

RAFFAELE MANNA, ANTONIO PASCUCCI, JOHANNA MONTI

## La misurazione stilistica della Falsificazione.

### I comunicati delle Brigate Rosse

### The Measure of Forgery. The statements of the Red Brigades

In Italy, during the so-called *Years of Lead*, the Red Brigades disseminated several written statements, in which they summarized the ideological intentions and the plan for the *Armed Struggle*. In addition to the propaganda dissemination, the Red Brigades reported about crimes and kidnappings of personalities from industry and politics using the form of written communication. In 1978, the Red Brigades kidnapped Aldo Moro giving news through a first statement, which was followed by nine other statements, one of which represents a fake. In the present study we show the possibilities of linguistic investigation in forensic science using Computational Stylometry techniques. Taking as an example the statements that appeared during the Moro Case, it will be shown *a posteriori* to what extent the fake statement is different from the original ones. Furthermore, through the stylometric analysis of the statements, we show which are the stylistic and lexical characteristics that distinguish the group of Red Brigades statements from the fake one.

*Keywords:* Computational Linguistics, Computational Stylometry, Digital Forensics, Red Brigades, Text Analysis.

#### 1. Introduzione

Nel corso degli anni, l'applicazione di metodi e tecniche proprie della linguistica per l'analisi del linguaggio parlato e scritto è stata progressivamente riconosciuta come saliente per fornire ulteriori evidenze e prove da aggiungere ad altri elementi probatori in ambito giurisdizionale. Nello specifico, l'intersezione di linguistica e discipline forensi è stata nominata come Linguistica Forense o, in alcuni casi, Linguistica Giudiziaria (Romito, 2013: 178). In questa, relativamente nuova, disciplina sono state riconosciute una serie di direzioni di ricerca, in gran parte ancorate all'applicazione della metodologia linguistica in diverse dimensioni forensi. In Gibbons e Turell (2008: 1) vengono descritte cinque categorie principali: una prima pertinente all'analisi del linguaggio legislativo; una seconda al linguaggio parlato nelle aule giudiziarie; una terza relativa all'elaborazione di evidenze; una quarta sull'insegnamento e nella ricerca del linguaggio legale scritto e parlato e infine, un'ultima categoria, sulla traduzione e interpretazione del linguaggio in ambito legale.

Nello specifico, per quanto riguarda l'elaborazione di evidenze linguistiche a partire da testi o parlato per fini investigativi e giudiziari, larga parte della ricerca e della pratica investigativa del linguista come perito si è concentrata sull'attribuzio-

ne e sulla descrizione del profilo autoriale a partire dalle caratteristiche linguistiche presenti in un testo e sul riconoscimento del parlante in casi di registrazione vocale (Coulthard, Johnson, 2007). In questo contesto e non restringendo il campo alla sola analisi di dati testuali in pratiche forensi, le metodologie e le tecniche appartenenti alla Linguistica Computazionale e al Trattamento Automatico del Linguaggio Naturale sono risultate adeguate nell'estrazione di caratteristiche linguistiche da testi e affidabili nell'attribuzione di testi ad un autore o ad una serie di autori diversi di paternità incerta (Koppel, Schler & Argamon, 2009; Rocha, Scheirer, Forstall, Cavalcante, Theophilo, Shen, Carvalho & Stamatatos, 2016) come anche nell'estrazione di caratteristiche sociolinguistiche e caratteristiche demografiche dai testi per la conseguente attribuzione degli stessi ad un profilo autoriale (Kocher, Savoy, 2017; Weren, Kauer, Mizusaki, Moreira, De Oliveira & Wives, 2014). Queste tecniche si avvalgono di componenti statistiche per l'analisi e l'estrazione di caratteristiche linguistiche presenti all'interno del testo o dei testi e, spesso, comprendono l'uso di algoritmi di classificazione e *clustering* per identificarne l'autore, con un certo grado di probabilità, in un dato insieme di autori conosciuti e in lizza per l'attribuzione o, nel caso in cui il possibile autore o autori non siano conosciuti, per attribuire il testo ad un profilo autoriale. In particolare, in compiti di attribuzione e profilazione autoriale, vengono spesso utilizzati metodi computazionali in grado di formalizzare ed estrarre caratteristiche stilistiche capaci di descrivere lo stile autoriale, ovvero la marca stilistica con la quale un testo o un insieme di testi sono stati scritti da un certo autore (Daelemans, 2013). Questo insieme di caratteristiche stilistiche viene definito e descritto nella ricerca e nei metodi appartenenti alla stilometria computazionale, ovvero come la misurazione quantitativa delle caratteristiche di stile. In stilometria, queste caratteristiche stilistiche sono intesi come tratti di scrittura fondamentali caratterizzanti lo stile di uno specifico autore. La misurazione quantitativa di caratteristiche stilistiche avviene ad ogni livello di analisi linguistica, a partire da tratti morfologici, lessicali e sintattici fino a tratti strutturali con i quali l'autore organizza la materia linguistica nei testi. Sebbene queste caratteristiche linguistiche e stilistiche compongano la quasi totalità dei testi scritti in una data lingua, un insieme delle stesse è in grado di descrivere e definire lo stile specifico di un autore che lo rende distinguibile da altri (Grant, Baker, 2001). In letteratura, ci si riferisce all'insieme di caratteristiche linguistiche e stilistiche proprie di un autore con espressioni come *impronta stilistica* o *DNA autoriale* (Daelemans, 2013; Eder, 2011) e con il termine *idioletto* (Coulthard, 2004). Oltre a ciò, le stesse caratteristiche stilistiche uniche e proprie di un autore sono state talvolta oggetto di tentativi di imitazione e, spesso, si sono verificati casi di falsificazione e interpolazione di testi (Heiser, 2007; Pearl, Steyvers, 2012).

