. Roma: Editori Riuniti.
- [18] Eder, M. 2013. «Mind Your Corpus: Systematic Errors in Authorship Attribution». *Literary and Linguistic Computing* 28 (4): 603–14.
<https://doi.org/10.1093/llc/fqt039>.
- [19] Eder, Maciej. 2015. «Does Size Matter? Authorship Attribution, Small Samples, Big Problem». *Digital Scholarship in the Humanities* 30 (2): 167–82.
<https://doi.org/10.1093/llc/fqt066>.
- [20] Eder, Maciej. 2017. «Short Samples in Authorship Attribution: A New Approach». In *Digital Humanities Conference*.
<https://api.semanticscholar.org/CorpusID:7574620>.
- [21] Eder, Maciej, Jan Rybicki, e Mike Kestemont. 2016. «Stylometry with R: A Package for Computational Text Analysis». *The R Journal* 8 (1): 107.
<https://doi.org/10.32614/RJ-2016-007>.
- [22] Fofi, Goffredo, e Vittorio Giacomini, a c. di. 1998. *Prima e dopo il '68: antologia dei Quaderni piacentini*. 1. ed. Roma: Minimum fax.
- [23] Franzini, Greta, Mike Kestemont, Gabriela Rotari, Melina Jander, Jeremi K. Ochab, Emily Franzini, Joanna Byszuk, e Jan Rybicki. 2018. «Attributing Authorship in the Noisy Digitized Correspondence of Jacob and Wilhelm Grimm». *Frontiers in Digital Humanities* 5 (aprile):4.
<https://doi.org/10.3389/fdigh.2018.00004>.
- [24] Giuliani, Alfredo. 1967. «Le cerimonie sadiche della critica». *Quindici*, 1967.

- [25] Gray, Andrew. 2024. «ChatGPT “contamination”: estimating the prevalence of LLMs in the scholarly literature». arXiv. <https://doi.org/10.48550/ARXIV.2403.16887>.
- [26] Guerriero, Stefano. 2021. «Salotto, laboratorio, dipartimento: la rivista come istituzione letteraria nel secondo Novecento, da “Aretusa” a “Linea d’Ombra”». In *Passeurs. La letteratura italiana del SEcondo Novecento fuori d’Italia: ricezione e immaginario (1945-1989)*. Vol. 49. Liminaires. Bruxelles: Peter Lang.
- [27] Haaf, Susanne, Frank Wiegand, e Alexander Geyken. 2013. «Measuring the Correctness of Double-Keying: Error Classification and Quality Control in a Large Corpus of TEI-Annotated Historical Text». *Journal of the Text Encoding Initiative*, fasc. Issue 4 (marzo). <https://doi.org/10.4000/jtei.739>.
- [28] Holley, Rose. 2009. «How Good Can It Get?: Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs». *D-Lib Magazine* 15 (3/4). <https://doi.org/10.1045/march2009-holley>.
- [29] Italia, Paola. 2013. *Editing Novecento*. Salerno.
- [30] Jauss, Hans R. 1982. *Ästhetische Erfahrung und literarische Hermeneutik*. Suhrkamp.
- [31] Jones, Cameron R., e Benjamin K. Bergen. 2025. «Large Language Models Pass the Turing Test». <https://doi.org/10.48550/ARXIV.2503.23674>.
- [32] Kestemont, Mike. 2014. «Function Words in Authorship Attribution. From Black Magic to Theory?» In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, 59–66. Gothenburg, Sweden: Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-0908>.
- [33] Kichuk, Diana. 2015. «Loose, Falling Characters and Sentences: The Persistence of the OCR Problem in Digital Repository E-Books». *Portal: Libraries and the Academy* 15 (1): 59–91. <https://doi.org/10.1353/pla.2015.0005>.
- [34] Kjell, Bradley, W. Addison Woods, e Ophir Frieder. 1994. «Discrimination of Authorship Using Visualization». *Information Processing & Management* 30 (1): 141–50. [https://doi.org/10.1016/0306-4573\(94\)90029-9](https://doi.org/10.1016/0306-4573(94)90029-9).
- [35] Köbis, Nils, e Luca Mossink. 2020. «Artificial Intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry». arXiv. <https://doi.org/10.48550/ARXIV.2005.09980>.
- [36] Koppel, Moshe, e Yaron Winter. 2014. «Determining If Two Documents Are Written by the Same Author». *Journal of the Association for Information Science and Technology* 65 (1): 178–87. <https://doi.org/10.1002/asi.22954>.
- [37] Liang, Weixin, Zachary Izzo, Yaohui Zhang, Haley Lepp, Hancheng Cao, Xuandong Zhao, Lingjiao Chen, et al. 2024. «Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews». arXiv. <https://doi.org/10.48550/ARXIV.2403.07183>.

