



AIUCD 2022 | UNIVERSITÀ DEL SALENTO

CULTURE DIGITALI

INTERSEZIONI

FILOSOFIA

ARTI

MEDIA

TESTO

ARTI

FILOSOFIA

CONTENUTI

INTELLIGENZA

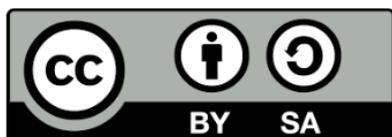
PROCEEDINGS

ISBN 9788894253566

Copyright ©2022 AIUCD
Associazione per l'Informatica Umanistica e la Cultura Digitale



Il presente volume e tutti i contributi sono rilasciati sotto licenza Creative Commons Attribution Share-Alike 4.0 International license ([CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)). Ogni altro diritto rimane in capo ai singoli autori.



This volume and all contributions are released under the Creative Commons Attribution Share-Alike 4.0 International license ([CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)). All other rights retained by the legal owners.

Fabio Ciraci, Giulia Miglietta, Carola Gatto (edd.), AIUCD 2022 - Culture digitali. Intersezioni: filosofia, arti, media. Proceedings della 11^a conferenza nazionale, Lecce, 2022. Fabio Ciraci, Giulia Miglietta, Carola Gatto (edd.), AIUCD 2022 - Digital cultures. Intersections: philosophy, arts, media. Proceedings of the 11th national conference, Lecce, 2022.

Salvo diversa indicazione, ogni link citato era attivo al 21 gennaio 2022. All links have been visited on 21th January 2022, unless otherwise indicated

Si prega di notificare all'editore ogni omissione o errore si riscontri, al fine di provvedere alla rettifica. Please notify the publisher of any omissions or errors found, in order to rectify them. [aiucd.segreteria \[at\] aiucd.org](mailto:aiucd.segreteria@aiucd.org)

I contributi pubblicati nel presente volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante *double-blind peer review* sotto la responsabilità del Comitato Scientifico di AIUCD 2022.

All the paper published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review process under the responsibility of the AIUCD 2022 Scientific Committee.

Il programma della conferenza AIUCD 2022 è disponibile online all'indirizzo/ The AIUCD 2022 conference program is available online all'apposito indirizzo

<http://aiucd2022.unisalento.it> <http://conference.unisalento.it/ocs/index.php/aiucd2022/index/pages/view/programma>

Comitato Scientifico:

General Chair: Fabio Ciraci (Università del Salento)

Local Chair: Mario Bochicchio (Università del Salento, Università di Bari)

Membri Comitato Scientifico: Marina Buzzoni (Presidentessa AIUCD, Uni. Venezia), Federico Boschetti (Ric. ILC-CNR); Federico Meschini (Uni. Tuscia); Roberto Rosselli Del Turco (Uni Torino); Rachele Sprugnoli (Ass. Ric. Univ. Cattolica); Donato Malerba (Università Bari);

Luca Bandirali, Daniela Castaldo, Francesco Ceraolo, Stefano Cristante, Domenico M. Fazio, Manolita Francesca, Marco Mancarella, Pietro Luigi Iaia, Massimiliano Rossi, Grazia Semeraro, Franco Tommasi, Luigi Patrono (Università del Salento)

Membri del Comitato di programma: Mario Bochicchio (Local Chair), Luca Bandirali, Daniela Castaldo, Marco Mancarella, Pietro Luigi Iaia, Federica Epifani (Responsabile Comitato di Programma), Ilenia Colonna, Patrizia Miggiano; Carola Gatto; Giulia Miglietta; Marco Giannotta; Alessia De Blasi, Isabella Hernandez.

Direttori di Area: Luca Bandirali; Mario Bochicchio; Fabio Ciraci; Roberto Rosselli Del Turco; Marco Mancarella; Grazia Semeraro.

Segreteria del Convegno: Dott.ssa Silvia Gravili

Resp. tecnico: Carlo Tafuro; web design: Dr.ssa Paola D'Amico; comunicazione: Dr.ssa Loredana De Vitis

Enti organizzatori / Organizing institutions:

AIUCD;

Università del Salento: Centro interdipartimentale in Digital Humanities in collaborazione con i corsi di laurea in Filosofia, DAMS, Beni Culturali e Digital Humanities; ISUFI, Scuola Placetelling.

Università degli Studi Aldo Moro, Dipartimento di Informatica

Sponsor

Regione Puglia; Provincia di Lecce; Città di Lecce; CINI – Consorzio Universitario Nazionale per l'Informatica; SFI-Società Filosofica Italiana; AFC - Apulia Film Commission, Teatro Pubblico Pugliese; Argo Software.

Lista dei revisori - List of the reviewers

Agnese Addone; Tommaso Agnoloni; Luca Bandirali; Nicola Barbuti; Andrea Bellandi; Armando Bisogno; Mario Alessandro Bochicchio; Andrea Bolioli; Federico Boschetti; Dominique Brunato; Paolo Buono; Dino Buzzetti; Marina Buzzoni; Luigi Catalani; Francesco Ceraolo; Daniele Chiffi; Simona Chiodo; Fabio Ciotti; Ilenia Colonna; Christian D'Agata; Elisa D'Argenio; Riccardo De Biase; Manuela De Giorgi; Daniela De Leo; Salvatore De Masi; Pierpaolo Del Coco; Angelo Mario Del Grosso; Francesca Di Donato; Giorgio Maria Di Nunzio; Federica Epifani; Daniela Fogli; Claudio Forziati; Greta Franzini; Francesca Frontini; Emiliano Giovannetti; Edmondo Grassi; Fabiana Guernaccini; Barbara Guidi; Pietro Luigi Iaia; Benedetta Iavarone; Fahad Khan; Maurizio Lana; Angelica Lo Duca; Donato Malerba; Marco Mancarella; Tiziana Mancinelli; Chiara Mannari; Valentina Marangi; Cristina Marras; Federico Meschini; Patrizia Miggiano; Giulia Miglietta; Paolo Monella; Giovanni Morrone; Serge Noiret; Deborah Paci; Antonio Pascucci; Enrico Pasini; Luigi Patrono; Igor Pizzirusso; Simone Rebora; Massimiliano Rossi; Daniela Rotelli; Enrica Salvatori; Eva Sassolini; Daria Spampinato; Rachele Sprugnoli; Enrico Terrone; Francesca Tomasi; Francesco Tommasi; Sara Tonelli; Gennaro Vessio; Marco Salvatore Zappatore.

Indice – Table of Contents

Prefazione	I
Sessione Arti 1 – Artemisia Gentileschi	2
La Comédie Virtuelle	4
Climate change & digital cultural impact, the Victoria & Albert Museum	9
La Digitalizzazione per una fruizione del Patrimonio Culturale in sito e da remoto: il caso studio della Pala Gozzi di Tiziano	12
Sessione Testi 1 – Claude Shannon	18
Verso la definizione di criteri per valutare soluzioni di scholarly editing digitale: il caso d'uso GreekSchools	20
HYLAS: A new metrical search tool for Greek and Latin poetry	26
Stylometry and Reader Response. An Experiment with <i>Harry Potter</i> Fanfiction	30
Sessione Intelligenza 1 – Alan M. Turing	35
Analisi e valorizzazione del patrimonio artistico mediante Intelligenza Artificiale	37
Un Oggetto Intelligente IoT per Migliorare le Visite Interattive di Siti di Interesse Culturale	42
Oxoce - Motore di ricerca tematico strutturato	46
Sessione Contenuti 1 – George Boole	49
Funzione ecosistemica e funzione storiografica della narrazione ambientale videoludica	51
Narrazioni mediatiche delle emergenze e processi di costruzione di <i>quest</i> : quali possibili analogie?	
L'incidente del "corrupted blood" in "World of Warcraft"	54
Narrazione e interazione	59
Sessione Testi 2 – Ada Lovelace	61
Web e social media come nuove fonti per la storia	63
Idee, persone, <i>realia</i> : un ambiente digitale per la Via della Seta	68
Visualizzazione del cambiamento d'uso del maschile e femminile nei titoli occupazionali	71
GenderedOntoComedy: Toward a Gendered Representation of Literary Characters in the Dante's Commedia	76
Sessione Filosofia 1 – Marisa Bellisario	81
Gli indici della prima modernità come strumento storiografico: questioni preliminari metodologiche e pratiche	83
Indici e mappe digitali per l'iter italicum di G. W. Leibniz	86
Ermeneutica digitale del testo filosofico. Problemi e opportunità	91
Human Enhancement e soggetto Post-Umano alla prova delle DH: come le tecnologie digitali ci trasformano	93

Sessione Testi 3 – Grace Murray Hopper	96
Conservazione e fruizione di banche dati letterarie: l'archivio della poesia italiana dell'Otto/Novecento di Giuseppe Savoca	98
«Le varianti della rosa». Per un prototipo di edizione digitale del <i>Nome della rosa</i> : interpretazione, didattica, annotazione	105
Online lexical resources for translators: where do we stand? A (possibly meaningful) case-study	111
 Sessione Filosofia 3 – Gilbert Simondon	 116
Governare le piattaforme. Cinque proposte su pluralismo e polarizzazione online	118
A Taxonomy of Depictive Representations: From Paintings and Sculptures to Virtual Reality	122
Paesaggi dell'incontro mediale on-demand	126
 Sessione Contenuti 2 – Marshall McLuhan	 129
Tra Public e Digital History: la soluzione ibrida dei registri parrocchiali di Monterosso on line	131
Una nuova mappatura digitale per i borghi delle aree interne	138
Intelligenza artificiale e archivi audiovisivi: potenzialità e sfide del progetto "PH-Remix"	141
 Sessione Intelligenza 2 – John von Neumann	 145
Un nuovo approccio per la descrizione e gestione del patrimonio culturale digitale relativo a MAB	147
Sulla funzionalità di un'ontologia della filosofia alto medievale. Il caso dei «Moralia in lob» di Gregorio Magno	151
La Visualizzazione Grafica di Sensi e Relazioni Semantiche di un Lessico Computazionale della Lingua Italiana	155
 Sessione Testi 4 – Hedy Lamarr	 161
Dalla codifica alla fruizione: l'edizione digitale Bellini Digital Correspondence	163
Dante e Petrarca allo (stesso) scrittoio. Per lo sviluppo di un'ontologia di IDP a partire dall'istanza manoscritti di Itinera	169
Il progetto 'epistolarITA' e una proposta di applicazione di algoritmi di prossimità testuale su documenti epistolari italiani (XV-XVII s.)	172
 Sessione Testi 5 – Hélèn Metzger	 177
Visualizing the genetic process of literary works	179
Analisi linguistica e pseudonimizzazione: strumenti e paradigmi	185
RePIM in LOD: semantic technologies to preserve knowledge about Italian secular music and lyric poetry from the 16th-17th centuries	193
 Sessione Filosofia 2 – Giulio Cesare Vanini	 196
Computare o comporre? Riflessioni sul rapporto tra poesia e digitalità alla luce di alcune considerazioni bachelardiane	198
Schemi, ipotesi e algoritmi. Approcci kantiani alla filosofia delle tecnologie digitali	203
Tra chair e empiriquement lo spazio topologico: contributo merleau-pontyano ai sistemi informatici	207

Sessione Testi 6 – Katherine Johnson	210
There and back again: what to expect in the next EVT version	212
XML-TEI: Un modello per la filologia d'autore	218
La svolta empirico-computazionale negli studi culturali e letterari: una nuova scienza della cultura	223
Poster	227
Wordforms and Meanings: un Updated Report on the LiLa Project	229
From Close to Distant Reading. Towards the Computational Analysis of “Liber Abbaci”	232
Citizen Humanities in Tyrol: a case study on historical newspapers	236
Un esperimento di visualizzazione grafica della terminologia del Talmud babilonese	239
Una edizione critica digitale per la cristianistica dell'antichità	242
Ritmi postumani: produzione poetica e machine learning	243
Argument-Checking: a critical Pedagogy Approach to Digital Literacy	245
“Nostra Signora Experience”: il Placetelling® in Ambiente Digitale	249

Prefazione

L'undicesima edizione del Convegno Nazionale dell'AIUCD-Associazione di Informatica Umanistica ha per titolo *Culture digitali. Intersezioni: filosofia, arti, media*. Nel titolo è presente, in maniera esplicita, la richiesta di una riflessione, metodologica e teorica, sull'interrelazione tra tecnologie digitali, scienze dell'informazione, discipline filosofiche, mondo delle arti e *cultural studies*. Per questo motivo, il Comitato Scientifico ha individuato cinque aree funzionali alla *call for paper*, in base alle quali selezionare i contributi da presentare in occasione del convegno nazionale. Tutte le aree sono connotate da un "+D" di digitale. Tale espressione non sta a indicare un addendo esornativo e accidentale, ammiccante e modaiolo, né un supplemento alle varie discipline umanistiche con funzione integrativa o sussidiaria; essa denota invece una contaminazione, profonda e trasformativa, delle discipline umanistiche con il digitale, intendendo quest'ultimo in senso ampio, come espressione di una trasformazione scientifica e tecnologica che investe e muta la cultura e la società. Alla luce dell'informatizzazione delle conoscenze e della digitalizzazione delle pratiche, che ridiscutono limiti e poteri delle discipline istituzionali, si tratta di comprendere il nuovo ruolo delle *humanities*. Si tratta di trasformazioni che pongono problematiche inedite, ma al contempo ampliano le possibilità di indagine nei campi della tradizionale ricerca umanistica. Fedeli alla massima di Terenzio – *homo sum humani nihil a me alienum puto* – siamo convinti che una tale contaminazione individui nell'umanista un interlocutore privilegiato. Siamo cioè dell'idea che i saperi si costruiscono reciprocamente, con mutua dipendenza e in maniera interrelata, travalicando i settori scientifici e le camicie di forza delle definizioni settoriali. In questo senso il *digital humanist* rappresenta una figura capace di un supplemento di conoscenza e di una visione interdisciplinare, è abilitato a una ricerca di confine spesso difficile da caratterizzare, sia in relazione agli aspetti più squisitamente teorici dell'informatizzazione, sia in riferimento agli effetti pratici e al loro portato sociale e culturale. A questa trasformazione partecipano a pieno titolo la filosofia e le arti, come discipline chiamate a riflettere sul digitale, non solo perché da sempre si interrogano sull'uomo e sul mondo, ma anche perché ambiscono a migliorare la realtà e governare il cambiamento.

Con l'intenzione, quindi, di coinvolgere la nostra comunità a riflettere sull'intersezione fra i saperi, nell'ottica di una pluralità di culture, il Comitato Scientifico ha individuato le seguenti aree di interesse: "Testo +D", che tesauroizza ed estende la tradizione di ricerca dell'AIUCD, rivolgendosi agli studi di linguistica computazionale, edizioni digitali, progetti ipertestuali, filologia ed ecdotica digitali; "Arti +D", relativa alle tecnologie digitali per il mondo dell'arte, *digital* e *cultural heritage*; "Filosofia +D", riguardante la filosofia dell'informazione, etica ed epistemologia del digitale; "Contenuti +D", con un focus su realtà virtuale e aumentata, contenuti multimediali e transmediali, ecosistemi narrativi e spazio dei media; "Intelligenza +D", orientata alla comunicazione mediata dal computer, apprendimento digitale e sistemi di traduzione automatizzata.

Per sviluppare al meglio le aree tematiche proposte per il convegno, nei mesi di ottobre e novembre 2021 il *Centro di ricerca in Digital Humanities* dell'Università del Salento, in collaborazione con l'AIUCD, ha organizzato il ciclo di seminari "Loading AIUCD2022", a cura di Fabio Ciraci e di Patrizia Miggiano, con sette incontri in modalità telematica, in cui numerosi accademici ed esperti del settore si sono confrontati sui seguenti temi: 20 ottobre 2021, *AI: quali rischi per l'autonomia dell'umano*, (Intelligenza + D), con relatori Angelo Alù, Mariagiovanna Gianfreda, Guglielmo Tamburrini, discussant Mario Bochicchio e moderatrice Ilenia Colonna; 27 ottobre 2021, *Immagini del passato, immagini del futuro* (Media + D), con relatori Malvina Giordana, Alma Mileto e Francesco Zucconi, discussant Luca Bandirali e moderatrice Isabella Hernandez; 9 novembre 2021, *Cultural Heritage & Digital Humanities: sfide di accessibilità* (Arte + D), con relatori Eva Degl'Innocenti, Lucio Tommaso De Paolis, Anna Maria Marras, Paola Moscati, discussant Grazia Semeraro e moderatrice Carola Gatto; 10 novembre 2021, *Textual scholarship: forme, strumenti, metodi* (Testo + D), con relatori Marina Buzzoni, Tiziana Mancinelli, Federico Meschini, Andreas Speer, discussant Fabio Ciraci e moderatrice Giulia Miglietta; 12 novembre 2021, *Politiche pubbliche per la costruzione di un ecosistema digitale* (Diritto + D), con relatori Bianca Bronzino, Mino Elefante, Claudia Morini, discussant Marco Mancarella e moderatore Marco Giannotta; 17 novembre 2021, *Tecnologia e umano: quale futuro per la conoscenza* (Filosofia + D), con relatori Simona Chiodo, Riccardo Fedriga, Cristina Marras e Viola Schiaffonati, discussant Fabio Ciraci e moderatrice Patrizia Miggiano; 24 novembre 2021, *Costruire mondi possibili: i videogiochi e le realtà sociali* (Media + D), con relatori Donata Bologna, Marco-Benoît Carbone, Riccardo Fassone e Pietro Luigi Iaia, discussant Luca Bandirali e moderatrice Alessia De Blasi.

La risposta alla *call for papers* è stata, ci pare, all'altezza delle aspettative: sono giunte 86 proposte, con una media di paper accettati del 77%, esattamente 18 paper e 5 poster accettati nell'area Testo+D, 3 paper accettati nell'area Arti+D, 11 paper e 2 poster in area Filosofia+D, 5 paper e 1 poster per Contenuti+D, infine 7 paper per Intelligenza+D, per un totale di 44 paper e 8 poster. Già da una rapida lettura dei titoli si evince non solo la molteplicità dei temi ma anche la varietà degli approcci metodologici, che attestano declinazioni interne anche alle medesime aree tematiche. Infine, per garantire una selezione dei contributi conforme alle aree di ricerca selezionate, abbiamo identificato la figura dei direttori di area, ai quali è stato assegnato il compito di individuare i revisori più adeguati ai temi dei contributi da revisionare, per un'analisi competente e puntuale: per l'area testo, Roberto Rosselli Del Turco; per l'area arti, Grazie Semeraro; per l'area filosofia, Fabio Ciraci; per l'area contenuti, Luca Bandirali e Marco Mancarella; per l'area intelligenza, Mario Bochicchio. A tutti loro va il ringraziamento del Comitato Scientifico e di AIUCD. Una tale suddivisione del lavoro e il supporto della piattaforma digitale *conference* hanno permesso di seguire con efficacia tutto il processo di selezione dei contributi: individuare i revisori idonei, confrontare le valutazioni e richiedere pareri ulteriori in caso di dubbio, controllare che le modifiche richieste agli autori in fase di revisione fossero correttamente apportate alla versione finale del *paper*, selezionare i contributi da presentare alla conferenza. Ciascun contributo è stato valutato da almeno due *referee* in caso di giudizio positivo, almeno tre in caso di giudizio incerto o di giudizi discordanti, o parere negativo. I 75 revisori hanno svolto un lavoro fondamentale di revisione che ha garantito una selezione seria e competente, assicurando al convegno dell'AIUCD la qualità delle proposte e il riconoscimento internazionale duramente conquistato dall'Associazione in questi undici anni di attività.

AIUCD2022 è patrocinato dalla Regione Puglia, dalla Provincia di Lecce e dalla Città di Lecce, la qual cosa è certo indice di una certa sensibilità territoriale ai temi della cultura e della innovazione. Inoltre, il convegno è stato sponsorizzato da: Dipartimento di Informatica dell'Università degli Studi di Bari, AFP – Apulia Film Commission, Il Teatro Pubblico Pugliese, CINI – Consorzio Interuniversitario nazionale per l'Informatica, SFI-Società Filosofica Italiana, il Teatro Pubblico Pugliese e Argo Software, che hanno generosamente sovvenzionato l'iniziativa.

Purtroppo, come nella scorsa edizione, nonostante il ricorso ai vaccini, anche quest'anno il *covid* ha ripreso a correre, improvvisamente, a poco più di una settimana dal Convegno, previsto per il 19-21 gennaio 2022, e ci ha costretti a rinviare il Convegno alla prossima estate. La scelta è stata sofferta e sicuramente ha determinato disagi, ma abbiamo inteso dare priorità alla sicurezza e alla salute pubblica, pur in assenza di decreti restrittivi o limitazioni governative all'attività convegnistica. Non abbiamo inteso proporre invece il convegno in modalità *online*, perché non abbiamo voluto rinunciare al nostro amato convegno in presenza. La virtuosa trasposizione in modalità digitale di AIUCD2021 offerta, in emergenza, per il Convegno di Pisa è stata sicuramente un esperimento riuscito. Tuttavia, dopo due anni di pandemia, il Comitato Scientifico, di concerto con il Direttivo AIUCD, ha reputato opportuno scegliere comunque di rinviare, per privilegiare il convegno in presenza, senza ovviamente rinunciare ai vantaggi offerti dalla modalità ibrida. Un ulteriore convegno solo in remoto avrebbe altrimenti gravato immancabilmente sugli aspetti sociali e relazionali, per nulla secondari, che costituiscono la vera sostanza del convegno nazionale, rendendolo un luogo di confronto vivo, un'insostituibile occasione di relazione e di partecipazione attiva. Siamo dell'opinione che il digitale debba rappresentare un'opportunità, non già una dimensione sostitutiva ed esclusiva, ma complementare e inclusiva.

Il Convegno previsto per il 19-21 gennaio 2022 indicava la partecipazione di prestigiosi studiosi che arricchivano la proposta tematica di AIUCD2022, che intendiamo confermare anche per il rinvio di giugno. Innanzitutto, i nostri *keynote*: Luciano Floridi – Professore Ordinario di filosofia ed etica dell'informazione presso l'Oxford Internet Institute e direttore del Digital Ethics Lab dell'Università di Oxford, nonché Professore di Sociologia della comunicazione presso l'Università di Bologna – inaugurerà il convegno con una lezione su *Semantic capital: its nature, value, and preservation*; Maurizio Ferraris – Professore Ordinario di filosofia teoretica presso la Facoltà di Lettere e Filosofia dell'Università degli Studi di Torino e noto studioso della *documerialità* – concluderà i lavori con una lezione intitolata *Webfare*. Si aggiungeranno gli *invited speaker* che, per ogni giorno della conferenza, sviluppano un tema specifico del convegno: Maria Grazia Mattei – umanista, critica d'arte e direttrice di *Meet the Media Guru* – si soffermerà sull'*Arte digitale: storia e panoramica attuale*; Gino Roncaglia – Professore Associato dell'Università Roma Tre, esperto di digitale e cultura del libro, consulente RAI – discuterà di *Simulismi*; Anna Bisogno – Professore Associato di Cinema Radio e Televisione dell'Università Telematica Mercatorum – analizzerà *La rete-visione. Televisione e schermi nell'era digitale*;

infine, Riccardo Fedriga – Professore Associato dell'Università di Bologna, esperto di editoria digitale, storico delle idee – esaminerà le *Fruttuose debolezze. Fragilità e indeterminismi digitali*.

Lavoreremo affinché il programma, così faticosamente costruito per gennaio, non subisca variazioni strutturali. Inoltre, al posto del consueto *Book of Abstracts*, per l'edizione del 2022 l'AIUCD ha scelto di pubblicare i *Proceedings*, come segno tangibile di un processo di aggiornamento continuo del Convegno Nazionale e di crescita intellettuale dell'Associazione. Essi vedono la luce nonostante il rinvio del convegno in presenza, per fornire una base alla discussione che si svolgerà questa estate, con la consapevolezza che gli studi pubblicati fotografano lo stato dell'arte, ma che la ricerca è in continua evoluzione. Quindi, in sede di convegno, faremo i conti con i progressi avvenuti nei mesi trascorsi dalla pubblicazione dei *Proceedings*, di cui terremo conto per l'eventuale pubblicazione dei *selected papers*.

Vorremmo chiudere la prefazione rivolgendo un particolare ringraziamento ai membri del Comitato Scientifico e, *last but not least*, esprimendo profonda gratitudine ai componenti del Comitato di programma, coordinati da Federica Epifani: tutte giovani e promettenti energie intellettuali a cui è dedicato il presente volume di *Proceedings*, non a caso edito a cura di Giulia Miglietta e Carola Gatto.

