

Un *corpus* online della letteratura secondaria (1872- 1890) del Verismo italiano

Denise Bruno¹, Giuseppe Canzoneri², Antonio Di Silvestro³,
Daria Spampinato⁴, Alessandro Zammataro⁵

¹Università di Catania, Italia – denise.bruno@phd.unict.it

²Università di Catania, Italia – giuseppe.canzoneri@unict.it

³Università di Catania, Italia – antonio.disilvestro@unict.it

⁴CNR Istituto di Scienze e Tecnologie della Cognizione, Catania, Italia – daria.spampinato@cnr.it

⁵Università di Catania, Italia – alessandro.zammataro@unict.it

ABSTRACT

Il contributo illustra le prime fasi del progetto COVerLeSS (*Corpus* Online del Verismo tra Letteratura, Storia e Società) che mira a conservare, valorizzare e analizzare, in un ambiente web integrato e open access, un archivio interrogabile della letteratura secondaria (recensioni, testi giornalistici e saggistici) relativa alla produzione letteraria del Verismo italiano. Il progetto non si limita però alla costruzione di un mero archivio statico, mirando semmai all'utilizzo dei dati acquisiti e indicizzati al fine di offrire una prospettiva di lettura diacronica dell'evoluzione del lessico del Verismo. Il vocabolario *Verbum*, che consente una ricerca delle forme notevoli del lessico "standard" della corrente verista, e la timeline *Ver-in-time*, che intende collegare la letteratura secondaria ospitata sul portale ai testi primari del Verismo disponibili in diverse biblioteche digitali, costituiscono gli strumenti principali attraverso cui condurre un'analisi semantica sul piano sincronico e diacronico. Nell'ottica della fruibilità e dell'inclusività, esso si colloca nella prospettiva di una fruizione del bene culturale che riflette sia il piacere dell'accostamento al documento (riprodotto nella sua veste originaria), sia la curiosità di "navigare" all'interno del documento rappresentato in formati testuali accessibili all'utente.

PAROLE CHIAVE

Secondary literature; Verism; Lexicography; Digital philology; Digital scholarly publishing.

1. INTRODUZIONE

Il progetto fornisce un'immagine "diversa" e innovativa del Verismo, grazie al suo riverbero nei testi giornalistici e saggistici del tempo e al modo in cui essi contribuirono a costruire un'immagine dell'Italia meridionale post-unitaria a livello sociale, culturale ed economico. Sul piano della conservazione, la necessità della creazione di questo *corpus* nasce dalla forte dispersione e frammentazione bibliografica delle fonti di letteratura secondaria, quasi mai fruibili a testo pieno. D'altronde, le caratteristiche dell'archivio fanno di esso un esperimento pilota per la costruzione di raccolte della letteratura secondaria, tuttora assenti dai repertori digitali per lo studio della letteratura italiana [1, 7, 10].

Per il conseguimento degli obiettivi si stanno seguendo tre macro-fasi, ciascuna dotata di una propria metodologia: 1) conservazione; 2) valorizzazione e fruizione; 3) analisi. La prima, svolta in accordo con le Linee Guida del Piano Nazionale di Digitalizzazione, si articola in ulteriori tre fasi:

Acquisizione e metadattazione. L'acquisizione e lo stoccaggio delle immagini avvengono mediante l'applicazione di tecnologie IIIF per la gestione e la visualizzazione. La metadattazione delle fonti digitalizzate consiste nella descrizione dei supporti e dei formati secondo gli schemi standard maggiormente riconosciuti (Dublin Core, MAG, XML/TEI) e integrati in formati METS per l'interscambio.

Trascrizione e codifica. Si sta procedendo alla trascrizione del *corpus* mediante metodologie e tecnologie di OCR (come Tesseract ed eScriptorium), con la metadattazione automatica del processo secondo lo standard XML-ALTO (*Analyzed Layout and Text Object*) [8]. Il testo così acquisito viene annotato da un punto di vista strutturale, filologico, linguistico, lessicale e semantico attraverso lo schema di codifica XML/TEI.

Gestione e organizzazione. L'organizzazione dei dati, per quanto riguarda le immagini, si basa sul protocollo IIIF, mentre per i metadati, trascrizione e annotazione, si appoggia su un'architettura di Database XML (eXistDB), con lo sviluppo di un'ontologia formale (RDF/XML, OWL) che estende la codifica XML/TEI, per quanto riguarda l'onomastica, la toponomastica e i titoli di opere e riviste, anche in modalità *Linked Open Data* (LOD) [6].

Questa fase consentirà di approntare, per la prima volta, la trascrizione dei testi delle riviste rendendola fruibile agli utenti mediante l'interfaccia web.

