

# Analisi stilometrica applicata alle capacità emulative di GPT-4

Marco De Cristofaro<sup>1</sup>, Mariangela Giglio<sup>2</sup>

<sup>1</sup>Université de Mons, C2W, European Postdoctoral programme Marie Curie Cofund Action, Belgio - marco.decrisofaro@umons.ac.be

<sup>2</sup>Università di Bologna, Italia - mariangela.giglio2@unibo.it

## ABSTRACT

La presente ricerca si propone di esplorare la capacità del Large Language Model GPT-4 di emulare lo stile di due riviste culturali italiane attive negli anni '60, «Quaderni Piacentini» e «Quindici». Utilizzando un corpus derivato dai primi numeri delle riviste, trascritti attraverso OCR e poi accuratamente rivisti, il lavoro vuole valutare se GPT-4 sia in grado di eludere l'analisi stilometrica, scrivendo un testo che rispecchi la strategia editoriale di una determinata rivista. Dopo aver utilizzato GPT-4 per generare testi che riflettessero lo stile delle riviste grazie a specifici prompt, è stata condotta un'analisi stilometrica per confrontare i testi artificialmente generati con il corpus originale. I risultati evidenziano le sfide e le potenzialità di GPT-4 nella creazione di scritti con stili complessi. Il confronto 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'utilizzo 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 o sull'indagine di piattaforme di scrittura, applicare queste metodologie all'ambito ristretto delle riviste consentirebbe più ampie considerazioni sulle strategie editoriali, sulle tendenze di lettura e sui processi di circolazione delle idee.

## PAROLE CHIAVE

Stylometry; Authorship attribution; Large Language Models; GPT-4; ChatGPT.

## 1. INTRODUZIONE

Le riviste letterarie, e più genericamente quelle culturali, hanno spesso rappresentato per autori e intellettuali un terreno di confronto e soprattutto un mezzo attraverso cui prendere posizione nel dibattito pubblico. Nel secondo Novecento esse si sono sempre più configurate come «istituzioni» [9] capaci di offrire non solo un canale di accesso al campo di produzione culturale [3], ma anche uno spazio di partecipazione in costante movimento. In un simile orizzonte, negli anni Sessanta in Italia un ruolo di primo piano è stato ricoperto da due riviste, capaci di coniugare la dimensione culturale con precise ambizioni politiche: «Quaderni Piacentini», fondata da Piergiorgio Bellocchio e Grazia Cherchi nel 1962<sup>1</sup>, e «Quindici», fondata a Roma nel 1967 dal Gruppo 63<sup>2</sup>. 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 loro pagine. Uno studio ad ampio raggio sullo stile degli articoli, sulle tematiche affrontate e sulle posizioni assunte dai diversi collaboratori fornirebbe non solo un quadro più completo sui punti di vista dei due giornali, ma anche una più profonda comprensione delle rispettive strategie editoriali, che avrebbe implicazioni sulle ricerche in merito alle modalità di lettura del pubblico. Il ricorso a metodologie stilometriche e a Large Language Models (LLM), come GPT-4, potrebbe così offrire un importante contributo, da una parte alle indagini sulla storia culturale italiana degli anni Sessanta, dall'altra alle ricerche più aggiornate sulla storia della lettura e della stampa.

Il presente lavoro si inserisce in questo contesto, con l'obiettivo di esplorare e valutare le capacità della stilometria nell'ambito dell'individuazione di specificità stilistiche e tematiche di ciascuna rivista, e nel contempo di riflettere sulle performance di GPT-4 applicate, più che allo stile di singoli autori, a un intero corpus editoriale. La ricerca si propone di andare oltre i tradizionali ambiti di applicazione della stilometria, concentrando l'attenzione sulla capacità degli LLM di replicare la somma stilistica di queste pubblicazioni.

