

Nuove Ricerche Umanistiche



NUOVE INCHIESTE SULL'EPISTOLA A CANGRANDE

Atti della giornata di studi
Pisa 18 dicembre 2018

a cura di Alberto Casadei
con la collaborazione di Elisa Orsi e Marco Signori

PISA
UNIVERSITY
PRESS

Nuove inchieste sull'epistola a Cangrande : atti della giornata di studi, Pisa 18 dicembre 2018 / a cura di Alberto Casadei ; con la collaborazione di Elisa Orsi e Marco Signori. - Pisa : Pisa university press, 2020. - (ILLA-Nuove ricerche umanistiche ; 2)

856.1 (WD)

Casadei, Alberto <1963- > II. Orsi, Elisa III. Signori, Marco 1. Alighieri, Dante - Epistola a Cangrande - Convegni

CIP a cura del Sistema bibliotecario dell'Università di Pisa

Collana ILLA - Nuove Ricerche Umanistiche

Responsabile: Rolando Ferri

Direzione: Alberto Casadei, Marina Foschi, Mauro Tulli

Comitato Scientifico: Albert R. Ascoli (Univ. Berkeley, Ca.), Simone Beta (Univ. Siena), Pietro U. Dini (Univ. Pisa), Francesca Fedi (Univ. Pisa), Maria Letizia Gualandi (Univ. Pisa), Juliane House (Univ. Amburgo), Mario Labate (Univ. Firenze), Irmgard Männlein-Robert (Univ. Tübingen), Guido Mazzoni (Univ. Siena), Paolo Pontari (Univ. Pisa), Biancamaria Rizzardi (Univ. Pisa), Emanuele Zinato (Univ. Padova)



Opera sottoposta a
peer review secondo
il protocollo UPI

In copertina: Miniatura a c.2r del ms.200 Inferno di Dante Alighieri conservato presso la civica Biblioteca Guarnieriana.

© Copyright 2020 by Pisa University Press srl

Società con socio unico Università di Pisa

Capitale Sociale € 20.000,00 i.v. - Partita IVA 02047370503

Sede legale: Lungarno Pacinotti 43/44 - 56126 Pisa

Tel. + 39 050 2212056 - Fax + 39 050 2212945

press@unipi.it

www.pisauniversitypress.it

ISBN 978-88-3339-333-9

layout grafico: 360grafica.it

impaginazione: 360grafica.it

L'Editore resta a disposizione degli aventi diritto con i quali non è stato possibile comunicare, per le eventuali omissioni o richieste di soggetti o enti che possano vantare dimostrati diritti sulle immagini riprodotte. Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume/fascicolo di periodico dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633. Le riproduzioni effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi - Centro Licenze e Autorizzazione per le Riproduzioni Editoriali - Corso di Porta Romana, 108 - 20122 Milano - Tel. (+39) 02 89280804 - E-mail: info@clearedi.org - Sito web: www.clearedi.org

INDICE

PREMESSA <i>Alberto Casadei</i>	5
NOTA AI TESTI	7
RIFLESSIONI BIOGRAFICHE E DI METODO SU UN VECCHIO PROBLEMA: DANTE A VERONA <i>Giuseppe Indizio</i>	11
MINOSSE, L'USCITA DALL'EGITTO, I "SENSI" DEL TESTO E IL PERCORSO "COMICO" DELL'ALLEGORIA <i>Piermario Vescovo</i>	27
QUALCHE PRECISAZIONE SUL SENSO ALLEGORICO NELL' <i>EPISTOLA A CANGRANDE</i> (E IN ALTRI COMMENTATORI) <i>Marco Signori</i>	49
ANCORA SULL' <i>EPISTOLA A CANGRANDE</i> , GUIDO, LANA: IL <i>SUBIECTUM</i> DELLA <i>COMMEDIA</i> <i>Fabrizio Franceschini</i>	77
DANTE <i>LIBERRIMUS VATUM</i> : APPUNTI SU TRADIZIONE ORAZIANA E VARIETÀ STILISTICA TRA <i>EPISTOLA A CANGRANDE</i> E <i>PRIMA EGLOGA</i> <i>Eugenio Refini</i>	105
"CANTICAM... OFFERO" E ALTRI PROBLEMI ESEGETICI <i>Alberto Casadei</i>	129
L' <i>EPISTOLA A CANGRANDE</i> AL VAGLIO DELLA <i>COMPUTATIONAL AUTHORSHIP VERIFICATION</i> : RISULTATI PRELIMINARI (CON UNA POSTILLA SULLA COSIDDETTA «XIV <i>EPISTOLA</i> DI DANTE ALIGHIERI») <i>Silvia Corbara, Alejandro Moreo, Fabrizio Sebastiani, Mirko Tavoni</i>	153

APPENDICE	193
IL MOTO DELLE “ETTERNE ROTE”: LA COSMOLOGIA DANTESCA TRA LA <i>COMMEDIA</i> E L’ <i>EPISTOLA XIII</i> A CANGRANDE DELLA SCALA <i>Susanna Barsella</i>	195
INDICE DEI NOMI <i>Elisa Orsi, Marco Signori</i>	225
ELENCO DEGLI AUTORI	235

**L'EPISTOLA A CANGRANDE AL VAGLIO DELLA
COMPUTATIONAL AUTHORSHIP VERIFICATION:
RISULTATI PRELIMINARI
(CON UNA POSTILLA SULLA COSIDDETTA
«XIV EPISTOLA DI DANTE ALIGHIERI»)***

Silvia Corbara, Alejandro Moreo, Fabrizio Sebastiani, Mirko Tavoni

Abstract

Questo lavoro applica tecniche automatiche di “Authorship Verification” (AV) al problema di riconoscere se l’“Epistola a Cangrande” sia un’opera autentica di Dante Alighieri o sia invece opera di un falsario. L’algoritmo di AV che viene utilizzato usa tecniche di “machine learning”: esso “addestra” un sistema automatico (un “classificatore”) a rilevare se un certo testo latino è di Dante o meno, esponendolo a un corpus di testi latini di Dante e di testi latini di autori coevi a Dante. L’algoritmo basa le sue ipotesi sull’analisi di un insieme di caratteristiche stilometriche, cioè di tratti linguistici legati allo stile, le cui frequenze d’uso tendono a rappresentare la “firma” inconscia di un autore. L’analisi condotta in questo lavoro suggerisce che, delle due parti in cui l’Epistola è tradizionalmente suddivisa, nessuna è di Dante. Esperimenti in cui lo stesso sistema di AV è stato applicato a ciascun testo del corpus suggeriscono che esso ha un grado di accuratezza abbastanza elevato, dando così credibilità alla sua ipotesi sulla paternità dell’Epistola. Nell’ultima sezione di questo lavoro applichiamo il nostro classificatore a quella che è stata ipotizzata essere la “14a Epistola di Dante”; il sistema rigetta, con grande sicurezza, l’ipotesi che questa epistola possa essere di Dante.

In this work we apply techniques from computational Authorship Verification (AV) to the problem of detecting whether the “Epistle to Cangrande” is an authentic work by Dante Alighieri or is instead the work

* Nell’ambito di una ricerca interdisciplinare concepita e sviluppata in comune, i §§ 1-2 sono di Fabrizio Sebastiani (ISTI-CNR), il § 3 di Alejandro Moreo (ISTI-CNR), il § 4 di Silvia Corbara (Scuola Normale Superiore), il § 5 e la *Postilla* di Mirko Tavoni (Università di Pisa, Dip. di Filologia, Letteratura e Linguistica). Ringraziamo Gabriella Albanese e Paolo Pontari per la preziosa consulenza filologica nell’allestimento dei corpora di *training*.

of a forger. The AV algorithm we use is based on “machine learning”: the algorithm “trains” an automatic system (a “classifier”) to detect whether a certain Latin text is Dante’s or not Dante’s, by exposing it to a corpus of example Latin texts by Dante and example Latin texts by authors coeval to Dante. The detection is based on the analysis of a set of stylometric features, i.e., style-related linguistic traits whose usage frequencies tend to represent an author’s unconscious “signature”. The analysis carried out in this work suggests that, of the two parts into which the Epistle is traditionally subdivided, neither is Dante’s. Experiments in which we have applied our AV system to each text in the corpus suggest that the system has a fairly high degree of accuracy, thus lending credibility to its hypothesis about the authorship of the Epistle. In the last section of this paper we apply our system to what has been hypothesized to be “Dante’s 14th Epistle”; the system rejects, with very high confidence, the hypothesis that this epistle might be Dante’s.

1. Introduzione

Il presente lavoro discute alcuni risultati preliminari, e le scelte metodologiche che a essi hanno portato, riguardanti l’applicazione di tecniche di *computational authorship verification* (AV) al problema della discussa paternità dantesca dell’*Epistola XIII*. L’AV è una sottodisciplina della cosiddetta *computational authorship analysis*, una (piccola) branca dell’informatica che si occupa di formulare ipotesi riguardanti le caratteristiche o l’identità dell’autore di testi di paternità ignota o controversa. Per AV si intende, più specificamente, la procedura volta a determinare se un dato testo di paternità incerta sia o meno da attribuirsi a un dato autore; ovviamente, nel nostro caso l’autore “candidato” è Dante Alighieri e il testo è l’*Epistola XIII*. La *computational authorship analysis* è una disciplina con una storia ormai cinquantennale (si veda ad esempio il fondamentale studio di Mosteller e Wallace (1964)), ma che affonda le sue radici negli studi tardo-ottocenteschi (ovviamente non computazionali) di Mendenhall (1887), che per primo affrontò gli studi di autorialità mediante tecniche di stilometria quantitativa, secondo quello che, con Ginzburg (1989), potrebbe definirsi “paradigma evidenziale”.

Il presente lavoro affronta il problema della AV usando le metodologie dell’apprendimento automatico (*machine learning*), una importante branca della cosiddetta “intelligenza artificiale” che si occupa di sviluppare algoritmi che apprendono a svolgere determinati compiti “dall’esperienza”, e cioè dall’esame di casi in cui il compito sia stato eseguito correttamente; nella fattispecie di nostro interesse, l’algoritmo

apprende a distinguere testi danteschi da testi non danteschi dall'esame di testi la cui paternità dantesca o non dantesca sia acclarata.

Il presente articolo è strutturato come segue. Nel § 2 presentiamo una breve introduzione alla *computational authorship analysis*, di cui la *authorship verification* (AV) è una sottodisciplina, e ai suoi principi metodologici. Il § 3 descrive invece il nostro lavoro di applicazione delle metodologie AV al quesito se l'*Epistola XIII* sia o meno stata scritta da Dante. Di fondamentale importanza in questa sezione sono: (a) la costruzione del corpus di riferimento, suddiviso in opere di sicura paternità dantesca e opere di paternità sicuramente non dantesca, che fungerà da insieme di *training* (costituirà cioè l'insieme di esempi da cui l'algoritmo apprenderà a distinguere testi danteschi da testi non danteschi), e (b) la scelta dei tratti linguistici e stilistici (*features*) che l'algoritmo imparerà a correlare con la paternità dantesca o non dantesca di un testo. Il § 4 discute i risultati ottenuti. Tale sezione non si limita a riportare l'ipotesi formulata dal sistema riguardo all'*Epistola XIII*, ma illustra anche i risultati di una più vasta sperimentazione a cui abbiamo sottoposto il sistema software sviluppato, nella quale, per ciascun testo del corpus di riferimento e per ciascun autore presente in detto corpus, abbiamo nascosto al sistema la vera paternità del testo e chiesto di determinare se il testo sia stato scritto o meno dall'autore; la percentuale di determinazioni corrette è un'implicita indicazione di quanto la determinazione del sistema riguardo all'*Epistola XIII* sia credibile o meno. Il § 5 espone alcune osservazioni conclusive, discute i problemi lasciati aperti dal presente lavoro, e indica alcune linee secondo cui intendiamo svilupparlo nell'immediato futuro.

2. Cenni sulla (computational) authorship analysis

Per *computational authorship analysis* (AA) si intende un insieme di metodologie e tecniche per la progettazione e realizzazione di strumenti informatici atti a formulare ipotesi riguardanti le caratteristiche (*authorship profiling*) o l'identità (*authorship attribution*) dell'autore di un testo di paternità ignota o controversa. La AA si declina a sua volta in varie sotto-attività, quali ad esempio:

- *authorship verification* (AV): dato un candidato predefinito, determinare se è o meno l'autore del testo (Stamatatos, 2016);
- *closed-set authorship attribution*: dato un insieme predefinito di candidati che si assume contenere il vero autore del testo, identificare tale autore fra essi (Stamatatos 2009);
- *shared authorship*: dati due testi, determinare se sono o meno dello stesso autore (Koppel e Winter 2014);

- *native language identification*: dato un insieme predefinito di lingue (che può comprendere o meno la lingua in cui il testo è scritto), identificare quale di esse è la lingua materna dell'autore del testo (Tetreault *et al.* 2012).

L'AA ha svariate aree di applicazione, di cui sono esempi la *cybersecurity* (cioè la prevenzione di reati che potrebbero essere commessi con l'ausilio di mezzi digitali) e la *computational forensics* (cioè l'analisi di tracce digitali riguardanti reati già avvenuti). Entrambe queste aree di applicazione si rivolgono a testi contemporanei che generalmente non hanno valore culturale, quali messaggi minatori, lettere anonime, o comunicazioni epistolari fra sospetti. Diverso è il caso in cui i testi siano di valore letterario o storico, nel qual caso l'AA si configura come strumento di ausilio alla filologia; alcuni esempi su testi contemporanei si possono trovare nei lavori compiuti per "smascherare" Elena Ferrante (Cortelazzo 2015; Eder 2017; Tuzzi e Cortelazzo 2018; e cfr. Magli 2017); nel tentativo di identificare l'autore del 15° *Libro di Oz* (Binongo 2003); nella ricerca di testi giornalistici gramsciani originariamente pubblicati senza firma (Basile e Lana 2008; Degli Esposti 2015) e nell'analisi sull'autenticità del *Diario Postumo* di Montale (Italia e Canettieri 2013; Condello, Garulli e Tomasi 2016; e cfr. Magli 2017); mentre fra gli esempi riguardanti testi antichi si possono annoverare l'analisi sulla paternità del *Fiore* (Canettieri, 2016), quella sull'autenticità della lettera sui cristiani di Plinio il Giovane a Traiano (Tuccinardi 2017) e altri ancora (Kabala 2020; Kestemont *et al.* 2015).

