Umanistica Digitale – ISSN: 2532-8816 – n. 24, 2026
A. Re – *Applicazioni alle lingue e alle letterature classiche delle tecniche di analisi digitale*
DOI: http://doi.org/10.60923/issn.2532-8816/22980

Applicazioni alle lingue e alle letterature classiche delle tecniche di analisi digitale

Alessandro Re

Università degli Studi di Udine
alessandro.re@uniud.it

Abstract

Questo articolo costituisce una *survey* di alcuni recenti studi inerenti vari metodi di applicazione delle tecniche di *Natural Language Processing* alla filologia classica, soprattutto nell'ambito dell'epica latina del I sec. d.C. Sebbene concepiti per l'indagine delle lingue moderne, i vari algoritmi sviluppati dalla linguistica computazionale nell'ultimo trentennio – da *Latent Sematic Allocation*, a *Latent Dirichlet Allocation*, fino ai più recenti *Transformer* basati sulle moderne reti neurali – hanno rivoluzionato la ricerca degli intertesti attraverso la pratica della *Topic Extraction*. L'interesse di chi scrive si è focalizzato in particolare su quei casi volti a individuare corrispondenze situazionali pur in assenza di vocaboli co-occorrenti. Gli esempi discussi confermano la bontà di questo genere di approccio e la convergenza tra le tecniche 'tradizionali' e quelle linguistico-computazionali.

Parole chiave: epica latina; intertestualità; Natural Language Processing; Topic Extraction; Corrispondenze situazionali.

This article offers a survey of recent studies on the application of Natural Language Processing techniques to classical philology, with particular emphasis on Latin epic poetry of the 1st century AD. Although originally designed for the analysis of modern languages, the algorithms developed in computational linguistics over the past three decades – ranging from Latent Semantic Allocation and Latent Dirichlet Allocation to the more recent Transformer models based on Neural Networks – have transformed intertextual research through the practice of Topic Extraction. Particular attention is given to cases aimed at detecting situational correspondences even in the absence of co-occurring words. The examples confirm both the efficacy of this approach and the convergence between 'traditional' philological methods and computational-linguistic techniques.

Keywords: Latin epics; Intertextuality; Natural Language Processing; Topic Extraction; Situational correspondances.

1. Premessa

Gli ultimi cinque lustri hanno conosciuto un rapido incremento delle tecniche di analisi digitale dei testi attraverso l'elaborazione di *software* volti a impiegare il *machine learning* per compiere azioni

complesse aventi come oggetto l'indagine del linguaggio naturale (*NLP*, cioè *Natural Language Processing*; cfr. Russell e Norvig 2020, 669–931). Tali applicazioni di *NLP* sono usate per ricavare *insight* da dati di testo anche molto ampi, consentendo di accedere alle informazioni estratte per generare una nuova comprensione di tali dati.

La formazione di chi scrive ha portato a indagare le loro applicazioni alle lingue classiche e, in particolare, al latino: nel corso dei decenni, molteplici modelli sono stati implementati allo scopo di migliorare le capacità d'indagine nei diversi campi dell'analisi testuale (cfr. Sommerschield *et.al.* 2023). Nel tracciare questa *survey*, l'attenzione di chi scrive si soffermerà, tra le varie piste di ricerca, sulla ricerca intertestuale, aspetto che, sin dall'evo antico, ha conosciuto grande rilevanza negli studi filologici.

2. Intertestualità digitale

Nella teoria letteraria, con il termine 'intertestualità' sono definite le relazioni che legano un testo ad altri: da sempre, essa persegue svariate strategie che possono essere pienamente deliberate – *e.g.* citazione, calco, allusione, traduzione, parodia – oppure sono ottenute mediante interconnessioni tra opere simili che siano percepite soprattutto da un pubblico colto; inoltre, se certi riferimenti sono fatti in modo consapevole e sono intrinsecamente dipendenti dalla conoscenza del referente da parte dell'autore, in altri casi l'intertestualità è un fatto involontario. Ciò detto, essa riguarda qualsiasi testo, giacché nulla di quanto viene prodotto – sia nell'ambito strettamente letterario, sia nel più ampio raggio dell'arte figurativa (pittura, scultura, fotografia e cinematografia) e dei *media* digitali – può essere del tutto indipendente da quello che in precedenza era stato composto.

Queste considerazioni e il nome stesso di *intertextualité* derivano dalle riflessioni della pensatrice franco-bulgara Julia Kristeva (in particolare Kristeva 1969):[1] nel tentativo di armonizzare la teoria semiotica di Ferdinand de Saussure, circa il modo con cui i segni traggono il loro significato all'interno della struttura di un testo, con il dialogismo di Mikhail Mikhailovich Bakhtin, la cui teoria presuppone un rapporto continuo con altre opere e altri autori, ella riconosce in ogni testo molteplici significati che sono strettamente connessi con il complesso *background* da cui ciascuno è ispirato. Come ha correttamente rilevato Nozomu Okuda nella sua recente tesi di dottorato,

> Kristeva's position on how literary texts are produced leads into the crux of intertextuality. The key player here is " the 'literary word' ". As Kristeva herself emphasized, the key concept that differentiates intertextuality from structuralism is the " *intersection of textual surfaces* ", in contrast to " a *point* (a fixed meaning) ". Structuralism is about finding the one point and discerning the fixed meaning of the literary word. Intertextuality is about looking for the many points and appreciating the fluid meanings of the literary word (Okuda 2022, 86–87).

Da Bakhtin, la studiosa ha derivato l'idea secondo la quale i testi sono prodotti non da un semplice modello ma da un processo di relazione; di conseguenza, le parole di un testo

[1] Per un inquadramento generale del concetto d'intertestualità si faccia riferimento ad (Alfaro 1996; Martin 2011) e soprattutto ad (Allen 2021); in opposizione alla teoria di Kristeva si veda invece (Irwin 2004).

consentono molte interpretazioni, per cui i vari fruitori dello scritto possono trarre significato in base a circostanze ed esperienze pregresse. Dunque, l'intertestualità consiste nella qualità di un passo i cui molteplici sensi sono il risultato della relazione che le parole hanno tra loro. Nello specifico, un intertesto è composto da un *locus* testuale che si accoppia a un altro cosicché il significato del primo è informato da quello del secondo.

La genesi degli intertesti è anzitutto propria dell'autore: componendo l'opera, egli li produce scegliendo le parole sulla base di come esse sono collegate al modello. Una volta che il testo viene diffuso, anche i suoi fruitori, leggendolo, produrranno intertesti in quanto collegheranno i vocaboli di questo a quelli esterni. Lo scrittore li può includere consciamente o inconsciamente: per le inclusioni coscienti egli s'interfaccia direttamente con un modello; per quelle inconsce, invece, consente ad altri testi d'indirizzare la scelta delle parole indipendentemente dall'influsso esercitato. Ugualmente, dalle opere che leggono, i lettori possono produrre consciamente o inconsciamente intertesti: nel primo caso, istituiscono connessioni tra le parole del testo e quelle di un altro; nel secondo, consentono ad altre opere d'influenzare il significato che attribuiscono al lessico. Tanto nella produzione quanto nel riconoscimento degli intertesti è imprescindibile il contesto culturale, che condiziona il modo in cui scrittori e lettori si approcciano alla scrittura o alla lettura: quando il lettore o lo scrittore produce o riconosce consapevolmente un intertesto, entrambi i passaggi erano già nella mente del creatore dello stesso; quando gli stessi producono o riconoscono inconsciamente un intertesto, più intertesti che hanno attinenza con lo specifico passaggio si compenetrano in proporzioni variabili.

Le riflessioni di Kristeva devono essere inserite entro la più ampia trattazione semantica sviluppata nella scuola linguistica francese del secondo dopoguerra. In *Sémantique structurale* (Greimas 1966), si afferma che il senso di un vocabolo non esiste in modo 'isolato': esso è generato a partire da strutture logiche profonde che possono essere graficamente espresse con il "quadrato semantico", la cui più remota formulazione si trova nel "quadrato delle opposizioni" già attestato nella logica aristotelica. Nei capitoli IV e V si analizzano i due livelli che concorrono alla definizione del significato: uno è il livello semiologico che stabilisce i semi nucleari – i tratti minimi e costanti della parola –, un altro è il livello semantico che definisce i semi contestuali – o classemi – indipendenti dall'oggetto ma legati al contesto in cui è inserito. Quindi, nel capitolo VI, è introdotto il concetto di semema: il significato di una parola nella frase non è mai fisso, risultando dall'unione tra nucleo stabile e influenza del contesto; la disambiguazione del significato di un atto comunicativo dipende dalla ridondanza dei classemi da cui consegue la coerenza testuale. In sintesi, con Greimas l'attenzione non è più rivolta alle singole parole: ogni testo è un 'microcosmo' con le sue regole, le sue gerarchie di valori e il suo modo di organizzare i semi; analizzarlo significa ricostruire la struttura di questo 'piccolo mondo chiuso'.

Tali teorie sono riprese e ulteriormente approfondite in *Sémantique interprétative* (Rastier 1987): da Greimas lo studioso francese riprende i concetti di sema come unità minima del significato di una parola e di isotopia che, da principio che spiega l'unità concettuale di un discorso, diventa l'elemento fondamentale dell'analisi testuale. La principale differenza tra i due pensatori risiede nella differente prospettiva adottata per spiegare la genesi del significato: se in Greimas esso è generato a partire da strutture profonde e astratte che vanno dal logico al testuale, in Rastier il senso si costruisce nell'interpretazione del testo concreto come effetto della lettura e del contesto. Alla rigidità del primo sono sostituite le molecole semiche, tratti che si muovono e si trasformano nel testo in modo più fluido e meno prevedibile.

Si diceva precedentemente che capitale importanza nella definizione della teoria intertestuale elaborata da Kristeva assume il riferimento alla lezione saussuriana circa il rapporto – arbitrario, eppur costante – tra *signifiant* e *signifié*: ora, nel considerare la forma letteraria in ottica di *NLP*, è possibile dire che ogni *token* è strettamente collegato con la sua rappresentazione vettoriale (*e.g.*

Turian *et al.* 2010; Blei 2012; Almeida e Xexéo 2019; Günther *et al.* 2019); questo approccio, quindi, richiama indirettamente il sistema relazionale tra significante e significato descritto da Saussure. Dunque, nei modelli di *Word embedding*, i *token* equivalgono ai significanti e le rappresentazioni vettoriali ai significati: ogni *token* si riferisce a qualcosa, fungendo da significante; ciascuna rappresentazione vettoriale – il *Word embedding* – codifica le informazioni sul *token*, agendo da significato. In un certo senso, si potrebbe dire che il moderno *NLP* abbia dato una 'veste matematica' alle intuizioni che i citati studiosi di semiotica avevano formulato in termini teorici: come il significato è una combinazione di tratti minimi – i semi –, così nei *Word embedding* ogni parola è rappresentata come un vettore in uno spazio multidimensionale; ogni dimensione può essere vista come un 'sema latente' e, per quanto non sempre sappiamo che cosa significhi, essa cattura un certo tratto distintivo che, insieme agli altri, definisce il senso della parola in relazione alle altre.

