

# *IL VELTRO*

RIVISTA DELLA CIVILTÀ ITALIANA



**ESTENSIONE ON LINE – FASCICOLO 1/2 2025**

ITALIA NEL MONDO intende promuovere, in Italia e fuori, la consapevolezza della tradizione e del presente della società italiana; delle sue affermazioni ideali, creative, umanitarie; dei valori e dei problemi che ne hanno orientato il corso storico; delle relazioni con altri Paesi, culture, società.  
Intende particolarmente favorire la partecipazione italiana alla ricerca contemporanea di prospettive originali e di tematiche innovatrici.



Sul frontespizio:

Piccolo levriero dalla stampa di  
S. Gioacchino di Wolfgang Huber  
(1480-1549)

IL VELTRO  
RIVISTA DELLA CIVILTÀ ITALIANA  
Organo di ITALIA NEL MONDO  
Rivista fondata nel 1957  
da Aldo Ferrabino e Vincenzo Cappelletti.

•  
COMITATO SCIENTIFICO:  
Mario Boffo; Vinicio Busacchi; Americo Cicchetti; Guido  
Cimino; Renato Cristin;  
Lorenzo Franchini; Paolo Garbini;  
Francesco Guida; Danijela Janjic';  
Cristiana Lardo; Giuseppe Manica; Ida Nicotra; Bernardo  
Piciché; Giovanni Pocaterra;  
Paolo Puppa; Roberto Rossi; Fabio Sattin; Paolo  
Tondi

REDAZIONE:  
Giovanni Barracco, Capo redattore  
letteratura e filosofia;  
Camilla Tondi, Capo redattore arte,  
scienze mediche e biologiche;  
Veronica Tondi, Capo redattore  
diritto ed economia.  
Coordinamento redazionale: Camilla Tondi

CLAUDIA CAPPELLETTI  
Direttore

VIRGINIA CAPPELLETTI  
Direttore responsabile

Simone Bocchetta, Responsabile editoriale

DIREZIONE, REDAZIONE,  
AMMINISTRAZIONE  
Via Giuseppe Gioachino Belli, 86  
00193 Roma info@ilveltrorivista.it  
ilveltrorivista.eu

Tutti i contributi pubblicati che afferiscono alle  
discipline per le quali la rivista *Il Veltro* è  
classificata nelle fasce ANVUR vengono  
sottoposti a un procedimento di revisione tra  
pari a doppio cieco (*double blind*).

• Abbonamento ordinario:

Italia € 90,00,  
Europa € 120,00, Altri  
Paesi € 160,00,  
Sostenitore € 200,00.  
Conto corrente postale 834010.

•  
© 2025

Edizioni Studium

Per informazioni sugli abbonamenti:  
abbonamenti@edizionistudium.it

ISSN 0042-3254

Autorizzazione del Tribunale di Roma  
N. 5643 in data 12-2-1957

Stampa: Marchesi Grafiche Editoriali Via  
dell'Artigianato, 19  
00065 Fiano Romano (Roma)

Trimestrale - Poste Italiane S.p.A. Spedizione in  
abb. post. D.L. 353/2003 (conv. in L.  
27/02/2004 n. 46)  
art. 1 comma 1 CN/FC

## SOMMARIO

### PERCORSI DEL PENSIERO SCIENTIFICO BIOLOGICO E MEDICO

di Vincenzo Cappelletti

Introduzione di Guido Cimino

Unità e storia della scienza (1983)

Sapere specialistico e sapere storico (1987)

Un percorso della ragione scientifica (2010)

Sulla dinamica dei paradigmi scientifici (1986)

Duplici rivoluzioni della scienza (2000)

Scienza dell'umanesimo e scienza illuministica (1977)

Evoluzionismo, creazionismo, neodarwinismo (2009)

Ontogenesi della vita (2007)

Il genoma umano: panorama storico e problemi etici (1998)

Morgagni e Virchow (1987)

Momenti della biologia tedesca: da Virchow a Driesch (1982)

Biomedicina del XX secolo (2003)

Medicina scientifica e medicina applicata (1999)

Sommario della Estensione online del Fascicolo 1-2/2025

**LETTERATURA**

Collaborazione redazionale di Massimo Castiglioni e Alessandro Gerundino

**DOSSIER**

**VERISMO IN RETE. VERGA, CAPUANA, DE ROBERTO TRA LESSICOGRAFIA, FILOLOGIA E CRITICA**

A cura di Antonio Di Silvestro e Liborio Pietro Barbarino

Marina Paino, Andrea Manganaro, Antonio Sichera, Antonio Di Silvestro, Liborio Pietro Barbarino, Introduzione 8

A. LETTURE

Liborio Pietro Barbarino, Scrivere su un margine virtuale. Per un commento digitale ai *Malavoglia* di Giovanni Verga (capitolo I) 14

Ottavia Branchina, La colica, il cane, il corvo. Commentare il *Mastro-Don Gesualdo* tra carta e digitale 34

Christian D’Agata, Per una lettura apocalittica di *Viceré e Imperio* di Federico De Roberto: Risorgimento, crisi, fine del mondo 57

Eliana Vitale, Di mutrie, mutismi e male parole: la parola antirelazionale nei *Viceré* di Federico De Roberto 79

B. VARIANTI

Mariagiusti Polizzi, Appunti per una nuova edizione de *Il Marchese di Roccaverdina* di Luigi Capuana 108

Miryam Grasso, Il laboratorio compositivo del Capuana “fantastico”: *Il Dottor Cymbalus* dalla rivista all’edizione a stampa 121

Elisa Conti, La ricerca di una nuova lingua in *C’era una volta*. Per uno studio del laboratorio variantistico di Luigi Capuana 143

Denise Bruno, La «concretezza» e il «fantastico»: la dàrsena filologica del Capuana per ragazzi 159

C. LESSICO

Antonio Di Silvestro, Per un dizionario tematico del Verismo: storia di *bozzetto* 175

Gabriella Alfieri, Stephanie Cerruto, Marco Biffi, Giovanni Salucci, Verso il vocabolario digitale dell’italiano verista (VIVer): *corpus*, metodi e prospettive 198

**DOSSIER**

**IL ROMANZO DI FAMIGLIA ITALIANO: NUOVE INDAGINI E PROSPETTIVE**

A cura di Giovanni Barracco e Lorenzo Mecozzi

Giovanni Barracco, Lorenzo Mecozzi, Introduzione 224

Mauro Distefano, *I Malavoglia*: romanzo familiare tra modernità e attualità 228

Andrea Sartori, Genealogie familiari. *I Viceré* (1894) ‘dopo’ *I Buddenbrook* (1901) 250

Luigi Gussago, Genio e sregolatezza. Percorso narrativo di una famiglia disgregata in *I divoratori* (1911) di Annie Vivanti 270

Emanuele Delfiore, Elisa filologa romanzesca: l'epistolario di Anna ed Edoardo in <i>Menzogna e sortilegio</i>	289
Lucia Faienza, Ricostruire l'albero. Il romanzo di famiglia di Natalia Ginzburg, tra dissolvimento e connessioni intertestuali	305
Silvia Annavini, <i>Homely/Unhomely</i> : il perturbante familiare. Natalia Ginzburg tra spazio domestico e scrittura minore	319
Alessandro Gerundino, La famiglia e le case: <i>Althénopis</i> di Fabrizia Ramondino	337
Marco Marzi, Aria di famiglia nel contesto brigatista	357
Giuseppe D'Angelo, «Nessuna resa dei conti». Il <i>family novel</i> di Antonio Franchini	374
Sonia Glauser, <i>L'abusivo</i> e <i>Il fuoco che ti porti dentro</i> di Antonio Franchini: un raffronto tra famiglie e generi	395
Serena Cianciotto, Romanzi multigenerazionali oggi	414
<b>ALTRA CRITICA</b>	
Paolo Puppa, Abramo in scena	435
Antonella De Blasio, Due romanzi post-millennial di Sally Rooney	451
Elena Grazioli, Finzioni biografiche e pubbliche conferenze: la ricezione della Beatrice dantesca nell'Ottocento	472
<b>STORIA DELLA DIPLOMAZIA</b>	
Massimo Spinetti, La cultura e la lingua italiana nell'azione diplomatica di Costantino Nigra	491
<b>CULTURA E SOCIETÀ</b>	
Elisabetta Vaccarone, Franco Pistono, Valerio Ciarocchi, Musica, mito, ambiente e intelligenza artificiale: una riflessione	504
<b>CINEMA</b>	
Enrico Procentese, Tra assurdo e assenza: L'eclisse e l'attesa di Godot. Intervista a Gianni Massironi	521
<b>RECENSIONI</b>	
<b>GEOPOLITICA</b>	
Mario Boffo, <i>Houti – Vengono da lontano, guardano al futuro</i> (di Athanasia Andriopoulou)	532
<b>LETTERATURA</b>	
Gabriele d'Annunzio, <i>Il fuoco</i> (di Giovanni Barracco)	535
Angelo Conti, <i>La beata riva. Trattato dell'oblio</i> . Preceduta da un «Ragionamento» di Gabriele d'Annunzio (di Giovanni Barracco)	540

## **LETTERATURA**

**VERISMO IN RETE.  
VERGA, CAPUANA, DE ROBERTO  
TRA LESSICOGRAFIA, FILOLOGIA  
E CRITICA**

a cura di Antonio Di Silvestro e Liborio Pietro Barbarino

## C. LESSICO

## VERSO IL VOCABOLARIO DIGITALE DELL'ITALIANO VERISTA (VIVER): CORPUS, METODI E PROSPETTIVE

*Il VIVer è un progetto di lessicografia digitale con intenti storico-letterari e storico-linguistici il cui obiettivo è la descrizione organica dell'italiano verista, anche nella prospettiva di colmare un vuoto nell'umanistica digitale restituendo un corpus unitario che documenti la testualità del Realismo italiano. Per questo incorpora vari generi testuali (narrativa, teatro, trattati di poetica e di sociologia, poesia, critica, letteratura educativa) atti a documentare organicamente l'italiano del Verismo nelle sue diverse articolazioni estetiche e intellettuali. In questo contributo si presentano le caratteristiche strutturali e tecniche del vocabolario, reticolare e dinamico, dell'italiano dei veristi, basate su quelle del VoDIM dell'Accademia della Crusca, evidenziandone le specificità, come la fraseologia, e producendo alcuni esempi di “voci”.*