La seconda fase prevede lo sviluppo del sistema di gestione, indicizzazione e il *Search Engine* progettati e implementati secondo le buone pratiche di ingegneria del software e dei modelli formali per la gestione di documenti testuali [5]. In un'ottica di *long term preservation*, in linea con i principi FAIR¹ [23] e TRUST² [14], si depositeranno i dati in infrastrutture di ricerca quali CLARIN e/o DARIAH.

L'ultima fase prevede lo sviluppo di strumenti per un'analisi dei dati attraverso alcuni algoritmi di *Natural Language Processing* e *Text Mining*.

2. IL CORPUS RIVER: MOTIVAZIONI STORICO-CRITICHE E ORGANIZZAZIONE

Il ventennio circa di cultura letteraria italiana che, ai fini della presente ricerca, si è scelto di comprendere tra i due estremi cronologici del 1872 e del 1890, si caratterizza per un dibattito straordinariamente partecipato da parte di letterati, scrittori e giornalisti, che si confrontano e riflettono [12], forse per la prima volta su scala nazionale, intorno ad aspetti basilari di tipo politico-economico e antropologico di una società ancora disomogenea e in via di formazione. Il tema centrale, non solo letterario, è la fedeltà al principio del “vero” in arte, alla rappresentazione della realtà nella sua forma più cruda e “naturale”, che si esprime soprattutto nel “coraggio” dell'artista di scandagliare i temi scabrosi della sofferenza umana nei suoi aspetti più degradanti, quelli della povertà materiale, psicologica e morale.

Nel 1871, in prossimità dell'inizio della *timeline* contemplata dal progetto, iniziava la storia del “romanzo naturalista” con il ciclo dei Rougon-Macquart di Émile Zola [16]: un'analisi della società francese ritratta nei suoi ambienti reali – dai mercati ai sobborghi operai, dai cenacoli degli artisti alle miniere e alle campagne – e nelle sue dinamiche quotidiane. Subito dopo in Italia Luigi Capuana pubblicava la raccolta di saggi intitolata *Il teatro italiano contemporaneo* (1872), aprendo la stagione del Verismo [18]. Non a caso nel 1879 sarà ancora Capuana, con *Giacinta* (dedicato a Zola), ad inaugurare la stagione delle grandi opere veriste, anticipando, insieme a Navarro della Miraglia, che nello stesso anno pubblicava *La nana*, il Verga maggiore di *Vita dei campi* (1880), de *I Malavoglia* (1881), delle *Novelle rusticane* (1883), per arrivare fino a *Mastro Don Gesualdo* [1], che nel 1889 chiude la stagione del Verismo più produttivo e vitale.

Le opere letterarie dei veristi “maggiori” (Verga, Capuana e De Roberto), ma anche quelle che negli stessi anni andavano pubblicando i cosiddetti “minori” (Onufrio, Navarro della Miraglia, Ragusa Moleti, Chelli, Scarfoglio, Zena, Pratesi), coprono un arco cronologico lungo il quale si snoda una produzione parallela di saggi, recensioni e contributi critici, che a volte entra direttamente in dialogo con le opere letterarie (come nel caso delle recensioni), mentre in altre converge insieme a queste in un'analisi della realtà postunitaria italiana focalizzata soprattutto sulle pesanti tare economiche e culturali del Mezzogiorno, che creavano un enorme divario tra la qualità della vita delle masse contadine meridionali e quella degli abitanti del Nord Italia. Sulla base di questa riflessione preliminare, una prima parte del *corpus* di testi analizzati nell'ambito del progetto è, quindi, costituita dalla letteratura secondaria coeva alle opere del Verismo, che annovera tra i testi di maggior rilievo gli interventi critici di Francesco De Sanctis (*Il principio del realismo* del 1876 e *Studio sopra Emilio Zola* del 1879), di Felice Camerini e di Federico De Roberto, solo per citarne alcuni. Si tratta spesso di scrittori che parteciparono in veste di critici e di cronisti al grande dibattito nazionale, in un continuo scambio di ruoli e di prospettive.

Questa produzione comprende tre tipologie di pubblicazioni: 1. i saggi sul Verismo in volume; 2. gli articoli e le recensioni pubblicate sui periodici dell'epoca e raccolti in volume dagli stessi autori; 3. le recensioni e gli articoli rimasti dispersi nelle pagine delle riviste e dei quotidiani dell'epoca [11, 17]. Le prime due categorie hanno una consistenza complessiva di circa 500 pagine e quasi equivalente è la consistenza della terza.