- [38] Liang, Weixin, Mert Yuksekgonul, Yining Mao, Eric Wu, e James Zou. 2023. «GPT detectors are biased against non-native English writers». arXiv. <https://doi.org/10.48550/ARXIV.2304.02819>.
- [39] Luperini, Romano. 1981. *Il Novecento*. Torino: Loescher.
- [40] Mangoni, Luisa. 1974. *L'interventismo della cultura. Intellettuali e riviste del fascismo*. Roma-Bari: Laterza.
- [41] Mitchell, Melanie, e David C. Krakauer. 2023. «The Debate over Understanding in AP's Large Language Models». *Proceedings of the National Academy of Sciences* 120 (13): e2215907120. <https://doi.org/10.1073/pnas.2215907120>.
- [42] Muraca, Giuseppe. 1990. «Cronistoria dei 'Quaderni piacentini'». In *Da Il Politecnico a Linea D'ombra*. Poggibonsi: Lalli.
- [43] Patat, Alejandro, e Brigitte Poitrenaud-Lamesi, a c. di. 2021. *Passeurs: La letteratura italiana del Secondo Novecento fuori d'Italia: ricezione e immaginario (1945-1989)*. Brussels: Peter Lang.
- [44] Piazzoni, Irene. 2023. *Il Novecento dei libri. Una storia dell'editoria in Italia*. Roma: Carocci.
- [45] Pontremoli, Giacomo. s.d. *I «Piacentini». Storia di una rivista (1962-1980)*. Edizioni dell'Asino.
- [46] Press, Gil. 2016. «Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says». *Forbes*. 23 marzo 2016. <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>.
- [47] Rebora, Simone. 2023a. «GPT-3 vs. Delta. Applying stylometry to large language models». In *La memoria digitale. Forme del testo e organizzazione della conoscenza*, 292–97.
- [48] Rebora, Simone. 2023b. «Short Texts with Fewer Authors. Revisiting the Boundaries of Stylometry», giugno. <https://doi.org/10.5281/ZENODO.8107473>.
- [49] Rossi, Cesare, e Renuccio Tirelli. 1962. «Due interventi». *Quaderni Piacentini*, Aprile: 25-30.
- [50] Redazione Quaderni Piacentini. 1962. «Editoriale». *Quaderni Piacentini*, marzo.
- [51] Redazione Quindici. 1967. «Editoriale». *Quindici*, giugno.
- [52] Scalia, Gianni, a c. di. 1960. *La cultura italiana del '900 attraverso le riviste*. Torino: Einaudi.
- [53] Sharma, Nishchal, e Ajay Kumar. 2024. «Deep Learning for Stylometry and Authorship Attribution: a Review of Literature». *International Journal for Research in Applied Science and Engineering Technology* 12 (9): 212–15. <https://doi.org/10.22214/ijraset.2024.64168>.
- [54] Spitzer, Leo. 1954. *Critica stilistica e storia del linguaggio*. Gius. Laterza & Figli.

- [55] Stamatatos, Efstathios. 2009. «A Survey of Modern Authorship Attribution Methods». *Journal of the American Society for Information Science and Technology* 60 (3): 538–56. <https://doi.org/10.1002/asi.21001>.
- [56] Stokel-Walker, Chris. 2024. «Chatbot Invasion». *Scientific American* 331 (1): 16. <https://doi.org/10.1038/scientificamerican072024-7bVpuVjnKZbJ8lQrkqMTlR>.
- [57] Van Der Lee, Chris, Albert Gatt, Emiel Van Miltenburg, e Emiel Krahmer. 2021. «Human Evaluation of Automatically Generated Text: Current Trends and Best Practice Guidelines». *Computer Speech & Language* 67 (maggio):101151. <https://doi.org/10.1016/j.csl.2020.101151>.
- [58] Varnum, Michael E. W., Nicolas Baumard, Mohammad Atari, e Kurt Gray. 2024. «Large Language Models Based on Historical Text Could Offer Informative Tools for Behavioral Science». *Proceedings of the National Academy of Sciences* 121 (42): e2407639121. <https://doi.org/10.1073/pnas.2407639121>.
- [59] Ward, Joe H. 1963. «Hierarchical Grouping to Optimize an Objective Function». *Journal of the American Statistical Association* 58 (301): 236–44. <https://doi.org/10.1080/01621459.1963.10500845>.
- [60] Wickham, Hadley. 2014. «Tidy Data». *Journal of Statistical Software* 59 (10). <https://doi.org/10.18637/jss.v059.i10>.
- [61] Zaitsu, Wataru, e Mingzhe Jin. 2023. «Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis». *PLOS ONE* 18. <https://doi.org/10.1371/journal.pone.0288453>.
- [62] Zanotti, Paolo. 2011. *Dopo il primato. La letteratura francese dal 1968 a oggi*. Roma-Bari: Laterza.
- [63] Zinato, Emanuele. 2017. «Editoria e critica». In *Modernità italiana. Cultura, lingua e letteratura dagli anni settanta a oggi*, 75–115. Roma: Carocci.
- [64] Zinato, Emanuele. 2018. «Ciclostilati in proprio: la critica dei “Quaderni piacentini”». In *Sistema periodico. Il secolo interminabile delle riviste*, 173-188. Bologna. Pendragon