Il saggio prende spunto dallo studio di Simone Rebora [16], che ha esaminato la capacità dei modelli linguistici di grande scala, come GPT-3, di «ingannare» le tecniche stilometriche nell'attribuzione autoriale. Tuttavia, mentre Rebora si è concentrato sulla capacità emulativa di singoli autori, questa ricerca si distingue per l'analisi di un fenomeno più specifico, tentando di esaminare la profondità e l'efficacia con cui GPT-4 può assimilare e riflettere lo stile editoriale di riviste culturali.

<sup>1</sup> Sui «Quaderni Piacentini» si vedano almeno le due antologie: «*Quaderni piacentini*» 1962-1968 [2] e *Prima e dopo il '68* [8]. Per una ricostruzione storica delle vicende della rivista si rimanda a [14, 15].

<sup>2</sup> Per le coordinate storiche di «Quindici» si rimanda all'antologia *Quindici: una rivista e il Sessantotto* [1].

## 2. CREAZIONE DEL CORPUS DI RIFERIMENTO

Una prima complessità all'interno del nostro studio riguarda la scelta del corpus di riferimento. È evidente, infatti, come, in virtù del loro ruolo chiave nella diffusione delle idee e nel dibattito pubblico, le riviste italiane del Novecento siano tanto numerose quanto variegate. Dal momento che spesso esse condividono scopi politico-culturali, forme espressive (racconti, reportage, articoli di critica ecc.) e approcci (commistione tra testi e immagini, utilizzo di traduzioni, coinvolgimento dei lettori) ci siamo chiesti se fosse possibile delimitare linee di demarcazione tra le diverse identità delle singole riviste. Da questa prima necessaria verifica discende la seconda criticità del nostro studio: trattandosi di testi di autori diversi, è possibile identificare elementi ricorrenti tra i vari contenuti di una sola rivista tali da determinarne la strategia editoriale? L'individuazione dei due giornali di riferimento dovrebbe rispondere, dunque, a questi specifici criteri di indagine. La scelta è ricaduta su «Quaderni Piacentini» e «Quindici» per diverse ragioni. In primo luogo, entrambe condividono, come emerge fin dalle dichiarazioni programmatiche dei rispettivi numeri iniziali, l'ambizione a considerarsi ad un tempo una rivista culturale ma anche uno spazio di intervento politico. Il secondo motivo è la prossimità temporale: è vero, infatti, che il primo numero di «Quaderni Piacentini» esce nel 1962 e quello di «Quindici» nel 1967, ma entrambi affrontano tematiche riconosciute come rappresentative del contesto sociale degli anni Sessanta prima del 1968. Inoltre, la somiglianza delle tematiche avrebbe garantito una maggiore centralità dell'analisi prettamente stilistica quale tratto dirimente della polarizzazione tra le due riviste: l'indagine stilometrica, in altre parole, sarebbe stata maggiormente valida, perché basata su aspetti stilistici e non influenzata da un'eccessiva distanza degli argomenti trattati. Infine, abbiamo deciso di concentrarci sui numeri iniziali delle riviste perché essi contengono dichiarazioni programmatiche, idealmente finalizzate a offrire al lettore una linea editoriale.

### 2.1 OCR e accuratezza

Una volta identificati i testi di riferimento si è provveduto alla trascrizione degli stessi: per quanto riguarda «Quaderni Piacentini» ci si è basati sull'antologia a cura di Luca Baranelli e Grazia Cherchi; per «Quindici» ci siamo affidati al volume curato da Nanni Balestrini. Sono stati dunque trascritti nell'interezza i primi tre numeri di ciascuna rivista per un totale di circa 25mila parole per «Quindici» e 22mila circa per «Quaderni Piacentini»<sup>3</sup>. Per garantire livelli di accuratezza quanto più possibile elevati è stato scelto di combinare tecniche OCR e forme di controllo manuale. 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 [11] l'accuratezza media dell'OCR senza correzione può variare fra il 71% e il 98,02%. Al livello 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) per le applicazioni di ricerca è di almeno il 99,95%. La trascrizione a quattro mani garantisce un'accuratezza del 99,997% [10]; 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 [12].

