DOI: 10.17469/O2106SLI000017

MIRKO TAVONI

# Allestimento, fruizione e prospettive di *DanteSearch*

La relazione illustra le funzionalità, la storia e le attuali linee di sviluppo della risorsa *DanteSearch* (https://dantesearch.dantenetwork.it/), corpus completo delle opere volgari e latine di Dante con annotazione linguistica in formato XML-TEI. Nato all'Università di Pisa all'inizio degli anni Duemila, *DanteSearch* si è sviluppato attraverso successivi progetti di ricerca e mette a disposizione funzionalità uniche di ricerca morfologica e sintattica. *DanteSearch* è una risorsa utilizzata nell'ambito di progetti di ricerca contigui quali il *Vocabolario Dantesco* (http://www.vocabolariodantesco.it/), il *Vocabolario Dantesco Latino* (http://www.vocabolariodantescolatino.it/) e il progetto ERC *LiLa-Linking Latin* (https://lila-erc.eu/#page-top). Nell'ambito del progetto *HDN-Hypermedia Dante Network* (https://hdn.dantenetwork.it/) *DanteSearch*, insieme con *DanteSources* (https://dantesources.dantenetwork.it/), viene ripensato in ottica di web semantico, al fine di costituire una base di conoscenza fondata su logiche calcolabili (RDF, OWL) superando i limiti della singola struttura gerarchica imposta da XML.

*Parole chiave*: Dante Alighieri, linguistica dei corpora, storia della grammatica italiana, italiano antico, latino medievale.

## 1. *Introduzione*

Sono onorato dell'invito a tenere questa relazione plenaria e felice di avere con ciò l'occasione di presentare anche a linguisti non italianisti uno strumento di ricerca – *DanteSearch: corpus delle opere volgari e latine di Dante con annotazione morfologica e sintattica* – a cui hanno lavorato con me nell'arco di vent'anni giovani ricercatori di varie generazioni: un'impresa in divenire passata attraverso diverse stagioni tecnologiche che si sta evolvendo nella logica del web semantico.

*DanteSearch* ha la sua ormai lontana origine nell'ambito del progetto di ricerca di interesse nazionale (PRIN 1999) di cui sono stato coordinatore nazionale, intitolato *La memoria testuale. Edizioni,*

*studi e strumenti per l'analisi computazionale del patrimonio italiano*, che portò alla creazione di molte collezioni e strumenti di filologia digitale presso diverse Università italiane, fra cui il nucleo di quella che diventerà poi la Biblioteca italiana, la più grande biblioteca digitale di testi della letteratura italiana (http://www.bibliotecaitaliana.it/).

L'archivio testuale dantesco prodotto allora (una prima presentazione in Tavoni 2005) presentava diversi motivi di interesse: il corpus era completo; includeva le opere latine, molto meno reperibili on line di quelle volgari; i testi digitali garantivano un'affidabilità ben diversa dalla massa dei testi incontrollatamente reperibili in rete; e soprattutto erano stati non solo lemmatizzati ma anche arricchiti di una marcatura morfologica molto capillare, che consentiva di svolgere sull'*opera omnia* di Dante ricerche linguistiche fino ad allora impossibili.

Quella prima versione è stata successivamente ampliata e perfezionata da diversi punti di vista, in particolare nell'ambito del PRIN 2009 *Morfosintassi e corpora informatici dell'italiano antico* coordinato da Lorenzo Renzi e successivamente da me.

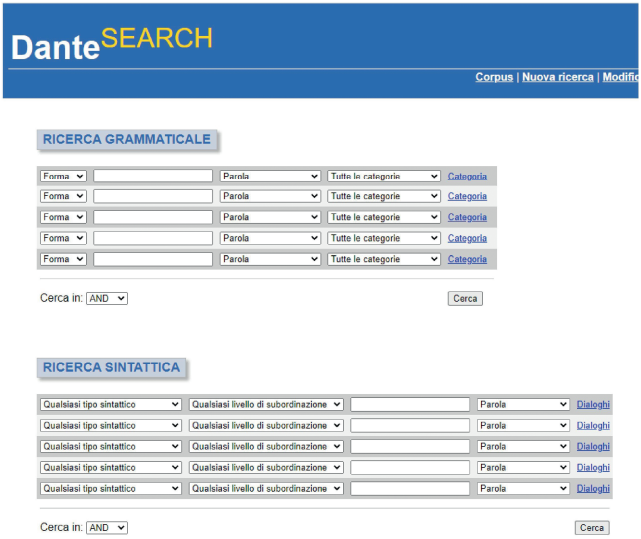
La squadra dei codificatori grammaticali del progetto originario era coordinata da Samuela Brunamonti, e il motore di ricerca XCDE (XML Compressed Document Engine: http://pages.di.unipi.it/ferragina/Libraries/xcde/index.htm) era stato ideato da Paolo Ferragina, brillante algoritmista, allora giovane ricercatore di Informatica e docente di Information retrieval nel neonato corso di laurea in Informatica umanistica.

È ovviamente superflo ricordare che la TEI-Text Encoding Initiative (http://www.tei-c.org/index.xml), «is a consortium which collectively develops and maintains a standard for the representation of texts in digital form». La conformità allo standard internazionale XML-TEI è – o almeno era, all'epoca – un requisito imprescindibile per qualunque impresa di filologia digitale: per rendere le codifiche indipendenti dalle tecnologie, hardware e software, in continua evoluzione, e così renderle permanenti nel tempo; scambiabili fra progetti di ricerca diversi che usano tecnologie diverse; riusabili anche a fini diversi da quelli per i quali sono state prodotte. La codifica grammaticale secondo lo standard XML-TEI fu curata da Elena Pierazzo, che aveva appena terminato il suo dottorato in Filologia italiana alla Scuola Normale Superiore ed era una colonna del nostro già ricordato corso di laurea in Informatica umanistica, ed era destinata a una bril-

lante carriera nel Regno Unito (King's College London) e in Francia (Université Grenoble 3, Université de Tours-CESR), e ad assumere ruoli di responsabilità, appunto, nel Consorzio TEI, e a divenirne, dal 2012 al 2014, Chair.