Fabio Ciraci

Mario Bochicchio

Sessione Arti 1
Artemisia Gentileschi

La Comédie Virtuelle

Luca Izzo

Università della Campania “Luigi Vanvitelli”, Italia, luca.izzo@unicampania.it

ABSTRACT

L’oggetto della ricerca è una performance di arte contemporanea – e, in più, un’*experience* di visita virtuale al nuovo edificio del Teatro di Ginevra – che rappresentano un caso unico nel loro genere, perché realizzate attraverso tecnologie di realtà virtuale e con i performer che sono ognuno in una città diversa, ma tutti contemporaneamente collegati e “presenti”, insieme ai fruitori e grazie alla rete, nel ciberspazio digitale della VR. Obiettivo della ricerca è stato quello di analizzare l’utilizzo del medium, inteso sia come strumento, sia come trans-ambiente e sia come trans-luogo della cognizione e della percezione del fruitore, ciò per individuare eventuali punti di forza ed eventuali punti di debolezza dell’utilizzo delle ICT al servizio della performance di arte contemporanea in tempo reale e con l’immersività sensoriale dei *performers* e dei fruitori, operazione ancora inedita su scala globale e battezzata proprio dall’*experience* in esame. La metodologia di ricerca è stata condotta attraverso la fruizione diretta dell’*experience* e, in seguito, procedendo con l’analisi secondo i criteri e le dinamiche unanimemente condivise nel panorama della letteratura scientifica internazionale. In particolare lo studio è stato svolto secondo la corrente di ricerca – in ambito psicologico – detta *Embodied Cognition*, nata in parte in opposizione ma anche in parte in integrazione, nei confronti delle scuole di pensiero e di indagine tradizionali, dal *Comportamentismo* al *Cognitivismo* e fino al *Connessionismo* ([24];[23]). Inoltre, alla *Cognizione incarnata* la ricerca ha affiancato gli ultimi risultati sugli studi dei cosiddetti *neuroni specchio* ([21]), costituenti una zona cerebrale che si attiva in direzione dell’emulazione di ciò che si percepisce visivamente, cosicché l’*experience* in VR rischia di connotarsi di un potenziale comunicativo ed empatico raro, offrendo un canale di comunicazione, fruizione e valorizzazione di contenuti non trascurabile. Seguendo i criteri dei sopradetti approcci scientifici, sono stati analizzati i rapporti utente-interfaccia, fruitore-transluogo, comunicazione-destinatario, in modo da individuare possibili caratteristiche per la maturazione di *best-practices* per gli obiettivi sopradetti.

PAROLE CHIAVE

Digitale, arte contemporanea, realtà virtuale, psicologia della fruizione, media.

INTERVENTO

1. FOCUS E ANALISI

L’oggetto della ricerca è una performance di arte contemporanea e un’*experience* di visita ad un nuovo edificio museale, realizzata dall’artista Cie Gilles Jobin (ballerino, coreografo e regista svizzero) in collaborazione con il Teatro di Ginevra e *La Comédie de Genève*. Questa operazione insieme artistica, architettonica, di design, progettazione e ingegneria informatica ricrea un mondo virtuale tutto in disegno digitale e, grazie alle ICT (*Information and Communications Technology*), riesce a connettere e mettere in relazione l’edificio con i diversi visitatori del teatro, gli artisti e i fruitori della performance, tutti connessi attraverso la rete internet – e in modalità VR – ognuno da qualsiasi luogo geografico si possa trovare. L’*experience* è formata da due possibili attività interattive: la visita alla nuova sede teatrale (che ogni utente può fare in propria libertà e in qualsiasi momento) e la fruizione della performance artistica (che, necessitando della partecipazione in tempo reale del team artistico di Jobin, prevede appuntamenti precisi). Per fruire di ognuna di queste opzioni, gli utenti devono indossare il visore VR e prendere i due *controller touch*. Partita l’applicazione, lo spettatore si ritrova – in modalità VR - nel giardino del nuovo edificio del Teatro di Ginevra (fig.1). L’interazione con l’interfaccia e l’usabilità del media risultano fortemente trasparenti¹: puntando il *controller* verso una direzione e tenendo premuto il tasto

¹ Un media digitale si definisce “trasparente” quanto più riesce ad avere i comandi, e quindi l’usabilità, semplice, intuitiva e, di conseguenza, non richiedente le energie attenzionali del fruitore, cosicché il fruitore, viceversa, potrà dirigere tutte le proprie energie e capacità attenzionali al transluogo in fruizione e agli eventi e agli oggetti in questo contenuti. Di contro, un media è indicato come “opaco” quanto più, per attuare i comandi e far procedere l’usabilità, esige le capacità e le energie attenzionali del fruitore, con la conseguenza che tali dinamiche attenzionali, non solo saranno private alla fruizione dell’oggetto dell’interesse, ma, per di più, faranno da richiami attenzionali al fruitore per percepire lo stato di alterazione della presenza fisica in un determinato luogo reale (con la necessità di utilizzare lo strumento materiale del media) e la presenza percettivo-psichica-attenzionale in un ambiente virtuale ([20]).

per pochi secondi, apparirà un raggio verde che va nella direzione scelta, luogo nel quale il nostro avatar² viene immediatamente trasferito.



Fig. 1. Nuovo edificio del Teatro di Ginevra ricostruito in digitale: esterno.

Con questa modalità di interazione è possibile spostarsi ovunque si vuole all'esterno e all'interno del teatro, potendo esperire una visita completa del nuovo edificio del Teatro di Ginevra. L'edificio e tutto il contesto sono realizzati in disegno digitale in 3D ed a 360°, permettendo diversi livelli di interazione, ovvero i 6 DOF (*Degrees of freedom*).

Nella visita al nuovo teatro non si è soli: Jobin ha voluto animare le sale con personaggi virtuali di ogni tipo: ballerini mentre provano, pagliacci, giocolieri, figure fantastiche, ecc., realizzati anche su diverse scale proporzionali (fig.2).



Fig. 2. Nuovo edificio del Teatro di Ginevra ricostruito in digitale: interno ingresso.

La visita al teatro è un'*experience* molto ben sviluppata. L'utente può riuscire a percepire un buon senso di presenza³ nel ciberspazio virtuale, che, a sua volta, riesce a fornire – come conseguenza dell'*experience* esperita – una sufficiente sensazione di aver visitato il nuovo edificio e di aver preso coscienza, in buona parte, di come questo si presenta.

Si individuano chiaramente alcune caratteristiche positive che concorrono alla riuscita di questi buoni esiti, e ovvero: la piena libertà di spostamento nel mondo virtuale; la piena libertà di direzionare lo sguardo, anche con scorci e inclinazioni molto accentuati sia verso l'edificio che verso i personaggi presenti; la possibilità di andare ovunque nell'edificio, dalla reception alle sale più remote dei vari livelli.

Tra gli aspetti realizzati negativamente, per la riuscita percettiva e cognitiva della sensazione di aver avuto un'esperienza in un determinato luogo, si possono individuare alcune caratteristiche: lo spostamento nell'ambiente non è stato realizzato con una modalità psico-simulativa del reale ma, altresì, con uno raggio verde di fantasia che inevitabilmente restituisce feedback di astrazione del mondo virtuale e delle anomalie percettive del fruitore; il fruitore è dentro il mondo virtuale solo con le capacità attenzionali e con le capacità percettive visive e uditive (altri sensi non ne vengono sollecitati); del suo corpo non esiste rappresentazione, né totale né parziale, né realistica né in disegno digitale, cosicché, se questi prova a guardare le proprie mani e/o i propri piedi, sarà costretto a percepire un vuoto della sua persona, altra caratteristica che, inevitabilmente, richiama ad una presa di coscienza del fruitore di vivere un'esperienza fortemente limitata e parziale, molto lontana dagli effetti di un'esperienza di visita reale.

Per quanto riguarda la fruizione della performance di arte contemporanea, questa è programmata in determinati orari. Da qualsiasi luogo (virtuale) ci si trova nel teatro, si ascolta uno speaker che annuncia l'inizio della performance nella sala numero X e che è possibile seguire le linee gialle sul pavimento per essere guidati alla sala. Giunti nella sala apposta,

² Per "avatar" si intende la rappresentazione data all'utente – visibile o non visibile (se non visibile allora intuibile e/o deducibile in base al punto di vista e la posizione) – nel transluogo virtuale, che sia in 2D, in 3D o in 4D.

³ Per i diversi livelli di sensazione di presenza nell'ambiente virtuale consultare ([19]).

questa si presenta come una grande palestra rettangolare, in cui – in confronto all’entrata – verso la parete in fondo c’è un grosso schermo nel quale si può vedere rappresentata un’area all’aperto (fig.3), mentre a sinistra, sul pavimento, c’è disegnata una griglia a scacchiera bianca, che corrisponde all’area della performance (fig.3) e, sulla parete di sinistra, più vicina alla zona della performance, si possono vedere tre grossi schermi con accanto, per ognuno di questi, i nomi delle città dalle quali si collegheranno i performer (le tre città sono Ginevra, Bengaluru e Sidney). In questa circostanza non si potrà non notare la presenza di altri fruitori, anch’essi collegati attraverso la rete e i device di VR. Questi sono rappresentati, nel transluogo virtuale, da sottili lineette colorate all’altezza del busto, due cerchi colorati al posto delle mani, un cerchio colorato all’altezza della testa e il loro nome scritto sopra, in scrittura digitale con caratteri bianchi (fig.3). L’evento artistico inizia con il collegamento, visibile nei tre schermi della parete di sinistra, dei 5 performer: 3 da Ginevra, 1 da Bengaluru e 1 da Sidney. Questi, con le loro sembianze reali (stabilendo, quindi, un collegamento non solo verso altre città ma, contemporaneamente, anche tra mondo virtuale e mondo reale), salutano gli spettatori e iniziano il riscaldamento poi, quando ha inizio la performance, i 5 performer appaiono, con i loro avatar disegnati in digitale ma dalle sembianze umane, nell’area a scacchiera bianca della sala, tra i fruitori in attesa (fig.3).



Fig. 3. Nuovo edificio del Teatro di Ginevra ricostruito in digitale: inizio della performance *La Comédie Virtuelle*.

Ha inizio la performance: da questo momento in poi ogni movimento realizzato da ognuno degli artisti dovrebbe essere sincronizzato sia con la loro immagine rappresentata negli schermi sulla parete di sinistra (che rappresenta il mondo reale), sia con i loro avatar digitali nell’ambiente virtuale. Gli artisti – e i loro avatar – danzano: coreografie libere che si intrecciano, si susseguono, ecc. I fruitori possono osservare restando all’esterno della griglia bianca delimitante l’area della performance, oppure anche entrando dentro l’area di esibizione, avvicinandosi ai performer e muovendosi tra loro. Nel grande schermo in fondo alla sala, invece, gli avatar dei performer sono rappresentati a danzare (sincronizzati) in un luogo all’aperto, che a tratti diventa l’esterno del nuovo edificio del Teatro di Ginevra.

Il trans-ambiente virtuale è stato realizzato così completamente sul piano della ricostruzione di un “nuovo mondo” (il mondo virtuale) che permette la visita al teatro – e alla performance – in totale libertà utente per utente, cosicché, se anche un fruitore accedesse alla sala della performance dal primo piano, questi la potrà ammirare dal suo peculiare punto di vista. Gradualmente la performance va ad astrarsi sempre più e i ballerini cominciano a diventare come figure gonfiate ad area che si sollevano dal suolo. L’architettura intorno al fruitore comincia a sparire, lasciando il posto a un cielo stellato: le figure di sagome gonfiate dei performer si ritrovano a danzare in un universo artefatto.

In conclusione, ai fruitori viene offerta la possibilità di entrare in un “mondo nuovo”, un mondo virtuale realizzato tutto in disegno digitale 3D e a 360°, che fa dimenticare le proprie limitazioni in quanto offre un’esperienza molto ricca: la visita ad un edificio con la libertà di poter andare ovunque si desidera; la fruizione di diverse figure “virtualmente vive” che circondano, ovunque, lo spettatore (le tante figure che animano gli spazi del teatro); la fruizione di una performance di arte contemporanea, unica nel suo genere, realizzata grazie alle ultime tecnologie di VR, in grado di connettere e mettere in relazione spazio-temporale e realtà-virtuale più persone, anche se stanti in diverse città del mondo, collegate dai loro ambienti reali e immessi in uno stesso trans-ambiente virtuale. Come precedentemente detto, sul piano progettuale di disegno digitale e architettonico, tutta l’operazione gode veramente di superba fattura. Nella progettazione visivo-percettiva si evidenzia un solo errore, anche se significativo, e ovvero che gli avatar dei performer appaiono in scala leggermente ridotta in confronto alla proporzione del punto di vista – e quindi della presunta virtuale fisicità – dello spettatore. Questa lieve disarmonia occupa un ruolo importante nell’ostacolare un profondo senso di “immersività” psico-percettivo del fruitore, in quanto fa da continuo feedback nel sollecitare la percezione di un mondo artefatto. Sul piano della realizzazione complessiva dell’*experience*, invece, non mancano punti critici molto significativi. Diverse superficialità grafico-organizzative ostacolano una piena immersività percettivo-psicologica del fruitore, fornendo continui feedback di percezione di un mondo immateriale, digitale, nutrendo un senso di astrazione e di estraniamento dell’utente. Fattori quali:

i fruitori non hanno rappresentazione corporeo-digitale attraverso un avatar, ma sono rappresentati come semplici linee colorate con un puntino al posto della testa e due puntini al posto della mani; la mancata previsione, da parte dei performer, di non far incrociare i loro avatar nei rispettivi movimenti, causa, nel trans-ambiente virtuale, momenti in cui gli avatar si trapassano l'uno con l'altro; le proporzioni leggermente più piccole del naturale degli avatar dei performer. Tali caratteristiche, come predetto, hanno un ruolo nel denunciare continuamente l'immaterialità del luogo virtuale nel quale il fruitore è immerso, quindi, di conseguenza, danno continua visibilità ai limiti dell'esperienza. Inoltre, purtroppo, bisogna segnalare anche la presenza di una criticità ben più grave e che – pur se restano validi i concetti alla base della realizzazione di Jobin e del Teatro di Ginevra – annulla, però, tutto il vero e straordinario senso della performance (ovvero la possibilità di collegare - in diretta e simultaneamente - diverse persone da diverse parti del mondo, performer e fruitori, dai loro mondi reali ad un unico e condiviso trans-luogo): se si guarda contemporaneamente i movimenti degli avatar e i movimenti delle persone reali connesse attraverso i tre monitor nella sala, purtroppo si noterà palesemente che i movimenti dei performer e dei loro relativi avatar non corrispondono, facendo crollare la credibilità di una diretta rappresentazione/presenza degli artisti dal loro mondo reale, e dalle loro diverse città, ad un unico e condiviso trans-ambiente.

2. OBIETTIVI DELLA RICERCA

Obiettivo della ricerca è stato quello di analizzare l'utilizzo del medium, inteso sia come strumento, sia come trans-ambiente e sia come trans-luogo della cognizione e della percezione del fruitore, ciò per individuare eventuali punti di forza ed eventuali punti di debolezza dell'utilizzo delle ICT al servizio delle arti contemporanee, in particolare, in questo caso, quando il medium è utilizzato per creare, in tempo reale, un transluogo di incontro capace di permettere l'immersività sensoriale dei performer e dei fruitori (operazione ancora inedita su scala globale e battezzata proprio dall'experience in esame). L'analisi delle dinamiche di fruizione offerte da questa esemplare *experience* di visita virtuale e performance di arte contemporanea, è stata svolta, inoltre, allo scopo di riconoscere come modelli di riferimento alcune *best-practices* per eventuali future applicazioni miranti a simili obiettivi, nonché alla conoscenza ed alla consapevolezza del potenziale e delle possibilità offerte dalle ICT all'espressioni di arte contemporanea.

3. METODOLOGIA

L'analisi dell'esperienza di fruizione è stata sviluppata attraverso i principali criteri che costituiscono la corrente di ricerca – in ambito psicologico – detta Cognizione incarnata, caratterizzata da tratti di ispirazione merleau-pontiana, nata negli anni Ottanta (in parte in opposizione e in parte in integrazione alle scuole di pensiero e di indagine tradizionali, dal comportamentismo al cognitivism e fino al connessionismo) e progressivamente elaborata e sviluppata fino ad oggi e in ambiti diversi, quali le scienze cognitive, la neuropsicologia, la psicologia percettiva, la psicologia dei media, ecc. Secondo questa linea di studi le capacità cognitive di un essere vivente, e l'universo cognitivo di ogni momento della sua vita, non dipendono esclusivamente dal cervello (come, altresì, indicano le teorie di indagine psicologica tradizionali) ma sono il frutto del risultato di una complessa serie di combinazioni riguardanti sia il cervello che il corpo, con tutto il suo apparato sensoriale, quindi di una condizione definita *incarnata*; tali capacità senso motorie non sono certo isolate, ma anch'esse dipendono e sono incluse in un contesto più ampio, come quello ambientale, culturale, mediale, ecc. ([24];[23]). Secondo tale scuola di pensiero, i processi cognitivi sono imprescindibili dai processi sensoriali e dai processi motori; la cognizione è incarnata nell'intero organismo e, con questo, situata nel mondo: il realizzarsi di un'esperienza negli infiniti modi possibili, il significato che questa stimolerà e riceverà dal soggetto stesso, non può essere riconducibile alle sole strutture cerebrali ma, viceversa, sarà sempre il prodotto di continue e reciproche relazioni fra il corpo (di cui il cervello è parte) e l'ambiente, compreso gli strumenti e/o i media che questo include ([13]). Tale ambito di ricerca, è stata adottata integrando i suoi diversi ma interagenti percorsi di sviluppo e di pensiero, indicati come le *4E della cognizione*, ovvero *embodied, extended, embedded, enactive*. La *Cognizione incarnata* appare particolarmente significativa nell'approcciarsi all'analisi dell'experience in VR, esaltandone il potenziale comunicativo ed esperienziale (in quanto medium che *incarna* il fruitore in un corpo digitale e, in questo, lo trasporta in un nuovo ambiente, digitale e immersivo) e mettendone in luce le criticità di fruizione (dato che l'utente viene a trovarsi in una singolare bilocazione, ovvero percepisce il suo vero corpo nella realtà – e ciò può distogliere l'attenzione dall'interazione con il mondo artefatto digitale che lo circonda, con la conseguente caduta dell'attenzione verso quanto si esperisce – ma, nello stesso tempo, si percepisce come corpo virtuale nel transluogo digitale). Inoltre, alla *Cognizione incarnata* la ricerca ha affiancato gli ultimi risultati sugli studi dei cosiddetti *neuroni specchio* ([21]), costituenti una zona cerebrale che si attiva in direzione dell'emulazione di ciò che si percepisce visivamente, cosicché l'experience in VR rischia di connotarsi di un potenziale comunicativo ed empatico raro, offrendo un canale di comunicazione, fruizione e valorizzazione di contenuti non trascurabile.

Seguendo i criteri dei sopradetti approcci scientifici, sono stati analizzati i rapporti utente-interfaccia, fruitore-transluogo, comunicazione-destinatario, in mira di individuare possibili caratteristiche per la maturazione di *best-practices* per gli obiettivi sopradetti.

4. CONCLUSIONI

Lo studio riconosce in *La Comédie Virtuelle* una performance di arte contemporanea – e un’*experience* in VR di visita virtuale – particolarmente all’avanguardia, per la singolarità di usare il medium digitale per abbattere i limiti di spazio e di tempo, consentendo a più autori e fruitori di ritrovarsi, in presenza (digitale e percettivo-attentionale) in un trans-ambiente condiviso, e poter sviluppare e fruire una performance di arte contemporanea, nonostante i singoli individui siano, in realtà fisico-temporale, ognuno nella sua città e in continenti diversi.

Ciò nonostante, la realizzazione digitale in VR presenta diverse criticità, purtroppo influenti sulle modalità e qualità della fruizione, ostacolando una piena immersività sensoriale dei fruitori. L’immersività spaziale è realizzata ottimamente, permettendo tutti i 6 gradi di libertà, ciò che viene intralciata è l’immersività attentionale, viste le *affordance* cognitive offerte di riallaccio alla dimensione reale.

BIBLIOGRAFIA

- [1] Balzola, Andrea, e Anna Maria Monteverdi. 2004. *Le arti multimediali digitali. Storia, tecniche, linguaggi, etiche ed estetiche delle arti del nuovo millennio*. Garzanti Libri S.p.a.
- [2] Bordini, Silvia. 2006. *Storia del panorama. La visione totale nella pittura del XIX secolo*. Edizioni Nuova Cultura.
- [3] Bucher, John. 2018. *Storytelling for Virtual Reality: Methods and Principles for Crafting Immersive Narratives*. Routledge.
- [4] Cappelletto, Chiara. 2009. *Neuroestetica. L'arte del cervello*. Edizioni Laterza.
- [5] Capucci, Pier Luigi. 1996. *Arte e tecnologie. Comunicazione estetica e tecnoscienze*. Edizioni dell’Ortica.
- [6] D’Aloia, Adriano. 2013. *La vertigine e il volo. L’esperienza filmica fra estetica e neuroscienze cognitive*. Fondazione ente dello spettacolo.
- [7] Dalpozzo, Cristiano, Federica Negri, e Arianna Novaga. 2018. *La realtà virtuale. Dispositivi, estetiche, immagini*. Mimesis.
- [8] Dogramaci, Burcu, e Fabienne Liptay. 2016. *Immersion in the visual arts and media*. Brill-Rodopi.
- [9] Eugeni, Ruggero. 2015. *La condizione postmediale*. La scuola editrice.
- [10] Fidotta, Giuseppe, e Andrea Mariani. 2018. *Archeologia dei media. Temporalità, materia, tecnologia*. Meltemi Editore.
- [11] Fontana, Andrea. 2009. *Manuale di storytelling*. Rizzoli Etas.
- [12] ———. 2016. *Storytelling d’impresa. La guida definitiva*. Hoepli.
- [13] Hillis, Ken. 1999. *Digital Sensations. Space, Identity and Embodiment in Virtual Reality*. University of Minnesota Press.
- [14] Izzo, Filomena. 2017. *Musei e tecnologie: valorizzare il passato per costruire il futuro*. Wolters Kluwer.
- [15] Montani, Pietro, Dario Cecchi, e Martino Feyles. 2018. *Ambienti mediali*. Meltemi.
- [16] Parisi, Francesco. 2019. *La tecnologia che siamo*. Codice Edizioni.
- [17] Perniola, Mario. 1997. *L’estetica del Novecento*. Il Mulino.
- [18] Polidoro, Piero. 2008. *Che cos’è la semiotica visiva*. Carocci Editore.
- [19] Riva, Giuseppe. 2004. *Psicologia dei nuovi media*. Il Mulino.
- [20] ———. 2019. *Realtà virtuali. Gli aspetti psicologici delle tecnologie simulate e il loro impatto sull’esperienza umana*. Giunti Editore.
- [21] Rizzolatti, Giacomo, e Corrado Sinigaglia. 2006. *So quel che fai. Il cervello che agisce e i neuroni specchio*. Raffaello Cortina.
- [22] Rogers, Ariel. 2013. *Cinematic appeals. The experience of new movie technologies*. Columbia University press.
- [23] Shapiro, Lawrence. 2019. *Embodied cognition*. Routledge.
- [24] Varela, Francisco J., Evan Thompson, e Eleanor Rosch. 1992. *La via di mezzo della conoscenza. Le scienze cognitive alla prova dell’esperienza*. Feltrinelli.

Climate change & digital cultural impact, the Victoria & Albert Museum

Mauro De Bari

University of Bari Aldo Moro, Italy – mauro.debari@uniba.it

ABSTRACT

Climate change is modifying the daily life of the communities, influencing the proposals for the enjoyment of the cultural heritage. In this unconventional situation, the digital is helping cultural institutions to address this issue, offering environmental-friendly solutions. This paper exposes the contemporary state-of-the-art, proposing examples of good practice to follow, identified in the Victoria & Albert Museum of London (V&A).

KEYWORDS

Climate change, cultural institutions, digital, cultural heritage.

TALK

1. INTRODUCTION

Climate change affects society daily ([9]), affecting cultural scenarios because the environment and heritage are strictly connected. Since 1972, UNESCO and representative members have adopted the Convention for the Protection of the World Cultural and Natural Heritage to stimulate the awareness of current and future generations ([11]).

This assumption has grown in importance, manifesting the necessity to launch some guidelines on the argument presented in 2003's Convention for the Safeguarding of the Intangible Cultural Heritage ([11]). Thus, especially after the COVID-19 pandemic, people seem to have understood the fragility of the ecosystem and the necessity to preserve it tangibly ([8]). Moreover, the cultural scene is strongly influenced by environmental changes ([6]). For this reason, many cultural institutions have planned proposals to address this issue correctly, especially after the two years of stasis provoked by the pandemic ([12]).

Recently, Europe has invested many resources to consolidate a European Green Deal plan¹. Politicians such as Ursula von der Leyen and Frans Timmermans consider this innovative plan as the element to restart proactively from the pandemic period. Inevitably, the call challenged culture and cultural institutions to contribute. As a result, Europa Nostra², a partner with Climate Heritage Network, has produced a book titled *European Heritage Green Paper*³, a *vademecum* for professionals of the cultural system. This challenging contribution will reinforce the underpinnings of a functional Green Deal plan. In this context, the most respectful strategy that creates a linkage between culture and preservation of the environment suggests digital: digital is the resource to invest in terms of possibilities to offer to people.

