



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

RELAZIONE

Lessico delle lirica del duecento-trecento:

Sondaggi Statistici

Candidato: *Johanna Zambrano*

Relatore: *Mirko Tavoni*

Correlatore: *Maria Simi*

Anno Accademico 2010-2011

“La lingua è la veste del pensiero.”

Samuel Johnson

Indice

Introduzione.....	3
--------------------------	----------

Parte Prima

1. Breve Storia della Lirica Italiana delle Origini.....	5
1.1 Scuola Siciliana.....	5
1.2 Rime di Dante.....	6
1.3 Dolce Stil Novo.....	7
1.4 Scuola Bolognese.....	8
1.5 Rime di Cino da Pistoia.....	9
2. Strumenti e concetti.....	10
2.1 Gattoweb.....	10
2.2 DanteSearch.....	12
2.3 IntraText.....	13
2.4 Google Books.....	14
2.5 Biblioteca Italiana.....	15
2.6 Biblioteca dei Classici Italiani.....	16
3. Metodologia proposta.....	17
3.1 Selezione degli autori e delle opere.....	17
3.2 Divisione del Corpus in sottocorpus.....	17
3.3 Ottenimento della lista della forme grafiche ed eliminazione delle parole vuote.....	19
3.4 Costruzione del Database MySQL.....	21
3.5 Inserimento dei dati nel Database.....	23
3.6 Ricerca delle forme grafiche comuni ai 5 sottocorpus.....	25

Parte Seconda

4. Aspetti quantitativi del corpus.....	29
4.1 Analisi dei dati quantitativi del corpus.....	29
4.2 Analisi dei dati quantitativi dei 5 subcorpora.....	30
5. Forme e occorrenze significative.....	33

5.1	Mente, umile e spirito.....	33
5.2	Viso e cera.....	36
5.3	Gentile e vile.....	38
5.4	Fina.....	40
5.5	Tempo.....	41
5.6	Amor/amore, cor/core, donna e occhi.....	42
5.7	Luce.....	44
5.8	Sicilianismi, provenzalismi e latinismi.....	45
5.9	Par e pare.....	47
5.10	Servire.....	48
Conclusioni.....		49
Bibliografia.....		51
Sitografia.....		54

Introduzione

La lirica italiana delle origini nasce e si sviluppa tra il duecento e il trecento. Si tratta di un periodo di lotte feroci, di grandi cambiamenti e scoperte, di rivolte religiose e rivoluzioni per l'affermazione della borghesia come nuova classe sociale. Tuttavia anche se si trattava di anni non facili, oltre alla prevalenza del volgare sul latino come lingua letteraria nascono figure importanti per la tradizione poetica italiana: in primis Dante Alighieri, Guido Guinizzelli, Giacomo da Lentini e Guido Cavalcanti. Questi autori sono riconosciuti tuttora come innovatori della loro epoca e addirittura oggi il loro pensiero viene considerato molto attuale.

L'insieme delle opere di questi poeti rappresenta una fonte incalcolabile di dati e di informazioni dal punto di vista linguistico, filologico e storico. Infatti sono moltissimi gli studi e le ricerche che riguardano la letteratura italiana del XIII e XIV secolo, attraverso i quali è stato possibile analizzare gli aspetti metrici, lessicali, sintattici o stilistici dei testi dell'epoca.

In base a quanto detto prima, questa ricerca si propone di ottenere il lessico che hanno in comune 5 gruppi di poeti e analizzare le forme più significative osservandole nel loro contesto. Nell'ambito dell'informatica Umanistica, uno dei modi migliori per effettuare questo tipo di analisi è attraverso l'uso degli strumenti informatici e delle risorse messe a disposizione dalla linguistica computazionale. Precisamente la combinazione dei dati quantitativi con l'analisi qualitativa è uno degli aspetti della linguistica computazionale più sfruttato nello studio di un testo. Si tratta di un approccio molto utilizzato in campi come la psicolinguistica, l'ingegneria del linguaggio o, come nel caso di questa ricerca, la stilometria. Infatti in ambito letterario, attraverso la stilometria è possibile calcolare e identificare lo stile caratteristico di un testo o di un autore. Il motivo di questo è dovuto alla peculiarità con cui viene utilizzato il linguaggio da ogni autore per quanto riguarda gli aspetti della sintassi, del lessico o della scrittura. La caratteristica analizzata con maggior frequenza sono i dati che riguardano il lessico dei testi insieme all'ipotesi che ogni autore aggiunge una proprio impronta stilistica in ognuna delle sue opere. Solitamente nel calcolo dello stile sono ritenuti dati misurabili la lunghezza delle parole, l'uso delle parole vuote o grammaticali, la distribuzione della punteggiatura, le frequenze di forme o di strutture sintattiche, ecc.

Quindi la proposta qui presentata si sviluppa in base a obiettivi ben precisi: combinare in modo chiaro ed effettivo la sfera umanistica a quella informatica utilizzando metodi statistici per ottenere il lessico in comune di 5 scuole poetiche e analizzarlo in modo da capire i tratti stilistici di ognuna di queste.

Inizialmente, nel primo capitolo, viene presentato un breve riassunto storico per introdurre il contesto storico della ricerca così come gli oggetti di studio, ovvero le 5 scuole poetiche.

Successivamente, nel capitolo 2 sono descritti gli strumenti adoperati nelle varie fasi del procedimento. In particolare si è deciso di basare la ricerca solo sui testi presenti nel corpus del TLIO (Tesoro della Lingua Italiana delle origini) e di usare come supporto linguistico i corpora di DanteSearch e di IntraText. Inoltre gli archivi elettronici, come Biblioteca Italiana e Google Books, sono stati strumenti importanti per reperire le opere che non sono più disponibili in modo cartaceo e anche perché in molti casi permettono una consultazione rapida che fa risparmiare tempo. Persino la Biblioteca dei classici italiani è uno strumento molto utile in quanto contiene oltre ai testi anche un glossario delle parole dell'epoca, il quale in alcuni casi ha permesso di capire i vari significati e uso delle forme studiate.

Nel terzo capitolo è illustrata passo dopo passo la metodologia adottata in tutte le fasi del procedimento. Inoltre vengono spiegate le ragioni delle varie scelte fatte in modo di spiegare chiaramente tutta la prassi.

Nel quarto capitolo sono presentati i risultati ottenuti dalle analisi effettuate sull'intero corpus e sui 5 subcorpora studiati. In particolare il corpus è analizzato dal punto di vista della ricchezza lessicale e della frequenza media. Invece per i 5 subcorpora sono state svolte analisi anche in base al numero di componenti di cui sono formati.

Infine nel quinto capitolo sono analizzate dal punto di vista stilistico alcune delle forme del lessico in comune ai 5 subcorpora. L'obiettivo è quello di comprendere l'impiego e il significato di queste forme nei vari autori attraverso la presenza nei diversi componenti.

1.

Breve storia della lirica italiana delle origini

La formazione delle varietà vernacolari nella penisola italiana si sviluppa dal latino rustico, parlato dalle popolazioni rurali durante la decadenza dell'impero romano. Nonostante ciò, il latino rimane la lingua dominante per molto tempo come lingua parlata e soprattutto come lingua di cultura scritta. Infatti queste varietà dialettali producono le loro opere letterarie molto tardi rispetto alla nascita di letterature volgari di altre zone della Romania¹, dato che l'uso del volgare come lingua letteraria non era stato ancora concepito. Soltanto verso la fine del secolo XII compaiono i primi testi volgari in metrica rudimentale.

La nascente poesia lirica italiana si vede fortemente influenzata dall'arrivo e dalla diffusione della lirica provenzale, grazie anche al continuo fiorire di poesie trobadoriche nelle corti italiane. A questo punto cresce notevolmente la necessità di rendere la cultura disponibile a un ampio numero di persone utilizzando il volgare come lingua letteraria. Quindi si può parlare di letteratura italiana soltanto quando successivamente all'uso prevalentemente pratico del volgare si congiunge un intento artistico.

1.1.

La Scuola Siciliana

La storia della letteratura italiana inizia intorno al 1230 in Sicilia presso la corte di Federico II di Svevia. Molti dei poeti della Magna Curia sono siciliani di nascita, altri invece provengono dal continente e sono funzionari e cortigiani, uomini di elevata cultura e conoscitori del latino, dell'arabo e di altre lingue. Tra i maggiori esponenti spicca in primis il notaio Giacomo da Lentini per la sua sorprendente produzione e per essere stato l'iniziatore della scuola siciliana. Dopo di lui, Guido delle Colonne, Stefano Protonotaro, Odo delle Colonne, Mazzeo di Ricco, Pier della Vigna e molti altri. Lo stesso imperatore non solo favorisce la creazione poetica ma ne partecipa attivamente lui stesso così come suo figlio Re Enzo.

La poesia siciliana prende come modello nella forma e nel contenuto la lirica provenzale, la quale insegna le maniere dell'amore cortese. La tematica principale è quella del "fino amor" e della sua

1. Per Romania s'intende oggi l'area in cui si sono sviluppate le lingue romanze.

fenomenologia. Tuttavia la struttura sociale del Regno di Sicilia porta i poeti di questa scuola a cambiare il rapporto di vassallaggio tra donna-signore; quindi la difficoltà di raggiungere l'amata diventa un momento di sofferenza amorosa e di disdegno. Per quanto riguarda le forme metriche vengono conservate quelle tipiche come la canzone di argomento aulico, la canzonetta con temi narrativi e frequentemente dialogati, mentre viene eliminato l'accompagnamento musicale e introdotte nuove forme metriche, tra cui il sonetto e in pochi componimenti anche il discordo.

La scelta linguistica dei poeti di questa scuola è il volgare siciliano in una forma nobilitata sotto il profilo del lessico, ricca di artifici retorici ma priva di forme popolari o gergali e fortemente influenzata dal latino e dal provenzale. Questa lingua colta però più che essere una lingua aperta alla comprensione di tutti, è sostanzialmente una lingua d'*élite* sociale e culturale che non ha somiglianza nemmeno con il popolo siciliano. Inoltre presenta caratteristiche morfologiche diverse rispetto alle varietà nascenti al centro della penisola e sarà proprio questo a segnare il modo in cui saranno trasmessi i componimenti siciliani. Quando il dominio federiciano ha fine e la cultura meridionale subisce una rapida decadenza le poesie siciliane vengono raccolte e trascritte in antologie in terra toscana. I copisti medievali però pensano alla trasmissione dei testi come alla loro interpretazione, quindi quelle forme che non trovano una corrispondenza nella loro varietà linguistica vengono modificate e, consapevolmente o no, il testo originale muta. Infatti le antologie di poesia siciliana trasmesse da copisti toscani sono state fortemente toscanizzate. Le uniche testimonianze rimaste della poesia siciliana in volgare locale sono "*Pir meu cori allegrari*"² di Stefano Protonotaro, un frammento di "*Allegru cori plenu*"³ e le ultime due stanze della canzone "*S'eo trovasse pietanza*"⁴ entrambe di Re Enzo.

1.2.

Le Rime di Dante

Tra il 1283 e il 1305 circa Dante concretizza un'ampia produzione lirica. Durante questi anni compone, oltre a una canzone e a un serventese andati perduti, 88 poesie in lingua volgare, 31 delle quali sono state inserite nella *Vita Nuova* e 3 nel *Convivio*. Le 54 poesie escluse da tali opere si raccolgono oggi sotto il nome di Rime e a queste si aggiungono circa una ventina di rime di dubbia

2. *Poeti del Duecento*, a cura di G. Contini, Milano-Napoli 1960, I, pp. 130-3.

3. *Ibid.*, pp. 157-9.

4. A. Monteverdi, Per una canzone di Re Enzo in "*Studi e saggi sulla letteratura italiana dei primi secoli*". Ricciardi, Milano-Napoli 1954, pp. 97-9

attribuzione per le quali mancano le prove di appartenenza alla produzione dantesca. Queste composizioni non sono state sistemate dal poeta in una raccolta organica ragion per cui non si può parlare di un'opera unitaria.

Le Rime sono caratterizzate dalla varietà dei temi, delle forme espressive e dello stile che sono proprio gli elementi che permettono di seguire l'evoluzione del pensiero e della poetica di Dante. Inoltre la maggior parte delle Rime non ha una datazione certa né un contesto preciso quindi generalmente sono catalogate per forme metriche, per ordine alfabetico o per affinità linguistiche. Le più famose sono le rime petrose, di denominazione allegorica perché in queste 4 composizioni si canta l'amore sensuale per una donna dura come la pietra. Un altro gruppo è quello formato dalle rime di stampo stilnovistico in cui domina la tematica amorosa dovuto all'influsso di Cavalcanti. Oltre a questi due gruppi c'è la tenzone con Forese Donati di genere comico e molto realistico, le rime dell'esilio e quelle allegoriche e dottrinali.

1.3.

Il Dolce Stil Novo

«O frate, issa vegg' io», diss' elli, «il nodo
che 'l Notaro e Guittone e me ritenne
di qua dal dolce stil novo ch'i' odo!»

(*Purgatorio*, XXIV, vv. 49-57)

E' proprio da questa terzina dantesca che si risale al nome della corrente lirica che si sviluppa tra il 1280 e il 1330. I concetti basi erano stati formulati da Guido Guinizzelli inizialmente a Bologna e il motivo non era stato casuale. Bologna era sede della famosa università che attirava intellettuali di ogni provenienza e in più era centro filosofico e scientifico in quel momento. Da questo intreccio, Guinizzelli riesce a concepire una nuova forma poetica di cui getta le fondamenta nella sua composizione “*Al cor gentil rempaira sempre amor*”. Essenzialmente le sue novità consistono nella sublimazione religiosa della donna ad angelo e nell'intensificazione dell'amore come segno di nobiltà di cuore invece che di nobiltà di nascita.

Lo Stilnovo successivamente si trasferisce a Firenze dove viene accolto da un gruppo di giovani poeti, i quali cercano di proseguire in maniere diverse l'esperienza poetica di Guinizzelli. Di questo gruppo spiccano Guido Cavalcanti, senza dubbio l'iniziatore della nuova forma poetica fiorentina, e

lo stesso Dante, primo intellettuale fiorentino a teorizzare la concezione guinizzelliana. L'incontro tra queste due figure di forte ingegno è decisivo per lo Stilnovo fiorentino. In particolare tra i due si svolge uno scambio di sonetti che alimenta la nuova poetica, tra cui il famoso “*Guido i' vorrei*”. Oltre ai grandi protagonisti, sono fiorentini anche quei poeti considerati “minori”: Lapo Gianni, Gianni degli Alfani, il cosiddetto “amico di Dante”, Dino Frescobaldi e altri ancora. Per queste figure lo Stilnovo funziona da miraggio irraggiungibile, ponendosi quasi come un ostacolo nel cammino molto più modesto delle loro esperienze poetiche.

Lo Stilnovo dal punto di vista storico riveste una grande importanza per la sua influenza sulla poesia italiana dell'epoca. Infatti proprio grazie a questa corrente poetica il toscano raggiunge il rango di lingua volgare e di modello sovra-regionale come lingua letteraria in tutta la penisola.

1.4.

La Scuola Bolognese

Bologna è il centro culturale per eccellenza tra la fine del duecento e l'inizio del trecento. Vi è attiva una rispettabile schiera di poeti locali che mantiene viva la tradizione lirica con tratti della poesia trobadorica e siciliana. I più importanti sono Onesto da Bologna, Guido Ghislieri, Polo Zoppo, Picciol da Bologna, Raniero dei Samaritani, Semprebene.

Tra questi si trova anche la figura di Guinizzelli, iniziatore del Dolce Stil Novo. La sua esperienza lirica si trova tra l'antico e il nuovo stile e quindi aggiunge delle novità alla sua poesia ma allo stesso tempo conserva elementi tipici della tradizione precedente come in *Donna, l'amor mi sforza*⁵.