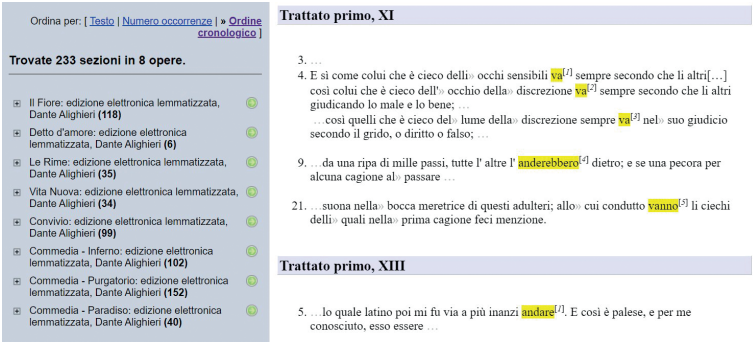
Chi si colleghi oggi al sito di *DanteSearch* (https://dantesearch.dantenetwork.it/), dopo aver cliccato su "Nuova ricerca" si trova davanti a una schermata divisa in due parti: "Ricerca grammaticale" (nel senso di morfologica) e "Ricerca sintattica" (Fig. 1).

Figura 1 - *Maschera di ricerca iniziale di DanteSearch*




Nella "Ricerca grammaticale" il primo campo permette di scegliere tra ricerca per "Forma" o per "Lemma"; il secondo tra ricerca per "Parola" (s'intende parola intera), "Sottostringa" (cioè qualunque parte di parola), "Prefisso" o "Suffisso" (naturalmente non in senso morfologico, ma nel senso di parte iniziale o finale di parola), o "Espressione regolare", o "Tutte le occorrenze". Per esempio, posso digitare nel primo campo la stringa 'fortuna' come "Forma" e come "Parola", il che mi darà come risultato 44 occorrenze della parola intera *fortuna* in 11 opere, sia in volgare sia in latino. Posso digitare 'andare' come "Lemma" e come "Parola", ottenendo 586 occorrenze in 8 opere, evidentemente solo in volgare (Fig. 2).

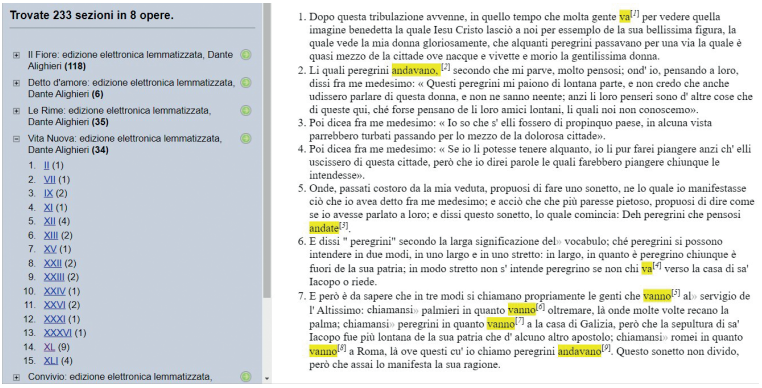
Figura 2 - *Risultati di ricerca per il lemma* andare*, visualizzazione ristretta*




E posso digitare parti di parole (parti qualunque, o iniziali, o finali), come forme o come lemmi. Il risultato di una ricerca di questo tipo appare inizialmente, come vediamo nella Fig. 2, in forma di elenco dei testi che contengono la parola, o la parte di parola, o l'insieme di parole richiesti, col relativo numero di occorrenze; questo elenco può essere ordinato, cliccando sulla riga in alto, o per "Testo" (cioè in ordine alfabetico per titolo di testo), o per "Ordine cronologico" (cioè secondo la successione delle date di composizione dei testi), o per "Numero occorrenze" (cioè ordinando i testi per numero decrescente di occorrenze contenute in ciascun testo). A partire da un elenco di questo tipo, si possono scegliere due diverse visualizzazioni delle occorrenze. La prima si ottiene cliccando sul tondino con freccetta a fianco del singolo testo, e consiste nella visualizzazione di tutte le occorrenze entro il testo, in ordine di testo, evidenziate in giallo entro contesti di poche righe ritagliati dal testo, come vediamo nella Fig. 2.

La seconda visualizzazione, che vediamo nella Fig. 3, si ottiene cliccando su un titolo di testo dell'elenco. Si apre allora nel menu a tendina l'articolazione interna delle "Sezioni" (cioè canti, capitoli, ecc.) di quel testo in cui è presente almeno una occorrenza della stringa ricercata: p.es. qui i capitoli della *Vita nova*. Cliccando su una di queste, nella finestra a destra compare il testo integrale della sezione, nel quale le occorrenze della stringa ricercata sono evidenziate in giallo, come vediamo.

Figura 3 - *Risultati di ricerca per il lemma* andare*, visualizzazione estesa*

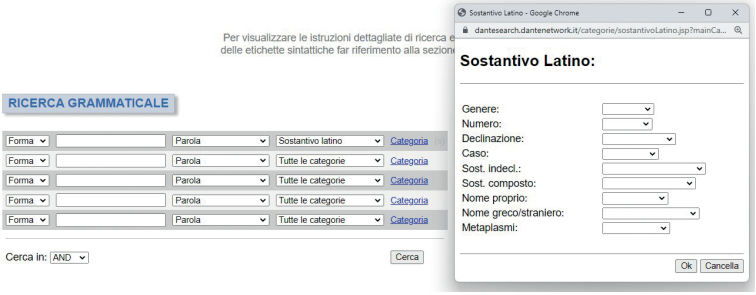



Dunque, tornando alla maschera di ricerca iniziale (Fig. 1), i primi due campi nella riga di ricerca (Lemma/Forma e Parola/Sottostringa, ecc.) servono per la ricerca lessicale. Il terzo campo introduce la vera novità: la ricerca per categorie grammaticali. Ogni occorrenza di ogni parola nel testo, infatti, è stata etichettata come appartenente a una parte del discorso, e più specificamente come caratterizzata dai corrispondenti possibili tratti morfologici. Il terzo campo della riga, che appare inizialmente occupato dalla dizione "Tutte le categorie", permette di scegliere fra Volgare e Latino e all'interno dell'uno o dell'altro fra Verbo, Sostantivo, Aggettivo, ecc.; operata questa selezione, cliccando sul link a destra "Categoria" si apre una finestra che consente di selezionare i tratti morfologici propri di quella categoria ovvero parte del discorso.

