

«La sua chiarezza séguita l'ardore»

**Studi di linguistica e filologia
offerti a Paola Manni**

Tre implementazioni informatiche per il *Vocabolario Dantesco* e alcune considerazioni a margine*

Salvatore Arcidiacono

A oltre cinque anni di distanza dalla prima presentazione del *Vocabolario Dantesco* (VD)¹, occasione nella quale apparvero già chiari a molti i meriti del progetto diretto da Paola Manni e Lino Leonardi², è possibile oggi riconoscere a quest'opera una precisa collocazione nel panorama della lessicografia elettronica in Italia, determinata, col senno di poi, anche dalla recente nascita di altre imprese che del VD hanno seguito l'impostazione tecnica. Il *Vocabolario Dantesco* è infatti stato il primo di una serie di vocabolari elettronici ad affidare la propria realizzazione alla piattaforma *online* di redazione e di pubblicazione Lexicad³: una

* I tre gruppi di funzioni che qui si presentano sono stati concepiti con la volontà di associare all'offerta di questo contributo alcune nuove applicazioni per la piattaforma del VD. Dopo aver avviato la scrittura del codice, è apparso subito chiaro che sarebbe stato pressoché impossibile conciliare la pubblicazione a stampa con il consueto ciclo di sviluppo, test e rilascio del *software*. Inoltre, urgenti esigenze redazionali hanno richiesto l'immediata installazione di alcuni aggiornamenti, cosicché, a dispetto delle intenzioni originarie, che qui mi preme quantomeno dichiarare, i gruppi di funzioni descritti di seguito sono stati resi disponibili in rete ben prima della stampa di queste pagine.

1 <http://www.vocabolariodantesco.it>

2 «*S'ì ho ben la parola tua intesa*», Atti della giornata di presentazione del *Vocabolario Dantesco*, Firenze, Accademia della Crusca, Villa Medicea di Castello, 1° ottobre 2018, a cura di Paola Manni, Firenze, Accademia della Crusca, 2020.

3 Nel caso specifico del VD, si tratta di una versione personalizzata del *software* Lexicad-Pluto (*Piattaforma Lessicografica del Tesoro delle Origini*), sviluppato da chi scri-

simile procedura è stata poi successivamente utilizzata, oltre che dal TLIO⁴, che ha inizialmente promosso lo sviluppo della piattaforma, anche dal VDL, dal VEV⁵ e dal VSM⁶; a questi dizionari si aggiungerà a breve, nell'ambito del progetto QM⁷, il DESN⁸.

La tentazione di provare a inquadrare questa risorsa in una ideale mappa di progetti digitali su Dante è molto forte: del resto, la storia dell'informatica umanistica – una disciplina che dimostra un incessante bisogno di riflettere sulla propria identità, costruendola anche attorno alle proprie origini – si potrebbe rileggere attraverso una precisa serie di tappe marcate dalle imprese dedicate all'opera dantesca.

La nascita delle *digital humanities*, come molto spesso piace ricordare, è certamente da rintracciare nella realizzazione dell'*Index Thomisticus* a opera di padre Roberto Busa S. J.⁹. Tuttavia, ciò può dirsi

ve presso l'Istituto CNR - Opera del Vocabolario Italiano (<http://www.ovi.cnr.it/Il-Software.html#pluto>).