Un secondo blocco di testi, notevolmente più consistente di quello appena citato, è costituito da un gruppo di riviste dell'epoca, selezionate per l'impegno nella diffusione dei principi del naturalismo francese e soprattutto nella teorizzazione di una versione italiana di tale corrente letteraria: «Il Momento Letterario-Artistico-Sociale» composto da circa 880 pagine [21] e pubblicato a Palermo dal 1883 al 1885 in 72 numeri; «La Fronda», 90 pagine ca. [20], che pur pubblicata a Firenze per soli 7 numeri, tra gennaio e febbraio del 1880, ospitò articoli, recensioni e opere dei maggiori letterati italiani, tra i quali Capuana e Verga; «La Farfalla», che, dopo una prima parentesi cagliaritano (1876-1877), venne trasferita dal fondatore Angelo Sommaruga, a Milano, dove venne pubblicata fino al 1883 per 43 numeri totali (520 pagine ca.) [2]. Quest'ultima rivista si propose come una delle voci più attive, e spesso polemiche, della Scapigliatura milanese, centro di ritrovo e di confronto tra i suoi esponenti e i veristi meridionali. L'ultima delle riviste facenti parte del *corpus* è la «Rassegna Settimanale di politica, scienze, lettere ed arti», fondata da Leopoldo Franchetti e da Sidney Sonnino e pubblicata per 213 numeri dal 1878 al 1882 (4260 pagine ca.)³, rivista che riuniva in un progetto finalmente unitario il

¹ Findability, Accessibility, Interoperability, Reusability.

² Transparency, Responsibility, User focus, Sustainability, Technology.

³ <https://rassegnasettimanale.animi.it/>

dibattito sui temi politici, economici, sociali e letterari, dedicando particolare attenzione alle problematiche delle popolazioni del Mezzogiorno d'Italia [15].

3. VERBUM

Dopo la fase di acquisizione dei testi secondari del Verismo – mediante metodologie e tecnologie di OCR – si sta conducendo l'annotazione dei testi tramite lo schema di codifica XML/TEI, da un punto di vista strutturale, filologico, linguistico, lessicale e semantico. In particolare, nel *teiHeader* vengono indicati tutti i metadati, dai responsabili della digitalizzazione, della codifica e delle revisioni alla descrizione delle fonti bibliografiche; nel corpo del documento, invece, attraverso i tag <div>, <p>, <head> vengono codificati gli elementi strutturali, e con i tag <persName>, <placeName>, <orgName>, <title>, <seg>, <term>, <distinct>, <interp> vengono segnalati i nomi dei personaggi, dei luoghi e delle organizzazioni, i titoli delle opere citate e le annotazioni semantiche. Verrà poi implementata, con le tecnologie del Semantic Web, un'ontologia formale delle entità nominate e delle informazioni bibliografiche contenute nei file XML/TEI e che colleghi tutto il *corpus* con i repository online in modalità LOD. I lemmi, sintagmi e le strutture fraseologiche, annotati opportunamente con i tag lessicografici, formano il vocabolario *Verbum*, che permetterà di indagare i fenomeni linguistici, lessicali e semantici più rilevanti del *corpus*, con l'obiettivo di rappresentare un lemmario di base, ricercabile per forme notevoli e per varianti inerenti a un vocabolario "ideale" della corrente verista. Dunque, tale vocabolario si fonda sulla ricorsività di lemmi, di combinazioni sintagmatiche (ad es. lemma base + un attributo), o di eventuali espressioni fraseologiche, che verranno appositamente marcate per valorizzare le particolari accezioni semantiche del lemma base. Per la marcatura dei lessemi si è scelto di servirsi del tag <term>, da specificare con l'attributo @type per individuare le varie tipologie di lessico o le combinazioni sintagmatiche e fraseologiche funzionali a *Verbum*. Un esempio della marcatura lessicografica a cui sono sottoposti i testi si può desumere dalla recensione di Filippo Filippi sulla rivista «La Perseveranza» del 2 ottobre 1880, poco dopo l'uscita della raccolta *Vita dei campi*. Filippi attinge al campo semantico della vista e delle arti visive, utilizzando espressioni come *colore locale*, *osservazione continua*, *varietà dei contorni e dei colori*, *scolpiti in bronzo*. Quest'ultima espressione sarà più volte ripresa dai critici, tra tutti Capuana, il quale sottolineava (nella recensione a *Vita dei campi* apparsa sul «Corriere della Sera» del 20-21 settembre 1880) come il «bronzo della lingua letteraria» fosse calato entro la «forma sempre fresca» del dialetto. Frequenti sono, inoltre, forme verbali quali *ritratto* e *vedere*, quest'ultimo nell'accezione del visualizzare nell'immaginazione del lettore i luoghi descritti. Un'altra recensione di particolare interesse lessicale è quella di Francesco Torraca sulla rivista «Il Diritto» del 9 maggio 1881, in cui spiccano lessemi ed espressioni che appaiono come varianti di un lessico di base paradigmatico, introdotto da Verga soprattutto nei suoi pochi testi "programmatici" o nella corrispondenza privata. Esempi ne sono, nella recensione di Capuana a *I Malavoglia* edita sul «Fanfulla della domenica» del 29 maggio 1881, espressioni come *dipinte con colori caldissimi*, *pennellate*, *minute particolarità*, *gran quadro*, *schizzò quei stupendi bozzetti*, *crudesse di toni*, *di mezze tinte*, *di sfumature* – con ripresa delle *mezze tinte dei mezzi sentimenti* che Verga aveva utilizzato nella Prefazione a *I Malavoglia* (vd. Fig. 1)⁴.