## 3. 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. Se da un lato sistemi di chatbot come ChatGPT possono essere validi alleati per il ricercatore, d'altro canto le problematiche legate a un loro uso massiccio emergono sempre con maggior chiarezza. ChatGPT (basato sulla tecnologia *Generative Pre-trained Transformer*) è in grado di generare testi simili a quelli umani in un'ampia varietà di stili e di lingue, con una fluency tale che la maggior parte delle persone non riesce a distinguere tra testi generati da AI e quelli scritti da esseri umani [13, 17]. Si è lungamente discusso della difficoltà di identificare con certezza testi scritti da AI, tanto che OpenAI il 31 gennaio 2023 ha rilasciato un classificatore di testi AI. È la stessa OpenAI, tuttavia, ad ammettere i limiti dell'accuratezza di rilevamento di questo classificatore, che presenta oltretutto prestazioni più basse per le lingue diverse dall'inglese. Sembra particolarmente interessante, dunque, valutare l'effettiva capacità emulativa di GPT in relazione non tanto al singolo autore (come già esplorato da Rebora [16]) ma rispetto alla

<sup>3</sup> Trattandosi di un lavoro che rientra in una ricerca più ampia, il corpus revisionato sarà reso accessibile, il prima possibile, all'interno del seguente repository github: <https://github.com/MarcoDeCristofaro/RivisteAnni60.git>.

somma stilometrica di un'intera rivista, seppure con uno stile peculiare e ampiamente riconoscibile. Si è scelto di utilizzare, a questo fine, il modello GPT-4 per la sua aumentata capacità di calcolo unitamente alla possibilità di allegare blocchi testuali direttamente nell'interfaccia di chat.

#### 4. DEFINIZIONE DEL TEST SET TRAMITE GPT-4

La definizione del test set ha rappresentato una fase cruciale per valutare l'abilità di GPT-4 nell'attività emulativa. Il corpus delle riviste è stato accuratamente suddiviso in blocchi di testo, ciascuno contenente un minimo di 5.000 parole. La soglia è stata stabilita in base a criteri stilometrici consolidati [6], che identificano questa lunghezza come necessaria per garantire analisi significative e affidabili. Ogni blocco di testo è stato presentato a GPT-4 tramite l'interfaccia di chat. L'intento era di sfruttare le avanzate capacità di elaborazione del linguaggio del modello per generare articoli che non solo rispecchiassero lo stile, ma anche le tematiche intrinseche dei testi 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 suggestionare eccessivamente l'elaborazione da parte di GPT, è stato necessario modificare leggermente i parametri per impedire sia l'esatto ripetersi dei risultati sia la creazione di testi poco coerenti. Seguendo le raccomandazioni della stessa *API section* di OpenAI<sup>4</sup> si è operato principalmente e in maniera alternata sulla temperatura<sup>5</sup> e top p<sup>6</sup>. Per il primo test-set di testi si è settata una temperatura di 0.8. Questo parametro, tuttavia, creava dei testi eccessivamente simili tra loro, difficilmente utilizzabili per le analisi stilometriche. Dopo una serie di prove si è scelto dunque di mantenere la temperatura consigliata di 1.0, la stessa utilizzata nell'interfaccia online di GPT 4. 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 giudicare 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 si è cercato di 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 adress similar topics, take the same political and cultural positions taken in the articles you analized. Try to emulate the articles and mirror the style, stylistic feature as much as possible. Language of the results "Italian" lenght of the results: "1000 words" minimum*».

Si è già lungamente discussa l'incapacità di GPT di contare (non soltanto i token), ma si è scelto comunque di mantenere un'indicazione di massima all'interno del prompt. La generazione testuale è stata ripetuta fino a raggiungere i testi 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. La risposta a una simile criticità è in parte imputabile al sistema di funzionamento di GPT, che sfrutta un approccio di tipo probabilistico per la generazione delle risposte. 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.

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.

<sup>4</sup> <https://platform.openai.com/docs/api-reference/chat/create>

<sup>5</sup> 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.