- 4 *Tesoro della Lingua Italiana delle Origini*, fondato da Pietro G. Beltrami, poi diretto da Lino Leonardi, oggi diretto da Paolo Squillaciotti, consultabile in rete all'indirizzo <http://tlio.ovi.cnr.it>.
- 5 *Vocabolario storico-etimologico del veneziano*, diretto da Lorenzo Tomasin e Luca D'Onghia, consultabile in rete all'indirizzo <http://vev.ovi.cnr.it>.
- 6 *Vocabolario del Siciliano Medievale*, diretto da Mario Pagano, consultabile in rete all'indirizzo <http://artesia.unict.it/vsm>.
- 7 Il progetto QM - *Quattrocento Meridionale. Il futuro dell'italiano antico* (PRIN 2020) si propone di avviare una continuazione modulare della ricerca lessicografica dell'O-VI, rimediando alla sottorappresentazione del Mezzogiorno continentale nel Corpus TLIO (<http://tlioweb.ovi.cnr.it>), dovuta a una documentazione volgare che per alcune aree meridionali è significativa solo dopo il *terminus ad quem* del corpus (1375). Oltre all'unità principale dell'Opera del Vocabolario Italiano (P.I. Pär Larson), partecipano al progetto l'Università degli Studi di Napoli Federico II, coordinata da Nicola De Blasi, e l'Università degli Studi di Catania, coordinata da chi scrive.
- 8 *Dizionario Etimologico e Storico del Napoletano*, diretto da Nicola De Blasi e Francesco Montuori. Le prime voci del DESN sono state appena pubblicate in *Voci dal DESN - Dizionario Etimologico e Storico del Napoletano*, a cura di Nicola De Blasi e Francesco Montuori, Firenze, Cesati, 2023.
- 9 *Index Thomisticus*, a cura di Roberto Busa, Stuttgart-Bad Cannstatt, Frommann-Holzboog, 1974-1980, 56 voll.

vero solo per la definizione e la preparazione del progetto, che risale al 1946¹⁰, ma non trova riscontro nelle prime sperimentazioni concrete, né, tantomeno, nella data di pubblicazione dei risultati: a ben vedere, i primi saggi di spoglio automatico del gesuita, effettuati con macchine elettromeccaniche che operavano su schede perforate, non hanno riguardato san Tommaso ma i versi di Dante; i risultati di queste sperimentazioni sono stati resi noti da Busa già nel 1951¹¹, in quella che può ritenersi la prima discussione metodologica sull'automazione del trattamento del testo (la stampa in volume dell'*Index* verrà invece avviata solo nel 1974)¹². Negli stessi anni, tra l'altro, all'interno del laboratorio di padre Busa – il Centro per l'Automazione dell'Analisi Linguistica (CAAL) di Gallarate – stavano anche prendendo forma le procedure e i programmi che furono impiegati negli spogli per l'*Enciclopedia Dantesca*¹³.

Quando nel 1956 le attività del CAAL confluirono nel CNUCE di Pisa (Centro Nazionale Universitario di Calcolo Elettronico), l'inaugurazione del nuovo Centro fu impreziosita dalla presentazione delle prestigiose concordanze della *Divina Commedia* a cura di Carlo Tagliavini¹⁴; la

- 10 ROBERTO BUSA, *The Annals of Humanities Computing. The Index Thomisticus*, in «Computers and the Humanities», XIV, 1980, pp. 83-90: 83.
- 11 ID., *Sancti Thomae Aquinatis hymnorum ritualium varia specimina concordantiarum. Primo saggio di indici di parole automaticamente composti e stampati da macchine IBM a schede perforate*, Milano, Bocca, 1951.
- 12 Come si afferma in JULIANNE NYHAN, MARCO PASSAROTTI, *One Origin of Digital Humanities. Fr. Roberto Busa in His Own Words*, Basel, Springer, 2019, è il lavoro del 1951, piuttosto che l'*Index Thomisticus*, a rendere Roberto Busa il padre ideale dell'informatica umanistica.
- 13 ANTONIO ZAMPOLLI, *Le principali attività dell'Istituto di Linguistica Computazionale: il punto di vista del Direttore*, in *Computational Linguistics in Pisa – Linguistica Computazionale a Pisa (Linguistica Computazionale, Special Issue, 16-17)*, a cura di Antonio Zampolli, Nicoletta Calzolari e Laura Cignoni, Pisa-Roma, IEPI, pp. XVII-LXX: XV.
- 14 *La Divina Commedia, testo, concordanze, lessici, rimario, indici*, prefazione di Carlo Tagliavini, nota tecnica di Giorgio Caldara, Milano, IBM Italia, 1965. La scelta della *Commedia* rientra anche nelle celebrazioni del centenario dantesco che ricorreva quell'anno. Le concordanze di Tagliavini vennero pubblicate a pochi mesi di distanza dall'indice di frequenza della *Commedia* elaborato con metodi mecano-

portata scientifica di quest'opera andava ben oltre il semplice intento celebrativo, perché costituì la necessaria premessa alla realizzazione del *Lessico di frequenza della lingua italiana contemporanea*¹⁵ (anche questa affidata a Tagliavini dalla IBM) che, sul modello dei dizionari di frequenza realizzati negli anni Sessanta da Alphonse Juilland presso il MIT, riuscì a colmare gli evidenti limiti della lista di Knease¹⁶, primo indice di frequenza generale per l'italiano, nonché unico all'epoca disponibile, ma allestito con «fini didattici e [...] criteri incerti e discutibili»¹⁷.