Contrariamente Onesto da Bologna è fermamente discordante con gli esperimenti lirici effettuati dalla nuova scuola. Rimprovera spesso questo modo di poetare sia nella sua corrispondenza con i guittoniani che in quella con Cino da Pistoia. A ogni modo però non può essere definito un purista della tradizione in opposizione allo Stilnovo dato che la sua poesia porta tracce, anche se inconsapevoli, di Guinizzelli e in molti casi ricorda lo stesso Cino da Pistoia come in *La partenza che fo dolorosa* o *Se li tormenti e dolor ch'omo à conti*⁶.

I restanti poeti del gruppo bolognese presentano nelle loro opere artifici di carattere guittoniano sebbene in alcuni, come Polo Zoppo, si possano osservare elementi guinizzelliani.

5. G. Contini, *opacità*, op. 21-34.

6. *I rimatori bolognesi del secolo XIII*, edizione critica a cura di G. Taccagni, Milano, Società Editrice Vita e Pensiero, 1933, pp. 29-30.

Guardando ai contributi fatti da questi rimatori si può dedurre come i volgari bolognese e siciliano partecipino alla formazione della lingua nazionale. Dante stesso nel *De Vulgari Eloquentia*⁷ loda il bolognese definendolo “leggiadra loquela” perché sviluppatosi dai volgari delle città circostanti, anche se paragonandolo al fiorentino e al toscano non lo considera illustre o di registro alto.

1.5.

Le Rime di Cino di Pistoia

La fama di Guittoncino de' Sighibuldi, noto come Cino da Pistoia, non si limita soltanto a quella come giurista. Egli infatti annovera una copiosa produzione lirica costituita da 165 rime (tra ballate, sonetti e canzoni) e in aggiunta circa 20 rime di dubbia attribuzione. La tematica predominante nelle rime del canzoniere ciniano è quella amorosa, e nella maggior parte canta il suo amore per una donna di nome Selvaggia. Questo amore, di carattere drammatico e violento, nasce dalla bellezza della donna e si manifesta nel desiderio continuo di godere della vista dell'amata, motivo che è soprattutto causa di tormento. Il suo stile dolce, lodato da Dante nel *De Vulgari Eloquentia*⁸, non raggiunge il livello filosofico e drammatico di Cavalcanti. Questo però non gli vieta di sperimentare forme metriche e stilistiche come la sua tipica costruzione retorica o di ricorrere alla variazioni dei *topoi* propri della lirica siciliana e provenzale usando motivi mitologici che si trovano anche in ambito giuridico.

Le opere di Cino, oltre a quella amorosa, toccano altre tematiche come la morte di Arrigo VII o quella di Dante e la satira contro Napoli. Inoltre nelle rime di corrispondenza che mantiene con Onesto da Bologna, Cacciamonte e altri poeti bolognesi minori sono presenti registri linguistici medi che cambiano, a seconda del momento, in satira, invettiva o burla.

7. Cfr. *DVE*, I, XV, 2-6.

8. Cfr. *DVE*, I, X, 2.

2.

Strumenti e concetti

Oggigiorno la rete è popolata da una grande quantità di informazione digitale (testi, immagini, audio, video, metadati, ecc), caratteristica che la rende una risorsa necessaria nella ricerca e nello studio in generale. Nell'ambito di un'investigazione, lo sconfinamento del web offre molte possibilità in più rispetto a una biblioteca, soprattutto perché migliora e facilita il compito degli studiosi.

Una di queste possibilità è rappresentata dai Corpus o raccolte di testi ordinati in base a metodi specifici che permettono studi linguistici sia da un punto di vista storico che comparativo. I Corpus sono impiegati moltissimo nelle ricerche proprio per la loro particolare funzione rappresentativa e per la loro organizzazione strutturale secondo criteri precisi. Un'altra possibilità viene dalle biblioteche digitali ovvero quelle risorse telematiche che contengono testi digitalizzati che possono essere consultati on-line. In tutto il mondo sono moltissimi gli enti che aiutano e facilitano la ricerca come il *Childe's Corpus*, il *British National Corpus*, il *Progetto Manuzio*, l'*Archivio Mediceo il Principato* dell'Archivio di Stato di Firenze e anche giornali come il *New York Times* o il *Corriere della Sera*.

Ovviamente, la scelta del web come strumento utile per trovare fonti storiche dipenderà dal materiale reso disponibile in rete e di come questo venga pubblicato.

Per questa ricerca è stato scelto di usare il *Corpus Tlio* e di affiancargli il *Corpus DanteSearch* per quanto concerne le rime di Dante. Inoltre tutto il materiale organico è stato scaricato dal sito *IntraText* e le informazioni ottenute sono state supportate non solo da materiale cartaceo ma anche da testi e documenti reperiti presso la *Biblioteca Italiana*, *Google Books* e il progetto della *Biblioteca dei Classici Italiani*.

2.1.

TLIO – Tesoro della lingua italiana delle origini

Si tratta di un database presente nell'OVI (Opera del Vocabolario Italiano) nato nell'ambito del progetto del vocabolario storico avviato nel 1995 dall'Accademia della Crusca con il finanziamento del CNR.

Il Corpus contiene testi in volgare italiano a partire dal primo documento, l'*Indovinello Veronese* del

secolo IX, fino ai testi scritti prima del 1375 (anno della morte di Boccaccio). In realtà i testi di cui si occupa sono databili ben oltre quella data.

Il TLIO è implementato nel software GATTO⁹ da Domenico Iorio-Fili e accessibile online in Gattoweb (v. fig. 2.1) all'indirizzo <http://gattoweb.ovi.cnr.it/>. E' un progetto in corso d'opera che attualmente contiene 2316 testi, 466.551 forme grafiche distinte e 116.411 lemmi.



Fig. 2.1 Homepage del Corpus OVI dell'Italiano antico.

Il Corpus ha quasi tutte le forme grafiche lemmatizzate e sono molto poche quelle rimaste ancora non interpretate. L'operazione di lemmatizzazione è molto complessa perché richiede lo studio dei vari documenti prendendo in considerazione: i numerosi contesti, le diverse varietà volgari, i tipi di testi così come i vari ambiti (lirico, giuridico, epistolare, medico, mercantile, ecc) e i differenti criteri filologici e di affidabilità.

Essendo un vocabolario storico presenta le parole così come sono nel testo antico, in un particolare periodo o in tutta la storia della lingua in cui sono scritte. Questo lo rende uno strumento molto utile nelle ricerche filologiche, linguistiche o sulla storia della lingua italiana.

⁹Acronimo per Gestione degli Archivi Testuali del Tesoro delle Origini. E' il software usato per costruire, lemmatizzare e interrogare in rete locale i corpus testuali interrogabili via Internet con Gattoweb. L'indirizzo web della risorsa é: [http://gattoweb.ovi.vnr.it/\(S\(eocsm145udxkqk45jzpp3q45\)\)/helpGattoWeb/Glossario.html](http://gattoweb.ovi.vnr.it/(S(eocsm145udxkqk45jzpp3q45))/helpGattoWeb/Glossario.html)

2.2

DanteSearch

DanteSearch nasce nell'unità di ricerca di Pisa, sotto la supervisione del Prof. Mirko Tavoni e la collaborazione di vari ricercatori. Il corpus contiene le opere in volgare e in latino di Dante Alighieri (v. fig. 2.2). Tutti i testi sono lemmatizzati e annotati grammaticalmente. La codifica sintattica è stata effettuata soltanto per le Rime, il Convivio e la Commedia secondo la Grande grammatica italiana di consultazione di Renzi, Salvi e Cardinaletti e sui fondamenti presenti nell'Appendice all'Enciclopedia Dantesca. Inoltre ogni parola è dotata di un'etichetta che ne specifica le informazioni grammaticali e morfosintattico. Sia per l'analisi grammaticale che per la lemmatizzazione è stato necessario uno studio approfondito del lessico di Dante così come confronti con testi latini e volgari medievali nei casi in cui un termine non fosse stato ancora attestato.

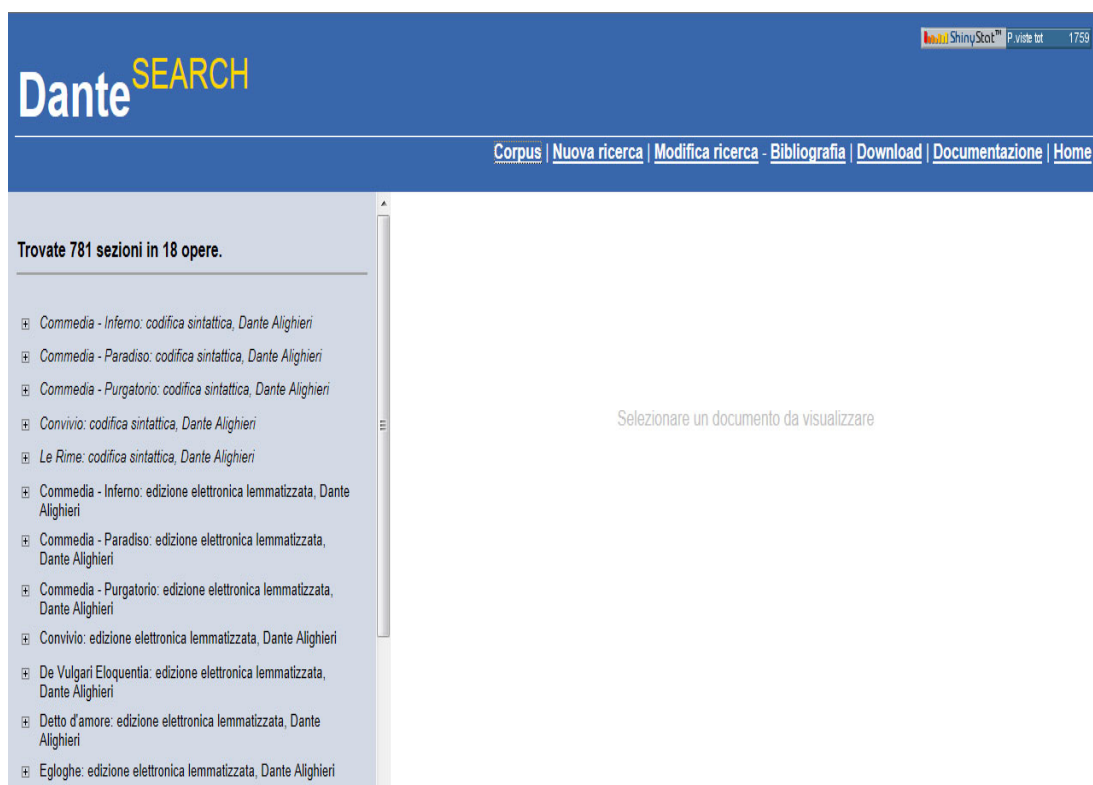


Fig. 2.2 Testi presenti nel Corpus DanteSearch.

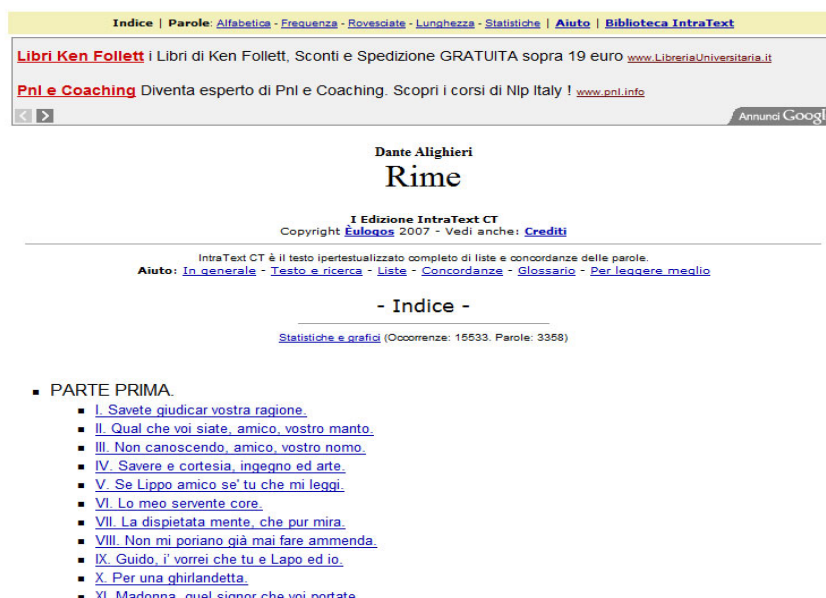
Il linguaggio utilizzato nelle varie codifiche è il metalinguaggio XML (Extensible Markup Language) sulla base delle Guidelines della TEI (Text Encoding Initiative).

DanteSearch si avvale anche di un motore di ricerca implementato da Claudio Corsi, che permette di compiere sia ricerche grammaticali che sintattiche, per forme o per lemmi e oltre a questo usare ulteriori filtri di ricerca o combinazioni con operatori booleani (AND, OR, NOT, NEAR). In aggiunta, la ricerca grammaticale e la ricerca sintattica possono essere accostate l'una accanto all'altra nella parte destra della pagina.

2.3. IntraText

Questa risorsa telematica di ricerca testuale on-line (<http://www.intratext.com>), è stata realizzata grazie al sostegno dell'azienda romana Eulogos Spa. Comprende sia opere letterarie antiche che opere contemporanee ed è accessibile gratuitamente .

Il progetto *IntraText* ha la doppia funzione di biblioteca digitale e di corpus: infatti combina la pubblicazione di documenti con cura scientifica (come testi, periodici, archivi) con l'ipertestualizzazione lessicale e il collegamento testo-concordanze con l'obiettivo di permettere di leggere i testi come ipertesti. L'individuazione dei documenti può essere svolta attraverso alcuni filtri come secolo di riferimento, autore o titolo e inoltre si possono confrontare le concordanze lessicali di un testo con altri testi (v. Fig. 2.3).



The screenshot displays the IntraText interface for Dante Alighieri's 'Rime'. At the top, there is a navigation bar with links for 'Indice', 'Parole', 'Alfabetica', 'Frequenza', 'Rovesciate', 'Lunghezza', 'Statistiche', 'Aiuto', and 'Biblioteca IntraText'. Below this, there are two promotional banners: one for 'Libri Ken Follett' and another for 'Pnl e Coaching'. The main content area features the title 'Dante Alighieri Rime' and indicates it is the '1 Edizione IntraText CT' with a copyright notice for Eulogos 2007. A footer section provides an 'Aiuto' menu with links to 'In generale', 'Testo e ricerca', 'Liste', 'Concordanze', 'Glossario', and 'Per leggere meglio'. Below the footer, there is a section titled '- Indice -' with a link for 'Statistiche e grafici' (Occorrenze: 15533, Parole: 3358). The main list of rime is titled 'PARTE PRIMA' and includes eleven numbered items, each with a blue hyperlink to the specific text.

Indice | Parole: [Alfabetica](#) - [Frequenza](#) - [Rovesciate](#) - [Lunghezza](#) - [Statistiche](#) | [Aiuto](#) | [Biblioteca IntraText](#)

Libri Ken Follett i Libri di Ken Follett, Sconti e Spedizione GRATUITA sopra 19 euro www.LibreriaUniversitaria.it

Pnl e Coaching Diventa esperto di Pnl e Coaching. Scopri i corsi di Nlp Italy ! www.pnl.info

Annunci Google

Dante Alighieri
Rime

1 Edizione IntraText CT
Copyright Eulogos 2007 - Vedi anche: [Crediti](#)

IntraText CT è il testo ipertestualizzato completo di liste e concordanze delle parole.
Aiuto: [In generale](#) - [Testo e ricerca](#) - [Liste](#) - [Concordanze](#) - [Glossario](#) - [Per leggere meglio](#)

- Indice -

[Statistiche e grafici](#) (Occorrenze: 15533, Parole: 3358)

■ PARTE PRIMA.

