



Come estrarre la conoscenza dai testi. Il data-mining e il topic modeling applicati a un repertorio dantesco

di Matteo Maselli
(Università di Macerata)

TITLE: *How to Extract Knowledge from Texts. Data-mining and Topic Modeling Applied to a Dantean Collection*

ABSTRACT: Il presente lavoro intende illustrare una metodologia d'indagine applicata a un repertorio digitale dantesco – il *Database Allegorico Dantesco* (DAD) – realizzato nell'ambito della mia ricerca dottorale e contenente centinaia di testi inerenti all'allegorismo della *Divina Commedia*. Rivolgendo alle opere archiviate specifiche teorie informatiche e ricorrendo all'uso congiunto di più strumentazioni digitali verrà descritta una procedura di studio che, facendo riferimento a principi statistici di scienze applicate, permetterà di giungere a dei risultati oggettivi in rapporto alle discussioni teoriche maturate intorno all'allegoria dantesca nei secoli XX-XXI. Nello specifico verrà mostrato come sia possibile sottoporre il *corpus* di testi scelti, poiché già trascritti in XML per l'archiviazione nel DAD, a una modellizzazione tematica (*topic modeling*) con la quale individuare gruppi di semantemi riferibili ad autori, motivi e temi ricorrenti nelle letture allegoriche del poema da utilizzare poi come linee guida per un esame diretto e approfondito (*close reading*) dell'oggetto della ricerca. In questo modo i *tool* digitali verranno adoperati attivamente per l'estrazione automatizzata di differenti tipologie di dati (*data-mining*) corrispondenti a nuclei tematici da sviluppare per tracciare un profilo critico dell'allegorismo della *Commedia*.

ABSTRACT: This paper intends to illustrate an academic methodology applied to a digital Dante repertory – the *Database Allegorico Dantesco* (DAD) – developed within



the framework of my doctoral research and containing hundreds of texts related to the allegorism of the *Divine Comedy*. By addressing specific computer theories to the archived works and by resorting to the joint use of several digital instruments, a study procedure will be described that, by referring to statistical principles of applied sciences, will allow for objective results in relation to the theoretical discussions developed around Dante's allegory in the 20th-21st centuries. Specifically, it will be shown how it is possible to subject the corpus of selected texts, since they have already been transcribed in XML for archiving in the DAD, to topic modelling in order to identify groups of semantemes referable to authors, motifs and recurring themes in the allegorical readings of the poem to be used as guidelines for a direct and in-depth examination (close reading) of the object of the research. In this way the digital tools will be actively used for the automated extraction of different types of data (data mining) corresponding to thematic nuclei to be developed in order to trace a critical profile of the allegorism of the *Comedy*.

PAROLE CHIAVE: Dante Alighieri; allegoria; *data-mining*; *topic modelling*; *information retrieval*

KEY WORDS: Dante Alighieri; allegory; *data-mining*; *topic modelling*; *information retrieval*

INTRODUZIONE

Nonostante le dinamiche allegoriche alla base della *Commedia* siano state indagate con una certa continuità, allo stato attuale degli studi più di una questione rimane aperta. Le difficoltà a ottenere risultati unanimi non sono da imputare alla sola e ovvia complessità della materia trattata ma riguardano anche le metodologie di studio adottate. Ritenendo urgente intervenire soprattutto in riferimento a quest'ultime si propone dunque di riflettere, illustrandoli, su alcuni recenti metodi d'indagine sorti in ambito informatico e ancora in parte estranei ai *Dante Studies*.¹

In ciò che segue si mostrerà come la cooperazione interdisciplinare tra una tradizionale esegesi testuale e una più pionieristica ermeneutica digitale possa aprire nuovi percorsi d'indagine per l'esame dell'allegorismo dantesco con i quali poter confermare o confutare ipotesi più volte discusse in sede accademica.

L'obiettivo che ci si pone è quello di produrre uno studio criticamente ragionato sull'allegorismo della *Commedia* applicando a una selezione di testi archiviati nel *Database Allegorico Dantesco* (DAD)² specifiche teorie informatiche e ricorrendo all'uso

¹ Di recente si sono tuttavia susseguite una serie di pubblicazioni che riflettono su approcci in parte prossimi a quelli qui discussi. Cfr. Bonera-Bardazzi; Marchesi-Patton-Qiu-Matukhin 298-301; Van Peteghem 229-252.

² Il *Database Allegorico Dantesco* è un *repository* di testi accademici sull'allegorismo della *Divina Commedia* realizzato come componente digitale del progetto dottorale del sottoscritto. Per un quadro



congiunto di più strumentazioni digitali. Poiché si farà riferimento a principi statistici di scienze applicate si ritiene possibile giungere a dei risultati oggettivi in rapporto alle discussioni teoriche maturate sull'allegoria dantesca nei secoli XX-XXI.³

Nello specifico un *corpus* di testi scelti,⁴ poiché già trascritti in XML (*eXtensible Markup Language*) per l'archiviazione nel DAD, verrà sottoposto a una modellizzazione tematica (*topic modeling*) con la quale individuare gruppi di semantemi riferibili ad autori, motivi e temi ricorrenti nelle letture allegoriche del poema da utilizzare poi come linee guida per un successivo esame diretto e approfondito dell'oggetto della ricerca (*close reading*).⁵ Gli strumenti digitali non saranno dunque limitati alla sola mediazione del rapporto con i testi raccolti, ma verrà favorito un loro impiego attivo per l'estrazione automatizzata di differenti tipologie di dati (*data-mining*) corrispondenti a nuclei tematici da sviluppare per tracciare un profilo critico dell'allegorismo della *Commedia*.⁶

IL PROCESSO ESTRATTIVO DELLA CONOSCENZA

Per le operazioni di analisi digitale si sono utilizzati *tool* quali *Sketch Engine*, *jsLDA*, *Voyan Tools*. Soprattutto quest'ultimo è risultato adatto alle prove computazionali di seguito descritte in quanto si propone come strumento *open source* che riesce adeguatamente a bilanciare l'intuitività dell'impiego con una varia casistica di valutazione e visualizzazione dei dati, con particolare attenzione a parametri centrali per un esame della ricorrenza semantica di concetti in un *corpus* controllato quali la frequenza e la distribuzione dei termini.

L'avvio dell'*information retrieval* sul *corpus* dell'area Sezione monografica del DAD ne ha comportato una prima scomposizione in unità costituenti (circa 2 milioni e

informativo sul DAD cfr. Maselli, *Database Allegorico Dantesco* 165-170; Maselli, *Struttura e funzioni* 241-260.

³ Si è scelto questo *range* temporale poiché è quello in cui si concentra un alto numero di testi che, oltre a essere autorevoli, sono anche ricorsivi (citazioni dirette e/o indirette tra gli stessi) permettendo così l'insorgenza di un *networking* sull'allegoria in Dante. Si tratta, cioè, di un orizzonte di dialogo tra intellettuali che non ha eguali nella storia della dantistica in termini di nomi coinvolti e qualità dei risultati ottenuti.