In Italia, in questa fase storica caratterizzata dalla sperimentazione sui nuovi sistemi elettronici, si osserva una certa compattezza nelle ricerche, che sembrano più sensibili alla celebre esortazione alle concordanze di Contini¹⁸ che non all'egemonia generativista rintracciabile

grafici presso l'Università di Utrecht in collaborazione con l'Accademia della Crusca (cfr. MARIO ALINEI, *La lista di frequenza della Divina Commedia*, in *Miscellanea dantesca*, edita dalla Società Dantesca dei Paesi Bassi, Utrecht-Antwerpen, Het Spectrum, pp. 138-270); questo lavoro rientra nel progetto per un *Inventario Linguistico dell'Italiano delle Origini*, che successivamente fornirà le prime concordanze della *Commedia* sull'edizione Petrocchi (ID., *Spogli Elettronici dell'Italiano delle Origini e del Duecento*, II. 5. Dante Alighieri. *La Commedia*, Bologna, il Mulino, 1971).

- 15** UMBERTA BORTOLINI, CARLO TAGLIAVINI, ANTONIO ZAMPOLLI, *LIF - Lessico di frequenza della lingua italiana contemporanea*, Milano, IBM Italia, 1971.
- 16** TACIE MARY KNEASE, *An Italian Word List From Literary Sources*, Toronto, The University of Toronto Press, 1933.
- 17** TULLIO DE MAURO, *Alcuni aspetti quantitativi della lingua della Commedia*, in Atti del Convegno di studi su Dante e la Magna Curia, Palermo, Catania, Messina, 7-11 novembre 1965, a cura del Centro di studi filologici e linguistici siciliani, Palermo, Centro di studi filologici e linguistici siciliani, 1967, pp. 519-525: 519. Cfr. anche ID., *La Commedia e il Vocabolario di base dell'italiano*, in *La funzione Dante e i paradigmi della modernità*, Atti del XVI Convegno internazionale della MOD LUMSA, Roma, 10-13 giugno 2014, a cura di Patrizia Bertini Malgarini, Nicola Merola e Caterina Verbaro, Pisa, ETS, 2015, pp. 17-24: 17.
- 18** Apparsa per la prima volta nel 1951 sulla rivista «Paragone», poi in GIANFRANCO CONTINI, *Preliminari sulla lingua del Petrarca*, in ID., *Varianti e altra linguistica. Una raccolta di saggi (1938-1968)*, Torino, Einaudi, pp. 169-192. Come noto, Contini ripeté l'esortazione nel 1960, in occasione della relazione tenuta al Convegno di studi di Filologia italiana di Bologna, per la quale cfr. ID., *Esperienze d'un antologista del Due-*

altrove; è evidente, inoltre, una strettissima connessione disciplinare tra filologia digitale, linguistica computazionale e lessicografia elettronica, supportata da un'altrettanto fitta rete di collaborazioni tra i maggiori attori di quel periodo, in cui l'Accademia della Crusca, presso la quale si stavano preparando gli spogli lessicali per il *Vocabolario storico*, rappresentava uno dei nodi principali¹⁹.

Limitatamente all'Italia, il crepuscolo della stagione delle concordanze può essere fatto coincidere con la pubblicazione della celebre *Letteratura Italiana Zanichelli* (LIZ)²⁰. Anche per Dante si può percepire il cambio di paradigma: come osserva lucidamente Claudio Marazzini nella sua recensione²¹, la svolta rappresentata dalla LIZ non riguarda solo gli autori che non hanno ricevuto le dovute attenzioni in precedenza, perché le possibilità offerte dal mezzo elettronico migliorano radicalmente le modalità di consultazione rispetto alle concordanze a stampa pubblicate nel periodo precedente. Il motore che ha reso pos-

cento poetico italiano, in ID., *Frammenti di Filologia romanza. Scritti di ecdotica e linguistica* (1932-1989), a cura di Giancarlo Breschi, Firenze, SISMEL-Edizioni del Galluzzo, I, pp. 155-187.