- [I. Savete giudicar vostra ragione.](#)
- [II. Qual che voi siate, amico, vostro manto.](#)
- [III. Non canoscendo, amico, vostro nome.](#)
- [IV. Savere e cortesia, ingegno ed arte.](#)
- [V. Se Lippo amico se' tu che mi leggi.](#)
- [VI. Lo meo servente core.](#)
- [VII. La dispietata mente, che pur mira.](#)
- [VIII. Non mi poriano già mai fare ammenda.](#)
- [IX. Guido, i' vorrei che tu e Lapo ed io.](#)
- [X. Per una qhirlandetta.](#)
- [XI. Madonna, quel signor che voi portate.](#)

Fig. 2.3 Raccolta delle Rime di Dante in IntraText

2.4.

Google Books

Google Books (<http://books.google.it/>) è un database offerto da Google che contiene testi digitalizzati, indicizzati e messi a disposizione nei risultati di Google o direttamente dalla pagina di Google Books (v. Fig. 2.4).

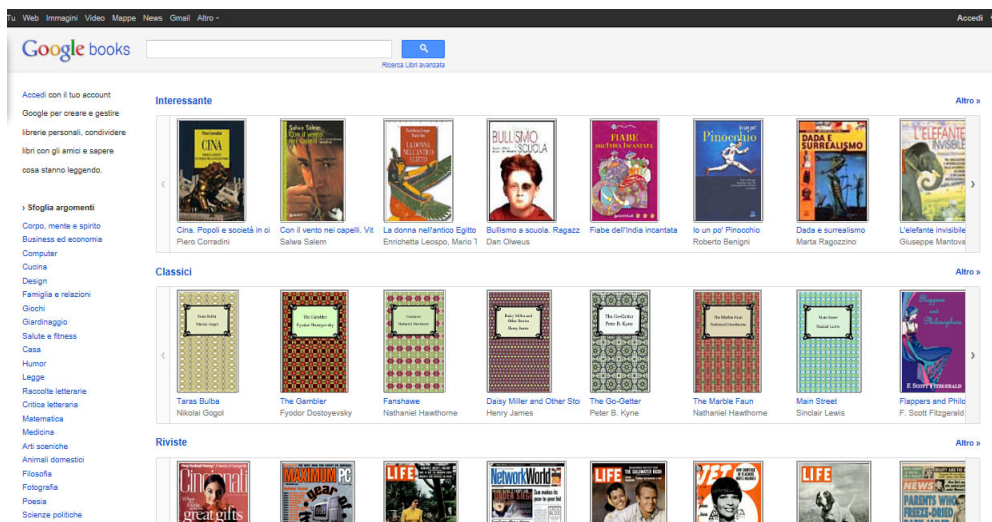


Fig. 2.4 Interfaccia di Google Books

Il progetto, iniziato nel 2004, si è sviluppato grazie alla collaborazione di numerose e importanti biblioteche oltre a università e case editrici di tutto il mondo. Ogni anno cresce il numero di partnership con il conseguente aumento dei libri digitalizzati. Infatti oggi si parla di più di 7 milioni di testi resi disponibili online.

La digitalizzazione dei testi è realizzata attraverso il riconoscimento ottico dei caratteri (OCR) e la varietà di questi è molto ampia: si va dai testi antichi ai testi contemporanei, dai libri alle riviste o ai saggi. In Google Books si possono effettuare ricerche in base sia ai dati bibliografici dei testi desiderati che a una parola chiave o a una stringa di testo. Se esiste una corrispondenza tra un testo e i dati immessi, viene visualizzato dall'utente il link del libro o della pagina in cui è presente la parola chiave o stringa cercata. Se il testo è protetto dai diritti d'autore si può leggere una piccola percentuale del testo (molte volte, addirittura poche righe). Se invece il testo non ha restrizioni (in

alcuni casi quelli più antichi) possono essere letti completamente in modo gratuito e, dove possibile, scaricati in formato PDF.

2.5.

Biblioteca Italiana

La *Biblioteca Italiana* è una biblioteca digitale promossa dal “Centro interuniversitario Biblioteca Italiana Telematica” (ciBit), di cui fanno parte molte università italiane, e dal progetto “Biblioteca Digitale Italiana” (BDI)¹⁰ del Ministero per i beni e le attività culturali.

Nel suo database sono presenti testi antichi e contemporanei che possono essere letti direttamente sul sito o scaricati gratuitamente in formato XML. Si tratta di oltre 1700 testi riguardanti la cultura e la letteratura italiana. Per quanto riguarda le ricerche dei testi desiderati, questa risorsa telematica consente di effettuarle per campi, autore, per genere, per periodo, per testi (v. Fig. 2.5).

The screenshot shows the search interface of the Biblioteca Italiana website. At the top, the header includes the site name 'Biblioteca Italiana' and the affiliation 'Università degli Studi di Roma "La Sapienza"'. A navigation menu on the left lists 'Biblioteca', 'Catalogo', 'Scorri per autore', 'Scorri per genere', 'Scorri per periodo', 'Ricerche testuali', 'Partner & Sponsor', and 'Servizi'. The main content area is titled 'Catalogo' and features a search form with the following elements:

- Ricerca per campi:** Two input fields for 'Autore' and 'Titolo', each with a dropdown menu and an 'elenco termini' button. A 'Help' link is also present.
- Limita la ricerca per periodo:** A row of checkboxes for years: 200, 300, 400, 500, 600, 700, 800, 900.
- Limita la ricerca per genere:** Two columns of checkboxes for genres: Narrativa, Letteratura teatrale, Documenti, Lettere ed epistolari, Memorialistica, Poesia, Oratoria, Testi storici e storiografici, Traduzioni e volgarizzamenti, and Trattati.
- Results:** A dropdown menu set to '15' and the text 'risultati per schermata', along with 'Invia richiesta' and 'Azzera' buttons.

At the bottom of the search area, there are three icons with labels: 'Autori', 'Periodo storico', and 'Genere'.

Fig. 2.5 Interfaccia dell'area di ricerca di Biblioteca Italiana

10. Consultabile all'indirizzo <http://www.iccu.sbn.it>.

La *Biblioteca Italiana* basa la sua architettura in buona parte su un sistema informativo aperto per l'archiviazione (Open Archival Information System, OAIS) e in aggiunta usa un sistema per la gestione delle risorse digitali oltre a un sistema per la gestione dei metadati associati a ogni documento. Invece per la descrizione completa della fonte cartacea del testo digitalizzato, adopera un indice SBN che è collegato alla fonte analogica descritta nell'OPAC SBN.

2.6.

La Biblioteca Dei Classici Italiani

Biblioteca dei Classici Italiani è un archivio digitale online (<http://www.classicitaliani.it>) che contiene testi estratti da vecchie edizioni (oggi di pubblico dominio) di grandi autori letterari italiani e di vario genere: si va da biografie, introduzioni, commenti e bibliografie a singoli testi letterari (v. Fig. 2.6).

Biblioteca dei Classici Italiani
di Giuseppe Bonghi da Lucera
... tra questa Immensità s'annega il pensier mio: E il naufragar m'è dolce in questo mare.

[Opere](#) [Biografie](#) [Anonimi](#) - [A](#) - [B](#) - [C](#) - [DEF](#) - [GHI](#) - [L](#) - [MNO](#) - [P](#) - [QRS](#) - [TUVZ](#) [Introduzioni](#) [Fonti](#)

Progetti			Unità culturali	
Alfieri	Duecento	Statuti italiani dell'Ottocento	Aforismi dai Classicitaliani	Il Contrasto di Cielo d'Alcamo
Alighieri Dante	Trecento	Medioevo	Visioni medioevali	Leggende di Giuda e Pilato
Aretino	Quattrocento	Appunti letterari italiana	I poeti del Dolce stil novo	Biblioteca rusticale
Ariosto	Cinquecento	Umanesimo - Rinascimento	tirannide - repubblica - democrazia - re	Il mondo dei villani
Bembo	Seicento	Illuminismo	serventesi - poemetti - cantari	Il diavolo nella letteratura
Boccaccio	Settecento	Romanticismo	Aforismi - Proverbi - Massime - Detti	Novella del Grasso legnaiuolo
Carducci	Ottocento	Verismo	Documenti di storia del Novecento	
Foscolo	Novecento - Poesia	Decadentismo	Motore di ricerca dei Classici della Letteratura <input type="text"/> <input type="button" value="Cerca"/>	
Goldoni	Novecento - Prosa	Neorealismo		
Latini Brunetto		Poesia religiosa in letteratura		
Leopardi				

Fig. 3.6 Home page della Biblioteca dei Classici Italiani

Oltre a questo, si trova un'area dedicata alla storia della letteratura italiana e un glossario di parole dell'italiano antico del 200/300.

Questo archivio digitale, nato nel 1996, fa parte di un progetto personale del Prof. Giuseppe Bonghi.

3.

Metodologia proposta

In questo capitolo è illustrata la metodologia adottata per ottenere il lessico in comune di 5 scuole poetiche del duecento/trecento. Nel procedimento complessivo è stato sfruttato il rapporto tra gli strumenti informatici e il testo. Nello specifico si è usufruito del Corpus dell'Italiano Antico e degli strumenti disponibili su Gattoweb così come di alcuni strumenti informatici ideali per la gestione dei dati linguistici: i database e i linguaggi di interrogazione.

3.1.

Selezione degli autori e delle opere

Dopo aver definito il marco storico e le scuole poetiche coinvolte, il passo successivo è quello di consultare i riferimenti bibliografici di cui dispone Gattoweb con l'obiettivo di definire gli autori e le opere di ogni scuola. Si procede in questo modo perché non tutti gli autori o i testi antichi sono stati ancora inseriti nella base di dati di Gattoweb. Infatti l'inserimento dei testi nel corpus richiede molto tempo perché comprende non solo la loro traduzione in formato elettronico e le eventuali codifiche ma uno studio filologico precedente attraverso il quale vengono controllati alcuni criteri editoriali o eseguita una vera e propria verifica testuale.

Nell'archivio di catalogazione (v. fig. 3.1) ogni testo del corpus è associato a un insieme di informazioni di tipo bibliografico (autore, titolo, edizione, sigla, genere, datazione, ecc) e questi testi a loro volta sono associati a dei contesti che possono essere estratti e visualizzati a seguito di una ricerca.

3.2.

Divisione del Corpus in Subcorpora

Quindi selezionati gli autori e le opere, vengono utilizzati le sigle dei testi per definire dinamicamente dei subcorpora all'interno del corpus e restringere a uno o più di essi le operazioni di

ricerca (v. fig. 3.2).

gattoweb.ovi.cnr.it/(S(0e12vg2acqbp4zkebh3r45))/CatForm25.aspx

Corpus OVI dell'Italiano antico - dati bibliografici dei testi

Chiudi pagina Riavvia GattoWeb Guide..

testi
1980 in 40 pag. pag. corrente 27

scegli un'altra pagina

ordinamento attuale:
genere (a-z)
ordina per..
A - Z Z - A

testi inclusi nella tabella
 corpus sottocorpus A sottocorpus D
 sottocorpus B sottocorpus E
 sottocorpus C sottocorpus F

sigla	anno iniziale	anno finale	check note	autore	titolo	titolo abbreviato	edizione	curatore	area generica	area specifica	indice di qualità	generetipoforma	data descrittiva	collocazione interna	inser
vz	1270	1300		Cavalcanti, Guido	Rime	Guido Cavalcanti (ed. Contini), 1270-1300 (fior.)	Poeti del Duecento, a cura di Gianfranco Contini, Milano-Napol..	Contini, G.	tosc.	fior.		lir. O V	1270-1300E/VZ		1997 00:00
vw	1291	1300		Cavalcanti, Guido (?)	Ballate (Due)	Guido Cavalcanti (?), Due ball., XIII ex. (fior.)	Rime, a cura di Guido Favati, Milano-Napol..	Favati, G.	tosc.	fior.		lir. O V	sec. XIII ex. E/VW		1997 00:00
b2	1276	1276		Guinizelli, Guido	Rime	Guinizelli (ed. Baudi di V.), a. 1276 (tos.)	Del reggimento e costumi di donna di messer Francesco Barber.	Baudi di Vesme, C.	tosc.	tosc.		lir. O V	a. 1276 E/B2		1997 00:00
b3	1276	1276		Guinizelli, Guido	Rime	Guinizelli (ed. Contini), a. 1276 (tos.)	Poeti del Duecento, a cura di Gianfranco Contini, Milano-Napol..	Contini, G.	tosc.	tosc.		lir. O V	a. 1276 E/B3		1997 00:00
jf	1330	1330		Guido Novello da Polenta	Rime	G. N. da Polenta, Rime, a. 1330 (ravenn.>ven.)	Domenico De Robertis, Il Canzoniere Escalonense e la tradiz..	De Robertis, D.	romagn.	ravenn.>ven.		lir. O V	a. 1330 E/JF		1997 00:00
bbl	1290	1304		Orlandi, Guido	Rime	Guido Orlandi, 1290/1304 (fior.)	Valentina Pollidori, Le Rime di Guido	Pollidori, V.	tosc.	fior.		lir. O V	1290/1304E/BBL		1997 00:00

Fig. 3.1 Archivio contenente i dati bibliografici dei testi.

gattoweb.ovi.cnr.it/(S(cuxyfw55uvyioi55roznzce5))/CatForm14.aspx

Corpus OVI dell'Italiano antico - definizione dei sottocorpora

Opera selezione.. Statistiche.. Azzera impostazioni Salva Chiudi pagina Riavvia GattoWeb Guide..

sottocorpus
sottocorpus A
sottocorpus B
sottocorpus C
sottocorpus D
sottocorpus E
sottocorpus F

selezione sulla base dei metadati

sigla kz escludi
titolo escludi
edizione escludi
area generica escludi
indice di qualità escludi
tipo escludi

autore escludi
titolo abbreviato escludi
curatore escludi
area specifica escludi
genere escludi
forma escludi

selezione cronologica
periodo
anno iniziale anno finale

testi
7 in 1 pag. pag. corrente 1

scegli un'altra pagina

sigla	anno iniziale	anno finale	check note	autore	titolo	titolo abbreviato	edizione	curatore	area generica	area specifica	indice di qualità	generetipoforma	data descrittiva	collocazione interna	inser
ar	1251	1300		Verzellino	Una piacente donna cōnta e bella /sonetto/	Verzellino, XIII sm. (fior.)	Poeti del Duecento, a cura di Mario Marti, Firenze, ..	Marti, M.	tosc.	fior.		lir. O V	sec. XIII s.m. E/AR		1997 00:00
vz	1270	1300		Cavalcanti, Guido	Rime	Guido Cavalcanti (ed. Contini), 1270-1300 (fior.)	Poeti del Duecento, a cura di Gianfranco Contini, Milano-Napol..	Contini, G.	tosc.	fior.		lir. O V	1270-1300 E/VZ		1997 00:00
vw	1291	1300		Cavalcanti, Guido (?)	Ballate (Due)	Guido Cavalcanti (?), Due ball., XIII ex. (fior.)	Rime, a cura di Guido Favati, Milano-Napol..	Favati, G.	tosc.	fior.		lir. O V	sec. XIII ex. E/VW		1997 00:00
vv	1291	1310		Alfani, Gianni	Rime	Gianni Alfani, XIII/XIV (fior.)	Poeti del Duecento, a cura di Gianfranco Contini, Milano-Napol..	Contini, G.	tosc.	fior.		lir. O V	sec. XIII/XIV E/VV		1997 00:00
ij	1291	1316		Frescobaldi, Dino	Rime	Dino Frescobaldi, Poeti del Duecento	Poeti del Duecento	Contini, G.	tosc.	fior.		lir. O V	sec. XIII ex. E/IJ		1997 00:00

Fig. 3.2 Definizione del subcorpus del Dolce Stil Novo.