In this paper, an example of good practices that tempt users to visit and enjoy the heritage will be outlined, respecting the basic for respectful fruition of inheritance and environment. Notably, two years after the pandemic spread, the V&A museum launched a significant event on this topic under the name of *Digital Design Weekend 2021*⁴. The aim is to blow people's minds, sensitise them to climate change, and use, as aforementioned, the digital as liaison.

2. GOOD PRACTICES IN THE VICTORIA & ALBERT MUSEUM

Cultural Heritage represents human creative and cultural expressions. It maintains a unique and indisputable value for human beings, communities and societies. However, climate change is provoking the deterioration of cultural heritage, necessitating a safeguarding strategy⁵.

In such a delicate scenario, museums play specific roles among cultural institutions as the significant exponents of cultural experiences and mediators with the user. One of the last points of discussion debated by the International Committee for

¹ https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_it.

² <https://www.europanostra.org/>.

³ https://issuu.com/europanostra/docs/20210322-european_cultural_heritage_green_paper_fu.

⁴ <https://www.vam.ac.uk/event/EY7xyG3q/digital-design-weekend-2021>.

⁵ <https://erc.europa.eu/projects-figures/stories/eu-research-and-innovation-for-more-resilient-cultural-heritage>.

Museums and Collections of Modern Art (CiMAM) was based on the challenge of museums conversion into a new approach that contrasts climate change. Fascinating is the point that argues the necessity to “*Produce virtual materials for families, learners, the disabled, and everyone to access online*” ([4]).

Many museums are moving to green with digital-environmentally friendly solutions ([10]). CiMAM reports that, in the last period, 600 professionals from 86 countries have contributed to the cause. Some museum realities in Germany, France, and the UK are shifting in this sense. However, “*European funds earmarked for digital and ecological transformation will accelerate a process that keeps most of the national museums in a critical situation (16 state institutions and 60 with transferred management) in Spain*”⁶ mainly.

Among all, The V&A of London represents the avant-garde on access online of artistic production. As previously mentioned, the museum has dedicated an entire event to sensitise people on climate and environmental issues, involving designers, artists, and humanists. In agreement with the CiMAM and ICOM⁷ proposes, V&A has spotted a strong advertising campaign ([3]). It is not the first pioneering action moved by V&A, considering its approach to balance touch-free experiences with traditional ones.

However, it is a long process, and it needs many competencies necessary to attract users, promoting unconventional and unique experiences ([1]).

The museum aims to revolutionise the entire traditional approach of cultural experiences provocatively, shifting as fast it can to a museum of the Digital Era. However, digital resources are not stable and permanent but are influenced by digital transformation (DT) changes. For this reason, “*it is becoming increasingly clear that preserving born-digital objects frequently requires a certain degree of reinterpretation and re-creation of the object, as opposed to more traditional approaches to conservation work aimed at stabilising the object in a fixed condition*” ([1]). Therefore, conservation and safeguarding are a challenge that must be considered and improved by all the cultural institutions, not only museums.

The museum aims “*to improving discoverability of the V&A’s collections online has been to better understand our users’ behaviours and to support their needs by opening journeys that were previously closed*” ([7]). V&A knows consciously that the online experience is not the same as the traditional one. In standard cultural experiences, users follow a route; by contrast, users choose to start from a specific point moving in a digital world in the digital alternative. So, for example, they could decide to reduce the entire cultural offer to enjoy just a digital resource or the whole, managing their time and in an eco-friendly way. Hence, the traditional visiting experience could be overturned and modified towards a touch-free approach ([3]), respecting the environment. Nevertheless, the premise remains to generate a strong emotional impact in users by involving them actively in the experience of interaction with goods, and it is this perspective according to which operating various international cultural institutions.

Aiming at an “unconventional” application of touch-free methods and tools and VR reconstructions of cultural heritage, the models developed by V&A present innovative elements that incorporate many of the previously highlighted issues, evolving the concept of “digital entity” into *museum representation meta-real* ([2]).

REFERENCES

- [1] Arrigoni, Gabi. 2021. “Born-Digital Objects and the Not so Digital Challenges for Museum Collections.” *V&A Blog* (blog). September 13, 2021. <https://www.vam.ac.uk/blog/projects/born-digital-objects-and-the-not-so-digital-challenges-for-museum-collections>.
- [2] Barbuti, Nicola, Giuliano De Felice, Annalisa Di Zanni, Paolo Russo, and Altheo Valentini. 2019. “Creating Digital Culture by Digitizing Cultural Heritage: The Crowddreaming Living Lab Method.” In *Didattica e Ricerca al Tempo Delle Digital Humanities. Teaching and Research in Digital Humanities’ Era*, 113–16. Udine.
- [3] Barto, Marc. 2021. “Digital Design Weekend 2021.” *V&A Blog* (blog). September 23, 2021. <https://www.vam.ac.uk/blog/design-and-society/digital-design-weekend-2021>.
- [4] Bethan, Ross. 2020. “Is This the End of Touchscreens in Museums? The Use of Touchless Gesture-Based Controls.” *Science Museum Group Digital Lab* (blog). June 18, 2020. <https://lab.sciencemuseum.org.uk/is-this-the-end-of-touchscreens-in-museums-the-use-of-touchless-gesture-based-controls-ee3f3c3f37ce>.
- [5] “CiMAM Toolkit on Environmental Sustainability in the Museum Practice.” 2021. [https://cimam.org/documents/159/CiMAM Toolkit on Environmental Sustainability in the Museum Practice. 2021.pdf](https://cimam.org/documents/159/CiMAM%20Toolkit%20on%20Environmental%20Sustainability%20in%20the%20Museum%20Practice.%202021.pdf).
- [6] Colette, Augustin. 2007. “Cases Studies on Climate Change and World Heritage.” 4. UNESCO World Heritage Centre.
- [7] Craig, Jack. 2021. “Making the V&A’s Collections More Discoverable Online.” *V&M Blog* (blog). March 31, 2021. <https://www.vam.ac.uk/blog/digital/making-the-vas-collections-more-discoverable-online>.
- [8] Crawford, Alec. 2020. “The Environmental Consequences of COVID-19 in Fragile State.” *International Institute for Sustainable Development* (blog). June 9, 2020. <https://www.iisd.org/articles/impact-covid-fragile-states>.

⁶ <https://cimam.org/news-archive/museums-want-to-go-green/>.

⁷ <http://www.icom-italia.org/icom-at-g20-culture-president-garlandinis-speech/>.

- [9] Nash, Nick, Lorraine Whitmarsh, Stuart Capstick, et al. 2020. "Local Climate Change Cultures: Climate-Relevant Discursive Practices in Three Emerging Economies." *Climatic Change* 163: 63–82.
- [10] Pop, Izabela L., Anca Borza, Anuta Buiga, Diana Ighian, and Rita Toader. 2019. "Achieving Cultural Sustainability in Museums: A Step Toward Sustainable Development." *Sustainability* 11 (4): 970.
- [11] Sesana, Elena, S. Alexandre Gagnon, Chiara Ciantelli, et al. 2021. "Climate Change Impacts on Cultural Heritage: A Literature Review." *Wiley Interdisciplinary Reviews: Climate Change* 12 (4).
- [12] Sutton, Sarah. 2020. "The Evolving Responsibility of Museum Work in the Time of Climate Change, Museum Management and Curatorship." *Museum Management and Curatorship* 35:6: 618–35.

La Digitalizzazione per una fruizione del Patrimonio Culturale in sito e da remoto: il caso studio della Pala Gozzi di Tiziano

Paolo Clini¹, Renato Angeloni¹, Mirco D'Alessio¹, Elena Marinelli¹

¹ Università Politecnica delle Marche, Italia

ABSTRACT

La Digitalizzazione costituisce un'azione imprescindibile per rendere il Patrimonio Culturale accessibile ovunque e da chiunque, permettendone una documentazione scientificamente rigorosa e lo sviluppo di soluzioni efficaci per la sua fruizione. Il presente contributo si inserisce in questo ambito, descrivendo il processo di riproduzione digitale della Pala Gozzi di Tiziano e la realizzazione di un'applicazione *mobile* finalizzata alla sua fruizione secondo due possibili modalità: da remoto, sfruttando diversi contenuti multimediali per la narrazione del dipinto nella sua assenza fisica, ed in sito, ampliando il godimento dell'originale grazie alla realtà aumentata (*Augmented Reality*, AR). Il caso di studio affrontato ha permesso dunque di definire un protocollo speditivo e sostenibile per la riproduzione digitale di dipinti, volta in particolare ad una loro disseminazione, personalizzabile nella modalità e nella narrazione secondo le finalità auspiccate dal conservatore dell'opera.

PAROLE CHIAVE

Riproduzione digitale, fruizione da remoto, fruizione in sito, applicazione *mobile*, realtà aumentata.

INTERVENTO

1. INTRODUZIONE

L'avvento del digitale ha generato un fondamentale cambio di paradigma nello scenario culturale globale, spostando il focus dell'istituzione museale dall'oggetto in sé al visitatore. Ricorrendo a strumenti divenuti di uso comune, come *smartphone* o *tablet*, questa tecnologia offre infatti la possibilità di disseminare il Patrimonio Culturale tramite la sua riproduzione digitale, permettendo di raggiungere in maniera capillare nuovi pubblici e di declinare l'offerta culturale in ragione di questi. Tale copia diviene così elemento fondante per nuove forme di narrazione e modalità di accesso ai contenuti culturali, definendo soluzioni centrate sull'utente in grado di superare barriere fisiche e temporali, nonché di connettere comunità diverse e dialogare con le nuove generazioni. Le potenzialità della riproduzione digitale permettono così di sviluppare tanto una fruizione da remoto dell'opera, individuata nel presente contributo come preludio all'esperienza dell'originale, quanto in sito, offrendo una mediazione nel rapporto con il dipinto finalizzata a guidare il visitatore nella sua lettura.

2. LA DIGITALIZZAZIONE COME STRUMENTO DI FRUIZIONE

Per mettere in atto un'efficace strategia di trasformazione digitale nei musei è necessario promuovere una cultura del digitale che permetta di ripensare in chiave strategica i beni culturali. La sfida risiede dunque nel comprendere come costruire un nuovo patrimonio e come crearne nuove forme di fruizione e nuove narrazioni, partendo dal presupposto che la presenza fisica delle opere non è sufficiente a raggiungere l'obiettivo fondamentale dell'istituzione museale: trasmettere cultura ai visitatori ([1]). L'ampliamento del pubblico e il miglioramento dell'accessibilità al Patrimonio passano dunque attraverso lo sviluppo di nuove strategie comunicative e fruizioni innovative, in grado di coniugare in maniera efficace educazione e intrattenimento, digitale e analogico ([8]). L'esigenza di accelerare su questi temi è stata ulteriormente evidenziata dal diffondersi della pandemia Covid-19, che ha costretto i luoghi della Cultura a ricorrere al digitale come unico canale di comunicazione e fruizione. Dalle indagini condotte a livello nazionale ed europeo durante il *lockdown* del 2020 è però emerso un quadro piuttosto sconsolante: solo il 43,6% delle collezioni museali è stato digitalizzato e di questo addirittura meno della metà è disponibile online ([5]), una situazione prevedibile se si considera che nel periodo pre-pandemia solo il 24% dei musei aveva redatto un piano strategico dell'innovazione digitale (il 6% come documento dedicato e il 18% all'interno di un più generale piano strategico) ([6]). Come individuato nel Piano Triennale per la Digitalizzazione e l'Innovazione dei Musei, occorre dunque ripartire dalla strutturazione di una diffusa campagna di digitalizzazione del Patrimonio Culturale, la ricostruzione di luoghi e oggetti mediante modelli digitali si presta infatti ad

una pluralità di obiettivi, coniugando la necessità di documentazione alle esigenze di fruizione e comunicazione – anche emozionale – per il pubblico più in generale ([7]). La Comunità scientifica deve dunque supportare le Istituzioni museali definendo processi che, mantenendo rigore metodologico e misurabilità dei risultati, permettano anche a non esperti di elaborare esatte riproduzioni digitali del Patrimonio Culturale ([2]). A partire da queste, numerose sono poi le modalità tecnologiche di fruizione che le stesse Istituzioni potranno offrire ([3]) a seconda ad esempio della scelta di proporre una fruizione da remoto, come previsto ad esempio per il Palazzo Ducale di Urbino dall'app *Ducale*¹, o in sito, come per i Musei Reali di Torino tramite l'app *MRT*². Il presente contributo si inserisce così in questo ambito definendo in primis un processo di acquisizione ed elaborazione dati eseguibile anche da operatori museali non esperti di digitalizzazione, quindi proponendone una soluzione di fruizione a misura del pubblico dei visitatori.

3. LA DIGITALIZZAZIONE DELLA PALA GOZZI

La Pala Gozzi di Tiziano, prima opera datata e firmata dal Maestro veneto ([11]) è oggi conservata presso la Pinacoteca Civica F. Podesti di Ancona (**Figura 1**).



Figura 1 Tiziano, Pala Gozzi. Immagine dell'opera dal Catalogo Beni Culturali della Regione Marche³.

Il dipinto «costituisce il punto fermo, quasi una svolta dell'attività giovanile di Tiziano, che subito dopo dà il via ad una attività frenetica ed insieme di incredibile creatività in continuo superamento» ([10]). Commissionata per la chiesa di San Francesco ad Alto dal mercante raguseo Alvise Gozzi, la raffigurazione si presta ad una chiara lettura in chiave politica: la Vergine, dipinta sopra la veduta di San Marco, rappresenta la Serenissima, cui rendono omaggio San Francesco per la città di Ancona e San Biagio protettore di Ragusa; una celebrazione della superiorità di Venezia rispetto ai due importanti porti del medio-Adriatico ([9]).

La digitalizzazione dell'opera è stata condotta ricorrendo esclusivamente ad acquisizioni fotografiche in luce visibile diffusa, finalizzate ad una ricostruzione tridimensionale della struttura dell'oggetto e alla sua *texturizzazione* ad alta risoluzione, sia per il fronte che per il retro. Le operazioni di ripresa sono state quindi pianificate ed eseguite con la finalità di garantire una documentazione scientificamente corretta della forma dell'oggetto, della sua morfologia superficiale e delle sue qualità cromatiche, inserendo nella scena dei riferimenti metrici ed un *colorchecker*.

Viste le dimensioni dell'opera (312x215 cm), al fine di ottimizzare i tempi di acquisizione, sono state pianificate due distinte fasi di ripresa (**Tabella 1**). La prima è consistita in un'acquisizione multi-scatto realizzata illuminando l'intera pala con pannelli led dotati di diffusori, per un totale finale di circa 150 immagini per lato, integrate con ulteriori scatti dei

¹ <https://play.google.com/store/apps/details?id=it.univpm.dii.Ducale&hl=it&gl=US>.

² <https://play.google.com/store/apps/details?id=it.coopculture.app.mrt&hl=it&gl=US>.

³ http://sirpac.cultura.marche.it/sirpacinraweb/storage/label/1409/384/00115783_022.jpg.

fianchi della cornice. Tale ripresa è stata eseguita utilizzando una fotocamera full-frame posizionata a 2 m dal dipinto, ottenendo una dimensione del pixel nella realtà (*Ground Sampling Distance*, GSD) e quindi un dettaglio massimo pari a 0,133 mm.

Dimensioni sensore		Dimensioni Pala Gozzi			
Larghezza	24 mm	Larghezza	2,15 m		
Altezza	36 mm	Altezza	3,12 m		
Progetto di presa Acquisizione n.1					
Distanza di presa	2 m	Sidelap	60%	Totale foto nadirali	110
Distanza focale	90 mm	Overlap	60%	Totale foto oblique	32
Ricoprimento		Spostamento		Dimensioni pixel	
Larghezza	0,533 m	Orizzontale	0,213 m	Larghezza	0,1333 mm
Altezza	0,8 m	Verticale	0,320 m	Altezza	0,1333 mm
Progetto di presa Acquisizione n.2					
Distanza di presa	0,56 m	Sidelap	25%	Totale foto nadirali	870
Distanza focale	90 mm	Overlap	25%	Totale foto oblique	-
Ricoprimento		Spostamento		Dimensioni pixel	
Larghezza	0,096 m	Orizzontale	0,072 m	Larghezza	0,024 mm
Altezza	0,144 m	Verticale	0,108 m	Altezza	0,024 mm

Tabella 1 Progetti di presa relativi alle due acquisizioni fotografiche eseguite.

La seconda fase di ripresa è stata invece eseguita posizionando la fotocamera a 56 cm di distanza dal dipinto, garantendo così un rapporto di ingrandimento di 1:4. La medesima fotocamera, dotata di obiettivo macro a distanza focale fissa pari a 90 mm e filtro polarizzatore, è stata montata su uno slider fotografico di lunghezza 240 cm, atto a garantire lo spostamento orizzontale automatico tra scatti successivi. A tale movimento è stato reso solidale un sistema di illuminazione costituito da due pannelli led posti ai lati dell'obiettivo, dotati anch'essi di filtro polarizzatore. Si sono così acquisite circa 870 immagini per ciascun lato, per una GSD finale pari a 0,024 mm.

Il primo set di immagini è stato quindi impiegato nel processo di ricostruzione tridimensionale della pala. Ricorrendo al software *RawTherapee*⁴, i singoli scatti, acquisiti in formato RAW, sono stati post-prodotti eseguendone la correzione del colore grazie al profilo elaborato mediante *ColorChecker Camera Calibration*⁵, e quindi salvati in formato JPG. I file ottenuti sono stati importati nel software per l'elaborazione fotogrammetrica di immagini digitali *Meshroom*⁶, ottenendo tramite il *workflow* di elaborazione automatica il modello 3D della pala. Grazie ai riferimenti metrici inseriti nella scena, è stato possibile scalare il modello portandolo alla dimensione dell'oggetto reale. Si è quindi proceduto all'ottimizzazione di tale primo output, al fine di garantirne un'adeguata fruizione anche all'interno di applicazioni per dispositivo mobile, ricorrendo al software di modellazione 3D e rendering *Blender*⁷ si sono quindi eseguite le operazioni di pulizia, *retopology*, *unwrap* e *baking* delle normali, passando da un modello 3D da 1,5 milioni di triangoli ad una da 100 mila quadrati.

Il secondo set di immagini è stato invece impiegato per la *texturizzazione* ad alta risoluzione, del modello precedentemente elaborato ed ottimizzato. Anche in questo secondo caso le immagini sono state post-prodotte e salvate in JPG. Quindi processate nel medesimo software per l'elaborazione fotogrammetrica in modo da essere orientate nello spazio e riferite al medesimo sistema di coordinate spaziali del modello 3D precedentemente realizzato. La proiezione di tali immagini sulla geometria ottimizzata ha quindi permesso di texturizzarne la superficie e di elaborare due distinte ortoeffimmagini, una raffigurante il fronte, l'altra il retro della pala. (**Figura 2**).

4. UNA PROPOSTA DI FRUIZIONE PER LA VISITA DA REMOTO E IN SITO

La proposta di fruizione per la Pala Gozzi di seguito presentata si basa sulle potenzialità della riproduzione digitale per il racconto dell'opera d'arte, qui veicolato tramite un'applicazione *mobile*. Modello 3D e immagini digitali sono stati impiegati come punto di partenza per l'elaborazione di contenuti di qualità, scientificamente corretti e adatti ad un pubblico generalista. La narrazione è stata strutturata in maniera svincolata dalle logiche didascaliche museali, coniugando il potere coinvolgente dello *storytelling* al rigore disciplinare scientifico, ed immaginando due distinti scenari di fruizione: un'esperienza da remoto ed una in sito. Selezionando nella schermata principale *Sei a casa?*, sarà avviata la visita virtuale

⁴ https://rawpedia.rawtherapee.com/Main_Page

⁵ <https://xritephoto.com/CameraCalibration>

⁶ <https://alicevision.org/#>

⁷ <https://www.blender.org/>

dell'opera articolata in tre sezioni: *Esplora*, *Scopri* e *Virtual Tour* (Figura 3). *Esplora* permette di navigare in totale libertà la riproduzione digitale della Pala, di apprezzarne i dettagli dipinti e la tridimensionalità. La sezione *Scopri* costituisce invece la principale proposta interattiva di carattere narrativo, strutturando un racconto coinvolgente ed evocativo grazie a brevi video realizzati a partire dalle ortoimmagini elaborate e focalizzati sui personaggi presenti nella scena, sulla sua



Figura 2 Workflow seguito per la digitalizzazione della Pala Gozzi, vista frontale del modello 3D e dettaglio dell'ortoimmagine sul volto del Gesù Bambino, la cui visualizzazione a grandezza reale presenta una risoluzione di oltre 1000 dpi.

composizione geometrica, sul paesaggio e i suoi dettagli, sulla tecnica esecutiva e i colori utilizzati. La sezione *Scopri* costituisce invece la principale proposta interattiva di carattere narrativo, strutturando un racconto coinvolgente ed evocativo grazie a brevi video realizzati a partire dalle ortoimmagini ad alta risoluzione precedentemente elaborate e focalizzati sui diversi personaggi presenti nella scena, sulla sua composizione geometrica, sul paesaggio e i suoi dettagli, sulla tecnica esecutiva e i colori utilizzati. Infine, *Virtual Tour* permette di esplorare a 360° la Sala Zampetti, luogo in cui l'opera è oggi esposta al pubblico, offrendo al visitatore la possibilità di scoprire brevi notizie anche di altri dipinti, in particolare la Crocifissione, opera dello stesso Tiziano esposta di fronte alla Pala Gozzi in un interessante dialogo tra la sua arte in giovane età e quella della sua maturità.

La selezione *Sei al museo?* permette invece di avviare un'esperienza di fruizione in AR del dipinto, concepita come strumento digitale di supporto per la visita fisica presso la pinacoteca. L'utilizzo del dispositivo e della tecnologia AR non è pensato per sostituire l'occhio umano durante la visita, volendo preservare quel circuito emozionale in cui il visitatore entra di fronte all'artefatto fisico, il godimento di quell'*hic et nunc*, quella sua esistenza irripetibile nel luogo in cui si trova ([4]). L'obiettivo di questa sezione è invece quello di fornire uno sguardo aggiuntivo, di accompagnare l'utente nella comprensione dell'opera permettendo di individuare e svelare dettagli che normalmente sfuggirebbero all'occhio umano. Il rapporto tra utente e opera rimane dunque centrale, venendo però ulteriormente supportato da contenuti multimediali che si stratificano al dipinto.

5. CONCLUSIONI E SVILUPPI FUTURI

La soluzione digitale presentata si propone come contributo per ampliare e implementare le possibilità di fruizione di una tra le più importanti opere custodite presso la Pinacoteca Civica di Ancona.

Nel suo complesso essa nasce dal concertamento di azioni messe in atto da professionalità afferenti a diversi ambiti disciplinari: il processo di digitalizzazione condotto da esperti di acquisizione ed elaborazione fotogrammetrica, quello di strutturazione della proposta di fruizione e realizzazione dei contenuti ad opera di esperti di storytelling e storici dell'arte ed infine lo sviluppo informatico eseguito da figure competenti in tale ambito.

Per il processo di riproduzione digitale, si è definito un workflow basato esclusivamente sull'uso di software gratuiti, una soluzione ripetibile e applicabile ad altre opere della medesima collezione, eseguibile in maniera autonoma anche dal personale interno della Pinacoteca e che permetterà quindi di elaborare una serie di riproduzioni digitali utili a strutturare un *corpus* di racconti digitali destinato in primis alla Comunità cui appartiene quel Patrimonio.

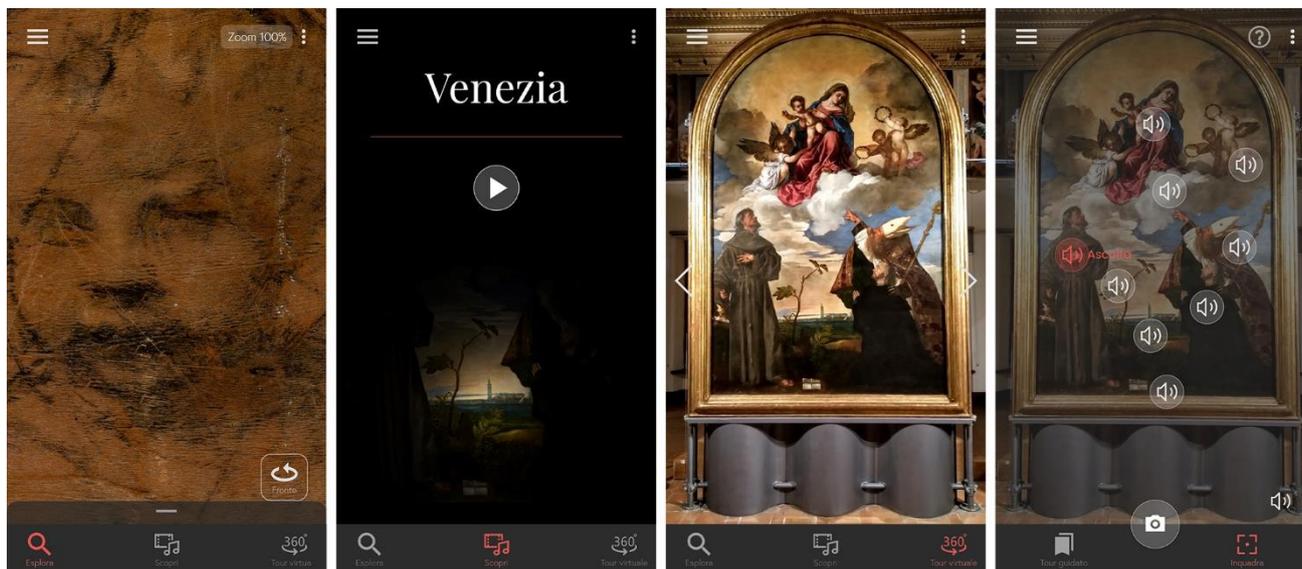


Figura 3 Vista delle schermate relative alle 3 sezioni dedicate alla fruizione da remoto e alla schermata di fruizione in AR.