Un'indagine mediante queste tecniche riguardante l'*Epistola XIII* ricade dunque sotto la fattispecie dell'*authorship verification*, in quanto il compito consiste nel formulare un'ipotesi se l'autore di detta epistola sia o meno Dante Alighieri¹. Lo scopo che ci poniamo in questo lavoro è però non solo quello di sviluppare un algoritmo in grado di formulare una tale ipotesi, ma soprattutto quello di far sì che tale ipotesi sia suffragata da presupposti solidi, e sia quindi credibile. Il metodo con cui cercheremo di garantire credibilità all'ipotesi che l'algoritmo formulerà sarà quello di sperimentare l'algoritmo su un corpus di testi anch'essi in latino medievale. Tale sperimentazione consisterà nel chiedere all'algoritmo di verificare se, dati un testo qualsiasi e un autore qualsiasi del nostro corpus, tale testo sia stato scritto o meno da tale autore, ovviamente

¹ Più specificamente, potremmo definire questo un esempio di *adversarial AV*, in quanto il compito consiste nel formulare un'ipotesi se l'autore dell'*Epistola XIII* sia Dante Alighieri o *un altro autore che si finge Dante Alighieri* (essendo *adversarial* un aggettivo che, nell'informatica, qualifica un compito che deve essere svolto "in ambiente ostile").

dopo aver nascosto all'algorithmo il nome del vero autore dello scritto. Se il corpus consiste di n testi, e se nel corpus sono rappresentati m autori, potremo ripetere questo test per un totale di $n*m$ volte, e calcolare la percentuale dei casi in cui l'algorithmo restituisce una risposta esatta; più alta sarà questa percentuale, maggiore sarà la credibilità dell'algorithmo, e maggiore potrà dunque essere la nostra fiducia che la risposta dell'algorithmo riguardante la paternità dantesca o meno dell'*Epistola XIII* sia corretta. Naturalmente, tale sperimentazione sarà tanto più convincente quanto più gli autori rappresentati nel corpus saranno "vicini" a Dante Alighieri (in altre parole, quanto maggiore sarà la difficoltà nel riconoscere essi, e non Dante Alighieri, come gli autori delle rispettive opere), e quanto più sarà alto il numero n dei testi e il numero m degli autori.

2.1 Authorship verification e stilometria

La *authorship verification*, come d'altronde tutta la *computational authorship analysis*, si basa non su un'analisi degli eventi descritti o dei concetti espressi nel testo, e/o della loro veridicità o plausibilità, ma unicamente sullo studio stilistico quantitativo (*stilometria*) della lingua con cui tali eventi vengono descritti e tali concetti espressi. L'idea alla base della AA è che esistano dei tratti inconsci (dei "tic compositivi") nell'uso che un autore fa della lingua, e che questi tratti costituiscano di fatto un'involontaria firma dell'autore; tali tratti sono comunemente detti *writer invariants*. La AA si basa sul tentativo di catturare questi tratti inconsci, e per farlo si appoggia prevalentemente alla stilometria, cioè allo studio quantitativo delle frequenze con cui determinati tratti sono presenti nel testo.

La storia della AA può essere suddivisa in tre fasi. La prima fase, che si fa convenzionalmente terminare con la pubblicazione del fondamentale studio di Mosteller e Wallace (1964) sull'autorialità dei *Federalist Papers*, è caratterizzata dal tentativo di riconoscere l'autore del testo in base a un unico tipo di tratto linguistico; un esempio di tale classe di metodi è lo studio propugnato da Mendenhall (1887), che basava la sua analisi unicamente sulle frequenze delle parole di determinate lunghezze. La seconda fase, che possiamo identificare con il periodo 1964-1999, è caratterizzata dall'impiego in contemporanea di diversi tipi di tratti linguistici per ipotizzare l'autorialità dello scritto, utilizzando metodologie statistiche per combinare queste diverse fonti. La terza e ultima fase si fa convenzionalmente iniziare nel 2000 ed è caratterizzata dalla sostituzione delle tecniche statistiche adottate in precedenza con modelli matematici e tecniche provenienti dall'"apprendimento automatico" (*machine learning*), un'importante sottodisciplina dell'informatica che si occupa della realizzazione di algoritmi che imparano a portare a termine determinati compiti a partire da esempi in cui detti compiti sono stati eseguiti cor-

rettamente. In altre parole, le prime due fasi si distinguono fra loro per l'eterogeneità o meno dei tratti linguistici da analizzare, mentre la terza si distingue dalle prime due per il tipo di modelli matematici usati per analizzare i tratti prescelti. La soluzione che cercheremo di dare al problema dell'autorialità dell'*Epistola XIII* sarà del terzo tipo.

2.2 Authorship verification, classificazione di testi e apprendimento automatico

La “madre” di tutti i compiti che vengono affrontati mediante apprendimento automatico è la classificazione di dati. Per *classificazione* nell'informatica si intende il compito di assegnare un dato a una classe scelta fra un insieme di classi finito e predefinito, ove tale compito sia non deterministico (per esempio, assegnare un numero naturale a una delle due classi *NumeriPrimi* e *NumeriNonPrimi* non è classificazione, perché tale assegnazione può essere effettuata deterministicamente, cioè senza possibilità di errore) e basato su un'analisi del contenuto del dato stesso. Molti compiti di interesse pratico possono essere formulati come problemi di classificazione: ad esempio, l'insieme di un certo numero di dati climatologici riguardanti la giornata odierna può essere classificato in una delle due classi *DomaniPioverà* e *DomaniNonPioverà*, e una radiografia toracica può essere classificata in una delle due classi *PresenzaDiLocalizzazioniNeoplastiche* e *AssenzaDiLocalizzazioniNeoplastiche*. Anche l'AV può essere formulata in termini di classificazione: il problema che ci interessa è infatti il classificare l'*Epistola XIII* in una delle due classi *Dante* e *NonDante*, che indicano rispettivamente la paternità dantesca e non dantesca di un testo. Si noti che qualsiasi problema di classificazione, esattamente a causa della sua natura non deterministica, ammette un margine di errore.

L'AV è dunque un'istanza della *classificazione di testi*, una classe di modelli matematici e tecniche in cui l'apprendimento automatico incontra l'analisi automatica del testo (Sebastiani, 2002). La classificazione automatica di testi può riguardare diverse dimensioni del testo fra esse ortogonali: il caso più frequente è quello della classificazione per argomento, in cui le classi rappresentano i temi di interesse di cui il testo può trattare, mentre un altro caso di crescente importanza è quello della classificazione per opinione (*sentiment classification*), in cui le classi di interesse rappresentano il tipo di opinione (e.g., *OpinionePositiva*, *OpinioneNegativa*) che il testo esprime su un determinato oggetto del discorso (e.g., un candidato politico, un prodotto). La AV rappresenta dunque una delle molte dimensioni (quella dell'autorialità) in base a cui un testo può essere classificato.

Nella classificazione di testi, un algoritmo di apprendimento automatico (il *trainer*) “allena” un sistema automatico (il *classificatore*) a classificare correttamente testi all'interno delle categorie di interesse (*Dante* e

NonDante, nel nostro caso), esponendolo a un insieme di testi (l'insieme di *training*) di cui sia nota l'appartenenza alle classi di cui sopra. Tali tecniche sono anche dette di apprendimento "supervisionato", in quanto il *trainer* gioca il ruolo del supervisore della fase di apprendimento.

In altre parole, esaminando gli esempi di *training* il classificatore impara quali sono i tratti linguistici che caratterizzano i testi di ciascuna classe di interesse, e sarà quindi in grado di applicare quanto appreso nel momento in cui gli venga richiesto di classificare nuovi testi, la cui appartenenza alle classi di interesse sia sconosciuta. Di fatto, quello che il classificatore impara è la correlazione statistica fra tratti linguistici e classi; in particolare, esso impara quali tratti sono fortemente correlati con una delle classi in gioco (e sono quindi utili al processo di classificazione) e quali invece non esibiscono correlazioni significative con alcuna delle classi (e sono quindi sostanzialmente inutili).

Due sono le scelte fondamentali che chi vuol costruire un sistema di AV deve compiere: la prima è la scelta dell'algoritmo di apprendimento automatico da utilizzare, la seconda è la scelta dei tratti linguistici che l'algoritmo deve esaminare. Dall'oculatezza di queste due scelte dipende ovviamente l'accuratezza che il classificatore dimostrerà una volta messo all'opera.

La prima di queste due scelte è importante, ma probabilmente meno della seconda: infatti, mentre è vero che esistono svariati algoritmi di apprendimento automatico, e che l'accuratezza di classificazione può variare sensibilmente da algoritmo ad algoritmo, è anche vero che ne esistono alcuni che hanno dimostrato un'ottima performance in quasi tutti i contesti a cui sono stati applicati. La nostra scelta, che discuteremo nelle prossime sezioni, è caduta su uno di questi.

La seconda di queste scelte, cioè i tratti linguistici che l'algoritmo deve esaminare, è invece di fondamentale importanza, e deve essere ben ponderata. Ad esempio, è evidente che scegliere come tratto linguistico la punteggiatura non sarebbe di grande aiuto se dovessimo eseguire una classificazione per argomento, dato che la frequenza con cui i segni di interpunzione vengono usati è sostanzialmente irrelata all'argomento che un testo tratta. È invece altrettanto evidente come il medesimo tratto linguistico potrebbe rivelarsi utile nella AV (ovviamente per quelle lingue che dei segni di interpunzione fanno uso, e per quelle situazioni storiche nelle quali sono gli autori a scegliere i segni d'interpunzione), perché diversi autori hanno diverse modalità di uso di tali segni, e tali modalità tendono a essere invarianti attraverso la produzione linguistica di un autore, quindi fanno parte della cifra stilistica dell'autore stesso. Risulta quindi evidente come chi costruisce un sistema di AV debba compiere un'oculata scelta dei tratti linguistici che chiederà all'algoritmo di analizzare, e che questa scelta debba cadere su

quegli elementi che, oltre a essere facilmente identificabili da un algoritmo, si ritengano fortemente correlati con il fenomeno dell'autorialità. La scelta (importante) dell'algoritmo di apprendimento automatico da utilizzare richiede quindi competenze informatiche, mentre la scelta (ancora più importante) dei tratti linguistici che l'algoritmo deve analizzare richiede competenze linguistiche e filologiche.

Del processo che la AV segue si può dare una interessante interpretazione geometrica (figura 1). Una volta che il testo sia stato convertito in un insieme di frequenze dei tratti linguistici prescelti (tale insieme è, nella matematica dell'apprendimento automatico, un "vettore"), per l'algoritmo tale testo è null'altro che un punto in un piano cartesiano, in cui a ciascun tratto corrisponde un asse del piano cartesiano² e alla frequenza di quel tratto nel testo corrisponde la coordinata che il punto ha per quell'asse. Nell'esempio in figura 1 i punti corrispondono a testi di *training* di quattro autori diversi, dove i simboli "#", "+", "*", "@" indicano i punti corrispondenti a testi di Autore1, Autore2, Autore3, e Autore4, rispettivamente. Per l'algoritmo di *machine learning*, allenare un classificatore per la AV significa, in prima approssimazione, individuare una linea che separi gli esempi di *training* dell'autore di interesse (che nell'esempio supponiamo essere Autore1) dagli esempi di *training* di altri autori (nell'esempio: Autore2, Autore3, Autore4): tale linea corrisponde al classificatore³. Nel momento in cui debba classificare un testo di paternità sconosciuta, l'algoritmo lo convertirà in un punto nel medesimo spazio cartesiano (in figura 1: il punto indicato da un piccolo cerchio); da quale parte della linea esso cadrà corrisponderà alla decisione del classificatore (in figura 1, la determinazione è che trattasi di testo di Autore1). La distanza di tale punto dalla linea può essere interpretata come il "grado di certezza" che il classificatore ha in questa determinazione, dove maggiore distanza corrisponde a maggiore certezza che il classificatore ha nella propria scelta.

² Per facilità di illustrazione faremo qui l'ipotesi che vengano considerati solo due tratti linguistici, il che dà origine al familiare piano cartesiano bidimensionale. Nella realtà, nella AV si considerano dalle centinaia alle centinaia di migliaia di tratti, il che dà origine a piani cartesiani fortemente multidimensionali, i quali (pur trattabili matematicamente al pari di quelli bidimensionali) non sono rappresentabili geometricamente nel modo che ci è consueto.

³ Un principio euristico (detto di *structural risk minimization*) adottato da alcuni algoritmi di apprendimento automatico è quello di selezionare, tra tutte le possibili linee che separano gli esempi di *training* dell'autore di interesse dagli esempi di *training* degli altri autori, quella che mette fra sé e gli esempi la maggior distanza. Un altro principio euristico è quello di scegliere, fra tutte le possibili linee che abbiano la caratteristica sopra descritta, quella che abbia la forma matematicamente più semplice (e.g., una linea retta); questo principio è una istanziazione pratica del "rasoio di Occam".

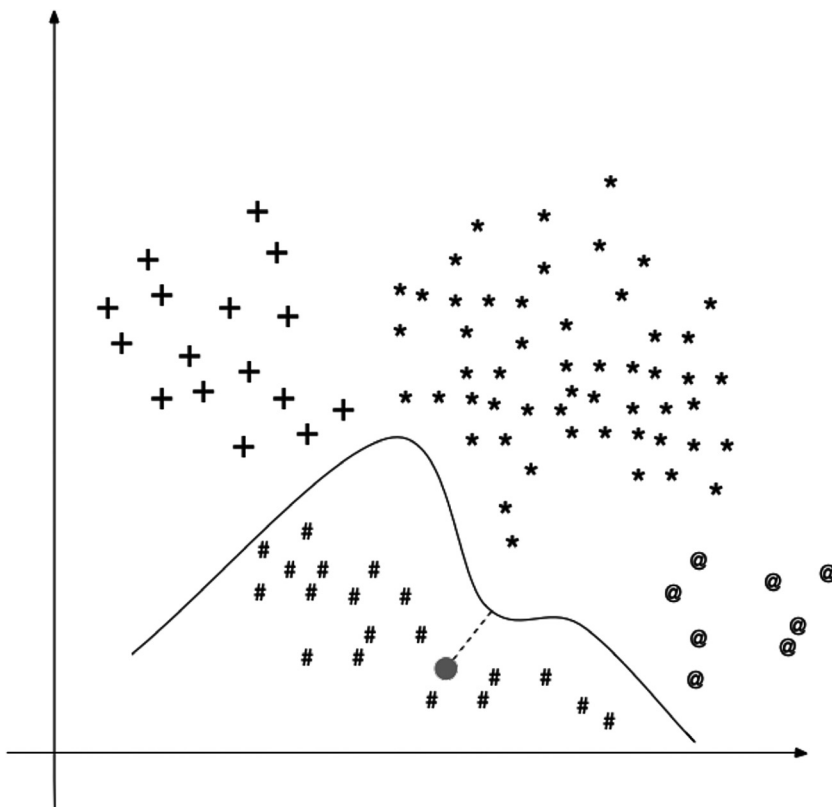


FIGURA 1. Interpretazione geometrica del processo di *authorship verification*.