La teoria di Rastier è probabilmente quella che più si avvicina alla logica dei moderni modelli di linguaggio: l'idea che il senso nasca dal contesto e dall'afferenza richiama l'ipotesi distribuzionale di Firth che sta alla base dei *Word embedding* (Firth 1957; Lenci 2008; Sahlgren 2008). Nelle più moderne tecniche di *NLP* – per esempio i *Transformer* (cfr. *infra*) – non si usa più un unico vettore per parola giacché quello di X cambia valore se si trova vicino a Y piuttosto che a Z:[2] ciò è l'esatta traduzione algoritmica del concetto secondo cui il contesto attiva quei tratti semantici – i semi afferenti – che trasformano il significato di base. Anche il concetto di isotopia trova un corrispettivo tecnico nella similarità del coseno: come un testo è coerente qualora le parole che lo compongono condividano isotopie, così in un *cluster* di documenti la coerenza è misurata calcolando quanto i vettori delle parole siano vicini tra loro nello spazio semantico; in termini matematici, un'isotopia è un'alta densità di vettori che puntano nella stessa direzione.

L'arbitrarietà gioca un ruolo importante nelle prime rappresentazioni vettoriali di *Word embedding*: essendo pratica comune per le reti neurali inizializzare il modello con valori casuali, anche le rappresentazioni vettoriali saranno inizialmente casuali e, solo dopo aver addestrato il modello con debiti processi di *training* e di *fine tuning*, esse si stabilizzeranno in forme utili. Si potrebbe definire un *Transformer* come una potente 'macchina rastieriana' la quale, pur non conoscendo le definizioni del dizionario, intuisce il senso di un testo solo guardando come i tratti semantici si raggruppano in esso, applicando una semantica interpretativa automatizzata secondo una scala che Rastier poteva solo teorizzare.

Questo sintetico quadro teorico è alla base della seguente *survey* degli studi riguardanti l'indagine degli intertesti su basi informatiche nell'ambito della filologia classica, latina in modo particolare.

3. Una breve storia degli studi

Risale all'anno 2008 uno studio di David Bamman e Gregory Crane circa la rilevazione delle allusioni nella poesia latina classica (Bamman e Crane 2008). Anzitutto i due autori passano in rassegna le diverse tecniche allusive proprie della tradizione antica: il tipo di allusione più esplicito

[2] Per esempio, si pensi al differente valore che *capo* assume nelle espressioni *capo di abbigliamento* e *capo ufficio*.

e inequivocabile è un riferimento sotto forma di citazione letterale[3] o indiretta;[4] si possono però individuare anche altre tipologie. Un primo caso riguarda l'*ordo uerborum* e la somiglianza sintattica, quando – come nel caso enunciato alla nota 4 – oltre a vocaboli identici (*bella* e *-que*) o che presentano un chiaro rapporto sinonimico (*cano* in Virgilio ed *edere* in Ovidio), l'autore riprende anche la struttura argomentale dell'espressione linguistica. Nell'ambito strettamente poetico, si riscontra non solo il recupero dello stesso metro ma anche di fenomeni prosodici particolari (cfr. Forstall *et al.* 2011): per rimanere sempre al citato legame tra gli *incipit* di *Eneide* e *Amores*, la somiglianza è molto più stretta dal momento che le prime sette sillabe di entrambi i versi (due dattili seguiti dall'arsi del terzo) sono metricamente identiche; inoltre, la sillaba finale *ō* prima della cesura pentemimere è la stessa in entrambe le frasi. L'ultimo caso che Bamman e Crane individuano è costituito dalla somiglianza sul piano semantico/contenutistico: nello stesso esempio, sia Virgilio sia Ovidio parlano di guerra e di strumenti bellici; però l'aspetto più interessante riguarda il fatto per cui

> [w]ith semantic similarity we can also group another very important variable – cross-language semantic information in the form of translation equivalents. This is extremely important given the reception of these texts across cultures and distant eras. Classical Roman poets themselves are especially fond of borrowing from Homer and Hellenistic poets (par. 3).

Questo aspetto, che la filologia classica ha ampiamente trattato sin dall'evo antico (*e.g.* Cannizzaro 2023), può certamente conoscere nuovi impulsi grazie alle recenti tecniche di *NLP*: alla base della ricerca automatica delle allusioni sta la trasformazione delle variabili sopra elencate in parametri 'numerici', trattabili sul piano computazionale; bisogna essere capaci di definire il grado preciso in cui due passaggi sono vicini per poter confrontare quantitativamente quali coppie di passaggi sono più simili tra loro rispetto ad altre. Ora, esistono diversi metodi per giudicare la somiglianza di due documenti. Quelli più utilizzati assegnano un punteggio di rilevanza basato sulla registrazione della variazione di tf–idf (*term frequency – inverse document frequency*):[5] due documenti sono tanto più simili se contengono parole che meno frequentemente ricorrono nella raccolta generale di testi giacché, più voci insolite condividono, maggiore sarà la somiglianza. Una delle più comuni tecniche atte a stabilire la somiglianza tra la rappresentazione vettoriale di due testi (o loro sezioni) si fonda sul coseno di similitudine (cfr. Morrelli *et al.* 2020).

La seconda parte dell'articolo di Bamman e Crane ha invece carattere sperimentale. Stabilito un ristretto *corpus* di autori (Catullo, Virgilio, Orazio, Properzio, Ovidio), vi vengono applicate tecniche d'indagine computerizzata riguardanti i lemmi identici, l'*ordo uerborum* e la somiglianza sintattica: mentre i primi due aspetti possono essere verificati anche su testo non strutturato, si

[3] Per esempio, Ov. *am.* 1, 13, 39-40: *at si, quem mauis, Cephalum conplexa teneres, | clamares: "lente currite, noctis equi!"*, rispetto a Christopher Marlowe, *Faust*, atto V, scena 2: «Stand still, you ever-moving spheres of heaven, | That time may cease, and midnight never come: | Fair Nature's eye, rise, rise again and make | Perpetual day; or let this hour be but | A year, a month, a week, a natural day, | That Faustus may repent and save his soul! | *O lente, lente, currite noctis equi!*».

[4] Per esempio, Verg. Aen. 1, 1: Arma uirumque cano […], rispetto a Ov. am. 1, 1-2: Arma graui numero uiolentaque bella parabam | edere […].

[5] Si tratta di una funzione utilizzata nel campo dell'*information retrieval* per valutare l'importanza di un vocabolo rispetto a un singolo documento o a una collezione di documenti: il valore è direttamente proporzionale al numero di occorrenze del termine e inversamente proporzionale alla frequenza del termine nella collezione; così si dà maggiore rilievo ai termini che compaiono nel documento ma che in generale sono poco frequenti (Manning *et al.* 2009, 117–19).

ha bisogno di dati 'etichettati' sintatticamente in relazione al terzo. Pertanto si è proceduto ad addestrare un *dependency parser* (cfr. McDonald *et al.* 2005) sui dati contenuti nella *Latin Dependency Treebank* onde impiegarli nell'analisi del *corpus* sopra citato (cfr. Celano *et al.* 2023). Circa gli esiti si afferma:

> The results are encouraging: while a detailed quantitative evaluation must await the creation of a test corpus of canonical allusions, we can at least now provide a list of the closest matches for all sentences in our collection. For any given sentence, further research will of course be necessary to discern whether it represents a real allusion, but the highest scoring pairs in our experiment tend to be strong examples (par. 5).

Dai dati raccolti emerge con chiarezza il fatto che l'uso della sintassi come metodo per stabilire la somiglianza tra le frasi dà risultati assai buoni, evidente testimonianza che le tecniche allusive riguardano livelli anche molto diversi: in questo campo il ruolo del *computer* potrà fornire una messe di dati sempre più 'raffinati'.

Di grande interesse e massima utilità nel compiere ricerche computerizzate di carattere intertestuale è l'ampia trattazione che si legge in (Coffee *et al.* 2012): l'articolo elabora un accurato confronto tra i metodi tradizionali e un approccio digitale all'intertestualità, indagando con dovizia di particolari in che modo la rilevazione automatica degli intertesti abbia significative ricadute pratiche e consenta anche l'apertura di nuovi orizzonti teorici. Il fulcro della trattazione riguarda l'impiego di *Tesserae*, uno strumento *online* gratuito creato da un'*équipe* della University at Buffalo coordinata dallo stesso Neil Coffee e accessibile all'URL https://tesserae.caset.buffalo.edu/.[6] Gli autori affermano con soddisfazione che

> the *Tesserae* tool recovers approximately a third of the parallels captured by traditional commentators, and adds a third not previously recorded. These results show that the tool can find valuable new parallels in even the best-studied authors, and suggest that it might be particularly useful in identifying intertexts in works that have received less attention. In combination with information from commentators, our method allows us to draw conclusions about the overall intertextual artistry of ancient authors, as well as the intertextual reading habits of scholars (Coffee *et al.* 2012, 386).

Il funzionamento di *Tesserae* si concentra sulla ricerca dei paralleli testuali caratterizzati da almeno due lemmi identici: come si è visto in precedenza, questo modo di procedere è alla base della 'tradizionale' identificazione dei *loci similes* che considera coppie di due parole come la forma più elementare e comune d'intertestualità; sono invece tralasciate le interazioni su piccola e larga

[6] Come recita un *banner* accessibile all'indirizzo citato, al momento della stesura di questo saggio – settembre 2025 – «[w]e have updated *Tesserae* to a new version (Version 5). The update is meant to improve the user experience with a more modern interface. It also offers a corpus view that allows users to see all the texts in the *Tesserae* corpus and choose which ones to compare. It includes a web API that allows other online services to request the *Tesserae* server to run searches. […] The new version is written in Javascript and Python, and uses a MongoDB database. This new basis should allow for easier updates and improvements going forward. Version 5 is a work in progress and we ask for your patience with any issues as we make fixes and improvements». Tuttavia, la precedente versione 3, quella usata in occasione della stesura del citato articolo, risulta ancora accessibile a questo URL: https://tesseraev3.caset.buffalo.edu/. Sul funzionamento di *Tesserae* si vedano anche (Coffee *et al.* 2013; Okuda *et al.* 2022).

scala che non possono essere identificate solo attraverso la semplice somiglianza lessicale. Come banco di prova per la verifica del funzionamento del *software*, i ricercatori si sono soffermati sulla comparazione tra il Libro I del *Bellum ciuile* di Lucano e l'*Eneide* di Virgilio: la scelta è caduta su queste due opere perché sufficientemente lunghe da fornire risultati rappresentativi e perché così ben studiate da consentire il confronto anche con gli approcci tradizionali. Il duplice *test* è condotto con algoritmi diversi: nella versione 1 si sono cercati sintagmi minimi con due parole identiche in tutto e per tutto sul piano flessivo, senza considerare l'ordine con cui appaiono nel testo e separate da non più di quattro altri vocaboli;[7] nella versione 2, invece, sono state abbinati sintagmi minimi costituiti da due parole identiche in base ai lemmi del dizionario, ignorando dunque la flessione e in qualsiasi ordine di attestazione.[8]

Dopo aver raccolto i dati, si è proceduto a una classificazione dei risultati allo scopo di valutare in che modo il calcolatore abbia associato gli ipotetici *loci similes*: tale valutazione è stata compiuta manualmente confrontando i dati raccolti con le analogie rilevate dai commentatori. Il punteggio ha cinque gradazioni: i paralleli più calzanti sono stati valutati da 5 a 3, mentre hanno ricevuto una valutazione di 2 o 1 quei casi in cui la somiglianza è troppo labile da risultare significativa sul piano intertestuale (Coffee *et al.* 2012, 392–98). In termini numerici, l'indagine computerizzata ha permesso di rilevare un numero considerevole di paralleli assai significativi per l'esegesi delle opere: *Tesserae* ne ha aggiunto 279 ai 364 già segnalati dai commentatori, con un aumento del 43%, per un totale di 643 *loci similes* tra il Libro I del *Bellum ciuile* e l'*Eneide*. Al riguardo, gli autori osservano che

> [a]lthough the *Tesserae* search captured only a quarter of the high interest type 5–4 parallels found by the commentators, this recovery rate nevertheless shows the capacity of automatic search to replicate traditional methods. Equally compelling is the fact that *Tesserae* search returned a high proportion of previously undiscovered parallels. This result suggests that the systematic application of fixed criteria can detect parallels traditional criticism may tend to overlook (Coffee *et al.* 2012, 400).