- 19** L'Accademia si propone di stringere i rapporti con il CAAL e con l'IBM nel 1961; nel 1964 viene formalizzata una collaborazione con il progetto SEIOD - *Spogli Elettronici dell'Italiano delle Origini e del Duecento* di Mario Alinei; nel 1965 il Consiglio direttivo decide di prendere accordi con il CNUCE, e ufficializza una collaborazione con Antonio Zampolli, da tempo prezioso elemento di raccordo tra l'Accademia e il CAAL. Per il progetto digitale del *Vocabolario storico della lingua italiana* e la continuità con la lessicografia digitale di oggi si rimanda a SALVATORE ARCIDIACONO, *Lessicografia elettronica e italiano delle origini*, Palermo, Centro di studi filologici e linguistici siciliani, 2022.
- 20** Su questa periodizzazione concordano, per esempio, PIETRO BELTRAMI e ANDREA BOCCELLARI, *Banche dati e dizionari on-line. Il Tesoro della Lingua Italiana delle Origini e la banca dati dell'italiano antico dell'Opera del Vocabolario Italiano*, in *Nuovi media e lessicografia storica*, Atti del colloquio in occasione del settantesimo compleanno di Max Pfister, a cura di Wolfgang Schweickard, Tübingen, Max Niemeyer, pp. 3-14: 3, e ROSARIO COLUCCIA, *Strumenti informatici e riflessi sul lavoro del lessicografo*, in *Nuovi media e lessicografia storica*, cit., pp. 59-70: 59.
- 21** CLAUDIO MARAZZINI, recensione a: *LIZ - Letteratura Italiana Zanichelli*, CD-Rom dei testi della letteratura italiana, a cura di Pasquale Stoppelli ed Eugenio Picchi, in «Lettere italiane», XLVII, 2, 1995, pp. 312-319.

sibili questi risultati è stato il performante DBT - *Database Testuale*, sviluppato da Eugenio Picchi presso l'ILC-CNR. Su DBT furono indicizzati anche i *Commenti danteschi* in CD-Rom a cura di Paolo Procaccioli²², ma, a quest'altezza cronologica, le risorse digitali su Dante erano già da tempo giunte a un evidente livello di maturità con le iniziative di Robert Hollander, il *Dartmouth Dante Project* (DDP)²³ e il *Princeton Dante Project* (PDP)²⁴, che tutt'oggi rappresentano due risorse di riferimento.

In linea generale, è possibile osservare che, sul finire del Novecento, le discipline genericamente afferenti all'informatica umanistica avevano intrapreso un elevato livello di specializzazione e percorsi divergenti, mentre i grandi pionieri della generazione precedente avevano consolidato le proprie linee di ricerca autonome. Da questo punto in poi, grazie anche alla nascita del *World Wide Web*, le iniziative digitali dedicate a Dante si moltiplicano ulteriormente; resistendo alla tentazione di intraprendere un censimento impossibile, si menzioneranno almeno le iniziative della Società Dantesca Italiana²⁵ e l'*Hypermedia Dante Network* (HDN)²⁶, che si propone di perfezionare ed estendere alla *Commedia* i risultati del precedente progetto *Per una enciclopedia dantesca digitale*²⁷, che, con la biblioteca digitale *Dante Sources*, è stato premiato ai DH Awards nel 2015 nella categoria *Best DH Tool or Suite of Tools*²⁸. Sul versante della filologia digitale, l'edizione della *Commedia*

22 *I commenti danteschi dei secoli XIV, XV e XVI*, a cura di Paolo Procaccioli, Roma, Lexis Progetti Editoriali, 1999.

23 <https://dante.dartmouth.edu>.

24 <https://dante.princeton.edu>.

25 Per le quali si rinvia ad ANDREA BOZZI, *I progetti digitali della Società Dantesca Italiana*, in «Griseldaonline», xx, 2, 2021, pp. 21-29.

26 <https://hdn.dantenetwork.it>.