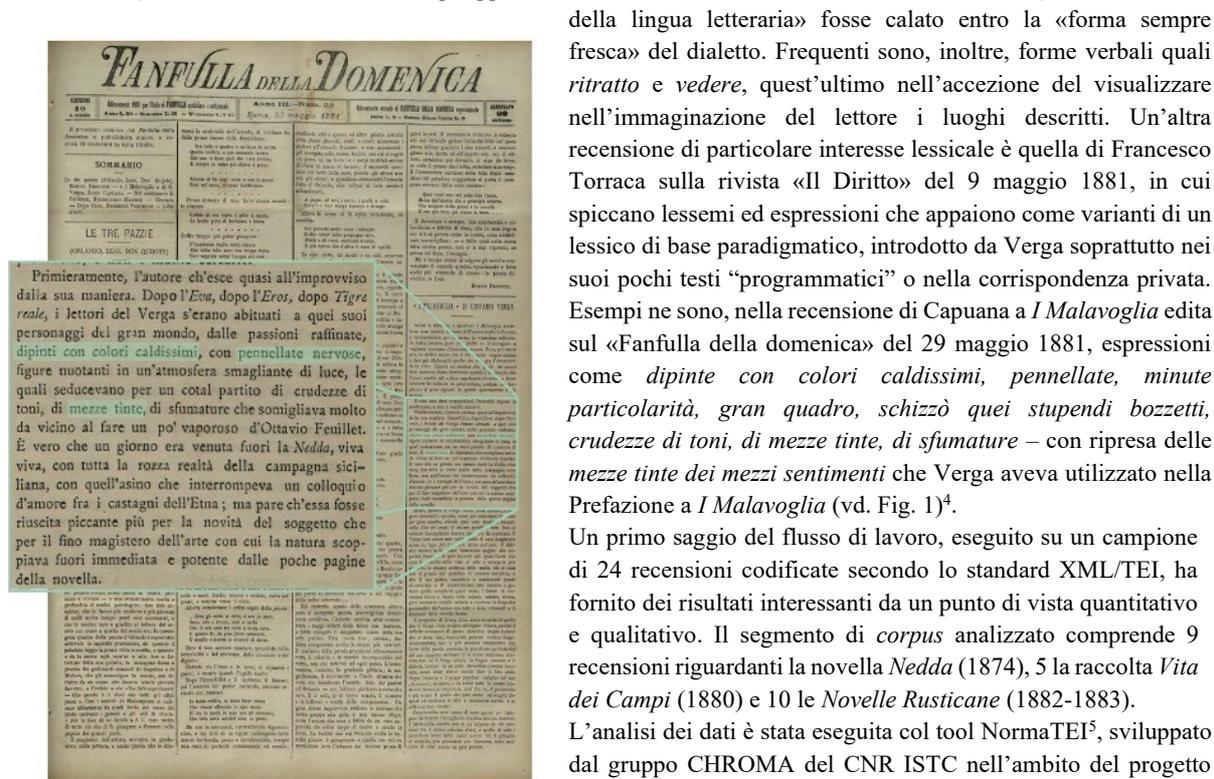


Figura 1. Recensione di Capuana a *I Malavoglia*, «Fanfulla della domenica», 29 maggio 1881

Un primo saggio del flusso di lavoro, eseguito su un campione di 24 recensioni codificate secondo lo standard XML/TEI, ha fornito dei risultati interessanti da un punto di vista quantitativo e qualitativo. Il segmento di *corpus* analizzato comprende 9 recensioni riguardanti la novella *Nedda* (1874), 5 la raccolta *Vita dei Campi* (1880) e 10 le *Novelle Rusticane* (1882-1883). L'analisi dei dati è stata eseguita col tool NormaTEI⁵, sviluppato dal gruppo CHROMA del CNR ISTC nell'ambito del progetto BellinInRete [4]. Questo tool permette di analizzare in modo complessivo tutti i file XML/TEI di un'edizione, e in esso è

⁴ Immagine con licenza CC0 1.0 Public Domain da <http://www.bncrm.beniculturali.it/it/32/biblioteca-digitale>

⁵ <https://github.com/pierpaolosichera/NormaTEI>

possibile effettuare sia ricerche rapide (per individuare eventuali refusi o effettuare normalizzazioni delle codifiche), che complesse, attraverso delle interfacce intuitive. Completano il tool un ambiente statistico e sistemi di esportazione dati (txt, xls). In questo caso NormaTEI è stato utilizzato per esemplificare un campione descrittivo di *Verbum*.