Per esempio, nella Fig. 4 si vedono quali sono i tratti morfologici selezionabili per il "Verbo volgare"; nella Fig. 5 quali sono i tratti morfologici selezionabili per il "Sostantivo latino".

Figura 4 - *Tratti morfologici ricercabili per "Verbo volgare"*


Figura 5 - *Tratti morfologici ricercabili per "Sostantivo latino"*



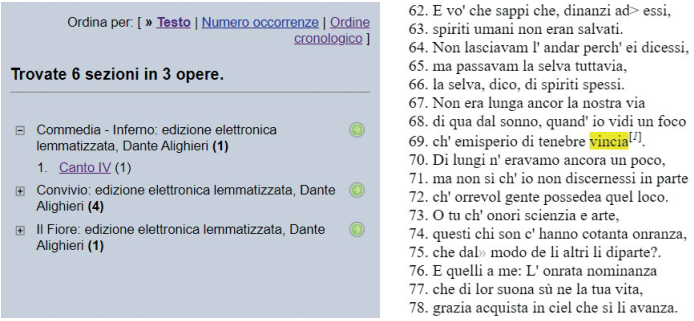

La ricerca lessicale e quella grammaticale possono essere attivate l'una indipendentemente dall'altra, oppure in combinazione. Se voglio cercare tutte le occorrenze che rispondono a una certa definizione grammaticale, senza porre nessuna restrizione di tipo lessicale, sceglierò, nel secondo campo, "Tutte le occorrenze", il che neutralizza la selezione lessicale. Per esempio posso cercare tutti i pronomi dimostrativi volgari (che sono 2.995 in 8 opere); se seleziono solo i femminili singolari si riducono a 550; solo i maschili plurali 306, e così via.

Oppure posso porre a questa ricerca grammaticale una restrizione lessicale. In questo caso non selezionerò, nel secondo campo, "Tutte le occorrenze" ma per esempio "Prefisso", per restringere la ricerca dei pronomi dimostrativi a quelli che iniziano con – poniamo – *cost-*: che risultano essere 148 (*costui*, *costu'*, *costei*, *coste'*, *costoro*, *costor*) in 7 opere.

Un esempio euristico di ricerca combinata per tratti grammaticali e stringa lessicale consiste nella ricerca delle III persone singolari degli imperfetti indicativi della III, o in alternativa della II coniugazione,

con la restrizione che si tratti di forme terminanti in *-ia* (restrizione che si ottiene digitando *ia* come "Forma" e come "Suffisso", cioè semplicemente come parte finale della parola). Ne risulta che gli esempi della III coniugazione sono molto numerosi: 82 occorrenze distribuite su tutte le opere volgari tranne il *Detto d'Amore*. Invece gli esempi della II coniugazione (Fig. 6) risultano essere pochissimi.

Figura 6 - *III p.s. imperfetto indicativo II coniugazione in* -ia

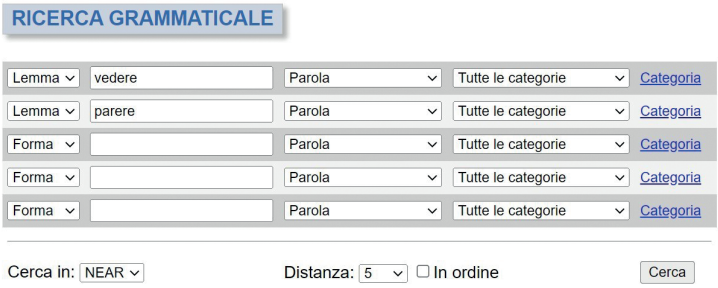



L'unica occorrenza nella *Commedia* è in *If* IV 69 «ch'emisperio di tenebre vincia» (R: *tuttavia* : *via*), ma solo perché il codificatore, seguendo la maggioranza degli interpreti, ha interpretato *vincia* come voce del verbo *vincere* e non di *vincire*, 'avvincere', interpretazione minoritaria. Ma questo quadro quantitativo d'insieme depone fortemente a favore di questa seconda interpretazione, perché non c'è in tutta la *Commedia*, oltre a questo caso dubbio, neanche un caso sicuro di imperfetto della II coniugazione in *-ia*, per sicilianismo (Tavoni 2011: 592-593; 2020b: 163-168).

Fin qui abbiamo visto singole ricerche, di natura lessicale e/o grammaticale, esprimibili su una sola riga della maschera di ricerca. Ma questa, come si vede in Fig. 1, ne annovera ben cinque. Su ogni riga si può formulare una ricerca, e due o tre o anche quattro o cinque ricerche si possono combinare con i classici operatori booleani AND, OR, NOT, NEAR. Da notare che la ricerca AND dà come risultato qualunque compresenza di parole all'interno della "Sezione" dell'opera (un intero canto della *Commedia*, un intero capitolo del *Convivio*, ecc.). Per cercare, come accade più spesso di voler fare, risultati compresenti a breve distanza, bisogna usare l'operatore NEAR. Il quale, una volta selezionato, fa aprire il box "Distanza", cioè a quante