In questo articolo, si discuteranno e si mostreranno le possibilità di applicazione di metodi e tecniche di stilometria computazionale in uno storico caso di interesse forense risalente a quelli che in Italia sono tristemente noti come gli *anni di piombo* (*fine anni Sessanta – inizio anni Ottanta*). Si prenderanno come caso di studio i nove comunicati delle Brigate Rosse (BR) insieme al falso comunicato, apparso il

18 Aprile 1978, diffusi durante il caso Moro. Si mostreranno *a posteriori* quali siano le caratteristiche stilistiche e linguistiche che distinguono il gruppo dei comunicati autentici delle BR rispetto al comunicato falso (denominato di seguito FBR7), utilizzando tecniche e metriche della stilometria computazionale. Inoltre, tramite un'analisi lessicale, si mostrerà quanto il falso comunicato attesti una lontananza e deviazione nel progresso ideologico e stilistico-linguistico rispetto alle intenzioni programmatiche delle BR.

Il presente articolo è strutturato come segue: nel secondo paragrafo si fornisce una breve narrazione della vicenda storica seguita dalla descrizione delle prime analisi condotte da linguisti ed esperti sui comunicati delle BR; nel terzo paragrafo si presenterà il corpus dei comunicati delle BR, insieme a statistiche descrittive; nel quarto paragrafo si descriverà la metodologia utilizzata per condurre le analisi e si mostreranno le caratteristiche stilometriche rilevate durante la comparazione linguistica del gruppo dei comunicati autentici a confronto con il comunicato falso; poi il paragrafo cinque sarà dedicato alle differenze lessicali tra due gruppi di comunicati autentici delle BR, tenendo come elemento separatore il comunicato falso; infine l'articolo si chiuderà con le conclusioni e la discussione delle possibilità offerte dalle stilometria computazionale in analisi testuali forensi.

## *2. I Comunicati delle Brigate Rosse e le prime analisi linguistiche*

Durante la mattina del 16 Marzo 1978, due auto furono aggredite da una banda di terroristi in via Fani a Roma. I terroristi non risparmiarono i passeggeri della seconda auto e stessa sorte toccò ad alcuni passeggeri della prima auto. Su questa, viaggiava il Presidente della Democrazia Cristiana (DC) ed ex Presidente del Consiglio Aldo Moro che fu catturato. Per i successivi 55 giorni, Aldo Moro venne tenuto ostaggio dai rapitori in un luogo segreto.

Il 18 Marzo del 1978 venne diffuso a Roma un comunicato nel quale l'organizzazione terroristica di estrema sinistra, le Brigate Rosse (BR), rivendicò il rapimento di Aldo Moro, dando il via alla lotta armata e allo "scontro di classe".

Fin dal giorno seguente l'apparizione del primo comunicato, la macchina investigativa linguistica si attivò. Il primo a condurre un'analisi di tipo linguistico-filologico fu Tullio De Mauro con un articolo dal titolo "Tentativo di lettura filologica del messaggio BR. Non è come gli altri: sembra tradotto dal francese" pubblicato il 19 Marzo 1978 sul quotidiano *Paese Sera* (Marchetti, 2017: 53). In questo articolo, il celebre linguista ipotizza l'influsso della lingua francese con calchi e interferenze nel lessico usato dalle BR. Successivamente a questo primo articolo di analisi linguistica, un secondo articolo firmato da Arrigo Levi apparve sul quotidiano *La Stampa* dal titolo "L'analisi del volantino dei 'brigatisti' trovato a Roma. C'è un'ipotesi latino-americana" in data 21 Marzo 1978. Come titola quest'ultimo articolo, in questo viene avanzata un'ipotesi di interferenza linguistica con la lingua spagnola.

Durante i giorni del sequestro, a questo primo comunicato seguirono altri nove comunicati, tra i quali uno falso e poi smentito dalle stesse BR. Infatti, il comunicato

numero 7 del 18 Aprile 1978 dava notizia dell'avvenuta esecuzione di Aldo Moro e forniva un'indicazione dettagliata ed esatta del luogo nel quale la salma giaceva, ovvero nei fondali del Lago della Duchessa. Un paio di giorni dopo, il 20 Aprile 1978, un altro comunicato numero 7 apparve a Roma, Genova, Milano e Torino. In questo comunicato, le Brigate Rosse presero le distanze dal comunicato 7 del 18 Aprile e smentirono l'avvenuta uccisione di Aldo Moro allegando una foto dello stesso in vita e con una copia di un giornale datato 19 Aprile. In seguito a questo comunicato, le BR diffonderanno altri tre comunicati. Il 5 Maggio 1978 apparve l'ultimo comunicato, il numero 9, nel quale le BR diedero notizia dell'avvenuta condanna di Aldo Moro, destabilizzante per l'apparato statale e la DC e chiusero così la trattativa con le parti coinvolte. Il 9 Maggio 1978, alla segreteria della DC tramite una telefonata venne riferito il luogo esatto nel quale trovare il cadavere di Aldo Moro. Il cadavere era stato adagiato in una Renault 4 parcheggiata in via Caetani a Roma.