⁴ Le cifre qui elencate riguardano delle prove di *data-mining* eseguite nei primi mesi del 2023 per la predisposizione del presente prospetto che hanno coinvolto un *corpus* di 361 testi (225 per il XX secolo; 136 per il XXI) inseriti nella Sezione monografica del DAD.

⁵ Il *close reading* non è estraneo alla dantistica, come indicato da Ascoli in merito agli studi nord-americani sull'*Epistola XIII*: "[...] il *close reading* all'americana, adoperato con cautela, può essere a volte il modo più diretto per arrivare a una lettura 'storicizzata' [dell'*Epistola* a Cangrande]" (Ascoli 156). Lo stesso Ascoli propone di "affrontare una lettura dettagliata – un *close reading* – di *Convivio II, i*" (157).

⁶ È comunque necessario poi uno studio tradizionale poiché, come ha giustamente intuito Bellomo discutendo delle possibilità offerte dalla rete nell'analisi di *Inf. V*, "[t]oday, electronic concordances allow us to gather, with feeble effort (perhaps too feeble), all or almost all the passages of Latin literature involving the crane, but they do not exempt the modern commentator from having to analyze and make a hierarchy of the search results" (Bellomo 97). La macchina, cioè, mette nella condizione di avere una completezza di dati, ma spetta poi al critico discernere i livelli di priorità degli stessi in relazione agli obiettivi prefissati con la sua ricerca.



200mila parole) delle quali si sono considerate solamente le forme di parole uniche (circa 140mila). Di quest'ultime sono poi state privilegiate quelle che ricorrono almeno due volte (circa 70mila), condizione necessaria della ricorsività tematica. Tra queste è stato necessario selezionare i soli termini ritenuti significativi (poche centinaia), ovvero quegli *item* che ben si radicano con alte percentuali di manifestazione rispetto alla totalità dei testi considerati e che dunque risultano particolarmente indicativi per il discorso allegorico. In base al numero delle loro apparizioni nel *corpus* – si tratta di *token* di parole – è stata istituita una gerarchia delle occorrenze come prima attestazione dei nodi tematici centrali da tenere in considerazione per lo studio dell'allegoria in Dante. Per la configurazione di tale tracciamento non si è tenuto conto della sola attestazione diretta dei termini, poiché ciò avrebbe comportato una catalogazione incompleta e fuorviante. Si è invece prestata attenzione anche al rilevamento di possibili derivativi attraverso un diretto esame del contesto semantico di maggiore prossimità alle voci individuate e a un'analisi dei rapporti statisticamente più frequenti che queste hanno con altre entità del *corpus* (*network analysis* che rileva reti di relazioni tra nodi concettuali legati da *ties* o *link*). Per chiare le procedure di cui si è detto, si propone di seguito un esempio operativo riferito all'oggetto 'Macrobio'.

ESEMPIO OPERATIVO: IL CASO DI MACROBIO

Ricercando nel *corpus* stabilito i termini di maggiore attinenza con MACROBIO, che è evidentemente significativo poiché ritorna con una certa costanza nei testi archiviati, si ottiene il seguente schema:

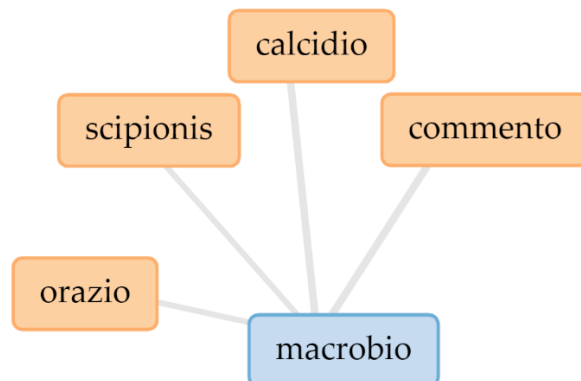


Fig. 1. Termini correlati alla voce MACROBIO

I rapporti tra ORAZIO-MACROBIO e CALCIDIO-MACROBIO sono poco rilevanti ai fini del rinvenimento di termini aggiuntivi per l'indicizzazione finale di MACROBIO poiché Orazio e Calcidio sono entità esplicite non soggette a più interpretazioni.



SCIPIONIS e COMMENTO, essendo di fatto generici, possono invece assumere più di un significato in base al rapporto con altre voci. Ricercando i termini più frequenti (*fetch collocates*) di riferimento a SCIPIONIS si ricava quanto segue:

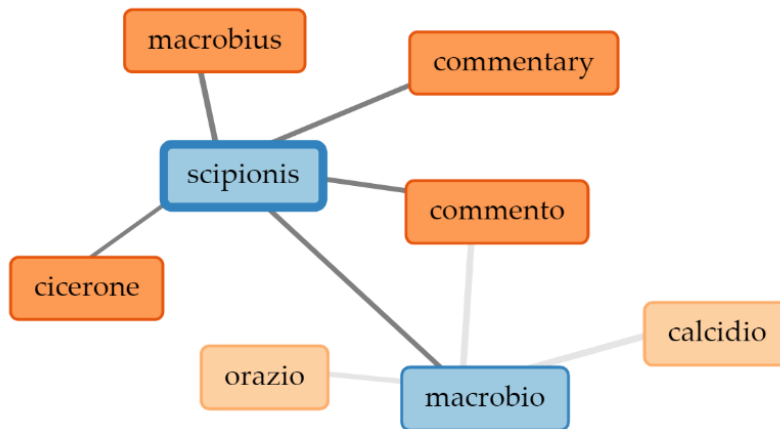


Fig. 2. Estensione dei termini correlati alla voce MACROBIO

SCIPIONIS è in rapporto con CICERONE, MACROBIUS e COMMENTARY. Già a questo livello di lettura è possibile un'esemplificazione utile a snellire le fasi successive dello studio. Si può infatti escludere la voce MACROBIUS poiché si è già tenuto in considerazione l'unità di partenza MACROBIO. Inoltre, poiché COMMENTARY è una variante linguistica di COMMENTO – SCIPIONIS è infatti in rapporto diretto con entrambe – la si può anche non considerare. Rimangono dunque gli *item* CICERONE, SCIPIONIS e COMMENTO. A questo punto, all'operatività della macchia bisogna affiancare l'intuizione dell'operatore. Bisogna, ovvero, ricavare un oggetto finale e ricapitolativo ottenuto mettendo in relazione i termini superstiti e che sia al contempo in grado di soddisfare appieno il rapporto tra gli stessi. Ci si deve, cioè, chiedere quale sia il prodotto dell'attività di commentatore (COMMENTO) di Macrobio che riguarda CICERONE e che implica un riferimento all'aggettivazione SCIPIONIS. L'ovvia risposta è i *Commentarii in Somnium Scipionis*, lavoro che rese Macrobio una vera e propria autorità scientifica e filosofica in tutto il Medioevo (Curtius 622). Ciò implica la non autoreferenzialità del peso di MACROBIO all'interno del *corpus*, poiché è dipendente anche dalle occorrenze dei *Commentarii*.