parole di distanza al massimo devono cooccorrere le parole ricercate, con la possibilità di indicare se devono comparire nell'ordine in cui è stata formulata la richiesta oppure no. Per esempio, nell'ambito di una ricerca sulla fenomenologia della visione, mi può interessare cercare quanto spesso cooccorrono, e quanto da vicino, e in quali opere, i verbi *vedere* e *parere* (che in italiano antico ha notoriamente un valore più forte, fisico: non 'sembrare', ma 'apparire', con impatto visivo). Scriverò dunque questi due verbi, come "Lemma", in due righe successive, e li cercherò con la funzione NEAR (Fig. 7). Posso ripetere la ricerca variando la distanza, e troverò che le cooccorrenze sono numerosissime, e in tutte le opere volgari: a distanza 20 sono 322, a distanza 10 sono 201, a distanza 5 – come qui – sono 114. In latino, invece, i lemmi *appareo* e *video* non cooccorrono mai, nemmeno a distanza 20. Questa "mutua attrazione" del vedere e dell'apparire è il portato di un'ossessione stilnovistica, cavalcantiana, che resta viva per tutta la carriera poetica, ma solo poetica, di Dante, e non compare mai in un universo di discorso profondamente diverso come quello della trattatistica latina.

Figura 7 - *Ricerca di* vedere *e* parere *cooccorrenti entro distanza di 5 parole*




Passiamo ora alla ricerca sintattica. La strada per questa risorsa, completamente nuova, è stata aperta da Sara Gigli, che alla codifica sintattica esaustiva della Commedia, con tutti i problemi interpretativi che essa evidentemente comporta, ha dedicato la propria tesi di dottorato (Gigli 2004; e cfr. 2003, 2007, 2015). Sara Gigli, prendendo a riferimento l'impianto teorico-descrittivo della *Grande grammatica italiana di consultazione* di Renzi-Salvi-Cardinaletti (1988-1995), quindi della *Grammatica dell'italiano antico* di Salvi-Renzi (2010), si è posta il pro-

blema di come applicare quell'apparato di categorie e nozioni sintattiche al testo della *Commedia*, e di come tradurlo in un sistema di codifica capace di descrivere il testo per intero, in tutta la sua complessità e varietà di aspetti. Con la collaborazione di Elena Pierazzo per quanto riguarda il formalismo XML-TEI, Sara Gigli ha portato a termine in modo eccellente un compito tanto oneroso quanto irto di problemi interpretativi, filologici e linguistici, e con ciò ha messo a disposizione della comunità scientifica uno strumento unico e sofisticatissimo.

Come si vede nei campi 3 e 4, anche la ricerca per categorie sintattiche può essere combinata con restrizioni di tipo lessicale. La differenza è che il terzo campo ammette solo la ricerca per "Forma"; che può essere specificata, nel quarto campo, come "Parola" o "Sottostringa". Se invece voglio fare una ricerca esclusivamente per categorie sintattiche, senza restrizioni lessicali, anche qui, come nella ricerca grammaticale, nel quarto campo selezionerò "Tutte le occorrenze". I campi per la ricerca sintattica sono il primo, dedicato al tipo di frase, e il secondo, dedicato al grado di subordinazione.

Nel primo campo, il sistema permette di interrogare il corpus su due livelli: per tipi sintattici (una trentina, articolati in frasi principali e coordinate a una principale, subordinate e coordinate a una subordinata, psudocoordinate, parentetiche e cordinate a una parentetica) e per sottotipi (più di 300: dichiarativa illocutiva, coordinata congiuntiva esclamativa, interrogativa di tipo x, ecc. ecc.). La prima ricerca, per tipi, consente di ricercare (e conteggiare) tutte le relative insieme, tutte le consecutive insieme, tutte le ipotetiche insieme, ecc.; la seconda consente di ricercare sottotipi di frase in modo estremamente analitico.

Il secondo campo permette di specificare il grado di subordinazione al quale si vuole restringere la ricerca del sottotipo di frase definito nel campo precedente. Per esempio voglio vedere le interrogative alternative, ma solo quelle che siano subordinate di III grado. Ce ne sono solo 3, di cui una nell'*Inferno*: «Dintorno mi guardò, come talento / avesse di veder s'altri era meco», (X 55-56). Si aggiunga la possibilità di usare gli operatori AND e NEAR, che consentono di ricercare cooccorrenze di tipi di frase, a una determinata distanza l'uno dall'altro ed eventualmente nell'ordine voluto – per esempio una causale immediatamente seguita da una finale (come «Qual si lamenta perché qui si moia / per viver colà su, non vide quivi...», *Pd* XIV 25-26) – e si percepirà il numero vertiginoso di combinazioni di

ricerca sintattica che si possono costruire, per rispondere ai più vari percorsi mentali del ricercatore, alimentando una catena virtuosa di risposte a curiosità di ricerca, a loro volta passibili di stimolare nuove intuizioni e nuove domande.

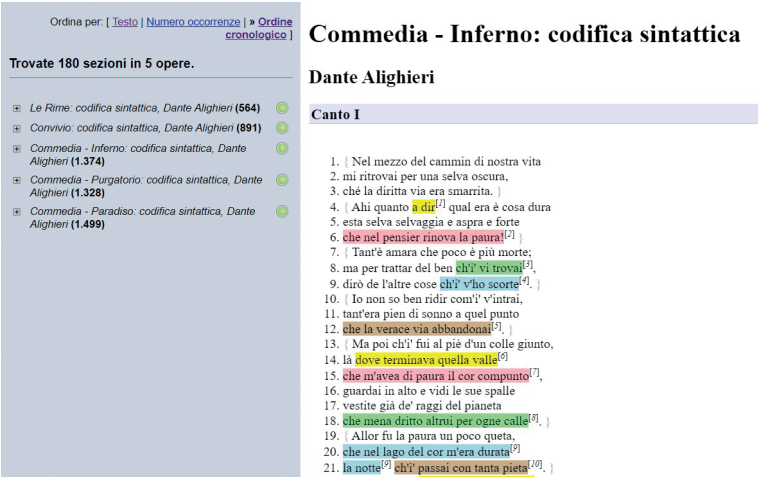
Vediamo ora un solo esempio di una ricerca sintattica generalissima, finalizzata a una verifica molto specifica. In un mio ormai antico lavoro (Tavoni 2002) ho proposto un argomento sintattico che non era mai stato addotto a proposito della vessatissima questione di «forse cui Guido vostro ebbe a disdegno» (*If* X 63): cioè il principio che i grammatici generativi chiamano "insularità delle relative", secondo il quale nessun elemento appartenente a una frase relativa può essere dislocato a sinistra del pronome relativo che la introduce – e quindi *forse* NON può appartenere alla relativa, cioè significare '...a colui che forse il vostro Guido ebbe a disdegno'. Questa era l'interpretazione preferita da Contini, ma a mio giudizio è agrammaticale, dunque impossibile. All'epoca non disponevo della ricerca sintattica di *DanteSearch*. Oggi è possibile sottoporre la mia tesi a una verifica esaustiva, semplicemente richiamando insieme tutte le proposizioni relative in tutta la *Commedia*. Lancio dunque la ricerca sintattica: "Relativa" (senza ulteriore specificazione) – "Tutte le occorrenze".

