



# La mappatura semantica delle fonti: una nuova chiave per gli studi danteschi

Michelangelo Zaccarello, professore ordinario di filologia della letteratura italiana presso l'Università di Pisa

Fra i settori degli studi danteschi che hanno conosciuto un maggiore progresso negli ultimi anni, un posto di rilievo spetta alle ricerche sul contesto sociale e culturale in cui il Poeta si è formato, fra gli ambienti fiorentini delle grandi biblioteche, come S. Croce e S. M. Novella, e quelli bolognesi delle *disputationes* laiche o religiose. Prima ancora dell'apporto decisivo degli ambienti e dei contatti dell'esilio, Dante aveva immagazzinato, grazie anche alla prodigiosa memoria, uno straordinario bagaglio di *auctoritates* che lo accompagnerà per tutta la vita, nella complessa stratigrafia del poema come nelle opere minori. La mole di conoscenze espresse dal secolare commento, se adeguatamente mappata e interrogata, permette di estendere e integrare l'analisi della multiforme cultura dantesca, dalle fonti classiche alle compilazioni scientifiche, dalle raccolte sapienziali ai cicli cavallereschi romanzi, per rispondere a domande sempre più articolate e complesse sulla *Commedia* e i suoi vari contesti di riferimento.

Nello studio degli autori specie medievali, la ricerca sulle fonti e sui vari fenomeni d'intertestualità è una prospettiva di estremo interesse, tanto per la definizione della cultura dello scrittore quanto per l'interpretazione della sua opera. Negli anni Sessanta e Settanta del secolo scorso, è stata messa a punto una importante impalcatura teorica per la definizione dei fenomeni dell'intertestualità (si pensi a *Palinsesti* di Gerard Genette, Einaudi 1997, ma anche a studi precedenti di Julia Kristeva e Roland Barthes). Ciò nonostante, le concorrenti spinte di fattori storici e contestuali (restrizioni legate al genere letterario o ai committenti) e l'adozione di codici espressivi di particolare complessità (come il linguaggio allegorico o profetico) rendono sempre difficile tracciare le filigrane culturali di un testo.

Nel caso di Dante, poi, lo straordinario ritmo di pubblicazione delle ricerche sui testi e l'estrema varietà delle proposte d'interpretazione rendono particolarmente arduo orientarsi in questo *mare magnum*: è per questo che gli studi danteschi sono sempre stati all'avanguardia nelle applicazioni delle *Digital Humanities*, che hanno offerto un supporto insostituibile alla 'navigazione' dei dati disponibili. Ad esempio, fin dagli anni Ottanta del secolo scorso è stata allestita una biblioteca digitale, il *Dartmouth Dante Project* (DDP)<sup>1</sup>, che permette di interrogare – per verso, terzina o porzioni più ampie – il secolare commento alla *Commedia*, confrontando in modo rapido e diretto l'annotazione antica e moderna, mettendone subito in rilievo le frequenti dissonanze.

Presto seguito da analoghe iniziative, il DDP è ancor oggi un punto di riferimento degli studi al più alto livello, affiancato da risorse maggiormente orientate alla didattica e alla multimedialità: il *DanteLab*, sempre sviluppato dal Dartmouth College, che usa la tecnologia *split screen* per creare visualizzazioni sinottiche dei testi, dei commenti e delle traduzioni. Presso la Columbia University, è stato sviluppato il *tool* Digital Dante sotto la direzione di Teodolinda Barolini, che vi ha integrato il proprio commento: attraverso una serie di link, il sito permette di evidenziare la complessa intertestualità del poema ed accedere direttamente ai passi paralleli<sup>2</sup>.

Benché appaia oggi offline, il sito *Mapping Dante* della University of Pennsylvania mette il poema in relazione con il contesto geografico e cosmologico, permettendo di visualizzarlo in modo intuitivo alla luce delle coeve conoscenze in materia. L'ultima generazione delle risorse digitali, non solo dantesche, verte infatti sulle possibilità di interrogazione e visualizzazione. Nel caso della *Commedia*, poi, l'estremo accumulo e la dispersione della conoscenza espressa dai commenti antichi e moderni rende di fatto obbligatorio il ricorso agli strumenti informatici e ai motori di ricerca.

Considerate le molteplici stratificazioni retoriche e semantiche della poesia di Dante, e la diversa portata delle sue implicazioni allegoriche, un fattore chiave per stabilire l'interpretazione di un dato brano è il riconoscimento delle sue varie fonti, che spaziano dall'intertestualità diretta a forme più ampie di riecheggiamento delle conoscenze enciclopediche medievali, tenendo conto del ruolo chiave giocato dall'oralità e dalla memoria nella circolazione delle idee, che spesso risale a fonti non scritte quali dispute e discorsi pubblici: *quaestiones* e *disputationes* di ambito accademico o religioso, come quelle citate in *Convivio* II xii 7 (le "scuole dei religiosi e le disputazioni dei filosofanti"); le varie forme di predicazione nelle pubbliche piazze.