Il tag scelto per rappresentare i termini all'interno di *Verbum* è <term>, il quale contiene una designazione composta da una sola parola, più parole o simbolica, con la quale viene identificato il termine tecnico, che nello specifico afferisce a un'area semantica fortemente caratterizzata del linguaggio verista.

L'utilizzo del tool ha consentito di isolare – all'interno delle 24 recensioni che costituiscono il mini-corpus preso in esame – 671 tag <term>, di cui ben 170 (circa il 25%) appartengono all'area semantica delle arti visive, e in generale all'atto della narrazione come rappresentazione. Nello specifico, all'interno di questo campo semantico sono state isolate una serie di “parole dominanti” (circa il 40%), le quali presentano un numero di occorrenze piuttosto elevato: esempi ne sono *bozzetto*, *colorito*, *descrizione*, *fotografare*, *quadro* – accompagnato dal ricorrente diminutivo *quadretto* – *ritrarre*, ecc. Questi “lemmi dominanti” presentano spesso delle specificazioni (ossia espansioni consistenti in aggettivi, complementi di specificazione, di modo, ecc.), che ricorrono a volte come costanti all'interno del corpus e altre volte come “varianti” all'interno di specifici sintagmi; è il caso di espressioni come:

Bozzetto	bozzetto dal vero
	bozzetto siciliano
Fedele	quadro fedele
	fedele pittura
Colorito	potenza di disegno e di colorito
	colorito del paesaggio
	colorito locale
	intonazione del colorito
Ritrarre	vivacità di colorito
	ritrarre la vita
	ritrarre i dialoghi
	ritrarre sentimenti
Quadro/quadretto	quadro di costumi
	quadri veri
	quadretti locali
	quadretti della vita siciliana
	quadrettino di genere campestre
Dipingere	quadri variati
	dipinge dal vero
Colore	delineare a colori vivi i contorni
	tratteggiare coi colori più smaglianti della sua tavolozza
	colore del vero
	colori vivi
	vivezza di colore
	ricchezza del colore

Per dare un'idea della ricorsività di questa area semantica, limitata alle “parole dominanti”, abbiamo sottoposto la nostra selezione di 24 recensioni a un'analisi con *Voyant Tools*⁶.

Il grafico che ne risulta (vd. Fig. 2) mette in evidenza le parole “pilota” già individuate con NormaTEI. Inoltre, i “picchi” delle frequenze relative (certamente ascrivibili al numero ridotto dei testi qui utilizzati), forniscono una prima, ancor grezza ma promettente, rappresentazione dell'area semantica scelta, che insieme alle altre man mano ricostruite con lo spoglio progressivo dei testi, offrirà una ricca cartografia lessicale e semantica di un eccezionale fenomeno del nostro post-risorgimento letterario quale è stato il Verismo.

⁶ <https://voyant-tools.org/>



Figura 2. Andamenti delle parole “pilota” ricavati con Voyant Tools

La Figura 3, per esempio, restituisce graficamente la situazione delle occorrenze del lessico afferente all’area semantica delle arti visive, prendendo in considerazione il numero di occorrenze per ciascun lemma tra quelli dell’area semantica selezionata.

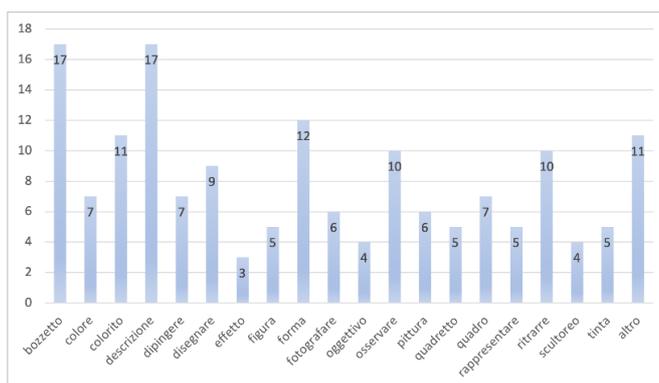


Figura 3. Lessico delle arti visive

Nella Figura 4, invece, le medesime occorrenze lessicali vengono raggruppate per affinità di significato, secondo il criterio della “parola dominante”. Per esempio, le parole ‘colore’ (7 occorrenze) e ‘colorito’ (11 occorrenze) sono raccolte sotto il lemma ‘colorito’. In questo grafico l’asse delle ascisse rappresenta le recensioni vagliate e facenti riferimento alle seguenti opere di Verga: *Nedda* (sigla N, recensioni pubblicate da giugno ad agosto 1874), *Vita dei campi* (VC, settembre 1880), *Novelle rusticane* (NR, dicembre 1882-marzo 1883); sull’asse delle ordinate, invece, è riportato il numero assoluto di occorrenze per singola recensione. In questo modo risulta evidente, dall’andamento delle linee di colore associate a ciascuna “parola dominante”, l’oscillazione nell’utilizzo da parte dei recensori di Verga del lessico delle arti visive, in una progressione cronologica che va dalla prima recensione a *Nedda* del 15 giugno 1874, all’ultima recensione alle *Novelle rusticane* del 30 marzo 1883.