27 È da ricondurre a questo progetto anche *DanteSearch*, diretto da Mirko Tavoni e accessibile all'indirizzo <https://dantesearch.dantenetwork.it>. *DanteSearch* è utilizzato come corpus di riferimento dal *Vocabolario Dantesco Latino* grazie a un sistema *software* in grado di trasferire automaticamente i risultati del motore di ricerca sulla piattaforma lessicografica basata su Lexicad.

28 <http://dhawards.org/dhawards2015/results>.

di Prue Shaw e Peter Robinson è stata recentemente oggetto di una migrazione sul *web*²⁹, segnalata tra le edizioni più rilevanti del noto *Catalogue of Digital Scholarly Editions* di Patrick Sahle³⁰.

Il VD aggiunge un vocabolario elettronico a questo vivace panorama, riunendo l'Accademia della Crusca e l'Istituto CNR - Opera del Vocabolario Italiano (OVI) in una nuova impresa digitale. Come anticipato, il VD è il primo progetto a mettere in esercizio il sistema Lexicad / Pluto, nella specifica implementazione denominata PlutoVD³¹. Nella sua prima versione, il sistema poteva dirsi perfettamente compiuto per quanto concerne la maschera di redazione (l'interfaccia che utilizzano i redattori per redigere le voci) e il *template* "di base" di presentazione delle voci *online*, ispirato all'interfaccia del TLIO e dotato di tutte le funzioni di questa. Il livello di sofisticazione della parte dedicata alla consultazione delle voci era invece molto più elementare rispetto a quello raggiunto nell'interfaccia di gestione dedicata ai redattori; del resto, la priorità, in quel momento, era la redazione del vocabolario e la quantità di dati immessi nel sistema non avrebbe comunque permesso l'attivazione di modalità di esplorazione e ricerca più sofisticate. I tre nuovi "pacchetti" di funzioni che si descriveranno di seguito – due dedicati al lemmario e uno alla visualizzazione della scheda – sono finalizzati a fornire le prime necessarie funzioni di filtro, ricerca e organizzazione dei dati.

Con il progressivo incremento delle voci redatte, il primo punto a richiedere un intervento era certamente il lemmario (accessibile dalla sezione *Il Vocabolario > Schede pubblicate*). L'elenco con le prime schede pubblicate *online*, date le poche entrate inizialmente disponibili, era

29 L'edizione digitale della *Commedia* (DANTE ALIGHIERI, *Commedia. A Digital Edition*, edited by Prue Shaw and Peter Robinson, Saskatoon-Firenze, Inkless Editions-Fondazione Ezio Franceschini, 2010) pubblicata per la prima volta in DVD-Rom, è stata recentemente messa *online* in una versione ampliata all'indirizzo <https://dante.commedia.it>.

30 <http://digitale-edition.de>.

31 Le caratteristiche di base della piattaforma sono state fornite in SALVATORE ARCIDIACONO, «Forse tu non pensavi ch'io löico fossi!»: metodi computazionali al servizio del VD, in «S'ì ho ben la parola tua intesa», cit., pp. 81-92.

perfino sprovvisto di un algoritmo per suddividere in diverse schermate le voci (la cosiddetta paginazione dei risultati, finalmente attiva a partire da questo aggiornamento). Il lemmario sollecita inoltre alcune considerazioni metalessicografiche: la macrostruttura di un dizionario può essere considerata, in prima istanza, un dispositivo di accesso alle voci, ragione per la quale rientra di diritto tra i cosiddetti elementi *medium-dependent*³²: tavole e indici dei dizionari a stampa vanno intesi, in questo senso, come espedienti messi in atto dal libro per permettere strategie di accesso alternative rispetto a un'organizzazione principale. Nei dizionari elettronici, la macrostruttura è un elemento totalmente dinamico che non esiste in quanto tale nel *database* ma viene rigenerato, in accordo ai parametri impostati nella richiesta, ricavando le informazioni in tempo reale dagli elementi microstrutturali codificati nella banca dati (il campo *lemma*, il campo *varianti*, il campo *rinvio* ecc.; in altre parole, in termini di dato codificato, esiste solo la microstruttura).