Un ulteriore aspetto sul quale si appunta l'attenzione degli estensori di questo articolo riguarda la provenienza virgiliana del materiale ripreso da Lucano nel Libro I: in generale, sia i commentatori sia *Tesserae* hanno rilevato significative variazioni quantitative circa le connessioni intertestuali con i diversi libri dell'*Eneide*, a significare una precisa scelta delle fonti da parte dell'autore più tardo rispetto a quello anteriore; in tal senso, proprio come Servio dedica una parte molto maggiore del suo commento alla prima esade rispetto alla seconda, nella composizione del Libro I Lucano sembra dimostrare un maggiore interesse per la prima metà dell'*Eneide* (Coffee *et al.* 2012, 404–13).

Se poi si prendono in considerazione non solo i paralleli interpretabili (valutazione 5 o 4) ma si aggiungono anche i tipi 3 e 2, s'individua un'altra tendenza: tanto i commentatori tradizionali quanto la ricerca automatica hanno riconosciuto una percentuale significativamente più elevata

[7] Per esempio, viene individuato il parallelo tra Verg. *Aen.* 9, 381-382 *dumis* […] | *horrida* e Lucan. 1, 28 *horrida* […] *dumis*: questa coppia consente di osservare come *Tesserae* riconosca paralleli anche se non è rispettato l'*ordo uerborum*.

[8] In tal caso sono abbinati Verg. *Aen.* 12, 942 *notis fulserunt cingula bullis* con Lucan. 1, 244 *notae fulsere aquilae*: sorprendentemente, annotano gli autori che «[t]he last example above was found by our lemma search, but does not appear in any of the major commentaries on Lucan, including the ample new volume on *BC* 1 by Roche»; il riferimento è al commento di Paul Roche (Roche 2009).

d'intertesti fondati sia sul «pure language reuse/code model language» (tipo 3) sia sul linguaggio ordinario (tipo 2). La crescente individuazione di paralleli per un verso è diretta conseguenza dell'esame più minuzioso del *Bellum ciuile* da parte dei commentatori successivi, per un altro deriva anche dal crescente utilizzo della ricerca digitale; sotto questo punto di vista, il volume di Roche costituisce un significativo punto di svolta rispetto ai precedenti esegeti giacché egli afferma di aver fatto uso di queste nuove tecniche digitali a supporto della propria indagine (Coffee *et al.* 2012, 403, nota 40).

In conclusione, gli autori riconoscono che, attraverso l'impiego di *Tesserae*, sono stati confermati gli assunti ampiamente noti agli studi esegetici circa l'intertestualità occorrente tra Virgilio e Lucano in termini non solo di ripresa stilistica ma anche di opposizione ideologica tra i due:

> A large-scale view allows us to see how this oppositional tenor is articulated in *BC* 1 through local strategies. Beyond amplifying and trumping the opening of the *Aeneid* in the first book of his own epic, Lucan creates a series of transformative identifications. Civil war recapitulates the destruction of Troy (*Aen.* 2). Prophecies of progress (*Aen.* 3) turn to forebodings of catastrophe. The madness and panic of Carthage (*Aen.* 4) spreads to all of Rome. Crimes worthy of the underworld (*Aen.* 6) appear on earth. Fate as a guiding force in war (*Aen.* 7) is shown to be hollow. The purpose of military honor, even as hesitantly posed by Vergil (*Aen.* 11), is questionable. Caesar is assimilated to Turnus while Pompey appears weaker than Aeneas (*Aen.* 12). These are just some of the ways in which Lucan refers to individual books of the *Aeneid* to demonstrate that civil war is the latest and most devastating repetition of previous mythical struggles, and that Vergil's honorable mythology of Rome's foundation is but a shadow of the truth (Coffee *et al.* 2012, 413–14).

Il principale 'limite' di *Tesserae* risiede, per così dire, nel fatto di rilevare i rapporti intertestuali solo se i due passi presentano vocaboli comuni. Tuttavia, come si accennava in precedenza, già sul finire del XX secolo erano state sviluppate tecniche linguistico-computazionali finalizzate a identificare parole e gruppi di parole semanticamente correlati così da scovare sinonimie e antinomie e identificare contesti tematici simili.[9] In ogni caso, tra i meriti di questa ricerca è lo sviluppo della *Lucan-Vergil Benchmark* i cui dati sono accessibili tra i *Collected Benchmark Sets* all'URL https://tesserae.caset.buffalo.edu/blog/collected-benchmark-sets-updated/.

Un altro approccio allo studio degli intertesti nella poesia esametrica latina mediante *Tesserae* è tracciato in (Bernstein *et al.* 2015). Considerata la pluralità di forme espressive – epica, satirica e didascalica – che la produzione in esametri ha assunto, la ricerca muove dalle seguenti domande:

> Is it possible to quantify the verbal cohesiveness and distinctiveness of these genres? What other general factors affect text reuse across the entire hexameter tradition? Can the well-known influence of Vergil and Ovid on their epic successors be quantified? In particular, can it be determined how frequently one predecessor's text is reused compared to another's? For example, is Statius' *Thebaid* more "Vergilian" in terms of text reuse than

[9] Un esempio d'implementazione delle tecniche di *topic modelling* è la ricerca svolta sul diario di Martha Ballard (cfr. Blevins 2010); per le applicazioni all'ambito classico si veda anche (Scheirer *et al.* 2016).

> another contemporary epic poem, Silius Italicus' *Punica*? […] Which works in the classical hexameter tradition provide the most significant verbal resources for the hexameter epics of late antiquity? (Bernstein *et al.* 2015, n. 7)

Attraverso un'opportuna scelta dei parametri descritta ai nn. 8-18 (§ 2), i dati raccolti e opportunamente campionati secondo una scala di valori numerici consentono di affermare che, innanzitutto, l'influenza più importante nell'intensità di riutilizzo del testo dipende dalla paternità dell'opera: in tal senso, Virgilio e Claudiano mostrano la maggior frequenza d'intertesti all'interno del proprio *corpus*, mentre Orazio e Stazio sembrano 'reimpiegare' le proprie poesie con meno intensità. Un'influenza secondaria sulla frequenza degli intertesti è il genere letterario: gli autori che praticano lo stesso genere mostrano affinità maggiori rispetto a quelli la cui produzione diverge tematicamente; tali risultati confermano le aspettative dell'esegesi 'tradizionale', provando che epica e satira sono assai distanti tra loro tanto per stile quanto per argomenti. Infine, la collocazione temporale sembra non avere alcuna influenza sulla frequenza degli intertesti: ciò non sorprende affatto, dal momento che i modelli tecnici ed estetici della poesia esametrica scoraggiano significativi cambiamenti nella dizione.

Accanto a queste osservazioni di ordine generale, possono essere aggiunte altre puntualizzazioni. Anzitutto, il *De rerum natura* di Lucrezio non è mai stato una risorsa importante per gli autori successivi. Sul versante dell'epica, l'*Eneide* dimostra una grande influenza sulla poesia di tutti i periodi; Lucano, Valerio Flacco, Silio Italico e Stazio appaiono strettamente correlati mentre i poeti tardo-antichi, secondo le previsioni, riusano ampiamente il materiale delle opere precedenti. La congruenza di questi risultati con gli studi tradizionali supporta dunque la tesi di partenza, mentre i risultati inattesi possono essere uno spunto per ulteriori ricerche: per esempio, le *Georgiche* e l'*Iliade Latina* hanno ottenuto punteggi elevati in termini d'intensità di rapporti intertestuali; ancora, l'*Ars poetica* di Orazio sembra avere inaspettatamente influenzato gli *Astronomica* di Manilio, suggerendo il fatto che la sensibilità didattica può trascendere i singoli generi; infine, la *Mosella* di Ausonio, solitamente considerata di natura virgiliana, mostra stretti legami anche con le *Metamorfosi* di Ovidio (Bernstein *et al.* 2015, nn. 19-40).

Come nel caso precedente, anche questo articolo dimostra come la maggior parte dei risultati sia conforme agli assunti ricavati dagli esegeti con metodi 'tradizionali': Virgilio e Ovidio sono le risorse principali; al converso, l'incidenza della *satura* è fortemente minoritaria rispetto agli altri generi. Se si accetta che l'alto livello di correlazione tra questi risultati numerici e le valutazioni qualitative della tradizione accademica siano conferme della bontà di questa metodologia, indirettamente si può comprendere anche il significato di certi risultati non pienamente attesi.

Nell'ambito dell'indagine intertestuale va ascritto anche lo studio di Patrick J. Burns circa la ricerca di riferimenti tra il *Bellum ciuile* e le elegie di Tibullo, Properzio e Ovidio mediante *Tesserae* (Burns 2017). Partendo dalle osservazioni sull'autore di Cordoba in (Coffee *et al.* 2012), dalle riflessioni generali sulla poesia esametrica in (Bernstein *et al.* 2015) e dalle ricerche sui rapporti tra Lucano e Orazio in (Gross 2013), lo studioso americano ritiene che

> [b]oth Farrell and Conte argue that the relationship between two texts can be drawn to some degree by the volume of potential intertexts and their consistent presence throughout a target text. A collocation tool like *Tesserae*, by algorithmically determining and reporting a complete collection of correspondences, offers a formalization of Farrell's system and Conte's model. Moreover, the data collected from *Tesserae* results can be used to

> formalize Helbig's observation about frequency and distribution as implicit signposts for unmarked intertextuality (Burns 2017, par. 2.1).[10]

Un accurato settaggio dei parametri di *Tesserae* restituisce dati che consentono di ottenere interessanti osservazioni riguardo i rapporti tra Lucano e l'elegia (Burns 2017, par. III–IV). Se non sorprende che i rapporti tra Virgilio e Lucano, misurati in base alla frequenza delle parole e alla densità delle frasi, siano superiori a quelli con gli elegiaci, tuttavia si possono individuare casi esemplari d'intertestualità: particolarmente significativo è il rapporto con le *Heroides*, chiara prova del fatto che le osservazioni già note ai filologi sono confermate anche da prove di carattere numerico; in generale, mappando tutte le corrispondenze tra elegia e Libro I del *Bellum ciuile*, si rileva che i legami si distribuiscono in modo diffuso, così dimostrando che l'elegia funziona anche come codice modello. Pur modificando la finestra d'indagine dal singolo verso a 25 versi il risultato non cambia: le corrispondenze sono distribuite nell'intero testo; inoltre, come spunto per ulteriori studi, sono evidenziati i *loci* lucanei maggiormente influenzati dalla poesia elegiaca. Così, se alcuni casi confermano il lavoro dei precedenti studiosi,[11] in altre sezioni la ricerca circa l'interazione dei due generi non era mai stata verificata.[12]

Come già nei precedenti esempi, Burns conclude che

> [e]xamination of the peaks and troughs of the text maps of allusive distribution forms an excellent starting point for a more "sensitive assessment" of the intertextual relationship between Lucan and the love elegists in the future and also provides a more empirical method for investigating the "generic enrichment" […] found in Latin poetry (Burns 2017, par. VI).