Parole chiave: *Verismo, Realismo, lessicografia digitale, vocabolario dinamico, corpus verista*

*VIVer is a digital lexicography project with literary-historical and historical-linguistic intentions, which aims at an organic description of the Italian verism, also with a view to filling a gap in the digital humanities by returning a unitary corpus that documents the textuality of Italian Realism. To this end, it incorporates various textual genres (fiction, theatre, treatises on poetics and sociology, poetry, criticism,*

*educational literature) to organically document the Italian of Verismo in its various aesthetic and intellectual articulations. In this contribution are presented the structural and technical features of the vocabulary, reticular and dynamic, of the Italian of the veristi, based on those of Accademia della Crusca's VoDIM. I will therefore highlight some of its features, such as phraseology, and produce some examples of 'voices'.*

Keywords: *Verism, Realism, digital lexicography, dynamic vocabulary, verist corpus*

## 1. Un vocabolario realista e dinamico<sup>1</sup>

Il VIVer è un progetto di lessicografia digitale finalizzato alla descrizione organica dell'italiano verista, con intenti ben definiti di natura storico-letteraria, storico-linguistica e storico-culturale<sup>2</sup>. Sul fronte letterario si mira ad attuare il programma di ricerca che la Fondazione Verga sta perseguendo negli ultimi anni: rivisitare e reintegrare il Verismo nel quadro sovranazionale del Realismo letterario, condividendo col Centre Zola il presupposto del *Naturalisme monde*, per cui solo una lettura trasversale garantirà un'interpretazione globale della narrativa realista, senza perciò obliterare le specificità nazionali<sup>3</sup>.

Sul versante storico-linguistico la produzione letteraria, teatrale, critica e metadiegetica del Realismo va altresì rapportata alla letteratura educativa (in senso lato) e alla pubblicistica prodotte dal *nation building* etico dell'Italia unita, che aveva come ricaduta sociocomunicativa la formazione di un linguaggio ispirato al modello di parlato-scritto o scritto-parlato teorizzato da Manzoni. Questa tipologia di lingua, testimoniata da una mole di scritti articolati in una cospicua varietà di generi testuali, ha trovato un'adeguata sede analitica in chiave pragmatico-testuale nel progetto PRIN 2022 denominato ARDIPS (Archivio Digitale dell'Italiano Parlato Scritto)<sup>4</sup>. Una descrizione lessicografica dell'italiano verista all'interno dell'italiano postunitario, come quella che Accademia della Crusca e Fondazione Verga stanno perseguendo in sinergia con numerose università italiane<sup>5</sup> realizzando il VIVer, può considerarsi pertanto un obiettivo del tutto pionieristico e innovativo, anche nell'ambito del verismo digitale.

Uno spoglio dell'italiano verista impostato seguendo le attuali tendenze della lessicografia informatizzata costituirà per gli storici della lingua una fonte preziosa per conoscere – e spesso retrodatare – un patrimonio lessicale e fraseologico ancora in gran parte ignorato.

Il VIVer, ideato dalla Fondazione Verga nel 2015 e avviato nel 2017 grazie a un finanziamento del MIC, si incentra sulla testualità letteraria e documentaria prodotta a ridosso

dell'Unità e subito prima del ventennio fascista: saranno inseriti nel corpus testi pubblicati dal 1850 (data convenzionale di diffusione della letteratura rustica) al 1922, anno della morte di Giovanni Verga. I testi inseriti nel corpus appartengono ad aree geoculturali di tutta la Penisola, conformemente al carattere regionale del Verismo ma anche alla politica culturale dello Stato postunitario, in cui al drastico monocentrismo fiorentino del ministero Broglio si erano affiancati progetti per un'unificazione linguistica 'federativa', fondata sulla progressiva convergenza verso il toscano di esperienze letterarie, filologiche e lessicografiche di origine regionale<sup>6</sup>.

Dopo aver illustrato i presupposti storico-culturali e storico-linguistici del VIVer e in un certo senso la sua natura di vocabolario realista, atto cioè a rappresentare il Realismo linguistico, passiamo a chiarirne la natura di vocabolario dinamico, sul piano tipologico e non certo su quello tecnico che sarà approfondito nel paragrafo 3.2.3.

Innanzitutto, va sottolineato che il VIVer si riconnette idealmente all'autorevole vocazione lessicografica della linguistica computazionale italiana<sup>7</sup>, nonché alla vocazione prioritariamente semantica della lessicografia<sup>8</sup>. Fondandosi su un corpus plurigenere, di cui si dirà più avanti, il VIVer contemplerà ambiti linguistico-semantici finora inesplorati negli altri vocabolari digitali, con l'eccezione del VoDIM<sup>9</sup>, che effettivamente ha introdotto in Italia un modello di vocabolario dinamico inteso come sistema lessicale digitale perennemente implementabile. Tale prerogativa consentirà di realizzare uno degli intenti sopra enunciati: verificare le risultanze della politica di unificazione 'federativa' dei ministeri postunitari nella scrittura verista, i cui testi potranno via via essere immessi e marcati nel corpus.

Primi sondaggi in questa importante direzione erano stati svolti in un importante convegno sui Verismi organizzato a Catania dalla Fondazione Verga nel 1992<sup>10</sup>. Dagli scandagli stilistici erano emerse scelte espressive concomitanti in tutti gli autori, indipendentemente dalla provenienza geolinguistica e dalla reciproca e rispettiva conoscenza: proverbi, modi di dire, codice gestuale, similitudini convenzionali, sintassi modellata sul parlato costituivano un codice condiviso e oggettivamente percepibile<sup>11</sup>. A queste acquisizioni storico-critiche fa riferimento l'impresa lessicografica del VIVer.

Il corpus della testualità verista italiana è concepito come corpus dei corpora dinamico, incrementabile in ogni momento e non rigidamente legato al canone letterario verista.

## 1.1 Un corpus dei corpora

Il corpus del VIVer è composto innanzitutto da testi rappresentativi del patrimonio

testuale dei veristi italiani, selezionati, come si diceva, a partire dagli autori studiati nel citato Convegno su *I verismi regionali*; a questo primo nucleo si sono via via aggregati testi afferenti alla produzione postunitaria gravitante intorno al Realismo: narrativa, teatro, melodramma, produzione critica o pubblicistica sociologica che forniva spunti tematici al Verismo.

Il progetto si fonda su basi filologiche, assumendo come testi di riferimento le edizioni critiche o, in assenza di queste, autorevoli edizioni commentate o ancora le prime edizioni delle opere, e si articola in generi testuali accorpati per ambiti geolinguistici secondo il seguente schema, in cui si indicano tra parentesi le sigle di marcatura:

1. Narrativa: novelle e romanzi (Narr)
2. Poesia verista (Poe)
3. Teatro (Tea)
4. Realismo sociale (RSO)<sup>12</sup>
5. Verismo documentario (Gio)<sup>13</sup>
6. Testi teorici e manifesti poetici (Est)<sup>14</sup>
7. Letteratura per l'infanzia (Fia)<sup>15</sup>

In prospettiva, per completare il quadro descrittivo dell'italiano verista, si includeranno nel corpus gli epistolari di Verga, Capuana, De Roberto e altri/altre autori/autrici, la cui stretta inerenza con il linguaggio narrativo è già stata dimostrata per Verga<sup>16</sup>.

In ordine al raccordo tra Verismo e Realismo europeo, è prevista in futuro l'integrazione delle traduzioni italiane, coeve agli autori veristi italiani, di Zola, Auerbach, Hardy, ecc.

La scelta di includere nel corpus i paratesti e i metatesti (prefazioni, recensioni e, in prospettiva, carteggi d'autore rilevanti ai fini della teorizzazione estetica) contribuirà a descrivere e storicizzare il sottocodice della poetica realista e verista, di cui la «poetica disseminata»<sup>17</sup> di Verga costituisce l'archetipo.

Il corpus è formato attualmente da oltre 200 testi, di cui 52 già revisionati e in corso di marcatura. I testi sono rappresentativi sul piano geoculturale e tipologico-testuale e riflettono le caratteristiche tendenziali dell'italiano verista: idiomatilità spiccata e conguaglio normativo sul toscano post-manzoniano.

### 1.1 **Marcatura dei testi e tracciato del VIVer**

I tag linguistici, dopo una lunga riflessione teorica<sup>18</sup>, sono stati scelti seguendo le

costanti emerse da importanti studi, che costituiscono la base del progetto VIVer, come gli atti del già citato convegno *I verismi regionali*<sup>19</sup>:

- ricorso al dialetto, il più delle volte italianizzato, o all'italiano regionale, e ricorso al toscano come lingua superdialettale di conguaglio;
- uso dei tecnicismi per la caratterizzazione socio-ambientale;
- ampio uso di strutture iperlinguistiche e antropologiche (similitudini convenzionali, espressioni fraseologiche, proverbi, sentenze, codice gestuale, ecc.)<sup>20</sup>.