3. Authorship verification per l'Epistola a Cangrande

In questa sezione descriviamo i passi che abbiamo seguito per costruire un sistema di AV orientato a formulare un'ipotesi di paternità, dantesca o meno, per l'*Epistola XIII*. Discuteremo la scelta dell'algoritmo di apprendimento automatico e quella dei tratti linguistici su cui si appunerà l'attenzione di tale algoritmo e, una volta allenato, quella del classificatore.

3.1 Scelta dell'algoritmo di apprendimento automatico

L'algoritmo di apprendimento supervisionato che abbiamo scelto è l'algoritmo di "regressione logistica" (*Logistic Regression* – LR). Si tratta di un algoritmo che genera un "classificatore lineare", e cioè un classificatore che, nella rappresentazione grafica della figura 1, corrisponde a una linea retta; un classificatore, quindi, estremamente semplice, in os-

sequio al principio euristico ispirato al “rasoio di Occam” e richiamato nella nota 3, che si è però spesso dimostrato molto accurato, quantomeno in applicazioni in cui gli oggetti della classificazione siano testuali.

Dato un testo di paternità sconosciuta, il classificatore restituisce un valore fra 0 e 1, che indica la probabilità che il classificatore soggettivamente attribuisce al fatto che il testo appartenga alla classe di interesse (nel nostro caso, alla classe *Dante*). Un valore vicino a 1 indica che il classificatore ritiene, con alto grado di certezza, che il testo appartenga alla classe *Dante*, mentre un valore vicino a 0 indica che il classificatore ritiene, con alto grado di certezza, che il testo appartenga alla classe *NonDante*; infine, un valore prossimo a 0.5 mostra che il classificatore è estremamente incerto sulla risposta da fornire al problema di autorialità.

La precisa formulazione matematica del problema di apprendimento e l’algoritmo per risolverlo esulano dallo scopo di questo articolo; rimandiamo il lettore interessato al manuale di Bishop (2006, pp. 205-206) per una trattazione più approfondita sul tema.

3.2 Scelta dei tratti linguistici da considerare

I tratti linguistici che l’algoritmo deve considerare (detti, nella letteratura dell’apprendimento automatico, *features*) devono riuscire, nel loro insieme, a catturare la dimensione di interesse (nel nostro caso: l’autorialità). In altre parole, se un fenomeno rilevante per la risoluzione del problema dell’autorialità non è rappresentato dai tratti prescelti, il classificatore non sarà in grado di discriminare correttamente tra istanze positive e istanze negative della classe di interesse (i.e., fra testi danteschi e testi non danteschi), indipendentemente dalla qualità dell’algoritmo di apprendimento automatico utilizzato. I tratti linguistici che abbiamo utilizzato sono largamente usati nella letteratura della AV e mirano a catturare i *writer invariants* dell’autore; in altre parole, l’ipotesi di base (e la speranza) è che questi tratti linguistici siano in grado di mettere a nudo *pattern* di utilizzo della lingua che siano ragionevolmente stabili all’interno della produzione di uno stesso autore, in quanto usati presumibilmente in maniera inconscia, e possano invece variare in maniera sensibile fra autori diversi.

Rappresentiamo ognuno di questi tratti rilevando la frequenza di occorrenza del tratto all’interno di ogni testo. Ad esempio, se il tratto linguistico è la “lunghezza delle parole” (le cosiddette *Mendenhall features*), avremo tanti tratti quante sono le diverse lunghezze delle parole presenti nel corpus, e per un dato testo il tratto “lunghezza n ” sarà valorizzato come il rapporto fra il numero di parole contenute nel testo che hanno lunghezza n e il numero totale delle parole presenti nel testo medesimo.

I tipi di tratti linguistici che abbiamo usato sono i seguenti:

- *Character n-grams*. Dato un testo, i suoi *character n-grams* sono definiti come tutte le sequenze di n caratteri consecutivi estraibili dal testo, eventualmente costituite da parti di parole consecutive. Ad esempio, dalla sequenza “Nomen omen” possiamo estrarre i seguenti *character 3-grams*:

Nom | ome | men | en_ | n_o | _om | ome | men

(dove “_” indica lo spazio). In questo lavoro abbiamo usato come *features* tutti i *character n-grams* di lunghezza 3, 4, e 5. L'ipotesi che sta alla base dell'uso dei *character n-grams* è che un autore possa avere una predisposizione a usare particolari suoni, che nel linguaggio scritto si materializzano in sequenze di lettere. Nel nostro caso, abbiamo ignorato la punteggiatura.

- *Word n-grams*. Dato un testo, i suoi *word n-grams* sono definiti come tutte le sequenze di n parole consecutive estraibili dal testo, eventualmente costituite da parti di frasi consecutive. Ad esempio, dalla sequenza “Per aspera ad astra” possiamo estrarre i seguenti *word 2-grams*:

Per aspera | aspera ad | ad astra

In questo lavoro abbiamo utilizzato *word n-grams* di lunghezza 1 e 2, evitando di spingerci a valori di n maggiori, in quanto in tal caso i *word n-gram* risultanti sarebbero un numero enorme, e quasi sempre scarsamente significativi.

- Parole funzione. Si tratta di parole grammaticali e utilizzate per la sola costruzione sintattica della frase, come preposizioni e congiunzioni. Abbiamo usato un insieme di 74 parole funzione della lingua latina, che sono ovviamente fra le parole più comuni nei testi da noi utilizzati.
- Desinenze verbali. Abbiamo considerato le terminazioni verbali regolari del latino (per un totale di 245 diverse terminazioni). L'inclusione di queste caratteristiche nasce con la finalità di catturare i tratti sintattici dovuti alle possibili preferenze dell'autore per certi modi e tempi verbali.
- Lunghezza delle parole. Tali *features* mirano a caratterizzare l'autore tramite la sua tendenza a utilizzare parole più o meno lunghe (Mendenhall, 1887). Nel nostro caso, la i -esima fra queste *features* rappresenta la frequenza relativa delle parole con lunghezza minore o uguale a i , con i compresa tra 1 e 23.
- Lunghezza delle frasi. Tali *features* mirano a caratterizzare l'autore tramite la sua tendenza a utilizzare frasi più o meno lunghe (Yule, 1938). Nel nostro caso, la i -esima fra queste *features* rappresenta la frequenza relativa delle frasi con lunghezza (determi-

nata dal numero di parole contenute nella frase stessa) minore o uguale a i , con i compresa tra 3 e 70.

La molteplicità delle *features* utilizzate rappresenta un punto di forza rispetto a sistemi di AV che utilizzano *features* di un solo tipo. Peraltro, non è detto che ciascuna di queste *features* sia significativa per il problema della AV, cioè sia correlata con la nostra classe di interesse, anzi è ipotizzabile che molte di esse non lo siano, o lo siano in minima parte. Ma un algoritmo di apprendimento automatico è in grado di considerarle una a una, di calcolare (a partire dall'insieme di *training*) per ciascuna di esse il grado di correlazione fra essa e la classe di interesse, e dunque di basare la propria determinazione solo su quelle *features* che esibiscono un grado di correlazione sufficientemente determinante. In altre parole, il classificatore combina svariate fonti di evidenza, delle quali sarebbe impossibile prevedere *a priori* quali si riveleranno più o meno significative, ma la cui maggiore, o minore, o nulla significatività emerge nel corso del procedimento.

3.3 La suddivisione dell'Epistola XIII e i corpora di training

Come è noto, l'*Epistola XIII* si compone di due parti ben distinte per scopo e, di conseguenza, stile. La prima (d'ora in poi indicata come Ep I, e consistente dei paragrafi 1-13) presenta tratti di tipo propriamente epistolare, mentre la seconda (Ep II, paragrafi 14-90) contiene un commento parziale alla *Commedia* e, in particolare, una *expositio* dei primi versi del *Paradiso*. Molti studiosi hanno dato giudizi di autorialità diversi sulle due sezioni; in particolare, convinti della natura spuria della seconda parte dell'*Epistola XIII*, ammettono però che la prima parte possa essere ascritta a Dante Alighieri.

Per questi motivi è sembrato imprescindibile nel nostro studio suddividere il nostro problema di classificazione in due sottoproblemi di AV distinti, uno per ciascuna delle due parti dell'*Epistola XIII*. Conseguenza di ciò è la necessità di creare un corpus di testi separato per ciascuna delle due parti, in modo da generare due classificatori distinti. In entrambi i casi, se per la classe *Dante* la soluzione è obbligata, in quanto i testi in latino attribuiti con certezza a Dante sono pochi e ben conosciuti, per la classe *NonDante* si pone un problema di scelta: essendo la classe di autori non danteschi potenzialmente infinita, è necessario individuarne un sottoinsieme sufficientemente rappresentativo. L'approccio che abbiamo scelto è quello di selezionare autori il più possibile "vicini" (in senso storico-linguistico) a Dante, e cioè autori la cui produzione sia caratterizzata da tratti linguistici simili a quelli di Dante, e reperirne adeguate edizioni, rendendole in formato digitale qualora questo non fosse

Tabella 1. Corpus per Ep I.	
Autore	Testo
Dante Alighieri	<i>Epistole</i> , ca. 1304-15, ed. Frugoni e Brugnoli 1996 ⁴ , Biblioteca Italiana (12)
Clara Assisiensis	<i>Epistola ad Ermentrudem</i> , ca. 1240-1253, ed. Menestò e Brufani 1995 (1)
	<i>Epistolae ad sanctam Agnetem de Praga I, II, III</i> , ca. 1234-1253, ed. Menestò e Brufani 1995 (3)
Guido Faba	<i>Epistole</i> , 1239-1241, ed. Gaudenzi 1971, ALIM (78)
Pietro della Vigna	<i>Epistolario di Pietro della Vigna</i> , ca. 1220-1249, ed. D'Angelo 2014 (146)
Autori vari ⁵	Dalla collezione di epistole di cancelleria di Petrus de Boate-riis, ca. 1250-1315, ed. Schneider 1926 ⁶ (30)
Giovanni Boccaccio	<i>Epistole e lettere</i> , ca. 1340-75, ed. Auzzas 1992, Biblioteca Italiana (24)

già disponibile. Il motivo di questa scelta, che può di per sé apparire ovvio, è che, se i testi non danteschi utilizzati per il *training* fossero linguisticamente molto diversi dai testi di *training* danteschi, un qualsiasi testo anche vagamente simile alla produzione dantesca verrebbe riconosciuto come inequivocabilmente dantesco, essendo il classificatore non allenato a operare distinzioni sottili. Ci si potrà invece aspettare risultati migliori se il classificatore verrà allenato a operare distinzioni sottili fra la produzione dantesca e quella non dantesca. Abbiamo quindi creato uno *screening* a larga scala di autori che abbiano scritto in latino in un periodo storico abbastanza ampio che include gli anni di Dante.

Per quanto riguarda Ep I, i testi che abbiamo inserito nel corpus di *training* sono elencati nella tabella 1. Come si vede, si tratta di testi di carattere epistolare (data la natura nuncupatoria di Ep I) databili fra il XIII e il XIV secolo, per un totale di 294 epistole (12 dantesche e 282 non dantesche). Nonostante si tratti di raccolte di epistole, ciascuna epistola è considerata come un testo di *training* singolo, data la natura

⁴ Solo l'Epistola XIII è tratta, anziché dall'ed. Frugoni e Brugnoli 1996, dall'ed. Azzetta 2016.

⁵ Federico II, Manfredi, Corrado d'Antiochia, Giordano d'Anglano, Urbano IV, Pietro Capocci, cardinale Simone, Carlo d'Angiò, Rolando di Brunforte, Michele VIII Paleologo, Carlo II di Napoli, Pietro III d'Aragona, Cangrande della Scala, Doffo da Panico, Enrico VII di Lussemburgo, Filippo II di Francia, vari anonimi.

⁶ Da questa raccolta abbiamo escluso l'Epistola V, di Cangrande ad Arrigo VII (pp. 265-266), per i motivi che saranno esposti nella *Postilla*.