L'indagine quantitativa impostata sulla correlazione strumento digitale-testualità ha così restituito una costante nel discorso sull'allegoria in Dante – i *Commentarii* – sulla quale bisognerà interrogarsi. Al supporto della strumentazione tecnologica subentra quindi un esame diretto dei testi, l'unico in grado di dare una 'voce' effettiva al riferimento statistico e di contestualizzarlo storicamente e criticamente nel quadro dell'allegoria dantesca.



Il *Commento al Sogno di Scipione* di Macrobio contribuì alla riflessione sull'uso dell'allegoria da parte dei poeti e più nello specifico al concetto di *integumentum*, trasmettendolo poi a tutto il Medioevo. A tal proposito, un confronto automatizzato con il materiale catalogato nel DAD può essere molto indicativo delle trame semantiche tra temi/autori in cui è coinvolto l'oggetto MACROBIO.

Esplicitando ulteriormente i legami che si dipanano dalla voce COMMENTO, chiaro riferimento a quello scritto per il *Sogno di Scipione*, si evince come questa sia uno snodo che rapporta Macrobio a Guglielmo di Conches, i cui studi sull'*integumentum* furono non casualmente influenzati proprio dai *Commentarii* (non per nulla, INTEGUMENTUM è un *item* di prossimità a Guglielmo di Conches):

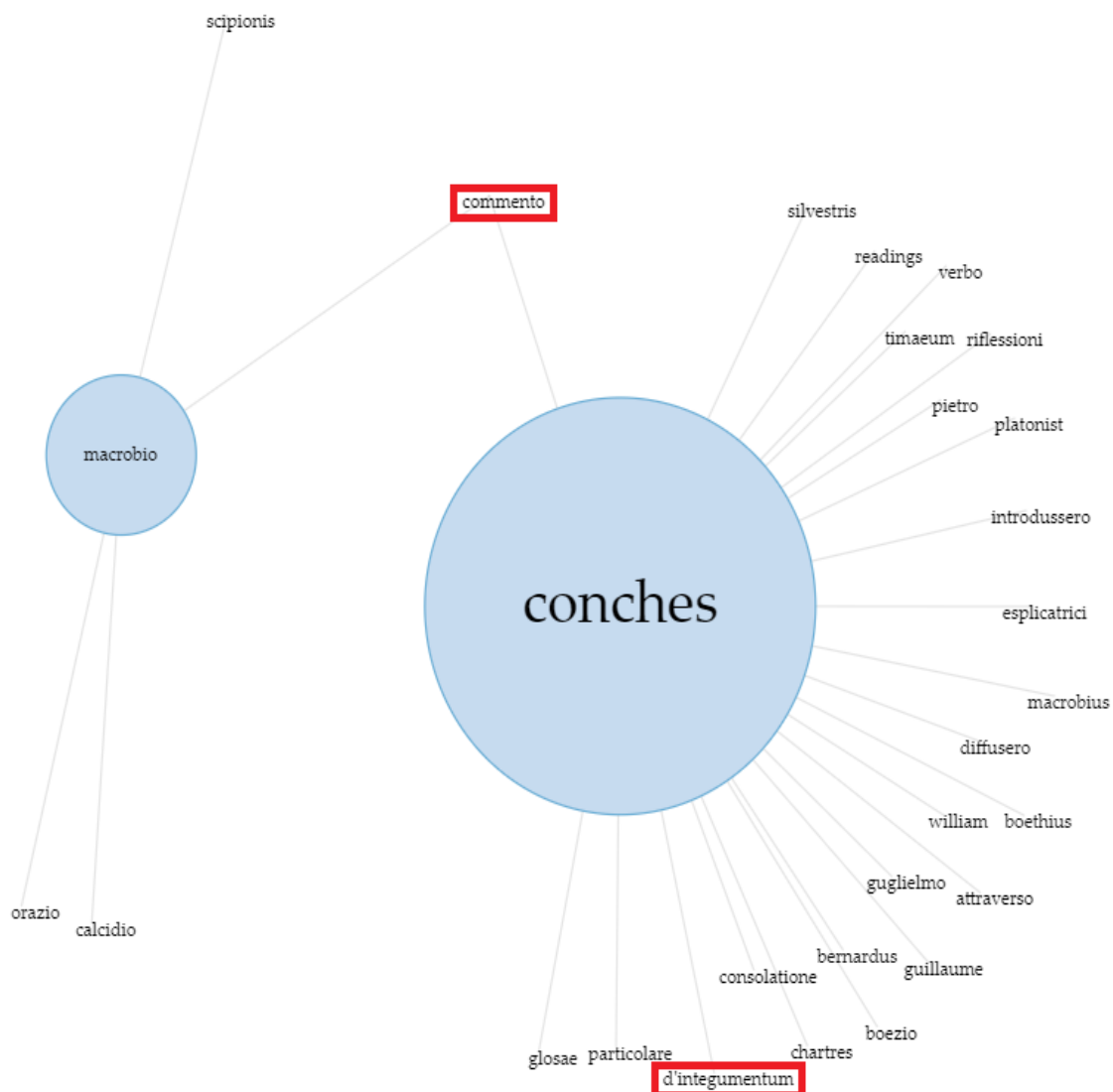


Fig. 3. Termini correlati tra la voce MACROBIO e [Guglielmo di] CONCHES



Per dimostrare la non superficialità dei processi deduttivi adottati a partire dai dati elaborati dalla macchina, si prende ora in considerazione il recente commento di Fioravanti e Giunta (Alighieri, *Convivo* ed. Fioravanti-Giunta) al passo del *Convivio* (II, i, 1-2) in cui Dante discute degli aspetti dai quali sono scaturiti gli schematismi succitati. A scanso d'equivoci, poiché avrebbe potuto falsare *a priori* il procedimento qui descritto, si specifica che il commento del *Convivio* menzionato non è stato usato come parte del *corpus* elaborato per via automatica poiché non lo si è ritenuto, da un punto di vista tipologico, pienamente affine ai criteri di selezione adoperati per la configurazione della Sezione monografica del DAD, che è dichiaratamente dedicata alla *Commedia*.

In merito al condizionamento esercitato da Macrobio sulla tradizione medievale dell'allegorismo poetico, Fioravanti e Giunta ritengono che:

Il testo di Macrobio aveva trasmesso questa teoria [la distinzione tra *fabula* e *narratio fabulosa* come base epistemologica dell'applicazione dell'allegoria poetica] al Medioevo, che ne avrebbe fatto uso soprattutto nel XII secolo con Bernardo Silvestre, Guglielmo di Conches e Alano di Lilla (Alighieri, *Convivo* ed. Fioravanti-Giunta 125).

Già nei contesti semantici prossimi a Guglielmo, di seguito maggiormente estesi rispetto alla figura 3, vi è una prima conferma di quest'osservazione in quanto vi è un significativo tasso percentuale di rapporto con Bernardo Silvestre come indicato dalla doppia voce BERNARDUS e SILVESTRIS (figura 4). Se poi si integrano anche le voci di Alano di Lilla e dello stesso Bernardo si otterrà un prospetto che conferma appieno come il *Commento* di Macrobio e la nozione di *integumentum* siano degli anelli di congiunzione tra Bernardo Silvestre e Guglielmo di Conches⁷ – è ovvero il *Commento* del latino in relazione al concetto di allegoria poetica ad aver influenzato i due –, e come quest'ultimo, mediato da Bernardo,⁸ sia in relazione diretta con Alano di Lilla (figura 4).