Sulla base di questi principi sono stati predisposti alcuni selettori visualizzati sulla destra dell'elenco delle voci (il pannello *Ricerca e filtra*)³³. Il campo *Filtra per lemma* permette di effettuare una ricerca sul lemmario, a partire da una chiave di ricerca immessa dall'utente, suscettibile di essere ulteriormente raffinata per mezzo del menu a tendina associato al campo; la chiave di ricerca può essere così ricercata come sottostringa di caratteri contenuta dal lemma (opzione *Contiene*), come *Stringa esatta* o per specificare le lettere iniziali e finali che compongono il lemma (*Inizia per* o *Finisce per*).

Proseguendo dall'alto verso il basso, la tendina *Ordina* permette all'utente di intervenire sulla visualizzazione dei risultati, secondo un ordine alfabetico ascendente o discendente.

I due selettori successivi permettono di filtrare le entrate del VD, intese come insieme di tutti i *record* presenti nella tabella delle voci nella base di dati, discriminando tra diversi tipi di elementi. Per chiarire

32 ROBERT LEW, AGNIESZKA SZAROWSKA, *Evaluating Online Bilingual Dictionaries. The Case of Popular Free English-Polish Dictionaries*, in «ReCALL», 29, 2017, pp. 138-159: 139.

33 <http://www.vocabolariodantesco.it/lemmario.php>.

meglio questo concetto si partirà dal secondo controllo (*Mostra rinvii / Nascondi*) che permette di intervenire su entrate che rinviano a un'altra (per esempio, *aborrire* v. rinvia ad *abborrare* v.). Sulla piattaforma Lexicad il rinvio è stato concepito come una voce al pari di tutte le altre: per fornire qualche dettaglio tecnico, il rinvio viene archiviato nella tabella delle voci, il suo oggetto *software* viene istanziato ricorrendo alla classe "voce" e dispone di una scheda completa nella maschera di redazione del tutto identica a quella di una normale voce. La differenza sostanziale nei comportamenti tra voce e rinvio viene determinata registrando, in un punto dedicato della microstruttura, l'identificativo univoco di una voce alla quale si intende rinviare; se il campo *rinvio* viene popolato con una "chiave esterna", il sistema interpreta il *record* come un'entrata di rinvio. Da questo momento, tutte le informazioni presenti nella scheda vengono considerate rilevanti esclusivamente ai fini della redazione (e diventano quindi accessibili solo dall'interfaccia dedicata ai redattori); nelle entrate sui lemmari destinati all'utente finale, invece, il *link* alla voce verrà disattivato e il sistema provvederà a generare un collegamento cliccabile al lemma *target*. Con questo nuovo selettore il lettore può intervenire sulla visualizzazione nel lemmario del VD e decidere di visualizzare solo le voci normali escludendo i rinvii³⁴.

Il comportamento del primo selettore è leggermente più complesso, perché riguarda la gestione delle varianti. Il *Vocabolario Dantesco*, come noto, tiene conto anche di una selezione della *varia lectio* ricavata dagli apparati delle edizioni di riferimento; in particolare, «la redazione del VD valuta e accoglie nel trattamento lessicografico solo le varianti sostanziali (si potrebbe dire, lessicali) e, in subordine, le eventuali forme che di esse l'apparato Petrocchi registra, con le specifiche che lo stesso editore ha a suo tempo limpidamente dichiarato»³⁵.

34 Questa funzione è disponibile anche sui lemmari principali del TLIO, del VSM e del VEV.

35 FRANCESCA DE BLASI, BARBARA FANINI, CRISTIANO LORENZI BIONDI, VERONICA RICOTTA, *Nell'officina del VD: gli strumenti e il lavoro di redazione*, in «S'i ho ben la parola tua intesa», cit., pp. 17-80: 59. Al momento dell'avvio del progetto, l'apparato di Petrocchi è stato scelto perché più ricco di varianti.