Uno degli obiettivi dell'azione proposta è infatti quello di incuriosire e riavvicinare la cittadinanza alle proprie opere d'arte e per questo si tenterà in futuro di attivare ulteriori sinergie mirate a rendere la Comunità stessa partecipe del processo creativo di creazione dei contenuti.

Il presente lavoro proseguirà quindi con la ricerca di soluzioni per lo sviluppo informatico di modalità di fruizione digitale realizzabili anche da non esperti, tentando una standardizzazione supportata da una più approfondita valutazione della *user experience*, per rendere il personale della Pinacoteca in condizione di raccontare il Patrimonio custodito sfruttando tecnologie e modalità che si riterranno maggiormente appropriate a perseguire gli obiettivi dell'Istituzione.

6. RINGRAZIAMENTI

Si ringrazia il Comune di Ancona e la Pinacoteca Civica "F. Podesti": l'Assessore alla Cultura, al Turismo e alle Politiche Giovanili Paolo Marasca, il Curatore Stefano Zuffi e la Dott.ssa Maria Vittoria Carloni, per aver messo a disposizione l'opera e la sala, permettendo le necessarie riprese fotografiche.

BIBLIOGRAFIA

- [1] Antinucci, Francesco. 2014. *Comunicare nel museo*. Nuova edizione riveduta e Aggiornata. Roma; Bari: Laterza.
- [2] Apollonio, Fabrizio Ivan, Filippo Fantini, Simone Garagnani, e Marco Gaiani. 2021. «A Photogrammetry-Based Workflow for the Accurate 3D Construction and Visualization of Museums Assets». *Remote Sensing* 13 (3): 486.
- [3] Bekele, Mafkereseb Kassahun, Roberto Pierdicca, Emanuele Frontoni, et al. 2018. «A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage». *Journal on Computing and Cultural Heritage* 11 (2): 36.
- [4] Benjamin, Walter. 1936. *The Work of Art in the Age of Mechanical Reproduction*.
- [5] «Digitisation and IPR in European Museums». 2020. NEMO. https://www.nemo.org/fileadmin/Dateien/public/Publications/NEMO_Final_Report_Digitisation_and_IPR_in_European_Museums_WG_07.2020.
- [6] «Musei, monumenti e aree archeologiche italiani che hanno formalizzato un Piano Strategico per l'Innovazione Digitale, 2020». 2020. <https://www.osservatori.net/it/prodotti/formato/grafici/musei-monumenti-aree-archeologiche-italiani-piano-strategico-innovazione-digitale-dettaglio-nuove-opportunita-esigenze-2021-grafico>.
- [7] «Piano Triennale per la Digitalizzazione e l'Innovazione dei Musei». 2018. Direzione generale Musei. <http://musei.beniculturali.it/wp-content/uploads/2019/08/Piano-Triennale-per-la-Digitalizzazione-e-l%E2%80%99Innovazione-dei-Musei.pdf>.
- [8] Pietroni, Eva. 2019. «Experience Design, Virtual Reality and Media Hybridization for the Digital Communication Inside Museums». *Applied System Innovation* 2 (4).

- [9] Zampetti, Pietro. 1988. «Tiziano Fino al 1520». In *Tiziano. La Pala Gozzi Di Ancona. Il Restauro e Il Nuovo Allestimento Espositivo*, 11–22. Bologna: Grafis Edizioni.
- [10] Zuffi, Stefano. 2008. *Tiziano*. Milano: Mondadori Electa.
- [11] ———. 2017. *Tiziano. Sacra Conversazione*. Milano: 24 Ore Cultura.

Sessione Testi 1
Claude Shannon

Verso la definizione di criteri per valutare soluzioni di scholarly editing digitale: il caso d'uso GreekSchools

Simone Zenaro¹, Angelo Mario Del Grosso², Federico Boschetti³, Graziano Ranocchia⁴

¹ Istituto di Linguistica Computazionale “A. Zampolli” - CNR, Italia - simone.zenaro@ilc.cnr.it

² Istituto di Linguistica Computazionale “A. Zampolli” - CNR, Italia - angelo.delgrosso@ilc.cnr.it

³ Istituto di Linguistica Computazionale “A. Zampolli” - CNR & VePDH, Italia - federico.boschetti@ilc.cnr.it

⁴ Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Italia - graziano.ranocchia@unipi.it

ABSTRACT

Il contributo propone alcuni criteri di valutazione per l'analisi degli approcci all'editing testuale. Dopo averli presentati brevemente, si mostra come sono stati applicati alle attività finora svolte nel contesto del progetto ERC-885222 GreekSchools per quanto riguarda la progettazione e lo sviluppo di una piattaforma web collaborativa che fa uso di linguaggi specifici di dominio (DSL) finalizzata allo studio scientifico e alla pubblicazione di testi papiracei mediante edizioni critiche digitali.

PAROLE CHIAVE

DSL, linguaggi specifici di dominio, filologia computazionale, ingegneria del software, papirologia digitale, Domain Driven Design.

INTERVENTO

1. INTRODUZIONE

La filologia tradizionale ha definito i propri metodi per lo studio dei testi che, nel tempo, si sono cristallizzati in processi e metodologie rigorose seppure con le differenze dovute al contesto di riferimento. D'altro canto, l'avvento della filologia computazionale sta cambiando il paradigma di lavoro tradizionale proponendo nuove modalità di analisi: ne è un esempio la codifica dei testi tramite le linee guida TEI. In questo contesto è interessante indagare il rapporto fra il rigore dei metodi tradizionali e gli approcci computazionali. Spesso i filologi avvertono il cambio di paradigma come un'imposizione e gli strumenti tecnologici disponibili come inadeguati o frustranti ([4];[13];[16]). Questa separazione fra le necessità del filologo e la disponibilità di strumenti, non sempre flessibili, ostacola la sinergia tra conoscenza, metodi e strumenti. Ci proponiamo di contribuire alla convergenza di modelli a effettivo supporto per il lavoro dei filologi attraverso la realizzazione di una piattaforma per lo studio scientifico del testo. Le nostre attività di ricerca si sviluppano in seno al progetto “ERC 885222-GreekSchools, The Greek philosophical schools according to Europe's earliest history of philosophy” il cui oggetto di studio è un prezioso trattato: *la Rassegna dei filosofi* di Filodemo di Gadara (75-50 a.C.) – conosciuto anche come *Sýntaxis* – i cui papiri originali sono in cattive condizioni e le cui edizioni disponibili sono superate. Di conseguenza, GreekSchools si prefigge di realizzare una nuova edizione dell'intero trattato munita di trascrizione paleografica e letteraria, dei corrispondenti apparati, di commenti e traduzioni.

2. STATO DELL'ARTE

Nonostante le numerose iniziative in ambito DH, gli esperti di discipline filologiche non possiedono ancora strumenti digitali e computazionali del tutto efficaci. Alcuni esempi di strumenti sviluppati per la pubblicazione e l'analisi di edizioni digitali sono Textual Communities¹, TEI Publisher², EVT³, Voyant Tools⁴, MQDQ⁵ mentre iniziative quali Canonical Text Services⁶ (CTS) e CollateX⁷ hanno definito protocolli condivisi all'interno della comunità degli studiosi. Si sono

¹ <https://textualcommunities.org/>.

² <https://teipublisher.com/>.

³ <http://evt.labcd.unipi.it/>.

⁴ <https://voyant-tools.org/>.

⁵ <http://mqdq.it>.

⁶ http://cite-architecture.github.io/ctsum_spec/.

⁷ <https://collatex.net/>.

avvicinate nel corso degli anni iniziative quali Bambo⁸, Interedition⁹ e DiXiT¹⁰, con l'obiettivo di sviluppare metodi e strumenti utili agli studi umanistici. Oggigiorno, lo sforzo maggiore, finalizzato a organizzare e gestire strumenti, risorse, servizi nonché attività di formazione, ricade su iniziative infrastrutturali quali DARIAH¹¹, CLARIN¹², Parthenos¹³. Tra queste, particolare interesse richiama il progetto SSHOC (Social Sciences & Humanities Open Cloud)¹⁴ poiché si colloca all'interno degli obiettivi dell'European Open Science Cloud. Tra le iniziative di riferimento nell'ambito della papirologia digitale, ricordiamo il progetto Papyri.info¹⁵, che non solo fornisce utili strumenti per la redazione di testi a supporto del processo editoriale con successiva gestione della persistenza dei dati, ma richiama soprattutto un'attitudine progettuale e di sviluppo volta ad includere la comunità di riferimento, rispondendo puntualmente alle sue necessità. Fra le caratteristiche notevoli di Papyri.info ci sono il tracciamento degli accessi e delle modifiche al documento, l'utilizzo di un Linguaggio Specifico di Dominio (DSL) ([15];[14];[8];[2]) per l'edizione dei testi (Leiden+) e un avanzato motore di ricerca. Nonostante ciò, crediamo che ci siano margini di miglioramento per quanto riguarda: (a) la possibilità di elaborare dati testuali, (b) il supporto alla collaborazione in editing, (c) l'utilizzo di molteplici DSL diversi da Leiden+ (e.g. per la codifica degli apparati), (d) l'integrazione delle fonti facsimilari, (e) l'usabilità del software.

3. COPHIEDITOR

Al fine di supportare il lavoro dei papirologi nella realizzazione dell'edizione della *Syntaxis* stiamo sviluppando uno strumento denominato CoPhiEditor: una piattaforma web collaborativa basata su DSL per la creazione di edizioni scientifiche digitali papirologiche. Un DSL è un linguaggio definito da una grammatica formale (usualmente una *Context-Free Grammar*) e progettato per un ambito circoscritto di conoscenza o di attività. I costrutti e il lessico dei linguaggi (o metalinguaggi) general-purpose, come Python, o di tipo descrittivo, come XML, si allontanano spesso notevolmente dal linguaggio naturale o da formalismi familiari agli specialisti di discipline umanistiche ([11]). Accade così che il filologo digitale si trovi a codificare i testi in modo alquanto lontano dalle pratiche del filologo tradizionale: si pensi per esempio a un apparato critico rappresentato in XML-TEI e a un apparato critico redatto da un papirologo.

Uno dei vantaggi maggiori di un DSL è dunque la familiarità con i formalismi adottati all'interno di un dominio di conoscenza che vanta una propria tradizione di studi che ha permesso, nel corso di decenni se non di secoli, di ottimizzare la rappresentazione dell'informazione pertinente all'oggetto di studio. Ciò comporta la compattezza di un DSL rispetto a un linguaggio general-purpose, perché – come stabilisce la teoria dell'informazione di Shannon – ciò che occorre più frequentemente può essere codificato con un numero minore di caratteri (tramite simboli o abbreviazioni perfettamente comprensibili e familiari allo specialista) mentre ciò che occorre più di rado richiede una codifica più verbosa. Abbiamo deciso di utilizzare un approccio basato su DSL dopo aver analizzato diverse soluzioni per l'editing testuale (cfr. Fig. 1) rispetto a questi sei differenti dimensioni: familiarità (*familiarity*), grado minimo di formazione tecnica (*transparency*), completezza (*completeness*), compattezza (*compactness*), coerenza (*consistency*), supporto all'elaborazione dei dati (*actionability*). Queste dimensioni sono, in parte, riprese dai concetti espressi in ([3];[17];[5]). Con *familiarity* ci riferiamo alla possibilità per lo studioso di mantenere il suo paradigma di lavoro abituale, mentre *transparency* indica la quantità di conoscenze tecniche necessarie per adottare l'approccio proposto (in ([10]) si descrive un lavoro di indagine preliminare rispetto a questi due temi). La *completeness* si riferisce alla quantità d'informazione che è possibile esprimere (per approfondimenti si veda anche ([12])). Mentre la *compactness* si riferisce al rapporto fra la *completeness* e l'estensione della formalizzazione (per riferimenti puntuali si rinvia a ([1];[18])). Con *consistency*, invece, valutiamo la coerenza nel descrivere o trattare gli stessi fenomeni nel medesimo modo; ciò implica che la rappresentazione di una stessa tipologia di informazione sia univoca e quindi non ambigua. Questo non è sempre garantito da tutti le soluzioni considerati, ad esempio è possibile codificare un'abbreviazione in XML-TEI in diversi modi nello stesso contesto (adottando gli elementi *choice*, *ex*, *expan*, etc.) senza la garanzia di utilizzare la medesima convenzione editoriale. Infine, la capacità di estrarre o dedurre informazioni a partire dai dati viene indicata dalla dimensione *actionability* (caratteristica intrinseca nei linguaggi formali, descritti da una grammatica formale e corredati comunemente da altri componenti per l'elaborazione del codice quali il lexer e il parser). Come mostrato in Fig.1, solo basandosi sui DSL si massimizzano tutte e sei le dimensioni. La Fig.1 potrebbe far pensare che adottare DSL sia la scelta più ovvia in ogni contesto e risolva ogni limite presente negli altri

⁸ <https://www.projectbamboo.org/>.

⁹ <http://www.interedition.eu/>.

¹⁰ <https://cordis.europa.eu/project/id/317436>.

¹¹ <https://www.dariah.eu/>.

¹² <https://www.clarin.eu/>.

¹³ <http://www.parthenos-project.eu/>.

¹⁴ <https://sshopencloud.eu/>.

¹⁵ <https://papyri.info/>.

approcci. Ovviamente non è così, i limiti dei DSL, nel nostro caso d'uso, rappresentano degli ostacoli marginali pur mantenendo tutte le caratteristiche che riteniamo importanti in una piattaforma di editing efficace per testi papirologici. Sembra dunque opportuno confrontare alcuni punti deboli dei DSL rispetto agli altri approcci. Un DSL, essendo espresso in plain text, deve rinunciare alla semantica attribuita alla formattazione come ad esempio il grassetto, adottato in molte pratiche editoriali dai filologi che usano Word Processor (ad esempio per indicare le *litterae suppositae* nell'apparato diplomatico). Una soluzione, meno elegante ma altrettanto efficace, è l'uso di segni minimi di apertura e chiusura, come avviene nel linguaggio markdown (ad es. *grassetto*). Un DSL apparentemente fornisce meno controllo sull'inserimento del testo, tuttavia un buon componente di editing per linguaggi formali (ad es. Monaco, Ace, CodeMirror) non solo è in grado di segnalare gli errori sintattici, ma fornisce suggerimenti per la loro risoluzione e meccanismi di autocompletamento. Infine, un DSL ha un'espressività più limitata rispetto a uno schema di marcatura complesso come XML-TEI. Tuttavia questo è un punto debole solo in apparenza. Infatti un DSL nasce dall'esigenza di individuare un contesto specifico in cui il linguaggio permetta di descrivere agevolmente, in maniera compatta e coerente tutte le peculiarità del proprio dominio, ma può essere esteso con moduli aggiuntivi così come accade con XML-TEI.

	Word Processor	Structured Text	GUI-centric	DSL
Familiarity	✓			✓
Transparency	✓		✓	✓
Completeness		✓	✓	✓
Compactness	✓			✓
Consistency			✓	✓
Actionability		✓	✓	✓

FIG. 1: Analisi degli approcci all'editing testuale

In GreekSchools, il DSL viene progettato dai ricercatori a stretto contatto con gli esperti papirologi, in quanto il linguaggio deve rappresentare, senza ambiguità ma in modo vicinissimo alle pratiche tradizionali, le informazioni necessarie a redigere l'edizione diplomatica e l'edizione letteraria del testo dei papiri e dei rispettivi apparati. Questa forma di co-design richiede continui incontri durante tutto il processo di sviluppo della piattaforma.

Per questo scopo abbiamo deciso di adottare il processo di progettazione e di sviluppo software Domain Driven Design (DDD)¹⁶. Riteniamo, infatti, che i principi e le prassi suggerite dal DDD siano in grado di definire con rigore e in maniera soddisfacente le specificità del metodo filologico tradizionale, arricchendolo, senza forzature, dei vantaggi derivanti dall'impiego delle nuove tecnologie. In particolare il processo di definizione e disambiguazione dei concetti si realizza in un lessico condiviso (*ubiquitous lexicon*) tra esperti di dominio con profilo umanistico e figure tecniche.

Sono previsti quattro DSL per la gestione della trascrizione diplomatica, di quella letteraria e dei rispettivi apparati paleografico e critico. Utilizzare un DSL in questo contesto permette di mantenere il focus principale sul testo senza perdere i vantaggi di un approccio strutturato per la rappresentazione delle informazioni testuali in cui tutti i dati vengono resi espliciti (ad esempio attraverso la codifica XML). Inoltre, la rappresentazione derivata dall'interpretazione automatica del DSL apre alla possibilità di adoperare tecniche complesse di elaborazione dei dati (e.g. analisi statistiche e classificatori semantici, verifiche di coerenza sintattica, etc.).

Nel nostro caso di studio, i dati risultanti dal parsing sono alla base del controllo di coerenza e della verifica di errori dovuti, ad esempio, alla mancata conformità alle convenzioni editoriali, o alle regole di corrispondenza fra trascrizione diplomatica e letteraria, le quali rimarranno fortemente legate, nell'edizione, a criteri di spazialità e interdipendenza. La Fig. 2 mostra la rappresentazione ad albero di un frammento di apparato in cui la parentesi tonda chiusa viene riconosciuta come un errore in quanto non conforme alle regole editoriali definite nella grammatica del DSL. In questo modo è possibile notificare all'editore gli errori perché li risolva, assicurando un risultato più conforme alle sue intenzioni.

¹⁶ Il Domain-Driven Design (DDD) è un processo di progettazione e sviluppo software descritto nel 2003 da Eric Evans ([7]). Il processo si propone di realizzare applicazioni software a partire da modelli che rappresentino nel modo più accurato possibile i concetti e le attività del dominio di riferimento. Per fare questo, l'approccio prevede il continuo coinvolgimento degli esperti di dominio durante le varie fasi dello sviluppo dell'applicazione, concentrandosi sulla definizione di un linguaggio comune chiamato *ubiquitous language*.

Un altro scenario di utilizzo della nostra proposta riguarda l'implementazione di un motore di ricerca avanzato e la possibilità di tradurre automaticamente le informazioni in altri formati (ad esempio XML/TEI) al fine di garantire persistenza e interoperabilità.

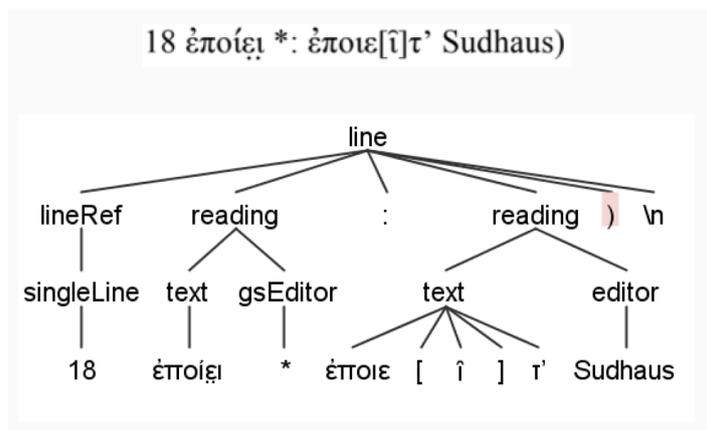


FIG. 2: Immagine DSL su Apparato.

Dal punto di vista del supporto al processo di edizione, CoPhiEditor implementa un workflow basato su sessioni di lavoro gestite dall'editore in cui collaboratori esterni possono commentare i testi e proporre congetture. Tali proposte verranno valutate e, possibilmente, integrate nel testo, realizzando così un processo di revisione continuo e collaborativo (Fig. 3). Nei casi in cui sia possibile proporre automaticamente congetture, la piattaforma sottoporrà una lista di candidati al giudizio dell'editore, implementando modelli computazionali della lingua greca e addestrando opportuni sistemi di predizione.

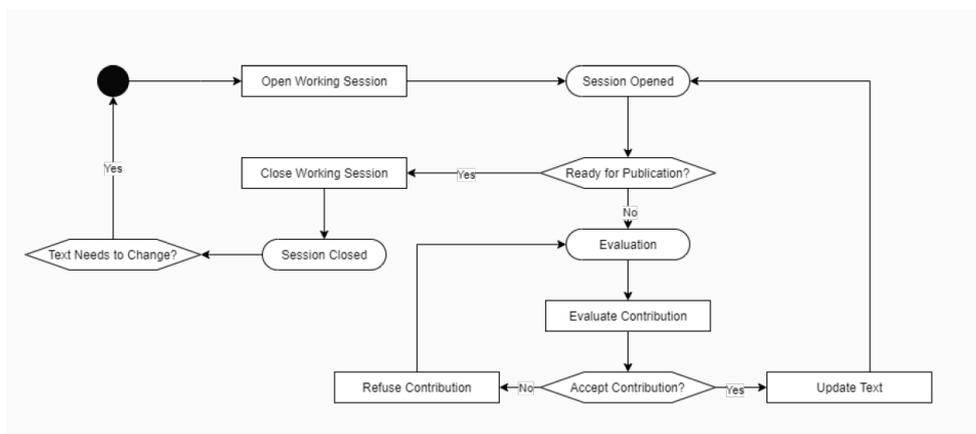


FIG. 3 Revisione continua e collaborativa.

Dal punto di vista dell'interfaccia grafica (GUI), il dialogo con gli esperti di dominio ha permesso di identificare una modalità di lavoro familiare che consiste nella giustapposizione delle fonti con il testo in via di costituzione e il frequente confronto con gli altri testi. Per questo motivo, come mostra il wireframe¹⁷ in Fig. 4, la GUI permette di affiancare qualsiasi numero di fonti ai testi in editing, lasciando l'utente libero di deciderne la disposizione nello spazio. La gestione delle immagini è progettata nel rispetto del protocollo IIF affiancando strumenti di supporto alla loro lettura (e.g. modificando le proprietà di luminosità e contrasto). In questo modo la GUI si adatterà alle abitudini dell'editore. Inoltre, la navigazione della struttura del testo permette di focalizzarsi su qualsiasi livello di granularità (colonna, testo, apparato, riga, entrata d'apparato, etc).

¹⁷ La rappresentazione grafica iniziale utilizzata per identificare la struttura degli elementi di una pagina web.

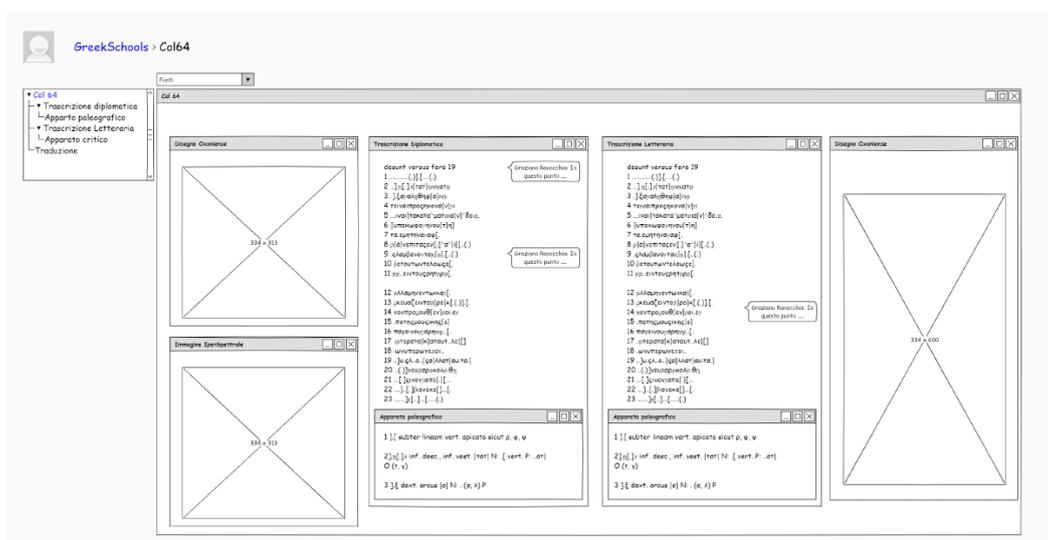


FIG. 4 GUI Risorse integrate e struttura edizione.