⁷ Stessa osservazione vale ovviamente anche per la dicitura "Dante", che riguarda però più direttamente la prospettiva della critica che discute delle derivazioni degli autori menzionati nell'Alighieri.

⁸ "Bernardo Silvestre, commentatore del XII secolo di cui Dante sembra conoscere alcune delle allegorie psicologiche di Virgilio – direttamente o attraverso un intermediario" (Dronke, *Dante e le tradizioni* 296).



avanzare un giudizio critico qualitativamente prossimo alle valutazioni di Fioravanti e Giunta.

A saper ben leggere le relazioni suggerite da particolarità riscontrabili nel grafico¹¹ è possibile esplicitare anche percorsi di lettura secondari ma comunque utili a ricostruire il pensiero di Dante sull'allegoria. Nelle diramazioni che si irradiano da Guglielmo di Conches, a prescindere dal rapporto con il centro, si notano alcuni elementi che pur nella loro univocità nominale possono rapportarsi biunivocamente. È il caso di BOEZIO (addirittura latinizzato in BOETHIUS) e CONSOLATIONE (segnalati in azzurro nella figura 5) che stanno ovviamente a indicare "quello non conosciuto da molti libro di Boezio, nel quale, [Dante] consolato s'avea" (*Conv.* II, xii, 2), cioè il noto prosimetro della *De consolatione philosophiae*. Poiché la relazione autore-opera è visivamente rappresentata come proprietà distintiva di [Guglielmo di] CONCHES, ciò significa che la modellizzazione computerizzata sta suggerendo, privilegiando Boezio rispetto ad altre voci del grafico, di approfondire la figura di Guglielmo con riferimento esplicito alla *Consolatio*.

Procedendo in tal senso – si intende con ciò uno studio tradizionale della figura del filosofo francese, che essendo mirato al solo lavoro boeziano è dunque gestibile – emergerà come Guglielmo di Conches fosse convinto assertore delle possibilità che testi di finzione potessero contribuire alla diffusione, se opportunamente intesi, di precetti d'ordine morale. Guglielmo riteneva che ciò fosse possibile poiché le favole a cui ricorrevano i poeti altro non erano che coperture di verità soggiacenti che la sola interpretazione allegorica avrebbe potuto liberare e porre a riferimento di una sana condotta etica. Le menzogne introdotte a livello della lettera dai poeti erano dunque al servizio di lodevoli verità morali e Guglielmo espresse questo suo pensiero proprio nel commento scritto alla *Consolatione* di Boezio. La spia testuale boeziana, che ha incanalato l'indagine in un percorso specifico, trova così una sua giustificazione, non prima di aver arricchito l'ambito di ricerca sull'allegorismo dantesco con indicazioni esegetiche parallele ma pertinenti, nella natura di apologeta dell'allegoria *in verbis* di Guglielmo di Conches.

Nella riorganizzazione dei dati si segnala inoltre anche un'altra costante, che un occhio attento riesce subito a rilevare e che opera nuovamente un congiungimento tra *item*: GIOVANNI e GARLANDIA (in verde nella figura 5). Ancora una volta, dunque, la manifestazione computerizzata di unità minime di significato, processate come componenti di un quadro più esteso quale l'allegorismo della *Commedia*, ci avverte che,

¹¹ Non sono peregrini gli approcci che mettono in relazione parola e immagine, come l'"esegesi visiva", "un vero e proprio ambito di studi [...] impennato sulla centralità assegnata a figure e diagrammi in ordine alla comprensione di esegesi medievale della Scrittura" (Smalley 22). Così Mira Mocan definisce l'"esegesi visiva" del XII secolo praticata dai Vittorini: "Si costituisce [...] da una parte come un 'commento visivo' portatore di uno spessore ermeneutico proprio, dall'altra come una dinamica psicologica per *introiettare* il testo interpretato, attraverso la sua visualizzazione" (Mocan 79). Insomma, l'"esegesi visiva" riproduce quello stesso effetto che, all'infuori di motivi escatologici, si auspica possa essere conseguito con una rappresentazione ricapitolativa degli esiti del *data-mining* in quanto "[r]accogliendo e sintetizzando in un'immagine unica la pluralità e la stratificazione di significati propri del testo sacro, il diagramma diventa [...] 'un saggio di evocazione visuale, un tentativo di sfuggire al tempo e di inglobare lo spazio in un solo sguardo'" (86).



con ogni probabilità, anche Giovanni di Garlandia si pone nella stessa tradizione allegorica a cui appartiene Guglielmo di Conches; e infatti, da un raffronto con quanto Pompeo Giannantonio scrive in *Dante e l'allegorismo* (1969), testo che al pari del commento di Fioravanti e Giunta non è stato tenuto in considerazione nella computazione meccanica per evitare di falsarla in automatico,¹² si evince che

per spiegare questo genere di interpretazione allegorica [quella poetica], [Guglielmo di Conches] si serve del vocabolo "integumentum" equivalente al ciceroniano "involucrum" e al classico termine "allegoria", come dirà Giovanni di Garlandia: "Integumenta, id est allegoricas sententias super fabulas" (Giannantonio 169).

Anche in questo caso, la figurazione dei legami semantici formulati meccanicamente in corrispondenza di CONCHES non solo ha confermato l'attinenza del quadro tracciato da Giannantonio (lo fa persino nella terminologia usata, perfettamente presente nella figura 5), ma ha persino esplicitato l'esistenza di una relazione tra Guglielmo e Giovanni di Garlandia che nella citazione riportata non è istituita poiché Giannantonio si serve del passo degli *Integumenta Ovidii* come sola informazione ricapitolativa di quanto detto in precedenza.

LIMITI E POTENZIALITÀ DEL METODO

Se l'esercizio logico fin qui mostrato permette di verificare, attraverso l'interazione tra differenti *range* informativi, ipotesi altrimenti accolte senza controprova critica, e congiuntamente consente di definire dei nuovi nuclei tematici da sviluppare poi con uno studio impostato su principi tradizionali della ricerca, è altrettanto vero che si tratta di un processo per il quale potrebbero segnalarsi dei limiti a esso intrinseci. Le fasi di studio descritte a partire dell'esempio su MACROBIO impongono infatti a monte la conoscenza del *Commento* scritto per il *Sogno di Scipione*, pena l'inattuabilità della correlazione tra i termini reduci dall'esemplificazione compiuta a partire dalla figura 2. Per quanto auspicabile, è però anacronistico pensare di poter esplicitare autonomamente tutte le incognite semantiche alla base dei legami tra le varie unità di significato. Se questo può accadere con opere, teorie, autori particolarmente noti – ma anche in questo caso possono sfuggire gli aspetti non primari –, figure secondarie o non prioritarie nella formazione di chi legge i grafici sono escluse dagli automatismi come quelli visti in merito al caso di Macrobio. Ciò significa che la proposta di procedere a una modellizzazione computazionale sul tema dell'allegoria per il tramite di algoritmi impostati e operanti meccanicamente non ha motivo di essere perseguita? Si può, cioè, avviare a quello che sembra un limite che minaccia la fattibilità di questo tipo d'indagine?