Per la marcatura linguistica si è scelto pertanto di procedere con la creazione di macro-tag, volutamente generici e ampi, che rispecchiano queste caratteristiche, a cui se ne sono aggiunte altre emerse dai primi sondaggi sul corpus VIVer. I tag definitivi sono quindi:

- regionalismi: termini dialettali puri o adattati, regionalismi, voci appartenenti in un certo senso al “dialetto nascosto”<sup>21</sup>, toscanismi demotici e toscano-italianismi<sup>22</sup>;
- tecnicismi: tecnicismi socio-ambientali, ossia termini e polirematiche regionali o panitaliani<sup>23</sup> che pertengono ai linguaggi settoriali motivati dal contesto narrativo (lessico marinaresco, contadino, medico, economico, ecc.);
- tecnicismi metadiegetici: singoli termini ed espressioni che vengono usati per riferirsi alla corrente verista o per descriverne le caratteristiche<sup>24</sup>;
- gergalismi: termini o espressioni appartenenti a sottocodici mimetici, adoperati da gruppi sociali emarginati come i detenuti o i malavitosi, per rendersi riconoscibili all'interno e incomprensibili all'esterno;
- nomignoli: soprannomi etnici che hanno valore identificativo all'interno di comunità ristrette;
- formularità: unità “fisse” più caratterizzanti dal punto di vista stilistico, come similitudini convenzionali, espressioni idiomatiche, proverbi, sentenze, gestualità, vale a dire gesti codificati con valore socio-etnico-comunicativo.

Quest'ultima categoria e quella dei regionalismi hanno impegnato lungamente il gruppo fondante del VIVer in quanto includono tipologie idiomatiche eterogenee e costitutive dello stile verista: da un'iniziale scelta di individuare con singoli tag ogni singola tipologia, si è poi optato per rubricare sotto etichette ampie i tratti linguistici connotati da caratteri fortemente accomunanti.

Si è deciso, inoltre, di introdurre una marcatura per evidenziare altri fenomeni di interesse linguistico, vale a dire le neoformazioni d'autore e le retrodatazioni. Queste ultime consentiranno una stretta collaborazione con l'importante progetto lessicografico

ArchiDATA<sup>25</sup>.

L'aspetto forse più innovativo e si direbbe esclusivo del VIVER è la registrazione delle unità fraseologiche che finora la lessicografia ha trattato univocamente come appendici delle voci messe a lemma. Una prima rubricazione autonoma ma limitata alla paremiologia si ha ultimamente nella banca dati *Proverbi italiani* di cui è responsabile Marco Biffi<sup>26</sup>. Un'ulteriore novità del VIVER è poi quella di raccogliere e descrivere organicamente per la prima volta il lessico e le espressioni che ruotano intorno alla poetica del verismo, obiettivo che il VIVER condivide con un altro importante progetto dell'Università di Catania, COVerLeSS (Corpus Online del Verismo tra Letteratura, Storia e Società), diretto da Antonio Di Silvestro<sup>27</sup>.

Allo stato attuale la ricerca del VIVER si sta concentrando sulla marcatura strutturale e linguistica dei circa 200 testi che costituiscono il corpus. Dai dati finora emersi dalla marcatura è stato elaborato dal gruppo di ricerca della Fondazione Verga (Gabriella Alfieri e Stephanie Cerruto) e dell'Accademia della Crusca (Marco Biffi e Giovanni Salucci) un prototipo di tracciato delle future voci del vocabolario che saranno create a partire dall'elaborazione di concordanze e di un lemmario.

Nel prototipo, pensato sia per le singole parole sia per la formularità, il lemma sarà articolato in:

- entrata
- marca grammaticale o il tipo di formulazione se si tratta di un'espressione fraseologica
- (eventuale) glossa metalinguistica presente nel corpus
- (eventuale) traduzione per le voci dialettali
- anno di prima attestazione nel corpus + il testo della prima attestazione
- prima attestazione fuori corpus VIVER
- corrispondenze lessicografiche (comprende anche i riferimenti ad altri vocabolari e ad altri corpora esterni al VIVER, per cui si ha a disposizione la fondamentale Stazione lessicografica della Crusca<sup>28</sup>, che consente di consultare in contemporanea i dizionari elettronici esistenti e i corpora dell'italiano esistenti per l'arco cronologico postunitario, oltre ai dizionari dialettali e altre fonti utili alla compilazione della voce)
- nota (etimologica, linguistica, stilistica, filologica) + bibliografia

Il lemma del vocabolario sarà completato dalle forme con cui la voce o i costrutti fraseologici si presentano nel nostro corpus.

Nel tracciato seguono a questo punto la definizione e le eventuali definizioni subordinate accompagnate da:

- marca d'uso
- campo semantico
- distribuzione geografica.

Il significato o i significati saranno corredati dai contesti di occorrenza, per i quali saranno segnalati l'autore, l'opera, l'anno e la pagina di riferimento dell'edizione prescelta.

Data l'ampiezza del corpus e la prevedibile mole di occorrenze per ogni singola voce, è prevista nel tracciato una selezione dei contesti più rappresentativi, mentre l'elenco completo delle opere in cui compare il termine o l'espressione sarà fornito in un campo dedicato.

Per i campi *marca grammaticale/tipo di formulazione*, *marca d'uso*, *campo semantico* e *corrispondenze lessicografiche*<sup>29</sup> sono state predisposte delle liste di opzioni fra cui il compilatore del lemma sceglierà la più adeguata alla redazione della voce.

Presentiamo in questa sede la voce *càlia*, non ancora definitiva, nel formato iniziale del tracciato previsto dal VIVer:

- Entrata: *càlia*
- Marca grammaticale: sf
- Glossa metalinguistica: *come dicono laggiù*
- Traduzione: *i ceci cotti sotto la rena (la càlia, come dicono laggiù)*
- Anno di prima attestazione nel corpus: 1914 – Fojanesi, *In Toscana e in Sicilia*
- Prima attestazione fuori corpus VIVer: 1768 (*Capitoli del venerabile Monte della Pietà di questa felice, e fedelissima Capitale del Regno di Sicilia Palermo opera disposta da Vincenzo Parisi de' marchesi dell'Ogliastro*. Fonte: Google Libri)
- Corrispondenze lessicografiche: Caglià (1840), Biundi (1851), Pitre (1870, 1875), VS, GRADIT
- Nota: voce siciliana e calabrese, dall'arabo *qaliyya* 'fritto, arrostito' (VSES)
- Forme: *càlia*
- Significati:
  1. Ceci abbrustoliti

Marca d'uso: REG (regionalismo)

Campo semantico: cibo

Distribuzione geografica: Sicilia, Calabria

Contesti:

«Si rosicchiarono le fave abbrustolite e *i ceci cotti sotto la rena (la càlia, come dicono laggiù)*» (Fojanesi, *In Toscana e in Sicilia*, 1914: 54)

Un altro esempio, questa volta di un proverbio, ci mostra come il tracciato faccia emergere l'importanza delle varianti presenti nel corpus e i possibili rimandi all'interno del vocabolario:

- Entrata: *gente allegra il ciel l'aiuta*
- Tipo di espressione fraseologica: proverbio
- Anno di prima attestazione nel corpus: 1908 [1906] – Nieri, *Cento racconti popolari lucchesi*
- Prima attestazione fuori corpus VIVeR: av. 1750 (Goldoni, *Il teatro comico. Commedia di tre atti in prosa*, Monaco, coi tipi di Giorgio Franz, 1847, p. 40.

Fonte: Google Libri)

- Corrispondenze lessicografiche: RF (*Gente allegra Iddio/Dio l'aiuta*), GDLI (*Gente allegra, il ciel l'aiuta*), PI<sup>30</sup> (*Gente allegra Iddio/Dio l'aiuta*)

- Forme: *gente allegra il ciel l'aiuta; cuor contento il ciel l'aiuta*
- Significati:

1. Chi è ottimista e fiducioso è aiutato dal caso/da Dio

Marca d'uso: COL (colloquiale)

Campo semantico: positività

Distribuzione geografica: panitaliana

Contesti:

«Bravo Domenico, venite, passate di qua. Beviamo un bicchiere; bisogna stare allegri in questo mondo. Gente allegra, Dio l'aiuta. Lo sapete pure?» (Nieri, *Cento racconti popolari lucchesi*, 1908: 193)

«“Gente allegra il ciel l'aiuta” pensava lui e, con la Gegia, le cose sarebbero andate bene» (Fojanesi, *In Toscana e in Sicilia*, 1914: 8)

«- Cuor contento il ciel l'aiuta, soggiunse don Zua» (Ballero, *Don Zua*, 1894: 12)

Si può prevedere inoltre il rimando a un'altra espressione, *cuor contento*, attestata ne *I Malavoglia*<sup>31</sup> che sarà messa a lemma separatamente come espressione formulare.

Le funzionalità di questo tracciato sono ben visibili, ad esempio, nel caso di *parommella*: il campo *note* consente di aggiungere informazioni importanti su questo tecnicismo che si trova ne *I Malavoglia*<sup>32</sup>. Innanzitutto, si segnalerà al consultatore che Verga opta per la variante con raddoppiamento della nasale *parommella* rispetto al *paromella* dell'autografo; inoltre, si aggiungerà un'importantissima citazione dalle lettere verghiane in cui l'autore spiega al traduttore Rod il significato dell'espressione in cui il termine è inserito: «*Alare una parommella* vuol dire = tirare dal fondo del mare un ordigno che non so come si chiami con vocabolo marinaresco in francese. Per questo e per altri termini tecnici bisognerebbe consultare un Dizionario di Marina Francese Italiano, se c'è»<sup>33</sup>.

Uno dei problemi con cui ci siamo dovuti misurare, grazie anche alle opportune osservazioni e agli utili consigli ricevuti da parte di alcuni componenti delle unità, è stata l'eccessiva complessità della voce, costruita con una griglia estremamente dettagliata, che potrebbe richiedere tempi lunghi per l'effettiva realizzazione di schede per le migliaia di voci previste per il vocabolario. Tuttavia, si è ritenuto di non sacrificare alcun campo: la concezione teorica alla base del modello lessicografico qui proposto si articola nella definizione di una scheda caratterizzata da una struttura ricca e stratificata, che comprende sia campi di interesse generale sia campi di natura più prettamente specialistica. Tale densità informativa mira a fornire un quadro esaustivo e, con opportuni applicativi, promuoverà un'esperienza di consultazione personalizzabile, consentendo agli utenti di tracciare percorsi di ricerca specifici e di adattare alle singole esigenze l'accesso ai dati.