Come si legge nell'Introduzione del VD (§ 2)³⁶, l'apparato Petrocchi si affianca al Corpus VD nello spoglio, limitatamente alle «varianti potenzialmente adiafore, ovvero le lezioni relegate in apparato ma ammissibili al posto di quelle messe a testo nell'edizione Petrocchi» (criterio di selezione filologico) e che siano latrici di un lemma o di un significato non altrimenti attestato nel testo dell'edizione Petrocchi (criterio di selezione linguistico). Sulla base di quest'ultimo criterio, l'apporto delle varianti può quindi concretizzarsi in due modi:

1. le varianti possono dare luogo a un lemma non altrimenti attestato nel VD: questo tipo di variante viene chiamata «variante-nuovo lemma», in sigla [var.]. La «variante-nuovo lemma» viene inserita come una nuova voce ed è segnalata nelle entrate dei lemmari con la sigla [var.] seguita dal rinvio al lemma di cui è lezione alternativa nell'edizione Petrocchi (per esempio, *addentrare* v. [var.] → *addentare* v., *addolcare* v. [var.] → *addolciare* v., ecc.);
2. le varianti possono contribuire all'informazione lessicale dal punto di vista semantico o grammaticale: in questo caso si parla di «variante-nuova accezione»³⁷.

In entrambi i casi, il rapporto tra i due lemmi costituisce una relazione di tipo multi-a-molti, perché, chiaramente, un lemma a testo può avere più varianti in apparato, e la stessa variante può occorrere più volte in corrispondenza di diversi lemmi a testo. Per ciascuna relazione, PlutoVD registra la forma della variante, il luogo in cui occorre il lemma a testo della relazione, i manoscritti e le eventuali edizioni che accolgono la variante a testo.

³⁶ <http://www.vocabolariodantesco.it/introduzione.php>.

³⁷ «Alla prova dei dati, nella maggior parte dei casi presi in esame, le “varianti-nuovo lemma” dantesche aggiungono occorrenze a lessemi che nei *corpora* linguistici presentano un numero di attestazioni assai esiguo [...], quasi nullo [...] o nullo [...]. Inoltre, talvolta, l'accezione che la variante assume nel passo dantesco accresce la documentazione di un'accezione generalmente assai poco rappresentata in italiano antico» (FRANCESCA DE BLASI, BARBARA FANINI, CRISTIANO LORENZI BIONDI, VERONICA RICOTTA, *Nell'officina del VD*, cit., p. 67).

Osservando i dati codificati del VD, si osserva che la «variante-nuovo lemma» non è esplicitamente marcata come tale; il sistema riesce però a distinguerla in quanto voce che soddisfa due condizioni: non è una voce di rinvio e la somma delle sue occorrenze nella *Commedia* è pari a zero. Il selettore dedicato alle varianti sul lemmario usa questa regola per permettere di filtrare le entrate: l'utente potrà scegliere di visualizzare le «varianti-nuovo lemma» assieme ai lemmi normali, di visualizzare solamente «le varianti-nuovo lemma», oppure di escluderle dalla visualizzazione e dalle ricerche.

Il secondo aggiornamento dedicato al lemmario si avvale del gestore dei contesti, un dispositivo *software* che permette ai redattori di trasferire i contesti localizzati dal corpus in GATTO³⁸ alla maschera di inserimento delle voci. Lo strumento di importazione, oltre a facilitare l'inserimento degli esempi sotto le accezioni pertinenti, effettua automaticamente la schedatura di tutti i contesti rilevati, corredati dai metadati del corpus e da un riferimento permanente alla voce corrispondente. Le frequenze, che possono essere calcolate a partire da questi dati, sono riportate, nella rappresentazione di PlutoVD, su undici campi: due per ciascuna cantica – una per le occorrenze del lemma a testo più una per la segnalazione delle varianti considerate nello spoglio – e cinque per le altre opere volgari (*Vita nuova*, il *Convivio*, le *Rime*, il *Fiore* e il *Detto d'Amore*). Sulla base di questi conteggi è stato realizzato un indice o *Lemmario per frequenze*³⁹, che organizza le entrate per rango: l'utente è libero di scegliere l'ordinamento (crescente o decrescente) sulla base della frequenza di un lemma sull'intera *Commedia* o su una singola cantica. Oltre a stabilire quanti risultati per pagina visualizzare, anche questo lemmario può essere filtrato specificando una chiave di ricerca con le consuete modalità di confronto (*Contiene*, *Stringa esatta*, *Inizia per* e *Finisce per*). Inoltre, accanto a ciascuna entrata, viene riportato un prospetto sintetico delle frequenze di occorrenza nella *Commedia*, così da permettere un immediato confronto tra le distribuzioni del lessema all'interno dell'opera.