Infine, per quanto riguarda l'editing del testo (generalmente suddiviso in colonna), ciascun'area implementa un ambiente basato su Operational Transformation ([6];[9]), una tecnologia utilizzata per fornire editing parallelo e collaborativo, come ad esempio in Google Docs. Un sistema di commenti al testo agevolerà la discussione fra editori e collaboratori esterni. La collaborazione efficace fra studiosi richiede il tracciamento e l'attribuzione di responsabilità degli interventi, la sincronizzazione e la costante verifica della coerenza fra le parti modificate. Un ambiente collaborativo ha lo scopo proprio di agevolare e supportare queste modalità di lavoro nella maniera più trasparente possibile ai soggetti che interagiscono col sistema di editing mediante l'uso di interfacce grafiche dinamiche, flessibili ed ergonomiche progettate per ottenere, al tempo stesso, un'architettura generale ed un'esperienza utente personalizzata.

4. CONCLUSIONI

Abbiamo presentato dei criteri per l'analisi degli approcci all'editing testuale individuando sei dimensioni di valutazione: *familiarity*, *transparency*, *completeness*, *compactness*, *consistency*, *actionability*. La proposta di valutazione che ne deriva supporta la scelta di utilizzare un approccio basato su DSL per quanto riguarda il contesto delle edizioni papirologiche digitali, nel nostro caso di studio il progetto ERC-885222 GreekSchools.

Di conseguenza abbiamo descritto lo stato di avanzamento della progettazione di CoPhiEditor: una piattaforma web collaborativa basata su DSL per la creazione di edizioni scientifiche digitali papirologiche. L'adozione della metodologia DDD risulta di particolare rilievo in quanto si allinea con la volontà di rimanere fedeli ai metodi filologici tradizionali principalmente per quanto riguarda la definizione di un lessico comune ad esperti di dominio, ricercatori e sviluppatori software: l'*ubiquitous language*. Questa metodologia aiuta quindi a modellare e comprendere a fondo i concetti di dominio. Tra le altre caratteristiche della piattaforma abbiamo individuato: il supporto all'editing collaborativo e alla ricerca sui dati, la definizione di sessioni di lavoro, l'integrazione delle fonti testuali e facsimilari, il supporto alla gestione delle congetture e alla coerenza delle scelte editoriali.

Il lavoro presentato si svilupperà in due direzioni: la creazione di un prototipo della piattaforma che racchiuda tutte le caratteristiche elencate e la generalizzazione del framework di valutazione per approcci all'editing testuale digitale. L'interdipendenza tra ricerca umanistica e ricerca informatica è uno dei maggiori punti di forza di questo progetto, perché si basa sul concetto di coevoluzione, in cui tutte le parti in causa traggono reciproco vantaggio: il processo di edizione, pur rimanendo molto vicino alle pratiche tradizionali (*familiarity*), è potenziato dalle possibilità di esplorazione del testo fornite dai mezzi digitali sia in fase di stesura degli apparati che di studio successivo (*actionability*); il processo di progettazione e sviluppo del software, pur applicando le metodologie e le prassi allo stato dell'arte, è potenziato dallo studio e dal trattamento complesso del linguaggio naturale, dell'incertezza, dell'ambiguità e dei dati lacunosi.

BIBLIOGRAFIA

- [1] Bambaci, Luigi, e Federico Boschetti. s.d. «Encoding the Critical Apparatus by Domain Specific Languages: The Case of the Hebrew Book of Qohelet». In *La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*, a cura di Marco Passarotti, Cristina Marras, e Eleonora Litta. Quaderni di Umanistica Digitale. Milano: Università Cattolica del Sacro Cuore.

- [2] Boersma, Meinte. 2021. *Domain-Specific Languages Made Easy*. Manning Publications Co.
- [3] Boschetti, Federico, e Angelo Mario Del Grosso. 2020. «L'annotazione di testi storico-letterari al tempo dei social media». *Italica Wratislaviensia* 11 (1): 65–99.
- [4] Bradley, John. 2015. «How about Tools for the whole range of scholarly activities». Sydney, Australia.
- [5] Buzzetti, Dino. 2002. «Digital Representation and the Text Model». *New Literary History* 33 (1): 61–87.
- [6] Chengzheng, Sun, e Clarence Ellis. 1998. «Operational transformation in real-time group editors: issues, algorithms, and achievements». In *Proceedings of the 1998 ACM conference on Computer supported cooperative work (CSCW '98)*. Association for Computing Machinery, New York, NY, USA, 59–68.
- [7] Evans, Eric. 2004. *Domain-driven design: tackling complexity in the heart of software*. Addison-Wesley Professional.
- [8] Fowler, Martin. 2010. *Domain-Specific Languages*. Addison-Wesley Signature Series. Pearson Education.
- [9] Mehdi, Ahmed-Nacer, Pascal Urso, Valter Balegas, e Nuno Pregoça. 2014. «Merging OT and CRDT algorithms». Amsterdam: Association for Computing Machinery.
- [10] Melighetti, Francesco. 2021. «Domain-Specific Languages per facilitare annotatori inesperti di XML-TEI nella codifica di documenti storici». Pisa: Università di Pisa.
- [11] Mugelli, Gloria, Federico Boschetti, Riccardo Del Gratta, Del Grosso, Angelo Mario, Fahad Khan, e Andrea Taddei. 2016. «A User-Centred Design to Annotate Ritual Facts in Ancient Greek Tragedies». *Bulletin of the Institute of Classical Studies* 59 (2): 103–20. <https://doi.org/10.1111/j.2041-5370.2016.12041.x>.
- [12] Mugelli, Gloria, Giulia Re, e Andrea Taddei. 2020. «Annotazione digitale di testi antichi. Lingue antiche e Digital Humanities, tra ricerca e didattica». *Umanistica Digitale* 5 9 (gennaio): 35–60.
- [13] Nantke, Julia, e Frederik Schlupkothén. 2020. *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization*. De Gruyter.
- [14] Parr, Terence. 2013. *The Definitive ANTLR Reference: Building Domain-Specific Languages*. Pragmatic Bookshelf.
- [15] ———. 2014. *Language Implementation Patterns: Create Your Own Domain-Specific and General Programming Languages*. Pragmatic Bookshelf.
- [16] Robinson, Peter. 2017. «Some Principles for Making Collaborative Scholarly Editions in Digital Form» 2.
- [17] Zowghi, Didar, e Vincenzo Gervasi. 2003. «On the interplay between consistency, completeness, and correctness in requirements evolution». *Information and Software technology* 45 14: 993–1009.
- [18] Zunke, Saurabh, e Veronica D'Souza. 2014. «Json vs xml: A comparative performance analysis of data exchange formats». *IJCSN International Journal of Computer Science and Network* 3 4: 257–61.

HYLAS: A new metrical search tool for Greek and Latin poetry

Michael A. Tueller

Arizona State University, USA mike_tueller@asu.edu

ABSTRACT

Hylas (*hylas.org*) is an algorithmic scanning and search tool for Greek and Latin poetry. The paper outlines the development of the tool and its coding in Python, from the rules-based model usually followed by classical scholars to a simplified machine-learning process, which has produced a series of databases by which the program judges each possible choice for syllable combination, prosody, scansion, and word division. At each step, the code judges all possibilities in parallel, assigning and compiling a probability ‘score’ for each before making a final judgement. By this method, the algorithm has achieved 100% accuracy for two metrical schemes, as measured by more than unique 2,000 spot-checks in a corpus of more than 200,000 lines.

KEYWORDS

Classical poetry, algorithmic scansion, Python.

TALK

1. INTRODUCTION

The *Hylas* project has two primary objectives: 1) to create a search engine by which scholars can search through a large corpus (presently more than 200,000 lines) of Greek and Latin poetry for metrical patterns, and 2) to allow the discovery of new metrical ‘laws’ governing the composition of this poetry. While the project is still under development, both of these goals have been largely realized for two of the most common ancient meters (dactylic hexameter and elegiac pentameter), which can now be searched at *hylas.org*.

2. BEGINNINGS AND COMPARANDA

Beginning in the 3rd century BCE, Greek (and later Latin) poetry began to be consumed primarily through reading, rather than oral performance. As a result, the audience of that poetry could more easily refer to poetic predecessors, something they had been doing to some degree already: after all, for them, poetry was an agonistic enterprise, in which each new poet was expected to compete with poets in the past. Classical scholarship has shown that this competition could take place on highly technical grounds, including such particulars as the precise placement of certain words within lines. At one point, Richard Thomas ([14]) speculated that it might be possible that poets could compete without quoting a single word – simply by duplicating complex metrical patterns. This, however, was essentially impossible to prove without the ability to search through ancient poetry by metrical criteria. It was to meet this need that I created the *Hylas* tool.

There are some tools in various stages of outline or development that have already made inroads into some part of this objective ([9];[5];[12];[2];[3];[6];[17]). It may be helpful to introduce the most important of these.

David Chamberlain’s *Greek and Latin Meter* ([2]) began as an attempt at algorithmic scanning, but in its current form is largely hand-scanned. Surprisingly, given the effort required, its corpus is quite large. Dr. Chamberlain has primarily focused his efforts on markup and presentation, and in this respect his project is plainly superior to others. Still, as it does not include a search function or automated scanning, this project is not really comparable to *Hylas*.

Pede Certo ([3]), a project at Ca’Foscari, the University of Venice, has a corpus similar in size to that of *Hylas*, though in Latin only. This project does algorithmic scanning, apparently by a rules-based model (see section 3 below), which leaves it slightly short of perfect accuracy: a little more than 1% of the lines in the dataset are actually scanned by hand. The presentation of results to the user is well done, and its search function, while approximately similar to that of *Hylas*, is more complex and certainly more intuitive. In these last characteristics, *Pede Certo* represents a goal for which *Hylas* is still striving, though *Hylas* does distinguish itself by its inclusion of Greek and its accuracy (see section 4 below).

Most promising is *Chiron*, a project headed by Daniele Fusi, also in Venice ([6]; see also [5]). So far Fusi has no public-facing release of this project, which does not only encompass scansion, but is integrated into a large ecosystem of modules for working with text, including text editing and digital publishing. Fusi’s proposed modules process verse in a fashion that, as we will see, is in many respects quite similar to *Hylas*, with successive layers of analysis, e.g. a phonological layer

preceding scansion itself. As Fusi is aiming for comprehensiveness, there is a syntactic module as well, integrating Natural Language Processing into the scheme. *Hylas* at present does not include such a module; in the initial stages of its development, I judged that it would require a great deal of computing power for a small gain in accuracy. However, as we will see below (section 6), at least some rudimentary NLP will be required as the project moves ahead.

The above projects all originate in the study of ancient Greek and Latin poetry; others ([8];[15];[13]) have been attempted in modern languages. As these modern languages feature a stress-based meter, however, their initial phonological steps are quite different from those in Latin and Greek, and thus the digital classics community has not drawn on their work. More contact between these two lines of inquiry, however, can only be helpful in the future.

3. A RULES-BASED MODEL

Greek and Latin poetic meters are quantitative; that is, they take as their basic unit the actual length of a syllable, rather than its stress, as most modern poetic meters do. To briefly digest the findings of scholarly guidebooks ([10];[11];[7];[16]), the essential rules are as follows:

1. If a syllable has a long vowel or diphthong, it is long.
2. If a syllable has a short vowel followed by two or more consonants, it is long, *unless*...
3. those two consonants are a mute followed by certain semi-vowels, in which case it could go either way.
4. Short vowels followed by one consonant or no consonants are short.
5. The last syllable in the line always counts as long.

The conventions of Greek writing make it possible to determine most of the time whether any syllable is long or short from spelling alone, even without knowing the word. (In this respect Latin is quite different.) The initial code for this project did nothing more than follow these rules and then compare its findings against the known possibilities for dactylic hexameter. By this means it was able to scan about 75% of the lines in a small test corpus in Greek.

To do better required two adaptations. First, some words could not properly be scanned without a dictionary indicating vowel quantities. Though time-consuming, this was a welcome step, as I had always planned on expanding the tool to encompass Latin poetry, which requires such a dictionary to make any headway at all. Fortunately, these dictionaries do not need to be complete; for words that are not in the dictionary, the code has certain 'rules of thumb' to guess certain features of a word's prosody; even at present, the dictionary covers less than 10% of the words in the corpus.

The second adaptation was more cumbersome. Despite its use of measured quantities, poetry is not mathematics; it often makes certain adjustments that do not quite follow its own rules. To address this, I added ancillary rules and triggers by which they would be activated, but ultimately accuracy plateaued in the neighborhood of 98%.

As nearly as I can tell, most similar projects follow one or the other of these models: they rely on either an assemblage of rules ([9];[12];[3]) or a dictionary ([17]), with human intervention making up the difference. I decided to follow a different path.

4. A SIMPLIFIED TRAINED MODEL

Roughly, the scanning project was now broken down into three major tasks:

1. Divide a line into syllables, each marked with its known characteristics.
2. Decide when two syllables combine to make a single syllable.
3. Decide whether a syllable is long or short.

The first of these could be done without any difficulty in Greek. Latin poetry required a little more intervention (partly because some of its letters can be either vowels or consonants, and partly because some of its words are actually transliterated Greek, and are thus divided like Greek words rather than Latin ones), but the problem had already effectively been solved using the previous rules-based model with its accompanying small dictionary.

To approach the other two tasks, I relied on the correctly scanned parts of the corpus, and simply took probabilities from the percentages seen there. The two syllables 'ēī', for instance, combine into a single syllable 7 out of 635 times; the syllable 'ēt h' usually scans short, but is long 4 out of 1038 times. At first, the ability of this procedure to achieve accurate results was substantially less than the earlier model, but that improved as I gradually began to fill the dictionary with more words, which would often invalidate earlier guesses made by the code and require other words in the line to be re-scanned. The success of the model is based on two pillars: the first, as just noted, is the statistics compiled from previous scanning experience, which are used to approach each new line. The second is parallel processing. As illustrated by table 1 below, The code only rejects the impossible, not the improbable: at every point where a choice can be made, the code makes *both* choices, assigning to each choice a probability to carry with it to the next step. For instance, a line with an 'ēī' in it would emerge from the 'syllable combination' step as two possible lines: one with the combined syllable 'ēī', tagged with the

probability 2%, and the other with the ‘ě’ and the ‘ī’ as two individual syllables, tagged with the probability 98%. These two lines would then go into the step that assigns syllable length, at which point the probabilities for each choice in that step are multiplied by 2% and 98% respectively. While some lines can, along the way, be tossed out as not meeting the overall scheme of the dactylic hexameter, in the end the code will accept the line that reaches the finish line with the highest score.

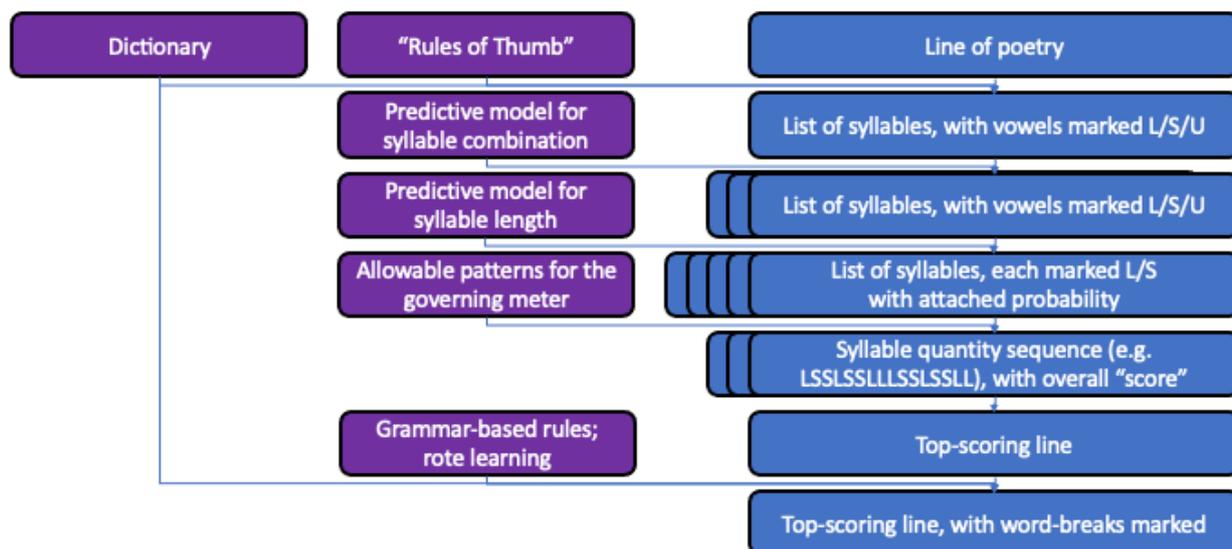


Table 1: Parallel processing in Hylas.

5. THE PROBLEM OF WORD BREAKS

The final step in table 1 is both more important and more difficult than it seems at first. While other meters are substantially more complex, a dactylic hexameter has only 32 different possibilities when seen simply as a combination of long and short syllables. Ancient metrics, however, also paid a great deal of attention to where word-breaks were placed. This results in countless possibilities for any given line. To rein in these possibilities, and to prove their talents against increasingly difficult challenges, many ancient poets created complex rules requiring or disallowing word breaks under certain circumstances.

Unfortunately, in neither Latin nor Greek does every space between words count as a break; some words are *proclitic*, effectively adhering to the word after them; others are *enclitic*, and adhere to the word before them¹. What is more, while scholars often make apparently definitive pronouncements about which words fall into which camp – proclitic, enclitic, or free-standing ([10];[11];[1]), the real situation is far murkier ([4];[16]); it is not easy to tell whether some words could inhabit more than one category, or whether some lines simply break the rules—in the end, both are likely to apply.

6. FUTURE DIRECTIONS

At present, the *Hylas* is still determining word breaks by a rules-based system, the inadequacy of which is becoming ever more apparent. As can be seen from the previous section, correcting this is not simply a matter of coding: the actual standards for -clitics first need to be firmed up. The next step for *Hylas*, then is an experiment I am currently constructing. Using two texts by quite regular poets (Apollonius of Rhodes’ *Argonautica* in Greek, and Vergil’s *Georgics* in Latin), I will begin with two spots in the hexameter line known for word breaks: the middle of the third foot, which nearly always has a break (under some circumstances it shifts to the fourth foot), and between the two short syllables of the fourth foot, which almost never has one. By surveying how these two poets treat these lines, we can discover which words tend to ‘break’ the rules: we may then suspect that, in fact, these words do not break the rules at all, but rather adhere to other words in predictable ways. Past scholarship indicates that these words will be distinguished by their lengths (mostly monosyllables) and grammatical functions (e.g. articles, prepositions, and the like), but, just as *Hylas* has discovered with metrical scansion itself, the results are likely to be more nuanced than a simple rule can express. These initial results will then be expanded to apply across the corpus, to determine how broadly they function, and how they must be refined.

¹ The reader acquainted with Greek grammar will note that the terms *proclitic* and *enclitic* are also used in that area. Though there is substantial overlap, metrical and grammatical-clitics are not equivalent.

It is already anticipated that some Natural Language Processing will be required to solve the word-break problem. This is because of homonyms that can be either proclitic, enclitic, or free-standing, with grammatical function making the determination. To cite only the most obvious example, The Greek word *oi* can be a proclitic nominative plural article or an enclitic dative singular pronoun. At present, I am already cleaning up the *Hylas* code to make room for this additional layer of processing, which must precede syllable division.

Dictionaries also must be expanded. *Hylas* succeeds with its relatively small dictionaries because the meters currently used are tightly constrained. The next meter that needs to be added, however, iambic trimeter, is much more flexible, and therefore offers far less guidance about whether a given syllable is long and short. Dictionaries must make up the difference. Here, too, Natural Language Processing will be helpful, as some words vary in pronunciation depending on their grammar. Latin *venimus*, for instance, is pronounced *vĕnĭmŭs* in the present tense but *vĕnĭmŭs* in the perfect. When iambic trimeter is added, the user interface for *hylas.org* also must be changed, as this far more flexible meter cannot be searched using the limited set of user choices currently offered.

Along the way, access to the tool will be broadened. My intention is that this tool be open access, but I am currently seeking advice from my university on licensing. When this issue is resolved, the code and corpora will be posted freely on GitHub. At present, there is no plan to include an open API; if this is included, it will be done in a later stage of the project.

REFERENCES

- [1] Bulloch, Anthony. 1970. "A Callimachean Refinement to the Greek Hexameter." *The Classical Quarterly* 20 2: 258–68.
- [2] Chamberlain, David. n.d. "Greek and Latin Meter." *Hypotactic* (blog). <http://hypotactic.com/latin>.
- [3] Colombi, Daniela. n.d. "Pede Certo." *Pede Certo* (blog). <http://www.pedecerto.eu/public/>.
- [4] Devine, A. W., and Laurence Stephens. 1978. "The Greek Appositives: Toward a Linguistically Adequate Definition of Caesura and Bridge." *Classical Philology* 73 4: 314–28.
- [5] Fusi, Daniele. 2015. "A Multilanguage, Modular Framework for Metrical Analysis: IT Patterns and Theoretical Issues." *Languages* 199, 41–66.
- [6] ———. n.d. "Chiron." *Fusisoft* (blog). <https://fusisoft.net/chiron/>.
- [7] Halporn, James W., Martin Ostwald, and Thomas G. Rosenmeyer. 1980. *The Meters of Greek and Latin Poetry, Revised Edition*. University of Oklahoma Press.
- [8] Hartman, Charles O. 2005. "The Scandroid." *Charlesohartman* (blog). 2005. <http://charlesohartman.com/verse/scandroid/index.php>.
- [9] Papakitsos, Evangelos C. 2011. "Computerized Scansion of Ancient Greek Hexameter." *Literary and Linguistic Computing* 26 1: 57–69.
- [10] Raven, D. S. 1962. *Greek Metre: An Introduction*. London: Faber & Faber.
- [11] ———. 1965. *Latin Metre: An Introduction*. Faber & Faber.
- [12] Schumann, Anne-Kathrin, Christoph Beierle, and Norbert Blößner. 2019. "A Finite-State Approach to Automatic Greek Hexameter Analysis." Utrecht.
- [13] Tanasescu, Chris, Bryan Paget, and Diana Inkpen. n.d. "Automatic Classification of Poetry by Meter and Rhym."
- [14] Thomas, Richard F. 1986. "Harvard Studies." In *Classical Philology*, 171–98. 90.
- [15] Van Wangenheim, Aldo, Luiz dos Santos Alckmar, and Adiel Mitmann. 2016. "A System for the Automatic Scansion of Poetry Written in Portuguese." Konya, Turkey.
- [16] West, Martin L. 1982. *Greek Metre*. Oxford University Press.
- [17] Winge, Johnan. 2015–2021. "A Latin Macronizer." *Alatius* (blog). <http://www.alatius.com/macronizer/>.

Stylometry and Reader Response. An Experiment with *Harry Potter* Fanfiction

Simone Rebora

Università degli Studi di Verona, Italy - simone.rebora@univr.it

ABSTRACT

The paper discusses the results of an experiment which presents readers with a typical problem in stylometry: the attribution of an anonymous text given the writing samples of candidate authors. Three *Harry Potter* fanfictions (average length ~ 1,500 words) were used for the experiment. Out of 142 participants, 60.56% were able to correctly attribute the anonymous text, while 87.11% of 12,793 stylometric analyses proved successful. A comparison is established between the motivations provided by readers and the working logic of stylometric methods, highlighting similarities and differences.

KEYWORDS

Stylometry, authorship attribution, reader response, Harry Potter fanfiction.

TALK

1. INTRODUCTION

In the context of “distant reading” ([12]), stylometry is currently one of the most advanced techniques, using statistical methods to automatically identify the “stylistic fingerprint” of a writer. Developed independently much before the advent of computers and Digital Humanities (the origins can be traced back to the middle of the XIX century) ([8]), it imposed itself as a computational method in literary studies after the success of approaches like Delta distance ([1]) in authorship attribution. Multiple studies have been dedicated to the application of stylometry to literary history and theory ([7]), while even more research has been focused on discovering its limits of applicability ([2]) and on understanding its core working logic ([4]). Still, one of the main mysteries surrounding this field of study is the very reason why it works so efficiently, which brought some scholars to term it as a kind of “black magic” ([9]). Explanations have been proposed, supported also by knowledge in cognitive science ([11]), but what is still lacking is an empirical validation of such proposals.

This paper aims at providing a contribution to this discussion, by setting its groundwork at the intersection between stylistics and reader response studies. Indeed, stylistics is one of the fields of research in humanities that has been most impacted by the success of stylometry ([6]). At the same time, as Whiteley and Canning recently stated, “stylistics is, and always has been, inherently and inescapably concerned with reader response” ([15]). Without aspiring at providing a “proof” for the reasons why stylometry works, this contribution will simply compare stylometric methods with the intelligence of common readers when attributing an anonymous text. This to highlight possible commonalities and differences: in particular, if there are elements in stylometry that escape the eye of the common reader (and, of course, also the opposite). Such a comparison will be made possible by an experimental setup ([14]), aimed at putting both humans and the computer in front of the same (rather simplified) authorship attribution challenge.