Tabella 2. Corpus per Ep II.	
Autore	Testo
Dante Alighieri	<i>De Vulgari Eloquentia</i> , ca.1304-06, ed. Tavoni 2011, DanteSearch
	<i>Monarchia</i> , ca. 1313-1319, ed. Nardi 1979, Biblioteca Italiana
Ryccardus de Sancto Germano	<i>Chronicon</i> , ca. 1216-1243, ed. Garufi 1937, ALIM
Guido Faba	<i>Dictamina rhetorica</i> , ca. 1226-1228, ed. Gaudenzi 1971, ALIM
Bene Florentinus	<i>Candelabrum</i> , ca. 1238, ed. Alessio 1983, ALIM
Boncompagno da Signa	<i>Liber de obsidione Ancone</i> , ca. 1198-1200, ed. Garbini 1999, ALIM
	<i>Palma</i> , ca. 1198, ed. Sutter 1894, ALIM
	<i>Rota Veneris</i> , ante 1215, ed. Garbini 1996, ALIM
	<i>Ysagoge</i> , ca. 1204, ed. Clark 1997, ALIM
Iulianus de Spira	<i>Vita Sancti Francisci</i> , ca. 1232-1239, ed. Menestò e Brufani 1995, ALIM
Iohannes de Plano Carpini	<i>Historia Mongalorum</i> , ca. 1247-1252, ed. Daffinà et al. 1989, ALIM
Guido de Columnis	<i>Historia destructionis Troiae</i> , ca. 1272-1287, ed. Griffin 1936, ALIM
Iacobus de Varagine	<i>Chronica civitatis lanuensis</i> , ca. 1295-1298, ed. Monleone 1941, ALIM
Iohannes de Appia	<i>Constitutiones Romandiolae</i> , ca. 1283, ed. Colini Baldeschi 1925-1926, ALIM
Raimundus Lullus	<i>Ars amativa boni</i> , ca. 1290, ed. Romano 2004, ALIM
Nicola Trevet	<i>Expositio Herculis Furentis</i> , ca. 1315-1316, ed. Ussani 1959, Biblioteca Italiana
	<i>Expositio L. Annaei Senecae Agamemnonis</i> , ca. 1315-1316, ed. Meloni 1953, Biblioteca Italiana
Giovanni del Virgilio	<i>Allegorie super fabulas Ovidii Methamorphoseos</i> , ca. 1320, ed. Cotza 2013
	<i>Ars dictaminis</i> , ca. 1320, ed. Kristeller 1961, ALIM
Graziolo Bambaglioli	<i>Commento all'Inferno di Dante</i> , ca. 1324, ed. Rossi 1999, Biblioteca Italiana
Guido da Pisa	<i>Expositiones et glose. Declaratio super Comediam Dantis</i> , ca. 1327-1328, ed. Cioffari 1974, DDP
Zono de' Magnalis	<i>Vita di Virgilio</i> , ca. 1340, ed. Stok 1991
Giovanni Boccaccio	<i>De vita et moribus d. Francisci Petracchi</i> , ca. 1342, ed. Fabbri 1992, Biblioteca Italiana
	<i>De mulieribus claris</i> , ca. 1361-1362, ed. Zaccaria 1967, Biblioteca Italiana
	<i>Genealogie deorum gentilium libri</i> , ca. 1360-1375, ed. Romano 1951, Biblioteca Italiana

Pietro Alighieri	<i>Comentum super poema Comedie Dantis</i> , ca. 1340-1364, ed. Chiamenti 1999, Biblioteca Italiana
Benvenuto da Imola	<i>Comentum super Dantis Aldigherij Comoediam</i> , ca. 1375-1380, ed. Brownlee e Hollander 2018, DDP
	<i>Expositio super Valerio Maximo</i> , ca. 1380, ed. Rossi 2002
	<i>Glose Bucolicorum Virgilii</i> , ca. 1380, ed. Mancuso 2015
Filippo Villani	<i>Expositio seu comentum super Comedia Dantis Alleghe-rii</i> , ca. 1391-1405, ed. Ferrario 1999, DDP

intrinsecamente unitaria e specifica di ogni epistola. Ogni riga della tabella 1 presenta l'autore, il nome della raccolta o dell'epistola, la datazione approssimativa, l'edizione da cui i testi o il testo sono stati tratti, l'archivio testuale online (qualora esista) dal quale sono stati scaricati, e il numero di epistole che la raccolta contiene.

La tabella 2 contiene invece i testi che abbiamo incluso nel corpus di *training* per Ep II: come si può vedere si tratta di testi di varia natura, soprattutto commenti a opere letterarie e trattati (data l'analogia natura di Ep II) databili tra inizio XIII e fine XIV secolo, per un totale di 30 testi (2 danteschi e 28 non danteschi). Ogni riga della tabella 2 presenta l'autore, il titolo del testo, la sua datazione approssimativa, l'edizione dal quale è tratto e, qualora esista, l'archivio testuale online da cui il testo è stato scaricato.

Abbiamo sottoposto tutti i testi (comprese le due parti dell'*Epistola XIII*) ad alcuni interventi meccanici necessari per la corretta lettura automatica del testo: abbiamo sostituito sempre il carattere *v-V* col carattere *u-U* (per neutralizzare la differenza fra edizioni di testi che distinguono o non distinguono *u* e *v* secondo l'uso moderno), convertito l'intero testo in minuscolo, nonché ovviamente rimosso tutte le indicazioni editoriali o meta-testuali (come titoli correnti e numeri di pagina). Inoltre, abbiamo delimitato le citazioni con asterischi (*...*), se in latino, o parentesi graffe ({...}), se in altra lingua (di solito: volgare); all'algoritmo viene successivamente richiesto di ignorare dette citazioni, per evitare di inquinare la rappresentazione di un autore con brani testuali di altro autore.

Oltre a ciò, abbiamo segmentato ogni testo (con l'esclusione delle due parti dell'*Epistola XIII*) in testi di lunghezza minore, in modo da aumentare il numero complessivo dei testi di *training*, pur riducendo la loro dimensione media. Ciò si è reso necessario in quanto i processi di apprendimento automatico richiedono un numero significativo di esempi di *training*, indipendentemente dalla loro lunghezza.

In particolare, per ogni testo:

- abbiamo identificato le frasi che compongono il testo (tramite il software NLTK)⁷; qualora una frase sia più corta di 8 parole, abbiamo accorpato tale frase alla frase successiva (o alla precedente, nel caso sia l'ultima frase del testo);
- abbiamo creato gruppi di 3 frasi consecutive, e considerato ciascuno di questi gruppi come un testo di *training*.

Al termine di questo processo, il numero dei testi di *training* a disposizione è significativamente aumentato, come si può osservare nella tabella 3 riassuntiva.

Per il corpus per Ep I il numero di *features* risultanti è di 16.094, il che vuol dire che ogni “vettore” che rappresenta un testo di quel corpus ha lunghezza 16.094, e che il piano cartesiano in cui tali testi sono rappresentati come punti ha 16.094 assi. Per il corpus per Ep II il numero di *features* risultanti è invece di 84.843.

4. Risultati sperimentali

In qualsiasi applicazione di algoritmi di AV è molto importante avere una stima dell'accuratezza del classificatore sviluppato, in modo da poter valutare quanto la sua determinazione di autorialità di un testo di paternità sconosciuta possa essere affidabile. Per ottenere una tale stima si può calcolare l'accuratezza del classificatore su un corpus di testi di paternità certa, come i due corpus di *training* che abbiamo descritto nella sezione precedente. Il procedimento che seguiamo per ottenere tale stima di accuratezza consiste nel chiedere al classificatore di determinare, per ogni testo t nel corpus e per ogni autore a ivi rappresentato, se a sia o meno l'autore di t (una volta che si sia ovviamente nascosta al classificatore la vera identità dell'autore di t). Se il corpus contiene n testi, per complessivi m autori, andranno effettuate $n*m$ determinazioni di autorialità, e una stima dell'accuratezza del classificatore potrà essere ottenuta calcolando quante di queste $n*m$ determinazioni si siano rivelate esatte. Naturalmente, quando si sperimenta il classificatore sul testo t , tale testo non può aver fatto parte dell'insieme di *training* su cui il classificatore è stato allenato, in quanto qualsiasi classificatore ha maggiori probabilità di catalogare correttamente un testo sul quale sia stato allenato, piuttosto che un testo mai incontrato precedentemente. La nostra sperimentazione consiste quindi nell'allenare, per ciascun testo t del corpus, un classificatore sui restanti $(n-1)$ testi, e usarlo per classificare t . Si noti comunque che i testi di *training* sono tutti i testi

⁷ Reperibile al link: <https://www.nltk.org/>.

Tabella 3. Corpora risultanti dalla segmentazione dei testi training.		
	Corpus per Ep I	Corpus per Ep II
Numero di testi di training iniziale	294	30
Numero di testi di training risultante (testi originali + segmenti)	1.310	12.312

risultanti dall'operazione di segmentazione descritta al termine della sezione precedente, mentre gli n testi che classifichiamo sono solo i testi originali (escludendo quindi i segmenti). Il motivo di quest'ultima scelta è che, poiché al nostro sistema di AV viene chiesto di verificare l'autorialità di testi interi, e non di loro frammenti, è necessario sperimentare la sua accuratezza su testi di natura analoga, al fine di ottenere una stima attendibile.

Dato un autore a , misuriamo l'accuratezza del sistema nel determinare quali degli n testi sia stato o meno scritto da a con la funzione F_i , la funzione matematica standard per la misura dell'accuratezza di classificazione. I valori di F_i variano fra 0 (accuratezza pessima) e 1 (accuratezza ottima). Dato un corpus in cui siano rappresentati m autori (nel nostro caso: 5 autori per il corpus relativo a Ep I, 20 autori per quello relativo a Ep II)⁸, esistono due modi alternativi per calcolare F_i su tutto il corpus. Il primo, detto Macro- F_i , consiste nel calcolare il valore di F_i autore per autore ed eseguire quindi la media aritmetica fra gli m valori risultanti. Il secondo, detto Micro- F_i , consiste nel calcolare F_i globalmente sulle $n*m$ determinazioni di autorialità⁹. In questo studio presenteremo i risultati per entrambe le misure.

La tabella 4 riporta i valori così ottenuti per le due parti dell'*Epistola XIII*, ove i risultati fra parentesi sono invece i valori medi che otterrebbe un ipotetico sistema di AV che determinasse l'autorialità in modo completamente casuale.

⁸ Si noti che per gli autori rappresentati nella collezione di epistole di cancelleria di Petrus de Boateris (Schneider, 1926) non abbiamo generato alcun sistema di AV; infatti, essendo tale collezione una miscellanea di epistole di vari autori, ciascuno di essi è rappresentato da un numero troppo esiguo di testi per poter ragionevolmente apprendere per esso un sistema di AV. Di conseguenza, il test sul corpus Ep I è stato effettuato su $n=294$ testi e $m=5$ autori, per un totale di $294*5=1.470$ coppie testo-autore. Similmente, per Ep II sono stati effettuati test solo su autori con più di un testo positivo, quindi per 6 autori, da cui $30*6=180$ coppie testo-autore.

⁹ Per un approfondimento sull'argomento si veda Manning *et al.*, 2008, pp. 259-260.

Tabella 4. Valori di accuratezza nella <i>authorship verification</i> ottenuti sui corpora di tabella 3.		
	Macro- F_1	Micro- F_1
Ep I	0.88 (0.22)	0.98 (0.26)
Ep II	0.68 (0.15)	0.77 (0.15)

I risultati mostrano come i classificatori che abbiamo allenato, e in particolare quello per Ep I, siano caratterizzati da un'accuratezza piuttosto alta, il che conferisce un buon livello di credibilità alla determinazione che, mediante i medesimi metodi, viene fatta per l'*Epistola XIII*. Il confronto tra il valore di accuratezza del nostro classificatore e quello che verrebbe ottenuto mediante una determinazione assolutamente casuale dimostra la bontà dell'intuizione secondo cui i tratti linguistici che abbiamo studiato costituiscono un'importante fonte di evidenza per la determinazione di autorialità.

Gli esperimenti precedenti erano volti ad accertare, sul complesso degli autori rappresentati nel nostro corpus, la bontà delle tecniche di classificazione che abbiamo utilizzato. Con la tabella 5 concentriamo invece la nostra attenzione unicamente sui risultati che abbiamo ottenuto sull'autore Dante, e in particolare sulla dibattuta paternità dell'*Epistola XIII*.

La colonna $F_1(Dante)$ indica il livello di accuratezza ottenuto dal classificatore limitatamente all'attribuzione dei testi nei nostri due corpora alla classe *Dante*; mentre tale accuratezza è indubitabilmente buona per quanto riguarda il corpus Ep I, essa è decisamente più bassa sul corpus Ep II¹⁰. La colonna *Ipotesi* indica il risultato dell'attribuzione effettuata dal classificatore sulle due parti dell'*Epistola XIII*: in entrambi i casi il classificatore propende per l'ipotesi che il testo non abbia paternità dantesca. La colonna *Probabilità* indica (si veda il § 3.1) la probabilità che il classificatore soggettivamente attribuisce al fatto che il testo appartenga alla classe *Dante*; in entrambi i casi il fatto che tale probabilità sia inferiore a 0.5 indica per l'appunto che il classificatore propende per l'ipotesi non dantesca, con un maggior grado di certezza per Ep I (0.24) che per Ep II (0.39).

¹⁰ Nel caso del corpus Ep II la funzione F_1 penalizza il classificatore per non aver attribuito a Dante la *Monarchia*. Per quanto riguarda i 28 testi non danteschi, per 26 di essi il classificatore correttamente riconosce la paternità non dantesca, mentre erroneamente attribuisce a Dante le *Genealogie deorum* di Boccaccio e il commento di Pietro Alighieri.

Tabella 5. Risultati di *authorship verification* Dante vs. *NonDante*.

	$F_1(\text{Dante})$	Ipotesi	Probabilità(Dante)
Ep I	0.95 (0.08)	NonDante	0.24
Ep II	0.40 (0.12)	NonDante	0.39

Nel complesso, i dati sperimentali sostengono con maggior forza l'ipotesi formulata dal classificatore riguardo a Ep I che non quella riguardante Ep II. Nel caso di Ep I il classificatore ha un buon grado di certezza nella propria determinazione (0.24 – dove 0 indica il massimo grado di certezza per l'attribuzione *NonDante*), e soprattutto dimostra un ottimo grado di accuratezza nel classificare i testi del corpus Ep I (0.88 in Macro- F_1 e 0.98 in Micro- F_1 – dove 1 indicherebbe il 100% di determinazioni corrette). Più interlocutorio appare invece il risultato riguardante Ep II, dato che il classificatore ha una maggiore incertezza nella propria determinazione (0.39) e ha dimostrato un minor grado di accuratezza nel classificare i testi del relativo corpus (0.68 in Macro- F_1 e 0.77 in Micro- F_1).

Un ulteriore interessante esperimento che abbiamo condotto è stato quello di chiedere al classificatore, per entrambe le parti dell'*Epistola XIII*, per ciascun paragrafo p della relativa parte e per ciascun autore a rappresentato nel relativo corpus, di valutare la probabilità che a sia l'autore di p . I risultati sono visualizzati nelle figure 2-3, la prima per Ep I e la seconda per Ep II, dove ogni coppia autore-paragrafo è rappresentata da una casella di tonalità graduata fra il bianco e il nero. Più il tono di grigio è scuro, più è probabile che il paragrafo sia stato scritto dall'autore; più il tono di grigio è chiaro, meno è probabile. La prima riga di ciascuna figura (etichettata "Full") riporta l'analogo dato per il documento (Ep I o Ep II) nel suo complesso. Entrambe le figure 2-3 suggeriscono che, nonostante il classificatore ritenga (come riportato in precedenza) che né Ep I né Ep II abbiano paternità dantesca, esso indica Dante, per entrambi i testi, fra gli autori "meno improbabili". Per l'Ep I, in testa alla "classifica di minor improbabilità" troviamo Clara Assisiensis, seguita da Dante e (a maggior distanza) Giovanni Boccaccio, mentre il classificatore ritiene altamente improbabile che Ep I sia della mano di Guido Faba o di quella di Pier della Vigna. Dante risulta invece di gran lunga l'autore "meno improbabile" di Ep II; curiosa è la similarità rilevata dal classificatore fra alcuni paragrafi di Ep II e la produzione rappresentata nel corpus di Filippo Villani e Guido da Pisa.