Tutti gli studi sin qui esaminati hanno fondato la ricerca d'intertesti sull'impiego di strumenti informatici che rilevano la ripetizione di vocaboli o di particolari sintagmi: è il caso del pluricitato *Tesserae* o di *Diogenes*.[13] Tuttavia, più recentemente, l'articolo di (Burns *et al.* 2021) intende superare questo 'limite' attraverso l'ottimizzazione di modelli di *Word embedding* per il latino così da poterli applicare a problematiche d'intertestualità (cfr. Bamman 2014). Considerata la quasi totale mancanza di tecniche di *NLP* applicate alle lingue classiche, gli autori si propongono di addestrare strumenti predisposti per altre lingue – in particolare l'inglese – su un *corpus* latino: già erano stati fatti tentativi di generazione e conseguente valutazione di *Word embedding* su testi non pre-processati mediante *Word2Vec*, *FastText* e *BERT* (*e.g.* Bamman 2012; Bjerva e Praet 2015; Grave *et al.* 2018; Devlin e Chang 2018; Devlin *et al.* 2019; Bamman e Burns 2020; Lendvai e Wick 2022); tuttavia i più recenti studi hanno evidenziato come il progresso dei risultati sia

[10] In queste righe l'autore fa riferimento agli studi di (Conte 1974, 1986; Helbig 1996; Farrell 2005).

[11] «For example, we see pronounced upticks in the average number of matches at the end of book 5 for Pompey and Cornelia's meeting before Pharsalus (Lucan 5.722-815), and at the beginning of book 10 for the appearance of Cleopatra (Lucan 10.53-106, 172-192)» (Burns 2017, par. 5.2).

[12] «[F]or example, in the book 1 proem (1.1-32) or Cato's speech to his troops in book 9 (9.222-283)» (*ibidem*).

[13] *Diogenes* è un *software* elaborato da un'*équipe* diretta da Peter Heslin del Department of Classics and Ancient History presso la Durham University e accessibile a questo URL: https://d.iogen.es/d/index.html. In *home page* si legge questa descrizione della risorsa: «*Diogenes* is a desktop/laptop application for searching and browsing the legacy databases of texts in Latin and ancient Greek that were once published on CD-ROM by the TLG and the PHI».

strettamente collegato all'impiego di un *corpus* accuratamente lemmatizzato e morfologicamente annotato.[14] Inoltre, condizione necessaria per una soddisfacente valutazione dei risultati, sono stati introdotti anche *database* per la selezione e la valutazione dei rapporti di sinonimia (cfr. Sprugnoli, Moretti, *et al.* 2020; Spinelli 2023). Un primo risultato di questa ricerca è espresso in questi termini:

> For the synonym search task, we consider the number of correct matches found in the top 1, 10, and 25 results by cosine similarity, as well as the mean reciprocal rank (MRR). We find that our models achieve state-of-the-art performance on both tasks compared to the five published models. The improvement in performance may be due to the combination of training on lemmatized text, which Sprugnoli *et al.* (2019) identified as an important optimization for Latin, and use of a lower-quality but much larger training corpus (1.38 billion tokens, compared to 1.7 million tokens in the curated LASLA corpus) (Burns *et al.* 2021, 4901).

Queste considerazioni portano alla necessità di operare un confronto tra le situazioni di intertestualità scovate attraverso le tecniche di *NLP* e quelle già registrate dai commenti: per compiere questa azione si è dunque costruito un *database* nel quale sono registrati i *loci similes* tra il Libro I delle *Argonautiche* di Valerio Flacco da un lato, e l'*Eneide* di Virgilio, le *Metamorfosi* di Ovidio, il *Bellum ciuile* di Lucano e la *Tebaide* di Stazio dall'altro (Dexter *et al.* 2023).[15]

Dunque, se i diversi metodi di ricerca computazionale sinora applicati alla lingua latina si basano sulla corrispondenza lessicale di parole correlate, gli autori di questo saggio presentano un approccio alternativo in cui i potenziali intertesti vengono classificati mediante *Word embedding*: secondo questo metodo, è valutato ogni bigramma rispetto a tutti quelli di un altro testo purché la distanza tra le due voci non superi un intervallo fisso, determinato dal numero di parole interposte tra quelle che compongono il bigramma d'interesse.[16] Il punteggio di somiglianza per ogni coppia di bigrammi è calcolato mediante il coseno di similitudine rispetto a ognuna delle quattro coppie di parole;[17] in tal modo la valutazione di un intertesto notato dai commentatori viene classificata rispetto a tutti gli altri bigrammi nel testo in questione, la cui dimensione è fissata al singolo libro di cui si compone il poema che è equivalente al testo su cui si basa il *database* sopra menzionato. Posti questi parametri, è operata una comparazione tra il risultato ottenuto mediante gli strumenti messi a punto dagli autori dell'articolo e quello prodotto attraverso *Tesserae* dal raffronto tra il Libro I di Valerio Flacco con i quattro *target text* precedentemente menzionati. Il giudizio al riguardo è espresso con queste parole:

> For the complete set of *Tesserae* results, the recall is 33.9%, and the precision is 0.97%; with $k = 250$, our method achieves a comparable precision (1.4%)

[14] È merito di (Sprugnoli, Passarotti, *et al.* 2020) aver posto l'attenzione sulle diverse tecniche di *PoS tagging* (*e.g.* Bacon 2020; Celano 2020; Stoeckel *et al.* 2020; Straka e Straková 2020).

[15] Gli intertesti sono raccolti sulla base dei tre più recenti commenti alle *Argonautiche* di Valerio Flacco (Spaltenstein 2002; Kleywegt 2005; Zissos 2008).

[16] La scelta dei bigrammi come unità di base è conforme alla norma poetica secondo cui le allusioni sono costituite da una coppia di vocaboli, pur non essendo escluse anche parole singole o frasi più lunghe: su questa considerazione si basa anche l'algoritmo di *Tesserae*.

[17] Dati per ipotesi i due bigrammi A B e A C, sono messe a confronto le quattro coppie A~A, A~B, A~C, B~C.

> but higher recall (82.4%). An important advantage of the *Tesserae* tool, however, is that it searches for similar phrases in parallel and does not require a list of specific queries as input. As such, the results aggregated for this comparison come from a much smaller number of *Tesserae* searches than the 945 embedding-based searches we run. For this reason, *Tesserae* is likely to be more suitable than our method for applications in which the user does not have predetermined phrases of interest (Burns *et al.* 2021, 4902–3).

L'ultima affermazione testimonia una grande affidabilità di *Tesserae* nella ricerca intertestuale, ma – come già si era evidenziato in precedenza – ripropone la problematica connessa alla condizione necessaria data dalla presenza di una coppia minima di lemmi su cui il *software* possa focalizzare la propria attenzione. Invece, il ricorso al *Word embedding* garantisce una soluzione che consenta l'identificazione di quei *loci similes* che non soddisfino tali condizioni così da integrare i metodi esistenti: i due esempi forniti dagli autori sulla base del *database* degli intertesti scovati dai commentatori per il Libro I delle *Argonautiche* mostrano come la tecnica allusiva non sia solo dipendente dalla presenza di vocaboli comuni ma anche da rapporti di sinonimia – talvolta di antonimia – che non potrebbero essere identificati da un'applicazione incapace di riconoscerli.[18] In tal senso, tale strumento è un utilissimo complemento a *Tesserae* o a *Diogenes* perché è in grado di risolvere la complessità dei richiami di cui è intrisa ogni forma di espressione letteraria.

Fra i più recenti studi intertestuali mediante l'impiego delle tecniche di *NLP* non può essere tralasciata la dissertazione di Nozomu Okuda intitolata *Deep Learning for Intertext Discovery in Latin Epic: Latin SBERT and the Lucan-Vergil Benchmark*, discussa nel giugno 2022 presso la State University of New York at Buffalo. In estrema sintesi, l'autore afferma che

> [f]our main questions guided the work of this dissertation:
> • How can we use deep learning to detect intertexts in Latin Epic?
> • How good is deep learning at detecting intertexts in Latin Epic?
> • Why would deep learning be able to detect intertexts in Latin Epic?
> • How might deep learning become better at detecting intertexts in Latin Epic? (Okuda 2022, 5)

A queste domande cercano di rispondere i quattro capitoli nei quali si articola la tesi dottorale.

Nel Capitolo 1 (*Points of Comparison*, pp. 28-63), dopo aver dato una definizione generale di intertestualità, l'attenzione dello studioso si concentra su un *case study* costituito dalle relazioni occorrenti tra il Libro I del *Bellum ciuile* di Lucano e il massimo poema virgiliano, muovendo dalla messe di dati raccolta nella citata *Lucan-Vergil Benchmark*, in cui sono rilevati 3.410 paralleli ricavati mediante *Tesserae* (cfr. *supra*). Scopo di questo primo capitolo è dare una valutazione qualitativa degli intertesti rilevati mediante questa procedura: essa è svolta attraverso una serie di

[18] «The phrases *e clausis* [*antris*] ("from the enclosed [cave]," *Arg.* 1.417) and *circum claustra* [*fremunt*] ("[they roar] around the gate," *Aen.* 1.56), for example, contain no words in common but have similar syntax (the prepositions *e* and *circum*) and semantics (words indicating enclosure). Similarly, the phrases *Phlegethontis operti* ("hidden Phlegethon," *Arg.* 1.735) and *Acherontis aperti* ("open Acheron," *Theb.* 11.150) both refer to rivers in the underworld and contain near-antonymic adjectives» (Burns *et al.* 2021, 4902–3).

test i cui dati sono accuratamente discussi nei §§ 1.2.2-1.5.2.[19] L'autore esprime il risultato di questa prima fase della ricerca con queste parole:

> These findings suggest that there is a relationship between a combination of *Tesserae* scores and human-judged intertext quality ratings, at least when considered in aggregate. Perhaps this relationship could be used to implement a software tool that, by sorting significant intertexts from insignificant ones, aids in intertext discovery. […] Given that humans presumably rely on words to judge the intertextual relation of two passages, it was surprising to find that models trained with word information from the parallel passages performed poorly. Among various reasons for poor model performance, the lack of word relationship and word frequency information in the word representation used in these experiments appear to be the most relevant deficiencies (Okuda 2022, 63).

La questione sollevata riguardo la relazione tra le parole e la loro frequenza è specificamente discussa nel Capitolo 2 (*Intertext Discovery with Transformers*, pp. 64-84): l'indagine intertestuale viene implementata mediante il ricorso ai trasformatori. Come già enunciato, il grande vantaggio di un'architettura fondata su *Word2Vec* riguarda il fatto che l'attenzione della macchina non si focalizza più solo sulla singola parola, bensì prende in considerazione anche il contesto all'interno del quale il vocabolo appare (cfr. Manjavacas *et al.* 2019; Burns *et al.* 2021). Un ulteriore vantaggio nella creazione e nel confronto di *Word embedding* è ottenuto mediante il ricorso a *BERT*: sviluppato in origine per la lingua inglese, una sua particolare versione è stata messa a punto da Bamman e Burns per operare nello specifico sul latino (Okuda 2022, 66–67; cfr. Bamman e Burns 2020; Huang *et al.* 2022); in §§ 2.2-2.3 l'autore illustra tutti gli esperimenti svolti su *Latin BERT*.[20] Il passo successivo è dato dall'elaborazione di *Latin SBERT* sul modello di *Sentence BERT*, in quanto

> [t]he *SBERT* paper showed that *BERT* performs poorly on sentence level tasks when comparing two meanpooled sentence vectors. Based on the *Latin BERT* results with the mean-pooled cosine similarity scheme just presented, it seems that *Latin BERT* inherits this weakness of *BERT* as well. […]
> *SBERT* takes two sentences as input and returns a similarity score as output. It does so by first obtaining the *BERT* vectors of each sentence separately. Then, it takes the mean of each sentence's vectors. Finally, the two mean vectors are compared, and a similarity score is computed. Adapting *SBERT* to *Latin BERT* yielded *Latin SBERT*. Using *Latin BERT* as the initial encoder model, *Latin SBERT* was trained on the *Lucan-Vergil benchmark* (Okuda 2022, 77).[21]

[19] I dati 'nudi' sono disponibili all'URL https://github.com/nOkuda/fusion-experiments.