### 3. **Aspetti informatici**

In linea con le opportunità offerte da Internet, il progetto del VIVer si propone di realizzare una piattaforma complessa, progettata e realizzata all'interno del Centro informatico dell'Accademia della Crusca<sup>34</sup>, che prevede un ambiente riservato di schedatura, catalogazione e digitalizzazione (*back-office*), e un portale integrato per la consultazione e le ricerche (*front-end*). Questa tipologia di infrastruttura informatica consente di gestire agevolmente risorse e materiali raccolti e digitalizzati nel corso del progetto per la conservazione in un repository interno, ma anche di censire e collegare risorse esterne, consentendo una maggiore valorizzazione dei materiali e degli studi realizzati in altri progetti.

Il progetto prevede l'uso sistematico di formati descrittivi in linguaggio XML, che consentono un più facile interscambio di dati, e l'adozione di appropriate procedure e linee

guida (Open Standard) che garantiscano un livello adeguato della qualità dei risultati ottenuti, certificandone la scientificità in linea con gli standard previsti. L'adozione di protocolli standard consente inoltre la diffusione e il recupero in Internet dell'informazione degli archivi e delle biblioteche digitali, nel rispetto della tutela della privacy e della proprietà dei soggetti conservatori.

Il portale del VIVer prevede l'accesso a contenuti e servizi nativi e a risorse esterne e personalizzazione per gli utenti. Per quanto l'obiettivo specifico del progetto sia la realizzazione di un dizionario, il risultato finale prevede anche la presenza di una banca dati testuale verista, che non solo costituirà il corpus rappresentativo di partenza per i lessicografi che redigeranno le voci, ma offrirà essa stessa uno strumento di consultazione di alto livello che permetterà, oltre alle indagini di tipo lessicale, in certa misura anche quelle di tipo grafico, morfologico e sintattico.

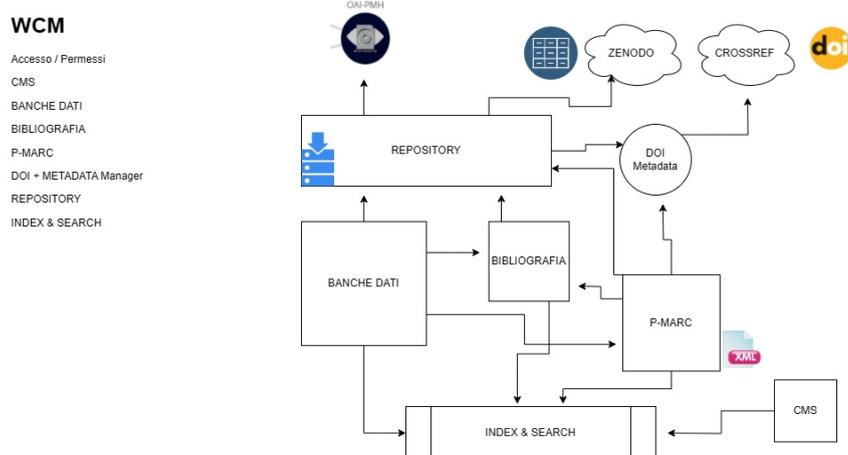
### 3.1 **Il *back-office*: l'officina della piattaforma**

Per la realizzazione della piattaforma è stata adottata la piattaforma WCM<sup>35</sup>, una piattaforma gestionale sviluppata da Progettinrete per la valorizzazione dei prodotti della ricerca nel settore dell'Umanistica digitale opportunamente configurata e personalizzata per le specifiche del progetto. La scelta di adottare questa soluzione è in continuità con altri progetti realizzati dal Centro Informatico dell'Accademia della Crusca, ma dovuta anche al fatto che WCM assicura l'integrazione di moduli per la digitalizzazione, la gestione avanzata e marcatura di prodotti e strumenti di ricerca (dizionari, glossari, fondi archivistici, corpora di varia entità e di implementare motori di ricerca e monitorare le statistiche di misurazione dell'impatto).

I moduli di WCM implementati in questo progetto sono:

- WCM P-MARC (per la gestione e la marcatura dei testi del verismo)
- WCM BancheDati (per la costruzione del dizionario del verismo e della bibliografia)
- WCM Index+Search plugin per la indicizzazione e la ricerca
- WCM CMS per la gestione dei contenuti del sito
- WCM Repository (per la realizzazione del repository e la preservazione digitale)

Nello schema seguente si illustra la presenza e il collegamento fra i vari moduli del *back-office*.



L'installazione del WCM parte dal modulo centrale, il cosiddetto “WCM core” che consente di definire l'architettura dell'applicazione, attivare i moduli previsti e le unità di ricerca coinvolte. La piattaforma consente infatti di partizionare i dati tra gruppi di lavoro che condividono la piattaforma, ma che possono avere controllo esclusivo su parti di informazione o su flussi e procedure.

Una volta definite le unità di lavoro (nel caso del VIVer sono stati attivati oltre 10 gruppi di lavoro indipendenti) è possibile aggiungere operatori, assegnando a ciascuno l'appartenenza a uno o più sedi, l'operatività sulla piattaforma e i relativi permessi per ciascun modulo (trascrizione e marcatura dei testi; revisione; validazione ecc.). La possibilità di configurare operatori con permessi differenziati, così come di aggiornare i permessi e le funzionalità in tempo reale, è uno dei vantaggi dell'utilizzo di questa piattaforma; inoltre, le operazioni svolte dai differenti operatori sono tracciate e in parte anche storicizzate, così da permettere di recuperare informazioni in caso di errori o malfunzionamenti della rete.

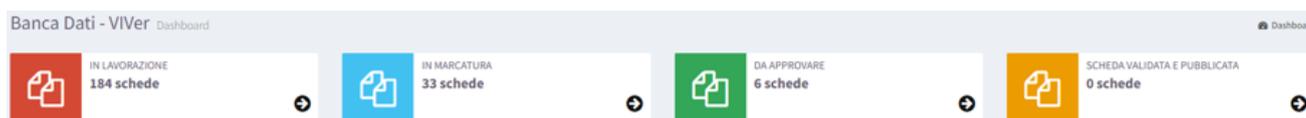
I moduli WCM P-MARC e BancheDati sono deputati alla gestione, all'inserimento e alla modifica dei dati e testi delle opere del corpus, che costituiscono l'ossatura del progetto VIVer, e alla predisposizione delle schede lessicografiche.

La struttura è organizzata in funzioni accessibili tramite un menù laterale presente in tutta l'applicazione: questo permette una più chiara e intuitiva organizzazione del lavoro per macroaree.

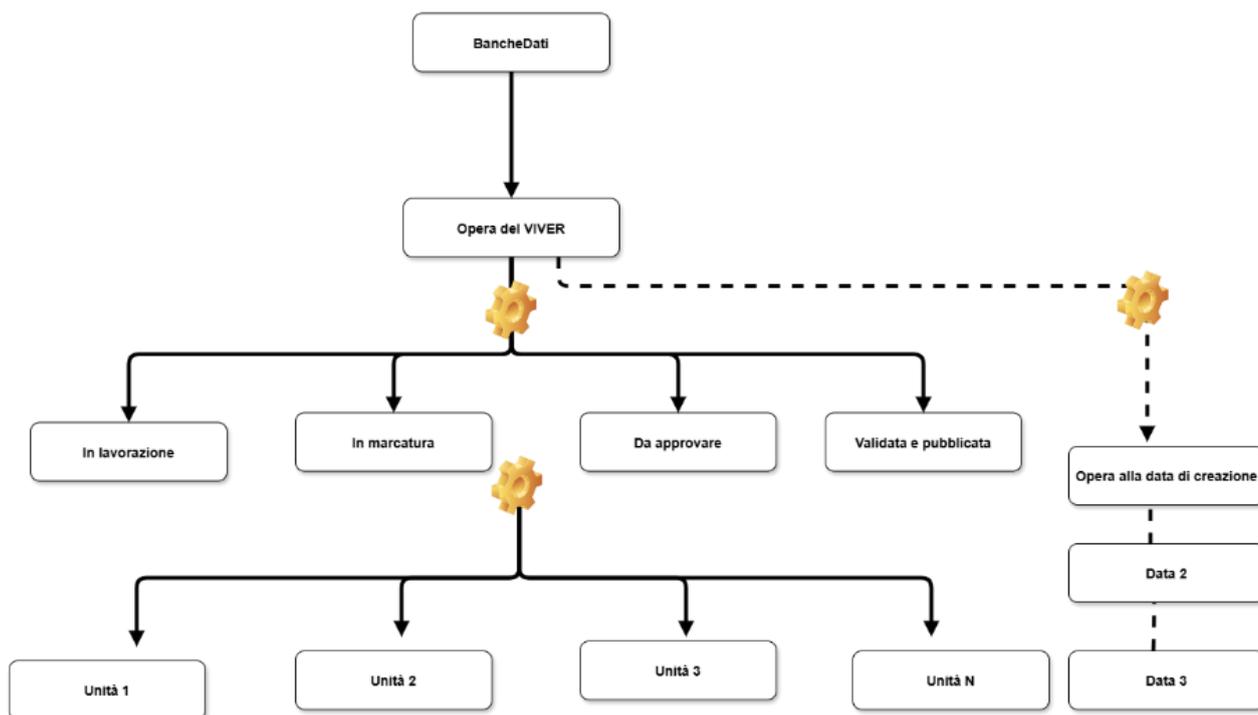
Nella sezione “Opere” è possibile gestire le opere che costituiscono il corpus del vocabolario, attraverso i dati e metadati descrittivi consueti. Le informazioni da

gestire sono eterogenee, a seconda della loro natura e delle finalità con le quali si è inteso utilizzarle negli strumenti che il progetto mette a disposizione dell'utente finale: liste con valori controllati, che dunque permettono di impostare filtri interessanti, oltreché decisivi, in termini di statistiche e di indagini sui materiali raccolti; valori numerici, che permettono di catalogare ciascuna opera entro limiti ben definiti e interrogabili; campi di testo ad inserimento libero di lunghezza contenuta, che, pur esaurendo l'informazione necessaria alla descrizione del singolo campo, talvolta impongono una condensazione informativa necessaria alla consistenza del dato in essere. Non mancano campi ad inserimento libero non vincolati da limiti di lunghezza, che permettono l'immissione di testo di approfondimento formattato a piacere.