Un'ulteriore visualizzazione è quella offerta nelle figure 2-3 che seguono, dove i vari paragrafi dell'*Epistola XIII* sono stati classificati uno

a uno dal sistema di AV, ed evidenziati in termini di maggiore o minore similarità, sulla base dei tratti linguistici da noi analizzati, con la produzione dantesca: più il paragrafo tende allo scuro più l'algoritmo lo ritiene dantesco, più il paragrafo tende al chiaro meno l'algoritmo lo ritiene dantesco.

- Ep I:

P1: Magnifico atque uictorioso domino domino Cani Grandi de la Scala, sacratissimi cesarei principatus in urbe Uerona et ciuitate Uincencie Uicario generali, deuotissimus suus Dantes Alagherii, Florentinus natione non moribus, uitam orat per tempora diuturna felicem et gloriosi nominis perpetuum incrementum. **P2:** Inclita uestre magnificentie laus, quam fama uigil uolitando disseminat, sic distrahit in diuersa diuersos, ut hos in spem sue prosperitatis attollat, hos exterminii deiciat in terrorem. Huius quidem preconium, facta modernorum exsuperans, tanquam ueri existentia latius, arbitratur aliquando superfluum. **P3:** Uerum, ne diuturna me nimis incertitudo suspenderet, uelut Austri regina Ierusalem petiit, uelut Pallas petiit Elicona, Ueronam petii fidis oculis discussurus audita, ibique magnalia uestra uidi, uidi beneficia simul et tetigi; et quemadmodum prius dictorum ex parte suspicabar excessum, sic posterius ipsa facta excessiua cognoui. Quo factum est ut ex auditu solo cum quadam animi subiectione beniuolus prius exstiterim, sed ex uisu postmodum deuotissimus et amicus. **P4:** Nec reor amici nomen assumens, ut nonnulli forsitan obiectarent, reatum presumptionis incurrere, cum non minus dispares connectantur quam pares amicitie sacramento. Nam si delectabiles et utiles amicitias inspicere libeat, illis persepius inspicienti patebit preheminentes inferioribus coniugari personas. **P5:** Et si ad ueram ac per se amicitiam torqueatur intuitus, nonne illustrium summorumque principum plerunque uiros fortuna obscuros, honestate preclaros, amicos fuisse constabit? Quidni, cum etiam Dei et hominis amicitia nequaquam impediatur excessu? **P6:** Quod si cuiquam quod asseritur nunc uideretur indignum, Spiritum Sanctum audiat amicitie sue participes quosdam homines profitentem. Nam in Sapientia de sapientia legitur, quoniam. **P7:** Sed habet imperitia uulgi sine discretione iudicium; et quemadmodum solem pedalis magnitudinis arbitratur, sic et circa mores uana credulitate decipitur. Nos autem, quibus optimum quod est in nobis noscere datum est, gregum uestigia sectari non decet, quin ymo suis erroribus obuiare tenemur. Nam intellectu atque ratione degentes, diuina quadam libertate dotati, nullis consuetudinibus astringuntur; nec mirum, cum non ipsi legibus, sed ipsis leges potius dirigantur. **P8:** Liquet igitur quod superius dixi, me scilicet esse deuotissimum et amicum, nullatenus esse presumptum. **P9:** Preferens ergo amicitiam uestram quasi thesaurum carissimum,

prouidentia diligenti et accurata sollicitudine illam seruare desidero. **P10:** Itaque, cum in dogmatibus moralis negotii amicitiam adequari et saluari analogo doceatur, ad retribuendum pro collatis beneficiis plus quam semel analogiam sequi michi uotiuum est; et propter hoc munuscula mea sepe multum consexi et ab inuicem segregauit nec non segregata percensui, dignius gratiusque uobis inquirens. **P11:** Neque ipsi preheminentie uestre congruum magis comperi quam Comedie sublimem canticam que decoratur titulo Paradisi; et illam, sub presenti epistola, tanquam sub epigrammate proprio dedicatam, uobis ascribo, uobis offero, uobis denique recomendo. **P12:** Illud quoque preterire silentio simpliciter inardescens non sinit affectus, quod in hac donatione plus dono quam domino et honoris et fame conferri potest uideri. Quidni cum eius titulum iam presagiam de gloria uestri nominis ampliandum? **P13:** Satis actenus uidebar expressisse quod de proposito fuit; sed zelus gratie uestre, quam sitio quasi uitam paruipendens, a primordio metam prefixam urget ulterius. Itaque, formula consumata epistole, ad introductionem oblatis operis aliquid sub lectoris officio compendiose aggrediar.

- Ep II:

P14: Sicut dicit Philosophus in secundo Metaphysicorum; cuius ratio est quia ueritas de re, que in ueritate consistit tanquam in subiecto, est similitudo perfecta rei sicut est. **P15:** Eorum uero que sunt, quedam sic sunt ut habeant esse absolutum in se; quedam sunt ita ut habeant esse dependens ab alio per relationem quandam, ut ea quorum esse est ad aliud se habere ut relatiua: sicut pater et filius, dominus et seruus, duplum et dimidium, totum et pars, et huiusmodi in quantum talia. **P16:** Propterea quodque esse talium dependet ab alio, consequens est quod eorum ueritas ab alio dependeat; ignorato enim dimidio, nunquam cognoscitur duplum, et sic de aliis. **P17:** Uolentes igitur aliqualem introductionem tradere de parte operis alicuius, oportet aliquam notitiam tradere de toto cuius est pars. Quapropter et ego, uolens de parte supra nominata totius Comedie aliquid tradere per modum introductionis, aliquid de toto opere premitendum existimaui, ut facilius et perfectior sit ad partem introitus. **P18:** Sex igitur sunt que in principio cuiusque doctrinalis operis inquirenda sunt, uidelicet subiectum, agens, forma, finis, libri titulus et genus philosophie. De istis tria sunt in quibus pars ista quam uobis destinare proposui, uariatur a toto, scilicet subiectum, forma et titulus; in aliis uero non uariatur, sicut apparet inspicienti; et ideo circa considerationem de toto ista tria inquirenda seorsum sunt: quo facto, satis patebit ad introductionem partis. **P19:** Deinde inquiremus alia tria non solum per respectum ad totum, sed etiam per respectum ad ipsam partem oblatam. **P20:** Ad eui-

dentiam itaque dicendorum sciendum est quod istius operis non est simplex sensus, ymo dici potest polisemos, hoc est plurium sensuum; nam primus sensus est qui habetur per litteram, alius est qui habetur per significata per litteram. Et primus dicitur litteralis, secundus uero allegoricus, siue moralis siue anagogicus. **P21:** Qui modus tractandi, ut melius pateat, potest considerari in istis uersibus. Nam si ad litteram solam inspiciamus, significatur nobis exitus filiorum Israel de Egipto tempore Moysis; si ad allegoriam, significatur nobis nostra redemptio facta per Christum; si ad moralem sensum, significatur nobis conuersio anime de luctu et miseria peccati ad statum gratie; si ad anagogicum, significatur exitus anime sancte ab huius corruptionis seruitute ad eternam glorie libertatem. **P22:** Et quomodo isti sensus mystici uariis appellantur nominibus, generaliter omnes dici possunt allegorici, cum sint a litterali siue historiali diuersi. Nam allegoria dicitur ab alleon grece, quod in latinum dicitur alienum, siue diuersum. **P23:** Hiis uersibus, manifestum est quod duplex oportet esse subiectum, circa quod currant alterni sensus. Et ideo uidendum est de subiecto huius operis prout ad litteram accipitur; deinde de subiecto prout allegorice sententiatur. **P24:** Est ergo subiectum totius operis, litteraliter tantum accepti, status animarum post mortem simpliciter sumptus; nam de illo et circa illum totius operis uersatur processus. **P25:** Si uero accipiatur opus allegorice, subiectum est homo prout merendo et demerendo per arbitrii libertatem iustitie premiandi et puniendi obnoxius est. **P26:** Forma uero est duplex: forma tractatus et forma tractandi. Forma tractatus est triplex, secundum triplicem diuisionem. Prima diuisio est, qua totum opus diuiditur in tres canticas. Secunda, qua quelibet cantica diuiditur in cantus. Tertia, qua quilibet cantus diuiditur in rithimos. **P27:** Forma siue modus tractandi est poeticus, fictiuus et descriptiuus, digressiuus, transumptiuus, et cum hoc diffinitiuus, diuisiuus, probatiuus, improbatiuus et exemplorum positiuus. **P28:** Libri titulus est. Ad cuius notitiam sciendum est quod comedia dicitur a comos quod est uilla et oda quod est cantus, unde comedia quasi uillanus cantus. **P29:** Et est comedia genus quoddam poetice narrationis ab omnibus aliis differens. Differt ergo a tragedia in materia per hoc, quod tragedia in principio est admirabilis et quieta et in fine seu exitu fetida et horribilis; et dicitur propter hoc a tragos, quod est hircus, et oda, quasi cantus hircinus, idest fetidus ad modum hirci, ut patet per Senecam in suis tragediis. Comedia uero inchoat asperitatem alicuius rei, sed eius materia prospere terminatur, ut patet per Terentium in suis comediis. Et hinc consueuerunt dictatores quidam in suis salutationibus dicere loco salutis. **P30:** Similiter differunt in modo loquendi: elate et sublime tragedia, comedia uero remisse et humiliter, sicut uult Oratius in sua Poetria, ubi licentiat aliquando comicos ut tragedos loqui, et sic e conuerso. **P31:** Et per hoc patet quod Co-

media dicitur presens opus. Nam si ad materiam respiciamus, a principio horribilis et fetida est, quia Infernus, in fine prospera et desiderabilis et grata, quia Paradisus; si ad modum loquendi, remissus est modus et humilis, quia locutio uulgaris, in qua et muliercule comunicant. Et sic patet quare Comedia dicitur. **P32:** Sunt et alia genera narrationum poeticarum, sicut carmen bucolicum, elegia, satira, et sententia uotiuua, ut etiam per Oratium patere potest in sua Poetria; sed de istis ad presens nichil dicendum est. **P33:** Potest amodo patere quomodo assignandum sit subiectum partis oblate. Nam si totius operis litteraliter sumpti sic est subiectum, status animarum post mortem non contractus sed simpliciter acceptus, manifestum est quod hac in parte talis status est subiectum, sed contractus, scilicet status animarum beatarum post mortem. **P34:** Et si totius operis allegorice sumpti subiectum est homo prout merendo et demerendo per arbitrii libertatem est iustitie premiandi et puniendi obnoxius, manifestum est in hac parte hoc subiectum contrahi, et est homo prout merendo obnoxius est iustitie premiandi. **P35:** Et sic patet de forma partis per formam assignatam totius; nam si forma tractatus in toto est triplex, in hac parte tantum est duplex, scilicet diuisio cantuum et rithimorum. **P36:** Non eius potest esse propria forma diuisio prima, cum ista pars sit prime diuisionis. **P37:** Patet etiam libri titulus seu de libri titulo. Nam titulus totius libri est, et cetera; titulus autem huius partis est. **P38:** Inquisitionis hiis tribus in quibus uariatur pars a toto, uidendum est de aliis tribus in quibus nulla uariatio est a toto ad partem. Agens igitur totius et partis est ille qui dictus est, et totaliter uidetur esse. **P39:** Finis totius et partis esse posset et multiplex, scilicet propinquus et remotus; sed, ommissa subtili inuestigatione, dicendum est breuiter quod finis totius et partis est remouere uiuentes in hac uita de statu miserie et perducere ad statum felicitatis. **P40:** Genus uero phylosophie sub quo hic in toto et parte proceditur est morale negotium, siue ethica, quia non ad speculandum sed ad opus inuentum est totum et pars. **P41:** Nam si in aliquo loco uel passu pertractatur ad modum speculatiui negotii, hoc non est gratia speculatiui negotii, sed gratia operis; quia, ut ait Phylosophus in secundo Methaphysicorum, aliquando. **P42:** Hiis itaque premissis, ad expositionem littere secundum quandam prelibationem accedendum est, et illud prenuntiandum quod expositio littere nichil aliud est quam forme operis manifestatio. **P43:** Diuiditur ergo ista pars, seu tertia cantica, que Paradisus dicitur, principaliter in duas partes, scilicet in prologum et partem executiuam. Pars secunda incipit ibi, quasi in medio primi. **P44:** De parte prima sciendum est quod, quamuis comuni ratione posset dici exordium, proprie autem loquendo non debet dici nisi prologus; quod Phylosophus in tertio Rhetoricorum uidetur innuere, ubi dicit quod. **P45:** Est etiam prenotandum quod preuiatio ista, que comuniter exordium dici potest,

aliter fit a poetis, aliter fit a rhetoribus. **P46:** Rethores enim consueuere prelibare dicenda ut animum comparent auditoris; sed poete non solum hoc faciunt, quin ymo post hec inuocationem quandam emittunt. **P47:** Et hoc est eis conueniens, quia multa inuocatione opus est eis, cum aliquid contra comunem modum hominum a superioribus substantiis petendum est, quasi diuinum quoddam munus. **P48:** Ergo presens prologus diuiditur in partes duas, quia in prima premittitur quid dicendum sit, in secunda inuocatur Apollo; et incipit secunda pars ibi. **P49:** Propter primam partem notandum quod ad bene exordiendum tria requiruntur, ut dicit Tullius in Noua Rethorica, scilicet ut beneuolum, attentum et docilem reddat aliquis auditorem; et hoc maxime in admirabili genere cause, ut ipsemet Tullius dicit. **P50:** Cum ergo materia circa quam uersatur presens tractatus sit admirabilis, et propterea ad admirabile reducenda, ista tria intenduntur in principio exordii siue prologi. Nam dicit se dicturum ea que qui uidit retinere non potuit in primo celo; in quo dicto omnia illa tria comprehenduntur. **P51:** Nam in utilitate dicendorum beneuolentia paratur, in admirabilitate attentio, in possibilitate docilitas. Utilitatem innuit, cum recitaturum se dicit ea que maxime allectiua sunt desiderii humani, scilicet gaudia Paradisi; admirabilitatem tangit, cum promittit se tam ardua, tam sublimia dicere, scilicet condiciones regni celestis; possibilitatem ostendit, cum dicit se dicturum que mente retinere potuit: si enim homo ipse, et alii poterunt. **P52:** Hec omnia tanguntur in uerbis illis ubi dicit se fuisse in primo celo, et quod dicere uult de regno celesti quicquid in mente sua quasi thesaurum potuit retinere. Uiso igitur de bonitate ac perfectione prime partis prologi, ad litteram accedatur. **P53:** Dicit ergo quod gloria primi Motoris, qui Deus est, in omnibus partibus uniuersi resplendet, sed ita ut in aliqua parte magis et in aliqua minus. **P54:** Quod autem ubique resplendeat, ratio et auctoritas manifestat. Ratio sic: omne quod est, aut habet esse a se, aut ab alio; sed constat quod habere esse a se non conuenit nisi uni, scilicet primo seu principio, qui Deus est, cum habere esse a se arguat per se necesse esse, et per se necesse esse non competat nisi uni, scilicet primo seu principio, quod est causa omnium; ergo omnia que sunt, preter unum, habent esse ab aliis. **P55:** Si ergo accipiatur ultimum in uniuerso, non quodcunque, manifestum est quod id habet esse ab aliquo; et illud a quo habet, habet a se uel ab aliquo. Si a se, sic est primum; si ab aliquo, et illud similiter uel a se uel ab aliquo. Et cum non aliud esset sic procedere quam procedere in infinitum in causis agentibus, ut probatur in secundo Metaphysicorum, erit deuenire ad primum, qui Deus est. **P56:** Et sic, mediate uel inmediate, omne quod habet esse habet esse ab eo; quia ex eo quod causa secunda recipit a prima, influit super causatum ad modum recipientis et reddentis radium, propter quod causa prima magis est causa. **P57:** Et hoc dicitur in libro De Causis quod. Sed