<sup>6</sup> 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.

## 5. ANALISI STILOMETRICA

Una prima indagine è stata dunque effettuata su un insieme di otto testi, suddivisi in quattro testi per le due riviste. Occorre precisare che i testi non corrispondono a uno specifico articolo ma rappresentano una somma di articoli differenti. Considerando, infatti, che il numero minimo di parole per un'analisi stilometrica affidabile è di circa 5000, abbiamo ritenuto necessario raccogliere gli articoli dei primi due numeri di ciascuna rivista in quattro differenti insiemi. In questo modo abbiamo ottenuto un testo sufficientemente lungo per l'analisi stilometrica e allo stesso tempo capace di rispecchiare una visione editoriale specifica e non frammentata della rivista.

Per la prima fase ci siamo affidati all'applicazione di tecniche stilometriche attraverso il linguaggio di programmazione R, sulla base della metodologia sviluppata da Maciej Eder e Jan Rybicki [7]. La metrica di riferimento iniziale è stata "Delta" [5]. Attraverso l'utilizzo di una serie di *features* differenti, siamo stati in grado di individuare possibili traiettorie distintive tra le due riviste. Per la visualizzazione dei primi risultati ci siamo affidati alla creazione di dendrogrammi [18], che permettono di raccogliere i testi 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, abbiamo portato avanti un confronto tra le riviste. Applicando la Zeta Analysis [5], abbiamo indagato la probabilità con cui determinate scelte lessicali possano rappresentare la visione delle sedi di pubblicazione.

### 5.1. Un confronto stilistico-tematico: «Quaderni Piacentini» e «Quindici»

Una prima analisi è stata necessaria a rilevare il numero di *most frequent words* (MFW), capace di offrire i risultati più stabili. È stata impostata la *feature* di incremento progressivo "mfw.incr=50", individuando nelle 150mfw il valore di maggiore stabilità.