Assodata l'irrinunciabile sollecitazione di intuizione e spirito d'osservazione, utile ai fini del superamento di questa *impasse* sarà un'operatività congiunta tra più risorse

¹² Il testo di Giannantonio non fa parte del *corpus* esaminato per via informatica poiché con esso l'autore propone un quadro riassuntivo della formazione dell'allegoria nel Medioevo e non specificamente in Dante.



digitali. Per esplicitare meglio questa necessità, si considerino i legami tematici in relazione ad ALANO [di Lilla], il *Doctor universalis* del XII secolo:

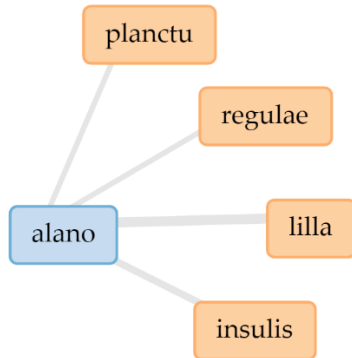


Fig. 6. Termini correlati alla voce ALANO

La coppia ALANO-LILLA (la più frequente nel *corpus* come indicato da un maggiore spessore della linea di congiunzione tra lemmi) non richiede spiegazioni aggiuntive poiché riferimento esplicito al luogo di nascita del teologo francese. Lo stesso dicasi per INSULIS – Alano è noto anche come Alano delle Isole, latinizzato in Alanus ad Insulis – che però è meno immediato della controparte italianizzata. Tuttavia, una veloce e neanche approfondita ricerca è più che sufficiente a chiarire questi aspetti. Per giustificare la presenza e il significato di PLANCTU e REGULAE si necessita invece di ulteriori passaggi.¹³

Esplicitando i termini più frequenti di prossimità a PLANCTU si nota quanto segue:

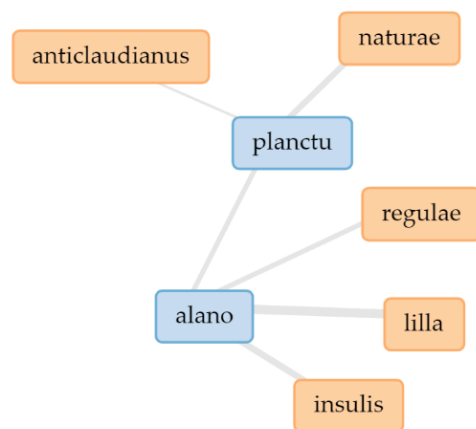


Fig. 7. Estensione dei termini correlati alla voce ALANO

¹³ Quanto di seguito detto deve intendersi come dimostrazione pratica dei procedimenti e delle potenzialità del modello qui considerato poiché è ovvio che anche nel caso di PLANCTU e di REGULAE basta un veloce raffronto biografico per risalire alle corrispettive opere di Alano.



L'istituzione di un successivo albero di parole esteso al contesto d'uso di PLANCTU nel *corpus* della Sezione monografica del DAD evidenzia come tra NATURAE e ANTICLAUDIANTUS solo il primo ricorre come termine associato (*collocates*):

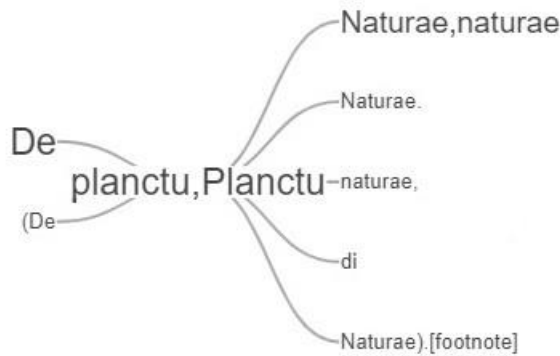


Fig. 8. Fetch collocates della voce PLANCTU

È dunque evidente che, senza procedere in ulteriori approfondimenti, PLANCTU NATURAE abbia un peso notevole in correlazione al nome ALANO, e infatti si tratta di un'opera di matrice allegorica e di stampo cosmologico che il monaco cistercense scrisse intorno alla metà del XII secolo.¹⁴ Poco conta che a quest'ultima conclusione ci si arrivi eventualmente per vie esterne al grafico – basta un semplice esame in un qualsiasi motore di ricerca online – poiché l'obiettivo che questo si pone è quello di mettere a disposizione dei dati grezzi (*raw data*) su cui riflettere, delle *keywords* mostrate senza porsi il problema se l'utente che dovrà leggerle ha la conoscenza per riconoscerne subito il significato o se invece necessita di passaggi aggiuntivi, ma comunque facili da espletare, per sopraggiungere a una completezza informativa.

Ragionamento simile vale per REGULAE, per il quale all'interno del *corpus* viene segnalato un unico significativo riferimento contiguo (ma più volte ripetuto):



Fig. 9. Fetch collocates della voce REGULAE

Alano fu, infatti, autore delle *Regulae theologicae*, un significativo trattato sulla fede cristiana. Un uso parallelo del DAD – sempre nell'ottica dell'impiego congiunto di

¹⁴ L'opera è un prosimetro fortemente influenzato, tra le altre cose, anche dalla *Cosmographia* di Bernardo Silvestre, motivo in più di contatto tra quest'ultimo e Alano come indicato nella figura 5. Sempre in merito ai contatti Alano-Bernardo, Dronke considera l'*Anticlaudianus* una continuazione, sull'impiego dell'*integumentum*, della *Cosmographia* (Dronke, *Il secolo XII* 291).



più *tool* – è una valida modalità per decodificare il significato dei rapporti tra termini e azzerare quasi del tutto le minacce di inattuabilità della ricerca prima menzionate. Riprendendo in considerazione l'esempio precedente, il grafico segnala che REGULAE è presente 5 volte nel *corpus* della Sezione monografica.¹⁵ Ricercandolo all'interno di quest'ultima vengono infatti restituiti 5 risultati, tutti riferiti a THEOLOGICAE e con sottolineatura dell'attività scrittoria di Alano:

e tanto più importante perché impiegato per la prima volta nel «figurare», «secondo la infallibile veritate», Beatrice come simbolo numerico - (sulla scorta, come si vedrà, delle Theologicae Regulae di Alano da Lilla).