[20] I codici che consentono di replicare gli esperimenti condotti sono disponibili all'URL https://github.com/nOkuda/bert-experiments.

[21] Lo «*SBERT* paper» cui si fa riferimento al principio della citazione è (Reimers e Gurevych 2019); si veda anche (Thakur *et al.* 2021).

La valutazione comparativa dei risultati precedentemente ottenuti attraverso *Latin BERT* e quelli ricavati da *Latin SBERT* porta a dire che il secondo funziona meglio nel riconoscimento degli intertesti rispetto al primo; comunque, in entrambi i casi, si ha la prova che le tecniche di *NLP* danno risultati superiori rispetto a quelli ottenuti con il solo *Tesserae*.

Per rispondere alla terza domanda, nel Capitolo 3 (*Literary Theory and Natural Language Processing*, pp. 85-101) sono anzitutto esposti i fondamenti teorici del concetto d'intertestualità. In seguito l'attenzione si sposta sul versante propriamente filologico-classico in considerazione del fatto che elementi strutturali della produzione letteraria sono, da un lato, l'impiego di una forma linguistica particolare e, dall'altro, quel rapporto di *imitatio* ed *aemulatio* che costantemente mette in relazione un autore con chi l'ha preceduto.[22] Un aiuto concreto nella determinazione degli intertesti deriva dunque dal ricorso a *Latin SBERT* in quanto esso è in grado di valutare il significato del singolo lemma senza trascurare il contesto cui esso appartiene, il quale svolge una funzione decisiva nel definire l'accezione che il singolo *token* assume con una precisione superiore rispetto al semplice *Latin BERT* che invece 'appiattisce' i lemmi, non essendo in grado di valorizzarne il contesto. Afferma Okuda:

> *Latin SBERT* embraces and eschews aspects of deconstruction. One way in which *Latin SBERT* embraces deconstruction is accounting for the multiplicity of meanings for every token. Whereas the earlier described word embedding approach to intertext discovery made the assumption that the vector representations of the input words would remain static no matter what context those words came from, *Latin SBERT* explicitly considers context of each input token by passing them through *Latin BERT*. Of course, the final vector representations used for each input passage are definitive stances on the single meaning of the two passages, which defies deconstruction in theory (Okuda 2022, 101).

Il Capitolo 4 (*Differences in Human and Computer Judgments*, pp. 102-135) si articola in due sezioni complementari. Nella prima (§§ 4.1-4.2) sono messi a confronto i principi in base ai quali l'essere umano formula i propri giudizi riguardo alle situazioni d'intertestualità comprese nella *Lucan-Vergil Benchmark* con quelli prodotti da *Latin SBERT*. Nella citata banca dati, gli intertesti hanno ricevuto una classificazione numerica espressa dalle unità comprese tra 1 (*not meaningful parallels*) e 5 (*meaningful parallels*) in base alle valutazioni date da quattro laureati frequentanti un corso seminariale professato da Neil Coffee, il lavoro dei quali è stato supervisionato da Sarah Jacobson. Okuda enuncia i principi di classificazione in questo modo:

> There were a variety of criteria to determine how high a rating a parallel would receive. The more criteria a parallel fulfilled, the higher the rating would be. The criteria Jacobson highlighted were: word order, morphology, content, distance, syntax, and context. These criteria were assessed against the matching words that *Tesserae* or commentators of Lucan noted. In the terms of the published criteria, it appears that "formal similarity" deals with word order, morphology, content, distance, and syntax. The "context"

[22] Per esempio, in Catull. 101, 1 è evidente il richiamo a Hom. *Od.* 1, 1-4 come stabiliscono le corrispondenze lessicali tra *multas gentes* e πολλῶν δ' ἀνθρώπων in riferimento alle molte popolazioni incontrate, tra *aequora* e πόντῳ e, parimenti, tra *vectus* e πλάγχθη come luogo paradigmatico del viaggio pericoloso (Conte 1974).

> referred to in the published criteria corresponds with the context Jacobson remembers (Okuda 2022, 104).

In particolare, quale sia la differenza tra gli intertesti di grado 5 rispetto a quelli di grado 4 può essere illustrato dal confronto di Lucan. 1, 8-9 rispettivamente con Verg. *Aen.* 11, 732-733 e 12, 500-503a.

All'inizio del poema di Lucano, la voce narrante si interroga circa le cause della guerra civile che sconvolse il territorio italico dopo che Giulio Cesare aveva varcato in armi il *pomerium* segnato dal fiume Rubicone, ponendosi nella condizione di *hostis publicus* (cfr. Roche 2009, 109–14).

> Lucan. 1, 8-9:
> *Quis furor, o ciues, quae tanta licentia ferri*
> *gentibus inuisis Latium praebere cruorem?*
>
> Quale follia, o cittadini, quale sfrenato abuso delle armi offrire il sangue latino alle genti nemiche?

Il modello virgiliano che restituisce il più grande grado d'intertestualità si rileva nel discorso che, nel corso del Libro XI, il capo etrusco Tarconte rivolge ai propri compagni d'armi (cfr. Horsfall 2003, 397–98):

> Verg. *Aen.* 11, 732-733:
> *quis metus, o numquam dolituri, o semper inertes*
> *Tyrrheni, quae tanta animis ignauia uenit?*
>
> Quale terrore vi prese, o voi che non vi lamentate mai di nulla, o sempre inerti Tirreni? Quale grande viltà invase gli animi?

Il punteggio 5 assegnato a questo parallelo si spiega sulla base dei seguenti criteri. Anzitutto, rispetto all'*ordo uerborum*, nel primo piede e mezzo di entrambi gli esametri l'interrogativo *quis* – impiegato come aggettivo – e la prima interiezione *o* occupano le stesse posizioni forti; da un'analisi ulteriore dei due passi si nota che la tesi del primo dattilo è costituita dai bisillabi *furor* in Lucano e *metus* in Virgilio, due voci isosillabiche che, pur non potendo considerarsi sinonimi, tuttavia afferiscono alla sfera del sentimento. E sebbene solo l'*Eneide* reiteri l'interiezione *o*, così conferendo un tono marcatamente patetico alla duplice invocazione che Tarconte rivolge ai *Tyrrheni* rimproverati per le continue esitazioni in battaglia, l'altro evidente elemento di ripresa è costituito dal sintagma interrogativo *quae tanta* che introduce gli astratti *licentia* e *ignauia*, voci isosillabiche collocate subito dopo la cesura eftemimere. Dissimile è, tuttavia, il contesto che provoca la duplice domanda: nel *Bellum ciuile* il narratore del poema si rivolge ai cittadini di Roma chiedendo le motivazioni delle guerre intestine tra Cesare e Pompeo, mentre nell'*Eneide* l'interrogazione è posta da Tarconte allo scopo di suscitare l'ardore nei propri commilitoni. Quindi, pur non essendo i due scenari completamente coordinati, la struttura generale presenta somiglianze tali da consentire l'attribuzione del punteggio 5 a questo parallelo.

Osservazioni non dissimili possono essere svolte per il seguente passo tratto dal Libro XII dell'*Eneide* (cfr. Tarrant 2012, 219–21).

> Verg. *Aen.* 12, 500-503a:
> *Quis mihi nunc tot acerba deus, quis carmine caedes*
> *diuersas obitumque ducum, quos aequore toto*

> *inque uicem nunc Turnus agit, nunc Troius heros,*
> *expediat?*
>
> Ora quale nume mai mi esporrà tante sventure, chi in poesia [narrerà] stragi crudeli e uccisioni di capi, quanti per tutta la pianura ora incalza Turno ora l'eroe Troiano?

In termini di *ordo uerborum*, come in Lucano *quis* viene prima di *quae*, anche in Virgilio l'interrogativo precede il relativo *quos*; tuttavia, nell'*Eneide* è ripetuto *quis* anche nella seconda parte del v. 500, subito dopo la cesura eftemimere, a conferire un marcato *pathos* alle parole del narratore che si chiede quale divinità ispiratrice della poesia potrà dargli la forza di narrare il duello campale tra Enea e Turno.[23] È però sul piano morfo-sintattico dove si avverte una maggiore divaricazione tra i due passaggi: se infatti *quis* è presente in entrambi, *quae* e *quos* svolgono funzioni assai differenti, interrogativo il primo e relativo il secondo. Circa i contenuti, il brano del *Bellum ciuile* allude con *cruorem* al sangue sparso durante le lotte fratricide tra cesariani e pompeiani, mentre l'*Eneide* accenna alla morte di tanti valorosi condottieri con la dittologia *caedes* e *obitum*. Infine, l'ultimo aspetto degno di nota è la maggiore ampiezza del testo virgiliano: in termini di distanza sono interposte solo tre parole tra *quis* e *quae* nel brano lucaneo, mentre se ne contano cinque tra il secondo *quis* e il relativo *quos* nel massimo poema dell'autore mantovano. Tali differenze spiegano perché i due passaggi siano certo in relazione tra loro ma ricevano un punteggio inferiore di un'unità rispetto a quello precedentemente esaminato.

Quello che emerge da questi esempi d'intertestualità e che può essere replicato per tutti i materiali della *Lucan-Vergil Benchmark* è la soggettività del giudizio:

> The frank wording of the article, the data collection method Jacobson remembers, and the fact that Jacobson had ultimate say in the rating that was recorded in the Lucan-Vergil benchmark all show how important human subjectivity was to the process of creating the benchmark. Further evidence of human subjectivity is in the class discussions about how to distinguish parallels of one rating from parallels of another rating. This indicates that raters had disagreements about how to rate a parallel. And while Jacobson had the ultimate say in what rating now appears in the Lucan-Vergil benchmark, she shared that she would rate some parallels differently now than she did in the past. This shows that even for the same rater, a parallel may be rated differently at a different time (Okuda 2022, 106).

Al di là di queste considerazioni, è sempre possibile spiegare le valutazioni che, di volta in volta, vengono date alle situazioni d'intertestualità contenute nel database citato. Ben più complesso è, invece, comprendere in che modo *Latin SBERT* processi il materiale linguistico: se pure conosciamo *input* e *output*, non è possibile dire quasi nulla di tutta la fase compresa tra ingresso e uscita, onde la definizione di *BERT* come «black-box model» (cfr. Aken *et al.* 2020); tuttavia è possibile ricercare un parallelo tra i criteri seguiti da Sarah Jacobson per il materiale della *Lucan-Vergil Benchmark* e i risultati generati dal *Transformer*. Anzitutto bisogna considerare che, in

[23] Ettore Paratore considera *quis mihi* un «singolare modo d'introdurre una nuova invocazione relativa a un episodio, quale quelle a IX 525 e a X 163, come ricorda il Forbiger. Ma ora il poeta non si rivolge più alle Muse o ad Apollo, quasi che la strage fatta da Enea sia così eccezionale che gli dei del canto non siano più in grado d'ispirargli la forza di descriverla» (Paratore 1983, 239).

generale, le tecniche di *NLP* non sono in grado di elaborare considerazioni relative alle parole co-occorrenti, in particolare circa l'*ordo uerborum*, la distanza e la sintassi; tanto meno è contemplato il contesto letterario. Inoltre, fatto che non deve mai essere trascurato, il nostro giudizio è sempre in qualche senso mediato da una tradizione di studi che, per gli autori classici, ha raggiunto una stratificazione molto complessa: almeno in linea teorica, un odierno *Large Language Model* ha nelle proprie disposizioni una tale quantità di dati derivanti anche dalla letteratura secondaria che, attraverso opportune tecniche di *refinement* o di allineamento, dovrebbe essere in grado di formulare valutazioni complesse, testualmente motivate e riconducibili a differenti tradizioni di studio (cfr. Ghizzota *et al.* 2025). Pertanto, rispetto alle più 'primitive' procedure di *NLP*, grazie allo sviluppo di un meccanismo di attenzione, i *Transformer* costruiscono le loro rappresentazioni vettoriali anche sulla base di quelle circostanti: questo fatto, che può essere visto come un indiretto mezzo di conoscenza del contesto letterario, è l'elemento che più avvicina il criterio di giudizio del *computer* all'intelligenza umana.