In seguito alla tragica conclusione dei fatti, l'interesse per la materia linguistica contenuta sia nei comunicati sia nelle lettere inviate da Moro nel periodo della prigionia, non si affievolì. Infatti, Leonardo Sciascia propose una lettura e un'analisi delle lettere di Moro ipotizzando l'intenzione del prigioniero di camuffare nel codice linguistico alcune indicazioni sulla posizione della sua prigionia (Sciascia, 1983). Sciascia, poi, fu chiamato a far parte della commissione parlamentare d'inchiesta costituita per il sequestro e l'omicidio di Moro. Proprio alla commissione d'inchiesta, nel 1981, venne inviato un dossier contenente un'analisi linguistica sui comunicati delle BR apparsi durante la prigionia di Moro. L'autore del dossier era Renzo Rota, diplomatico e consigliere all'ambasciata italiana a Mosca dal 1965 al 1972. Nel dossier, Rota ipotizzò che i primi due comunicati brigatisti fossero stati scritti da un sovietico adducendo come prove linguistiche determinati calchi fraseologici dal russo molti comuni nella retorica di alcuni quotidiani sovietici, ovvero *Pravda* e *Kommunist*. Anni dopo, nel 1990, apparve un articolo di Mario Medici (Medici, 1990) nel quale, oltre alla discussione sui diversi comunicati delle BR apparsi prima e dopo il caso Moro, lo studioso propose una breve discussione sul possibile numero di autori e sugli autori stessi dei comunicati ed espose un'analisi sui tipi e sulle principali strategie retoriche messe in atto. Oltre a ciò, propose un'analisi più incentrata sul tipo di linguaggio utilizzato nei comunicati, illustrando alcuni esempi di linguaggio tecnico e settoriale per quanto riguarda i comunicati apparsi durante il caso Moro.

Con riferimento a ricerche apparse nell'ultimo decennio, un articolo (Re, 2013) propone un'analisi linguistica dei comunicati delle BR durante il caso Moro. In questo articolo vengono sottolineate sia caratteristiche di continuità lessicale che di struttura morfosintattica ricorrenti nei comunicati, osservando che i documenti diffusi dalle BR sono dotati di una struttura unitaria. Di unitarietà linguistica e stilistica attraverso i comunicati si discute anche in (Marchetti, 2017). In questo articolo, vengono offerte alcune analisi sul registro linguistico con il quale i comunicati sono scritti e su alcune figure retoriche utilizzate per organizzare l'argomentazione. Inoltre, si propongono alcune analisi lessicali basate sulla frequenza nei comunicati,

ribadendo sia l'uniformità linguistica che l'abilità comunicativa nel gestire toni e registri formali, tecnici ed espressivi.

### 3. *Il Corpus dei comunicati*

Il corpus testuale analizzato in questo articolo è composto dai nove comunicati autentici inviati dalle BR durante il sequestro Moro con l'aggiunta del comunicato 7, prodotto di una falsificazione.

Di seguito, si farà riferimento ai singoli comunicati utilizzando una sigla di riferimento composta da BR seguita dal numero crescente per indicare l'ordine di apparizione dei comunicati (BR1, BR2, ...BR9), mentre (come detto già in precedenza) si indicherà il comunicato falso come FBR7. I testi dei comunicati sono stati estrapolati manualmente<sup>1</sup> e divisi in file testuali singoli con titolo composto dalla sigla di riferimento del comunicato e il numero separato da trattino basso (br\_1, br\_2, etc.). Per tutti i comunicati, sono state conservate le caratteristiche formali, ovvero parole in maiuscolo e spazi tra i paragrafi.

I comunicati sono stati poi analizzati utilizzando il pacchetto *Stylo*<sup>2</sup> (Eder, Rybicki & Kestemont, 2016) disponibile per il linguaggio R. Questo pacchetto offre gli strumenti e le metriche necessarie per condurre analisi e comparazioni di tipo stilometrico su testi.

Il corpus indagato in questo articolo è composto da un totale di 9.332 parole. Nella Tabella 1 si riportano le lunghezze testuali in termini di parole relative ad ogni comunicato, tenendo in considerazione solamente stringhe composte da caratteri alfabetici.

Tabella 1 - *Lunghezza testuale dei comunicati*

<i>br_1</i>	<i>br_2</i>	<i>br_3</i>	<i>br_4</i>	<i>br_5</i>	<i>br_6</i>	<i>br_7</i>	<i>br_8</i>	<i>br_9</i>	<i>fbr_7</i>
739	1.342	1.023	<b>1.355</b>	675	952	1.136	910	1.051	<b>149</b>

In Tabella 1 sono evidenziate, in grassetto, la lunghezza maggiore raggiunta dal comunicato BR4 e la lunghezza minore riferita al FBR7. Già da questa analisi preliminare risulta una disparità in termini di parole utilizzate nel gruppo dei comunicati autentici e il falso comunicato. Infatti, oltre al BR4, anche gli altri comunicati prodotti dalle BR risultano essere composti da più parole rispetto al comunicato falso.