Oltre alle strutture sintagmatiche e fraseologiche presenti e annotate nei testi, *Verbum* sarà arricchito di sinonimi, macro e microcategorie e parole chiave, ricavati sulla base di cicli di revisione del vocabolario stesso, al fine di essere utilizzato per una indicizzazione semantica dei documenti del *corpus*, rendendoli così ricercabili nell’interfaccia di ricerca con *keywords* avanzate e accessibili nelle maschere di visualizzazione. Per fare ciò, saranno testati strumenti basati su metodologie di *Machine Learning* per la classificazione di documenti di fine Ottocento anche senza parole chiave. Alcuni documenti saranno classificati da esperti del dominio allo scopo di creare un set di apprendimento sufficientemente grande per addestrare il classificatore. In un secondo momento, si procederà alla categorizzazione di tutti i documenti del database. *Verbum*, inoltre, verrà collegato semanticamente con le opere primarie disponibili online e con altre risorse in

LOD⁷, in modo da consentire interrogazioni complesse e composite su autore, anno, parole chiave, opera, titolo, abstract [3]. Inoltre, le risorse testuali oltre a essere accessibili e ricercabili, saranno rese disponibili per il download in modalità open access.

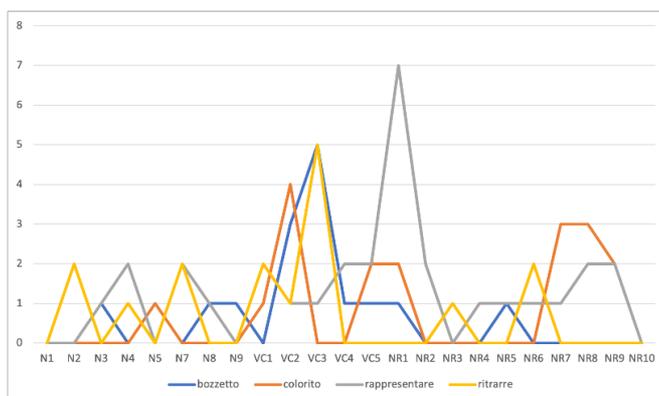


Figura 4. Lessico delle arti visive per "parole dominanti"

4. RISULTATI ATTESI

Nell'ambito del progetto, la valorizzazione e disseminazione dei risultati attesi è affidata a un portale web costituito da diversi moduli che integrano strumenti di consultazione e interrogazione avanzata. L'obiettivo è di ridefinire la fruizione e l'interpretazione della letteratura secondaria del Verismo attraverso l'impiego di strumenti di esplorazione, analisi e divulgazione del *corpus* in modo semplice ed efficace. Il primo modulo si avvale di una piattaforma web basata sul modello offerto dal software TEI Publisher⁸, che consente la visualizzazione delle riproduzioni digitali delle riviste in parallelo alle loro trascrizioni. La codifica del testo è conforme allo standard XML/TEI e incorpora tag che forniscono informazioni riguardanti i metadati, le date citate all'interno dell'opera, nomi di personaggi, luoghi, incluse alcune note di carattere filologico, linguistico e lessicale (vd. Fig. 5).

Il secondo modulo sarà costituito da un motore di ricerca avanzato che consente di interrogare i testi sulla base di diversi tag utilizzati per l'annotazione del testo XML, quali autore, periodo, opere citate, nomi, luoghi e parole chiave [22].

Il terzo modulo farà uso del *Natural Language Processing* e delle librerie Javascript per fornire agli utenti una visualizzazione dei dati sul modello dello *storytelling*. Questi strumenti digitali saranno funzionali alla costruzione di una linea del tempo con sviluppo diacronico, che abbiamo denominato *Ver-in-time*.

The screenshot shows the TEI Publisher interface for the article "IL PRINCIPIO DEL REALISMO" by Francesco De Sanctis. The left sidebar contains the following information:

- Summary:** Recensione allo scritto filosofico omonimo di von Kirchmann, dal quale l'autore prende spunto per la sua riflessione sul metodo scientifico nella rappresentazione artistica della realtà, che deve essere sì coerente ai limiti dell'osservabile, ma non priva del calore dell'immaginazione e del sentimento.
- Author:** Francesco De Sanctis (18 marzo 1817- 29 dicembre 1883)
- Keywords:** Realismo, Intellettuali, Filosofia
- Persons:** Francesco Bacone (12 gennaio 1561 - 9 aprile 1626), Cartesio (31 marzo 1596 - 11 febbraio 1650), Étienne Bonnot de Condillac (30 settembre 1714 - 3 agosto 1780), John Locke (29 agosto 1632 - 28 ottobre 1704)