La seconda sezione del Capitolo 4 (§ 4.3) presenta tre diversi *case studies* volti a comprendere meriti e demeriti di *Latin SBERT*: in essi si confrontano tre brani del Libro I del *Bellum ciuile* (§ 4.3.2 *Catalog of Tribes*, per Lucan. 1, 420-437; § 4.3.3 *Oak Tree Simile*, per Lucan. 1, 135-143; § 4.3.4 *Civil War Devastation*, per Lucan. 1, 24-32) con una serie di presunti intertesti virgiliani.

In estrema sintesi si può affermare che

> [t]he key findings in this analysis were that (1) compared to Jacobson's criteria, *Latin SBERT* is not as sensitive to word order; (2) compared to *Latin SBERT*, Jacobson's criteria are limited in the kinds of similarities they can find; and (3) Jacobson's criteria are driven by human subjectivity, while *Latin SBERT* operates with the mystery of complexity (Okuda 2022, 134).

Dunque, confrontando le valutazioni di *Latin SBERT* alla luce della logica sottesa alla *Lucan-Vergil Benchmark*, il *deep learning* sembra rilevare intertesti che la valutazione umana non sarebbe portata a considerare: in particolare, va tenuto presente quel fenomeno che Okuda ha chiamato *doppelgänger entries*, ossia

> the phenomenon where multiple entries in the Lucan-Vergil benchmark are indistinguishable to *Latin SBERT* in terms of their passages. It is caused primarily by the difference in passage extraction techniques between the benchmark and *Latin SBERT*. The benchmark used an assortment of passage extraction techniques on the *Bellum Civile* and the *Aeneid*, including the sparse quotation style of commentators as well as the text breaking algorithm of the version of *Tesserae* available at the time of the benchmark's creation. *Latin SBERT*, in contrast, uses the sentence splitting algorithm available in the Classical Languages Toolkit (CLTK) library. As a result, multiple shorter passages listed in the benchmark may be part of the same longer passage that *Latin SBERT* sees. If at least two of the shorter passages share an associated passage, then *Latin SBERT* will see two entries with the same pair of passages (Okuda 2022, 132).

Questo potrebbe avere effetti positivi o negativi sulle prestazioni di *Latin SBERT* nei casi in cui i paralleli non si trovino nello stesso blocco di convalida incrociata, com'è chiaramente esposto nella discussione delle tipologie intertestuali relative ai brani lucanei riferiti ai §§ 4.3.2-4.3.4. Per esempio, al principio del *Bellum ciuile*, Pompeo è paragonato a un enorme albero di quercia,

decorato con molti oggetti, ancora ben solidamente ancorato al terreno nonostante le sue radici abbiano perso parte del proprio antico vigore (Lucan. 1, 136-143a; cfr. Roche 2009, 185–89):

> *qualis frugifero quercus sublimis in agro*
> *exuuias ueteris populi sacrataque gestans*
> *dona ducum nec iam ualidis radicibus haerens*
> *pondere fixa suo est, nudosque per aera ramos*
> *effundens trunco, non frondibus, efficit umbram,*
> *et quamuis primo nutet casura sub Euro,*
> *tot circum siluae firmo se robore tollant,*
> *sola tamen colitur.*
>
> Quale in un fertile campo una quercia sublime, che regge le spoglie antiche di un popolo, e i consacrati doni dei duci; né già aderendo alla terra con valide radici vi resta fermamente unita per il peso suo, e, i nudi rami stendendo all'aria, col tronco, non con il fogliame dà ombra: e se pure, al primo soffio dell'Euro, ondeggi lì lì per cadere, e tanti alberi intorno si levino con saldo tronco, essa sola tuttavia è venerata.

Pur non mancando elementi contrastivi, questa similitudine trova un significativo parallelo nel paragone tra Enea e la solidità di una quercia sferzata dai venti a causa della propria risolutezza nel decidere la partenza da Cartagine nonostante l'insistenza di Didone (Verg. *Aen.* 4, 441-449; cfr. Fratantuono e Smith 2022, 643–54):

> *ac uelut annoso ualidam cum robore quercum*
> *Alpini Boreae nunc hinc nunc flatibus illinc*
> *eruere inter se certant; it stridor, et altae*
> *consternunt terram concusso stipite frondes;*
> *ipsa haeret scopulis et quantum uertice ad auras*
> *aetherias, tantum radice in Tartara tendit:*
> *haud secus adsiduis hinc atque hinc uocibus heros*
> *tunditur, et magno persentit pectore curas;*
> *mens immota manet, lacrimae uoluuntur inanes.*
>
> E come le tramontane alpine lottano fra loro, con raffiche ora qua ora là, per abbattere una robusta quercia dal fusto annoso; va uno stridore, e le alte fronde cospargono la terra intorno allo scosso tronco; quella s'abbarbica alle rocce, e quanto con la cima nell'aria del cielo, tanto con la radice si protende verso il Tartaro: non diversamente l'eroe è battuto qua e là da assidue voci, e sente nel grande cuore la pena; l'animo resta imperturbabile, vane scorrono le lacrime.

Nonostante, applicando in modo molto rigoroso i criteri esposti da Sarah Jacobson, la valutazione di 5 data nella *Lucan-Virgil Benchmark* possa risultare quasi eccessiva, è evidente la vicinanza tematica tra le due situazioni. Una riprova della bontà di questa valutazione è confermata anche da *Latin SBERT* il quale attribuisce un punteggio di 0,985 all'intertesto con una discrepanza di solo 0,015 rispetto al massimo possibile, corrispondente con l'unità intera: questo fatto testimonia come il *Transformer* abbia la possibilità di riconoscere una somiglianza tematica anche al di là di una stretta relazione sul piano dell'*ordo uerborum*, della vicinanza morfologica, dell'ampiezza del brano e della sua struttura sintattica. È ben evidente come in

questo caso la linguistica computazionale possa dare un'ulteriore conferma di come Lucano si sia esplicitamente rifatto al precedente virgiliano.

Curiosamente, Okuda sembra non conoscere il contributo di (Scheirer *et al.* 2016) nel quale si propone un calzante confronto anche con Verg. *Aen.* 2, 626-631 (cfr. Horsfall 2008, 448–52):

> *ac ueluti summis antiquam in montibus ornum*
> *cum ferro accisam crebrisque bipennibus instant*
> *eruere agricolae certatim, illa usque minatur*
> *et tremefacta comam concusso uertice nutat,*
> *uulneribus donec paulatim euicta supremum*
> *congemuit traxitque iugis auulsa ruinam.*
>
> Come sulle vette dei monti i boscaioli gareggiano ad abbattere con frequenti colpi di scure un antico frassino intaccato dal ferro; quello continuamente minaccia, e scossa la cima, oscilla tremante la chioma, finché vinto a poco a poco dai colpi emette un ultimo gemito, e schiantato rovina sui gioghi.

Ivi si paragona la città di Troia, ormai condannata alla rovina e devastata dai Greci, a un frassino moribondo che viene abbattuto da contadini laboriosi. Così affermano gli autori:

> We see a resemblance of theme: both texts describe a tottering old tree. Both passages also share a narrative function. In each case the tottering tree foreshadows the downfall of a hitherto stalwart bastion: the Trojan citadel in the *Aeneid*, the republican general Pompey in the *Civil War*. Indeed, the two events appear intertextually connected: the capture and destruction of mythological Troy anticipates the historical defeat and death of the Roman leader (Scheirer *et al.* 2016, 205).

Sebbene una sola voce comune – il verbo *nuto* – occorra sia nel citato passo lucaneo (1, 141) sia nell'ultimo *locus* virgiliano (*Aen.* 2, 629), tuttavia la somiglianza tra i due brani appare evidente anche solo a un lettore superficiale: il concetto di *ruina*, vocabolo significativamente posto a chiusura del v. 631, indirizza con chiarezza il fruitore del testo poetico che dalla lettura del *Bellum ciuile* potrà riconoscere nella caduta di Troia e nella morte di Priamo prolessi della morte di Pompeo, autentico 'eroe tragico' di Lucano.[24]

Un approccio diverso ma di grande interesse è costituito da *Musisque Deoque. Un archivio digitale di poesia latina, dalle origini al Rinascimento italiano* (https://www.mqdq.it/), sviluppato presso l'Università Ca' Foscari di Venezia sotto la direzione di Paolo Mastandrea a partire dall'anno 2005: esso consente di svolgere ricerche inerenti il lessico, la posizione delle parole nel verso, il rapporto tra parole e metrica, le varianti presenti nella tradizione manoscritta (cfr. Boschetti *et al.* 2021; Venuti *et al.* 2023; Venuti 2025). Come si può leggere sulla *home page* del progetto,

> [a]d oggi, le principali collezioni di classici sono state trasferite su supporto digitale ed esistono risorse, per lo più *online*, che rendono assai celere ogni ricerca lessicale. Nella grandissima parte dei casi, tuttavia, il motore di

[24] Al riguardo si vedano anche (Hinds 1998, 8–10) e, con maggior dettaglio, (Horsfall 2008, 421–23) sul rapporto tra Verg. *Aen.* 2, 557-558 e Lucan. 1, 685-686, accomunati dall'espressione *iacet truncus*; sui legami intercorrenti tra i libri II e IV dell'*Eneide* resta importante il contributo di (Fenik 1959).

> ricerca si limita a fornire semplici occorrenze lessicali all'interno di un testo fisso, 'autoritario'. *Musisque Deoque* si è proposto di superare questa limitazione, permettendo di reperire non solo le forme scelte e riportate dall'edizione di riferimento, ma anche le varianti presentate in apparato e selezionate sotto la responsabilità di un 'editore digitale'.

Dal 2019 questo nucleo iniziale si è espanso a formare quella che gli editori hanno chiamato *MQDQ Galaxy* (https://pric.unive.it/projects/mqdq-galaxy/home), un complesso del quale fanno parte i seguenti *corpora* testuali.

I *Carmina Latina Epigraphica* (https://www.mqdq.it/ce/presentazione) raccolgono i testi contenuti nell'edizione *Teuberniana* curata da Franz Bücheler ed Ernst Lommatzsch (1930) oltre a ulteriori 1235 iscrizioni rinvenute in epoca successiva: grande merito di questo progetto è l'aver unito in un solo spazio di facile consultazione e interrogazione una grande messe di testi precedentemente editi nei repertori più vari.