2. THE EXPERIMENTAL SETUP

The experiment was set up as follows: participants had to read two texts by two different authors and then attribute a third anonymous text to one of the two authors. This setup emulates the typical authorship attribution task in stylometry, where an algorithm is “trained” on texts of known authorship (to devise the stylistic fingerprints of different authors) and then applied to texts of unknown authorship ([13]). In order to allow a wide participation in the experiment, such a task had to be simplified substantially: instead of “training” readers on multiple long texts and asking them to attribute many anonymous documents, the experiment was limited to the reading of three, rather short texts. This posed an issue for the stylometric analysis, as the attribution of short texts generally proves less efficient. In this regard, it should be noted from the beginning that texts were selected after having verified that: (a) stylometric methods could attribute them correctly; (b) authorship was not blatantly evident via close reading. Such choices unavoidably added an element of arbitrariness in the experimental setup, but they appeared as a necessary compromise to allow the very feasibility of the experiment. Texts were downloaded from *Archive of Our Own*, a fanfiction platform already used for authorship attribution tasks ([10]).

A fictional frame was devised to present the experiment as a possible real case scenario, with two *Harry Potter* fanfiction writers (Anne and Rose) proposing a challenge to the participant (see Figure 1).

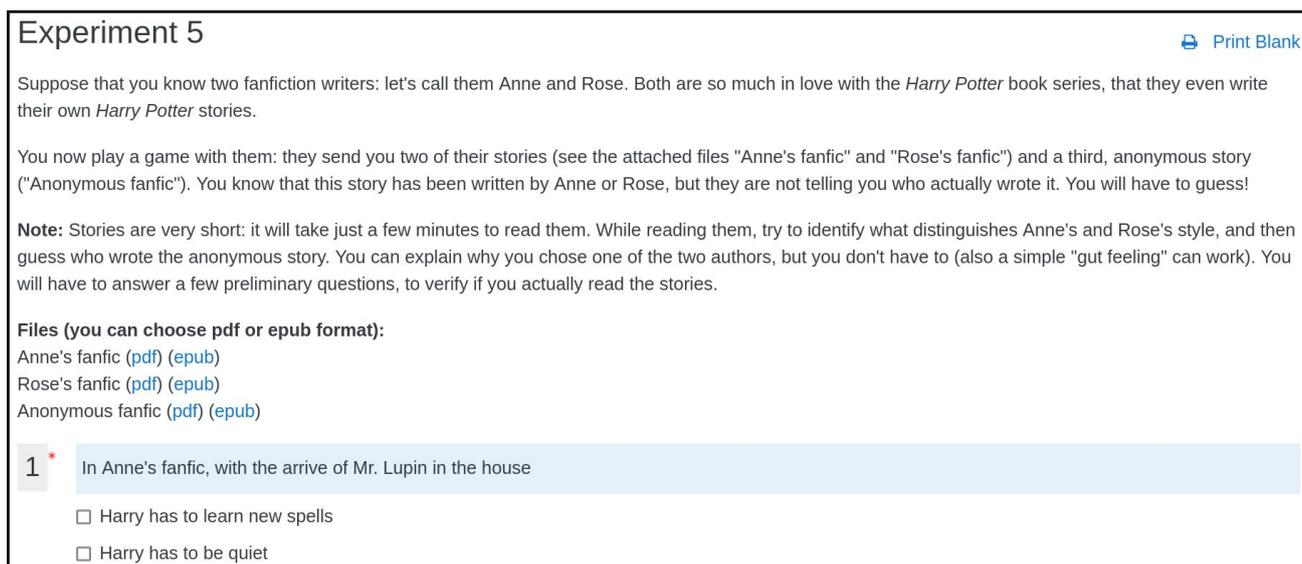


Figure 1. Screenshot of the questionnaire

The first text to be read (“Anne’s fanfic”, 1,525 words) presented a young Harry living together with Remus and Sirius, tentatively getting acquainted with a new host in the house, Mr. Lupin; the second text (“Rose’s fanfic”, 1,528 words) presented instead a middle-aged Harry catching up her daughter at the train station and gradually realizing how she had changed and grown up; the third text (“Anonymous fanfic”, 1,739 words) recounted the first encounter between three characters (Sirius, Remus, and Peter) and a newborn Harry, still in his mother’s arms. Participants had to answer three control questions to verify if they actually read the stories (see Figure 1 for an example), they had to vote for Anne or Rose as the author of the “Anonymous fanfic”, and were allowed to leave a comment to motivate their choice.

Table 1 reveals the correct answer to the central question, being Rose the actual author of the “Anonymous fanfic”.

Experiment title	Actual title	Author	Link
Anne’s fanfic	Patient	likehandlingroses	https://archiveofourown.org/works/19242301
Rose’s fanfic	Saying Hello	FloreatCastellum	https://archiveofourown.org/works/18672805
Anonymous fanfic	Godfather	FloreatCastellum	https://archiveofourown.org/works/20496251

Table 1. Questionnaire solution

3. RESULTS (READERS)

A total of 162 participants took part in the experiment, which was repeated five times between April 2020 and April 2021¹. Participants were both Bachelor’s students (two rounds, n = 56), Master’s students (two rounds, n = 73), PhD students and young researchers (one round, n = 33). Twenty of them made at least one mistake in the control questions, so their answers were not considered in the analysis. Out of the remaining 142 participants, 83 left a comment to motivate their choice.

As Table 2 shows, 60.56% of the participants gave the correct answer to the attribution question. Overall, it can thus be stated that the text was correctly attributed by a slight (but substantial) majority of the readers. It is also interesting to notice how the level of education does not seem to correlate with the attribution accuracy, with Bachelor’s students reaching the highest score and PhD/researchers placing themselves at the overall mean. Percentages do not even increase substantially when considering just the 83 participants who left a comment, suggesting how a more attentive reflection on the problem does not simplify its solution.

	All participants	Commenting participants
--	------------------	-------------------------

¹ The full dataset, together with the scripts for its analysis, is available on GitHub at the following link: https://github.com/SimoneRebora/Stylometry_ReaderResponse.

	<i>Rose</i>	<i>Anne</i>	<i>Rose</i>	<i>Anne</i>
Total	60.56 %	39.44 %	61.45 %	38.55 %
Bachelor	68.09 %	31.91 %	61.54 %	38.46 %
Master	54.69 %	45.31 %	60 %	40 %
PhD	61.29 %	38.71 %	64.71 %	35.29 %

Table 2. Results overview (readers)

A detailed analysis of the 83 comments was carried out by setting up their categorization into eight main categories²:

- “stylistics”, indicating a focus on linguistic or syntactic cues to attribute the text;
- “structure”, i.e. the general structure of the narrative;
- “dialogues” and “descriptions”, i.e. the high/low presence of such narrative devices³;
- “focus”, i.e. aspects like point of view and focalization;
- “knowledge”, i.e. knowledge of the *Harry Potter* fictional universe;
- “content” and “characters”.

Inevitably, some comments escape this categorization⁴, which imposes an oversimplification to their variance. Still, the categorization serves the goal of providing a synthetic overview of the main reasons that guided the attribution.

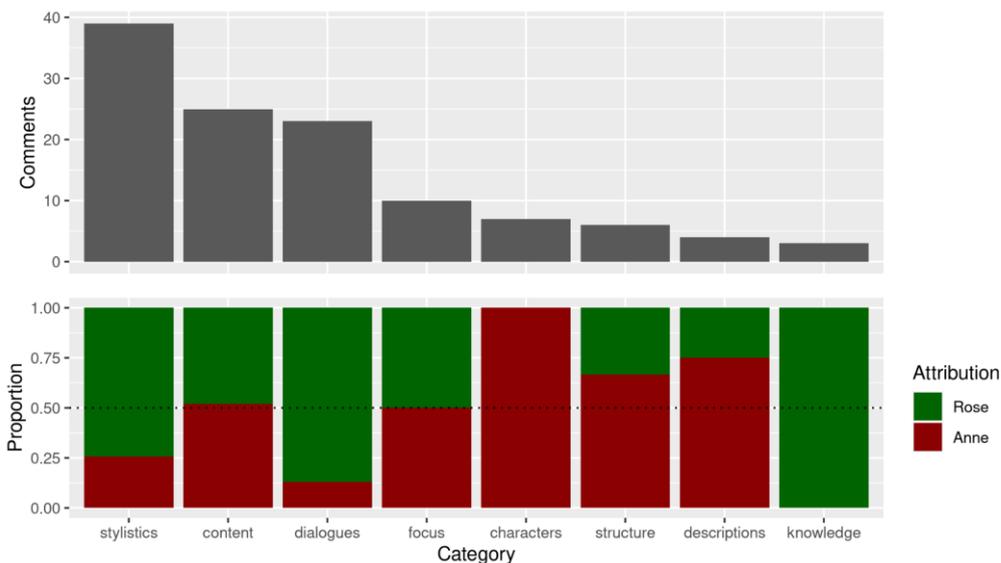


Figure 2. Comment categorization overview

As shown by Figure 2, “stylistics” is the dominant category (39 comments can be grouped under this tag), with a substantial majority (74%) of correct attributions. Other categories that drove a correct attribution are “dialogues” (when participants, even without deepening the analysis, noted the higher number of dialogues in both Rose’s and the anonymous fanfiction) and “knowledge” (which even reaches a 100% of accuracy, with participants able to discern the authors based on their different treatment of the *Harry Potter* source material). More problematic are then the “content” and “focus” categories, which caused an almost equal number of correct and incorrect attributions. As this case study suggests, relying on these aspects alone does not prove resolute when dealing with attribution issues. Clearly inefficient are then the “characters”, “structure”, and “descriptions” categories, with “characters” even reaching a 100% of inaccuracy (indeed, the main challenge for the attribution was the recurrence of some characters in both Anne’s and the anonymous fanfiction). “Structure” is more problematic, as it might also enclose the “dialogues” category (which drove a majority of correct attributions): still, comments falling under the latter category were much more generic, looking at more evident

² Comments providing multiple reasonings to support the attribution were categorized under multiple categories, with a maximum of three categories per comment.

³ Notice how these two categories can be considered as subcategories of “structure”. They were isolated because of their high frequency in the comments.

⁴ For example, a comment used the “emotional response” of the reader to (correctly) attribute the text. As such a response was still caused by the content of the narrative, the comment was (quite forcefully) categorized under “content”.

(quantitative) aspects of the text. As it seems, a deeper attention to more complex structural aspects (or the focus on less frequent devices like descriptions) caused instead a higher number of errors in the attribution.

4. RESULTS (STYLOMETRY)

To get an overview of the efficiency of stylometric methods on this case study, a series of analyses was performed using the *Stylo* R package ([3]), by combining the eleven distance measures available in the package with all the possible lists of most frequent words (from one to 1,163 MFW). This resulted in a total of 12,793 different analyses. Overall, 87.11% of the analyses provided the correct attribution. However, it should be noted that accuracy dropped from 100% only when considering more than 641 MFW, i.e. when taking into account words that appeared just once in the whole corpus. Such a selection of MFW is generally avoided in stylometry, which bases itself more on highly-frequent terms than on hapaxes ([9]).

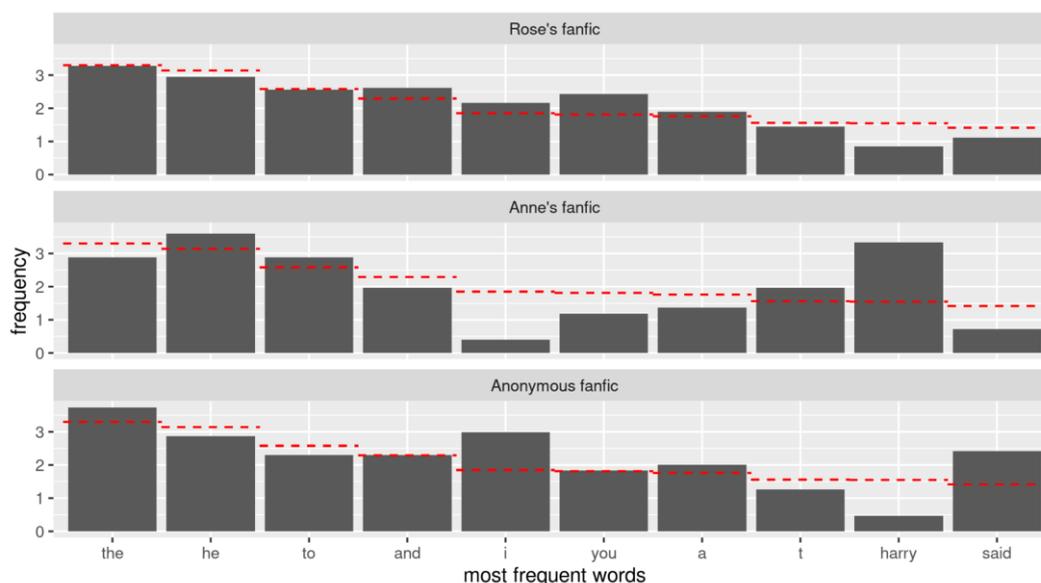


Figure 3. Frequency of the ten most frequent words in the three texts

A deeper understanding of the phenomena that caused such a result is allowed by Figure 3, which shows the frequency of the ten MFW in the corpus. The grey bars represent the frequency of each word in each text, while the red dashed lines represent their mean frequency in the whole corpus. When comparing the ten words, it becomes evident that in the majority of the cases there is a clear pattern connecting Rose’s text with the Anonymous text. See for example the word “he”: both Rose’s and the Anonymous fanfiction use it less than the average, while Anne’s fanfiction uses it more than the average. Stylometric distances like Burrows’s Delta automatize such a reasoning in a purely mathematical manner, by repeating this operation for all words in a list and calculating a unique distance measure that averages on all the observations. Visualizing their working logic can become problematic when considering thousands of words and hundreds of texts, but the basic approach is not different from the one exemplified here.

5. DISCUSSION

In Figure 3, the already-mentioned case of the word “he” is mirrored by the case of the word “t”, which is clearly overused in both Rose’s and the Anonymous fanfiction. It seems reasonable to suppose that such a correspondence was caused by a comparatively higher use of dialogues in the two texts: and while readers caught this overuse through a structural analysis of the text, stylometry catches it by just looking at the frequency of single words. In terms of modeling ([5]), it can be stated here that by representing style as similarity in patterns of word frequencies, stylometry also measures a wide range of phenomena that are commonly identified by readers, but it does it indirectly, by looking at how they affect the use of words.

Readers also partly “emulated” stylometry when looking at stylistic aspects of the text (e.g., by evaluating the wideness of the vocabulary, or the over-/underuse of specific parts of speech), and, at least in this case study, such an approach proved successful. Results of the experiment also show a general correspondence between the attention at high-level features (structure and characters in readers; low-frequency words in stylometry) and the unsuccessfulness of attributions. At the

same time, there are approaches in readers that cannot find a correspondence in stylometry. In particular, the attributions here labelled as “knowledge” appear as the most difficult to model: and while methods have been proposed that move towards this direction (by profiting from large databases to extract general conceptual knowledge) ([10]), it seems that such an approach still requires all the intelligence and skillfulness of the human reader to be implemented.

This paper has just set up the groundwork for such reflections, which can move even further but will also need much more analysis and experimentation to be fully developed. Still, none of them will be possible without a fruitful combination between literary theory, digital humanities, and empirical methods: an interdisciplinary collaboration that might open very promising pathways for future research.

REFERENCES

- [1] Burrows, John. 2002. “Delta’: A Measure of Stylistic Difference and a Guide to Likely Authorship.” *Literary and Linguistic Computing* 17 (3): 267–87. <https://doi.org/10.1093/lc/17.3.267>.
- [2] Eder, Maciej. 2013. “Does Size Matter? Authorship Attribution, Small Samples, Big Problem.” *Digital Scholarship in the Humanities* 30 (2): 167–82.
- [3] Eder, Maciej, Jan Rybicki, and Mike Kestemont. 2016. “Stylometry with R: A Package for Computational Text Analysis.” *The R Journal* 8 (1): 107–21.
- [4] Evert, Stefan, Thomas Proisl, Jannidis Fotis, et al. 2017. “Understanding and Explaining Delta Measures for Authorship Attribution.” *Digital Scholarship in the Humanities*.
- [5] Flanders, Julia, and Fotis Jannidis, eds. 2019. *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*. London: Routledge, Taylor & Francis Group.
- [6] Herrmann, J. Berenike, Christof Schöch, and Karina Van Dalen-Oskam. 2015. “Revisiting Style, a Key Concept in Literary Studies.” *Journal of Literary Theory* 9 (1): 25–52.
- [7] Jannidis, Fotis, and Gerhard Lauer. 2014. “Burrows’s Delta and Its Use in German Literary History.” In *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, Edited by Matt Erlin and Lynne Tatlock, 29–54. Rochester: Camden House.
- [8] Juola, Patrick. 2006. “Authorship Attribution.” *Foundations and Trends in Information Retrieval* 1 (3): 233–334.
- [9] Kestemont, Mike. 2014. “Function Words in Authorship Attribution. From Black Magic to Theory?” In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, 59–66. Gothenburg, Sweden: Association for Computational Linguistics.
- [10] Kestemont, Mike, Michael Tschuggnall, Efstathios Stamatatos, et al. 2018. “Overview of the Author Identification Task at PAN-2018: Cross-Domain Authorship Attribution and Style Change Detection.” In *In Working Notes Papers of the CLEF 2018 Evaluation Labs. A Cura Di Cappellato Linda et Al.*, 1–25. Avignon, France.
- [11] Lancashire, Ian. 2005. “Cognitive Stylistics and the Literary Imagination.” *Companion to Digital Humanities*, 397–414.
- [12] Moretti, Franco. 2013. *Distant Reading*. London: Verso.
- [13] Stamatatos, Efstathios. 2009. “A Survey of Modern Authorship Attribution Methods.” *Journal of the Association for Information Science and Technology* 60 (3): 538–56.
- [14] Van Peer, Willie, Jèmeljan Hakemulder, and Sonia Zyngier. 2012. “Scientific Methods for the Humanities.” In *Linguistic Approaches to Literature*. Vol. 13. Amsterdam; Philadelphia: John Benjamins Pub.
- [15] Whiteley, Sara, and Patricia Canning. 2017. “Reader Response Research in Stylistics.” *Language and Literature* 26 (2): 71–87.

Sessione Intelligenza 1

Alan M. Turing

Analisi e valorizzazione del patrimonio artistico mediante Intelligenza Artificiale

Giovanna Castellano¹, Gennaro Vessio²

¹Università degli Studi di Bari, Italia – giovanna.castellano@uniba.it

²Università degli Studi di Bari, Italia – gennaro.vessio@uniba.it

ABSTRACT

Soluzioni basate su Intelligenza Artificiale stanno già potenziando numerosi campi del sapere e dell'attività umana, inclusa l'arte. Infatti, i recenti progressi nel campo dell'Intelligenza Artificiale, insieme con la crescente disponibilità di collezioni di opere d'arte digitalizzate, stanno offrendo nuove opportunità agli specialisti in questi settori perché assistano la comunità artistica, ma anche semplicemente gli appassionati, con nuovi strumenti automatici e "intelligenti". Il presente articolo delinea attività di ricerca condotte nell'ambito del progetto PON AIM "Metodi computazionali per il patrimonio culturale", il cui fine è proporre nuovi metodi, basati su Intelligenza Artificiale, per una migliore analisi e valorizzazione del patrimonio artistico digitalizzato.

PAROLE CHIAVE

Digital Humanities, intelligenza artificiale, computer vision, deep learning, data science.

INTERVENTO

1. INTRODUZIONE

Il patrimonio artistico riveste un ruolo di importanza strategica per la crescita economica e culturale di un Paese. In tempi recenti, grazie al progresso tecnologico, è stato compiuto uno sforzo di digitalizzazione su larga scala, che ha portato alla crescente disponibilità di grandi collezioni di opere d'arte digitalizzate. Un esempio particolarmente noto è WikiArt, la versione "artistica" di Wikipedia. La disponibilità di tali basi di conoscenza, unita ai recenti progressi nel campo dell'Intelligenza Artificiale, ha aperto nuove opportunità agli specialisti in questi settori perché assistano storici dell'arte, curatori di musei, appassionati, ecc., nello studio e in una migliore fruizione del patrimonio artistico digitalizzato. La fruizione di tale patrimonio da parte di un pubblico sempre più vasto promuove la diffusione della cultura ([5]).

La capacità di riconoscere caratteristiche, similarità, ecc., in e fra opere d'arte digitalizzate ricade inerentemente nel dominio della percezione estetica umana. Giacché tale percezione è fortemente soggettiva, e influenzata da diversi fattori, fra cui la conoscenza storica relativa all'opera, o l'emozione che l'opera suscita in chi l'osserva, essa è estremamente difficile da codificare ([7]). Tuttavia, grazie a tecniche di apprendimento automatico e rappresentazione della conoscenza, come quelle su cui si basano modelli di reti neurali allo stato dell'arte ([10]), è oggi possibile dotare le macchine di capacità percettive, che permettano loro di estrarre automaticamente caratteristiche descrittive di opere d'arte dalla loro codifica elementare in pixel. Tali rappresentazioni possono essere di beneficio per automatizzare numerosi compiti rilevanti dal punto di vista artistico, come la categorizzazione di un dipinto in base ad artista, stile e genere, oppure il ritrovamento di opere simili a una data opera in ordine a caratteristiche visuali, testuali, ecc.

Il progetto PON AIM "Metodi computazionali per il patrimonio culturale", svolto all'interno del Laboratorio di Intelligenza Computazionale del Dipartimento di Informatica dell'Università degli Studi di Bari, sta contribuendo alla ricerca interdisciplinare in questo settore, oggi molto attiva e fertile, al fine di sviluppare e applicare nuovi metodi, basati su Intelligenza Artificiale, per una migliore analisi e fruizione del patrimonio artistico digitalizzato.

2. LINEE DI RICERCA

Il nostro contributo si è focalizzato su quattro linee di ricerca:

- il *visual link retrieval* e il *knowledge discovery* in database di opere d'arte digitalizzate;
- il *clustering* automatico di tali opere;
- l'integrazione di informazione "contestuale" in modelli di Computer Vision, codificata per mezzo di *knowledge graph* (KG);

- l'integrazione di tali modelli nel robot sociale Pepper per l'interazione uomo-robot all'interno di musei o gallerie d'arte.

Maggiori dettagli nei seguenti sottoparagrafi.

2.1 VISUAL LINK RETRIEVAL

Uno degli elementi costitutivi la maggior parte delle analisi nelle arti visive è la ricerca di relazioni di “similarità” tra opere di artisti e scuole pittoriche differenti. Queste relazioni possono aiutare gli storici dell'arte a scoprire o meglio comprendere le influenze artistiche e i cambiamenti culturali avvenuti da un movimento artistico a un altro. Infatti, gli esperti d'arte raramente analizzano le opere visive come creazioni isolate, ma studiano queste all'interno di contesti più ampi, che coinvolgono influenze e connessioni tra diverse scuole.

Tradizionalmente, questo tipo di analisi è condotta manualmente ispezionando grandi raccolte di foto annotate. Tuttavia, la ricerca manuale fra migliaia di immagini, distribuite tra diversi periodi e scuole pittoriche, è un processo difficile e oneroso. Lungo questa direzione, abbiamo proposto un metodo a supporto del *visual link retrieval*, che si basa sull'uso di una rete neurale convoluzionale *deep* per l'estrazione di “feature” significative dalle immagini delle opere digitalizzate e su di un meccanismo di *nearest neighbor* non supervisionato per ritrovare link tra tali opere ([2]). La ricerca di link “visuali” è completamente non supervisionata, rendendo il metodo particolarmente utile nei casi in cui eventuali metadati siano scarsi, non disponibili o difficili da raccogliere. Vale la pena notare che il metodo proposto non solo fornisce quelle immagini che sono più simili alla *query* in input (v. Fig. 1), ma consente anche di studiare *pattern* storici analizzando un “grafo delle influenze” costruito sui link ritrovati. Infatti, applicando misure di grafo sulla rete complessa costruita sui link ottenuti, il metodo proposto permette una forma di *knowledge discovery* sugli artisti.

Il metodo proposto può essere vantaggioso non solo per gli storici dell'arte. Gli appassionati, infatti, possono beneficiare del ritrovamento automatico di link durante la visita di collezioni digitalizzate di musei e gallerie d'arte online, favorendo la fruizione delle collezioni digitali.