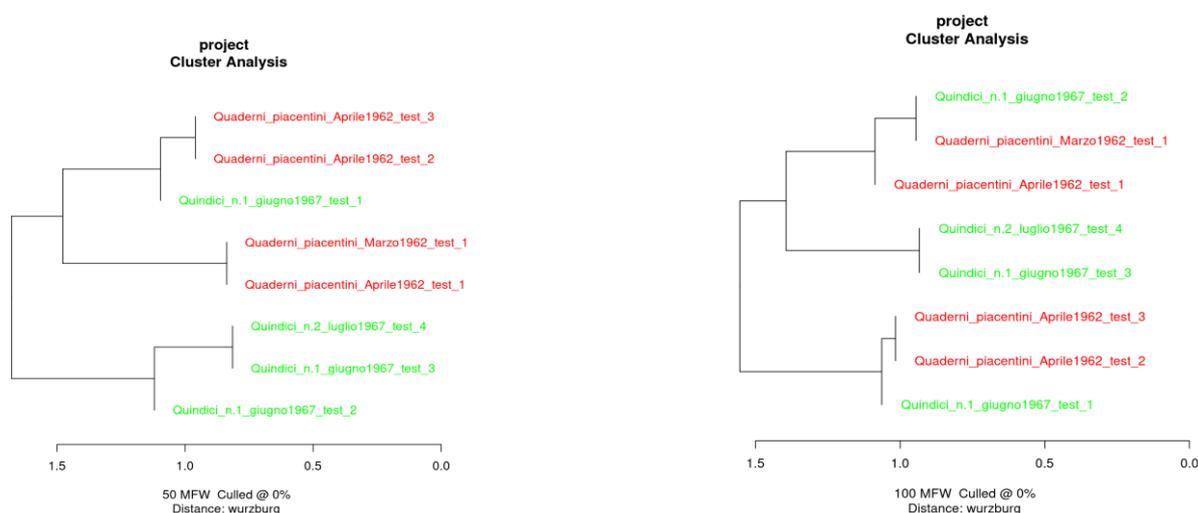


Figura 1. Cluster Analysis sul training set 50-100MFW

I risultati (vd. Fig. 1- Fig. 2) mostrano come le riviste si siano disposte sostanzialmente in due rami differenti evidenziando approcci distinti. Due casi significativi di contaminazione a livello stilometrico sono rappresentati dai testi estratti dalla seconda parte del secondo numero di «Quaderni Piacentini», pubblicato nell'aprile del 1962, e dalla seconda parte del primo numero di «Quindici», pubblicato nel giugno 1967. Procedendo con un'ulteriore analisi di comparazione con il metodo Zeta (vd. Fig. 3), che mette in opposizione a livello lessicale e in modo specifico i testi relativi alle due riviste, i risultati evidenziano una ben più chiara distanza rispetto ai termini maggiormente utilizzati. Nella Zeta Analysis le scelte lessicali individuano approcci ben più marcati, tracciando con maggior nettezza la posizione di ciascuna rivista. Di particolare rilievo in questo caso è la distanza tra i campi semantici preferiti dall'una e dall'altra rivista: mentre «Quaderni Piacentini» manifesta un maggior ricorso a termini di ambito politico come "lotta", "comunisti", "democrazia", "compagni", "socialista", "fascista", «Quindici» riflette gli interessi culturali dei suoi fondatori prediligendo una terminologia tipica della critica letteraria con parole quali "ipotesi", "lettore", "avanguardia".