The central text area displays the title "IL PRINCIPIO DEL REALISMO" and the beginning of the article, starting with "M'è venuto innanzi ultimamente l'Annuario di una Biblioteca filosofica, pubblicata a Lipsia fin dal 1868, e ricca già di sessantasette volumi. Mi ha fatto impressione vedere in quella raccolta certi nomi, che da un pezzo erano usciti di moda, Cartesio, Bacone, Locke, e fino Condillac. Mi pareva trovarmi in pieno secolo XVIII. Parimente stupii che in Germania gli studi filosofici fossero ancora coltivati con tanto ardore, che vi fosse possibile una così vasta pubblicazione, la quale presso di noi non avrebbe che pochissimi lettori. È la Biblioteca non è solo una ristampa, ma ci trovo prefazioni, dilucidazioni e commentari, i quali talora pigliano parecchi volumi, e rivelano la presenza di una mente attiva e direttrice. Questa mente è il presidente Kirchmann. Mi ricordai allora della Società filosofica di Berlino, di cui il Kirchmann è uno de' membri più distinti e più attivi. Conoscevo per udita la sua *Filosofia della*

The right sidebar shows a thumbnail of the original document page, which matches the title and content of the central text area.

Figura 5. TEI Publisher, F. De Sanctis, «Il principio del realismo»

⁷ Va segnalato, tra le risorse esterne a cui il portale intende collegarsi, VIVer (<https://testi.progettoviver.it/>), un *corpus* leggibile e interrogabile di testi della letteratura del Verismo esteso alle diverse espressioni regionali, in cui i testi presentano una marcatura di tipo linguistico e sintattico-fraseologico.

⁸ <https://teipublisher.com/exist/apps/tei-publisher-home/index.html>

I tre moduli saranno integrati all'interno del portale, che si prefigge di superare i limiti entro cui sono confinati i semplici archivi di risorse digitalizzate, offrendo un ambiente interattivo per l'esplorazione, l'analisi e la divulgazione del *corpus* della letteratura secondaria relativa al Verismo [9].

La digitalizzazione integrale dei periodici selezionati, che saranno resi disponibili sul portale, non solo preserverà un bene culturale materiale a rischio di deperimento, ma lo renderà accessibile a un pubblico globale, incentivando la lettura critica e l'interazione con documenti e testimonianze letterarie e storico-culturali "di prima mano", di cui la pubblicistica periodica è un eccezionale veicolo nell'Ottocento.

Dal punto di vista analitico, il progetto incorporerà e implementerà strumenti di *Text Mining* e *Natural Language Processing* per condurre analisi quali la *sentiment analysis*, il *topic modelling* e l'estrazione di entità nominate.

Incrociare, all'interno del portale e con gli strumenti di interrogazione da esso forniti, i dati dei testi della letteratura secondaria in modo semplice e immediato può favorire uno studio più profondo e complesso delle intricate ramificazioni dei temi, delle tendenze e dei contesti storici e culturali associati al Verismo. Questi strumenti analitici non solo arricchiranno la piattaforma con nuove modalità di esplorazione dei dati, ma contribuiranno anche alla ricerca accademica fornendo nuove prospettive di analisi sul materiale testuale.

La realizzazione di questi risultati sottolinea l'impegno del progetto nel rispettare i principi FAIR per la gestione dei dati di ricerca, assicurando così che le risorse prodotte siano di ampia fruibilità e sostenibilità.

In conclusione, la piattaforma realizzata non sarà solo un ambiente digitale (mera interfaccia per sfogliare online edizioni critiche nate come cartacee), ma riprenderà il concetto che vede nell'interazione dell'utente, nella possibilità di creare percorsi "personali" a seconda dei propri interessi, il cuore della fruizione di *corpora* ed edizioni realmente digitali, in linea con quegli «ambienti, o anche infrastrutture, che raccolgono testi, servizi, interfaccia e strumenti di accesso *in usum philologorum*» [13: 113].

5. RINGRAZIAMENTI

Si ringraziano Christian D'Agata e Angelo Mario Del Grosso per il supporto nella costruzione del modello di codifica proposto e Denise Maci e Martina Corti per la codifica delle recensioni. "Finanziato dall'Unione Europea – Next Generation EU".