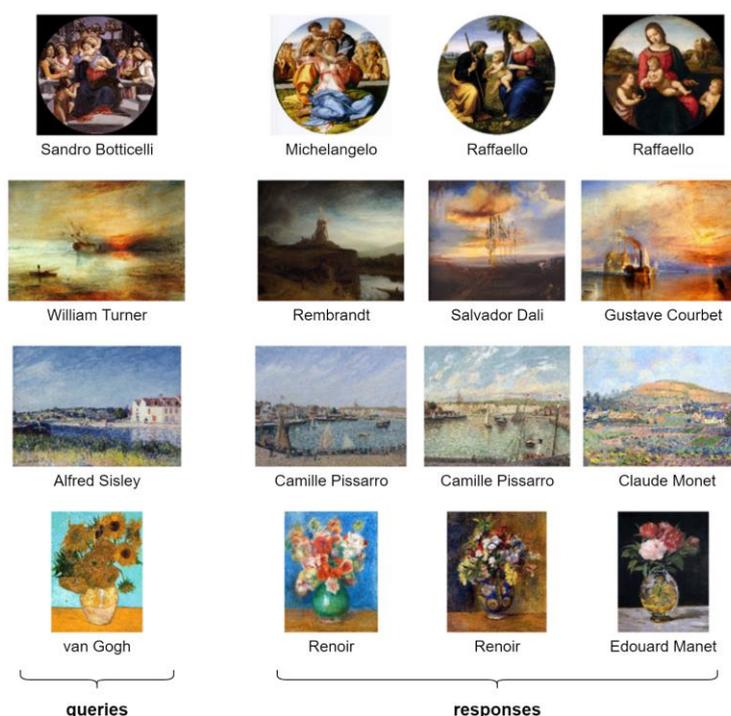


Figura 1. A sinistra, immagini in input al sistema; a destra, corrispondenti immagini restituite in output, considerate “visivamente” simili alle prime. L'algoritmo esclude, dal processo di *visual link retrieval*, opere prodotte dallo stesso artista della *query* in input.

2.2 ARTWORK CLUSTERING

Sebbene l'approccio descritto nel sottoparagrafo precedente sia adeguato per trovare opere d'arte collegate visivamente, esso non è efficace per raggruppare le opere appunto in gruppi ben distinti, poiché i dati appaiono distribuiti uniformemente all'interno di un singolo *cluster* omogeneo nello spazio delle feature. Avere un modello in grado di raggruppare le opere d'arte senza dipendere da etichette difficili da raccogliere o da un giudizio umano soggettivo può essere utile per molte applicazioni di dominio. Ad esempio, tale modello potrebbe essere utilizzato per scoprire periodi diversi nella produzione

di uno stesso artista. Allo stesso modo, potrebbe aiutare gli esperti del dominio a classificare l'arte contemporanea, che non può essere riccamente annotata.

A tal fine, abbiamo proposto un metodo che utilizza una rete neurale convoluzionale *deep* pre-addestrata per l'estrazione delle feature, ma che utilizza anche una componente di *deep clustering*, basata su di un *auto-encoder*, per effettuare il clustering ([6]). La scelta di tale soluzione è stata motivata dalla difficoltà di applicare metodi tradizionali di *clustering* o *dimensionality reduction* sia allo spazio dei pixel in input, dalla dimensionalità intrattabile, che allo spazio delle feature risultante dall'*embedding* di una rete neurale, specialmente quando le immagini in input sono immagini artistiche molto complesse.

Risultati sperimentali quantitativi e qualitativi hanno mostrato che il metodo proposto è in grado di trovare *cluster* ben separati sia quando si considera un insieme di dati ampio ed eterogeneo che abbracci diversi periodi, sia quando ci si concentra su opere prodotte da uno stesso artista (v. Fig. 2). In particolare, da un punto di vista qualitativo, sembra che il modello guardi non solo agli stilemi per raggruppare le opere, ma anche soprattutto agli attributi semantici relativi al contenuto della scena rappresentata; in altre parole, al genere dell'opera. Questa capacità sembra essere promettente per affrontare il noto problema della *cross depiction*, che rappresenta ancora una sfida aperta per la comunità scientifica ([9]). In effetti, questa capacità potrebbe essere sfruttata per trovare somiglianze tra opere d'arte indipendentemente dal modo in cui queste sono rappresentate.

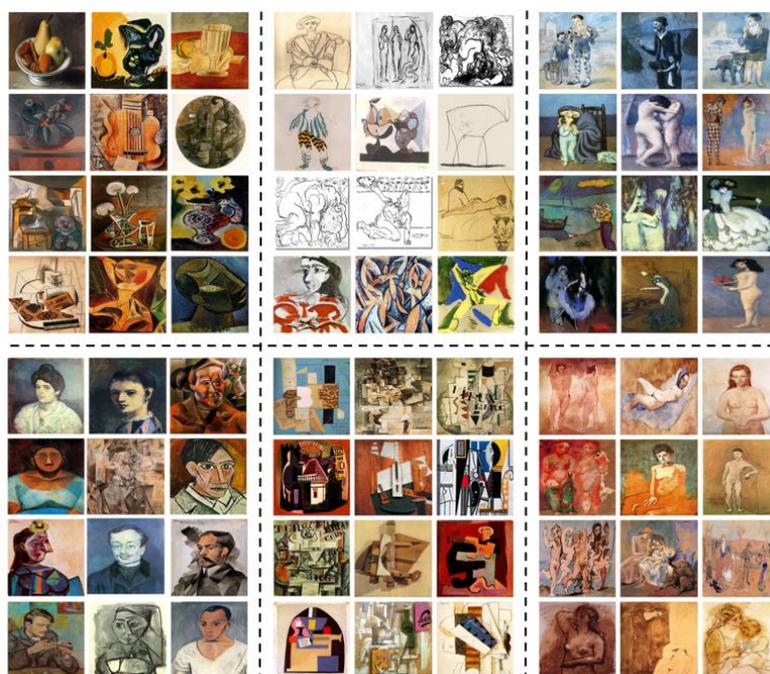


Figura 2. Cluster individuati automaticamente dal sistema tra opere prodotte da Pablo Picasso. I cluster contengono opere che condividono caratteristiche sia stilistiche che di contenuto.

2.3 KNOWLEDGE GRAPH

La ricerca ha poi mosso da un'altra considerazione: gran parte dei lavori in letteratura si basa esclusivamente sull'informazione codificata dai pixel caratterizzanti i dipinti digitalizzati. Sfortunatamente, questo approccio porta a ignorare una grande quantità di conoscenza di dominio, nonché relazioni e connessioni note tra opere d'arte e/o artisti, che potrebbero aumentare la qualità delle soluzioni esistenti ([8]). Le opere d'arte, infatti, non possono essere studiate solo sulla base del loro aspetto visivo, ma anche considerando vari altri fattori storici, sociali e "contestuali" che consentono di inquadrarle all'interno di un quadro più complesso. Pertanto, disporre di una base di conoscenza in cui non solo le opere d'arte, ma anche una ricca pletora di metadati, informazioni contestuali, descrizioni testuali, ecc., siano unificate all'interno di un unico framework strutturato può fornire una risorsa preziosa per sviluppare modelli più efficaci. Tale framework sarebbe utile non soltanto per gli utenti generici, che potrebbero sfruttare le informazioni codificate per navigare nella base di conoscenza, ma anche soprattutto per gli esperti d'arte, interessati a trovare nuove relazioni tra opere d'arte e/o artisti per una migliore comprensione dell'arte passata e contemporanea.

Per colmare questa lacuna, stiamo lavorando allo sviluppo di *ArtGraph*: un *knowledge graph* nel dominio artistico ([4]). Un KG fornisce una rappresentazione più espressiva e flessibile per incorporare relazioni di complessità arbitraria

tra entità concernenti l'arte, che non possono essere ottenute considerando solo il loro aspetto visivo. Il KG proposto integra le informazioni raccolte da WikiArt e DBpedia e sfrutta le potenzialità di un database NoSQL, Neo4j, che fornisce una modellazione basata su grafo altamente espressiva e un linguaggio di interrogazione molto potente. In questo modo, il database NoSQL aiuta già a fornire uno strumento per la scoperta di conoscenza senza addestrare esplicitamente un algoritmo di apprendimento. La conoscenza codificata in *ArtGraph* può essere poi integrata con feature apprese automaticamente per affrontare vari *task* tra cui la previsione di attributi, come stile e genere.

2.4 SOCIAL ROBOTICS

Man mano che le applicazioni degli algoritmi di Computer Vision a task nel dominio artistico diventano sempre più mature, una interessante implementazione di queste tecniche in contesti reali è incorporarle all'interno dei cosiddetti "robot sociali". Questi rappresentano un campo di ricerca emergente focalizzato sullo sviluppo di una "intelligenza sociale" che mira a mantenere l'illusione di avere a che fare con un essere umano ([1]). In questo contesto, i recenti progressi nella Computer Vision consentono ai ricercatori di dotare i robot di nuove e potenti capacità. Nella nostra ricerca stiamo usando un robot sociale, Pepper, come guida turistica di un museo ([3]). In particolare, stiamo sviluppando un approccio basato su visione per supportare i turisti durante la visita. Pepper è un robot semi-umanoide su ruote, dotato di diverse telecamere e sensori. Il modulo di visione consente a Pepper di percepire la presenza dei visitatori e di localizzarli nello spazio, stimandone età e genere. Inoltre, lo stesso modulo di *visual link retrieval* descritto in precedenza offre a Pepper la capacità di utilizzare l'immagine del dipinto osservato dal visitatore come una *query* visiva per cercare dipinti visivamente simili all'interno dello stesso museo. Il robot utilizza questi dati e altre informazioni acquisite durante il dialogo per fornire quindi ai visitatori consigli su opere d'arte simili a cui potrebbero essere interessati.

Progettare i comportamenti di un robot sociale che funga da guida museale richiede di dotarlo di diverse competenze che forniscano ai visitatori un'esperienza coinvolgente ed efficace durante la visita. Queste funzionalità hanno lo scopo di consentire al robot di rilevare e localizzare le persone nel museo, riconoscere l'opera d'arte che il visitatore sta guardando, profilare l'utente durante la visita al fine di generare raccomandazioni adeguate e, infine, coinvolgere le persone nell'interazione utilizzando adeguate capacità di conversazione. Abbiamo testato l'approccio proposto nel nostro laboratorio di ricerca ed esperimenti preliminari hanno dimostrato la sua fattibilità.

3. CONCLUSIONI

La crescente disponibilità di vaste collezioni di opere d'arte digitalizzate ha dato vita a una nuova intrigante area di ricerca in cui l'Intelligenza Artificiale e le arti visive si incontrano. Tale area di ricerca si inquadra all'interno dell'emergente Informatica Umanistica, che mira a unificare tecnologie digitali e discipline umanistiche. Le applicazioni sono innumerevoli e spaziano dal ritrovamento di informazioni in banche dati digitali alla generazione sintetica di nuove forme d'arte. Siamo fiduciosi che questo entusiasmante campo di ricerca sarà rafforzato sfruttando i rapidi progressi negli approcci di Deep Learning. Riteniamo che questi approcci continueranno a evolversi rapidamente, aprendo così la strada alla realizzazione di scenari sorprendenti in cui i sistemi informatici saranno in grado di analizzare e comprendere le "belle arti" autonomamente. Infatti, uno degli obiettivi finali di questa ricerca è la capacità delle macchine, quando opportunamente addestrate, di ricavare autonomamente una comprensione di ciò che la scena all'interno di un'opera rappresenta, qual è la metafora che ne è alla base, quali sono le possibili implicazioni storiche, ecc., senza alcuna supervisione umana.

Tuttavia, il dominio artistico è significativamente diverso dal dominio naturale/fotorealistico cui gli esperti di Intelligenza e Visione Artificiale sono avvezzi. In primo luogo, esiste una variabilità intrinseca tra le caratteristiche stilistiche e figurative dei due domini, nonché tra opere di artisti diversi appartenenti a uno stesso periodo, se non tra opere di uno stesso artista. Inoltre, i dataset con cui ora pre-addestriamo i modelli di Deep Learning sono affetti da "recentismo", e non sono cioè rappresentativi di situazioni, modi di essere/vestire, di scene iconografiche e mitologiche, ecc., del passato, che non sono mai esistiti oppure semplicemente non esistono più. In altre parole, il patrimonio artistico, dato il suo background storico evolutosi nel corso dei secoli, pone sfide scientifiche del tutto nuove e intriganti che, se affrontate, possono avanzare la comprensione automatica di una scena digitalizzata, per ora solo parzialmente ottenuta con i modelli attuali.

4. RINGRAZIAMENTI

Gennaro Vessio riconosce il sostegno finanziario del Ministero dell'Università e della Ricerca attraverso il progetto PON AIM 1852414.

BIBLIOGRAFIA

- [1] Castellano, Giovanna, Berardina De Carolis, Francesca D'Errico, et al. 2021. «PeppeRecycle: Improving Children's Attitude Toward Recycling by Playing with a Social Robot». *International Journal of Social Robotics* 13 (1): 97–111.
- [2] Castellano, Giovanna, Eufemia Lella, e Gennaro Vessio. 2021. «Visual link retrieval and knowledge discovery in painting datasets». *Multimedia Tools and Applications* 80 (5): 6599–6616.
- [3] Castellano, Giovanna, Nicola Macchiarulo, Bernardina De Carolis, e Gennaro Vessio. 2020. «Pepper4Museum: Towards a Human-like Museum Guide». *AVPCH@AVI*.
- [4] Castellano, Giovanna, Giovanni Sansaro, e Gennaro Vessio. 2021. «Integrating Contextual Knowledge to Visual Features for Fine Art Classification». *arXiv preprint* 2105.15028.
- [5] Castellano, Giovanna, e Gennaro Vessio. 2021. «Deep learning approaches to pattern extraction and recognition in paintings and drawings: an overview». *Neural Computing and Applications* 1 (20).
- [6] ———. 2021. «A deep learning approach to clustering visual arts». *arXiv* 2106.06234.
- [7] Cetinic, Eva, Sonja Grgic, e Lipic Tomislav. 2019. «A deep learning perspective on beauty, sentiment, and remembrance of art». *IEEE Access* 7.
- [8] Garcia, Nia, Benjamin Renoust, e Yuta Nakashima. 2020. «ContextNet: representation and exploration for painting classification and retrieval in context» *International Journal of Multimedia Information Retrieval* 9 (1): 17–30.
- [9] Hall, Peter, Cai Hongping, Wu Qi, e Tadeo Corradi. 2015. «Cross-depiction problem: Recognition and synthesis of photographs and artwork». *Computational Visual Media*, 1 (2): 91–103.
- [10] LeCun, Yann, Yoshua Bengio, e Geoffrey Hinton. 2015. «Deep learning». *Nature* 521: 436–44.

Un Oggetto Intelligente IoT per Migliorare le Visite Interattive di Siti di Interesse Culturale

Fabrizio Balducci¹, Paolo Buono¹

^{1,2} Università di Bari “A. Moro”, Italia,

¹fabrizio.balducci@uniba.it

²paolo.buono@uniba.it

ABSTRACT

Nell’Internet of Things, il mondo reale e quello virtuale sono connessi attraverso sensori che ricevono ed elaborano dati dall’ambiente circostante. Tali oggetti “intelligenti” permettono modalità di fruizione interattive e multimediali utili nei beni culturali per arricchire le esperienze dei visitatori. Si propone un sistema in cui un curatore museale pianifica e monitora visite tramite un’app che permette ai visitatori di svolgere attività che prevedono dei quesiti a cui dovranno rispondere con l’ausilio di una “bacchetta magica”. Tale oggetto interagisce con le opere del museo attraverso la tecnologia RFID. Test effettuati hanno mostrato che l’interazione tra oggetto IoT ed app mobile arricchisce l’esperienza dell’utente e del curatore museale che potrà usare la sua creatività ed esperienza per organizzare i contenuti del museo.

PAROLE CHIAVE

Beni culturali, IoT, android.

INTERVENTO

1. INTRODUZIONE E RELATED WORKS

Il lavoro qui presentato è sistema hardware-software che permette di coinvolgere visitatori di siti di beni culturali tramite interazione fisica con oggetti smart. L’Internet of Things ([3]) e la diffusione di dispositivi connessi sempre più “intelligenti” ed autonomi sta trasformando musei tradizionale in “smart-Museum”. Le fasi iniziali del progetto hanno visto diversi colloqui con curatori di musei per abbozzare prototipi ed individuare i reali fabbisogni degli esperti del dominio. Un mini-gioco quiz viene proposto durante una visita in punti individuati dal curatore. Il sistema è caratterizzato da una “bacchetta magica” con cui puntare fisicamente oggetti capaci di reagire a tale interazione, posizionati dal curatore.

Sono numerose le tecnologie nell’ambito del Cultural Heritage e relativi a visite di siti di interesse culturale. L’audioguida è lo strumento più familiare per un ascolto passivo di registrazioni ([5]); proiezioni a muro permettono anche interazioni gestuali ([6]). In ([7]) è illustrato l’uso di giochi per interagire con un luogo e con gli oggetti ivi contenuti. I QR code sono usati quotidianamente per trasferire testo, audio e video ([9]). Le tecnologie RFID hanno applicazioni diversificate e il loro uso ha semplificato la gestione dei dispositivi all’interno di musei dove è significativo l’accesso automatico ai contenuti ([4]).

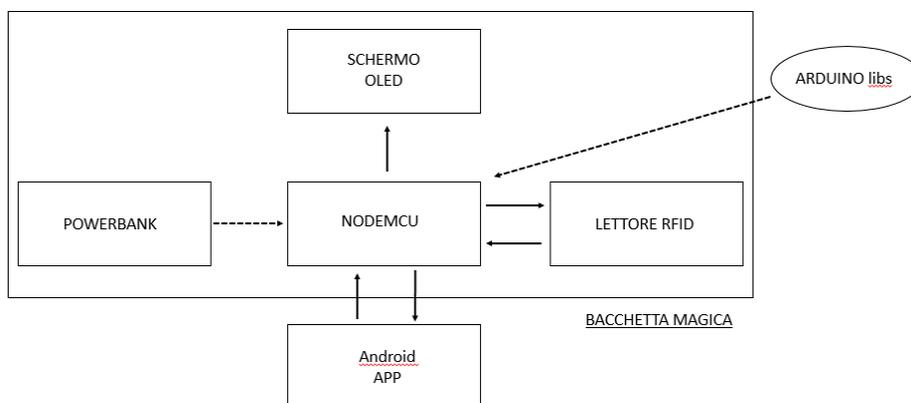


Figura 1: Architettura del sistema.

2. LA “BACCHETTA MAGICA”

La “bacchetta magica” (Fig.1) è uno strumento IoT con componenti hardware *Arduino* ([2]) che permette ad un visitatore di interagire con opere museali mentre il curatore monitora a distanza la visita tramite app Android. Le componenti utilizzate sono: *microcontrollore NodeMCU* che tramite wireless comunica con l’app Android; *lettore codici RFID* che scambia dati tramite radiofrequenza; *schermo OLED* per mostrare messaggi testuali; *power bank* per l’alimentazione. Il dispositivo effettua in ciclo tre operazioni in sequenza: i) controllo della connessione con dispositivi abbinati; ii) controlla ricezione messaggi; iii) lettura TAG di carte RFID il cui risultato (vero o falso) è visualizzato sullo schermo. Il prototipo è riportato in Fig.2.



Figura 2: Prototipo della ‘bacchetta magica’ inserita in una intelaiatura di alluminio.

3. APP

L’app è pensata per pianificare un percorso di visita tramite smartphone da dare a uno o più visitatori. Mentre in progetti come *Overtheview* ([8]) i contenuti sono concepiti per persone con disabilità visiva ed uditiva, in questo lavoro si presuppone una lettura standard di testo sullo schermo del device.

L’app realizzata prevede due modalità:

- i) configuratore di tappe della visita scegliendo punti di interesse da visitare a cui associare le domande che saranno proposte ai visitatori;
- ii) strumento interattivo dedicato ai visitatori che seguono un percorso prestabilito dalla guida e che nei punti d’interesse devono rispondere a quiz.

Il visitatore che impugna la bacchetta magica, vicino all’opera trova tre carte RFID, egli dovrà toccare con la bacchetta (che cela un lettore RFID) una di esse per visualizzare sullo schermo della bacchetta il risultato della scelta.

Le sezioni principali dell’app per il curatore sono:

- 1) *Home*, per la configurazione della bacchetta e la creazione di una visita (Fig.3-a) con un quiz (Fig.3-b) a cui associare il codice RFID della carta contenente la risposta esatta;
- 2) *Percorso*, utile per visualizzare le visite create (Fig.3-c) con dettagli e domande;
- 3) *Completate*, con i dati relativi alle sessioni concluse come: data della visita, numero di partecipanti e punteggi (Fig.3-d). I dati sono sincronizzati tra tutti i client e restano disponibili anche quando l’applicazione è offline.

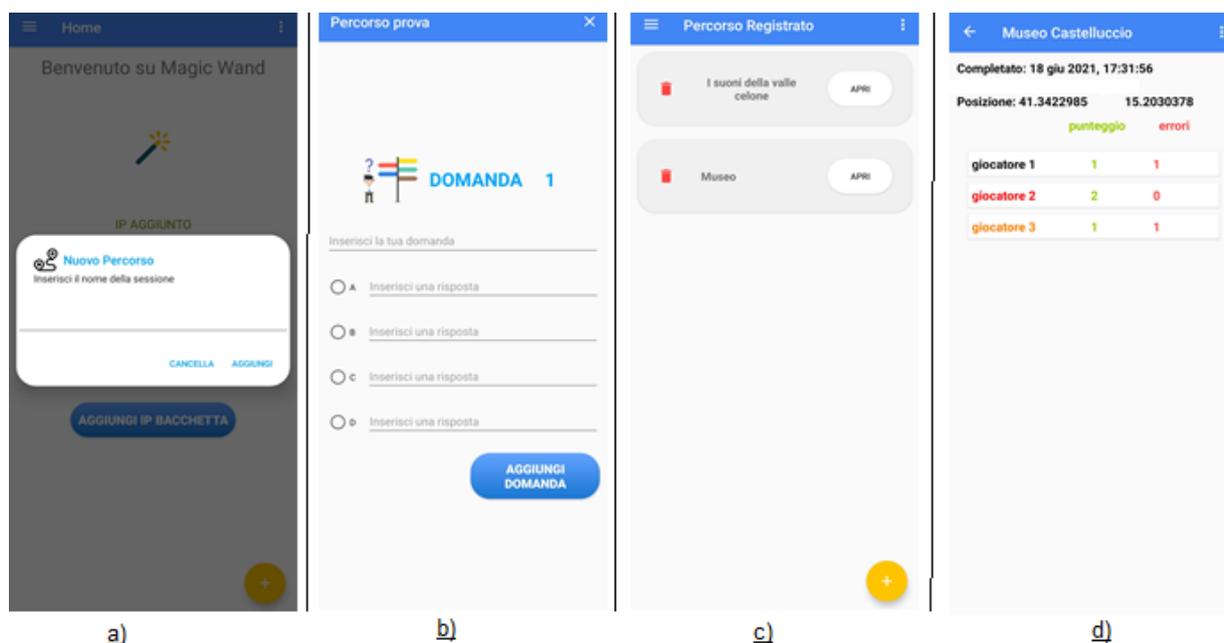


Figura 3: Schermate dell'app Android con il quale si può creare una visita (a), inserire un quiz educativo per un luogo di interesse (b), visualizzare i percorsi memorizzati (c), accedere ai dati su partecipanti e sulle attività eseguite nella visita museale (d).

4. SPERIMENTAZIONE

Si è effettuata una sperimentazione sul campo prevedendo due modalità di test (“curatore” e visitatore”) secondo l’approccio End-User Development (EUD) ([1]).

Nella modalità “curatore” la guida può definire i task che i visitatori effettueranno. Sono stati coinvolti 5 partecipanti (2 maschi e 3 femmine) con età media di 31 anni, di cui 3 con dimestichezza nell’uso di smartphone e app e 1 con dimestichezza su dispositivi intelligenti ed IoT. I task che i curatori hanno definito sono: 1) avvio dell’app e creazione di un account; 2) accesso all’applicazione e concessione dei permessi; 3) pianificazione di una visita interattiva inserendo le domande del quiz; 4) visualizzazione della visita creata; 5) creazione di un collegamento tra l’applicazione Android e la bacchetta magica; 6) posizionamento delle card RFID nei punti definiti e avvio visita; 7) monitoraggio sessione di visita; 8) controllo risultati della visita conclusa.

Due partecipanti non hanno riscontrato problemi nel completare tutti i task, uno (utilizzatore di smartphone non Android) ha avuto problemi con il task 5, uno con i task 2 e 5 (per distrazione) ed uno con i task 1, 2, 4 e 5 (scarsa dimestichezza con la tecnologia). Nel questionario sull’esperienza e le aspettative, tutti hanno espresso pareri positivi mostrando interesse nel progettare visite sfruttando la bacchetta magica ed interazioni con oggetti tangibili. Apprezzata anche la possibilità di usare il sistema senza connessione dati, evitando costi, problemi di poca copertura e velocità di connessione.

Per la modalità “visitatore”, dopo una breve spiegazione sul funzionamento della bacchetta magica e di come interagire con le carte RFID, 12 partecipanti con età compresa tra i 13 e i 69 anni hanno eseguito i task: 1) visualizzare i dati delle sessioni gioco sullo schermo della bacchetta; 2) eseguire la visita seguendo la guida museale ed interagendo con le carte RFID quando richiesto.

Tutti i soggetti sono riusciti a terminare i test e nel questionario sull’esperienza effettuata i pareri sono stati generalmente positivi; solo lo schermo OLED ha avuto giudizio negativo da tutti i partecipanti a causa delle sue dimensioni ridotte.