### 4. *Analisi dei Cluster e caratteristiche stilometriche*

In questa sezione si descriveranno gli esperimenti effettuati per analizzare il corpus dei comunicati delle BR e il comunicato FBR7. Nello specifico, si discuterà l'applicazione

<sup>1</sup> Il testo integrale di ogni comunicato è stato estrapolato dal sito web seguente: <http://www.archivio900.it/it/documenti/finestre-900.aspx?c=423>. Per ogni comunicato si è copiato solamente il testo escludendo la glossa iniziale a cura della redazione del sito.

<sup>2</sup> <https://cran.r-project.org/web/packages/stylo/index.html>

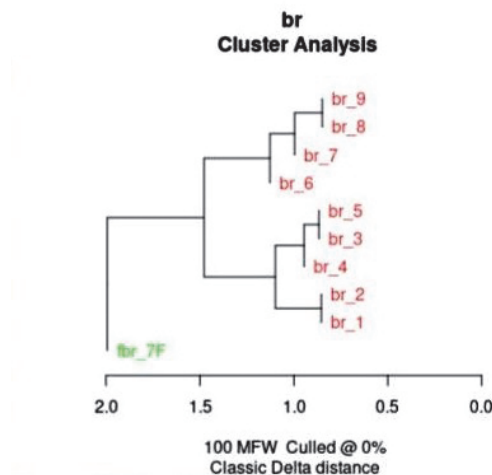
cazione di metodologie stilometriche e tecniche di *clustering* (che consentono di raggruppare i dati presi in analisi in classi omogenee in base alle similarità che i dati stessi presentano) con lo scopo di individuare le caratteristiche stilistico-linguistiche rilevanti per misurare e quantificare il grado di falsificazione e la distanza stilistica del falso comunicato 7 rispetto ai comunicati autentici.

Nel sottoparagrafo 4.1 si forniranno prove quantitative della falsificazione di FBR7 tramite l'utilizzo di singole caratteristiche stilometriche, mentre nel sottoparagrafo 4.2 si analizzeranno le co-occorrenze di parole per misurare la distanza tra il comunicato falso e i comunicati autentici.

#### 4.1 Analisi dei Cluster e caratteristiche stilometriche basate su singole parole

Questo tipo di analisi richiede innanzitutto una pre-elaborazione dei dati. Si è quindi proceduto alla trasformazione di tutti i caratteri in minuscoli, in modo da normalizzare i testi. Successivamente sono state calcolate le frequenze relative alle cento parole singole più comuni. Nessuna parola è stata rimossa in base ad una soglia di frequenza di occorrenza. In seguito, è stata effettuata un'analisi dei cluster utilizzando come caratteristiche testuali dei comunicati le frequenze delle cento parole singole più presenti estratte e come metrica il *Burrow's Delta* – Delta di Burrows – (Burrows, 2002) o *Classic Delta* in *Stylo*. Nello specifico, il Delta di Burrows misura quanto un testo di incerta attribuzione o anonimo, in alcuni casi, e un insieme di testi con autori certi e conosciuti divergono dalla media delle caratteristiche calcolata di questi stessi testi messi insieme e presi in esame. Inoltre, il metodo Delta assegna un peso uguale ad ogni caratteristica stilistico-linguistica che misura, evitando in questo modo che le parole più comuni inficino il risultato. L'analisi dei cluster così condotta è stata poi rappresentata tramite un dendrogramma dei raggruppamenti o cluster. Nella Figura 1 si mostra il risultato di questa prima analisi condotta sui testi.

Figura 1 - *Analisi dei Cluster sulle frequenze delle 100 parole più comuni*



Come riportato in Figura 1, il comunicato FBR7 sembra essere molto distante rispetto alla totalità dei comunicati autentici. Infatti, il dendrogramma mostra molto chiaramente la distanza stilistico-linguistica che esiste nei testi componenti il corpus. Oltre alla contrapposizione tra comunicati autentici e comunicato falso, in Figura 1 si può notare anche la formazione di due gruppi omogenei all'interno dei comunicati autentici: un primo che raggruppa i testi dei primi cinque comunicati, un secondo che avvicina i comunicati dal BR6 al BR9. Si descriverà con maggiori dettagli questo aspetto nel paragrafo 5.

L'ispezione manuale delle caratteristiche stilistiche e lessicali utilizzate durante questo tipo di analisi stilometrica, volta a misurare la distanza testuale tra i comunicati, ci mostra un insieme di caratteristiche rilevanti nel determinare il grado di falsificazione del comunicato FBR7. Nell'ispezione, oltre alla lunghezza minore in termini di parole del comunicato in esame rispetto ai comunicati autentici, già rilevata in precedenza, tra le caratteristiche stilistico-linguistiche che ci permettono di determinare il grado di falsificazione c'è la minore frequenza della congiunzione "e" rispetto ai comunicati autentici. Come anche rilevato in (Marchetti, 2017: 58), nei comunicati autentici sembra, infatti, prevalere una preferenza per la paratassi o coordinazione rispetto al comunicato falso. Come è evidente in Tabella 2, i valori relativi alla congiunzione *e* per il comunicato falso (segnalati in grassetto) sono distanti dal gruppo dei comunicati autentici. Nello specifico, in Tabella 2 si riportano i valori riguardanti sia la frequenza, calcolata come la "quota" appartenente ad ogni autore o comunicato per una data caratteristica per il numero totale di parole, sia quelli relativi allo *z-score* che rappresenta e descrive quanto sia distante una data caratteristica dalla norma dell'uso in un dato corpus in riferimento ad ogni autore o comunicato.