La ripartizione tematica dei materiali ben si presta alla loro diversa accessibilità in termini di permessi assegnati ai ricercatori assegnati al loro trattamento: le varie unità partecipanti al progetto possono modificare solo le schede inerenti alle opere di loro competenza. In questo scenario il singolo operatore può visualizzare e modificare solo le opere che gli sono state assegnate; coloro che invece hanno il ruolo di revisori possono permettere l'avanzamento dello stato dei lavori cambiando lo stato delle singole schede, come si vede nell'immagine sottostante che mostra una parte della dashboard dell'ambiente di schedatura:



Nel seguente schema, si presenta il flusso di avanzamento nella validazione dei materiali e nell'attribuzione dei permessi, così come configurati. L'icona dell'ingranaggio è posta in corrispondenza delle tre tipologie di accesso (stato di avanzamento, gruppo di lavoro, dati storicizzati).



Ogni opera viene gestita in base allo stato di avanzamento e all'unità di appartenenza, tenendo conto anche dello storico della schedatura come ulteriore criterio di classificazione dei permessi. Questo approccio consente una prima catalogazione interna del materiale, rivolta principalmente agli operatori del *back-office* (come vedremo più avanti). Successivamente, questa classificazione si riflette in una categorizzazione più ampia e dettagliata, finalizzata a migliorare l'accessibilità e la fruizione dei contenuti da parte dell'utente finale nel *front-end*.

I testi sono stati marcati secondo gli standard XML/TEI adottando una marcatura TEI leggera, utilizzando i tag più rispondenti agli obiettivi ricordati nel paragrafo 2. La marcatura, infatti, non è funzionale alla realizzazione di edizioni digitali, ma innanzitutto alla costruzione del corpus testuale di base, interrogabile all'interno della piattaforma lessicografica, per la redazione delle voci.

La piattaforma consente anche di importare testi già marcati con altri editor XML, sui quali vengono eseguiti automaticamente controlli di correttezza formale e di adeguatezza degli standard di caratteri (eventualmente ricondotti ad omogeneità per favorire le procedure di tokenizzazione, indicizzazione e interrogazione). Ma la procedura naturale e ideale è invece quella dell'inserimento del testo non lavorato, acquisito e collazionato, che, dopo essere stato sottoposto al controllo dei caratteri e alle procedure di omogeneizzazione, è reso disponibile

per la marcatura all'interno della piattaforma, nel corretto livello della struttura, e disponendo di una serie di pulsanti personalizzati in funzione dei tag specifici previsti nello *Schema* del progetto e quindi in una modalità estremamente funzionale. Questo secondo approccio garantisce la massima omogeneità e organicità del testo, della sua marcatura, della sua architettura.

La sezione “Vocabolario”, contenente la gestione del lemmario, rappresenta un vero e proprio passo in avanti nella gestione e redazione di un vocabolario elettronico come quello del VIVer. Le faticose operazioni di spoglio sono effettuate direttamente all'interno della piattaforma di schedatura, consentendo così di marcare contestualmente eventuali interessanti interrelazioni che, in un lavoro di spoglio tradizionale, tendono a risultare invisibili. Per ogni singola entrata sono previsti tutti i campi della struttura individuata per scheda lessicografica (vedi il paragrafo 2.); oltre a questi, a rinsaldare la reticolarità del vocabolario, sono compresi i rinvii ad altro lemma e la bibliografia.

Di particolare rilievo è l'integrazione del modulo lessicografico con quello di gestione e marcatura del corpus (WCM Index+Search): infatti, dall'interno del cruscotto lessicografico è possibile lanciare delle ricerche libere e mirate sull'intero corpus testuale, andando a recuperare i contesti che possono essere agganciati automaticamente nell'apposito campo previsto nello schema della voce; tale operazione non solo consente un risparmio di tempo da parte del redattore, ma crea quella circolarità tra testi e voci del dizionario che costituisce un punto di forza del VIVer.

Visto che il VIVer e la sua banca dati sono uno strumento aperto, e quindi soggetto ad incremento variabile e non preventivabile, si è introdotto anche un campo in cui si indica la data dell'ultimo controllo (e indicizzazione) del corpus, che offre la possibilità di uno studio diacronico del materiale introdotto nella piattaforma, in modo da poter gestire e mantenere lo storico dei dati nel gestionale, e quindi recuperare scelte effettuate dallo schedatore in corso d'opera, intervenendo laddove nuovi contesti siano pertinenti per le varie schede (ad es: data di prima attestazione, ecc.).

Il modulo WCM CMS consente la gestione dei contenuti testuali e multimediali, da visualizzare sul sito, nel *front-end*. Per ogni contenuto è possibile definire la categoria di appartenenza (scelta che si ripercuote sul posizionamento tematico all'interno del menu di navigazione), il titolo della pagina, eventuali immagini ad essa associate, documenti PDF o altro materiale, il contenuto testuale vero e proprio

(che può essere gestito attraverso editor HTML in modo da personalizzarne la formattazione tipografica e ipertestuale). Il sistema, facendo parte della stessa piattaforma unitaria, si integra con i dati gestiti in WCM BancheDati. In linea di massima, abbiamo previsto di destinare l'ambiente di schedatura (WCM BancheDati) a studiosi che si occupino della correttezza scientifica del materiale e l'ambiente di allestimento dei materiali per il sito web (WCM CMS) ad una redazione che si occupi della comunicazione sul sito web e sui social network, e quindi più predisposta alle tecniche di scrittura per il web e di posizionamento sui motori di ricerca.

L'ultimo modulo della piattaforma è il WCM Repository, che fa della preservazione e salvaguardia del materiale digitale la sua missione principale. Viene consentito il caricamento di materiali digitali nei formati tra i più diffusi (immagini, testi, apparati, ecc.), il tutto con un'ampia schedatura che rispetta gli standard scientifici internazionali.

In tandem con l'archiviazione su server proprietari (il progetto VIVer è ospitato sui server dell'Accademia della Crusca) è prevista una archiviazione cloud: il sistema scelto in questo caso è Zenodo, il più importante archivio Open Access per la ricerca, gestito dal CERN e sostenuto dalla Comunità Europea, che non a caso deve il suo nome a Zenotodus, il primo bibliotecario della celeberrima Biblioteca di Alessandria. WCM Repository è inoltre integrato con la registrazione DOI attraverso CrossRef (vedi paragrafo 3.3), scelto per garantire il massimo della diffusione in ambito accademico e universitario.

Tutta l'infrastruttura di gestione dei contenuti, della loro presentazione e preservazione è stata pensata all'interno di una filosofia "Internet per Internet": materiale nativo digitale che, presentandosi in maniera responsiva, cioè adattabile automaticamente su diversi dispositivi, a differenti risoluzioni, viene reso accessibile liberamente in qualsiasi momento da qualunque parte del mondo, realizzato secondo standard internazionali e preservato perché possa durare nel tempo.

### 3.2 Il *front-end*: il portale di navigazione e interrogazione

Per la progettazione del portale di consultazione si è tenuto in considerazione che questo prevede varie sezioni in dialogo tra loro: banca dati testuale, sala di lettura, dizionario dinamico. Ciascuna di queste sezioni ha le proprie specifiche esigenze, come illustrato di seguito, restituite nella realizzazione del *front-end* in base alle seguenti linee guida e specifiche tecniche comuni:

- studio grafico dell'interfaccia secondo le linee guida della usabilità e accessibilità;
- realizzazione dei template HTML per la pubblicazione del sito in lingua italiana con codifica dei caratteri universali Unicode e con grafica responsiva e quindi fruibile dai principali dispositivi a varie risoluzioni;
- ottimizzazione strutturale per i motori di ricerca Organic SEO (Search Engine Optimization) e l'integrazione con sistemi di monitoraggio delle statistiche di accesso;
- presenza dei principali profili di metadati per la condivisione sui social network e per favorire l'indicizzazione nei principali aggregatori di ambito accademico.

### 3.2.1 La banca dati testuale

Il primo aspetto a cui abbiamo prestato particolarmente attenzione è stata l'architettura della banca dati, che è fondamentale per riuscire a ottenere le informazioni necessarie alla nostra ricerca.

La banca dati di VIVer è organizzata su quattro livelli: corpus generale; sottocorpus specifico; testo; unità testuale. La divisione in quattro livelli è fondamentale per poter gestire in modo adeguato i risultati delle ricerche.

Una specifica attenzione merita il secondo livello, quello dei sottocorpora, che abbiamo gestito in modo dinamico, assegnando a ciascun testo più classificatori (di genere, di tipologia testuale, ecc.) in modo da consentire all'utente di circoscrivere gruppi di testo specifici in base alle proprie esigenze. Il quarto livello è invece più funzionale a una buona gestione del testo per consentire alla procedura informatica di interrogazione di raffinare in modo opportuno la ricerca (potendo contare su macrocontesti e microcontesti funzionalmente definiti, incluso l'anno).

L'indicizzazione finale (quella che è effettuata sui testi ritenuti definitivi e che sono quindi accessibili al pubblico) avviene secondo cicli periodici stabiliti dagli amministratori della piattaforma, e mette a disposizione per la consultazione e interrogazione i testi pubblicati.