Gli autori scelti vengono abbinati alle scuole in base alla loro provenienza geografica in modo da poter studiare il lessico secondo la varietà volgare che usavano. Così vengono formati 5 subcorpora, relativamente organizzati in ordine cronologico, come segue:

- Scuola Siciliana (1250 circa), questo subcorpus è composto da Giacomo da Lentini, Ruggieri d'Amici, Odo delle Colonne, Guido delle Colonne, Stefano Protonotaro, Mazzeo di Ricco, Federico II, Re Enzo, Tommaso di Sasso, Cielo d'Alcamo e Jacopo Mostacci.
- Rime di Dante (1283 - 1305 circa), ne fanno parte i 54 componimenti più quelli di dubbia attribuzione.
- Dolce Stil Novo (1251 - 1328 circa), formato dai poeti fiorentini Guido Cavalcanti, Dino Frescobaldi, Lapo Gianni, Gianni Alfani e Verzellino.
- Scuola Bolognese (1251 - 1300 circa), costituito dai poeti bolognesi Guido Guinizzelli, Onesto da Bologna, Picciol da Bologna, Polo Zoppo da Castello, Semprebene da Bologna e Ranieri de' Samaritani.
- Rime di Cino da Pistoia (1336), sebbene il suo canzoniere comprenda più di 160 sonetti in questo subcorpus sono stati riuniti solo 119 poesie e 19 rime di dubbia attribuzione. Questa scelta è stata necessaria perché le opere di Cino sono divise in due file su Gattoweb e ci sono rime che compaiono in entrambi i file, di conseguenza prenderli tutti e due comporterebbe la ripetizioni delle occorrenze di alcune forme.

3.3.

Lista della forme grafiche ed eliminazione delle parole vuote

Gattoweb permette di ottenere in maniera molto semplice formari completi ed esclusivi, lemmari completi ed esclusivi e incipit delle opere quando un determinato subcorpus viene definito. Nel caso di questa ricerca, dato che si tratta di 5 subcorpora, bisogna selezionarne uno per volta in modo di avere la lista delle forme grafiche separate (v. fig. 3.3).

Ottenuto il formario completo si può procedere all'eliminazione delle parole vuote “cioè quelle parole come congiunzioni, preposizioni e articoli, che formano il *tessuto connettivo* della lingua di un testo”¹¹. Tuttavia per semplificare questa fase si è fatto uso dello strumento “*salva*” presente in Gattoweb, scaricando così in un file .rft la lista delle forme. Poi con un semplice copia e incolla, sono

11. Alessandro Lenci, Simonetta Magni, Vitto Pirelli, *testo e computer: Elementi di linguistica computazionale*, Carocci Editore, Roma, p. 136.

state utilizzate le forme per una nuova ricerca nel subcorpus, questa volta però attraverso “ricerca di contesti: per forme (da un testo)”. L'obiettivo è quello di ottenere oltre alle forme e alle occorrenze di queste anche i lemmi di quelle forme che sono già state catalogate dall'OVI (v. fig. 3.4). Facendo così si possono eliminare più rapidamente alcune parole vuote che compaiono lemmatizzate nella nuova lista ottenuta, evitando in questo modo possibili problemi dovuti a forme ambigue. Ovviamente non tutte le forme presenti in Gattoweb sono lemmatizzate, perciò una gran parte delle parole vuote presenti nei 5 subcorpora va controllata nel proprio contesto per poterla eliminare dal formario. I criteri utilizzati per eliminarle manualmente vengono presi dai lavori effettuati da studiosi ed esperti dell'italiano antico e grazie al contributo offerto da DanteSearch, dato che le opere di Dante sono un modello letterario che permette oggi di avere un riferimento sulla lirica italiana delle origini.

The screenshot shows the Gattoweb OVI interface for the subcorpus 'Dolce Stil Novo'. The page title is 'Corpus OVI dell'Italiano antico: formario completo dei sottocorpora (sottoc. A)'. The search results are displayed in a table with columns for selection, form, and number of occurrences.

sel.	forma	n. occ.
<input type="checkbox"/>	'	1
<input type="checkbox"/>	a	198
<input type="checkbox"/>	a'	4
<input type="checkbox"/>	à	2
<input type="checkbox"/>	abandona	5
<input type="checkbox"/>	abandonaro	1
<input type="checkbox"/>	abbandona	2
<input type="checkbox"/>	abbandonati	1
<input type="checkbox"/>	abbi	1
<input type="checkbox"/>	abbia	1
<input type="checkbox"/>	abbiate	1
<input type="checkbox"/>	abello	1
<input type="checkbox"/>	abito	1
<input type="checkbox"/>	abbraccia	2
<input type="checkbox"/>	abbracciar	1
<input type="checkbox"/>	absalone	1
<input type="checkbox"/>	accende	3
<input type="checkbox"/>	accesa	1
<input type="checkbox"/>	accese	1
<input type="checkbox"/>	accesi	1
<input type="checkbox"/>	accidente	2
<input type="checkbox"/>	acciò	1
<input type="checkbox"/>	acciso	1
<input type="checkbox"/>	accoglienza	2
<input type="checkbox"/>	accompagnata	1
<input type="checkbox"/>	accompagnato	1
<input type="checkbox"/>	acconci	1
<input type="checkbox"/>	acconcio	2
<input type="checkbox"/>	accordellata	1
<input type="checkbox"/>	accorta	4
<input type="checkbox"/>	accorte	2

Fig. 3.3 Formario completo del subcorpus del Dolce Stil Novo.

Corpus OVI dell'Italiano antico: coppie forma-lemma localizzate - (sottoc. A)

Copia in accumulatore | Mostra solo forme | Sel./desel... | Salva... | Vai a... | Riavvia GattoWeb | Guide..

sel.	forma	lemma	cat. gr.	disambiguatore	n. occ.	commento
<input type="checkbox"/>	a				118	
<input type="checkbox"/>	a		prep.		31	
<input type="checkbox"/>	a	a	prep.] + inf.	8	
<input type="checkbox"/>	a	a	prep.] + pers.	23	
<input type="checkbox"/>	a	a	prep.] ^art.det.	18	
<input type="checkbox"/>	a'				1	
<input type="checkbox"/>	a'	a	prep.] ^art.det.	3	
<input type="checkbox"/>	à				2	
<input type="checkbox"/>	abandona				3	
<input type="checkbox"/>	abandona	abbandonare	v.		2	
<input type="checkbox"/>	abandonaro				1	
<input type="checkbox"/>	abbandona	abbandonare	v.		2	
<input type="checkbox"/>	abbandonati	abbandonare	v.		1	
<input type="checkbox"/>	abbi	avere	v.		1	
<input type="checkbox"/>	abbia				1	
<input type="checkbox"/>	abbiate	avere	v.		1	
<input type="checkbox"/>	abello	abele	n.p.	antr.::	1	
<input type="checkbox"/>	abito	abito	s.m.		1	
<input type="checkbox"/>	abbraccia				1	
<input type="checkbox"/>	abbraccia	abbracciare	v.		1	
<input type="checkbox"/>	abbracciar	abbracciare	v.		1	
<input type="checkbox"/>	absalone	assalonne	n.p.	antr.::	1	
<input type="checkbox"/>	accende				1	
<input type="checkbox"/>	accende	accendere	v.		2	
<input type="checkbox"/>	accesa	acceso	agg.		1	
<input type="checkbox"/>	accese	accendere	v.		1	
<input type="checkbox"/>	accesi				1	
<input type="checkbox"/>	accidente	accidente	s.m.		2	
<input type="checkbox"/>	acciò				1	
<input type="checkbox"/>	acciso	aucidere	v.	accidere	1	

Fig. 3.4 Coppie forma-lemma del subcorpus del Dolce Stil Novo.

3.4.

Costruzione del Database MySQL

L'eliminazione di tutte le parole vuote porta alla completezza dei formari contenenti le parole lessicali che servono per la ricerca degli elementi in comune tra i subcorpora. A questo proposito è stato creato un database MySQL che permette di organizzare e di relazionare un insieme di dati nelle varie tabelle di cui è composto e in questa maniera averne una veloce accessibilità. Anche se Gattoweb conta con uno strumento che permette di confrontare i subcorpora definiti, l'utilizzo di un database permette di ottenere:

- Il lessico in comune tra i 5 subcorpora e non soltanto di mettere a confronto tutte le forme presenti.
- La lista degli hapax¹² e il loro numero totale in ogni subcorpus. Il numero totale degli hapax permette di effettuare il calcolo della ricchezza lessicale¹³ e in Gattoweb non è possibile ottenere questi dati.

12. Forma linguistica che ricorre una volta soltanto all'interno di un testo, di un autore o di una lingua.

13. La ricchezza lessicale è il rapporto tra le parole tipo e le parole unità (type/ token ratio) I valori del rapporto vanno da 0 a 1.

- La possibilità di utilizzare i dati ricavati così come i risultati raggiunti e di presentarli attraverso l'uso del web e di altre applicazioni.

Il database MySQL è composto da tre tabelle per quel che riguarda la ricerca del lessico comune tra i subcorpora, una chiamata “scuole”, una chiamata “scuoleForme” e l'ultima chiamata “forme”.

La prima tabella contiene i dati di ogni scuola ed è composta da 7 campi:

- Il campo “id”, è la chiave primaria della tabella cioè l'identificatore univoco di ogni scuola. E' di tipo INT o intero.
- Il campo “nome”, contiene il nome di ogni scuola ed è di tipo CHAR e massimo di 50 caratteri.
- I campi “tokenTot”, “formeTot”, “parolePieneTot”, “ricchezzaLess” e “freqM”, contengono rispettivamente il numero totale dei token, delle forme e delle parole piene o lessicali così come il valore di ricchezza lessicale e di frequenza media.

La seconda tabella (detta “ponte”) serve a collegare la tabella “scuole” e la tabella “forme”. Le tabelle “ponte” si rendono necessari quando un record può essere relazionato a più di un record di una tabella e viceversa (relazione molti a molti). Quindi questo tipo di tabella costituisce un ponte tra le due da relazionare. I campi di cui è composta sono:

- Il campo “idScuola”, Foreign Key o identificatore della scuola alla quale appartiene la singola forma. E' proprio questa la chiave che relaziona le due tabelle e deve essere dello stesso tipo e dimensione della chiave a cui fa riferimento.
- Il campo “idforma”, anche in questo caso Foreign Key o identificatore della forma che relaziona la tabella “forme” a “scuoleForma”.
- Il campo “occorrenze” fa riferimento al numero di volte che una stessa forma appare in un subcorpus. E' di tipo intero (INT).

Questi 3 campi insieme costituiscono una UNIQUE KEY ovvero un vincolo che si applica a un campo o un insieme di campi di una tabella e impone che i valori dell'attributo (o le ennuple di valori sull'insieme di attributi) siano una Primary Key.

La terza tabella invece contiene i dati riguardanti le forme presenti in ogni subcorpora ed è composta da 3 campi:

- Il campo “id”, identificatore univoco di ogni forma e di tipo intero (INT).
- Il campo “forma” contiene la forma grafica così come si trova in ogni poesia.

Nello stesso database MySQL sono state create altre tabelle che permettono di collegare le informazioni di ogni poeta con le loro opere e successivamente con le scuole e le varie forme. Queste tabelle servono alla presentazione dei dati raggiunti grazie alla ricerca eseguita.

Una volta creato l'intero database MySQL si può passare alla fase successiva, quella di aggiungere i dati nel database.

3.5.

Inserimento dei dati nel Database

Solitamente per aggiungere una riga all'interno di una tabella di un database MySQL viene utilizzato il comando INSERT INTO con la seguente sintassi:

```
INSERT INTO nome_tabella (nome_colonne)
VALUES (valore_campo)
```

In questo modo viene mostrato a quale tabella e a quale colonna devono essere aggiunti i dati, insieme a una lista di valori. Quando invece devono essere inserite più righe all'interno di una sola tabella si possono usare tante istruzioni INSERT INTO quanto sono le nuove righe. Tuttavia MySQL permette gli inserimenti multipli anche con una sola istruzione INSERT INTO, in cui ogni VALUES deve essere separato dall'altro con una virgola (,).

Il problema con queste due operazioni è il calo delle prestazioni e soprattutto lo spreco di codice inutile. Nello specifico, in questa ricerca è necessario inserire 20.000 ennuple circa nella tabella “forme” del database MySQL e con il comando INSERT INTO diventa un processo molto lungo, soggetto anche a errori umani durante l'inserimento. Fortunatamente in MySQL esiste più di un metodo per importare molti dati in una volta sola. Difatti possono essere utilizzate senza distinzioni sia l'istruzione MySQLimport che l'istruzione LOAD DATA INFILE, perché con entrambe si ottengono gli stessi risultati. In questo caso è stato preferito il comando LOAD DATA INFILE, la cui sintassi completa è la seguente:

```
LOAD DATA [LOW_PRIORITY | CONCURRENT] [LOCAL]
INFILE 'nome_file'
[REPLACE | IGNORE]
INTO TABLE nome_tabella
[FIELDS [TERMINATED BY 'char']]
[[OPTIONALLY] ENCLOSED BY 'char']
[ESCAPED BY 'char' ]]
```

```
[LINES [TERMINATED BY 'char']  
[IGNORE num LINES]  
[(colonna_nome1,colonna_nome2,...)]
```

Il comando `LOAD DATA INFILE` presenta delle proprietà opzionali (tra parentesi quadre come `LOW_PRIORITY`, `CONCURRENT`, `REPLACE`, `IGNORE`, ecc) che ai fini di questa ricerca non vengono utilizzate¹⁴. Di conseguenza può essere semplificato quanto fornito dal manuale MySQL in questo modo:

```
LOAD DATA INFILE './path-to-file.csv'  
INTO TABLE miodatabase.tabella  
FIELDS TERMINATED BY ";"  
LINES TERMINATED BY "\n";
```

Le opzioni `FIELDS` e `LINES` indicano rispettivamente come sono strutturati i dati nel file ovvero indicano il separatore tra i campi e il termine della riga.

Il comando `LOAD DATA INFILE` può essere eseguito dal client MySQL o dall'interno di un'applicazione. Basta semplicemente esportare un foglio excel contenente i dati in un file di testo in formato CSV (Comma Separated Values). Questo tipo di file memorizza i dati in un formato i cui campi sono separati da un segno di interpunzione e le righe da un a capo. Da ricordare però che si devono tenere alcuni accorgimenti per caricare i dati con successo:

- Il file CSV deve avere lo stesso numero di colonne della tabella nella quale saranno inseriti i dati del file.
- Il set di caratteri del file CSV deve essere lo stesso del database. Per default il charset è Unicode UTF8.
- I separatori di campo e di testo nel file CSV devono essere il punto e virgola (;) e i doppi apici (“”).
- E ultimo ma molto importante, si devono disabilitare le FOREIGN KEY (se presenti) della tabella in modo di evitare la violazione delle regole di integrità referenziali¹⁵.

14. Per approfondimenti consultare Luke Welling, Laura Thomson, MySQL Tutorial, Milano, Pearson Education Italia S.r.l, pp. 73-75.

15. L'applicazione dell'integrità referenziale assicura che le relazioni definite tra le tabella non possano essere violate. Essa coinvolge due tabelle: una referenced table (tabella referenziata) che contiene la PRIMARY KEY (PK) e una referencing table (tabella referenziante) che si lega ad essa attraverso una FOREIGN KEY (FK). Più precisamente, la FK può far riferimento a una o più colonne su cui è definito un vincolo di unicità (UNIQUE CONSTRAINT), e in ogni caso

Quando il file CSV è pronto e le FOREIGN KEY della tabella sono state disabilitate, dall'interfaccia PHPMyAdmin dell'applicazione (come EasyPHP) si sceglie la tabella del database nella quale si vogliono caricare i dati. A questo punto non rimane che importare il file dalla pagina PHPMyAdmin e selezionare come formato del file importato “CSV usando LOAD DATA” (v. fig. 3.5).