³⁸ <http://www.ovi.cnr.it/Il-Software.html#gatto>.

³⁹ <http://www.vocabolariodantesco.it/indici.php>.

Il terzo e ultimo pacchetto di funzioni consiste in due algoritmi sperimentali che trasformano l'organizzazione dei significati grazie alla strutturazione granulare delle informazioni sulla base di dati: questa, in un *database* lessicale, può conferire una forte dinamicità alla microstruttura. Nella scheda del VD l'area della semantica⁴⁰ presenta come prima accezione il senso proprio registrato in italiano antico (spesso il più prossimo al valore etimologico) e, discostandovisi gradualmente, prosegue con una numerazione progressiva delle ulteriori accezioni. Qualora il redattore individui due diversi sememi, verranno riportati altrettanti significati di primo livello; i significati di livello inferiore verranno invece creati qualora si rilevi uno scostamento per alcuni semi rispetto al significato sovraordinato; i significati figurati ed estensivi vengono, di norma, isolati nella gerarchia⁴¹.

Nelle due visualizzazioni sperimentali, la numerazione originale – che veicola l'architettura semantica costruita dal redattore – viene mantenuta ed esplicitamente segnalata, ma l'interfaccia può riformulare l'ordine di presentazione delle accezioni sulla base di criteri alternativi. Le visualizzazioni sviluppate sono per il momento nascoste e per attivarle occorre intervenire manualmente sull'URL della pagina: concatenando all'indirizzo di qualsiasi voce la stringa “&m=luogo”⁴², l'albero dei significati verrà strutturato sulla base del luogo in cui il lessema occorre per la prima volta nella *Commedia*. Questa visualizzazione presenta quindi le accezioni secondo l'ordine in cui si rintracciano sul testo e allinea la microstruttura del VD a quella del VDL, che presenta le accezioni secondo l'ordine restituito dal motore di interrogazione *DanteSearch*. Attivata questa modalità, l'interfaccia di consultazione mostrerà, in cima all'area della semantica, un avviso che permetterà di ritornare all'ordinamento semantico predefinito o di passare all'or-

⁴⁰ Cfr. *La struttura della scheda* all'indirizzo <http://www.vocabolariodantesco.it/scheda.php>.

⁴¹ FRANCESCA DE BLASI, BARBARA FANINI, CRISTIANO LORENZI BIONDI, VERONICA RICOTTA, *Nell'officina del VD*, cit., pp. 27-28.

⁴² Per es. http://www.vocabolariodantesco.it/voce_tab.php?id=4997&m=luogo.

Tre implementazioni informatiche per il *Vocabolario Dantesco*

dinamento per frequenza⁴³. Quest'ultima opzione mostrerà invece le accezioni da quella più frequente in Dante a quella meno frequente. Anche questa soluzione è possibile grazie al sistema di importazione dei contesti e di una funzione, attivata al momento della programmazione della maschera di redazione, che consente di associare un contesto di occorrenza all'accezione pertinente anche senza visualizzarla nell'area della semantica.

Riassunto Dopo una breve introduzione sui progetti danteschi nelle *digital humanities*, dalle origini a oggi, l'articolo documenta l'applicazione di alcuni metodi di elaborazione dei dati lessicografici applicati al *Vocabolario Dantesco*. Vengono descritti tre gruppi di funzioni, due dedicati al lemmario e uno alla microstruttura, finalizzati a dotare il vocabolario elettronico di modalità aggiuntive e più sofisticate di esplorazione, ricerca e consultazione.

Abstract After a short introduction on Dante's projects in digital humanities, from its origins to today, the article describes the application of some lexicographic data processing methods applied to Dante's Vocabulary. Three groups of functions are described, two dedicated to the headword and one to the microstructure, aimed at providing the electronic dictionary with additional and more sophisticated methods of exploration, research and consultation.

⁴³ Analogamente a quanto fatto in precedenza è possibile attivare questa seconda visualizzazione concatenando all'URL di qualsiasi voce la stringa "&m=frequenza".