Nel caso di Dante, la quantità di informazioni disponibili su tali aspetti è imponente e soggetta all'estrema dispersione delle tradizionali pubblicazioni scientifiche: anzitutto i commenti, ma anche le riviste, le miscellanee, le enciclopedie e altri repertori generali. Con l'avvento dei media digitali, l'accesso a queste fonti è stato notevolmente accresciuto dall'indicizzazione e dalle ricerche testuali. Comunque, tali interrogazioni sono perlopiù orientate a ipotesi già formulate nella mente degli studiosi e/o nella biblio-

## Note

<sup>1</sup> Gli strumenti sviluppati nell'ambito del Dartmouth Dante Project sono ospitati dal sito <https://dante.dartmouth.edu>.

<sup>2</sup> Digital Dante è consultabile sul sito <https://digitaldante.columbia.edu>.

grafia progressiva; un reale progresso della conoscenza ha bisogno di superare la rigidità delle interrogazioni basate sul testo (e sulle marcature in linea incorporate nel testo) e di legare la conoscenza a un più vasto campo di classi e categorie, in connessione con parti e/o caratteristiche astratte dei testi primari più ampie. Tale evoluzione può essere agevolata dagli standard di descrizione e interrogazione del ‘web semantico’, un’estensione del normale web compiuta attraverso standard descrittivi che rendono le informazioni già disponibili sul web comprensibili al computer, mediante una descrizione formale e schemi classificatori chiamati ‘ontologie’.

Queste ultime possono essere definite come una ‘rappresentazione semantica dell’informazione’<sup>3</sup>, espressa in un formato comprensibile e interrogabile dalla intelligenza artificiale, che può a sua volta espandere tale rappresentazione mediante le varie possibili inferenze logiche estraibili dalle asserzioni formulate in input dai collaboratori del progetto, con il risultato di costruire una rete semantica espandibile. Applicate alla conoscenza specifica espressa dai vari domini, disponibili da repertori tematici o progettati in modo specifico, le ontologie sono un fattore chiave per governare la difficile transizione al web semantico. Le annotazioni già disponibili (per esempio nella marcatura ormai pressoché generalizzata XML, cioè *eXtensible Markup Language*, che è però vincolata alla sequenza verbale e ai normali rapporti di contiguità e dipendenza sintattica) sono integrate da una marcatura offline, dove un motore di calcolo basato sulle ontologie aggiunge tutte le proprietà che possono essere dedotte o indotte da esse; così, la ricerca basata sulle ontologie può sfruttare le risorse web preesistenti e renderle più ‘comprensibili’ per le macchine, dando dunque risposta a interrogazioni più ampie e sofisticate.

Anche se risulta assai complicato definire degli standard che possano essere adattati alle molteplici finalità e caratteristiche dei testi annotati, esistono naturalmente grandi repertori di ontologie disponibili alla descrizione di risorse specie bibliografiche: il progetto europeo C-IDOC-CRM, lo standard di descrizione bibliografica *Functional Requirements for Bibliographic Records* (FRBR) sviluppato dalla *International Federation of Library Association* (IFLA) e soprattutto il *Dublin Core Metadata Initiative* sviluppato negli Stati Uniti<sup>4</sup>; tuttavia, è inevitabile dover progettare e sviluppare ontologie dedicate per venire incontro alle particolari caratteristi-

che che si intende evidenziare nella marcatura (e rendere quindi disponibili all’interrogazione).

Data la sua straordinaria vastità e complessità, la cultura dantesca si presta a una mappatura basata sulle ontologie, improntata a standard descrittivi quali la famiglia di linguaggi RDF (*Resource Description Framework*), e sono stati progettati strumenti specifici per esprimere gli aspetti più complessi e articolati della produzione letteraria dantesca, quali il suo uso di fonti bibliche, classiche e medievali. Basandosi sul *tool Dante Sources*, l’*Hypermedia Dante Network* (HDN) che qui si propone ha lo scopo di raffinare ed estendere le ontologie sviluppate per le opere minori di Dante al più complesso mondo della *Commedia*. Applicare i nuovi standard al poema sarà impegnativo perché a) il suo diverso campo di riferimenti richiederà un vero approccio ipermediale (per esempio un lavoro di rete collaborativo fra risorse comunicanti) e b) la sua sterminata bibliografia pone seri problemi nell’estrarre informazioni rilevanti, un compito che può essere svolto solo sviluppando strumenti per l’elaborazione semi-automatica delle fonti.