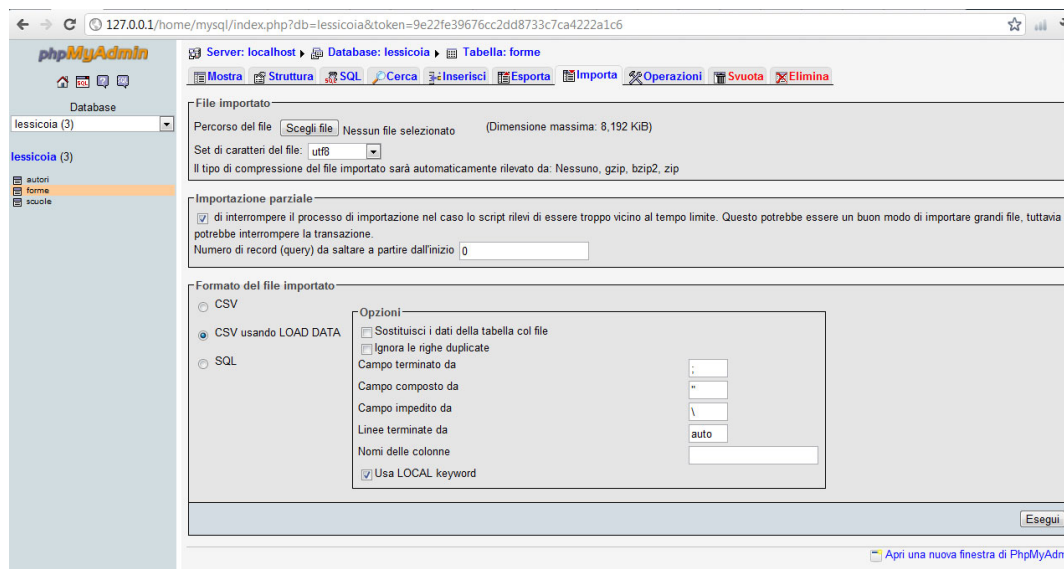


Fig. 3.5 Pagina PHPMyADMIN – Sezione Importa

Quando l'importazione è stata effettuata bisogna riabilitare le FOREIGN KEY della tabella, in modo da gestire i dati appena caricati nel database MySQL.

3.6.

Ricerca delle forme grafiche comuni ai 5 subcorpora

Il passo successivo è quello di collegarsi al database per ottenere le forme grafiche comuni ai 5 subcorpora. Per accedere con facilità ai dati presenti nel database è necessario eseguire delle query, ovvero interrogazioni che si fanno al database. Per formulare le query è fondamentale l'uso di un linguaggio di interrogazione come SQL (Structured Query Language) che consenta la manipolazione e la gestione dei dati memorizzati nel database.

occorre che il tipo di dato delle colonne in relazione sia identico (fatta eccezione per l'attributo di [NOT] NULL).

Per interrogare il database è necessario anche un linguaggio di programmazione in grado di effettuare le query SQL e di visualizzare i risultati. Per semplicità viene utilizzato il linguaggio PHP che si integra perfettamente con i database MySQL. PHP è un linguaggio di programmazione che viene interpretato da un server web apache e genera codice HTML dinamico. La sua sintassi è molto simile a quella dei linguaggi C e Perl.

A questo punto si passa all'elaborazione del codice PHP, composto dalla connessione al database, le query SQL e le varie funzioni e variabili. Nel caso delle query SQL, inizialmente è stata valutata la possibilità di sfruttare l'INTERSECT, l'operatore insiemistico dell'algebra relazionale. Infatti l'operatore INTERSECT restituisce i valori comuni a tutti gli insiemi coinvolti. Il problema è che l'operatore INTERSECT non è supportato da MySQL. In alternativa è stato scelto di usare l'operatore GROUP BY, la condizione HAVING e le funzioni aggregative COUNT e SUM, scelta che si è dimostrata efficiente e poco dispendiosa.

A questo punto, ecco il codice PHP finale della pagina formeComuni.php:

```
<?php
    include ("connessione.php");
    $count=0;
    $sql=" SELECT scuoleforme.idforma, forme.forma,
            SUM(scuoleforme.occorrenze) AS tot
        FROM scuoleforme JOIN forme.forma
            ON scuoleforme.idforma=forme.id
        GROUP BY scuoleforme.idforma
        HAVING (count(*) = 5)
        ORDER BY tot DESC";
    $res=mysql_query($sql,$conn)or die( "Errore orrore"
        .mysql_error());

    while ($records=mysql_fetch_row($res))
    {
        count=$count+1;
        $idforma=$records[0];
        $forma=$records[1];
        $tot=$records[2];
        echo" $count <a href='vediforma.php?idforma=$idforma'>$forma</a>
$tot
        </br>";
    }
```

?>

In pratica, il primo comando che viene eseguito è `include("connessione.php")`; che serve per includere il file contenente i parametri per accedere al database. Subito dopo viene creata una variabile che ha il compito di contare il totale delle forme comuni. Invece la query SQL, attraverso l'operatore GROUP BY, raggruppa tutte le forme grafiche in base al "idforma" della tabella "scuoleforme". Successivamente attraverso la condizione nel HAVING sono selezionati soltanto i gruppi composti da 5 ennuple. In altre parole vengono estratti soltanto i gruppi in cui ci sono 5 ripetizioni dell'attributo "idforma". Infine nel SELECT sono recuperati i dati dell'idforma, delle forme e il totale delle occorrenze (o "tot") in tutto il corpus di ogni forma in comune. Quest'ultimo si ottiene grazie alla somma eseguita dalla funzione aggregativa SUM. Invece ORDER BY presenta le forme comuni in ordine decrescente in base al totale delle occorrenze.

Una volta che la query SQL è eseguita, i risultati ottenuti passano alla variabile \$res. Poi attraverso la funzione PHP `mysql_fetch_row` si ottiene un array() ordinato con tutti i dati di questa variabile, trasferiti riga per riga. I campi delle tabelle sono restituiti nell'ordine richiesto nella SELECT. Questo vuol dire che le variabile \$idforma, \$forma e \$tot contengono rispettivamente i dati dei campi "idforma", "forma" e "tot".

Poiché le occorrenze di ogni forma in comune devono essere analizzate nel dettaglio, si è scelto di visualizzarle singolarmente. A questo scopo, nel comando `echo` è stato inserito un link collegato a ogni forma ed è stata aggiunta una query string al link. Una query string è una parte dell'URL (Uniform Resource Locator) che inizia con un punto interrogativo (?), quindi in questo caso, la query string è `?idforma=$idforma`. La sua funzione è quella di passare dalla pagina `formeComuni.php` alla pagina `vediForme.php` i parametri all'interno della variabile \$idforma:

```
<?php
    $idforma=$_GET['idforma'];
    include("connessione.php");
    $sql="SELECT forme.forma, scuoleforme.occorrenze,
           scuole.nome, scuole.id
    FROM scuole JOIN scuoleforme
           ON scuole.id=scuoleforme.idscuola
           JOIN forme
           ON scuole.idforma=forme.id
    WHERE forme.forma='$forma'
    ORDER BY scuole.id ASC";
    $res=mysql_query($sql,$conn) or die( "Errore orrore "
```

```
.mysql_error());

while ($records=mysql_fetch_row($res))
{
    $forma=$records[0];
    $occorrenza=$records[1];
    $scuola=$records[2];

    echo"$scuola $forma ($occorrenza) </br>";
}
?>
```

Infatti con il metodo `$_GET['idforma'];` viene recuperato il parametro contenuto nella variabile `$idforma`. In questo modo può essere associato l'identificatore univoco della forma presente nella tabella "scuoleforme" a quello nella tabella "forme". Quindi il risultato finale che si ottiene è la visualizzazione di ogni forma singolarmente, accompagnata dalle occorrenze per scuola.

Aspetti quantitativi del corpus

In questo capitolo vengono presentati i dati quantitativi estratti dal testo che consentono di osservare la distribuzione delle parole al suo interno così come la sua struttura e la sua organizzazione. Lo stesso procedimento è stato svolto nel caso dei 5 subcorpora in modo di confrontare le varie misure lessicometriche.

4.1.

Analisi dei dati quantitativi del corpus

Ai fini di un'analisi statistica, che consenta di capire com'è organizzato il corpus, è importante conoscere alcune informazioni a livello quantitativo (v. Fig. 5.1):

Word Token ¹⁶ (N)	74151
Word Type (V)	9317
Type / Token Ratio (TTR) V/N	0,13
V ₁ /V ¹⁷	0,59
Freq.Media N/V	7,96

Fig. 5.1 Dati quantitativi del Corpus

La prima informazione ottenuta riguarda la voce *word token* (N), che indica il totale delle occorrenze presenti ovvero la dimensione del corpus. In questo caso la dimensione del corpus è N=74.151. In generale, quando un corpus si aggira tra le 70.000 e 100.000 occorrenze è considerato di dimensione medio-grande. *Word type* invece si riferisce al numero di forme grafiche diverse e permette di sapere la dimensione del vocabolario del corpus. Quindi la dimensione del vocabolario di questo corpus è

16. Un token è l'unità più piccola in un corpus, composto da una sequenza di caratteri delimitati da spazi. Per approfondimenti si consiglia Alessandro Lenci, Simonetta Magni e Vito Pirrelli, *Testo e Computer. Elementi di linguistica computazionale*, Carocci, 2005.

17. V₁ Rappresenta l'insieme delle forme che appaiono una sola volta (chiamate *hapax*) mentre V rappresenta il totale delle forme diverse.

V=9317.

In base a questi dati è possibile calcolare la *Type/Token Ratio* (TTR) o ricchezza lessicale del corpus. Si tratta di un rapporto tra il totale delle forme diverse (V) e il totale delle occorrenze del corpus (N) i cui valori sono compresi tra 0 e 1. Proprio la tendenza di questo rapporto verso i valori 1 o 0 indica se il vocabolario di un testo è più o meno ricco. Pertanto il vocabolario di questo corpus è meno ricco dato che la *Type/Token Ratio* (TTR) è $V/N = 0,13$.

Un'altra misura utile per calcolare la variabilità lessicale del corpus è attraverso il rapporto tra il numero di *hapax* (V_1) e il totale delle forme diverse che formano il vocabolario (V). Nel corpus il numero di *hapax* è uguale a 5484 e quindi il rapporto V_1/V è uguale 0,59. Questo vuol dire che le forme che compaiono nel testo una volta solo sono moltissime e che il 58% del vocabolario del testo è composto di queste.

L'ultimo parametro è riferito alla *frequenza media* generale (N/V). Questo indice inverso di ricchezza lessicale viene calcolato dal rapporto tra il totale delle occorrenze del corpus (N) e il totale delle forme diverse (V). Questo significa che maggiore sarà il totale delle forme del vocabolario, tanto minore sarà la loro frequenza media. Nel corpus la frequenza media è $N/V = 7,96$.

4.2.

Analisi dei dati quantitativi dei 5 subcorpora

Il corpus è suddiviso in 5 subcorpora per poter confrontarli e trovare il lessico comune a tutti. Di conseguenza è stato possibile ottenere alcune informazioni dal punto di vista quantitativo per ognuno di essi.

Oltre ai dati sulla dimensione del corpus, del vocabolario e sui vari rapporti lessicali, è stato aggiunto anche il numero di rime per scuola, in modo di valutare la loro percentuale nel corpus così come analizzare quale scuola poetica presenta i componimenti più lunghi (v. Fig. 4.2).

Come si può notare dalla tabella, il numero dei componimenti varia notevolmente da alcuni subcorpora ad altri. Il subcorpus di Cino da Pistoia (138 componimenti) e quello degli stilnovisti (101 componimenti) si presentano come quelli più “produttivi”. Successivamente si trovano quelli della scuola siciliana e di Dante con 76 componimenti entrambi. Invece il subcorpus bolognese (55

componenti) è quello più esiguo a confronto con gli altri 4. Questo perché non è stato possibile aggiungere tutti i componenti dei poeti bolognesi del periodo studiato, dato che finora non sono presenti nel database fornito dal TLIO.

	N° Rime	% Rime	Word Token (N)	Word Type (V)	Type/Token Ratio (TTR) V/N	V ₁ / V	Freq. Media N/V
Scuola siciliana	76	17,04%	17552	3732	0,21	0,63	4,70
Rime di dante	76	17,04%	14183	3132	0,22	0,62	4,53
Dolce Stil Novo	101	22,65%	16416	3296	0,20	0,59	4,98
Scuola Bolognese	55	12,33%	7702	2306	0,29	0,68	3,34
Cino da Pistoia	138	30,94%	18298	3276	0,18	0,58	5,59
	446	100,00%	74151				

Fig. 4.2 Dati quantitativi del Corpus suddiviso per scuola poetica.

Inoltre i dati quantitativi per ogni sottocorpus forniscono alcune informazioni interessanti. Precisamente, il subcorpus di Cino da Pistoia contiene il maggior numero di componenti e il maggior numero di token (18298) ma non il maggior numero di forme grafiche (3276). Infatti sorprende ancora di più che sia il subcorpus dei siciliani ad averne il maggior numero (3732) e che oltre a questo, con solo 76 componenti, abbia un maggior numero di occorrenze (17552) rispetto al subcorpus stilnovista (101 componenti, 16416 token e 3296 forme grafiche). Mentre nei subcorpora di Dante (14183 token, 3132 forme grafiche) e della scuola bolognese (7702 token, 2306 forme grafiche) tutto corrisponde alla norma. Queste informazioni permettono di dedurre che i componenti della scuola siciliana siano abbastanza più lunghi e con una maggiore ricchezza lessicale rispetto ai componenti delle altre scuole. Infatti in base al totale dei componenti e al totale dei token in ogni subcorpus, si può calcolare anche la lunghezza media dei componenti per scuola. Dai calcoli effettuati, i componenti della scuola siciliana hanno una lunghezza media di 230 token, seguita dai componenti danteschi con 186 token. I bolognesi, invece, anche se con un numero piuttosto basso di componenti, presentano una lunghezza media (140 token) abbastanza alta se messa a confronto con quelle degli stilnovisti (162 token) e Cino da Pistoia (133 token).

Riguardo i valori di varietà lessicale misurata dal Type/Token Ratio (V/N) nei 5 subcorpora, soltanto in due casi è stata rispettata la soglia consigliata di 0,20, specificamente nei casi di Cino da Pistoia (N/V=0,18) e degli stilnovisti (N/V=0,20). Invece nei casi dei subcorpora rimanenti, il valore è di

poco oltre la soglia nella scuola siciliana ($N/V=0,21$) e in Dante ($N/V=0,22$) mentre nella scuola bolognese ($N/V=0,29$) è stato ampiamente superato. Ovviamente per non superare tale soglia ci deve essere un rapporto proporzionale tra il totale delle occorrenze (N) e il numero delle forme grafiche (V). L'ultima considerazione legata alla ricchezza lessicale concerne gli *hapax* che nei 5 subcorpora superano abbondantemente lo 0,50. I valori sono molto alti ma allo stesso tempo questa caratteristica diventa rilevante, dato che in alcuni casi l'uso di un *hapax* si deve a scelte stilistiche come esigenza di parole in rima, innovazioni semantiche, creare parole nuove, ecc che diventano un marchio o un segno distintivo.

Infine in base ai valori della frequenza media in ordine decrescente si presenta al primo posto il subcorpus di Cino da Pistoia (5,59), seguito da quelli degli stilnovisti (4,98), della scuola siciliana (4,53), di Dante (4,53) e della scuola bolognese (3,34). Questi valori indicano che a seconda della dimensione del corpus o a seconda del numero totale delle forme del vocabolario la frequenza media può aumentare o diminuire.

Forme e occorrenze significative

Dai dati ottenuti risulta che le forme comuni delle 5 scuole poetiche sono 442 per un totale di 22.428 occorrenze. Visto che la dimensione dell'intero corpus è formato da 74.151 occorrenze, le forme comuni rappresentano il 30,24%.

Il numero delle forme comuni ottenuto non è molto grande ma si è deciso comunque di analizzare quelle forme che presentano occorrenze significative se messe a confronto tra i subcorpora. Inoltre la lista di forme comuni è stata ulteriormente ridotta eliminando gli aggettivi e gli avverbi deittici così come gli avverbi di negazione o i verbi ausiliari (come essere e avere).

5.1.