A riprova varrà per la Vita Nuova, l'impiego delle «Regulae theologicae» di Alano per elevare Beatrice da personaggio storico a simbolo. Questo passo di Alano riflette ovviamente le convinzioni dell'autore espresse in altro lavoro, in quelle Theologicae Regulae, ove il Doctor Universalis apertamente condanna quel termine di «involucrum» discusso sinonimo di «integumentum» che i Carnotensi, a partire da Guglielmo di Conches, introdussero e diffusero attraverso i loro Commenti soprattutto a Virgilio (Giovanni di Salisbury dirà che Virgilio nell'Eneide ha cantato «sub involucro fictitii

E quando in questo processo di ricupero dei miti, «Platone» e «Mosè» concorderanno, così come Mercurio e la Filologia, basterà far scattare ed applicare la formula paolina del figurismo perché gli Dei e gli eroi pagani siano trasformati in «figurae Christi», e quello stesso Alano, così severo nelle Theologicae Regulae, potrà anticipando Dante, nell'Anticlaudianus chiamare Dio «supremus Jupiter» e gli angeli «proceres Tonantis», e rappresentare il transito della creatura al Creatore, del naturale al soprannaturale come un transito da Apollo a Giove, dalla terra all'olimpico:

Come numero figurato Beatrice era quindi assunta a *typus Trinitatis gerens*, e nel far questo il Poeta ricorre «sottilmente» alle Theologicae Regulae di Alano, come già abbiamo affermato, e precisamente alla «cinquantesima»:

Fig. 10. REGULAE nel *corpus* della Sezione monografica.

La consultazione del DAD ha pertanto contestualizzato il riferimento a REGULAE come rimando a un lavoro di Alano di Lilla.

Per ANTICLAUDIANUS, invece, così come per qualunque altro lemma del quale non se ne coglie il valore a livello delle prime corrispondenze associative, può insistersi ulteriormente nell'esame delle voci correlate. Operando in questo modo c'è però il rischio di dover gestire termini eccessivamente generici che non aggiungono informazioni dirimenti per una contestualizzazione specifica della parola di partenza (si tratta di *function words* – preposizioni, articoli, congiunzioni, etc. – che di solito ricorrono con maggiore frequenza rispetto a parole con significato proprio). Un arricchimento tematico poco funzionale alla comprensione della voce analizzata implica che già con la forma ANTICLAUDIANUS si è raggiunta una quota di riferimento del valore della parola tale da permetterne un riconoscimento del peso rispetto ad ALANO. Non saranno per questo necessari approfondimenti ulteriori dei contatti con altri termini poiché si finirebbe per diradarne inutilmente il valore già precedentemente dato. Sempre adoperando le funzioni di ricerca integrate nel DAD, evitando dunque di rivolgersi necessariamente altrove per fugare eventuali dubbi, si evince la plausibile ascendenza che l'*Anticlaudianus* ebbe su Dante, non solo come fonte tematica ma anche come modello narratologico.

Tra i vari autori archiviati che lo evidenziano si segnalano solamente pochi casi scelti poiché indicativi di ulteriori sotto-tipologie di indagini che un uso ragionato del DAD può consentire all'utente che ne fa uso. A titolo d'esempio possono stabilirsi con estrema facilità impliciti rapporti derivativi tra testi – da non confondersi con i legami ipertestuali espliciti che la connettività ipermediale tra schede archiviate permetterà di sfruttare. Come caso significativo, si consideri la vicinanza teorica tra questo passo di Cerroni (13):

¹⁵ Tutti i dati riportati in questo testo sono aggiornati al mese di luglio 2023.



L'efficacia comunicativa e persuasiva dell'allegoria si misura, nel medioevo, con l'eredità pagana e con due distinte condizioni culturali: la conoscenza del mondo secondo il modello cristiano e l'attivazione di un nuovo universo mitico. La tradizione allegorica fiorita tra IV e XIV secolo, infatti, è disseminata di opere esegetiche che leggono il mondo attraverso i testi sacri e i testi pagani e di opere letterarie che propongono nuovi miti o ne riattivano di tramontati. L'allegoria interpretativa, o allegoresi, e l'allegoria creativa sono dunque i prodotti di una stessa storia e di uno stesso processo culturale. La Commedia fiorì sull'innesto di questi due rami principali, secondo l'esempio della grande tradizione dei poemi allegorico-scientifici mediolatini: la *Cosmographia* di Bernardo Silvestre, il *De Planctu Naturae* e l'*Anticlaudianus* di Alano di Lilla furono senza dubbio tra le fonti primarie dell'allegoria dantesca, particolarmente in virtù dell'ideazione di un immaginario al tempo stesso persuasivo e argomentativo, produttore di miti e interprete veridico del mondo.

Fig. 11. Cerroni.

e il seguente di Mazzali (130):

Il mondo antico - ha osservato Ernst Robert Curtius nel grande libro *Europäische Literatur und lateinisches Mittelalter* cit., - si realizza attivamente nella cultura romanica attraverso le codificazioni, gli apparati, i repertori, i lessici e le epitomi propri della cultura latina medioevale; e lo stesso Curtius avanza con fortissime probabilità l'ipotesi che Dante conoscesse, fra i testi latini medioevali, l'*Anticlaudianus* di Alain de Lille (e cfr. ora i riferimenti al testo di Alano suggeriti da Raimondi in *Metafora e storia* cit., pp. 50, 52, 59, 72, 99, 112, 115, 118, 140).

Fig. 12. Mazzali.

Così come non appare erroneo mettere quest'ultimo in rapporto diretto con alcune osservazioni, di due anni prima, di Padoan (31-32), quantomeno a fronte della stessa indicazione bibliografica avanzata a supporto della tesi espressa:

si vuole ora, contro l'interpretazione che fu cara al Foscolo, mettere soprattutto in luce per la Comedia il carattere fondamentale di invenzione poetica, di «romanzo teologico», di «itinerarium» filosofico, sottolineandone quasi esclusivamente i valori letterari o teologici; ed anzi si tende a sviluppare un accenno (utile se mantenuto nei suoi limiti) di Ernst Robert Curtius, nel volume, estremamente ricco e prezioso, *Europäische Literatur und lateinisches Mittelalter*, che mira ad inscrivere la Comedia, almeno in parte, nell'ambito culturale di quella letteratura filosofeggiante ed allegoristica francese che produsse l'*Anticlaudianus* di Alano da Lilla: soprattutto per qualche affinità (non esclusiva però di altre fonti) tra l'itinerario dantesco a Dio e il viaggio di Fronesis attraverso le sfere celesti fino alla sfera della Trinità trascendente, con la guida della Ragione, poi della Teologia ed infine della Fede (secondo la consueta triplice distinzione dei gradi della conoscenza umana). Alano avverte i suoi lettori che nell'*Anticlaudianus* il senso letterale è il meno importante e che la verità è celata nell'allegoria (avvertimento del resto, in quel caso, quasi ovvio); ed ecco dunque che anche il senso letterale della Comedia è persa a non pochi studiosi mera fictio in funzione del vero che è nell'allegoria.

Fig. 13. Padoan.