Potendo contare su alcune delle più importanti istituzioni scientifiche per gli studi su Dante, quali la Società Dantesca Italiana di Firenze, l’Accademia della Crusca e la Scuola Normale Superiore di Pisa, HDN si baserà su competenze specialistiche, essenziali per amministrare le fasi iniziali della ‘popolazione’ di ontologie specificamente designate, per esempio identificare ed estrarre istanze di classe ontologica dai testi archiviati. Mentre i testi sono dapprima analizzati per estrarre e codificare tutte le informazioni che possono essere elaborate automaticamente da un apposito software, questo processo deve essere seguito da una integrazione ‘manuale’ dei dati, codificati da specialisti, e dalla loro pubblicazione sul web come *Linked Data*, standard aggiornato che facilita la connessione e il riuso di dati strutturati.

In altri termini, HDN ha lo scopo di costituire un ambiente collaborativo per il commento delle opere dantesche, con particolare attenzione alle questioni di lingua, stile e soprattutto intertestualità, andando a costituire una importante integrazione dei *tool Dante Search* – che delle opere dantesche indaga la dimensione propriamente linguistica mediante una marcatura dedicata agli aspetti morfosintattici – e *DaMA. Dante Medieval Archive*, che punta a ricostruire “le principali fonti classiche, tardoantiche e medievali, latine e volgari, delle opere dantesche e i commenti oggetto della riflessione teorica e letteraria di Dante”<sup>5</sup>. Usando fonti

## Note

3 Cfr. Metilli, Daniele, *Gli strumenti informatici per lo studio di Dante. Prospettive di sviluppo e integrazione*, consultabile online su <http://www.labcd.unipi.it>.

4 Cfr. <http://dublincore.org>.

5 Dante Medieval Archive è ospitato sul sito <https://dama.dantenetwork.it>.

primarie affidabili in formato XML, la biblioteca digitale HDN risponderà a una vasta gamma di interrogazioni, grazie a una descrizione delle risorse basata sulle ontologie e alla più facile condivisione di dati con risorse preesistenti.

Indirizzato all'uso didattico e scientifico ma disponibile a un'utenza più generale, HDN avrà ampia accessibilità grazie a interfacce di facile utilizzo e potrà dunque essere utilizzabile anche per importanti scopi educativi. Per esprimere efficacemente le conoscenze condivise sugli autori citati o allusi nel testo, la biblioteca digitale HDN parte da simili applicazioni precedenti di linguaggi descrittivi basati sulle ontologie, quali lo strumento *Dante Sources*, dedicato alle fonti individuate dai commenti moderni nelle opere minori di Dante.

Vediamo adesso di entrare dentro all'articolazione del *tool* HDN, nel modo cioè in cui le asserzioni di un dato commento vengono interpretate dalla marcatura semantica e rese dunque disponibili a una serie di operazioni svolte dal motore di ricerca per future interrogazioni. Si è detto che, a partire dai repertori disponibili, il nostro team – in particolare, l'unità di ricerca pisana assistita dal CNR – ha sviluppato un'ontologia dedicata. La classe al centro dell'ontologia HDN è il 'riferimento', che può essere definito nel modo seguente:

Il testo A (il commento dantesco, detto 'sorgente' del riferimento) dice che la conoscenza sul testo B (un brano della *Commedia*, detto 'soggetto' del riferimento) si può arricchire grazie al testo C (la fonte di Dante, detta 'oggetto' del riferimento).

Per citare un esempio dal canto proemiale dell'*Inferno*:

Poeta fui, e cantai di quel giusto / figliuol d'Anchise che venne di Troia, / poi che 'l superbo Ilíon fu combusto (*Inf.* I, 73-75).

In proposito, la chiosa dell'Anonimo Fiorentino recita: "Qui si manifesta Virgilio all'Autore, et dice ch'elli fu poeta, et che fece versi d'Enea, il quale venne di Troja, et fu figliolo d'Anchise et di Venere. Chiamalo giusto per seguitare Virgilio in ogni sua opera; onde Virgilio nel primo dell'Eneida: Rex erat Eneas nobis, quo justior alter Nec pietate fuit nec bello major et armis etc.". Dunque, ai nostri fini, possiamo sintetizzare:

l'Anonimo Fiorentino (sorgente del riferimento) dice che la conoscenza su *Inf.* I, 73-75 (soggetto del riferimento) si può arricchire grazie a *Aen.* I, 544-545 (oggetto del riferimento).