BIBLIOGRAFIA

- [1] Briganti, Alessandra, Camilla Cattarulla, e Franco D'Intino. *I periodici letterari dell'Ottocento: indice ragionato (collaboratori e testate)*. Milano: F. Angeli, 1990.
- [2] Chemello, Adriana. «*La Farfalla*» di Angelo Sommaruga. *Storia e indici*. Roma: Bulzoni, 1977.
- [3] Cristofaro, Salvatore, Christian D'Agata, Antonio Di Silvestro, Giuseppe Palazzolo, Pierpaolo Sichera, e Daria Spampinato. «DEMOTICOn. Per un'edizione semantica dei Malavoglia». In *AIUCD2021 Book of Extended Abstracts*, (a cura di) Federico Boschetti, Angelo Mario Del Grosso, e Enrica Salvatori, 471–73. Quaderni di Umanistica Digitale, 2021. <https://doi.org/10.6092/unibo/amsacta/6712>.
- [4] Cristofaro, Salvatore, Angelo Mario Del Grosso, Laura Mazzagufu, Pierpaolo Sichera, e Daria Spampinato. «Bellini Digital Correspondence: A Model for Making Collaborative Digital Scholarly Editions», 615–20. Agadir - Essaouira, Morocco: IEEE, 2023. <https://doi.org/10.1109/CiSt56084.2023.10409920>.
- [5] Cristofaro, Salvatore, e Daria Spampinato. «Aspetti funzionali e implementativi del Museo epigrafico digitale EpiCUM». *Umanistica Digitale* 4, fasc. 9 (2020): 61–77. <https://doi.org/10.6092/issn.2532-8816/9973>.
- [6] Daquino, Marilena, Francesca Giovannetti, e Francesca Tomasi. «Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini». *Umanistica Digitale* 3, fasc. 7 (2019). <https://doi.org/10.6092/issn.2532-8816/9091>.
- [7] De Pasquale, Andrea. «The Italian national digital newspaper library». *Bibliothecae.It* 7, fasc. 2 (2018): 348–70. <https://doi.org/10.6092/issn.2283-9364/8951>.
- [8] Del Grosso, Angelo Mario, Andrea Bellandi, Emiliano Giovannetti, Simone Marchi, e Ouafae Nahli. «Scanning is Just the Beginning: Exploiting Text and Language Technologies to Enhance the Value of Historical Manuscripts». In *IEEE 5th International Congress on Information Science and Technology*, 2018. <https://doi.org/10.1109/CIST.2018.8596373>.
- [9] Di Silvestro, Antonio, Christian D'Agata, Giuseppe Palazzolo, e Pierpaolo Sichera. «Conservazione e fruizione di banche dati letterarie: l'archivio della poesia italiana dell'Otto/Novecento di Giuseppe Savoca». In *Atti del Convegno AIUCD2022*, 98–104, 2022.
- [10] D'Orsogna, Fabio, e Giulio Palanga. «Riviste digitali e digitalizzate italiane (RIDI): a reconnaissance for the national newspaper library». *JLIS.It* 13, fasc. 1 (2022): 374–89. <https://doi.org/10.4403/jlis.it-12734>.
- [11] Farinelli, Giuseppe. *La pubblicistica nel periodo della scapigliatura*. Milano: IPL, 1984.

- [12] *I verismi regionali. Atti del Congresso Internazionale di Studi, Catania, 27-29 aprile 1992*. 2 voll. Catania: Fondazione Verga, 1996.
- [13] Italia, Paola, e Francesca Tomasi. «Filologia digitale. Fra teoria, metodologia e tecnica». *Ecdotica* 11 (2014): 112–31. <https://doi.org/10.7385/99218>.
- [14] Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, et al. «The TRUST Principles for digital repositories». *Scientific Data* 7 (2020): Article number: 144. <https://doi.org/10.1038/s41597-020-0486-7>.
- [15] Melis, Rossana. *La bella stagione del Verga*. Catania: Biblioteca della fondazione Verga, 1990.
- [16] Pellini, Pierluigi. *Naturalismo e verismo. Zola, Verga e la poetica del romanzo*. Firenze: Le Monnier Università, 2010.
- [17] Rappazzo, Felice, e Giovanna Lombardo. *Giovanni Verga fra i suoi contemporanei*. Soveria Mannelli: Rubbettino, 2016.
- [18] Raya, Gino. *Bibliografia di Luigi Capuana (1839-1968)*. Roma: Ciranna, 1969.
- [19] Raya, Gino. *Bibliografia Verghiana (1840-1871)*. Roma: Ciranna, 1972.
- [20] Romano, Cinzia. *Emanuele Navarro della Miraglia: un percorso esemplare di secondo Ottocento*. Catania: Biblioteca della Fondazione Verga, 1998.
- [21] Saja, Giuseppe. «*Il Momento*». *Identità di una rivista di fine Ottocento con gli indici del periodico (1883-1885)*. Caltanissetta-Roma: Sciascia, 2004.
- [22] Spadini, Elena, Francesca Tomasi, e Georg Vogeler, (a cura di). *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*. Norderstedt: Herstellung und Verlag, 2021.
- [23] Wilkinson, Mark, Michel Dumontier, Ijsbrand J. Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. «The FAIR Guiding Principles for scientific data management and stewardship». *Scientific Data* 3, fasc. 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.