Mente, umile e spirito

*Mente, umile*¹⁸ e *spirito* sono forme che presentano occorrenze contrastanti da un subcorpus all'altro (v. Fig. 5.1).

forma	totale	siciliani	dante	stilnovo	bolognese	cino
mente	161	13	29	63	6	50
Umile	28	4	4	12	1	7
spirito	72	7	7	37	1	20

Fig. 5.1 Occorrenze di *mente, umile e spirito*.

In base alla tabella, il maggior numero di occorrenze di *mente* (63) si trova nel subcorpus degli stilnovisti così come quello di *spirito* (37 occorrenze) e di *umile* (12 occorrenze). Questi dati non sono sorprendenti se osservati nel subcorpus degli stilnovisti perché queste forme appartengono al lessico tipico di questa scuola. Sono ricorrenti soprattutto in Guido Cavalcanti che ne sfrutta le possibilità espressive con l'intento di rendere tangibili e visibili i moti dell'anima, trasformandoli in

18. Le occorrenze complessive di *umile* si riferiscono non solo a quella forma grafica ma anche a *umile* (con l'accento sulla seconda sillaba perché di origine provenzale).

fantasmi di persone con attenta introspezione psicologica. Nel pensiero cavalcantiano la forma *mente* non ha il significato di intelletto ma viene interpretato, secondo la terminologia di Averroé¹⁹, come l'immaginazione o la memoria. Quindi per Cavalcanti la mente non è la sede della conoscenza razionale ma “è lo spazio entro il quale si consuma il dramma dell'anima e del cuore” (De Robertis 1986, p. 28). Con questo nuovo uso di *mente*, Cavalcanti cambia la metafora canonica precedente in cui la figura della donna attraversava gli occhi e arrivava al cuore²⁰. Addirittura in questo senso raggiunge un elevato grado di innovazione con la sua famosa metafora pluviale in “S'io prego questa donna che Pietate” al v. 12-12: *Allor mi par che ne la mente piova / una figura di donna pensosa*. Anche la forma *Spirito* (così come *spirto e spirti*) ha una concezione tecnico-filosofica di corpi sottili grazie ai quali si svolgono le operazioni sensoriali dell'uomo (come la vista) o i gesti dell'anima (come il tremore). Con Cavalcanti, la fenomenologia degli spiriti entra definitivamente come tecnica espressiva nel tema amoroso del linguaggio poetico. Invece la forma *umile*, in senso stilnovistico, rappresenta la fonte del bene, la benevolenza ovvero una delle virtù ideali.

Il secondo subcorpus con il maggior numero di occorrenze è quello di Cino da Pistoia: *mente* (50 occorrenze), *umile* (7 occorrenze) e *spirito* (20 occorrenze). La poesia ciniana, attraverso l'uso di queste forme, fa una citazione importante dei concetti dottrinali del nuovo stile, in molti casi addirittura ingigantendoli. Un esempio di questo viene dato dall'uso di *spirito* in “Tutto mi salva il dolce salutare”: si passa dall'immagine dantesca di uno spirito che muove la bocca della donna all'immagine dello spazio pieno di spiriti d'amore, come se questi fossero quasi una caratteristica dell'aria. A Questa connotazione “atmosferica” manca però il carattere scientifico presente nella poetica di Cavalcanti.

Anche nelle rime di Dante sono frequenti *mente* (29 occorrenze), *umile* (4 occorrenze) e *spirito* (7 occorrenze). Queste 3 forme sono per lo più sotto l'influsso stilnovistico, in prevalenza quello di Cavalcanti. *Mente*, quella più usata da Dante, si trova principalmente nelle cosiddette rime stilnovistiche. Il suo valore semantico assume molti significati ma nella maggior parte dei casi è legata concettualmente alla memoria come nei vv. 20, 44, 87 nella rima “E' m'incresce di me sì duramente”.

Tuttavia la rilevanza di questi dati trova il suo peso nei dati forniti dalla scuola bolognese: *mente* (6 occorrenze), *umile* e *spirito* (entrambe 1 occorrenza). Il primo dato curioso si ottiene dall'iniziatore dello Stilnovo, Guinizzelli. Infatti *mente*, che è una parola-chiave nel nuovo stile, occorre soltanto 1 volta nei suoi componimenti, specificamente nel sonetto indirizzato a Guittone d'Arezzo “[O] caro

19. Averroè era un commentatore arabo di Aristotele, il cui obiettivo era quello di chiarire il significato autentico della filosofia aristotelica. Per approfondimenti: <http://www.unisi.it/ricerca/prog/fil-med-online/autori/htm/averroee.htm>

20. *Da Guido Guinizzelli a Dante: nuove prospettive sulla lirica del duecento*, a cura di Furio Brugnolo e Gianfelice Peron, Monselice, Il Poligrafo, 2004, pp. 212-213.

padre meo, de vostra laude”, v. 3: *ché 'n vostra mente intrar vizio non aude*. Questi versi mostrano come inizialmente Guinizzelli fosse ancora attaccato ad alcuni aspetti della tradizione precedente e ancora lontano ideologicamente da quello che diverrà lo Stilnovo. Comunque, anche se in questi versi *mente* è in sostanza lontana e isolata dalla concezione stilnovistica, molti studiosi ritengono che questo passaggio sia servito a Cavalcanti come spunto per cambiare la sede dell'amore dal cuore alla mente²¹.

In ogni caso l'informazione più interessante è senza dubbio fornita da Onesto da Bologna. Non solo ricorre nei suoi componimenti la forma *mente* (3 volte) ma è presente anche l'unica occorrenza di *umile* dell'intero subcorpus bolognese. Entrambe queste forme sono usate insieme in “Mente e umile e più di mille sporte”. Questo è uno dei 5 sonetti di corrispondenza indirizzato a Cino da Pistoia nei quali si svolge la forte diatriba tra i due poeti, dovuta alla concezione molto diversa che hanno entrambi della poesia. Nei versi di “Mente e umile e più di mille sporte” Onesto critica molto ironicamente gli elementi stilnovisti e li definisce monotoni e ripetitivi. Rivolge il suo rifiuto anche ai concetti filosofici come quello degli spiriti²² per i quali assume un tono burlesco accostandoli con sporte. Per Onesto da Bologna *'l'andar filosofando'* di Cino “ (...) riduce questa poesia ad una fredda e tecnica *'manera'* e gli fa rappresentare colloqui e soliloqui fittizii (...) ” (Libertini 1974, p.15). Le altre occorrenze di *mente* nelle opere di Onesto da Bologna sono presenti in “Quel che per lo canal perde la mescola” al v.14 e “ Se con lo vostro val mio dire e solo” al v.33, con il valore tradizionale di intelletto. Anche nella canzone “Prego, ch'audir ve piazza me, Pizòlo” ai vv. 39, 45 di Picciol da Bologna sono presenti 2 occorrenze di *mente* con lo stesso significato.

Per ciò che riguarda invece la scuola siciliana, le forme *mente* (13 occorrenze), *umile* (4 occorrenze) e *spirito* (7 occorrenze) sono presenti per lo più nei componimenti di Giacomo da Lentini e Guido delle Colonne. Per i siciliani, *mente* (la forma più interessante) vuol dire “contemplare con attenzione” o fissare l'immagine della donna amata come nella similitudine del pittore in “Uno disio d'amore sovente” al v.2 di Giacomo da Lentini. Quindi per questi poeti *mente* viene usata per affermare quanto sia mentale l'esperienza amorosa del poeta.

5.2.

21. Ibid., p. 213.

22. La forma utilizza da Onesto da Bologna in “Mente e Umile e più di mille sporte” è *spirti*, il cui lemma nel TLIO è quello di spirito.

Viso e cera

Le forme *viso* e *cera* presentano occorrenze interessanti all'interno dei subcorpora studiati (v. Fig. 5.2).

forma	totale	siciliani	dante	stilnovo	bolognese	cino
cera	25	11	1	2	2	9
viso	82	38	7	16	6	15

Fig. 5.2 Occorrenze di *viso* e *cera*.

Quello che colpisce maggiormente nei dati presenti nella tabella è quell'unica occorrenza di *cera* nelle rime di Dante. Questa forma si aggiudica un'importanza particolare proprio perché si tratta di un 'hapax dantesco' ma anche per i molteplici significati che possedeva nell'italiano antico. Precisamente una di queste accezioni è quella di “volto” nel sonetto “Di donne io vidi una gentile schiera” ai vv. 5-7. Nello specifico si tratterebbe di “ (...) occhi e volto di madonna abitati da Amore (...)” (Contini 1946, p.71).

Allo stesso modo le occorrenze della scuola siciliana forniscono dati interessanti. La forma *cera* ha 11 occorrenze nei poeti siciliani e presenta diversi significati a seconda del contesto. Infatti, secondo il TLIO²³, nella scuola siciliana *cera* ha il concetto di:

- “Parte esterna del corpo umano delimitata dall'attaccatura dei capelli e dalla linea della mandibola; il volto” come in Giacomo da Lentini, “Guiderdone aspetto avere” al v.48;
- “Pasta solida dalla bassa temperatura di fusione prodotta dalle api” come in Guido delle Colonne, “ La mia vit'è sì fort'è dura e fera “ al v.3;
- “Immagine” come in Giacomo da Lentini, “Madonna mia, a voi mando” al v.43;
- “Il messaggio più o meno volontariamente trasmesso dai movimenti, dalle espressioni e dall'apparenza del volto (...)” come in Guido delle Colonne , “Amor, che lungiamente m'hai menato” al v. 47.²⁴

Cera, intesa come volto o faccia, viene trasmessa dai siciliani e arriva fino a Cino da Pistoia. Dopo quello dei siciliani, il suo sottocorpus è quello che presenta più occorrenze di *cera* interamente con il

23. Alla pagina <http://tlio.oivi.cnr.it/TLIO/> è possibile anche consultare le voci del dizionario che sono state inserite e contestualizzate in base al significato.

24. Per ulteriori approfondimenti consultare <http://tlio.oivi.cnr.it/TLIO/> all voci *cera* (1) e *cera* (2).

senso di volto come in “Sì mi stringe l'amore” al v. 57. Anche se questa canzone presenta una tendenza stilnovistica spiccano elementi di stampo guittoniano e arcaico presenti nel testo tra cui appunto l'uso di questa forma.

Infatti *cera* con questo senso è quasi del tutto assente nel gruppo stilnovista. Ci sono soltanto 2 occorrenze, entrambe in Cavalcanti nelle rime “ Fresca rosa novella” al v. 23 e “ In un boschetto trova' pasturella” al v. 4. Anche il subcorpus dei bolognesi presenta solo 2 occorrenze di questa forma, nello specifico in Onesto da Bologna nella rima “Davante voi, madonna, son venuto” al v. 12 e in Semprebene da Bologna nella canzone “ Come lo giorno quand'è dal mattino ” al v. 36.

La seconda analisi riguarda le occorrenze della forma *viso*. Questa forma nei subcorpora di Dante (7 occorrenze), scuola bolognese (6 occorrenze), stilnovisti (16 occorrenze) e Cino da Pistoia (15 occorrenze) ha in alcuni casi il significato di “volto” mentre in altri ha valore di “vista” o di sguardo (atto di guardare) come in Guinizzelli “*Com'om che pinge bene / colora viso tale*” (Lo fin pregi' avanzato, vv. 18-19) e in Cavalcanti “*E non si po' conoscer per lo viso*” (Donna me prega, - per ch'eo voglio dire, v. 63) rispettivamente.

Invece le occorrenze di *viso* nella scuola siciliana forniscono informazioni stilistiche di grande rilevanza. In questo subcorpus, questa forma ha ben 38 occorrenze di cui 28 si trovano nei componimenti di Giacomo da Lentini. Dal punto di vista stilistico, il poeta siciliano fa un uso magistrale della forma *viso* in due dei suoi sonetti: “[E]o viso – e son diviso – da lo viso ” e “[L]o viso mi fa andare alegramente ”. Il primo sonetto è composto dalle molte ripetizioni della forma *viso* (ben 11 occorrenze) con significati diversi e dalle sue metamorfosi lessicali come nei vv. 1-5:

“[E]o viso – e son diviso – da lo viso,
e per aviso – credo ben visare;
però diviso – 'viso' – da lo 'viso',
c'altr'è lo viso – che lo divisare.”

Quasi come fosse un indovinello, questo sonetto è composto da parole con la stessa radice ma con significato diverso. In questi versi, ad esempio, *viso* assume sia valore verbale di “vedere” che sostantivale di “volto” e di “immagine”²⁵.

Nel secondo sonetto, quello in cui effettua la lode del viso, impiega questa forma 9 volte con diverse funzioni e con più significati. Inoltre sono molti gli stilemi che impiega intorno a questa forma come l'accostamento a un aggettivo di valore estetico attraverso il quale si individua il significato di volto.

Un esempio molto chiaro nei vv. 1-5:

25. Nel suo primo significato, *Viso* (termine lat.) aveva il significato non solo di faccia, volto ma anche di vista, sguardo e metonimicamente figura/immagine, da video, véggio (vedere). Cfr. Voce *Viso*, in *Dizionario Italiano Ragionato*, G. D'Anna Sintesi, Firenze, 1988, p. 1966.

“ Lo viso mi fa andare alegramente,
 lo bello viso mi fa rivegliare,
 lo viso mi conforta ispessamente,
 l'adorno viso che mi fa penare.”

5.3.

Gentile e vile

Dal punto di vista linguistico e filologico, le occorrenze di queste due forme rivelano informazioni molto importanti nei 5 subcorpora (v. Fig. 5.3).

forma	totale	siciliani	dante	stilnovo	bolognese	cino
gentil	96	5	19	21	12	39
gentile	41	1	9	14	4	13
vile	19	1	1	12	3	2

Fig. 5.3 Occorrenze di *gentile* e *vile*.

In ambito bolognese Guinizzelli usa le forme *gentile* (4 occorrenze) e *gentil* (12 occorrenze) per esprimere il superamento del concetto tradizionale di nobiltà di sangue con quello di nobiltà d'animo legata all'amore. Secondo questa nuova concezione, la nobiltà o gentilezza d'animo è una condizione che deriva dalla natura e rende l'uomo capace di provare sentimenti elevati. Quindi l'idea della tradizione feudale che la nobiltà appartenga solo agli aristocratici, perché acquisita per eredità, non ha spazio nella nuova visione comunale e borghese in cui gli uomini sono misurati in base alle loro qualità personali. Allo stesso tempo però, questa polemica implica anche che le doti umane e intellettuali di ogni individuo rappresentano un nuovo tipo di aristocrazia e sono la causa di un nuovo segno di disuguaglianza²⁶.

In Guinizzelli il tema della gentilezza è ancora messo in evidenza maggiormente con l'uso della forma *vile* (12 occorrenze) in contrapposizione alla forma *gentile* come in “Io voglio del ver la mia donna laudare” nei vv. 9-12:

26. Carlo Salinari, *La poesia lirica del duecento*, Napoli, Liguore Editore, 1968, pp. 164-166

*Passa per via adorna, e sì gentile
ch'abassa orgoglio a cui dona salute,
e fa 'l de nostra fé se non la crede;
e no· lle po' apressare om che sia vile;*

Il manifesto della nuova concezione è senza dubbio però “Al cor gentil rempaira sempre amore”, dove Guinizzelli usa la forma vile 2 volte e complessivamente 12 volte le forme gentile e gentil. In questo modo introduce il concetto dell'amore legato a quello della gentilezza, modernizzando il rapporto tradizionale della lirica cortese in cui il nesso era tra amore e cortesia.

Il significato guinizzelliano di nobiltà è presente negli stilnovisti attraverso le ricorrenze di *gentile* (14 occorrenze) e *gentil* (21 occorrenze) così come la forma *vile* (12 occorrenze) in contrapposizione. Oltre a ciò, in questi poeti c'è un uso ricorrente della rima in ILE che coinvolge forme come umile, sottile, gentile e vile come nel caso di “Posso degli occhi miei novella dire” nei vv. 12-15 di Guido Cavalcanti.