hoc quantum ad esse. **P58:** Quantum uero ad essentiam, probo sic: omnis essentia, preter primam, est causata, alias essent plura que essent per se necesse esse, quod est impossibile. Quod causatum, est causatum uel a natura uel ab intellectu, et quod a natura, per consequens causatum est ab intellectu, cum natura sit opus intelligentie; omne ergo quod est causatum, est causatum ab aliquo intellectu uel mediate uel immediate. **P59:** Cum ergo uirtus sequatur essentiam cuius est uirtus, si essentia intellectiua est causata, uirtus essentie intellectiue est tota et unius quod causat. Et sic, quemadmodum prius deuenire erat ad primam causam ipsius esse, sic nunc essentie et uirtutis. **P60:** Propter quod patet quod omnis essentia et uirtus procedit a prima, et intelligentie inferiores recipiunt quasi a radiante et reddunt radios superioris ad suum inferius ad modum speculorum. Quod satis aperte tangere uidetur Dionysius de celesti hierarchia loquens. **P61:** Et propter hoc dicitur in libro De Causis quod. Patet ergo quomodo ratio manifestat diuinum lumen, id est diuinam bonitatem, sapientiam et uirtutem, resplendere ubique. **P62:** Similiter etiam et scientius facit auctoritas. Dicit enim Spiritus Sanctus per Hieremiam; et in Psalmo, et cetera. Et Sapientia dicit quod. Et Ecclesiasticus quadragesimo secundo. **P63:** Quod etiam scriptura paganorum contestatur; unde Lucanus in nono. **P64:** Bene ergo dictum est cum dicit quod diuinus radius siue diuina gloria: penetrat quantum ad essentiam, resplendet quantum ad esse. **P65:** Quod autem subicit de magis et minus, habet ueritatem in manifesto, quoniam uidemus in aliquo excellentiori gradu essentiam aliquam, aliquam uero in inferiori; ut patet de celo et elementis, quorum quidem illud incorruptibile, illa uero corruptibilia sunt. **P66:** Et postquam premisit hanc ueritatem, prosequitur ab ea circumloquens Paradisum; et dicit quod fuit in celo illo quod de gloria Dei, siue de luce, recipit affluentius. **P67:** Propter quod sciendum quod illud celum est celum supremum: continens corpora uniuersa et a nullo contentum; intra quod omnia corpora mouentur, ipso in sempiterna quiete permanente; uirtuans et omnia sua contenta et a nulla corporali substantia uirtutem recipiens. **P68:** Et dicitur Empyreum, quod est idem quod celum igne sui ardoris flagrans; non quod in eo sit ignis uel ardor materialis, sed spiritualis, quod est amor sanctus siue caritas. **P69:** Quod autem de diuina luce plus recipiat, potest probari per duo: primo, per suum omnia continere et a nullo contineri; secundo, per sempiternam suam quietem siue pacem. **P70:** Quantum ad primum, probatur sic: continens se habet ad contentum in naturali situ sicut formatiuum ad formabile, ut habetur in quarto Physicorum; sed in naturali situ totius uniuersi primum celum est omnia continens; ergo se habet ad omnia sicut formatiuum ad formabile, quod est se habere per modum cause. Et cum omnis uis causandi sit radius quidam influens a prima causa que Deus est, manifestum est quod illud celum quod

magis habet rationem cause magis de luce diuina recipit. **P71:** Quantum ad secundum, probatur sic: omne quod mouetur, mouetur propter aliquid quod non habet, quod est terminus sui motus; sicut celum lune mouetur propter aliquam partem sui, que non habet illud ubi ad quod mouetur; et quia sui pars quelibet non adepto quolibet ubi, quod est impossibile, mouetur ad aliud, inde est quod semper mouetur et nunquam quiescit, et est eius appetitus. Et quod dico de celo lune intelligendum est de omnibus preter primum. Omne ergo quod mouetur est in aliquo defectu et non habet totum suum esse simul. **P72:** Illud igitur celum quod a nullo mouetur, in se in qualibet sui parte habet quicquid potest modo perfecto, ita quod motu non indiget ad suam perfectionem. Et cum omnis perfectio sit radius primi, quod est in summo gradu perfectionis, manifestum est quod celum primum magis recipit de luce primi, qui est Deus. **P73:** Ista tamen ratio uidetur arguere a destructione antecedentis, ita quod simpliciter et secundum formam arguendi non probat. Sed si consideremus materiam eius, bene probat, quia de quodam sempiterno, in quo potest defectus sempiternari: ita quod, si Deus non dedit sibi motum, patet quod non dedit sibi materiam in aliquo egentem. **P74:** Et per hanc suppositionem tenet argumentum ratione materie; et est similis modus arguendi ac si dicerem: Si homo est, est risibile; nam in omnibus conuertibilibus tenet similis ratio gratia materie. Sic ergo patet: cum dicit, intelligit circumloqui Paradisum, siue celum Emphyreum. **P75:** Premissis quoque rationibus consonanter dicit Phylosophus in primo De celo ubi dicit quod celum. **P76:** Ad hoc etiam posset adduci quod dicit Apostolus ad Ephesios de Christo. Hoc est celum deliciarum Domini, de quibus delitiis dicitur contra Luciferum per Ezechielem. **P77:** Et postquam dixit quod fuit in loco illo Paradisi per suam circumlocutionem, prosequitur dicens se uidisse aliqua que recitare non potest qui descendit. Et reddit causam dicens quod intellectus in tantum profundat se in ipsum desiderium suum, quod est Deus, quod memoria sequi non potest. **P78:** Ad que intelligenda sciendum est quod intellectus humanus in hac uita, propter connaturalitatem et affinitatem quam habet ad substantiam intellectualem separatam, quando eleuatur, in tantum eleuatur, ut memoria, post reditum, deficiat propter transcendisse humanum modum. **P79:** Et hoc insinuat nobis per Apostolum ad Corinthios loquentem, ubi dicit. Ecce, postquam humanam rationem intellectus ascensione transierat, quid extra se ageretur non recordabatur. **P80:** Hoc etiam insinuat nobis in Matheo, ubi tres discipuli, nichil postea recitantes, quasi obliui. Et in Ezechiele scribitur. Et ubi ista inuidis non sufficiant, legant Richardum de Sancto Uictore in libro De Contemplatione, legant Bernardum in libro De Consideratione, legant Augustinum in libro De Quantitate Anime, et non inuidebunt. **P81:** Si uero in dispositionem eleuationis tante propter peccatum loquen-

tis oblatrarent, legant Danielelem, ubi et Nabuchodonosor inuenient contra peccatores aliqua uidisse diuinitus, obliuionique mandasse. **P82:** Nam, aliquando misericorditer ad conuersionem, aliquando seure ad punitiorem, plus et minus, ut uult, gloriam suam quantumcunque male uiuentibus manifestat. **P83:** Uidit ergo, ut dicit, aliqua que referre nescit et nequit rediens. Diligenter quippe notandum est quod dicit: nescit quia oblitus, nequit quia, si recordatur et contentum tenet, sermo tamen deficit. **P84:** Multa namque per intellectum uidemus quibus signa uocalia desunt: quod satis Plato insinuat in suis libris per assumptionem metapho-rismorum; multa enim per lumen intellectuale uidit que sermone proprio nequiuit exprimere. **P85:** Postea dicit se dicturum illa que de regno celesti retinere potuit, et hoc dicit esse materiam sui operis; que qualia sint et quanta, in parte executiua patebit. **P86:** Deinde cum dicit, facit inuocationem suam. Et diuiditur ista pars in partes duas: in prima inuocando petit, in secunda suadet Apollini petitionem factam, remunerationem quandam prenuntians; et incipit secunda pars ibi. **P87:** Prima pars diuiditur in partes duas: in prima petit diuinum auxilium, in secunda tangit necessitatem sue petitionis, quod est iustificare ipsam, ibi. **P88:** Hec est sententia secunde partis prologi in generali; in speciali uero non exponam ad presens. Urget enim me rei familiaris angustia, ut hec et alia utilia reipublice derelinquere oporteat. Sed spero de magnificentia uestra ita ut alias habeatur procedendi ad utilem expositionem facultas. **P89:** In parte uero executiua, que fuit diuisa contra totum prologum, nec diuidendo nec sententiando quicquam dicitur ad presens, nisi hoc, quod ubique proceditur ascendendo de celo in celum, et recitatur de animabus beatis inuentis in quolibet orbe. Et quia illa uera beatitudo in sentiendo ueritatis principio consistit – ut patet per Iohannem ibi, et cetera; et per Boetium in tertio De Consolatione ibi –, inde est quod, ad ostendendum gloriam beatitudinis in illis animabus, ab eis tanquam uidentibus omnem ueritatem multa queruntur, que magnam habent utilitatem et delectationem. **P90:** Et quia, inuento principio seu primo, uidelicet Deo, nichil est quod ulterius queratur, cum sit Alpha et O, idest principium et finis, ut uisio Iohannis designat, in ipso Deo terminatur tractatus, qui est benedictus in secula seculorum.

5. Osservazioni conclusive

Abbiamo presentato quello che riteniamo ancora uno studio preliminare sull'applicazione di tecniche di *computational authorship verification* (AV) al problema dell'autorialità dell'*Epistola XIII*. Il meccanismo di AV che abbiamo approntato, e a cui abbiamo chiesto un'ipotesi riguardo alla paternità dantesca o meno delle due sezioni (nuncupatoria

ed esegetica) dell'*Epistola XIII*, ritiene entrambe le sezioni non dantesche (tabella 5), anche se con diversi gradi di certezza (maggiore per la prima sezione, minore per la seconda). Che all'ipotesi generata dal sistema e riguardante Ep I vada attribuito maggior credito che non a quella riguardante Ep II discende anche dal fatto che, dei due separati classificatori che abbiamo approntato (uno per ciascuna sezione), il primo si dimostra più accurato del secondo (con l'altissima accuratezza di 0.98, misurata mediante la funzione $\text{Micro-}F_1$) una volta applicato a verificare l'autorialità di tutte le coppie (autore, testo) comprese nel relativo corpus (tabella 4).

Da tutta la procedura di ricerca fin qui descritta saranno risultate evidenti le differenze macroscopiche che sussistono fra le ricerche sull'attribuzione sviluppate secondo i metodi filologici, in senso lato, e le ricerche sull'attribuzione, o sulla verifica, di autorialità fondate sul *machine learning*. Si tratta di procedure di ricerca radicalmente diverse.

Le ricerche di stampo filologico, in senso lato, si fondano su argomenti interni ai testi (per esempio, se una certa idea o affermazione presente nel testo sia coerente o compatibile con quanto sappiamo del pensiero dell'autore; o se una certa espressione presente nel testo sia coerente o compatibile con la lingua e lo stile dell'autore); o su argomenti che mettono in rapporto i contenuti e le forme del testo con il contesto o i contesti nei quali esso può essere stato composto, fruito, trasmesso (per esempio, se questo o quel dato del testo, o il testo nel suo insieme, o la tradizione del testo, siano compatibili con le circostanze biografiche nelle quali l'autore potrebbe averlo composto); o su analoghi argomenti qualitativi. La storia delle controversie sulla nostra *Epistola XIII* fornisce abbondantissima esemplificazione di questo tipo di argomenti filologici in senso lato, pro e contro la paternità dantesca: argomenti di varia natura, più di carattere locale o di portata generale, elaborati da una intelligenza umana, su vari livelli di astrazione e di complessità, in base a criteri logici espliciti o esplicitabili così da tendere – idealmente, auspicabilmente – a risultati “falsificabili”.

Invece le ricerche computazionali di *authorship analysis*, o *attribution*, o *verification*, come quella qui descritta, operano su caratteristiche oggettive dei testi, estese sugli interi testi, e preferibilmente su caratteristiche materiali di livello basico quali le sequenze ricorrenti di caratteri, la lunghezza delle parole o delle frasi, ecc., e rigorosamente quantitative, come quelle appunto elencate al § 3.2.

Si potrebbe osservare che caratteristiche come queste rientrano, in linea di principio, in una categoria ben nota al metodo filologico, quella di *usus scribendi*, ma si tratta di caratteristiche molto diverse da quelle

sulle quali il filologo è abituato a fermare il suo sguardo; anzi si tratta di caratteristiche non visibili a occhio nudo, perché emergono solo statisticamente.

Una prima conclusione che si può trarre da queste macroscopiche differenze è che le procedure e i risultati della *authorship analysis*, in senso lato, si collocano su un canale parallelo e – almeno inizialmente – non comunicante con le procedure e i risultati della discussione scientifica fondata su argomenti filologici, sempre in senso lato. In altre parole, le decisioni a cui perviene la *authorship analysis* né avvalorano né inficiano, nel nostro caso, la validità di alcun singolo argomento addotto pro o contro la paternità dantesca.

Ma – ragionando ora con la logica del filologo – che validità dobbiamo riconoscere alle decisioni a cui perviene la *authorship analysis*, fondate su una logica talmente diversa da quella che ci è abituale? Se guardiamo alle caratteristiche su cui essa opera (caratteristiche oggettive, materiali, basiche, meccaniche, quantitative) è legittimo il dubbio se e quanto esse davvero racchiudano “l'impronta digitale” dell'autore. E, se guardiamo alla logica con cui la macchina, o meglio il classificatore, basandosi sulle caratteristiche di cui sopra, trae le sue conclusioni – logica che non siamo sempre in grado di controllare fino in fondo – è più che legittimo il dubbio se e quanto tale logica, abbastanza imperscrutabile, sia affidabile.