Tabella 2 - *Preferenza di coordinazione nei comunicati autentici*

	<i>br_1</i>	<i>br_2</i>	<i>br_3</i>	<i>br_4</i>	<i>br_5</i>	<i>br_6</i>	<i>br_7</i>	<i>br_8</i>	<i>br_9</i>	<i>fbr_7</i>
e	3.51	2.75	5.37	4.72	4.44	4.51	3.34	4.95	4.94	<b>2.01</b>
e (z-sc)	-0.49	-1.18	1.20	0.60	0.35	0.41	-0.65	0.80	0.81	<b>-1.86</b>

In entrambi i casi, i valori di frequenza e *z-score* sono inferiori e si allontanano dalla norma d'uso dell'intero corpus dei comunicati BR. Tra i comunicati autentici, sembra mostrare una preferenza per la coordinazione il BR3 che supera di poco le 1.000 parole in lunghezza testuale. Il BR4, il comunicato più esteso testualmente, ha un valore leggermente minore rispetto ad altri due comunicati BR8 e BR9, il primo con una lunghezza testuale inferiore sia al BR9 che al BR4, il secondo con una lunghezza testuale superiore alle 1.000 parole.

Un secondo insieme di caratteristiche stilometriche utilizzate durante l'analisi fa riferimento ai connettivi di causa ed effetto come "quindi", all'avverbio di comparazione "come" e alla congiunzione avversativa "ma".

In Tabella 3 si riportano i valori di frequenza in riferimento a queste tre caratteristiche.

Tabella 3 - *Uso di connettivi, avverbi e congiunzioni*

	<i>br_1</i>	<i>br_2</i>	<i>br_3</i>	<i>br_4</i>	<i>br_5</i>	<i>br_6</i>	<i>br_7</i>	<i>br_8</i>	<i>br_9</i>	<i>fbr_7</i>
quindi	0.13	0	0.09	0.22	0	0.21	0	0.21	0.19	<b>0</b>
come	0.14	0.74	0.48	0.29	0.14	0.31	0.96	0.32	0.20	<b>0</b>
ma	0.27	0.22	0.58	0.36	0.59	0.21	0.52	0.33	0.38	<b>0</b>

Anche in Tabella 3 sono segnalati in grassetto i valori del comunicato falso. Come è evidente dai risultati, il FBR7 non mostra nessun utilizzo del connettivo “quindi”, dell’avverbio di comparazione “come” e della congiunzione avversativa “ma”. Ciò ancora a dimostrazione della lontananza stilistico-linguistica rispetto ai comunicati autentici. Infatti, in questi ultimi sembra esserci una maggiore attenzione e raffinatezza nella strutturazione della materia linguistica e delle argomentazioni sviluppate attraverso l’utilizzo di strumenti stilistici come connettivi, avverbi e congiunzioni. Come anche discusso da Marchetti (2007: 59), l’organizzazione dei periodi nei comunicati BR sembra riflettere una padronanza della lingua scritta e uno stile analitico nel discutere ed organizzare il programma ideologico brigatista.

Inoltre, nel comunicato FBR7 non vi è alcun utilizzo di parole proprie del vocabolario militante del programma brigatista come “movimento”, “partito”, “controrivoluzione”, “proletariato”, “governo”, “potere”, “classe”, “comunista”, “multinazionali”.

#### 4.2 Analisi dei Cluster e caratteristiche stilometriche basate su co-occorrenze di parole

Di seguito, sarà descritta una seconda analisi stilometrica effettuata sui testi del corpus preso in esame con lo scopo di rilevare ulteriori caratteristiche in grado di quantificare la falsificazione. In questo secondo caso, saranno estratte e misurate le coppie di parole più frequenti nei comunicati componenti il corpus. Nello specifico, sono state calcolate le co-occorrenze delle parole più comuni per misurare la distanza stilistica tra i comunicati. Oltre ad una differenza stilistica tra i comunicati, in questa analisi, data la finestra testuale più ampia, sono state anche rilevate le differenze lessicali e di contenuto tra i comunicati.

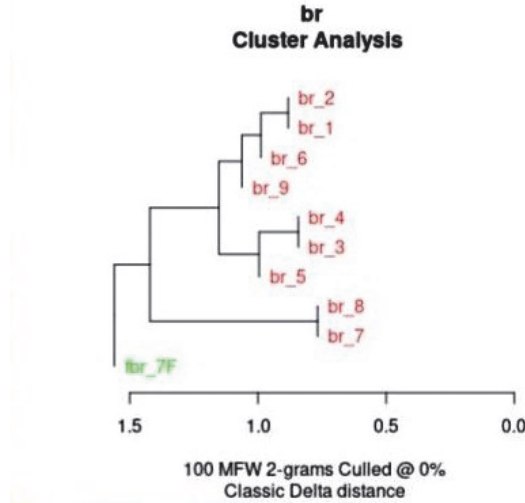
A tal fine, sono state calcolate ed utilizzate le prime cento coppie di parole (2-grams) più frequenti all’interno del corpus e nessun *culling* è stato applicato all’estrazione delle caratteristiche testuali. Anche in questo esperimento, è stata effettuata un’analisi dei cluster utilizzando come metrica il Delta di Burrows (Burrows, 2002).