Il risultato è un totale di 4.201 frasi relative nella *Commedia*. Purtroppo, al momento non è possibile chiedere al sistema di fare anche l'ultima operazione materiale per noi, cioè chiedergli di estrarre da queste 4.201 frasi quelle in cui eventualmente il pronome relativo non occupi la prima posizione. Non resta dunque che scorrere tutte queste frasi, dalla prima all'ultima: operazione che comunque non richiede più di qualche ora. Nella Fig. 8 appaiono, nel corso di un tale scorrimento di tutti i contesti dell'*Inferno*, quelli compresi nei primi 21 versi del I canto.

Come abbiamo già notato, nella maschera della ricerca sintattica il terzo campo prevede solo la possibilità di ricercare per forma, non per lemma. Questo significa che la lemmatizzazione non è fruibile all'interno della ricerca sintattica, come non lo è la marcatura morfologica, perché la lemmatizzazione-marcatura morfologica da una parte, e la marcatura sintattica dall'altra, sono state realizzate su due distinte copie (in partenza identiche) dello stesso corpus testuale. Sarebbe stato impossibile sovrapporre le due marcature l'una sull'altra: è questo un limite del formalismo di codifica XML che rende impossibile, allo

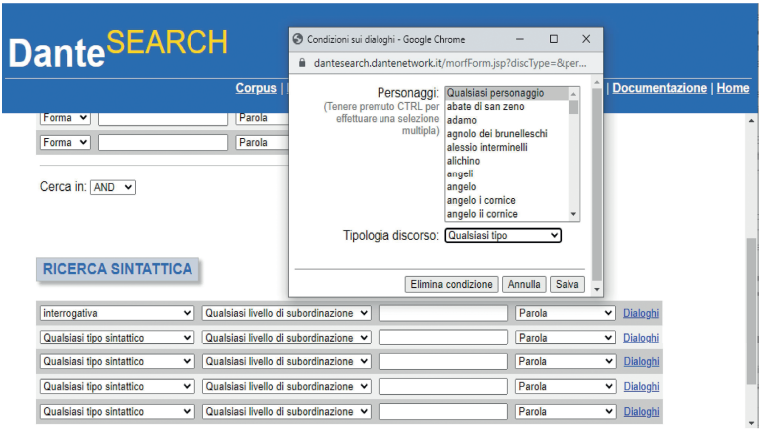
stato dell'arte, interrogare insieme per lemmi, per categorie morfologiche e per categorie sintattiche. Sarebbe invece molto interessante poter combinare questi diversi tipi di ricerca. Come superare questo limite di XML costituisce precisamente il problema che l'attuale fase di sviluppo di *DanteSearch* ha davanti, e su questo mi soffermerò alla fine della mia relazione.

Figura 8 - *Le frasi relative nel I canto dell'*Inferno



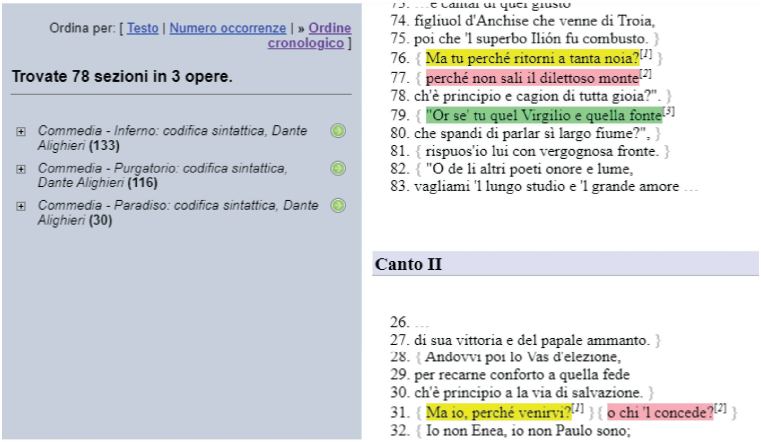

Infine, la ricerca sintattica sul parlato dei personaggi. Per dare un'idea delle possibilità offerte dalla ricerca sintattica limitata ai discorsi o pensieri pronunciati o pensati da personaggi della *Commedia*, codifica introdotta da Marta D'Amico sulla base della sua tesi di laurea magistrale dal suggestivo titolo *La sintassi dell'aldilà*, vediamo anzitutto che, cliccando su "Dialoghi" nella maschera di ricerca, si apre un box che permette di selezionare i "Personaggi" (tutti i personaggi della *Commedia* vi compaiono in ordine alfabetico) e/o la "Tipologia di discorso" (Fig. 9). Lanciamo dunque per prima cosa una ricerca sintattica scegliendo, nella Finestra "Dialoghi", "Qualsiasi personaggio" e "Qualsiasi tipologia di discorso", per andare a verificare la presenza e distribuzione di un certo tipo di frase – per esempio le principali interrogative – all'interno delle parti mimetiche del poema.