La piattaforma pubblica di interrogazione della banca dati testuale prevede un motore di ricerca per forma che consente l'impiego dei caratteri jolly, degli operatori booleani (anche a distanza fissata dall'utente), la ricerca espansa, e contempla anche il concorso di potenti strumenti di analisi statistica (in particolare in relazione alle co-occorrenze, per facilitare

l'individuazione di espressioni idiomatiche anche nascoste). Ma si sta valutando anche la possibilità di implementare una procedura di lemmatizzazione semiautomatica tarata sulla lingua del verismo italiano, che potrebbe essere un significativo valore aggiunto del nostro progetto. La ricerca di una parola fornisce l'accesso ai dati quantitativi a essa relativi e alle liste di concordanza, che prevedono contesti immediati (di lunghezza stabilita di volta in volta dal consultatore) e contesti allargati all'intera pagina con accesso al facsimile della stessa e la possibilità di sfogliare l'opera in avanti e indietro, in modo che la contestualizzazione della parola o del fenomeno ricercato sia potenzialmente aperta a qualunque misura si riveli necessaria. La lista dei risultati dei contesti immediati è dinamica e quindi ordinabile per rilevanza, titolo e anno. I risultati sono raggruppati per anno, per categorie e per opera, e da qui si possono aprire nuove ricerche trasversali o effettuare delle *subquery* grazie ai filtri disponibili (ad esempio: per autore, per categorie).

A una ricerca semplice, che permetterà la libera interrogazione del testo a prescindere dalla griglia di marcatura, si affiancherà una ricerca avanzata, attraverso la quale sarà possibile individuare colloquialismi, regionalismi ecc., ed effettuare ricerche mirate di forme all'interno delle porzioni di testo ricondotte a una certa categoria dallo specifico marcatore. A questo approccio, funzionale per chi sappia che cosa ricercare, si affiancheranno anche ricerche guidate che accompagnino il consultatore nelle specificità della banca dati; e ricerche guidate con specifici obiettivi didattici in modo che essa possa diventare anche un importante punto di partenza per le attività di insegnamento della lingua e della letteratura nella scuola secondaria e all'università.

### 3.2.2 La sala di lettura

Come sempre avviene per gli strumenti realizzati dal Centro informatico dell'Accademia della Crusca, alla banca dati testuale, che permette ricerche raffinate e approfondite, viene sempre affiancata anche una "Sala di lettura", che la trasforma di fatto in una biblioteca digitale, dai cui "scaffali" il lettore (sia esso un addetto ai lavori o appartenente al largo pubblico) potrà estrarre uno dei testi e leggerlo contando, oltre che sul testo elettronico, anche sulla riproduzione in facsimile dell'originale.

Nella sala di lettura sono previsti tre livelli per i vari testi (che corrispondono poi alle tre fasi di lavorazione): a) accesso al PDF (accompagnato da una scheda di introduzione); b) accesso alla versione elettronica collazionata e indicizzata (e quindi già disponibile anche per la modalità ricerca), con facsimile dell'originale; c) accesso alla versione marcata.

Un apposito database consentirà la ricerca per autore, titolo, data per poter così

raggiungere la scheda dell'opera desiderata, che sarà disponibile in uno dei tre formati previsti. La banca dati è aperta e dinamica e pertanto i testi che di volta in volta raggiungeranno livelli diversi di lavorazione “saliranno” alla categoria successiva (dal PDF alla versione indicizzata, dalla versione indicizzata a quella marcata).

L'impostazione modulare consente di mettere a disposizione dei consultatori i materiali anche se non hanno raggiunto una completa lavorazione. Ci è sembrata infatti una scelta preferibile quella di consentire l'accesso immediato al più alto numero possibile di testi, seppure a livelli di base o intermedi (comunque dichiarati), piuttosto che attendere lo stadio finale del completamento della marcatura dell'intero corpus. E la stessa filosofia è stata applicata anche agli strumenti di interrogazione.

### 3.2.3 Il dizionario dinamico

La banca dati testuale, descritta in precedenza e pubblicata sul portale come strumento autonomo a disposizione degli studiosi e più in generale di un pubblico largo, ha tuttavia come sua primaria funzione quella di fornire un corpus rappresentativo che serva di base per il dizionario. Di questo è stato elaborato il tracciato della voce, per il quale si rimanda al paragrafo 2 (in cui sono presentati anche alcuni prototipi di scheda lessicografica).

La struttura della voce è in linea con quella dei dizionari storici (come il TLIO), a vocazione storica (come il GDLI), o insieme sincronici e diacronici come il VoDIM<sup>36</sup>. Sono previsti quindi campi legati alla prima attestazione (con particolare attenzione all'ingresso in ambito verista e a eventuali specifiche accezioni), alla rete di corrispondenze, a indicazioni di tipo diatopico e diafasico, nonché agli ambiti d'uso. Sono poi individuati campi che rendano conto delle specificità del lessico verista, anche in chiave comparativa con le letterature affini delle altre lingue europee (individuando ad esempio *cluster* multilingui di parole corrispondenti allo stesso significato che potrebbero essere fornite come possibili traduenti ideali in contesto verista in un apposito campo del dizionario).

Il dizionario è stato poi pensato dinamico (qui il riferimento è di nuovo al VoDIM<sup>37</sup>), vale a dire pensato per garantire una consultazione di voci a struttura variabile. A partire dalla scheda completa che raccoglie tutti i campi previsti, saranno infatti messe a punto schede mirate in funzione del consultatore (ricercatore, curioso, studente, docente, consultatore straniero ecc.), e sarà prevista la possibilità di organizzare una voce personalizzata del dizionario scegliendo i campi di specifico interesse<sup>38</sup>.

Sono infine previsti collegamenti interni, reciproci, fra il dizionario e la banca dati testuale; e collegamenti esterni con corpora e dizionari disponibili in rete.

### 3.3 Interoperabilità e preservazione a lungo termine

Il VIVer si configura come un cantiere lessicografico digitale in cui le attività scientifiche hanno una forte integrazione con quelle informatiche. Nei progetti lessicografici digitali viene consigliato di utilizzare Identificatori persistenti e si è quindi deciso di utilizzare il DOI (Digital Object Identifier) come strumento per la diffusione e promozione, scegliendo Crossref come agenzia di registrazione<sup>39</sup>.

La prima operazione in tal senso prevede di assegnare e registrare un DOI per ogni edizione digitale delle nuove opere digitali del corpus prodotte in formato XML. Trattandosi a tutti gli effetti di nuove edizioni digitali, la registrazione del DOI consente di attestare e registrare la responsabilità di ciascun ricercatore, registrandone il ruolo nei metadati del DOI.

Ma la vera novità passa attraverso la assegnazione e registrazione di una serie di DOI anche per la banca dati del vocabolario, in corrispondenza dei vari livelli gerarchici con cui il database del VIVer è organizzato (database, collection, record, quest'ultimo corrispondente alla scheda del dizionario). Si prevede la compilazione di metadati di qualità e ricchi di informazioni, con l'obiettivo di favorire la diffusione del progetto e di massimizzarne l'impatto attraverso l'indicizzazione dei dati e metadati negli aggregatori accademici e nei motori di ricerca generalisti.

La creazione di metadati di qualità sarà un aspetto rilevante in tutto il progetto, sia nella definizione dei profili e dei tag da utilizzare, sia nella successiva cura da destinare alla attività di metadattazione oggetto per oggetto; i metadati, infatti, servono non soltanto per classificare e gestire gli oggetti digitali all'interno della collezione e repository, ma soprattutto nella prospettiva della interoperabilità con altri sistemi (vedi la costituenda Digital Library<sup>40</sup> prevista dal PNRR). In questo modo dati e metadati saranno distribuiti in Open Access sul portale e su altri aggregatori online.

Per descrivere questo aspetto decisamente innovativo, ci sembra utile approfondire e motivare le scelte effettuate; tenuto conto delle specificità del progetto, che oltre alla dimensione scientifica e lessicografica punta a un allargamento a un pubblico più ampio, nella scelta dei metadati si è optato per l'adozione di profili sia specialistici che generali. Innanzi tutto, abbiamo previsto l'utilizzo del Dublin Core<sup>41</sup>; oltre alla presenza del profilo CrossRef<sup>42</sup> (per la registrazione del DOI secondo il tracciato *dataset*) si è deciso anche per l'adozione di Highwire Press<sup>43</sup> per favorire l'indicizzazione in Google Scholar e negli aggregatori accademici. Infine, per favorire

la diffusione e la condivisione nelle reti social, si adottano i profili Open Graph<sup>44</sup> e Twitter:Cards<sup>45</sup>. Scelte simili sono già state effettuate in progetti analoghi, e hanno avuto un immediato e positivo riscontro nell'esito della indicizzazione su Google e sugli altri motori di ricerca, sia generalisti che accademici: già pochi giorni dopo il rilascio pubblico, le schede lessicografiche e le edizioni digitali sono state indicizzate e risultavano presenti nelle prime posizioni dei risultati per le parole chiave pertinenti.

Inoltre, saranno anche presenti procedure e metodi per la misurazione dell'impatto e dei risultati della diffusione, grazie ad un'interfaccia che permette in modo sistematico di monitorare gli accessi e il download dei materiali.

Tuttavia, per migliorare ulteriormente il posizionamento e mettere a fuoco le strategie digitali, occorre considerare che le metriche attualmente disponibili sui siti e sulle piattaforme social per misurare il coinvolgimento degli utenti non sono strumenti completamente efficaci. Metriche standard come "accessi", "visualizzazioni", "mi piace", "download", "condivisioni", ecc. sono certamente elementi da considerare, ma non tengono conto appieno della qualità di interazione aperta e partecipativa che il portale intende promuovere.

Per superare questo limite, abbiamo previsto di integrare le metriche standard con sondaggi e interviste, in modo particolare rivolte a target specifici di utenti, così da ottenere una comprensione più approfondita delle interazioni digitali.

Infine, per garantire la conservazione a lungo termine delle risorse digitali, utilizzando l'apposito modulo della piattaforma WCM, come indicato in precedenza, è stata predisposta l'archiviazione su Zenodo<sup>46</sup>.

### 3.4 **Impatto tecnologico e prospettive future**

Il progetto intende dare un decisivo contributo nella direzione della Open Science<sup>47</sup>, nella quale i documenti pubblicati sono associati a metadati che ne specificano il contesto informativo in un formato adatto all'interrogazione e all'interpretazione anche automatizzata.