La risposta che si può dare a entrambi questi dubbi è che, prima di applicare i suoi strumenti di decisione al caso dubbio, il classificatore li ha applicati a casi non dubbi – cioè a domande la cui risposta noi conosciamo ma abbiamo tenuta nascosta al classificatore – e i risultati a cui è pervenuto sono corretti, in misura maggiore o minore. E appunto la misura in cui questi risultati sono corretti – ciò che al § 4 è designato col termine “accuratezza”, ovvero la funzione F_1 (Macro- e Micro- F_1) – ci dà la misura dell'affidabilità del classificatore. La stessa affidabilità, dunque, dobbiamo assumere che il classificatore abbia anche nel darci la sua risposta alla domanda cruciale per la quale abbiamo attivato tutta la procedura. Anche se non sappiamo bene *perché* il classificatore propenda per l'ipotesi della paternità dantesca o non dantesca, dobbiamo prendere atto che, prima di darci, nel suo linguaggio numerico, questa risposta, il classificatore si è guadagnato una non trascurabile credibilità. Meglio: una credibilità onestamente quantificata. E il fatto che il classificatore tragga le sue conclusioni secondo una logica del tutto indipendente dalla nostra, e sulla base di dati empirici del tutto diversi dai nostri, aumenta il valore del controllo incrociato fra i nostri e i suoi risultati.

Una ulteriore conferma di quanto la logica della macchina sia indipendente dalla nostra logica di filologi – e quindi sia interessante,

perché ne possiamo ricavare stimoli impreveduti – sta nel fatto che la macchina, che propende per l'ipotesi *NonDante* sia per la prima parte che per la seconda parte dell'Epistola (Ep I e Ep II), propende per questa ipotesi più per Ep I che per Ep II. Volendo tradurre i responsi numerici dati al §4 in linguaggio umano colloquiale – e ci scusiamo se questo suona troppo informale – potremmo dire che, in entrambi i casi, la risposta alla domanda “È di Dante o no?” è: “Più no che sì”. Ma, delle due, la parte “meno” dantesca dell'*Epistola* è la prima.

Questo risultato è certamente impreveduto a partire dallo stato della discussione su base (sempre in senso lato) filologica. Infatti ci sono studiosi convinti che l'*Epistola XIII* sia tutta dantesca, studiosi convinti che sia tutta non-dantesca, e studiosi che tendono a ritenere la prima parte dantesca e la seconda no. Ma non c'è nessuno, ovviamente, che ritenga dantesca la seconda parte e non la prima, perché questa combinazione sarebbe storicamente del tutto implausibile. Il responso impreveduto della verifica computazionale su questo punto, dunque, potrebbe far propendere per la totale inautenticità dell'*Epistola*, ma qui ci interessa sottolineare il fatto che la verifica si conferma di natura non affine e non riducibile al modo di ragionare, in senso lato, di noi filologi.

Per finire, ribadiamo ancora una volta che si tratta di risultati preliminari, all'interno di una ricerca *in fieri* che potrà progredire in vari modi, per esempio:

- aumentando i testi di confronto, e costituendo quindi corpora più ampi e più rappresentativi su cui allenare e sperimentare i nostri classificatori;
- includendo nuove *features* che codifichino elementi linguistici di secondo livello, non immediatamente coincidenti con sequenze di caratteri, estraibili dai testi grazie all'applicazione di strumenti computazionali di *natural language processing* relativi alla lingua latina: per esempio, suffissi (nel senso di morfemi, non di sequenze materiali di caratteri), o marcatori delle parti del discorso (*POS tags*), o indici di frequenza delle parole;
- esplorando la bibliografia sul *machine learning* in cerca di strade nuove per superare lo stato dell'arte.

6. Postilla: la cosiddetta «XIV Epistola di Dante Alighieri»

Le parole tra virgolette sono tratte dal titolo dell'articolo di Paolo Pellegrini *La quattordicesima epistola di Dante Alighieri. Primi appunti per una attribuzione*, pubblicato nella rivista online «StEFI. Studi di Erudizione e di Filologia Italiana» VII (2018b [ma 2019]), pp. 5-20. La

tesi sostenuta da Pellegrini è quella dichiarata nell'abstract – ma il concetto di ipotesi espresso nell'abstract non compare nel titolo perentorio:

In summer 1312 Cangrande della Scala sent a very important epistle to the Emperor Henry VIIth: this epistle displays the same range and the same sequence of sources which can be found both in a document drafted in 1306 whose content is partially attributed to Dante's pen and in a couple of passages of the *Monarchia*. These elements, together with further evidences, corroborate the hypothesis that Cangrande's epistle was actually dictated by Dante himself.

Ecco il testo della breve epistola, tratto dall'edizione Zaccagnini 1926, pp. 212-213¹¹. I paragrafi sono ombreggiati secondo la medesima convenzione seguita nel § 4:

P1: Illustrissimo domino Henricho, inclyto Romanorum imperatori et semper augusto, C. capitaneus ueronensis deuotione fidelitatis continua semper insistere uotis suis. Cum serena pacis tranquillitas, decora genitrix artium et alumna, multiplicet et dilatet quam plurimum commoda populorum, cura uigili procurare tenetur cuiuslibet principatus intencio que sonoro laudis preconio desiderat predicari, ut inuiolatus permaneat status pacificus subiectorum. **P2:** Nam, ut lectio testatur diuina, illud imperium, illud regnum quod diuersis uoluntatibus intercisum in se non continet unionem, desolationem incurrit, nec in illo corpore sospitatis ilaritas perseuerat, cuius partes uel membra passionibus aliquibus singulariter affliguntur. **P3:** Quippe recenter uobis hoc notifico euenisse, quod quidam iniquitatis alumni, uasa scelerum et putei uitiorum, quorum propositum est clandestinum et nefandum, sub cuius effectus specie imperiale decus corruere moluntur, quod absit, inter uirum magnificum, dominum P., inclytum principem Achaie, et hominem excelse potentie, dominum G., comitem, quos in istis partibus prefeceratis in presides et rectores, malignis afflatibus seminauerunt de nouo semen et materiam uictiorum, ita quod uterque ipsorum, cum suorum comitiua sequacium, contentionum ardoribus concitatus, ad prouintiam alterius prorumpere iam presunt multotiens. **P4:** Itaque fere iam partis cuiuslibet acies concurrissent, conquassatis capitibus plurimorum, nisi forent quorundam

¹¹ Lo Zaccagnini aveva già studiato questa raccolta di lettere, contenuta nel ms. Magliabechiano II.IV.312 della Bibl. Nazionale di Firenze, in Zaccagnini 1924; essa venne contemporaneamente pubblicata anche da Schneider 1926, pp. 265-266, dal quale la riproduce Pellegrini.

magnatum fidelium imperii suadele que ad salutem et robur imperialis diadematis aspirantes, pro uiribus studuerunt extinguere iracundiam iam conceptam, quod nondum tamen efficaciter potuerunt, malignante diabolo, bonorum operum subuersore, propter quod prouincia Lombardorum tota concutitur tremebunda timore, ne causa huius scandali lanietur, quassantibus inimicis propter casum huiusmodi, dum ex hoc cogitant euenire quod iam pridem attentius desideratis affectibus cupiuerunt. **P5:** Studeat igitur imperatoria celsitudo sui maturitate consilii has radices amarissimas et pericula submouere. Nam si membra talia uestri gubernaculi tam excelsi sic inter se iam ceperint debaccari, quin contra se ipsa alia non insurgant non debet fore dubitabile menti uestre.

Pellegrini aveva già espresso la sua diagnosi di paternità dantesca in sede giornalistica, su *Alias*, supplemento culturale del *Manifesto*, in un pezzo intitolato “Lo stilista nascosto di Cangrande della scala”, nel quale la scoperta che «due passi ‘da’ Cassiodoro» contenuti nell’Epistola «sono spia dello stile dell’Alighieri» lo aveva condotto alla drastica conclusione che l’epistola è di Dante, con l’ulteriore perentoria conseguenza che «un capitolo intero della biografia dantesca avrà bisogno di una robusta riscrittura» (p. 7). È vero infatti, riconosce Pellegrini, che «nessun filologo anche mediamente preparato ignora che le *Variae* di Cassiodoro erano testo molto diffuso nel Medioevo», ma nonostante questo egli è certo che la ricerca ancora da fare «non farà altro che confermare quanto appare chiaro sin da questi primi assaggi» (Pellegrini 2018a, pp. 5 e 7).

Questo modo di procedere è stato scherzosamente definito “filologia degli annunci” («Abituati come siamo, almeno in Italia, alla politica degli annunci, non facciamoci mancare la filologia degli annunci»), e giudicato «interessante soprattutto per il metodo, perché innova l’ordine tradizionale ricerca-risultati della ricerca» (Tavoni 2018, p. 205). Le ragioni per le quali questa epistola dovrebbe essere attribuita a Dante, anticipate (è il caso di dire) da Pellegrini, sono poi state criticamente esaminate da Alberto Casadei nel suo ultimo libro, in un capitolo (2019, pp. 278-286) intitolato “Una nuova epistola di Dante? Ovvero: fare filologia al tempo delle fake news”.

Poiché tutto l’apparato per sottoporre a *authorship verification* l’*Epistola XIII*, dettagliatamente descritto nelle pagine precedenti, era già stato approntato, è bastato pochissimo per sottoporre a verifica, secondo gli identici criteri, anche l’“*Epistola XIV*”. Risulterà adesso chiaro perché, all’atto di inserire nel corpus epistolare di confronto per Ep I la raccolta di epistole cancelleresche di Pietro de’ Boattieri, abbiamo preventivamente tolto da tale raccolta (cfr. la n. 6) questa

epistola (che nell'ed. Schneider 1926 compare col n° V alle pp. 265-266): sia per non inserire fra i testi sicuramente non danteschi uno nel quale è stato proposto che ci sia la mano di Dante, sia per evitare che il testo che qui sottoponiamo al classificatore fosse usato per allenare il classificatore stesso, cosa che genererebbe un cortocircuito metodologico.

In questo caso, i valori Micro- e Macro- F_1 di riferimento (cioè la misura della "accuratezza", ovvero dell'affidabilità della classificazione) sono quelli ottenuti per Ep I (cioè 0.88-0.98: vedi tabella 4). E il responso del classificatore (prudenzialmente nominato, qui come sopra, "Ipotesi") è *NonDante*. La figura 4 riporta, con le medesime convenzioni grafiche adottate nel § 4, la probabilità (visualizzata paragrafo per paragrafo) che il classificatore assegna al fatto che il testo sia di pugno di uno dei cinque autori rappresentati nel corpus per Ep I, uno dei quali è Dante.

Il dato più rilevante è la probabilità che in questo caso il classificatore assegna alla propria "Ipotesi", illustrata nella tabella 6.

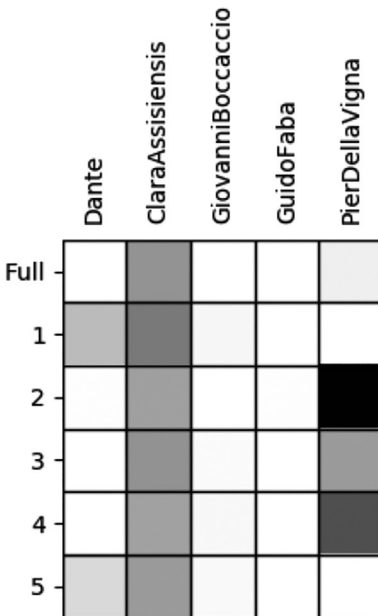


FIGURA 4. Probabilità, secondo il classificatore, che il paragrafo della "XIV Epistola" sia dell'autore indicato.

Nella prima riga della tabella 6 l'epistola, in quanto tale, è stata confrontata con le epistole di Dante e con il corpus delle epistole non dantesche, esattamente come al § 4 è stato fatto per Ep I, la parte nuncupatoria dell'*Epistola a Cangrande*. Così facendo, la probabilità che la cosiddetta *Epistola XIV* sia di Dante, secondo il classificatore, risulta dello 0.004 (quattro per mille).

Ma, poiché le ragioni addotte da Pellegrini chiamano in causa la somiglianza con un paio di passi della *Monarchia*, abbiamo provato a confrontare l'epistola anche con il corpus della trattatistica dantesca, inclusa la *Monarchia*, e non dantesca (riga 2), nonché col corpus dantesco e non dantesco risultanti dall'unione dei precedenti (riga 3). Così facendo, sia nel secondo che nel terzo confronto, la pro-

Tabella 6. Risultati di <i>authorship verification</i> per la “XIV Epistola”.		
	Ipotesi	Probabilità(Dante)
Corpus per Ep I	<i>NonDante</i>	0.004
Corpus per Ep II	<i>NonDante</i>	0.001
Corpus per Ep I + II	<i>NonDante</i>	0.001

bilità che l'epistola sia di Dante, secondo il classificatore, scende dallo 0.004 (quattro per mille) allo 0.001 (uno per mille).

Bibliografia

Fonti primarie

- Alessio, G.C. (a cura di) (1983). *Bene Florentini Candelabrum*. Padova, Antenore. [Online] Accessibile da: <https://bit.ly/2Xbl0pR> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Auzzas, G. (a cura di) (1992). *Epistole e lettere*. In: *Tutte le opere di Giovanni Boccaccio*. Milano, Mondadori. [Online] Accessibile da: <https://bit.ly/2I2VLjp> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].
- Azzetta, L. (2016). *Epistola XIII*. In: *Nuova edizione commentata delle opere di Dante*, vol. V: *Epistole. Egloge. Questio de aqua et terra*. Roma, Salerno Editrice: 271-487.
- Brownlee, K., Hollander, R. (a cura di) (2018). *Benevenuti de Rambaldis de Imola Comentum super Dantis Aldigherij Comoediam, nunc primum integre in lucem editum sumptibus Guilielmi Warren Vernon, curante Jacobo Philippo Lacaita. Florentiae, G. Barbèra, 1887*. [Online] Accessibile da: <https://bit.ly/2Dqj6du> (DDP) [Ultimo accesso: 28 maggio 2019].
- Chiamenti, M. (a cura di) (1999). *Comentum super poema Comedie Dantis (terza ed ultima redazione del 'Comentum')*, in *I commenti danteschi dei secoli XIV, XV e XVI*. Roma, LEXIS Progetti Editoriali. [Online] Accessibile da: <https://bit.ly/2ECcU2c> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].
- Cioffari, V. (1974). *Guido da Pisa's Expositiones et Glose super Comediam Dantis, or Commentary on Dante's Inferno*. Albany, N.Y., State University of New York Press. [Online] Accessibile da: <https://bit.ly/2FRhT0x> (DDP) [Ultimo accesso: 28 maggio 2019].
- Clark, E. (a cura di) (1997). *Magistri Boncompagni Ysagoge*. In: «*Quadrivium*» 8: 23-71. [Online] Accessibile da: <https://bit.ly/2MedBFc> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Colini, Baldeschi L. (1925-1926), *Le "Constitutiones Romandiolaë" di Giovanni d'Appia*. In: «*Nuovi studi medievali*», 2, 1: 221-252. [Online] Accessibile da: <https://bit.ly/30RRoAg> (ALIM) [Ultimo accesso: 28 maggio 2019].