Naturalmente, i riferimenti possono essere di diversi tipi, a seconda del rapporto instaurato dal testo sorgente fra soggetto e oggetto: l'importante è che tale rapporto sia sempre determinato dalla conoscenza espressa dal commento (e non, ad esempio, dalla nostra sensibilità moderna, magari più evoluta o con un più ampio accesso alle fonti). Si possono individuare tre tipologie fondamentali di riferimento:

- supporto esterno: il commentatore menziona un testo per supportare una propria interpretazione del passo dantesco;
- luogo parallelo: il commentatore menziona un testo in cui un concetto, uno stilema o un personaggio manifestano delle somiglianze con il testo dantesco, senza implicare una derivazione;
- citazione: il commentatore ritiene che Dante stia citando un testo.

Anche se non è sempre possibile stabilire il tipo di riferimento, anche in presenza di tipologie 'miste', l'esempio citato da *Inf.* I può agevolmente rientrare nel terzo tipo, con rinvio dichiarato a un'*auctoritas*. Sulla base dell'esperienza maturata da *Dante Sources* sulle opere minori, l'ontologia HDN è stata modellata per poter esprimere la tipologia della citazione, che può assumere tre forme fondamentali:

- citazioni esplicite: il commentatore ritiene che Dante citi in maniera esplicita una fonte;
- citazioni stringenti: il commentatore individua una citazione implicita ma riconducibile a un preciso brano fonte;
- concordanze generiche: il commentatore individua un concetto che si trova genericamente alla base di un brano dantesco.

Per fare un altro esempio, a *Inf.* III, 112-114 Dante usa una celebre similitudine:

Come d'autunno si levan le foglie / l'una appresso de l'altra, fin che 'l ramo / vede a la terra tutte le sue spoglie

Buona parte dei commenti, ad esempio quello assai diffuso di Umberto Bosco e Giovanni Reggio (1979), cita un passo virgiliano: "quam multa in silvis autumnus frigore primo / lapsa cadunt folia aut ad terram gurgite ab alto / quam multae glomerantur aves, ubi frigidus annus / trans pontum fugat et terris immittit apricis" (*Aen.* VI 309-312: 'Quante foglie cadono nelle selve ai primi freddi dell'autunno o quanti uccelli si riuniscono sulla terra venendo dal mare quando la fredda stagione

li mette in fuga attraverso l'oceano e li fa andare nelle terre soleggiate'). Come si vede, la citazione è implicita, perché Dante non nomina espressamente il passo dell'*Eneide*, ma la concordanza è stringente e può ricadere nel secondo tipo illustrato sopra.

Queste e altre specificazioni non sono, naturalmente, nuove nel panorama degli studi danteschi: nuovo è però il modo di esprimerle all'interno di un reticolo logico comprensibile all'intelligenza artificiale, che può dunque espanderne la portata e testarne il funzionamento, consentendo interrogazioni complesse e articolate di una tipologia non praticabile sulla base della tradizionale ricerca di concordanze verbali. L'immissione dati è laboriosa, data la grande mole di commenti coinvolti (in questa fase, 51 degli oltre 70 digitalizzati dal DDP), ma sono stati implementati sistemi di autocompletamento per sostenere il lavoro degli annotatori. Inoltre, per alcuni elementi del poema (ad esempio i personaggi e i luoghi) sono state importate e codificate le informazioni dell'*Enciclopedia Dantesca* e di *Wikidata*. Anche in questo caso è possibile, ove necessario, rettificare e ridefinire ogni dato alla luce degli studi più recenti.

## Bibliografia

Meghini, Carlo; Tavoni, Mirko; Zaccarello, Michelangelo, *Mapping the Knowledge of Dante Commentaries in the Digital Context: A Web Ontology Approach*, in "Romanic Review", CXII/1, 2021, pp. 138-157.

Tavoni, Mirko; Andriani, Paola; Meghini, Carlo; Bartalesi, Valentina; Metilli, Daniele, *L'esplorazione delle fonti dantesche attraverso la biblioteca digitale Dante Sources*, in "Sulle tracce del Dante minore. Prospettive di ricerca per lo studio delle fonti dantesche", a c. di Thomas Persico e Riccardo Viel, Bergamo, Società Dante Alighieri/ Sestante edizioni, 2017, pp. 29-52.

Tavoni, Mirko, *DanteSearch: Il corpus delle opere volgari e latine di Dante lemmatizzate con marcatura grammaticale e sintattica*, in "Lectura Dantis 2002-2009. Omaggio a Vincenzo Placella per i suoi settanta anni", a c. di Anna Cerbo et al., Napoli, Il Torcoliere, 2011, pp. 583-608.

Tavoni, Mirko, *DanteSearch: Istruzioni per l'uso. Interrogazione morfologica e sintattica delle opere volgari e latine di Dante*, in "Sintassi dell'italiano antico e sintassi di Dante", a c. di Marta D'Amico, Pisa, Felici, 2015, pp. 59-79.