Circa i servizi prestati per gli interessati, grazie alla presenza degli strumenti di interrogazione calibrati sui materiali lessicografici e sui testi letterari ma personalizzabili sulla base degli specifici interessi degli utenti stessi, saranno possibili ricerche testuali complesse e comparative, che permetteranno di costruire una rete personalizzabile di relazioni e connessioni tra entità, secondo criteri stabiliti dalle necessità dell'utente consultatore.

Il progetto prevede la possibilità che altre realtà istituzionali intervengano e mettano a disposizione le proprie risorse all'interno del repository, in modo che il portale VIVer raggiunga il suo scopo e non si esaurisca con l'acquisizione di materiali caricati a cura delle unità di ricerca partecipanti al VIVer.

Particolare attenzione sarà data alla possibilità di integrare, anche successivamente alla fine del progetto, i materiali presenti e le pubblicazioni attinenti alla ricerca attraverso la partnership con progetti analoghi ed Enti anche internazionali, collegati tramite portale.

Con il portale, gli studiosi avranno così l'opportunità di studiare i materiali da remoto e di avere a disposizione una selezione, in costante ampliamento, di testi e pubblicazioni strettamente legate all'ambito di ricerca sul verismo.

Al fine di promuovere la conoscenza e la valorizzazione del progetto a favore di un'utenza sempre più ampia e diversificata, il progetto segue quindi due direttive complementari:

- 1) un'alta formazione scientifica, con l'affinamento e l'acquisizione di risorse e strumenti rivolti a studiosi con elevate competenze lessicografico-letterarie, filologiche, linguistiche e archivistiche;
- 2) un accesso semplificato alle raccolte documentarie e una fruizione libera dei servizi di informazione e documentazione, rivolti anche a utenti non specialisti, al fine di migliorarne la conoscenza in ambito nazionale ed internazionale e di favorire la crescita culturale.

Inoltre, la creazione e la diffusione di risorse digitali progettate secondo i requisiti tipici di un archivio aperto istituzionale garantiscono un positivo impatto sulla società, contribuendo all'abbattimento delle barriere ambientali, architettoniche, culturali e sociali. Segnatamente, la pubblicazione di dati e di contenuti ad accesso libero rappresenta un significativo contributo al concetto dell'Open Science e alla realizzazione di una società più democratica e inclusiva, aperta alla condivisione di saperi, conoscenze e tradizioni interdisciplinari, internazionali e pluridentitarie.

La parziale pubblicazione di alcune collezioni come Open Data (unita all'adozione di adeguate *open data policies* ispirate ai principi FAIR<sup>48</sup>) fornirà libero accesso e libero utilizzo dei dati messi a disposizione con Licenza aperta (dunque senza limitazioni eccetto quelle costituite da diritto d'autore, se necessarie) e costituirà un primo passo per aumentare la visibilità delle collezioni stesse e per incoraggiarne l'utilizzo nel campo della ricerca da parte di una più ampia fascia di utenza.

#### 4. Conclusioni e prospettive

Il VIVer può in definitiva considerarsi un'impresa di lessicografia digitale 'ibrida' tra il vocabolario elettronico e l'archivio informatizzato: il patrimonio testuale dei veristi e di autori e autrici di Otto-Novecento, accomunati dalla ricerca del realismo linguistico e dalle pratiche scrittorie nei vari generi e sottogeneri sopra indicati che ne costituisce la base, rappresenterà di per sé un importante e innovativo corpus. Con adeguato metamotore, inoltre, si può arrivare a un'interrogazione unificata tra questo macrotesto e altri corpora già costituiti o in avanzata elaborazione<sup>49</sup>, sulla falsariga seguita dall'OVI per il TLIO (Tesoro della Lingua Italiana delle Origini) e dalla Crusca per le fonti testuali che costituiscono la banca linguistica della Stazione lessicografica dell'Accademia.

Dal punto di vista informatico, il progetto del VIVer intende assumere un ruolo di riferimento e di guida per le altre istituzioni nella diffusione di nuove metodologie e procedure per la valorizzazione e la comunicazione digitale del patrimonio lessicografico e letterario italiano, allo scopo da un lato di favorire la collaborazione e lo scambio di competenze scientifiche, dall'altro di definire e incentivare l'acquisizione di buone norme e pratiche condivise. Queste si realizzeranno non solo tramite la possibilità di accesso a fonti e risorse tramite una piattaforma integrata a libero accesso, ma anche attraverso nuovi servizi informatici rivolti al coinvolgimento diretto dell'utenza.

Per quanto riguarda poi l'aspetto strettamente linguistico, il VIVer consentirà un'interoperatività tra i vari sottocorpora che lo compongono, individuando per lessico e fraseologia concomitanze tra autori di varia provenienza regionale, nonché retrodatazioni di termini ed espressioni idiomatiche. In sostanza, la dimensione dell'interregionalità, già accertata per la lingua di Verga, potrebbe estendersi a tutti o a gran parte dei veristi.

In prospettiva storico-linguistica poi il VIVer possiede la potenzialità radicalmente innovativa di contribuire a una descrizione esauriente e ravvicinata del lessico e della fraseologia di una varietà tuttora poco conosciuta dell'italiano postunitario: il parlato-scritto, inteso sia come parlato scritto-narrato che come scritto extraletterario ma colloquiale. Su quest'ultimo fronte la diffusione di un italiano sciolto e pragmatico, legato a una dimensione comunicativa media indotta dalla divulgazione tecnico-scientifica ed educativa, troverà finalmente uno sbocco descrittivo statisticamente attendibile.

Sul fronte più spiccatamente letterario, infine, la circolazione di moduli linguistici e stilistici rilevabile dal confronto organico consentito dal VIVer permetterà di cogliere

rapporti reticolari tra le scelte espressive di narratori e drammaturghi. Su un orizzonte più vasto, poi, la registrazione lessicografica di unità lessicali, fraseologiche e idiomatiche rivelerà fermenti e modalità di realizzazione delle istanze naturaliste, non in un'analisi fine a sé stessa, ma in una rilettura comparativa estesa trasversalmente alla Francia e alle principali letterature europee, nella prospettiva globale del cosiddetto *Naturalisme monde*.

Proiettando il VIVer in una visione transnazionale si prospetta così la storicizzazione critica e filologica, finora praticabile solo a macchia di leopardo, dello stesso *Naturalisme monde*, la cui testualità sarebbe sottoponibile a una lettura sinottica e trasversale. Ne potrebbe risultare una costellazione semantico-culturale, un lessico emozionale e socioculturale realista, con ricadute modellizzanti sulla letteratura di consumo.

Da un punto di vista più generalmente stilistico, la lingua mutante del Realismo, interferita ma non prevaricata dallo stile popolare, lasciandosi alle spalle la lettura univoca e stereotipata del rapporto lingua-dialetto, si riqualificherebbe come strumento espressivo moderno in chiave nazionale.

GABRIELLA ALFIERI

MARCO BIFFI

STEPHANIE CERRUTO

GIOVANNI SALUCCI

## Note

<sup>1</sup> Il testo è stato condiviso da tutti gli autori. Le attribuzioni vanno così ripartite: Gabriella Alfieri parr. 1, 4; Stephanie Cerruto par. 2; Marco Biffi e Giovanni Salucci par. 3 (in particolare: Giovanni Salucci parr. 3.1, 3.3 e 3.4; Marco Biffi par. 3.2).

<sup>2</sup> Per una presentazione dettagliata del progetto si rinvia a G. Alfieri-S. Arcidiacono-M. Biffi-S. Cerruto-A. Di Silvestro-V. Puglisi-R. Sardo, *Il VIVer: vocabolario reticolare dell'italiano veristico*, in *Annali della Fondazione Verga*, XIV, 2021 [ma 2022], pp. 343-402; e per gli aspetti informatici a G. Alfieri-M. Biffi-S. Cerruto-G. Salucci, *Il VIVer (Vocabolario dell'Italiano Verista)*, in *Me.Te. Digitali. Mediterraneo in rete tra testi e contesti*, Proceedings del XIII Convegno Annuale AIUCD, Catania 28-30 maggio 2024, a cura di A. Di Silvestro-D. Spampinato, Università di Catania 2024, pp. 275-280 (<https://amsacta.unibo.it/id/eprint/7927/1/AIUCD2024-proceedings.pdf>).

<sup>3</sup> Cfr. AA.VV., *Les mondes naturalistes, II. Naturalismes du monde, les voix de l'étranger*, in *Les Cahiers Naturalistes*, XCIV, 2020, pp. 175-440.

<sup>4</sup> Il progetto, di cui sono stata P.I. fino al novembre 2024, è attualmente guidato da Fabio Rossi e coinvolge le Università di Catania, Messina e Milano, coordinate rispettivamente da Daria Motta, Fabio Rossi e Giuseppe Sergio. Il corpus è costituito da romanzi, drammi, galatei, libri per l'infanzia, scritti giornalistici, fumetti, lettere e si sta sottoponendo a marcatura e analisi linguistico-testuale e pragmatica.

<sup>5</sup> Dal 2023 al nucleo originario formato da Fondazione Verga, Università di Catania e Accademia della Crusca si sono aggiunte numerose unità di ricerca dislocate in atenei delle diverse regioni rappresentative del corpus: Università del Piemonte orientale, Università Statale di Milano, Università di Verona, Università di Genova, Università di Firenze, Università per Stranieri di Siena, Terza Università di Roma, LUMSA, Università di Chieti-Pescara, Università di Napoli Federico II, Università di Napoli L'Orientale, Università di Bari, Università di Messina, Università di Cagliari, Università di Sassari. Nel 2024 si è aggiunta anche la Corsica.

<sup>6</sup> G. Alfieri, *Non solo vocabolario: "mezzi" e "provvedimenti" "fattibili" nella proposta manzoniana*, in *Storia della lingua italiana e storia dell'Italia unita. L'italiano e lo Stato nazionale*, Atti del IX Convegno ASLI (Firenze, 2-4 Dicembre 2010), a cura di A. Nesi-S. Morgana-N. Maraschio, Firenze, Cesati 2011, pp. 53-85.