Figura 9 - *Ricerca sul parlato dei personaggi della* Commedia




Ed ecco il risultato (Fig. 10): 133 frasi principali interrogative nell'*Inferno*, 116 nel *Purgatorio*, 30 nel *Paradiso*. La sproporzione fra la terza cantica e le prime due è clamorosa, e rende evidente (ciò che resta invece invisibile "a occhio nudo") una diversa "condizione cognitiva" dei beati rispetto ai dannati e ai penitenti. Mentre a destra compare un esempio della ricorrenza di interrogative nei dialoghi fra I e II canto dell'*Inferno*.

Figura 10 - *Le frasi principali interrogative nel parlato della* Commedia



È solo un esempio. La ricerca di Marta D'Amico (2009 e 2015) esplora una grande messe di dati desumibili dal sistema di interrogazione e li combina con altri dati, come gli introduttori, i tempi e i modi verbali, la semantica delle frasi, ricavabili per altra via, traendone conclusioni molto interessanti circa la caratterizzazione linguistica in rapporto all'identità storica dei personaggi, individualmente e per categorie in senso lato sociali; circa i fenomeni di mimesi del parlato che l'autore sfrutta graduandoli sapientemente (su questo cfr. anche Tavoni 2020a); e circa i diversi, appunto, stili cognitivi che riflettendosi in strutture sintattiche privilegiate qualificano i discorsi delle tre cantiche.

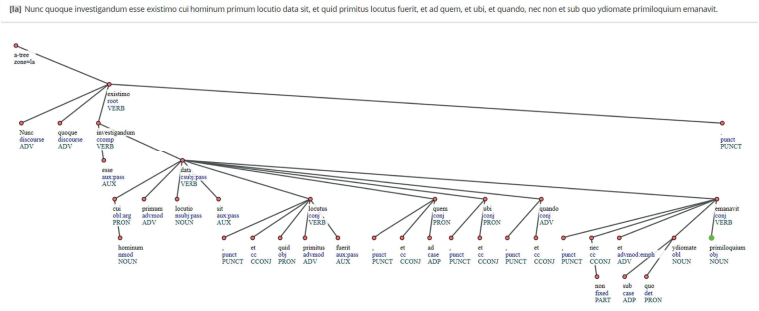
*DanteSearch* è dunque un sistema aperto. E per indicare in quali direzioni è attualmente aperto, indicherò tre progetti di ricerca con i quali collabora, con reciproca utilità, e la propria attuale prospettiva di sviluppo in direzione del web semantico.

I due primi progetti con cui *DanteSearch* collabora sono il *Vocabolario Dantesco* (http://www.vocabolariodantesco.it/), impresa congiunta dell'Accademia della Crusca e dell'Istituto CNR OVI – Opera del Vocabolario Italiano diretta da Paola Manni e Lino Leonardi, e il *Vocabolario Dantesco Latino* (http://www.vocabolariodantescolatino.it/), progetto parallelo e strettamente collegato al *Vocabolario Dantesco*, realizzato, oltre che dalla Crusca e dall'OVI, dalla Fondazione Franceschini e dalla SISMEL, dalla Società Dantesca Italiana, dall'ISTI-CNR e dal Dip. di Filologia Letteratura e Linguistica dell'Università di Pisa, e coordinato da Gabriella Albanese. I due progetti hanno il fine di arrivare a dare una rappresentazione lessicografica completa della cultura bilingue di Dante, e *DanteSearch* costituisce un ovvio strumento di lavoro quotidiano dei redattori di entrambi.

Un terzo progetto con cui *DanteSearch* collabora è il progetto ERC *LiLa-Linking Latin* (https://lila-erc.eu/#page-top), Principal Investigator Marco Passarotti, che si propone di «connect and ultimately exploit the wealth of linguistic resources and NLP tools for Latin created so far», in ottica di web semantico. *Vocabolario Dantesco Latino* e *DanteSearch* collaborano con *LiLa* alla codifica sintattica delle opere latine di Dante secondo lo standard del progetto mondiale Universal Dependencies (https://universaldependencies.org/). Ecco per esempio, alla Fig. 11, la rappresentazione sintattica di una frase del *De vulgari eloquentia*, codificata da Giulia Pedonese (la *Monarchia* è

codificata da Federica Favero, le *Epistole* da Elena Vagnoni, le *Egloge* da Veronica Dadà).

Figura 11 - *Codifica sintattica di* De vulgari eloquentia*, I iv 1*

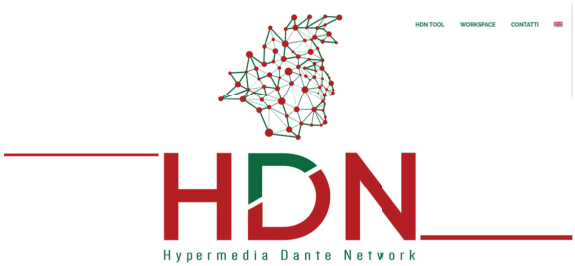



La codifica sintattica delle opere latine realizzata secondo questa grammatica a dipendenze integra così la codifica sintattica delle opere volgari secondo la grammatica a costituenti inserita in *DanteSearch*.

Lo sviluppo di DanteSearch in direzione del web semantico si realizza nell'ambito del PRIN *HDN-Hypermedia Dante Network* (https://hdn.dantenetwork.it/), Principal Investigator Michelangelo Zaccarello, professore di Filologia italiana dell'Università di Pisa; responsabile della realizzazione informatica Carlo Meghini, dirigente di ricerca dell'ISTI-CNR (Fig. 12).

Questa realizzazione informatica, in sintonia con il progetto *LiLa*, ha preso il nome di *LiDa-Linking Dante*, con l'obiettivo di digitalizzare le opere dantesche e la conoscenza a esse relativa, con ciò intendendosi anzitutto i commenti, e in particolare il secolare commento alla *Commedia*, creando una biblioteca digitale che rispetti i principi FAIR-Findable, Accessible, Interoperable and Reusable (https://www.go-fair.org/fair-principles/), e su cui si possano formulare e testare ipotesi scientifiche, eventualmente anche attraverso agenti digitali (Bartalesi-Meghini-Andriani-Tavoni 2015).