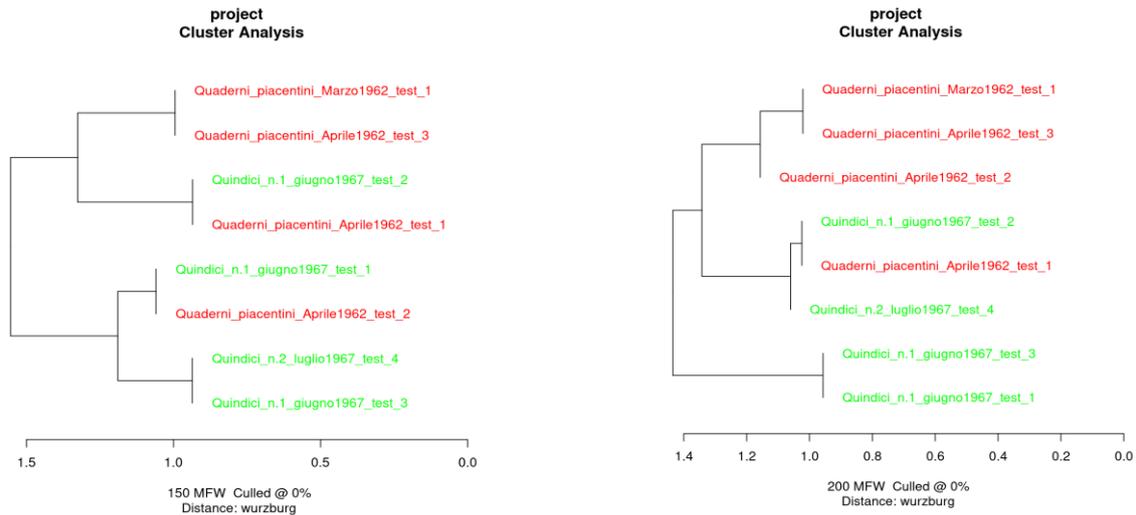


Figura 2. Cluster Analysis sul training set 150-200MFW

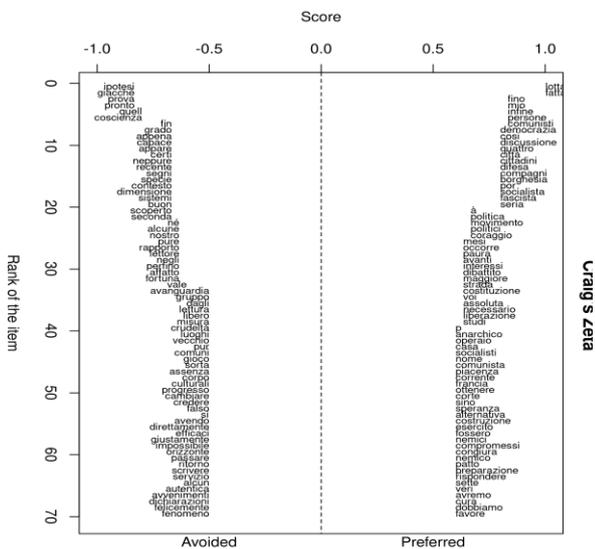


Figura 3. Zeta Analysis tra "Quaderni Piacentini" e "Quindici"

Se, dunque, un primo approccio complessivo sembra evidenziare una maggiore prossimità tra le due visioni politico-culturali, l'indagine lessicale esprime una più evidente contrapposizione, che sarà fondamentale tener presente al momento del confronto tra i testi originali e quelli prodotti da Large Language Models.

### 5.2. Stilometria su testi AI generated

Una volta appurata la specificità stilistica delle singole riviste, è stato possibile inserire nell'analisi anche i testi generati da GPT-4. Si sono dunque creati due blocchi testuali contenenti l'interezza degli articoli scritti da GPT per ogni rivista, in modo da poter efficacemente valutare il lavoro emulativo.

I risultati mostrano chiaramente la riconoscibilità della cifra stilistica dei testi prodotti con GPT-4. Alla bipartizione dell'albero già evidenziata in Fig. 1 si affianca un terzo ramo che pone in evidente separazione i testi AI generated, mantenendo visibile la contaminazione di cui sopra (vd. Fig. 4)

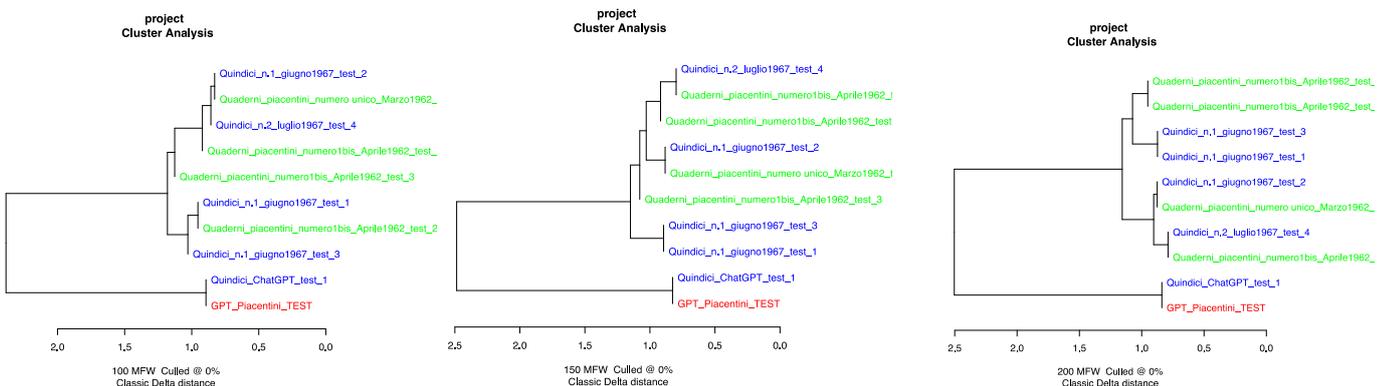


Figura 4. Cluster Analysis riviste e testi GPT-4

Aumentando il numero delle most frequent words si nota in modo ancora più evidente la corretta attribuzione. Non diversi sono i risultati con l'applicazione della Zeta Analysis (vd. Fig. 5). 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 a cui ricorre l'LLM troviamo: società, ruolo, sfide, cambiamento, contesto, arte, futuro, culturale, sociale. Si tratta di termini portatori di significato e dal grande peso

semantico. Se essi risultano tanto utilizzati da GPT in relazione alle riviste, si può ipotizzare che il modello, nel tentativo di imitare lo stile dei giornali culturali, si affidi maggiormente ai nuclei tematici dei testi più che alle function words. L'aspetto tematico di conseguenza prevale su quello stilistico nell'addestramento del modello ma è, allo stesso tempo, il principale artefice del suo fallimento imitativo: lo sforzo di concentrarsi, infatti, sulle parole significative non è premiato, anzi, appare come l'elemento che causa la distanza tra i due approcci alla descrizione della realtà circostante.