- Cotza, V. (2013). *Giovanni del Virgilio, Allegorie super fabulas Ovidii Methamorphoseos. Edizione critica e introduzione*. Tesi di Laurea Magistrale, Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa.
- D'Angelo, E. (2014). *L'epistolario di Pier della Vigna*. Centro Europeo di Studi Normanni, Rubbettino.
- Daffinà, P., Leonardi, C., Lungarotti, M.C., Menestò, E., Petech, L. (a cura di) (1989). *Giovanni di Pian di Carpine. Storia dei Mongoli*. Spoleto, Fondazione CISAM: 227-333. [Online] Accessibile da: <https://bit.ly/2WoZaSM> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Fabbi, R. (a cura di) (1992). *De vita et moribus d. Francisci Petracchi*, in *Tutte le opere di Giovanni Boccaccio*. Milano, Mondadori, 1992. [Online] Accessibile da: <https://bit.ly/2MdBUTV> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].
- Ferrario, F. (a cura di) (1999). *Expositio seu comentum super "Comedia" Dantis Allegherii*. [Online] Accessibile da: <https://bit.ly/2Uac1mS> (DDP) [Ultimo accesso: 28 maggio 2019].
- Frugoni, A., Brugnoli G. (a cura di) (1996). *Epistole*, in *Dante Alighieri - Opere minori*. Milano, R. Ricciardi. [Online] Accessibile da: <https://bit.ly/2JIPYTp> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].
- Garbini, P. (a cura di) (1996). *Boncompagnus de Signa - Rota Veneris*. Roma, Salerno editrice. [Online] Accessibile da: <https://bit.ly/2wrCTVS> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Garbini, P. (a cura di) (1999). *Boncompagno da Signa, L'assedio di Ancona - Liber de obsidione Ancone*. Roma, Viella: 110-162. [Online] Accessibile da: <https://bit.ly/2QuMMLw> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Garufi, C.A. (a cura di) (1937). *Ryccardi de Sancto Germano notarii Chronica*, in *Rerum Italicarum Scriptores*, 7, 2. [Online] Accessibile da: <https://bit.ly/2HHrE1W> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Gaudenzi, A. (a cura di) (1971). *Guidonis Fabe Dictamina rhetorica*, in *Guido Faba - Dictamina Rhetorica Epistole*. Bologna, Forni Editore. [Online] Accessibile da: <https://bit.ly/2I36vhI> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Gaudenzi, A. (a cura di) (1971). *Guidonis Fabe Epistole*, in *Guido Faba - Dictamina Rhetorica Epistole*. Bologna, Forni Editore, [Online] Accessibile da: <https://bit.ly/2WrxgWh> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Griffin, N.E. (a cura di) (1936). *Guido De Columnis - Historia destructionis Troiae*. Cambridge, Mass., The Mediaeval Academy of America. [Online] Accessibile da: <https://bit.ly/2EF02IR> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Kristeller, P.O. (1961). *Un' "Ars dictaminis" di Giovanni del Virgilio*, in *Italia medioevale e umanistica*. 4: 179-200. [Online] Accessibile da: <https://bit.ly/2MeCD7k> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Mancuso, S. (2015). *Benvenuto da Imola, Glose Bucolicorum Virgilii (ad Buc. I, IV, VI, X). Studi critici ed edizione*. Tesi di Laurea Magistrale, Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa.

- Meloni, P. (a cura di) (1953). *Nicolai Treveti Expositio L. Annaei Senecae Agamemnonis*. Palermo, Centro di studi filologici e linguistici siciliani. [Online] Accessibile da: <https://bit.ly/2KfRxYt> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].
- Menestò, E., Brufani, S. *et al.* (a cura di) (1995). *Fontes Franciscani*, Assisi, Edizioni Porziuncola, 1995.
- Nardi B. (a cura di), *Monarchia*, in *Dante Alighieri - Opere minori*. Milano, R. Ricciardi. [Online] Accessibile da: <https://bit.ly/2EEQQLg> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].
- Monleone, G. (1941). *Jacopo da Varagine e la sua Cronaca di Genova dalle origini al MCCXCVII*. Roma, FSI: 3-414. [Online] Accessibile da: <https://bit.ly/2MhKlxs> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Romano, M.M.M. (a cura di) (2004). *Raimundus Lullus - Ars amativa boni*, in *Corpus Christianorum Continuatio Mediaevalis CLXXXIII*. Turnhout, Brepols: 120-432. [Online] Accessibile da: <https://bit.ly/2wnBu2t> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Romano, V. (a cura di) (1951). *Giovanni Boccaccio - Genealogie deorum gentilium libri*. Bari, Laterza. [Online] Accessibile da: <https://bit.ly/2ZEvcZI> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].
- Rossi, L.C. (a cura di) (1999). *Commento all'Inferno di Dante*, in *I commenti danteschi dei secoli XIV, XV e XVI*. Roma, LEXIS Progetti Editoriali. [Online] Accessibile da: <https://bit.ly/2EzazoX> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].
- Rossi, L.C. (2002). “*Benevenutus de Ymola super Valerio Maximo*”. *Ricerca sull'Expositio*, in «*Aevum - Rassegna di scienze storiche linguistiche e filologiche*». 76: 369-423.
- Schneider, F. (1926). *Untersuchungen zur italienischen Verfassungsgeschichte: Staufisches aus der Formelsammlung des Petrus de Boateriis*. In: «*Quellen und Forschungen aus italienischen Archiven und Bibliotheken*» 18: 191-273.
- Stok, F. (1991). *La “Vita di Virgilio” di Zono de’ Magnalis*. In: «*Rivista di cultura classica e medioevale*» 33, 2: 143-181.
- Sutter, C. (1894). *Aus Leben und Schriften des Magisters Boncompagno*. Freiburg i.B, Akademische Verlagsbuchhandlung von J.C.B. Mohr. [Online] Accessibile da: <https://bit.ly/2ws07eg> (ALIM) [Ultimo accesso: 28 maggio 2019].
- Tavoni, M. (a cura di) (2011). *De vulgari eloquentia*, in *Dante Alighieri – Opere*. Milano, Mondadori. [Online] Accessibile da: <https://bit.ly/2HIjBRW> (DanteSearch) [Ultimo accesso: 28 maggio 2019].
- Ussani, Jr. V. (a cura di) (1959). *L. Annaei Senecae Hercules furens et Nicolai Treveti expositio*. Roma, Edizioni dell'Ateneo. [Online] Accessibile da: <https://bit.ly/2KccRxN> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].
- Zaccagnini, G. (1924). *Le epistole in latino e in volgare di Pietro de’ Boattieri* (estratto dal vol. VIII degli *Studi e memorie per la storia dell’università di Bologna*). Imola.

- Zaccagnini, G. (1926). *La vita dei maestri e degli scolari nello Studio di Bologna nei secoli XIII e XIV. Con due Appendici e trentatré tavole illustrative*. Genève, Olschki.
- Zaccaria, V. (a cura di) (1967). *De mulieribus claris*, in *Tutte le opere di Giovanni Boccaccio*. Milano, Mondadori. [Online] Accessibile da: <https://bit.ly/211VWv1> (Biblioteca Italiana) [Ultimo accesso: 28 maggio 2019].

Fonti secondarie

- Basile, C., Lana, M. (2008). *L'attribuzione di testi con metodi quantitativi: riconoscimento di testi gramsciani*. In: *Atti del Convegno Nazionale dell'Associazione Italiana Terminologia: Terminologia, analisi testuale e documentazione nella città digitale*. Cosenza, IT: 177-195.
- Binongo, J.N. (2003). *Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution*. In: «Chance» 16, 2: 9-17.
- Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. New York, Springer.
- Canettieri, P. (2016). *Chi non ha scritto il Fiore*. In: N. Tonelli (a cura di), *Sulle tracce del Fiore*. Firenze, Le Lettere: 121-134.
- Casadei, A. (2019). *Dante: altri accertamenti e punti critici*. Milano, Franco Angeli.
- Condello, F., Garulli, V., Tomasi, F. (2016). *Montale e pseudo Montale. Autopsia del diario postumo*. Bologna, Bononia University Press.
- Cortelazzo, M.A. (2015). *Quando la filologia e la matematica si alleano*, in *Treccani Lingua Italiana*. 2 dicembre. [Online] Accessibile da: http://www.treccani.it/magazine/lingua_italiana/speciali/attribuzione/Cortelazzo.html.
- Degli Esposti, M. (2015). *Riconoscere Gramsci con la matematica*, in *Treccani Lingua Italiana*, 2 dicembre. [Online] Accessibile da: http://www.treccani.it/magazine/lingua_italiana/speciali/attribuzione/Degli_Esposti.html.
- Eder, M. (2017), *Elena Ferrante: A Virtual Author*, in *Drawing Elena Ferrante's Profile - Workshop Proceedings*. Padova, Padova University Press: 31-45.
- Ginzburg, C. (1989). *Clues: Roots of an Evidential Paradigm*. In: *Id., Clues, Myths and the Historical Method*. Baltimore, Johns Hopkins University Press: 96-125.
- Italia, P., Canettieri, P. (2013). *Un caso di attribuzionismo novecentesco: il "Diario Postumo" di Montale*. In: «Cognitive Philology» 6. [Online] Accessibile da: <https://ojs.uniroma1.it/index.php/cogphil/article/download/11586/11448>.
- Kabala, J. (2020). *Computational Authorship Attribution in Medieval Latin Corpora: The Case of the Monk of Lido (ca. 1101-08) and Gallus Anonymus (ca. 1113-17)*. In: «Language Resources and Evaluation» [Online] Accessibile in modalità "online first" da: <https://doi.org/10.1007/s10579-018-9424-0>.

- Kestemont, M., Moens, S., Deploige, J. (2015). *Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux*. In: «Digital Scholarship in the Humanities» 30, 2: 199-224.
- Koppel, M., Winter, Y. (2014). *Determining if Two Documents are Written by the Same Author*. In: «Journal of the Association for Information Science» 65, 1: 178-187.
- Magli, A. (2017). *Il mondo degli algoritmi*. In: «Puntoinformatico» 30 gennaio (intervista a Mirko Degli Esposti sui Diari postumi di Montale e ad A. Tuzzi su Elena Ferrante), [Online] Accessibile da: <https://www.punto-informatico.it/il-mondo-degli-algoritmi/>.
- Manning, C.D., Raghavan, P., Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge, Cambridge University Press.
- Mendenhall, T.C. (1887). *The Characteristic Curves of Composition*. In: «Science» 9: 237-249.
- Mosteller, F., Wallace, D. (1964). *Inference and Disputed Authorship: The Federalist*. Stanford, US, CSLI Publications [2007].
- Pellegrini, P. (2018a). *Lo stilista nascosto di Cangrande della scala*. In: «Alias – il Manifesto», 16 settembre: 5-7.
- Pellegrini, P. (2018b). *La quattordicesima epistola di Dante Alighieri. Primi appunti per una attribuzione*. In: «StEFI. Studi di Erudizione e di Filologia italiana» 7: 5-20.
- Sebastiani, F. (2002). *Machine Learning in Automated Text Categorization*. In: «ACM Computing Surveys» 34, 1: 1-47.
- Stamatatos, E. (2009). *A Survey of Modern Authorship Attribution Methods*. In: «Journal of the American Society for Information Science and Technology» 60, 3: 538-56.
- Stamatatos, E. (2016). *Authorship Verification: A Review of Recent Advances*. In: «Research in Computing Science» 123: 9-25.
- Tavoni, M. (2018). *Dante e il 'paradigma critico della contingenza'*. In: E. Brilli (a cura di), *Forum Dante and Biography*. In: «Dante Studies» 136: 201-212.
- Tetreault, J.R., Blanchard, D., Cahill, A., Chodorow, M. (2012). *Native Tongues, Lost and Found: Resources and Empirical Evaluations in Native Language Identification*. In: *Proceedings of the 24th International Conference on Computational Linguistics*. Mumbai, IN: 2585-2602.
- Tuccinardi, E. (2017). *An Application of a Profile-based Method for Authorship Verification: Investigating the Authenticity of Pliny the Younger's Letter to Trajan concerning the Christians*. In: «Digital Scholarship in the Humanities» 32, 2: 435-47.
- Tuzzi, A., Cortelazzo, M.A. (2018). *What is Elena Ferrante? A Comparative Analysis of a Secretive Bestselling Italian Writer*. In: «Digital Scholarship in the Humanities» 33, 3, September: 685-702. <https://academic.oup.com/dsh/article-abstract/33/3/685/4818094>.

Yule, G.U. (1938). *On Sentence Length as a Statistical Characteristic of Style in Prose with Application to Two Cases of Disputed Authorship*. In: «Biometrika» 30: 363-390.

ELENCO DEGLI AUTORI

Susanna Barsella

Associate Professor of Italian, Fordham University

Alberto Casadei

Ordinario di Letteratura italiana, Università di Pisa

Silvia Corbara

Perfezionanda in Data Science, Scuola Normale Superiore

Fabrizio Franceschini

Ordinario di Storia della lingua italiana, Università di Pisa

Giuseppe Indizio

Ricercatore indipendente

Alejandro Moreo Fernández

Ricercatore ISTI-CNR, Pisa

Elisa Orsi

Dottoranda in Studi italianistici, Università di Pisa

Eugenio Refini

Assistant Professor of Italian, New York University

Fabrizio Sebastiani

Direttore di ricerca ISTI-CNR, Pisa

Marco Signori

Perfezionando in Filosofia, Scuola Normale Superiore

Mirko Tavoni

Già Ordinario di Storia della lingua italiana, Università di Pisa

Piermario Vescovo

Ordinario di Discipline dello spettacolo, Università di Venezia Cà Foscari