In *Musa Medievalis* (https://www.musamedievalis.it/) è contenuta un'ampia selezione di testi di poesia latina medievale la cui datazione è compresa tra la seconda metà del VII secolo e la prima del XIII. Ideale prosecuzione di tale progetto è *Poeti d'Italia in lingua latina tra Medioevo e Rinascimento* (https://www.poetiditalia.it/) che copre l'arco cronologico che va dalla nascita di Dante Alighieri alla metà del Cinquecento. Alla pagina indicata si legge che

> [t]ali testi possono essere indagati grazie a un motore di ricerca specializzato, che nel 2023 ha ampliato le sue potenzialità, permettendo un'indagine integrata ed estesa anche al *corpus* della poesia latina antica di *Musisque Deoque*. Nel 2025 questa possibilità è stata estesa anche ai testi di *Musa Medievalis*, oggi accessibile online e facilmente ricercabile. Viene con ciò ad estendersi grandemente il campo aperto a un'indagine intertestuale celere e sicura che spazia lungo diciotto secoli di produzione poetica, dalle remote origini della lingua letteraria di Roma sino al fiorente Umanesimo e Rinascimento italiano.

Della 'galassia' fa parte anche il prezioso strumento *PedeCerto. Metrica latina digitale* (https://www.pedecerto.eu/): messo a punto dall'Università di Udine nell'ambito del progetto coordinato da Emanuela Colombi "FIRB – Traditio Patrum" sulla tradizione testuale degli autori ecclesiastici latini, la sua applicazione a *MQDQ* ha consentito la scansione degli oltre 250.000 versi dattilici ivi contenuti.

Ancora a livello di prototipo – l'ultimo aggiornamento è datato 26 marzo 2025 – è invece *Hellenica. Un archivio digitale di poesia greca* (https://open.unive.it/hellenica/) che si propone di estendere i principi di *MQDQ* anche alla produzione in lingua ellenica.

Come si sarà inteso, *MQDQ* non è semplicemente una banca dati specializzata nella poesia latina (cfr. Milanese 2020, 108–16), piuttosto uno strumento utile anche nell'indagine intertestuale: in un recente saggio di Luca Mondin volto a ricercare le occorrenze di Marziale nella poesia cristiana, «[l]'indagine è stata condotta mediante l'uso della funzione *Co-occorrenze lessicali* della piattaforma *Musisque Deoque* […] ricontrollando i risultati mediante lo strumento *Ricerca* della stessa piattaforma e nelle banche dati testuali di *Brepolis* (http://www.brepolis.net/) mediante il *Cross Database Searchtool*» (Mondin 2024, 112). Secondo un criterio non dissimile da quello visto in (Coffee *et al.* 2012) o in (Okuda 2022), il lavoro dello studioso veneziano è completato da un *dossier* di oltre 360 riscontri testuali classificati secondo diversi gradi di probabilità.

4. Conclusioni e prospettive

In queste pagine è stato tracciato un percorso di carattere sia storico sia teorico al fine d'inquadrare l'evoluzione delle tecniche di *NLP*, con particolare riferimento alla ricerca delle relazioni intertestuali nella letteratura latina e, soprattutto, nell'epica di I secolo d.C.

In generale, mi sembra possibile affermare che, al livello attuale dello sviluppo, il ruolo dei *Transformer* nell'analisi letteraria sia ampiamente promettente: (Okuda 2022) – insieme ai molti altri presenti nell'ampia bibliografia compulsata prima di stendere queste pagine – dimostra una certa convergenza tra le forme di analisi 'tradizionale' e quelle di natura linguistico-computazionale. Non è escluso che, a dispetto dell'ampia tradizione di studi classici, il ricorso al *NLP* possa portare alla luce nuove forme di 'corrispondenza situazionale' sinora sfuggite: con questa definizione non si vuole porre tanto l'accento sulla co-occorrenza di singoli lemmi o di espressioni formulari, su cui la filologia sin dalle proprie origini ha sempre indagato, piuttosto rendere conto di strutture assai più ampie e generali, quei *topic* i quali forniscono anche indicazioni di tipo verbale per affinare sempre meglio la ricerca. Un recente e documentato esempio di come una ricerca di questo genere possa arricchire la critica testuale è costituito dalla menzionata monografia di (Cannizzaro 2023).

Alla luce della bibliografia esaminata, credo che un fertile ambito di applicazione delle tecniche di *NLP* possa essere indirizzato alla ricerca interlinguistica: una delle caratteristiche comuni a tutti i metodi di cui si è reso conto in questa rassegna è il fatto che lo strumento informatico non abbia una 'vera' conoscenza della lingua in cui sono scritti i testi destinati all'analisi, traducendoli in rappresentazioni di natura vettoriale. Come la filologia ha preso in esame i rapporti intercorrenti non solo all'interno della produzione latina ma si è sempre interrogata per un verso sull'influsso dei modelli greci nella letteratura di Roma e per un altro sulla ripresa delle opere classiche nelle letterature moderne, così le reti neurali potrebbero essere addestrate al riconoscimento di intertesti tra lingue diverse, implementando una nuova forma di ricerca volta a individuare inesplorati rapporti tra autori con cui mirare a un arricchimento sempre maggiore delle capacità interpretative nei confronti della produzione letteraria antica e moderna.

Bibliografia

Aken, Betty van, Benjamin Winter, Alexander Löser, e Felix A. Gers. 2020. «VisBERT: Hidden-State Visualizations for Transformers». *Companion Proceedings of the Web Conference 2020* (New York), 207–11. https://doi.org/10.1145/3366424.3383542.

Alfaro, María Jesús Martínez. 1996. «Intertextuality: origins and development of the concept». *Atlantis* 18 (1/2): 268–85. https://www.jstor.org/stable/41054827.

Allen, Graham. 2021. *Intertextuality*. 3a ed. Routledge.

Almeida, Felipe, e Geraldo Xexéo. 2019. «Word Embeddings: A Survey». https://doi.org/10.48550/arXiv.1901.09069.

Bacon, Geoff. 2020. «Data-driven Choices in Neural Part-of-Speech Tagging for Latin». In *Proceedings of LT4HALA 2020 – 1st Workshop on Language Technologies for Historical and Ancient Languages*, a cura di Rachele Sprugnoli e Marco Passarotti. European Language Resources Association (ELRA). https://aclanthology.org/2020.lt4hala-1.17.

Bamman, David. 2012. «11K Latin Texts». https://www.cs.cmu.edu/~dbamman/latin.html.

Bamman, David. 2014. «Intertextuality Beyond Words». The Tesserae Project, febbraio 17. https://tesserae.caset.buffalo.edu/blog/intertextuality-beyond-words-by-david-bamman/.

Bamman, David, e Patrick J. Burns. 2020. «Latin BERT: A Contextual Language Model for Classical Philology». https://doi.org/10.48550/arXiv.2009.10053.

Bamman, David, e Gregory R. Crane. 2008. «The Logic and Discovery of Textual Allusion». *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*. https://www.perseus.tufts.edu/~ababeu/latech2008.pdf.

Bernstein, Neil, Kyle Gervais, e Wei Lin. 2015. «Comparative rates of text reuse in classical Latin hexameter poetry». *Digital Humanities Quarterly* 2015 (9): 3. https://doi.org/10.63744/f2uy9vfafc5e.

Bjerva, Johannes, e Raf Praet. 2015. «Word Embeddings Pointing the Way for Late Antiquity». *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)* (Beijing), 53–57. https://doi.org/10.18653/v1/W15-3708.

Blei, David M. 2012. «Probabilistic topic models». *Commun. ACM* 55 (4): 77–84. https://doi.org/10.1145/2133806.2133826.

Blevins, Cameron. 2010. «Topic Modeling Martha Ballard's Diary». aprile 1. https://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/.

Boschetti, Federico, Angelo Mario Del Grosso, e Linda Spinazzè. 2021. «La galassia Musisque Deoque: storia e prospettive». In *Paulo maiora canamus. Raccolta di studi per Paolo Mastandrea*, a cura di Massimo Manca e Martina Venuti. Edizioni Ca' Foscari. http://doi.org/10.30687/978-88-6969-557-5/026.

Burns, Patrick J. 2017. «Measuring and Mapping Intergeneric Allusion in Latin Poetry using Tesserae». *Journal of Data Mining & Digital Humanities* Numéro spécial sur le traitement assisté par ordinateur de l'intertextualité dans les langues anciennes. https://doi.org/10.46298/jdmdh.3821.

Burns, Patrick J., James A. Brofos, Kyle Li, Pramit Chaudhuri, e Joseph P. Dexter. 2021. «Profiling of Intertextuality in Latin Literature Using Word Embeddings». In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, a cura di Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, *et al.* Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.389.

Cannizzaro, Francesco. 2023. *Sulle orme dell'Iliade. Riflessi dell'eroismo omerico nell'epica d'età flavia*. Studi e ricerche del Dipartimento di Lettere e Filosofia – Antichità e Filologia. Società Editrice Fiorentina. https://www.sefeditrice.it/catalogo/sulle-orme-dellemiliade-em/14607.

Celano, Giuseppe G. A. 2020. «A Gradient Boosting-Seq2Seq System for Latin POS Tagging and Lemmatization». In *Proceedings of LT4HALA 2020 – 1st Workshop on Language Technologies for Historical and Ancient Languages*, a cura di Rachele Sprugnoli e Marco Passarotti. European Language Resources Association (ELRA). https://aclanthology.org/2020.lt4hala-1.19.

Celano, Giuseppe G. A., Gregory R. Crane, e Bridget Almas. 2023. «The Ancient Greek and Latin Dependency Treebank». luglio 21. https://perseusdl.github.io/treebank_data/.

Coffee, Neil, Jean-Pierre Koenig, Shakthi Poornima, Christopher W. Forstall, Roelant Ossewaarde, e Sarah L. Jacobson. 2013. «The Tesserae Project: intertextual analysis of Latin poetry». *Literary and Linguistic Computing* 28 (2): 221–28. https://doi.org/10.1093/llc/fqs033.

Coffee, Neil, Jean-Pierre Koenig, Shakthi Poornima, Roelant Ossewaarde, Christopher W. Forstall, e Sarah L. Jacobson. 2012. «Intertextuality in the Digital Age». *Transactions of the American Philological Association* 142 (2): 383–422. https://doi.org/10.1353/apa.2012.0010.

Conte, Gian Biagio. 1974. *Memoria dei poeti e sistema letterario: Catullo, Virgilio, Ovidio, Lucano*. Ricerca letteraria 23. Einaudi.

Conte, Gian Biagio. 1986. *The Rhetoric of Imitation: Genre and Poetic Memory in Virgil and Other Latin Poets*. Tradotto da Charles Segal. Cornell Studies in Classical Philology 44. Cornell University Press.

Devlin, Jacob, e Ming-Wei Chang. 2018. «Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing». Google Research, novembre 2. https://blog.research.google/2018/11/open-sourcing-bert-state-of-art-pre.html.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, e Kristina Toutanova. 2019. «BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding». In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, a cura di Jill Burstein, Christy Doran, e Thamar Solorio, vol. 1. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423.

Dexter, Joseph P., Pramit Chaudhuri, Patrick J. Burns, *et al.* 2023. «A Database of Intertexts in Valerius Flaccus' Argonautica 1: A Benchmarking Resource for the Evaluation of Computational Intertextual Search of Latin Corpora». Harvard Dataverse. https://doi.org/10.7910/DVN/S6RD4M.

Farrell, Joseph. 2005. «Intention and Intertext». *Phoenix* 59 (1/2): 98–111. http://www.jstor.org/stable/25067741.

Fenik, Bernard. 1959. «Parallelism of Theme and Imagery in Aeneid II and IV». *The American Journal of Philology* 80 (1): 1–24. https://doi.org/10.2307/291884.