Nel caso di Cino da Pistoia c'è una ripresa di *gentile* (13 occorrenze) e *gentil* (39 occorrenze) dagli stilnovisti. In questo senso però, c'è un uso frequente di questi aggettivi, come avviene di solito con gli elementi stilnovistici riproposti da Cino. Inoltre in due casi utilizza l'accostamento tra *gentile* e *vile* in opposizione e in rima proprio come fanno i poeti del nuovo stile, come in “I no spero che mai per mia salute” nei vv. 41-46.

In Dante le forme *gentile* (9 occorrenze) e *gentil* (19 occorrenze) si riferiscono altrettanto a tutto ciò che è nobile di cuore e bello spiritualmente come in Guinizzelli. Questo concetto è soprattutto evidente in *Simili beni al cor gentile accosta*, in “Poscia ch'amor del tutto m'ha lasciato” nel v. 107, formula già citata nella *Vita Nuova*²⁷. Sempre in “Poscia ch'amor” ai vv. 9-12, si trova l'unica occorrenza della forma *vile* di tutte le rime. Quest'occorrenza è usata da Dante in contrapposizione al vocabolo leggiadria, che ha il significato di pregio:

*Ch'è nato in noi, di chiamare a ritroso
Tal che è vile e noioso
con nome di valore
cioè di leggiadria, ch'è bella tanto*

Invece in ambito siciliano sono poche le occorrenze delle forme gentile (1), gentil (5) e vile(1). Ad esempio, nel sonetto di Giacomo da Lentini, “Donna, eo languisco e non so qua speranza” nel v. 38, è presente la figura dell'ossimoro costituito dall'unica occorrenza di *gentile* e dal sostantivo *ira*. La funzione di questa figura retorica è quello di riappacificare i sentimenti contrasti nell'ambito

27. Cfr. Gianfranco Contini, *Dante Alighieri: Rime*, a cura di Gianfranco Contini, Torino, Einaudi Editore, 1946, p.104.

dell'amor cortese. Questo tipo di stilemi si arricchisce particolarmente in ambito siciliano perché è idonea a presentare il dramma dell'uomo come vittima del potere irrazionale dell'amore. Un altro hapax siciliano, la forma *vile*, si trova in “Misura provvidenza e meritanza” di Federico II per indicare che chi è vile non smette di esserlo nemmeno se possiede nobiltà o ricchezza²⁸.

5.4

Fina

La forma *fina* è presente nei subcorpora (v. Fig. 5.4) con due valenze, come verbo e come aggettivo.

forma	totale	siciliani	dante	stilnovo	bolognese	cino
fina	19	12	1	2	1	3

Fig. 5.4 Occorrenze della Forma Fina.

Nel corpus è usata soprattutto dalla scuola siciliana (12 occorrenze). Infatti in ambito federiciano era l'aggettivo appropriato per indirizzare i complimenti alla donna ed esaltarla²⁹. *Fina* è una forma tipicamente provenzale con il significato aggettivale di preziosa, bella, alta, nobile come in *Guiderdone aspetto avere*, v. 30 e 32, di Giacomo da Lentini o in *Gioisamente canto*, v. 40, di Guido delle Colonne. Sempre nella scuola siciliana, ha anche il significato verbale di finire come in *A pena pare ch'io saccia cantare*, v. 50, di Jacopo di Mostacci o in *Gioisamente canto*, v. 44, di Guido delle Colonne.

L'uso di questa forma in Cino da Pistoia (3 occorrenze), nei poeti bolognesi (1 occorrenza) e negli stilnovisti (2 occorrenze) è dovuta a una ripresa dei modelli dalla tradizione siciliana e provenzale. In Cino da Pistoia si trova con il significato verbale di finire nelle rime *Sì mi stringe l'amore* (v. 83), *L'anima mia, che si va peregrina* (v. 8) e *Cercando di trovar minera in oro* (v. 6). Invece negli stilnovisti è usata con valenza aggettivale nelle rime di Cavalcanti (*Fresca rosa novella*, v. 33) e di

28. G. Folena, *Cultura e poesia dei Siciliani*, in: SLI, diretta da E. Cecchi e N. Sapegno, vol. I, le Origini e il Duecento, Milano, 1965-69, pp. 287-288.

29. Carlo Salinari, *La poesia lirica del Duecento: corso di letteratura italiana tenuto nell'anno accademico 1967-68 presso il Magistero di Salerno*. Napoli, Liguori, 1992, p. 64.

Lapo Gianni (*Amore, i' non son degno ricordare*, v. 27). Per quanto riguarda i poeti bolognesi, l'unica occorrenza ha significato aggettivale e si trova in *Madonna, il fino amor ched eo vo porto*, v. 8, di Guido Guinizzelli.

Nella produzione dantesca l'unica occorrenza si trova nella rima di dubbia attribuzione *De gli occhi di quella gentil mia dama*, v. 6, con significato aggettivale.

5.5

Tempo

Tempo è un'altra forma usata nei 5 subcorpora in maniera non omogenea (v. Fig. 5.5) anche se per lo più è utilizzata in stilemi simili.

forma	totale	siciliani	dante	stilnovo	bolognese	cino
tempo	72	7	31	12	5	17

Fig. 5.5 Occorrenze della forma *tempo*.

In Dante c'è il maggior numero di occorrenze (31) ed è utilizzata soprattutto in varie metafore come in *Io son venuto al punto de la rota*, per indicare la primavera (v. 31, 38 e 67) o la giovane età (v. 39)³⁰. Anche nella rima *Amor, tu vedi ben che questa donna*, la forma *tempo* appare ben 14 volte, in metafore riferite alla vita lunga o felice (v. 33, 48, 51) e alla tempesta (v. 54) o per fare riferimento al giudizio universale (v. 58)³¹.

Come nelle rime dantesche, anche in Cino da Pistoia (17 occorrenze), *tempo* è usata spesso in metafore riguardanti la primavera o la vita. Nella rima *S'io ismagato sono ed infralito*, v. 52, la primavera (“il tempo gaio”) è in contrapposizione al “tempo freddo”. In questo modo Cino da Pistoia indica che è quello il tempo in cui la verità santa trionferà sull'errore e sul male del presente³².

Negli stilnovisti la forma *tempo* ha il significato di ora, spazio, come in *La forte e nova mia*

30. Cfr. Gianfranco Contini, *Dante Alighieri: Rime*, a cura di Gianfranco Contini, Torino, Einaudi Editore, 1946, p.185.

31. Ibid., 194.

32. Emilio Pasquini e Quaglio, Antonio Enzo. *Lo Stilnovo e la poesia religiosa*. Bari, Laterza, 1975, pp. 67-69.

disventura, v.23 di Cavalcanti e in *Tanta è l'angoscia ch' i' nel cor mi trovo*, v.6 di Frescobaldi. Anche Cavalcanti ricorre all'uso della forma *tempo* nello stilema metaforico per indicare la stagione primaverile nella rima *Fresca rosa novella*, v.15.

Nei bolognesi la forma *tempo* ha soltanto 5 occorrenze. Nella rima *Donna, l'amor mi sforza*, v. 16 di Guinizzelli è usata nella metafora “tempo torto” con il significato di maltempo. Invece in Onesto da Bologna, l'unica occorrenza di questa forma si trova in *One cosa terena quanto sale*, v. 14, con il significato di in poco d'ora; in questo modo riprende “un'ora” del v. 4 della stessa rima³³. In tutti gli altri poeti bolognesi, *tempo* ha il significato di epoca o momento.

Anche nei siciliani sono poche le occorrenze (7) e per lo più la forma *tempo* è usata in metafore come in *Guiderdone aspetto avere*, v.25 di Giacomo da Lentini, in cui “lo reo tempo” indica il cattivo tempo del quale il poeta se la ride sperando diventi un tempo migliore.

5.6 Amor/amore, cor/ core, donna, occhi

Questo gruppo di forme costituiscono le parole-chiavi proprie della lirica delle origini, da qui l'occorrenza elevata nel corpus (v. Fig. 5.6).

forma	totale	siciliani	dante	stilnovo	bolognese	cino
amor	469	93	74	104	52	146
amore	297	92	50	67	30	58
cor	339	47	42	94	30	126
core	265	79	35	61	23	67
donna	388	98	79	106	23	82
occhi	246	17	55	72	15	87

Fig. 5.6 Occorrenze delle forme *amor/amore, cor/core, donna e occhi*.

33. *Le rime di Onesto da Bologna*, a cura di Sandro Orlando, Firenze, Sansoni, 1974, p. 86.

Tutte queste forme riguardano il *topos* dell'amore cortese e dello sviluppo dell'innamoramento in cui l'immagine della donna passa attraverso gli occhi e da lì arriva fino al cuore infrangendolo. Inoltre la forma donna ha esclusivamente il significato originario di “signora” (del cuore).

Nei siciliani la fenomenologia amorosa viene spiegata attraverso il ricorso a questo *topos* e a queste forme. Gli occhi sono lo strumento che afferra l'immagine della donna e che permette che arrivi al cuore. Tuttavia nei siciliani anche la mente gioca un ruolo importante perché permette di immaginare il fantasma³⁴ della donna in sua assenza.

Nel caso degli stilnovisti questo è un motivo fondamentale e molto ricorrente che li separa nettamente dalla tradizione precedente siculo-toscana. In Cavalcanti soprattutto vi è presente ma nei suoi componimenti si trovano anche alcune varianti. Infatti il motivo amoroso cavalcantiano non è segnato solo dalle forme *cor*, *occhi* ma anche dalla forma *mente* come nei siciliani. La differenza però è che l'utilizzo cavalcantiano è in senso rigorosamente scientifico. Lo scopo di Cavalcanti è quello di oggettivare il sentimento amoroso con la rappresentazione della donna che attraversa gli occhi, arriva al cuore e desta la mente dal sonno. Per Cavalcanti *core* e *mente* sono termini strettamente connessi dato che segue la filosofia averroistica in cui si afferma che le radici delle virtù immaginative si trovano nel cuore e si svolgono nella mente³⁵.

Anche nel caso della scuola bolognese si ricorre a questo motivo soprattutto nel caso di Guinizzelli. Per quanto riguarda Onesto da Bologna sono presenti in alcune rime giochi etimologici che coinvolgono le forme amore e cor con *amaro* e *corpo*, probabilmente come ripresa dello stile guittoniano come in *Davante voi, madonna, son venuto*, v. 5³⁶.

In Dante il tema amoroso è centrale e lo rappresenta seguendo lo schema appartenente alla tradizione. Nella rima *Tre donne intorno al cor mi son venute*, vv. 1-3, viene sottolineata la presenza di amore nel cuore proprio come avviene presso gli stilnovisti. Inoltre Amore viene personificato secondo i canoni cortesi e stilnovistici. Un altro esempio è presente nel sonetto *Guido, i' vorrei che tu e Lapo ed io*, v. 12, nel quale l'amore diventa epicentro della seconda stanza.

E' però in Cino da Pistoia che si trovano tantissimi occorrenze di queste forme. Molti sonetti ciniani riportano al *topos* degli occhi come lo strumento che cattura amore e come messaggero del cuore. Nei suoi componimenti è evidente che Cino da Pistoia ripropone alcuni *topoi* stilnovistici ma in alcuni casi con l'aggiunta di motivi personali come l'uso ricorrente della forma *occhi*, come nella rima *La dolce vista e 'l bel guardo*. Nei componimenti ciniani è presente una concreta stilizzazione della tematica amorosa nella quale la donna tende a personalizzarsi dato che il tempo e lo spazio in cui si trova prendono caratteristiche più definite.

34. Nel medioevo l'immagine o ricordo della donna veniva definito in questo modo.

35. Emilio Pasquini e Quaglio, Antonio Enzo. *Lo Stilnovo e la poesia religiosa*. Bari, Laterza, 1975, pp. 57-64.

36. *Le rime di Onesto da Bologna*, a cura di Sandro Orlando, Firenze, Sansoni, 1974, p. 77.

5.7

Luce

La forma *luce* è un latinismo utilizzato nei 5 subcorpora (v. Fig. 5.7) come segue:

forma	totale	siciliani	dante	stilnovo	bolognese	cino
luce	41	1	21	14	1	4

Tabella 1: Fig. 5.7 Occorrenze della forma *luce*.

Sia nella scuola siciliana che in quella bolognese questa forma ricorre una volta sola. Nei siciliani è presente nella rima *Madonna à 'n sé vertute con valore*, v. 5 di Giacomo di Lentini con significato verbale di lucere. Invece nel caso dei poeti bolognese, si trova in *Dolente, lasso, già non m'asecuro*, v. 9 di Guinizzelli con valenza sostantivale di luce.

In Cino da Pistoia occorre soltanto 4 volte con la doppia valenza di verbo e di sostantivo. Nelle rime *Lo intelletto d'amor ch'io solo porto*, v. 13 e *Molte fiate Amor, quando mi desta*, v. 6 si riferisce al verbo lucere; mentre nella rima *Io son sì vago de la bella luce*, v. 1 appare come il sostantivo luce. L'ultima occorrenza si trova nella rima di dubbia attribuzione *Io maladico il dì ch'io vidi imprima*, v. 2, in cui si fa riferimento alla luce degli occhi.

Questa forma però acquisisce molta importanza in Dante e negli stilnovisti. Nelle rime dantesche l'uso della forma *luce* segue le idee della filosofia medievale della luce in cui questa appartiene al cielo e si riferisce alle cose divine e a Dio o nel senso di sede del sole e delle stelle. Per indicare invece l'illuminazione terrena, Dante fa affidamento alla forma *lume*. È probabile che quest'ideologia della luce proposta da filosofi come Roberto Grossatesta, Tommaso di Aquino o Alberto Magno si basasse su alcuni versi biblici in cui viene esposta quest'idea. È importante notare che Dante accosta alla forma *donna* le forme *luce* o *lume* a seconda se si riferisce a una figura femminile di provenienza divina o terrena. Inoltre nella rima *Amor, tu vedi ben che questa donna*, Dante usa ben 13 volte questa forma, in pochi casi con significato verbale e nella maggior parte come sostantivo. Sempre in questa canzone, al v. 46, ricorre alla forma *luce* nella figura retorica metonimica per indicare il giorno che si contrappone alla notte³⁷.

Negli stilnovisti, *luce* viene ripresa dai temi guinizzelliani da Cavalcanti come nella rima *Avete 'n vo'*

37. Gianpiero W. Doebler, *Non mi può far ombra: le distinzioni fra luce e lume nelle rime di Dante*. (<http://www.ucm.es/info/italiano/acd/tenzone/t7/Gianpiero-W-Doebler-Tenzone-7.pdf>).

li fior' e la verdura, v. 2, con il significato verbale di lucere, risplendere. Anche nei primi due versi di *Io voglio del ver la mia donna laudare* è evidente il calco guinizelliano in cui ritorna la luminosità del viso della donna³⁸. Un uso metaforico invece è presente nella rima *Donna me prega*, - *per ch'eo voglio dire*, v. 68, attraverso la quale si indica come amore in un ambiente buio escluda la luce³⁹. Nel caso di Frescobaldi e Gianni, la forma luce ha sia il significato di verbo che di sostantivo. Un esempio di uso sostantivale in cui la luce viene contrapposta alla parte oscura o al buio, lo si trova in *Voi che piangete nello stato amaro*, v.3, di Dino Frescobaldi.

5.8

Sicilianismi, provenzalismi e latinismi

Sul piano lessicale, i 5 subcorpora condividono alcune forme che provengono da altre lingue sia volgari e non, come il provenzale, il siciliano e il latino. L'uso di queste forme conferisce alla lingua della lirica del duecento-trecento caratteristiche che hanno permesso di definirla interregionale, anche se proprio questi elementi furono molto criticati da Dante e dagli stilnovisti.