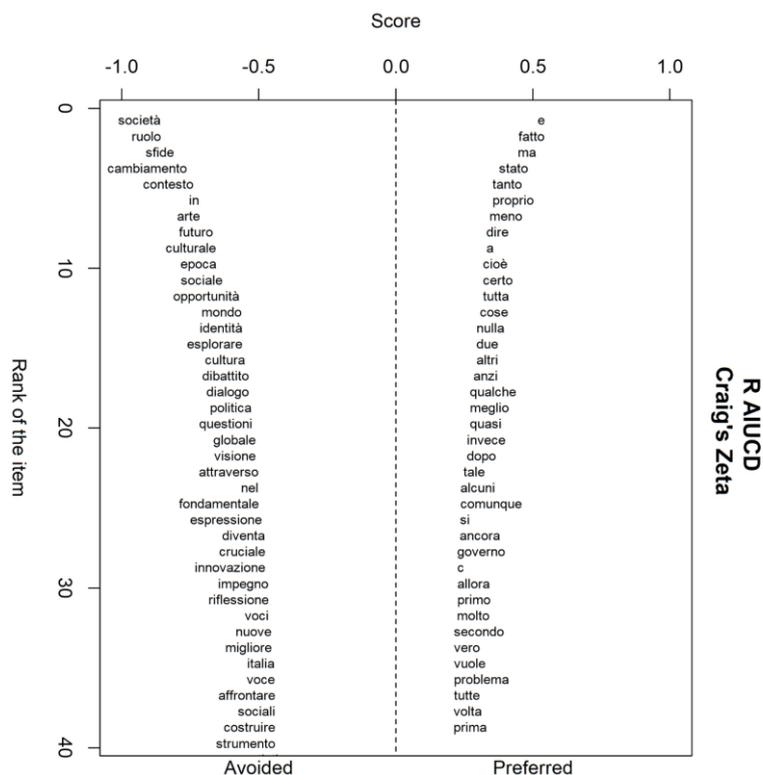


Figura 5. Zeta Analysis tra riviste e ChatGPT-4

LLM idealmente finalizzati alla pubblicazione su una sede – il giornale letterario appunto – potrebbe portare a percorsi di approfondimento diversi intorno alle modalità di addestramento degli LLM stessi, alle loro attuali capacità generative e all'insieme di tecniche adottate sul loro sviluppo: individuare e riprodurre una strategia editoriale, d'altro canto, rappresenta una sfida ancora complessa e un terreno di studio da sondare ed esplorare prima di tutto nel campo dei *cultural studies*. In questo senso, potrebbero venire in soccorso tecniche di analisi computazionale, capaci di riconoscere tratti dirimenti tra le diverse posizioni editoriali attraverso uno sguardo d'insieme. Ecco perché è auspicabile che l'analisi stilometrica si rivolga in futuro, come solo in parte ha già fatto, anche a miscellanee che vadano dalle riviste alle raccolte collettive di opere di diversi autori e si affianchi a un approfondimento parallelo della capacità imitativa degli LLM rispetto a insiemi complessi di testi: un simile approccio potrebbe, infatti, condurre a importanti novità in quei vasti campi di indagine che sono la storia della lettura e la teoria della ricezione.