In Figura 2 si riporta il dendrogramma generato da questa seconda analisi. Come si evince dalla Figura 2, anche nella misurazione della distanza testuale utilizzando coppie di parole, il comunicato FBR7 risulta discostarsi dal gruppo di comunicati autentici. Inoltre, da un punto di vista linguistico-stilistico, nella formazione dei gruppi omogenei testuali, i comunicati autentici risultano aggregati secondo una diversa organizzazione della materia linguistica rispetto all’organizzazione mostrata in Figura 1. Infatti, i risultati di questa seconda analisi non suggeriscono la stessa contrapposizione stilistica tra i primi cinque comunicati e gli ultimi quattro, come invece evidenzia l’analisi sulle parole singole (cfr. Figura 1), bensì si dispongono in



funzione di una diversa organizzazione dei temi trattati in ciascun comunicato. Ciò è rilevabile dall'ispezione manuale delle caratteristiche testuali utilizzate nell'analisi.

Figura 2 - *Analisi dei Cluster sulle frequenze delle 100 coppie di parole più frequenti*



In Tabella 4 si riportano le prime cinque coppie di parole estratte e utilizzate durante questa seconda analisi. I valori relativi al comunicato falso sono segnalati in grassetto.

Tabella 4 - *Co-occorrenze rilevanti nel determinare la falsificazione*

	<i>br_1</i>	<i>br_2</i>	<i>br_3</i>	<i>br_4</i>	<i>br_5</i>	<i>br_6</i>	<i>br_7</i>	<i>br_8</i>	<i>br_9</i>	<i>fbr_7</i>
stato imperialista	0.13	0.07	0.39	0.51	0.14	0.21	0.17	0.22	0.28	<b>0</b>
controrivoluzione imperialista	0.13	0.14	0.19	0.14	0.14	0.10	0	0	0.28	<b>0</b>
comunista combattente	0.13	0.14	0.19	0.22	0.29	0.10	0.08	0.11	0.19	<b>0</b>
borghesia imperialista	0.27	0.14	0.09	0	0	0.21	0	0.22	0.09	<b>0</b>
(stato) imperialista (delle) multinazionali	0.27	0.07	0.39	0.36	0.29	0.31	0	0	0.28	<b>0</b>
il movimento (rivoluzionario)	0.40	0.14	0.19	0.29	0.14	0.21	0.17	0.22	0.28	<b>0</b>
<b>la dc</b>	0.27	0.07	0	0.07	0	0.63	0.96	0.99	0.19	<b>0</b>
<b>della dc</b>	0.13	0.14	0	0.07	0.14	0.21	0.08	0.33	0.57	<b>1.35</b>
Aldo Moro	0.54	0.52	0.29	0.07	0.14	0.63	0.70	0.66	0.57	<b>2.02</b>

Nell'analisi delle co-occorrenze di parole si possono notare gli intenti programmatici e il linguaggio formulaico, come “stato imperialista”, ideologicamente motivato dei comunicati autentici. Proprio per la presenza e l'utilizzo di formule ed espres-

sioni ideologiche, come “movimento rivoluzionario” o “stato imperialista”, il comunicato FBR7 viene a trovarsi in una posizione distante rispetto ai comunicati più programmatici. Infatti, in questo falso non compare nessun riferimento al lessico militante diffuso nella totalità dei comunicati autentici. Come anche il riferimento, oggetto o soggetto, all'*establishment* rappresentato da “la DC”, ha una frequenza nulla rispetto ai comunicati autentici. Sebbene in questi “la DC” venga maggiormente rappresentata linguisticamente nel gruppo di comunicati che va dal BR6 al BR8 rispetto a quanto sia rappresentato nei primi comunicati, nel comunicato FBR7 il riferimento alla Democrazia Cristiana è presente sotto forma di complemento (“della DC”), con una frequenza nettamente superiore rispetto alla media attestata nei comunicati autentici.

Nel comunicato FBR7, pur non essendoci un uso del lessico militante votato alla propaganda rivoluzionaria, l'intenzione di diffusione di false notizie intorno al sequestrato come tematica principale di tutto il comunicato può essere letto nell'uso di “Aldo Moro”, maggiore rispetto alla media dei comunicati autentici.

In maggior misura, in questa seconda analisi, proprio l'assenza di un vocabolario marcato dal punto di vista ideologico e la conseguente interruzione di una narrazione militante nel testo del comunicato FBR7, allontana quest'ultimo dai comunicati apparsi in precedenza.