Nelle forme comuni alle 5 scuole ci sono pochi sicilianismi, dovuto principalmente al poco uso che ne facevano gli stilnovisti e Dante. Inizialmente però molti sicilianismi non adattati dai copisti toscani erano stati ripresi da parte dei poeti siculo-toscani in seguito alla fine della scuola federiciana. Tra i pochi sicilianismi comuni nei 5 subcorpora ci sono *foco*, *loco*, *core*, *aggio*, *miso*, *veggio* e *meo* (v. Fig. 5.8).

forma	totale	siciliani	dante	stilnovo	bolognese	cino
foco	63	25	9	4	15	10
loco	68	14	7	12	10	25
core	265	79	35	61	23	67
aggio	61	36	6	1	9	9
miso	14	8	1	1	3	1
veggio	45	7	4	13	1	20
meo	158	80	6	14	24	34

Fig. 5.8 Occorrenze delle forme siciliane.

38. Emilio Pasquini e Quaglio, Antonio Enzo. *Lo Stilnovo e la poesia religiosa*. Bari, Laterza, 1975, pp. 62.

39. Cfr. *Inf.* VIII, 118.

Per esempio *foco* viene utilizzata dai siciliani citando il motivo della poesia trobadorica in cui la salamandra attraversa il fuoco e ne esce indenne come in *Madonna dir vo voglio*, v. 28, di Giacomo da Lentini. La forma siciliana *miso*, per indicare messo, è utilizzata insieme alla forma *foco* da Dante, nella rima *Tre donne intorno al cor mi son venute*, v. 83. Il meridionalismo *loco* è utilizzato non solo in senso sostantivale di luogo ma anche come avverbio con il significato di lì, come in *Donna, l'amor mi sforza*, vv. 29 e 31 rispettivamente.

Nemmeno le forme provenzali comuni non sono molto numerose nei 5 subcorpora (v. Fig. 5.9):

forma	totale	siciliani	dante	stilnovo	bolognese	cino
canoscenza	16	3	2	6	3	2
valenza	12	4	1	3	1	3
gioia	69	32	2	9	11	15
orgoglio	24	14	3	1	2	4
coraggio	13	4	3	1	3	2
speranza	42	17	5	12	2	6
chiaro	12	2	2	2	1	5

Fig. 5.9 Occorrenze dei provenzalismi.

Queste forme richiamano le loro equivalenti in provenzale, soprattutto quelle che hanno un suffisso in -anza e -enza come *valenza*, *canoscenza* e *speranza*. Un altro esempio significativo riguarda la forma *coraggio* che indica il cuore. Questa connotazione semantica è nei siciliani una citazione dello stile proprio del provenzale cortese come in *Dal core mi vene*, v. 122 di Giacomo da Lentini.

Alcune forme comuni alle 5 scuole derivano dal latino come nel caso di *fortuna*, *viso* e *luce* (v. Fig. 5.10):

forma	totale	siciliani	dante	stilnovo	bolognese	cino
fortuna	9	1	3	2	1	2
viso	82	38	7	16	6	15
luce	41	1	21	14	1	4

Fig. 5.10 Occorrenze dei latinismi.

Esempi dei latinismi *fortuna* e *viso* sono presenti rispettivamente in *guido i' vorrei che tu e lapo ed io*, v. 5 e in *Tre donne* di Dante e in *Tre donne intorno al cor mi son venute*, v. 90 di Dante.

5.9

Par e pare

Queste due forme hanno valore verbale di parere e presentano molte occorrenze nei vari subcorpora (v. Fig. 5.11):

forma	totale	siciliani	dante	stilnovo	bolognese	cino
par	103	<u>24</u>	<u>23</u>	<u>29</u>	15	12
pare	52	<u>26</u>	3	8	7	8

Tabella 2: Fig. 5.11 Occorrenze delle forme *par* e *pare*.

Le forme *par* e *pare* nel medioevo avevano un significato molto diverso da quello attuale. Essenzialmente secondo Contini non avevano il significato di “sembra” e nemmeno quello di “appare” ma quello di “appare evidente, è o si manifesta nella sua evidenza”⁴⁰.

Nei componimenti dei poeti delle 5 scuole queste forme hanno il significato medievale e in molte rime vengono utilizzate in alcuni elementi stilistici come le ripetizioni o in ampliamenti delle sfere semantiche.

Nei siciliani le forme *par* (24 occorrenze) e *pare* (26 occorrenze) sono utilizzate in ripetizioni del verbo parere con variazioni semantiche come in *Meravigliosamente* di Giacomo da Lentini.

Anche Guinizzelli usa queste forme in ripetizioni proprio come i siciliani. Un esempio si trova in *Madonna, il fino amor ched eo vo porto*, vv. 3, 63, 70 e 77.

Invece in cino da Pistoia le forme *par* e *pare* sono collegate al campo semantico del verbo vedere dato che significano “si vede, appare” come nella rima *Sta nel piacer de la mia donna Amore*, v. 9 o nella rima *In disnor e 'n vergogna solamente*, v. 9.

40. Gianfranco Contini, “Esercizio di interpretazione sopra un sonetto di Dante”, in *Varianti e altra linguistica. Una raccolta di saggi (1938-1968)*, Torino, Einaudi, 1970, pp. 161-168

5.10

Servire

La forma *servire* presenta delle discrepanze dal punto di vista delle occorrenze tra i 5 subcorpora (v. Fig. 5.12):

forma	totale	siciliani	dante	stilnovo	bolognese	cino
<i>servire</i>	27	14	3	3	2	5

5.12 Occorrenze della forma *servire*.

Nei siciliani la forma *servire* è un riflesso del rapporto feudale di vassallaggio tra l'amante e la donna, il primo inteso come vassallo e la seconda come signore. Il tema della servitù d'amore è ricorrente nella poesia siciliana ed è collegato ad altre forme dal punto di vista semantico e dello schema della tradizione aulica, come penare, gioia, dolore, lontananza o talento.

Nella scuola bolognese, specificamente in Guinizzelli, con l'uso di questa forma c'è una ripresa dei contenuti tipici dei siciliani e del rapporto cortese feudale.

Per quanto riguarda Dante, Cino da Pistoia e gli stilnovisti, la forma *servire* è sinonimo di ubbidire. Infatti in tutta la tradizione cortese e di ispirazione stilnovistica, l'amore si presenta come un servizio in cui l'amante consegna se stesso alla donna, come prova della sua gentilezza e cortesia.

Conclusioni

Lo scopo della ricerca effettuata sui 5 subcorpora utilizzati era quello di ottenere il lessico comune e analizzare le forme presenti più significative. Grazie alla metodologia proposta è stato possibile raggiungere il primo traguardo e successivamente scegliere per l'analisi le forme di maggiore rilevanza dal punto di vista del contenuto.

Il lessico comune è composto da 442 forme, alcune delle quali una volta analizzate hanno fatto emergere diversi aspetti riguardanti il linguaggio tipico della poesia cortese così come vari aspetti extratestuali.

Forme come *Amor, donna, cor, occhi*, portano alla dimensione dell'amore cortese tipico della poesia tra il duecento e il trecento. Ovviamente ogni scuola poetica la sviluppa in modo personale a seconda dell'ambiente che lo circonda. Così nei componimenti siciliani l'amore è sinonimo del rapporto tipicamente feudale di vassallaggio mentre negli stilnovisti la tematica viene trattata su un piano più elevato per rappresentare la bellezza e la purezza della donna amata. Sempre le forme *cor* e *occhi* rappresentano i *topoi* della lirica delle origini, in cui gli occhi sono il mezzo attraverso cui viaggia l'amore per arrivare al cuore, il quale a sua volta è la sede dell'amore.

Invece altre forme permettono di cogliere riferimenti alle idee filosofiche del momento come nel caso di *mente, luce* o *spirito*. Idee filosofiche che, ad esempio, riguardano la tradizione aristotelica ripresa da Averroè oppure la filosofia medievale della luce basata su scritti biblici o su scritti da Tommaso d'Aquino, Roberto Grossatesta fra altri scrittori antichi.

Anche i cambiamenti radicali che stava subendo la società dell'epoca vengono riflessi nell'uso della forma *gentil/gentile*. Già in contesto feudale, addirittura prima della scuola siciliana, si erano avvertiti i primi sintomi di questa situazione attraverso il trattato *De Amore* di Andrea Cappellano. L'apice però si raggiunge nella sfera comunale, presso gli stilnovisti, nella quale il significato di gentilezza cambia completamente valore. Per rendere noto l'affermarsi della borghesia come nuova élite sociale, si passa da gentilezza intesa come nobiltà aristocratica ed ereditata attraverso il sangue a nobiltà d'animo basata sull'intelligenza e la spiritualità.

Uno degli aspetti di quest'analisi che spicca su tutto però è la precisa idea che ogni poeta aveva della poesia stessa e del pensiero che voleva trasmettere. Infatti in molti casi la presenza di alcune forme comuni nei componimenti sono motivate da riprese di contenuto. Con questa tecnica stilistica un determinato poeta voleva far sentire la sua concezione di poesia, criticare il modo di esprimersi di un altro o ironizzare su certi argomenti. Nei componimenti presi in esame sono presenti diversi esempi

come nel sonetto di Onesto da Bologna a Cino da Pistoia (“Mente e umile e più di mille sporte”) o in quello di Guinizzelli indirizzato a Guittone d'Arezzo (“[O] caro padre meo, de vostra laude”).

Allo stesso tempo, in altri casi le forme in comune sono usate semplicemente come citazioni di stile nei confronti di altri poeti o della tradizione a cui si ispirano o da cui traggono alcune tematiche. Un esempio è dato dalla scuola siciliana nell'uso della forma *coraggio* con connotazione semantica di cuori come nella poesia provenzale o come avviene in Cino da Pistoia e negli stilnovisti con la forma *fin* ripresa dalla tradizione siciliana e provenzale.

D'altro canto alcune forme possono essere intese come eredità linguistica. Forme come *gioia*, *speranza*, *talento*, *coraggio* derivate dalla poesia provenzale, inizialmente sono usate dalla scuola siciliana ma vengono mantenute in tutti gli altri gruppi poetici. Lo stesso concetto può essere applicato alle forme di origine siciliana come i sostantivi *loco*, *core*, *foco* e le forme verbali *veggio*, *aggio*, *vorria*, *saccio* o ad alcuni latinismi come *luce* o *viso*.

Evidentemente in questa ricerca non sono stati analizzati tutti i percorsi possibili perché questo argomento è molto vasto. Quindi dal punto di vista linguistico sarebbe possibile approfondire ulteriormente alcuni temi o svilupparne nuovi. Un'altra possibilità sarebbe quella di svolgere l'indagine statistica su unità linguistiche più complesse delle forme grafiche, come le categorie grammaticali, costruzioni sintagmatiche, ecc. Invece dal punto di vista informatico, si potrebbe ingrandire il database inserendo le tabelle autori e componimenti. In questo modo le forme comuni potrebbero essere collegate ai singoli autori e ai contesti a cui appartengono. L'obiettivo sarebbe quello di utilizzare con fini didattici i risultati ottenuti in un sito web o in un'applicazione web.

Bibliografia

Balbo, Cesare. *Sommario della Storia d'Italia: dalle origini fino ai nostri tempi*. Torino, Unione tipografico-editrice, 1860.

Da Guido Guinizzelli a Dante: nuove prospettive sulla lirica del duecento. A cura di Furio Brugnolo e Gianfelice Peron. Monselice, Il Poligrafo, 2004.

Dante Alighieri: Rime. A cura di Gianfranco Contini. Torino, Einaudi Editore, 1946.

Dalle Origini al Cinquecento. A cura di Loredana Chines, Giorgio Forni, Giuseppe Ledda ed Elisabetta Menetti. Milano, Bruno Mondadori, 2007.

Di Benedetto, Luigi. *Da Giacomo da Lentino a Francesco Petrarca*. Napoli, Astrea, 1949.

Dizionario Italiano Ragionato, G. D'Anna Sintesi, Firenze, 1988.

Cavalcanti, Guido. *Rime con le rime di Iacopo Cavalcanti*, a cura di Domenico di Robertis, Torino, Einaudi. 1986.

Emmer, Michele. *Matematica e Cultura 2007*. Milano, Springer, 2007.

Fiorino, Antonia. *Metri e temi della scuola siciliana*. Napoli, Liguori, 1969.

Folena, G. *Cultura e poesia dei Siciliani*, in: *SLI*. Diretta da E. Cecchi e N. Sapegno, vol. I, le Origini e il Duecento. Milano, 1965-69.

Gigliozzi, Giuseppe. *Introduzione all'uso del computer negli studi letterari*. A cura di Fabio Ciotti, Milano, Bruno Mondadori, 2003.

I rimatori bolognesi del secolo XIII. A cura di Guido Zaccagnini. Milano, Vita e Pensiero, 1933.

I rimatori bolognesi del secolo XIII. Edizione critica a cura di G. Taccagni. Milano, Società Editrice Vita e Pensiero, 1933.

Le rime di Cino da Pistoia. A cura di Guido Zaccagnini. Geneve, Olschki, 1925.

Lenci, Alessandro, Montemagni, Simonetta, Pirrelli, Vito. *Testo e computer: elementi di linguistica computazionale*. Roma, Carrocci, 2010.

Libertini, Vincenzo. *Cino da Pistoia*. Lanciano, Itinerari, 1974.

Monteverdi, A. Per una canzone di Re Enzo in “*Studi e saggi sulla letteratura italiana dei primi secoli*”. Milano-Napoli, Ricciardi, 1954.

Orlando, Sandro (a cura di). *Le rime di Onesto da Bologna*. Firenze, Sansoni, 1974.

Panvini, Bruno (a cura di). *Le rime della scuola siciliana*. Firenze, Olschki, 1962.

Pasquini, Emilio e Quaglio, Antonio Enzo. *Le origini e la scuola siciliana*, seconda edizione. Bari, Laterza, 1975.

Pasquini, Emilio e Quaglio, Antonio Enzo. *Lo Stilnovo e la poesia religiosa*. Bari, Laterza, 1975.

Poeti del Duecento, a cura di G. Contini, Milano-Napoli 1960.

Rimatori del Dolce Stil Novo: Guido Guinizzelli. A cura di Luigi di Benedetto. Bari, Laterza, 1939.

Russo, Luigi. *Studi sul Due e Trecento*. Roma, Edizioni italiane, 1946.

Salinari, Carlo. *La poesia lirica del Duecento: corso di letteratura italiana tenuto nell'anno accademico 1967-68 presso il Magistero di Salerno*. Napoli, Liguori, 1992.

Scardicchio, Andrea. *Letteratura e informatica: Problemi ed esperienze*. Melpignano (LE), Amaltea Edizioni, 2006.

Tomasi, Francesca. *Metodologie informatiche e discipline umanistiche*, Roma, Carocci, 2008.

Varvaro, Alessandro. *Linguistica Romanza: corso introduttivo*, seconda edizione. Napoli, Liguori, 2001.

Ventivogli, Bruno e Paola Vecchi Galli, *Filologia italiana*. Milano, Bruno Mondadori, 2002.

Welling, Luke e Laura Thomson. *MySQL Tutorial*. Milano, Pearson Education Italia S.r.l, 2004.

Siti web

Corso di Laurea in Informatica Umanistica, Università di Pisa, www.unipi.it

Corpus OVI dell'italiano antico, <http://gattoweb.ovi.cnr>

MySQL, www.mysql.com

PHP, <http://php.net>

Corpus di DanteSearch, <http://dante.di.unipi.it:8080/DanteWeb/>

TLIO (Tesoro della lingua italiana delle origini), <http://tlio.ovi.cnr.it/TLIO/>

Dizionario etimologico, <http://www.etimo.it/>

IntraText, <http://www.intratext.com/ita/>

Google Books, <http://books.google.it/>

Biblioteca Italiana, <http://bibliotecaitaliana.it/>

Biblioteca dei classici italiani, <http://www.classicitaliani.it/>

Scuola Online, <http://www.pubblicascuola.it/indice.php>

Sistema bibliotecario dell'Università di Pisa, <http://biblio.unipi.it/>

Progetto Duecento, <http://www.silab.it/frox/200/>

Dante Online, http://www.danteonline.it/italiano/home_ita.asp