## 7. RINGRAZIAMENTI

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

## BIBLIOGRAFIA

- [1] Balestrini, Nanni. *Quindici: una rivista e il Sessantotto*. Milano: Feltrinelli, 2008.
- [2] Baranelli, Luca, e Grazia Cherchi. *Quaderni piacentini 1962-1968*. Milano: Gulliver, 1977.
- [3] Bourdieu, Pierre. *Les Règles de l'art. Genèse et structure du champ littéraire*. Paris: Editions du Seuil, 1992.
- [4] Burrows, John. «All the Way Through: Testing for Authorship in Different Frequency Strata». *Literary and Linguistic Computing* 22, fasc. 1 (2006): 27-47. <https://doi.org/10.1093/lilc/fqi067>.

- [5] Burrows, John. «“Delta”: a Measure of Stylistic Difference and a Guide to Likely Authorship». *Literary and Linguistic Computing* 17, fasc. 3 (2002): 267-287.
- [6] Eder, Maciej. «Does Size Matter? Authorship Attribution, Small Samples, Big Problem». *Digital Scholarship in the Humanities* 30, fasc. 2 (2013): 167-182. <https://doi.org/10.1093/lc/fqt066>.
- [7] Eder, Maciej, Jan Rybicki, e Mike Kestemont. «Stylometry with R: A Package for Computational Text Analysis». *R Journal* 8, fasc. 1 (2016): 107. <https://doi.org/10.32614/rj-2016-007>.
- [8] Fofi, Goffredo, e Vincenzo Giacobini. *Prima e dopo il '68. Antologia dei Quaderni Piacentini*. Roma: minimum fax, 1998.
- [9] Guerriero, Stefano. «Salotto, laboratorio, dipartimento: la rivista come istituzione letteraria nel secondo Novecento, da “Aretusa” a «Linea d’Ombra». In *La letteratura italiana del Secondo Novecento fuori d’Italia: ricezione e immaginario (1945-1989)*, a cura di Alejandro Patat e Brigitte Poitrenaud Lamesi. Bruxelles: Peter Lang, 2021.
- [10] Haaf, Susanne, Frank Wiegand, e Alexander Geyken. «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* 4 (2013). <https://doi.org/10.4000/jtei.739>.
- [11] Holley, Rose. «How Good Can It Get?» *D-Lib Magazine* 15, fasc. 3/4 (2009). <https://doi.org/10.1045/march2009-holley>.
- [12] Kichuk, Diana. «Loose, Falling Characters and Sentences: The Persistence of the OCR Problem in Digital Repository E-books». *Libraries and the Academy* 1 (2015): 59-91. <https://doi.org/10.1353/pla.2015.0005>.
- [13] Köbis, Nils, e Luca D. Mossink. «Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry». *Computers in Human Behavior* 114 (2021): 106-553.
- [14] Muraca, Giuseppe. «Cronistoria dei ‘Quaderni piacentini’». In *Da “Il Politecnico” a “Linea D’ombra”*. Poggibonsi: Lalli, 1990.
- [15] Pontremoli, Giacomo. *I “Piacentini”. Storia di una rivista (1962-1980)*. Roma: Edizioni dell’asino, 2017.
- [16] Reborà, Simone. «GPT-3 vs. Delta. Applying stylometry to large language models». In *La memoria digitale: forme del testo e organizzazione della conoscenza. XII Convegno Annuale AIUCD, Siena, 5-7/06/2023. Proceedings*, a cura di Emmanuela Carbé, Gabriele Lo Piccolo, Alessia Valenti, e Francesco Stella, 292-297, 2023. <https://doi.org/10.6092/unibo/amsacta/7721>.
- [17] Van der Lee, Chris, Albert Gatt, Emiel Van Miltenburg, e Emiel Kraemer. «Human evaluation of automatically generated text: Current trends and best practice guidelines». *Computer Speech & Language* 67 (2021): 101151. <https://doi.org/10.1016/j.csl.2020.101151>.
- [18] Ward, Joe H. «Hierarchical Grouping to Optimize an Objective Function». *Journal of the American Statistical Association* 58, fasc. 301 (1963): 236-244. <https://doi.org/10.1080/01621459.1963.10500845>.