### 5. *Differenze lessicali e tematiche nei comunicati*

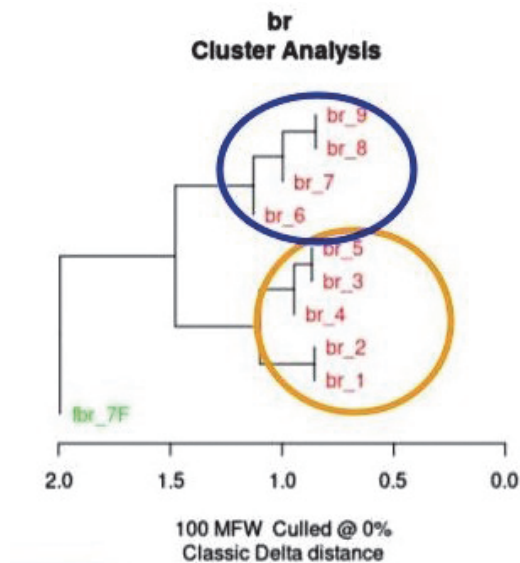
Nella Figura 1 è stato illustrato il dendrogramma generato dall'analisi delle caratteristiche stilometriche in riferimento alle cento parole più frequenti utilizzate nel corpus. Come anticipato nella sezione 4.1, nella suddetta figura si può notare la formazione di due gruppi omogenei e ben definiti di comunicati, accomunati da alcune caratteristiche lessicali ricorrenti nei due rispettivi gruppi. Ciò fornirà un'ulteriore misura della falsificazione e della lontananza ideologica del comunicato falso rispetto alle intenzioni programmatiche mostrate nei primi sei comunicati autentici (Re, 2013).

In questo paragrafo, si descriverà un metodo per misurare le differenze tra testi o gruppi di testi e, in generale, per misurarne la progressione temporale in termini lessicali. Per condurre ciò, utilizzando le caratteristiche lessicali estratte in precedenza in riferimento alle cento parole più frequenti, si è utilizzato un metodo basato sulla funzione *Craig's Zeta* (Burrows, 2007; Craig, Kinney, 2009).

Questa funzione non si basa sulla frequenza delle parole ma sul loro uso da parte di un autore in diversi testi. A tal fine, i testi vengono divisi in set di uguale dimensioni e grazie a questa funzione si confronta quante sezioni di ogni set contiene ciascuna parola. Il metodo combina la percentuale delle sezioni di un set di testi in cui compare una determinata parola con la percentuale delle sezioni di un set di testi in cui tale parola è assente in un'unica misura distintiva per ciascuna parola. In questo modo, si producono due liste di parole, la prima delle parole usate e preferite nel primo gruppo di testi e la seconda lista con le parole assenti nel primo gruppo di testi e presenti nel secondo gruppo.

Come mostrato in Figura 3, si prenderanno in esame i due gruppi segnalati e le rispettive differenze con il comunicato FBR7. I due gruppi segnalati sono disposti in ordine cronologico secondo la progressione temporale di apparizione dei comunicati. Infatti, il primo gruppo fa riferimento ai primi cinque comunicati, mentre il secondo gruppo si riferisce ai comunicati BR6 e successivi. Applicando la funzione Zeta di Craig, si rilevano le aree tematiche racchiuse intorno al lessico utilizzato nei diversi periodi di apparizione dei comunicati distinguendo così, dal lessico presente nei due gruppi, la progressione temporale e linguistica delle intenzioni e delle svolte programmatiche nelle BR.

Figura 3 - *Gruppi omogenei di comunicati*



Il risultato di questo metodo segnala due aree tematiche nel gruppo dei primi cinque comunicati e tre aree nel secondo gruppo. Il lessico utilizzato nella stesura dei primi cinque comunicati è nettamente più votato alla diffusione e alla chiamata verso una lotta e una rivoluzione contro l'establishment. Infatti, le parole maggiormente presenti nei primi cinque comunicati ruotano intorno all'area tematica coerente con azioni e concetti come: "Armata", "rivoluzionaria", "controrivoluzione", "proletariato", "iniziativa", "militari", "ristrutturazione", "popolo", "guerra", "terrore", da un lato; mentre la seconda area tematica si forma intorno all'oggetto contro il quale detta guerra rivoluzionaria deve essere svolta: "regime", "classe", "potere", "paese", "strutture", "multinazionali", "imperialista", "democristiani", "stampa".

Al contrario, nel secondo gruppo di comunicati che va dal BR6 al BR9, l'attenzione e le intenzioni delle BR si spostano sulla lotta alla contrattazione e all'accusa verso la classe politica del tempo come sicari e "complici" nella "condanna" di Aldo Moro. In questo gruppo di comunicati, dall'analisi vengono segnalate tre aree tematiche. Una prima sulle complicità nel sequestro e nella futura condanna: "complici";

“governo”, “morte”, “DC”, “condannato”, “Moro”, “sentenza”; una seconda sulle possibilità di accordo: “rilascio”, “libertà”, “risposta”, “Andreotti”; e una terza intorno alla liquidazione del sequestrato come problema non risolvibile: “scaricare”, “partito”, “prigioniero”, “problema”, “genocidio”, “condanna”.

## 6. Conclusioni

In questo articolo sono state presentate le analisi condotte con metodi di stilometria computazionale in grado di misurare la falsificazione a posteriori di un documento apparso durante il caso Moro. Durante l’analisi dei cluster, sono state presentate due applicazioni di metodi stilometrici utilizzate per quantificare le caratteristiche linguistico-stilistiche nei comunicati autentici e in quello falso, calcolando una finestra testuale differente con la metrica Delta di Burrows, al fine di estrarre e computare la distanza stilistica tra i comunicati e, di conseguenza, in grado di misurarne la falsificazione nel comunicato FBR7.