Figura 12 - *Homepage del progetto HDN-Hypermedia Dante Netwo*



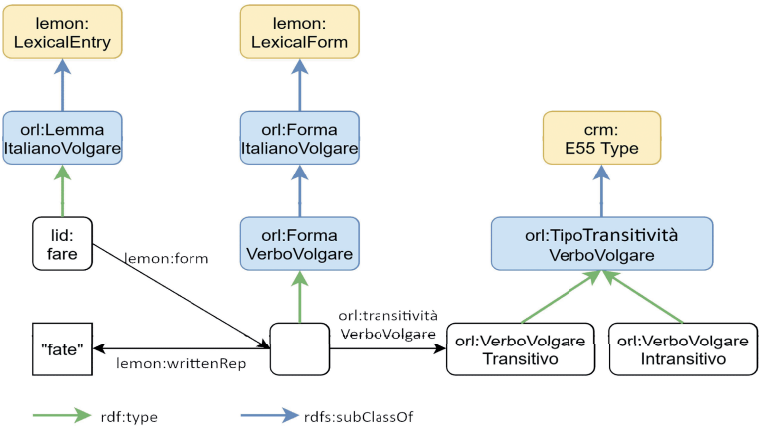

A questo punto bisogna almeno nominare il progetto *DanteSources* (https://dantesources.dantenetwork.it/), base di conoscenza sulle fonti citate nelle opere di Dante, che negli anni scorsi abbiamo sviluppato in parallelo a *DanteSearch* (vedi Bartalesi-Meghini-Metilli-Andriani-Tavoni 2017, Tavoni-Andriani-Meghini-Bartalesi-Metilli 2017). L'ambizioso obiettivo di Carlo Meghini e dei suoi collaboratori presso l'ISTI-CNR (Cesare Concordia, Chiara Paolini, Daniele Metilli, Luca Trupiano) è ora arrivare a creare una unica base di conoscenza aperta e sempre implementabile nella quale confluiscano sia *DanteSearch* sia *DanteSources*: cioè sia l'informazione linguistica (morfologica, sintattica, ma anche retorica) sia tutta l'informazione di diversissima natura sulle fonti e su quant'altro registrato dai commenti, adottando il paradigma del web semantico, e con ciò uscendo dai limiti, notati prima, della codifica in XML, perché l'eXtensible Markup Language (XML) non è adeguato alla doppia integrazione delle opere con i commenti e dei commenti tra di loro, dal momento che si fonda su una singola struttura gerarchica.

*LiDa*, dunque, farà uso di logiche calcolabili, che usano il Resource Description Framework (RDF) come linguaggio di base e ruotano intorno all'Ontology Web Language (OWL) per la codifica formale della conoscenza in ambiente web. L'uso di queste logiche permette di superare il problema della singola struttura gerarchica posto dall'XML, e di rappresentare così in modo formale non solo il testo dantesco, ma anche la sua esegesi e i legami che testo e commenti hanno fra loro.

In particolare, la reimplementazione di *DanteSearch* muove dal fatto che la lemmatizzazione e la marcatura morfologica non sono fruibili all'interno della ricerca sintattica, dato che XML non permet-

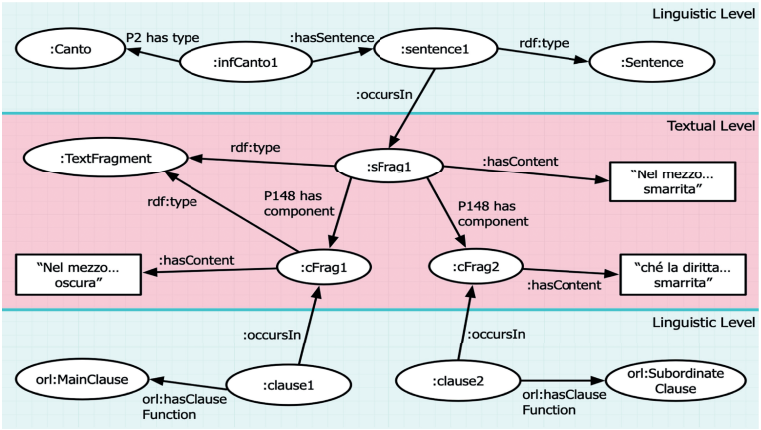
te di sovrapporre le due marcature, e si propone di rendere la lemmatizzazione fruibile non solo all'interno della ricerca sintattica ma anche di qualsiasi altra ricerca che riguardi un aspetto affrontato dal progetto *LiDa*. Per raggiungere questo scopo, le due marcature alla base di *DanteSearch* si trasformano in un unico grafo RDF, quindi in un insieme di triple, conformi a corrispondenti ontologie: un'ontologia morfologica e un'ontologia sintattica. La Fig. 13 rappresenta l'ontologia morfologica, e la Fig. 14 rappresenta l'ontologia sintattica.

Figura 13 - *L'ontologia morfologica del nuovo* DanteSearch




La trasformazione richiede dunque il prioritario sviluppo di dette ontologie, e successivamente lo sviluppo di un algoritmo di parsing dell'XML e di successiva emissione delle triple. Al momento, siamo in fase di testing dell'ontologia sintattica e di messa a punto dell'algoritmo di trasformazione della marcatura sintattica, mentre l'ontologia morfologica e la trasformazione della corrispondente marcatura è stata completata.

Figura 14 - *L'ontologia sintattica del nuovo* DanteSearch




Una volta completata anche la trasformazione della marcatura sintattica sarà possibile reimplementare le operazioni offerte dall'interfaccia *DanteSearch* come interrogazioni SPARQL al grafo risultante, sostituendo il back-end attuale di *DanteSearch* con il nuovo back-end. In questo modo, gli utenti di *DanteSearch* potranno continuare a usare la stessa interfaccia avendo in più la possibilità di combinare in modo arbitrario i due criteri di ricerca.