Firth, John Rupert. 1957. «A Synopsis of Linguistic Theory 1930-1955». In *Studies in linguistic analysis. Special volume of the Philological Society*. Basil Blackwell.

Forstall, Christopher W., Sarah L. Jacobson, e Walter J. Scheirer. 2011. «Evidence of intertextuality: investigating Paul the Deacon's Angustae Vitae». *Literary and Linguistic Computing* 26 (3): 285–96. https://doi.org/10.1093/llc/fqr029.

Fratantuono, Lee M., e R. Alden Smith, a c. di. 2022. *Virgil, Aeneid 4. Text, Translation, and Commentary*. Mnemosyne Supplements 462. Brill.

Ghizzota, Eleonora, Pierpaolo Basile, Lucia Siciliani, e Giovanni Semeraro. 2025. «The Meaning of Beatus: Disambiguating Latin with Contemporary AI Models». In *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, a cura di Cristina Bosco, Elisabetta Jezek, Marco Polignano, e Manuela Sanguinetti. CEUR Workshop Proceedings. https://aclanthology.org/2025.clicit-1.46/.

Grave, Edouard, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, e Tomas Mikolov. 2018. «Learning Word Vectors for 157 Languages». In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, a cura di Nicoletta Calzolari,

Khalid Choukri, Christopher Cieri, *et al.* European Language Resources Association (ELRA). https://aclanthology.org/L18-1550.

Greimas, Algirdas Julien. 1966. *Sémantique structurale. Recherche de méthode*. Librairie Larrousse.

Gross, Daniel. 2013. *Plenus litteris Lucanus: zur Rezeption der horazischen Oden und Epoden in Lucans Bellum Civile*. Litora classica 3. Verlag Marie Leidorf.

Günther, Fritz, Luca Rinaldi, e Marco Marelli. 2019. «Vector-Space Models of Semantic Representation From a Cognitive Perspective: A Discussion of Common Misconceptions». *Perspectives on psychological science* 14 (6): 1006–33. https://doi.org/10.1177/1745691619861372.

Helbig, Jörg. 1996. *Intertextualität und Markierung: Untersuchungen zur Systematik und Funktion der Signalisierung von Intertextualität*. Anglistische Forschungen 3. Winter.

Hinds, Stephen. 1998. *Allusion and Intertext: The Dynamics of Appropriation in Roman Poetry*. Cambridge University Press.

Horsfall, Nicholas, a c. di. 2003. *Virgil, Aeneid 11: a commentary*. Mnemosyne Supplements 244. Brill.

Horsfall, Nicholas, a c. di. 2008. *Virgil, Aeneid 2. Commentary*. Mnemosyne Supplements 299. Brill.

Huang, Austin, Suraj Subramanian, Jonathan Sum, Khalid Almubarak, e Stella Biderman. 2022. «The Annotated Transformer». https://nlp.seas.harvard.edu/annotated-transformer/.

Irwin, William. 2004. «Against intertextuality». *Philosophy and Literature* 28 (2): 227–42. https://doi.org/10.1353/phl.2004.0030.

Kleywegt, A. J., a c. di. 2005. *Valerius Flaccus, Argonautica, Book I: A Commentary*. Mnemosyne. Bibliotheca Classica Batava 262. Brill.

Kristeva, Julia. 1969. *Σημειωτική. Recherches pour une sémanalyse*. Éditions du Seuil.

Lenci, Alessandro. 2008. «Distributional semantics in linguistic and cognitive research». *Italian Journal of Linguistics* 20 (1). https://www.italian-journal-linguistics.com/app/uploads/-2021/05/1_Lenci.pdf.

Lendvai, Piroska, e Claudia Wick. 2022. «Finetuning Latin BERT for Word Sense Disambiguation on the Thesaurus Linguae Latinae». In *Proceedings of the Workshop on Cognitive Aspects of the Lexicon*, a cura di Michael Zock, Emmanuele Chersoni, Yu-Yin Hsu, e Enrico Santus. Association for Computational Linguistics. https://aclanthology.org/2022.cogalex-1.5.

Manjavacas, Enrique, Brian Long, e Mike Kestemont. 2019. «On the Feasibility of Automated Detection of Allusive Text Reuse». In *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, a cura di Beatrice Alex, Stefania Degaetano-Ortlieb, Anna Kazantseva, Nils Reiter, e Stan Szpakowicz. Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-2514.

Manning, Christopher D., Prabhakar Raghavan, e Hinrich Schütze. 2009. *An Introduction to Information Retrieval*. Cambridge University Press. https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf.

Martin, Elaine. 2011. «Intertextuality: An Introduction». *The Comparatist* 35: 148–51. https://doi.org/10.1353/com.2011.0001.

McDonald, Ryan, Fernando Pereira, Kiril Ribarov, e Jan Hajič. 2005. «Non-Projective Dependency Parsing using Spanning Tree Algorithms». In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, a cura di Raymond Mooney, Chris Brew, Lee-Feng Chien, e Katrin Kirchhoff. Association for Computational Linguistics. https://aclanthology.org/H05-1066.

Milanese, Guido Fabrizio. 2020. *Filologia, letteratura, computer. Idee e strumenti per l'informatica umanistica*. Vita & Pensiero.

Mondin, Luca. 2024. «Un classico inaspettato? Marziale nella poesia cristiana». *Poetica spolia. Il reimpiego del testo dei poeti nei generi letterari della tarda latinità. Atti del convegno di Napoli, sede centrale dell'Università Federico II, Aula Leone / Aula Guarino, 27-28 ottobre 2022* (Trieste), 109–220. https://doi.org/10.13137/978-88-5511-523-0/36090.

Morrelli, Massimiliano, Giacomo Pansini, Massimiliano Polito, e Arturo Vitale. 2020. «Similarità per la ricerca del dominio di una frase». https://doi.org/10.48550/arXiv.2002.00757.

Okuda, Nozomu. 2022. «Deep Learning for Intertext Discovery in Latin Epic: Latin SBERT and the Lucan-Vergil Benchmark». Tesi di dottorato, State University of New York at Buffalo. https://www.proquest.com/openview/6ad7e7b5f52cf5fec9cfed450ce36720/1?cbl=18750&diss=y&pq-origsite=gscholar.

Okuda, Nozomu, Jeffery Kinnison, Patrick J. Burns, Neil Coffee, e Walter J. Scheirer. 2022. «Tesserae Intertext Service». *Digital Humanity Quarterly* 16 (1). https://doi.org/10.63744/px6e7btus63x.

Paratore, Ettore, a c. di. 1983. *Virgilio. Eneide – Volume VI (Libri XI-XII)*. Tradotto da Luca Canali. Fondazione Lorenzo Valla – Arnoldo Mondadori Editore.

Rastier, François. 1987. *Sémantique interprétative*. Formes sémiotiques. Presses universitaires de France.

Reimers, Nils, e Iryna Gurevych. 2019. «Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks». *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (Hong Kong), 3982–92. https://doi.org/10.48550/arXiv.1908.10084.

Roche, Paul. 2009. *Lucan, De Bello Ciuili, Book I*. Oxford University Press. https://doi.org/10.1093/actrade/9780199556991.book.1.

Russell, Stuart, e Peter Norvig. 2020. *Artificial Intelligence: A Modern Approach*. 4a ed. Pearson. http://aima.cs.berkeley.edu/index.html.

Sahlgren, Magnus. 2008. «The distributional hypothesis». *Italian Journal of Linguistics* 20 (1): 33–53.

Scheirer, Walter, Christopher W. Forstall, e Neil Coffee. 2016. «The sense of a connection: Automatic tracing of intertextuality by meaning». *Digital Scholarship in the Humanities* 31 (1): 204–17. https://doi.org/10.1093/llc/fqu058.

Sommerschield, Thea, Yannis Assael, John Pavlopoulos, *et al.* 2023. «Machine Learning for Ancient Languages: A Survey». *Computational Linguistics* 49 (3): 703–47. https://doi.org/10.1162/coli_a_00481.

Spaltenstein, François. 2002. *Commentaire des Argonautica de Valérius Flaccus (livres 1 et 2)*. Collection Latomus. Latomus.

Spinelli, Tommaso. 2023. «The First Online Dictionary of Latin Near-Synonyms». luglio 7. https://doi.org/10.17630/3cf644e6-86b8-44d0-a50a-b33c7ca86072.

Sprugnoli, Rachele, Giovanni Moretti, e Marco Passarotti. 2020. «Latin embeddings». settembre 10. https://doi.org/10.5281/zenodo.4022817.

Sprugnoli, Rachele, Marco Passarotti, Flavio Massimiliano Cecchini, e Matteo Pellegrini. 2020. «Overview of the EvaLatin 2020 Evaluation Campaign». In *Proceedings of LT4HALA 2020 – 1st Workshop on Language Technologies for Historical and Ancient Languages*, a cura di Rachele Sprugnoli e Marco Passarotti. European Language Resources Association (ELRA). https://aclanthology.org/2020.lt4hala-1.16.

Stoeckel, Manuel, Alexander Henlein, Wahed Hemati, e Alexander Mehler. 2020. «Voting for POS tagging of Latin texts: Using the flair of FLAIR to better Ensemble Classifiers by Example of Latin». In *Proceedings of LT4HALA 2020 – 1st Workshop on Language Technologies for Historical and Ancient Languages*, a cura di Rachele Sprugnoli e Marco Passarotti. European Language Resources Association (ELRA). https://aclanthology.org/2020.lt4hala-1.21.

Straka, Milan, e Jana Straková. 2020. «UDPipe at EvaLatin 2020: Contextualized Embeddings and Treebank Embeddings». *Proceedings of LT4HALA 2020 – 1st Workshop on Language Technologies for Historical and Ancient Languages* (Marseille), 124–29. https://doi.org/10.48550/arXiv.2006.03687.

Tarrant, Richard, a c. di. 2012. *Virgil. Aeneid. Book XII*. Cambridge University Press.

Thakur, Nandan, Nils Reimers, Johannes Daxenberger, e Iryna Gurevych. 2021. «Augmented SBERT: Data Augmentation Method for Improving Bi-Encoders for Pairwise Sentence Scoring Tasks». In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, a cura di Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, *et al.* Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.28.

Turian, Joseph, Lev Ratinov, e Yoshua Bengio. 2010. «Word representations: a simple and general method for semi-supervised learning». *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (Stroudsburg [PA]), ACL '10, 384–94. https://dl.acm.org/doi/10.5555/1858681.1858721.

Venuti, Martina. 2025. «Res novae? Immo novissimae. Ultime evoluzioni degli archivi digitali di poesia latina (MQDQ Galaxy)». In *RES NOVAE. Il latino nella società postdigitale. Atti del Convegno della CUSL*, a cura di Maria Luisa Delvigo e Fabio Gasti. Palumbo Editore. https://www.classicocontemporaneo.eu/index.php/biblioteca/248-vol-17-res-novae-il-latino-nella-societa-post-digitale/684-res-novae-immo-novissimae-ultime-evoluzioni-degli-archivi-digitali-di-poesia-latina-mqdq-galaxy.

Venuti, Martina, Angelo Mario Del Grosso, Federico Boschetti, *et al.* 2023. «La 'Galassia MQDQ': un concetto di filologia tradizionale, digitale, sostenibile». *magazén* 4 (1): 71–120. https://doi.org/10.30687/mag/2724-3923/2023/07/003.

Zissos, Andrew, a c. di. 2008. *Valerius Flaccus' Argonautica, Book 1. Edited with Introduction, Translation, and Commentary*. Oxford University Press.