Inoltre, tramite l’applicazione di un’ulteriore misura stilometrica, nota come Craig’s Zeta, si è illustrato un metodo per comparare l’uso lessicale tra testi o gruppi di testi, in questo caso tra i comunicati, fornendo risultati che dimostrano le differenze lessicali tra testi e in grado di monitorare la progressione temporale stilistica e contenutistica in testi.

Il risultato di questi esperimenti condotti con tecniche stilometriche ha mostrato la capacità di estrarre tali evidenze testuali e, in generale, di analisi di dati testuali per l’identificazione e riconoscimento di caratteristiche linguistico-stilistiche in casi di falsificazione testuale.

## Ringraziamenti

Questa ricerca è stata svolta nell’ambito di un progetto di dottorato innovativo a caratterizzazione industriale finanziato dal fondo POR Campania FSE 2014/2020. Gli autori hanno contribuito alla stesura delle sezioni come segue: Raffaele Manna è l’autore dei paragrafi 3, 4 e 5. Antonio Pascucci è l’autore del paragrafo 1, mentre i paragrafi 2 e 6 sono in comune tra Raffaele Manna e Antonio Pascucci. Si ringrazia la Professoressa Johanna Monti per aver guidato la fase di ricerca e la stesura di questo articolo.

## Riferimenti bibliografici

- BURROWS, J.F. (2002). ‘Delta’: a measure of stylistic difference and a guide to likely authorship. In *Literary and linguistic computing*, 17(3), 267-287.
- BURROWS, J.F. (2007). All the way through: testing for authorship in different frequency strata. In *Literary and Linguistic Computing*, 22(1), 27-48.

- COULTHARD, M. (2004). Author identification, idiolect, and linguistic uniqueness. In *Applied linguistics*, 25(4), 431-447.
- COULTHARD, M. & JOHNSON, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. Londra-New York: Routledge.
- CRAIG, H. & KINNEY, A.F. (Eds.). (2009). *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge University Press.
- DAELEMANS, W. (2013). Explanation in computational stylometry. In *International Conference on Intelligent Text Processing and Computational Linguistics*. Berlin/Heidelberg: Springer, 451-462.
- EDER, M. (2011). Style-markers in authorship attribution: a cross-language study of the authorial fingerprint. In *Studies in Polish Linguistics*, 6(1), 99-114.
- EDER, M., RYBICKI, J. & KESTEMONT, M. (2016). Stylometry with R: a package for computational text analysis. In *The R Journal*, 8(1), 107-121.
- GIBBONS, J. & TURELL, M.T. (Eds.). (2008). *Dimensions of Forensic Linguistics*. Amsterdam Filadelfia: John Benjamins.
- GRANT, T. & BAKER, K. (2001). Identifying reliable, valid markers of authorship: a response to Chaski. In *Forensic Linguistics*, 8, 66-79.
- HEISER, M.S. (2007). The Majestic Documents: A Forensic Linguistic Report. <https://drmsb.com/michaelsheiser/MJ%20Test%20article.pdf/> Accessed 07.01.2022
- KOCHER, M. & SAVOY, J. (2017). Distance measures in author profiling. In *Information processing & management*, 53(5), 1103-1119.
- KOPPEL, M., SCHLER, J., & ARGAMON, S. (2009). Computational methods in authorship attribution. In *Journal of the American Society for information Science and Technology*, 60(1), 9-26.
- MARCHETTI E. (2017). "Boia imperialisti, spie di regime e corrotti buffoni". La lingua dei comunicati delle brigate rosse durante il Sequestro Moro. In *L'analisi linguistica e letteraria*, 2, 51-70.
- MEDICI, M. (1990). Nel ventre del mostro. Le caratteristiche stilistiche e le fonti di ispirazione dei volantini prodotti dal terrorismo italiano. In *Italiano e oltre*, 1(5), 17-30.
- SCIASCIA, L. (1983). *L'affaire Moro. Con aggiunta la Relazione Parlamentare*. Palermo, Sellerio.
- PEARL, L. & STEYVERS, M. (2012). Detecting authorship deception: a supervised machine learning approach using author writeprints. In *Literary and linguistic computing*, 27(2), 183-196.
- RE, M. (2013). Il linguaggio dei 55 giorni che cambiarono l'Italia. Analisi dei documenti scritti dalle Brigate Rosse durante il sequestro di Aldo Moro. In *Revista de la Sociedad Española de Italianistas*, 9, 251-76.
- ROMITO, L. (2013). *Manuale di Linguistica Forense*. Roma: Bulzoni.
- ROCHA, A., SCHEIRER, W.J., FORSTALL, C.W., CAVALCANTE, T., THEOPHILO, A., SHEN, B., CARVALHO, A.R.B., & STAMATOS, E. (2016). Authorship attribution for social media forensics. In *IEEE transactions on information forensics and security*, 12(1), 5-33.
- WEREN, E.R., KAUER, A.U., MIZUSAKI, L., MOREIRA, V.P., DE OLIVEIRA, J.P.M., & WIVES, L.K. (2014). Examining multiple features for author profiling. In *Journal of information and data management*, 5(3), 266-266.