<sup>7</sup> Basti per tutti il nome di Antonio Zampolli. Cfr. anche M. Biffi, *La lessicografia della Crusca in rete*, in AA.VV., *Una lingua e il suo vocabolario*, Firenze, Accademia della Crusca 2014, pp. 113-127.

<sup>8</sup> D'A.S. Avalle, *Al servizio del vocabolario della lingua italiana*, Firenze, Accademia della Crusca 1979.

<sup>9</sup> M. Biffi, *Progettare il corpus per il vocabolario postunitario*, in *L'italiano elettronico. Vocabolari, corpora, archivi testuali e sonori*, Atti della "Piazza delle Lingue" dell'Accademia della Crusca (Firenze, 6-8 novembre 2014), a cura di C. Marazzini-L. Maconi, Firenze, Accademia della Crusca 2016, pp. 259-280.

<sup>10</sup> Vd. AA.VV., *I verismi regionali*, Atti del Congresso Internazionale di Studi (Catania, 27-29 aprile 1992), Biblioteca della Fondazione Verga, 2 voll., Catania, Fondazione Verga 1996.

<sup>11</sup> Cfr. G. Alfieri, *Lingua e letteratura nei "Verismi": un intreccio o un intralcio?*, in AA.VV., *I verismi regionali*, cit., vol. II, pp. 815-823.

<sup>12</sup> Si pensi ad autori e autrici che scrivevano di questioni sociali del tempo o ad autori e autrici a metà fra realismo e verismo, come ad esempio Farina, Valera, Neera.

<sup>13</sup> Ad esempio gli scritti di Jessie White Mario e in generale testi giornalistici e inchieste sociali.

<sup>14</sup> *Est* sta per "estetica", in quanto il genere testuale include le prefazioni degli autori alle raccolte di novelle o romanzi, le recensioni di critici letterari, ecc.

<sup>15</sup> *Fia* sta per "fiabe" e, oltre a quelli di Capuana, il corpus include testi di altri autori, tra cui Maria Messina.

<sup>16</sup> Cfr. S. Cerruto, *«La competenza multipla» di Giovanni Verga tra scrittura epistolare e letteraria*, Fondazione Verga-Euno, Catania-Leonforte 2023.

<sup>17</sup> G. Alfieri, *Una poetica disseminata: idee di Verga sul realismo*, in *Verga nel realismo europeo ed extraeuropeo*, Atti del Convegno Internazionale di Studi (Catania, 5-7 dicembre 2022; 20-22 aprile 2023), a cura di G. Alfieri-G. Longo-A. Mangano, 2 voll., Euno, Leonforte 2023, vol. I, pp. 69-102.

<sup>18</sup> Un'ampia fase iniziale del progetto VIVer è stata dedicata all'elaborazione della griglia di analisi e degli strumenti informatici da parte del gruppo di lavoro composto da Gabriella Alfieri, Marco Biffi, Antonio Di Silvestro, Stephanie Cerruto, Giovanni Salucci, Rosaria Sardo.

<sup>19</sup> AA.VV., *I verismi regionali*, cit.

<sup>20</sup> G. Nencioni, *Antropologia poetica?*, in *Strumenti critici*, XIX, 1972, pp. 243-58, poi in Id., *Fra grammatica e retorica. Da Dante a Pirandello*, Einaudi, Torino 1983, pp. 161-75.

<sup>21</sup> Calcando il titolo della bella monografia di Enrico Testa (*L'italiano nascosto*, Einaudi, Torino 2014), si è pensato di definire come "dialetto nascosto" la categoria che include voci dialettali che trovano una marcata corrispondenza nell'italiano comune e letterario, come ad esempio *incignare* in Pirandello (cfr. A. Stussi, *Lingua, dialetto e letteratura*, Einaudi, Torino 1993).

<sup>22</sup> La categoria è stata adottata poi anche nella griglia del citato progetto ARDIPS e si riferisce a voci toscane diffuse su scala nazionale in seguito alla riforma manzoniana, poi uscite dall'uso, quali *desinare*, *uscio*, *cencio*, *seggiola*, ecc.

<sup>23</sup> Nel tracciato della voce del vocabolario si segnalerà l'eventuale origine regionale o dialettale.

<sup>24</sup> Ad esempio: *analisi, osservazione, quadro, racconto, realismo, studio, verità, colore locale, esser sinceri, mezze tinte, riproduzione del vero, forma inerente al soggetto*, ecc.

<sup>25</sup> <https://www.archidata.info>.

<sup>26</sup> <https://www.proverbi-italiani.org/index.asp>.

<sup>27</sup> Il progetto è finanziato con fondi MUR PRIN 2022 PNRR.

<sup>28</sup> <https://www.stazionelessicografica.it>

<sup>29</sup> La lista di opzioni nel campo *corrispondenze lessicografiche* riguarderà soprattutto le fonti lessicografiche più comuni, come ad esempio GDLI (S. Battaglia-G. Barberi Squarotti, *Grande dizionario della lingua italiana*, 21 voll. più 2 supplementi, UTET, Torino 1961-2009), TB (N. Tommaseo-B. Bellini, *Dizionario della lingua italiana*, Unione Tipografico-Editrice, Torino 1865-1879), RF (G. Rigutini-P. Fanfani, *Vocabolario italiano della lingua parlata*, Cennini, Firenze 1875), ecc.

<sup>30</sup> PI = Proverbi italiani (<https://www.proverbi-italiani.org/index.asp>).

<sup>31</sup> G. Verga, *I Malavoglia*, a cura di F. Cecco, Edizione Nazionale delle Opere di Giovanni Verga, Fondazione Verga – Interlinea, Novara 2014, pp. 33, 58, 218.

<sup>32</sup> Ivi, p. 19.

<sup>33</sup> Lettera di G. Verga a E. Rod, [Milano, ca. 8 dicembre 1881], in *Carteggio Verga-Rod*, a cura di G. Longo, Fondazione Verga, Catania 2004, p. 103.

<sup>34</sup> La piattaforma è stata configurata da Giovanni Salucci secondo le specifiche definite all'interno del Centro, diretto da Marco Biffi.

<sup>35</sup> <https://www.wcm.it/wcm-digital-humanities/>

<sup>36</sup> Cfr. M. Biffi, *Progettare il corpus per il vocabolario postunitario*, cit., pp. 269-269.

<sup>37</sup> Sulla natura dinamica del VoDIM si vedano M. Biffi, *Progettare il corpus per il vocabolario postunitario*, cit.; Id., *Strumenti informatico-linguistici per la realizzazione di un dizionario dell'italiano post-unitario*, cit.; P. Bertini Malgarini-M. Biffi-U. Vignuzzi, *Dal «Vocabolario storico della cucina italiana postunitaria» («VoSCIP») al «Vocabolario dinamico dell'italiano moderno» («VoDIM»): riflessioni di metodo e prototipi*, in *Studi di lessicografia italiana*, XXXVI, 2019, pp. 341-366; M. Biffi-A. Ferrari, *Progettare e ideare un corpus dell'italiano nella rete: il caso del CoLIWeb*, in *Studi di Lessicografia Italiana*, XXXVII, 2020, pp. 357-374.

<sup>38</sup> Si vedano come esempi i prototipi di voci dinamiche preparate per il VoDIM per quanto riguarda la cucina (cfr. P. Bertini Malgarini-M. Biffi-U. Vignuzzi, *Dal «Vocabolario storico della cucina italiana postunitaria»* cit., pp. 361-364) e parte (cfr. B. Patella, *Il Vocabolario dinamico dell'italiano moderno (VoDIM): proposta di schede lessicografiche per la lingua dell'arte*, in *Italiano digitale*, XIII, fasc. 2, 2000, pp. 122-170).

<sup>39</sup> G. Salucci, *Utilizzo del DOI (Digital Object Identifier) per la diffusione di progetti lessicografici digitali*, in *DILEF. Rivista digitale del Dipartimento di Lettere e Filosofia*, III, 2023, pp. 275-292, 10.35948/DILEF/2024.4327.

<sup>40</sup> <https://digitallibrary.cultura.gov.it/>.

<sup>41</sup> <https://www.dublincore.org/specifications/dublin-core/dces/>.

<sup>42</sup> <https://www.crossref.org/documentation/schema-library/metadata-deposit-schema-5-3-1/>.

<sup>43</sup> <https://scholar.google.com.au/intl/en/scholar/inclusion.html>.

<sup>44</sup> <https://ogp.me/>.

<sup>45</sup> <https://developer.twitter.com/en/docs/twitter-for-websites/cards/overview/markup>.

<sup>46</sup> <https://zenodo.org/>.

<sup>47</sup> Cfr. UNESCO, *UNESCO Recommendation on Open Science* (<https://unesdoc.unesco.org/ark:/48223/pf0000379949>).

<sup>48</sup> M.D. Wilkinson et al., *The FAIR Guiding Principles for scientific data management and stewardship*, in *Nature-Scientific Data*, III, 2016 (<https://doi.org/10.1038/sdata.2016.18>).

<sup>49</sup> All'interno del DISUM (Dipartimento di Scienze Umanistiche) dell'Università di Catania sono in corso importanti progetti di informatica umanistica: *Verismo digitale* (responsabile scientifico Marina Paino, componenti: Andrea Manganaro, Antonio Sicherà), svolto nell'ambito del progetto PNRR CHANGES SPOKE 3, e finalizzato alla creazione di un ecosistema digitale intorno alle opere di Verga, De Roberto, Capuana, di cui si realizzano edizioni digitali sinottiche, hyperedizioni, e concordanze; COVerLESS (Corpus on line del Verismo tra Letteratura, Storia e Società) coordinato da Antonio Di Silvestro, e il VERBUM a questo correlato, un vocabolario del lessico standard della poetica verista, nonché la Timeline Ver-intime, interrogabile trasversalmente con altre biblioteche digitali sul Verismo.