Queste parole di Carlo Meghini e collaboratori, che vi ho riportato, oltrepassano i limiti delle mie competenze. Hanno per me il sapore della Teoria del Tutto, che io conosco non da Stephen Hawking ma da Sheldon Cooper. E questo, in realtà, mi dà un grande senso di appagamento, perché è arrivato il momento in cui giovani informatici umanisti, che tu hai contribuito a formare, ti hanno superato, e questo ti dà la certezza che il tuo passaggio di alcuni decenni dall'Università non è stato inoperoso e non è stato inutile.

## *Riferimenti bibliografici*

Bartalesi, Valentina & Meghini, Carlo & Andriani, Paola & Tavoni, Mirko. 2015. Towards a Semantic Network of Dante's Works and Their Contextual Knowledge. *Digital Scholarship in the Humanities*, 30(1). 28-35.

Bartalesi, Valentina & Meghini, Carlo & Metilli, Daniele & Andriani, Paola & Tavoni, Mirko. 2017. DanteSources: a Digital Library for Studying Dante Alighieri's Primary Sources. *Umanistica Digitale* 1. 119-128.

D'Amico, Marta (a cura di). 2015. *Sintassi dell'italiano antico e sintassi di Dante. Atti del seminario di studi, Pisa, 15-16 ottobre 2011*. Pisa: Felici.

D'Amico, Marta. 2009. *La sintassi dell'aldilà. Studio sulla sintassi periodale dei discorsi diretti delle anime della* Commedia *di Dante.* Università di Pisa. (Tesi di laurea specialistica in Lingua e letteratura italiana.)

D'Amico, Marta. 2015. Le interrogative dirette non canoniche nei discorsi dei personaggi della *Commedia*: pragmatica e testualità. In D'Amico, Marta (a cura di), *Sintassi dell'italiano antico e sintassi di Dante. Atti del seminario di studi (Pisa 15/16 ottobre 2011)*, 125-139. Pisa: Felici.

Gigli, Sara. 2003. *Le proposizioni consecutive nella* Commedia*: osservazioni stilistiche. In Battaglia Ricci, Lucia (a cura di), Leggere Dante,* 329-344. Ravenna: Longo.

Gigli, Sara. 2004. *Codifica sintattica della* Commedia *dantesca*, Università di Pisa, Scuola di dottorato in Studi italianistici.

Gigli, Sara. 2007. Le subordinate concessive nella *Commedia* dantesca, *Studi Linguistici Italiani*, XXXIII(2). 161-190.

Gigli, Sara. 2015. *La codifica sintattica della* Commedia *di Dante*, in D'Amico (a cura di), *Sintassi dell'italiano antico e sintassi di Dante. Atti del seminario di studi (Pisa 15/16 ottobre 2011)*, 81-95. Pisa: Felici.

Meghini, Carlo & Tavoni, Mirko & Zaccarello, Michelangelo. 2021. Mapping the Knowledge of Dante Commentaries in the Digital Context: A Web Ontology Approach. *Romanic Review*, 112/1 (*The Pleasure of Dante's Text / Il piacere del testo dantesco*, H. Wayne Storey, Guest Editor). 138-157.

Renzi, Lorenzo & Salvi, Giampaolo & Cardinaletti, Anna (a cura di). 1988-1995. *Grande grammatica italiana di consultazione*, 3 voll., Bologna: il Mulino.

Salvi, Giampaolo & Renzi, Lorenzo (a cura di). 2010. *Grammatica dell'italiano antico*, Bologna: il Mulino.

Tavoni, Mirko & Andriani, Paola & Meghini, Carlo & Bartalesi, Valentina & Metilli, Daniele. 2017. L'esplorazione delle fonti dantesche attraverso la biblioteca digitale *DanteSources*. In Persico, Thomas & Viel, Riccardo (a cura di), *Sulle tracce del Dante minore. Prospettive di ricerca per lo studio delle fonti dantesche*, 29-52. Bergamo: Sestante Edizioni.

Tavoni, Mirko. 2002. Contributo sintattico al 'disdegno' di Guido (*If* X 61-63). Con una nota sulla grammaticalità e la leggibilità dei classici, *Nuova Rivista di Letteratura Italiana*, V(1). 51-80.

Tavoni, Mirko. 2005. Un nuovo strumento informatico per lo studio di Dante (con una proposta interpretativa per *Inf.* IV 69). In De Matteis, Giuseppe (a cura di), *Dante in lettura*, 217-229. Ravenna: Longo.

Tavoni, Mirko. 2011. *DanteSearch*: il corpus delle opere volgari e latine di Dante lemmatizzate con marcatura grammaticale e sintattica. In Cerbo, Anna (a cura di), *Lectura Dantis 2002-2009. Omaggio a Vincenzo Placella per i suoi settanta anni*, t. II: *Lectura Dantis 2004 e 2005*: 583-608. Napoli: Università degli Studi di Napoli L'Orientale – Officine Grafico-Editoriali "il Torcoliere".

Tavoni, Mirko. 2015. *DanteSearch*: istruzioni per l'uso. Interrogazione morfologica e sintattica delle opere volgari e latine di Dante. In D'Amico (a cura di), *Sintassi dell'italiano antico e sintassi di Dante. Atti del seminario di studi (Pisa 15/16 ottobre 2011)*, 59-79. Pisa: Felici.

Tavoni, Mirko. 2020a. Lingua parlata e lingua scritta in Dante: appunti metalinguistici e linguistici. In Orletti, Franca & Albano Leoni, Federico (a cura di), *L'antinomia scritto/parlato*, 89-115. Città di Castello: I libri di Emil.

Tavoni, Mirko. 2020b. Lessicografia ed esegesi dantesca. In Manni, Paola (a cura di), *«S'i'ho ben la tua parola intesa». Atti della giornata di presentazione del* Vocabolario dantesco. *Firenze, Villa Medicea di Castello, 1° ottobre 2018*: 157-168. Firenze: Accademia della Crusca (Quaderni degli *Studi di lessicografia italiana*).