L'inoltro della ricerca nel *corpus* della Sezione monografica del DAD impostata su 'Anticlaudianus' ha, cioè, determinato una selezione automatica di testi, e al loro interno di passi, tra i quali si è potuto procedere con facilità a un raffronto incrociato per portare alla luce una linea di pensiero comune che si rafforza nel lungo periodo (anni '70-anni 2000), tanto significativa perché coinvolge voci diverse per obiettivi di ricerca, destinatari dello studio ed epoca storica di riferimento. È chiaro che si tratta di una sottotrama presente a prescindere dallo strumento tecnologico usato per renderla manifesta. Eppure, senza quest'ultimo sarebbe stato necessario possedere fisicamente i testi elaborati in automatico dal DAD, leggerli nella loro interezza, percepire una certa affinità teorica tra alcuni passaggi e procedere manualmente a un loro confronto. È evidente che vi siano importanti limiti che ostacolano la buona riuscita di una simile procedura, per la quale il venir meno di uno solo degli *step* progressivi indicati implica di rimando l'impossibilità di approdare a risultati che il DAD permette invece di ottenere con minor sforzo e maggiore precisione.

Un caso concettualmente molto prossimo a quanto appena detto riguarda anche i legami espliciti tra testi.

Si tenga conto, per esempio, del seguente passo in cui Hollander (42) denuncia come incorretta l'ipotesi di Scott (558-591), che ritiene, sulla scorta di alcuni passaggi dell'*Anticlaudianus*, che Dante non sia stato il primo autore ad aver applicato a un'opera di finzione poetica i principi dell'allegoria biblica:



Scott, however, also asserts that two passages in the *Anticlaudianus* of Alain de Lille also reveal the adaptation of scriptural allegory to a secular work. The text of the first of these passages occurs in Alain's prose Prologus and runs as follows: "In hoc etenim opere litteralis sensus suavitas puerilem demulcebit auditum, moralis instructio perficientem imbuet sensum, acutior allegorie subtilitas proficientem acuet intellectum." Are these senses really "biblical"? My own view is that the passage shows only a marginal and general awareness of biblical exegesis, that the particular exegetical procedure involved, if it does refer to biblical exegesis, has to do only with the sense contained in the literal sense (about which more below), not with a historical sense by any means, and that at any rate Alain (as are others who make similar claims) is a good deal less precise about biblical allegory than even the Dante of the *Convivio*. Scott's second example is still more troubling, especially since he claims that it "is even more significant." He refers to "Alain's claim (in the *Summarium* of the *Anticlaudianus* - Bossuat, p. 201) that his subject is twofold, "una historialis, alia mystica." Had he read either the *Summarium*, which is a fairly confused piece of argumentation, or Bossuat's Introduction with care, he would likely have acceded to Bossuat's well-argued judgment that the *Summarium* is the work of a later and lesser hand. In short, it is more than likely that Alain does not claim "historicity" for his work, and that Scott and Lubac are incorrect.

Fig. 14. Hollander.

Qualora si fosse in possesso del solo saggio in cui Hollander presenta le sue motivazioni in difesa del primato dantesco, sarebbe facile se non addirittura naturale, con tutte le ripercussioni implicite di una coercizione alla scelta, accogliere come vera la sua posizione. Viceversa, se si consultasse solamente lo scritto di Scott, non sarebbe improbabile rimanerne persuasi. Grazie alle caratteristiche di progettazione delle schede del DAD e ai parametri che hanno guidato la selezione e la raccolta dei testi – gli stessi, cioè, che hanno portato tra l'altro al reperimento sia del lavoro di Hollander che quello di Scott – è invece possibile usufruire di collegamenti ipermediali tra i lavori citati (segnalati da una riconoscibile sottolineatura del testo di diverso colore – in questo caso lo 'Scott' ad apertura citazione) per un raffronto simultaneo tra più prospettive esegetiche. Il DAD mette, cioè, a disposizione dell'utente, nel limite del possibile, tutta la documentazione necessaria per predisporgli le condizioni più accademicamente democratiche in cui poter giudicare da sé, in base a criteri che si spera vengano stabiliti con cognizione di causa, quale sia la posizione più corretta in rapporto a una specifica problematica.

Un ultimo esempio riguarda nuovamente la ricchezza informativa del DAD e tiene conto dell'intenzione dell'utente di condurre su di essa ricerche che siano il più approfondite possibile. Interrogando i testi restituiti dal *software* in cui appare almeno un riferimento all'*Anticlaudianus* l'utente noterà con una certa facilità, poiché quasi unanimemente condivisa, la relazione che lega l'opera di Alano alla *Commedia* con riguardo specifico alla condivisione del poema dantesco di personificazioni come forme di allegorie poetiche. È indubbio che, come Alano, Dante abbia riservato un ruolo di primo piano all'ipostatizzazione di concetti attinenti alla sfera etica e religiosa dell'uomo. Tuttavia, se si passa in rassegna la totalità dei lavori segnalati del motore di ricerca del DAD si noterà che uno di essi puntualizza un tratto non considerato altrove:

Singleton's German contemporary, Erich Auerbach (1892-1957), also made recourse to ideas about the Bible's modes of signification. In particular, he utilized the concept of prefiguration to contextualize one of the Comedy's striking novelties: its realistic depiction of psychologically complex individuals in contrast to the stilted personification allegories found in Prudentius' *Psychomachia* (c.405) and allegorical-didactic poems like Alan of Lille's *Anticlaudianus* (1181-4). Auerbach suggested that the denizens of the afterlife were 'figural' fulfilments of their earthly selves, whereby their eternal existence reflected their earthly lives.

Fig. 15. Kriesel.

Il chiarimento storico-esegetico di Kriesel (114) qui riportato, a discapito della sua singolarità, risulta decisamente fondato, come una veloce indagine esterna, sempre necessaria per non demandare tutto alla macchina, conferma. Battaglia Ricci afferma, infatti, che



il confronto tra i due testi [l'*Anticlaudianus* e la *Commedia*] serve anche per mettere in rilievo vari elementi di divergenza: in particolare, a livello di organizzazione narrativa l'esplicitazione che, [...] il testo latino – ma non quello dantesco – offre e che trova un suo evidente corrispettivo nell'utilizzazione di semplici personificazioni contro la qualità 'figurale' delle tre guide dantesche: Virgilio, Beatrice, Bernardo (Battaglia Ricci 135-136).

Una ricerca impostata su una singola parola processata dal DAD – ANTICLAUDIANTUS – ha dunque permesso di cogliere un aspetto che ridefinisce totalmente nelle funzioni, nelle caratteristiche costitutive, nelle implicazioni culturali, storiche, sociologiche il rapporto tra generi, rivelando importanti dissonanze, come quelle che dividono la *Commedia* dai poemi allegorico-didattici personificati, anche laddove sembrava che non vi fossero. Pertanto, l'eterogenea disponibilità di materiali informativi catalogati nel DAD e la loro facile e diretta consultazione consente all'occorrenza di allargare esponenzialmente il tono della ricerca originariamente impostata su poche unità iniziali.

Ricapitolando quanto detto finora, parlare di PLANCTU NATURAE, REGUALE THEOLOGICAE, ANTICLAUDIANTUS implica allora il parlare di Alano di Lilla. Questi passaggi fanno capire come nel conteggio finale del numero di volte che Alano viene menzionato nel *corpus*, il suo peso non deve limitarsi al solo nome esatto ALANO, ma dovranno sommarsi anche i riferimenti all'accostamento di parole quali PLANCTU NATURAE, REGUALE THEOLOGICAE e ANTICLAUDIANTUS.

Stesso dicasi per i derivativi (es. 'aristotelico' per Aristotele), per le traduzioni in altre lingue (es. Gesù, Jesus, Yahweh) e per tutte le varietà o varianti flesse con cui un termine compare (si parla di 'forme' o 'type').

Bisogna poi fare attenzione a ulteriori specificità come i casi di omonimia. Per esempio, al cospetto della voce CHARITY bisogna compiere una disambiguazione tra Alan Clifford Charity e l'atto della benevolenza ('charity' tradotto dall'inglese significa appunto 'beneficenza'). È chiaro che ai fini del *profyling* tematico sull'allegoria dantesca gli studi di Charity, certamente importanti, non hanno motivo di essere paragonati a una manifestazione di carità.

Si deve poi tenere conto anche della sinonimia. Satana, Lucifero e Mefistofele, per esempio, rientreranno tutti in una sola categoria tematica.

Una volta che si è giunti a una indicizzazione completa delle voci più frequenti è però necessaria un'ulteriore scrematura, che potrebbe anche ridefinire totalmente le posizioni stilate fino a quel momento. È richiesta, ovvero, una particolare attenzione alla pertinenza delle voci rinvenute con il tema che ci si è prefissati di esaminare. Ciò significa che, per esempio, se Virgilio e lemmi affini compaiono 50 volte, ma solo in 20 casi riferiti al discorso allegorico mentre nei restanti 30 come rimando a questioni avulse da una specificazione dell'allegorismo della *Commedia* (es. "L'Eneide di Virgilio ha rappresentato un riferimento stilistico e tematico per l'*Inferno*..."), solo il primo blocco di termini (20) e non la sua interezza (50) dovranno essere tenuti in conto. L'eliminazione delle voci non pertinenti richiede nuovamente l'uso del DAD poiché un veloce raffronto critico dei passi che contengono l'*item* evidenziato, reso possibile con la funzione di ricerca integrata nel *database*, può facilmente escludere i termini, e dunque i concetti, non utili allo studio.



utilità come tratti concettuali oggettivi dell'allegorismo della *Commedia*, saranno usati per tracciare un completo profilo storico-critico di quest'ultimo (forme assunte, particolarità, aspetti canonici, etc.) da ritenersi attendibile tanto in una prospettiva storica quanto esegetica. L'interrogazione computazionale dei testi avrà infatti assicurato statistiche affidabili poiché pienamente corrispondenti a campioni numericamente e qualitativamente rilevanti essendo questi i migliori rappresentanti dei più eminenti studi che la dantistica internazionale ha prodotto negli ultimi 120 anni sul tema dell'allegoria nel poema dantesco così come venne formulandosi a partire dall'epoca medievale.¹⁷

BIBLIOGRAFIA

- Alighieri, Dante. *Convivo*, a cura di Franca Brambilla Ageno, Le Lettere, 1995.
- . *Convivo*, a cura di Giancarlo Fioravanti e Claudio Giunta, Mondadori, 2019.
- Ascoli, Albert Russell. "Tradurre l'allegoria: 'Convivio' II, i." *Critica del testo*, vol. XIV, 2011, pp. 153-175.
- Battaglia Ricci, Lucia. "Dante e la tradizione letteraria medievale." *Una proposta per la Commedia*, Giardini, 1983, pp. 135-136.
- Bellomo, Saverio. "How to read the early commentaries." *Interpreting Dante. Essays on the tradition of Dante commentary*, a cura di Paola Nasti e Claudia Rossignoli, University of Notre Dame Press, 2013, pp. 84-109.
- Bonera, Matteo, e Anna Bardazzi. *Data Visualization as a Tool to Experience the Legacy of Dante's Divine Comedy and its Influence on the Cultural Heritage*, www.thevisualagency.it. Consultato il 17 Ago. 2023
- Cerroni, Monica. "Li versi stranil." *Forme dell'allegoria nella Commedia di Dante*. ETS, 2003.
- Curtius, Ernst Robert. *Letteratura europea e Medio Evo latino*, a cura di Roberto Antonelli, Quodlibet, 2022.
- Dronke, Peter. *Dante e le tradizioni latine medievali*. Il Mulino, 2022.
- . "Il secolo XII." *Letteratura Latina Medievale. Un manuale*, a cura di Claudio Leonardi. SISMEL-Edizioni del Galluzzo, 2003, pp. 231-302.
- Pompeo, Giannantonio. *Dante e l'allegorismo*. Olschki, 1969.
- Hollander, Robert. "Dante 'Theologus-Poeta'." *Studies in Dante*, Longo, 1980, pp. 39-89.
- Kriesel, James C. "Allegories of the Corpus." *The Cambridge Companion to Dante's 'Commedia'*, a cura di Zygmunt G. Baranski e Simon Gilson, Cambridge University Press, 2019, pp. 110-126.
- Marchesi, Simone, et al.. "Literary Visualization. Towards a Visual Annotation of Dante's Comedy." *Bibliotheca Dantesca*, vol. 5, 2022, pp. 298-301.

¹⁷ Il *corpus* di testi processato comprende infatti lavori che coprono un arco temporale che va dal 1902 al 2023.



Maselli, Matteo. "Database Allegorico Dantesco (DAD)." *Umanistica Digitale*, vol. 13, 2022, pp. 165-170.

---. "Struttura e funzioni di un database dantesco. Proposta per un'inedita risorsa digitale." *L'altro Dante: processi di attualizzazione della Commedia*, a cura di Mario Cimini, Carabba, 2022, pp. 241-260.

Mazzali, Ettore. "La dottrina e le componenti del poema." *Dante, la vita, il pensiero, le opere*, Accademia, 1979, pp. 91-150.

Mocan, Mira. *Immagine, figura, astrazione. Le geometrie del testo nella "Commedia" di Dante*. Salerno, 2022.

Newman, Mark. "The Structure and Function of Complex Networks." *SIAM Review*, vol. 45, 2003, pp. 167-256.

Padoan, Giorgio. *La "mirabile visione" di Dante e l'Epistola a Cangrande*. Longo, 1977.

Scott, John A. "Dante's Allegory." *Romance Philology*, vol. XXVI, 1973, pp. 558-591.

Smalley, Beryl. *Lo studio della Bibbia nel Medioevo*. Edizioni Dehoniane, 2008.

Van Peteghem, Julie. "Exploring Dante's Sources Online: Interactive Reading, Visualizations, and the Study of Dantean Intertextuality in the Digital Age." *Bibliotheca Dantesca*, vol. 5, 2022, pp. 229-252.

Matteo Maselli è Dottore di ricerca in Umanesimo e Tecnologie presso l'Università di Macerata, dove ha lavorato su un progetto dedicato all'allegorismo dantesco e alle Digital Humanities. Si è formato all'Università di Bologna, dove attualmente è cultore della materia in Critica Letteraria e Letterature Compare, e ha studiato alla Oxford University e all'University of Notre Dame.

<https://orcid.org/0000-0002-3084-6651>

m.maselli2@unimc.it