5. CONCLUSIONI E SVILIPPI FUTURI

In questo lavoro è stato presentato un sistema che permette di pianificare visite in siti di beni culturali e di far vivere una esperienza interattiva con oggetti smart e tangibili. Attraverso la “bacchetta magica” il visitatore si diverte e apprende informazioni sugli oggetti esposti nei musei. I test effettuati sia con curatori che con visitatori del museo hanno evidenziato facilità di configurazione e di utilizzo. Sviluppi futuri includono l’aumento delle dimensioni del display OLED integrato, l’ampliamento delle attività che si possono condurre, la personalizzazione della visita in base al tipo di utente.

6. RINGRAZIAMENTI

Lavoro supportato dal progetto “Gestione di oggetti intelligenti per migliorare le esperienze di visita di siti di interesse culturale” del programma REsearch For INnovation (REFIN) CUP:H94I20000410008 cod. F517D521 POR Puglia FESR FSE 2014-2020. Si ringrazia Antonio Ricchetti per l’aiuto nell’implementazione dell’app.

BIBLIOGRAFIA

- [1] Ardito, Carmelo, Paolo Buono, Maria Francesca Costabile, e et al. 2012. «End users as co-designers of their own tools and products». *Journal of Visual Languages & Computing* 23 (2): 78–90.
- [2] «Arduino Products». s.d. Arduino. Consultato 13 novembre 2021. <https://www.arduino.cc/en/Main/Products>.
- [3] Ashton, Kevin. 2009. «That ‘internet of things’ thing». *RFID journal*, giugno, 97–114.
- [4] Caviglione, Luca, Mauro Coccoli, e Alberto Grosso. 2011. «A framework for the delivery of contents in RFID-driven smart environments». *IEEE International Conference on RFID-Technologies and Applications*, 45–49.
- [5] Fisher, Jennifer. 2004. «Speeches of Display: Museum Audioguides by Artists». *Aural Cultures*.
- [6] Geller, Tom. 2006. «Interactive tabletop exhibits in museums and galleries». *IEEE Computer Graphics and Applications* 26 (5): 6–11.
- [7] Malegiannaki, Irini, e Thanasis Daradoumis. 2017. «Analyzing the educational design, use and effect of spatial games for cultural heritage: A literature review». *Computers & education* 108: 1–10.
- [8] Paddeu, Gavino, Anna Maria Marras, Andrea Ferrero, Francesca Pintori, e Antonio Mura. 2021. «Il progetto Overtheview: schema di progettazione per l’accessibilità museale». In *Book of extended abstracts of the 10th national conference*, 51–55. Pisa.
- [9] Pérez-Sanagustín, Mar, Denis Parra, Renato Verdugo, Gonzalo García-Galleguillos, e Miguel Nussbaum. 2016. «Using QR codes to increase user engagement in museum-like spaces». *Computers in Human Behavior* 60: 73–85.

Oxoce - Motore di ricerca tematico strutturato

Tiago Luis Gil

Università di Brasilia, Brasile – tiagoluisgil@gmail.com

ABSTRACT

Lo scopo di questa presentazione è mostrare le funzionalità di “Oxoce”, un sistema automatizzato per la scansione, l’organizzazione e la strutturazione dei dati storici ricavati da una pluralità di fonti. Oxoce è il nome della divinità yoruba a cui sono associate la caccia, la pesca e l’abbondanza. Oxoce agisce alla stregua di un motore di ricerca per un periodo e una regione specifici: per essere più precisi come un motore di ricerca storico. Il progetto prevede – almeno durante la fase di test – di includere solamente contenuti inerenti alla storia del Brasile coloniale, soffermandosi in particolare sul XVIII secolo inserendo progressivamente con il consolidamento del sistema, altri periodi e aree geografiche. Lo strumento è in grado di passare in rassegna libri, articoli, tesi e fonti storiche pubblicate alla ricerca di nomi di persone, date, luoghi, temi (su due diversi livelli), ma anche riferimenti bibliografici, individuando le pagine delle opere in cui tali dati sono stati citati. Questi dati vengono successivamente interrelati e organizzati in un ampio database, che consente di ricercare proprio una persona, in un dato momento e relativamente ad alcuni argomenti specifici. Il sistema opera, quindi, con le stesse modalità di un motore di ricerca internet, ma con la possibilità di definire un periodo storico e una regione; consente inoltre ampie rilevazioni bibliografiche su periodi specifici, permettendo di separare, nei risultati, i riferimenti bibliografici dai dati storici.

PAROLE CHIAVE

Motore di ricerca, database, ricerca in storia.

INTERVENTO

1. INTRODUZIONE

La proposta del motore di ricerca “Oxoce” trae spunto da diverse ispirazioni nell’ambito della ricerca storica. L’obiettivo principale di questo progetto, la costruzione di un motore di ricerca incentrato su un segmento cronologico e spaziale, è un tema ancora pionieristico. Possiamo evocare l’esperienza del progetto “Isidore”, un motore di ricerca incentrato sulle pubblicazioni di scienze sociali che costituisce già un grande avanzamento dal momento che consente di lavorare con una selezione in grado di rendere la ricerca molto più pertinente ([6];[7]). Numerose sono invece le iniziative volte a utilizzare le tecnologie di *text mining* per i dati storici, facendo ricorso al *Topic Models* e alla *Named Entity Recognition* (NER), ma non solo ([2];[3];[4];[5];[8]). Queste iniziative, tuttavia, si concentrano sempre sull’uso di raccolte specifiche, sui luoghi citati in un ampio corpus documentale, ad esempio ([8];[5]) ma non esplorano la possibilità di integrare nomi di persone, luoghi, date e temi. Peraltro, la ricerca dei nomi di persona è un tema abbastanza trascurato negli studi storici, proprio per la difficoltà di ricostruire i nomi integralmente, date le innumerevoli possibilità di combinazione di questi “oggetti”.

Nel corso degli anni Novanta si diffusero grandi motori di ricerca come AltaVista e Yahoo, che tuttavia nei primi anni del nuovo millennio vennero surclassati da Google. Questi motori di ricerca, benché estremamente diversi tra loro, si rivolgono tutti al grande pubblico e sebbene permettano di scoprire svariati contenuti interessanti per la ricerca, sono molto limitati quanto alle possibilità di selezione delle variabili che importano al ricercatore. Google Scholar, benché rivolto specificamente al mondo accademico, non consente allo storico di ricercare un determinato tema in uno specifico periodo. Risulta impossibile, ad esempio, rintracciare opere che trattino di poteri locali limitatamente all’età moderna ed è ancora più difficile circoscrivere questo tema e questo periodo in relazione ad una regione in particolare. Una ricerca su opere accademiche che trattino di poteri locali nella Lombardia del XVII secolo restituirebbe infatti risultati incerti e parziali. La proposta di “Oxoce” è quella di mettere a disposizione dello storico tutte queste funzionalità.

2. DESCRIZIONE DEL SISTEMA

Sviluppato in linguaggio Python, “Oxoce” è composto da diversi moduli di raccolta dati. In questa fase di sviluppo l’implementazione avviene tramite file PDF che contengono prodotti della ricerca scientifica come libri, articoli e capitoli di collane; una procedura che avviene manualmente, osservando il comportamento del codice di programmazione per ogni opera che viene inserita. Una volta che un file PDF fa il suo ingresso nel sistema, viene sottoposto a una serie di processi,

che passeranno in rassegna il testo alla ricerca di informazioni in forma organizzata. Il primo processo attiene alla ricerca dei nomi di persona: è stato infatti realizzato un algoritmo appositamente per questo. La struttura nominale in portoghese è relativamente complessa e può essere composta, inclusi nomi e cognomi, da uno a otto sintagmi. Lo strumento è in grado, attingendo ad un ampio dizionario, di identificare nomi e cognomi maschili e femminili, di comporli per esteso e separandoli in una colonna specifica. Una volta raccolti e organizzati questi dati, la frequenza di determinati nomi sulle stesse pagine consentirà l'analisi delle reti sociali dei personaggi storici, anche se sotto forma di suggerimento, lasciando al ricercatore la lettura esatta dei testi presi in esame.

Il secondo trattamento porta all'individuazione delle date che compaiono nel testo, sia per quel che riguarda gli anni specifici che con riferimento ai decenni e ai secoli. Una volta raccolte, le date vengono analizzate in termini quantitativi e si ottengono le medie, le mode, le mediane e gli scarti tipo di ciascuna pagina e dell'opera nel suo complesso, il che consente al programma di indicare automaticamente il periodo storico coperto dall'opera, indicando il periodo maggiormente citato. Anche un'opera che si occupa di un intero secolo può concentrarsi maggiormente su un decennio: il codice è in grado d'identificare questa caratteristica.

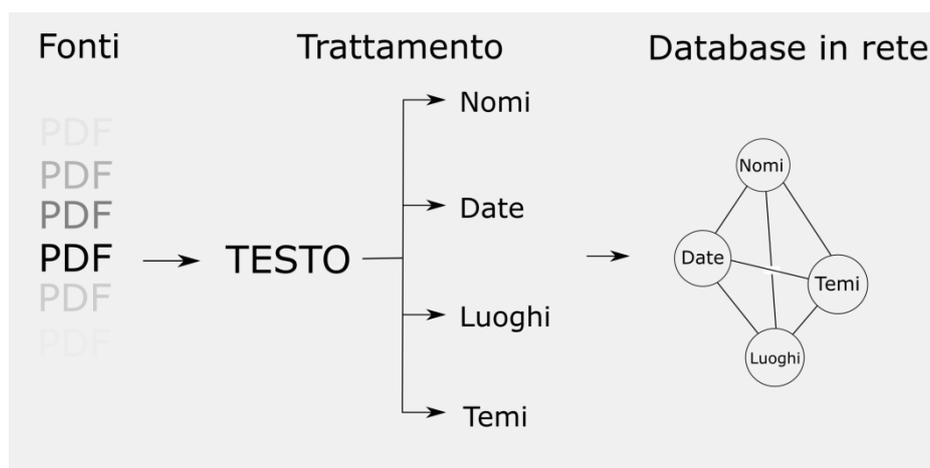


Figura 1: Struttura del Sistema Oxoce

Il terzo trattamento prende in esame le tematiche. Questa elaborazione si divide in due fasi: una induttiva, basata sui *topic models*, che elimina le *stopword*, lasciando solo il materiale residuo che viene quantificato e da cui, con vari procedimenti statistici, vengono definiti i soggetti più ricorrenti. Successivamente interviene l'elaborazione deduttiva, basata sulla NER: sono stati creati all'uopo decine di dizionari tematici che raccolgono gruppi di parole associate a ciascun argomento. Queste parole vengono ricercate nel testo e, a seconda del numero di volte in cui compaiono, consentono una maggiore o minore caratterizzazione del testo in base a quel tema.

Una quarto trattamento consente la ricerca le parole associate ai luoghi (prendendo in qualche misura spunto dalla NER), tanto nella forma in cui vengono indicate in portoghese, quanto attraverso l'impiego del dizionario topografico presente nel progetto "Atlas Digital da América Lusa" ([1]), che contiene migliaia di riferimenti nominativi alle località dell'epoca: tutto ciò consente di raccogliere e organizzare una grande quantità di luoghi nel database.

Il risultato di questi quattro processi di trattamento viene raccolto e organizzato per pagina, mettendo i dati in relazione tra loro. Possiamo così cercare le variabili separatamente, trovando opere consigliate da leggere, ma possiamo anche effettuare ricerche più complesse, ad esempio, le persone chiamate "Antonio" in pagine che menzionano una città specifica e un determinato arco di tempo. Il risultato non è specifico per quel personaggio, ma può costituire una buona approssimazione. Il numero di volte in cui le tre cose (persona, tempo e luogo) compaiono insieme può, successivamente, essere utilizzato per ottenere informazioni aggiuntive e consentire ricerche più efficaci.

Saranno infine raccolti tutti i riferimenti bibliografici utilizzati nel materiale preso in esame; essi permetteranno di associare a tali opere determinati temi, periodi e luoghi. Ciò consentirà un'indicazione più puntuale delle opere consigliate per determinate ricerche, oltre a consentire l'osservazione di reti di citazioni e tradizioni storiografiche. Permetterà inoltre di individuare autori e gruppi di autori interessati a determinati temi e periodi, un aspetto che potrà far luce anche sulla storiografia.

3. CONCLUSIONI

Riteniamo che questo strumento possa risultare estremamente utile per la ricerca d'informazioni di buona qualità quando si conduce una ricerca in campo storico, lasciando al ricercatore il compito di leggere e analizzare i dati ottenuti, nonché di valutarne i limiti.

BIBLIOGRAFIA

- [1] «Atlas Digital da América Lusa». s.d. Consultato 7 ottobre 2021. <http://lhs.unb.br/atlas>.
- [2] Brauer, René, e Mats Fridlund. 2013. «Historicizing topic models, a distant reading of topic modeling texts within historical studies». In *International Conference on Cultural Research in the context of "Digital Humanities"*. St. Petersburg: Russian State Herzen University.
- [3] Ehrmann, Maud, Ahmed Hamdi, Elvys Linhares Pontes, Matteo Romanello, e Antoine Doucet. 2021. «Named Entity Recognition and Classification on Historical Documents: A Survey». *arXiv* 2109.11406.
- [4] Graham, Shawn, Ian Milligan, e Scott Weingart. 2015. *Exploring big historical data: The historian's microscope*. World Scientific Publishing Company.
- [5] McDonough, Katherine, Ludovic Moncla, e Matje Van de Camp. 2019. «Named entity recognition goes to old regime France: geographic text analysis for early modern French corpora». *International Journal of Geographical Information Science* 33 (12): 2498–2522.
- [6] Pouyllau, Stéphane. 2011. *ISIDORE : une plateforme de recherche de documents et d'information pour les Sciences Humaines et Sociales*.
- [7] ———. 2012. «Les moteurs de recherche profitent aussi de la sémantique». *Documentaliste-Sciences de l'Information* 48 (4): 36–37.
- [8] Won, Miguel, Patricia Murrieta-Flores, e Bruno Martins. 2018. «Ensemble named entity recognition (ner): evaluating ner Tools in the identification of Place names in historical corpora». *Frontiers in Digital Humanities* 5 (2).

Sessione Contenuti 1
George Boole

Funzione ecosistemica e funzione storiografica della narrazione ambientale videoludica

Matteo Genovesi
mattjean@hotmail.it

ABSTRACT

Come argomenta Wolf, una componente indispensabile di un mondo immaginario risiede nella mappatura degli spazi, a prescindere dalle relazioni più o meno esplicite col mondo reale: le mappe, infatti, mettono in relazione una serie di luoghi entro i quali avvengono gli eventi e in cui si muovono determinati personaggi ([11]: 153-198).

Focalizzando l'attenzione sul medium videoludico, e in particolare nell'ultimo quindicennio, la mappatura degli spazi assume altresì un ruolo molto rilevante, non solo per la delimitazione dei tratti morfologici dei mondi virtuali, ma anche per la gestione della libertà esplorativa dell'utente e la sua potenziale funzione narrativa. Tramite le possibilità di movimento, infatti, l'utente può conoscere un racconto insito negli spazi tramite elementi ambientali che racchiudono importanti porzioni narrative secondo il processo dell'environmental storytelling ([5]).

La mia proposta qui presente, al momento di carattere preliminare in vista di una futura ricerca più ampia, intende evidenziare due possibili macro-funzioni dell'environmental storytelling. Da un lato, la funzione ecosistemica, evidenziabile in videogiochi che creano universi narrativi immaginari capaci di varcare i confini del medium d'origine ed essere distribuiti anche su altri settori d'intrattenimento secondo logiche transmediali; dall'altro lato, la funzione storiografica, evidenziabile in videogiochi ancorati al medium d'origine e focalizzati su avvenimenti storici e/o ambienti realmente esistenti, con l'obiettivo di fornire spunti di riflessione con cui comprendere il passato e l'oggi.

PAROLE CHIAVE

Funzione ecosistemica, funzione storiografica, transmedia storytelling, environmental storytelling.

INTERVENTO

1. INTRODUZIONE

Come sostiene Jenkins, gli sviluppatori che vogliono realizzare un'efficace narrazione ambientale videoludica devono calarsi nel metaforico ruolo di architetti narrativi: infatti, è importante che lo spazio venga concepito come un elemento funzionale alla contestualizzazione narrativa sfruttando le possibilità di movimento offerte all'utente, il quale dev'essere costantemente invogliato a ricercare indizi sparsi per gli ambienti ([5]). A partire dal saggio di Jenkins ci sono stati in seguito altri studi che hanno approfondito questo tema ([7];[3];[2]), sottoscrivendo sempre l'importanza della spazialità in quanto componente fondamentale per la strutturazione del racconto.

La delimitazione dei connotati morfologici degli spazi virtuali nel medium videoludico, quindi, serve sia per strutturare le possibili azioni dinamiche dell'utente così come per caricare potenzialmente le medesime azioni di uno scopo narrativo oltre che ludico. Navigando negli spazi, l'utente può infatti trovare oggetti, personaggi, strutture e altri elementi potenzialmente in grado di arricchire la sua conoscenza del mondo virtuale in cui si muove. Vasti mondi immaginari o spazi circoscritti possono poi spingere l'utente al di là della fruizione videoludica secondo due differenti funzioni, quella ecosistemica e quella storiografica.

2. FUNZIONE ECOSISTEMICA

La creazione di universi narrativi distribuiti su molteplici settori d'intrattenimento rappresenta una realtà ormai consolidata nel panorama mediale contemporaneo. Jenkins, con la terminologia di transmedia storytelling, indica una narrazione frammentata e distribuita su molteplici media, in cui ogni singolo medium, sfruttando le proprie caratteristiche espressive peculiari, offre il suo prezioso contributo allo sviluppo del racconto ramificato ([6]).

Gli studi di Jenkins hanno poi ispirato molti altri contributi accademici volenterosi di approfondire questi fenomeni, condotti talvolta con metodologie diversificate tra loro, come accaduto in Italia con l'approccio testualista di Carini ([1]) e quello ecosistemico di Innocenti e Pescatore ([4];[8]). A prescindere dai differenti metodi, ciò che emerge da questi studi è la considerazione del medium videoludico come parte integrante, ma non fondante, delle realtà transmediali

contemporanee. La situazione è pressoché la medesima negli studi internazionali sulla narrazione transmediale, che salvo pochi contributi ([9];[10]), si incentrano prevalentemente sull'analisi di universi narrativi che coinvolgono i videogiochi ma non nascono con essi.

La funzione ecosistemica, richiamando gli studi dei già citati Innocenti e Pescatore sugli ecosistemi narrativi, si esemplifica in quei videogiochi che, grazie alla loro narrazione ambientale, creano degli universi immaginari concretamente espansi su altri settori d'intrattenimento, sviluppando così dei racconti transmediali in cui il fruitore viene spinto a muoversi tra molteplici media per trovare ogni frammento narrativo.

Nell'ultimo quindicennio nel medium videoludico si sono registrati vari casi che hanno concretamente creato universi narrativi transmediali, avvalorando la loro soggiacente funzione ecosistemica. Tra i tanti esempi potenzialmente citabili, altresì meritevoli di diventare ampi casi di studio futuri, merita sicuramente considerazione la saga di *Mass Effect* (2007- in corso): attualmente formata da sei videogiochi, a partire dal 2007 delinea un'ampia cosmologia fantascientifica capace di coinvolgere sia la sfera letteraria, come dimostrano i sette romanzi e le dodici serie a fumetti, così come l'orizzonte cinematografico grazie a un film d'animazione.

3. FUNZIONE STORIOGRAFICA

Grazie alla sua continuativa evoluzione espressiva, il medium videoludico negli ultimi anni ha raggiunto una maturazione tale da poter non solo costituire infrastrutture per narrazioni ecosistemiche, ma anche per l'approfondita rappresentazione virtuale di eventi e ambienti concretamente esistenti.

La funzione storiografica della narrazione ambientale videoludica intende focalizzarsi su quei videogiochi basati su contesti narrativi legati ad avvenimenti realmente accaduti, in cui l'environmental storytelling serve per aumentare la conoscenza dei fatti rappresentati e non dimenticarne la loro portata al giorno d'oggi, stimolando dialoghi non solo tra gli utenti, ma anche tra utenti e sviluppatori.

Tra i tanti casi di studio potenziali, è lecito citare almeno: *Valiant Hearts: The Great War* (Ubisoft Montpellier, 2014), videogioco che segue le tragiche vicende di quattro personaggi durante la Prima Guerra Mondiale, tratteggiata con approfondita accuratezza grazie a oggetti collezionabili concretamente esistenti (come alcune foto d'epoca). *The Town Of Light* (LKA, 2016): realizzato in Italia dopo vari sopralluoghi al vecchio manicomio di Volterra corredate da molteplici ricerche sul trattamento sanitario dei pazienti, il videogioco si incentra su una donna che torna nell'edificio toscano per confrontarsi col proprio passato, finendo col trovare indizi negli ambienti che ricostruiscono nella sua memoria i maltrattamenti subiti.

4. RISULTATI ATTESI

Lo studio della funzione ecosistemica esige all'inizio un'analisi testuale strettamente videoludica, in modo da individuare i contenuti narrativi legati all'esplorazione ambientale preposti all'espansione su altri media, per poi indagare la conseguente costruzione transmediale e, infine, le effettive attività di fruizione multimediale degli utenti così come i loro scambi di opinioni.

Un processo di indagine simile riguarda anche la funzione storiografica, in cui, a seguito dell'analisi testuale mirata a individuare quegli elementi narrativi ambientali potenzialmente capaci di aumentare la conoscenza di fatti realmente accaduti, seguirà un focus sulle attività post-fruizione videoludica degli utenti, osservando i discorsi sociali tra di loro e con gli sviluppatori stessi.

Entrambe le funzioni della narrazione ambientale sinteticamente qui presentate, hanno quindi lo scopo di valorizzare non solo la concreta maturazione espressiva del medium videoludico, ma anche di sottolineare la sua costante presenza all'interno dei processi culturali e comunicativi contemporanei.

LUDOGRAFIA

- *Mass Effect* (saga), Bioware, 2007 – in corso;
- *The Town Of Light*, LKA 2016;
- *Valiant Hearts: The Great War*, Ubisoft Montpellier 2014.

BIBLIOGRAFIA

- [1] Carini, Stefania. 2009. *Il Testo Espanso. Il telefilm nell'età della convergenza*. Vita e pensiero.
- [2] D'Armenio, Enzo. 2014. *Mondi Paralleli. Ripensare l'interattività nei videogiochi*. Unicopli.
- [3] Fernández-Vara, Clara. 2011. «Game Spaces Speak Volumes: Indexical Storytelling». In *Proceedings of the 2011 DiGRA International Conference. Think Design Play*. Vol. 6.

- [4] Innocenti, Veronica, e Guglielmo Pescatore. 2017. «Narrative Ecosystems. A Multidisciplinary Approach to Media Worlds». In *World Building. Transmedia, Fans, Industries*, Marta Boni, 164–83. Amsterdam: Amsterdam University Press.
- [5] Jenkins, Henry. 2004. «Game Design as Narrative Architecture». In *First Person: New Media as Story, Performance, and Game*, Pat Harrigan, Noah Wardrip-Fruin, 118–30. The MIT Press.
- [6] ———. 2006. *Convergence Culture: Where Old and New Media Collide*. New York University Press.
- [7] Nitsche, Michael. 2008. *Video Game Spaces. Image, Play and Structure in 3D Worlds*. The MIT Press.
- [8] Pescatore, Guglielmo. 2018. *Ecosistemi narrativi*. Milano: Carocci.
- [9] Veugen, Connie. 2016. «Assassin’s Creed and Transmedia Storytelling». *Simulations* 8 (2): 1–19.
- [10] Wilk, Elisa. 2019. «More Interaction, More Story, More Lore: Motivations Related to Game-centric Transmedia». In *Proceedings of the 2019 DiGRA International Conference: Game, Play and the Emerging Ludo-Mix*.
- [11] Wolf J.P., Mark. 2012. *Building Imaginary Worlds: The Theory and History of Subcreation*. Routledge.

Narrazioni mediatiche delle emergenze e processi di costruzione di *quest*: quali possibili analogie? L'incidente del “corrupted blood” in “World of Warcraft”

Luigi Giungato¹, Patrizia Miggiano²

¹Università della Calabria, Italia – lgiungato@gmail.com

²Università del Salento, Italia – patrizia.miggiano@unisalento.it

ABSTRACT

Il presente contributo intende riflettere sulla dimensione narrativa della disinformazione dell'emergenze nelle società dell'informazione, attraverso il ricorso a un approccio narratologico e critico-semiotico, con specifico riferimento al contributo apportato dalle teorie narratologiche che analizzano la produzione ricorsiva industriale di schemi narrativi tipici dei *massive(ly) multiplayer online role playing gaming* – o *MMORPG* – e dei *serial* ([29];[4]). Attraverso questa architettura di ricerca e mediante l'analisi delle implicazioni – ludiche e reali – derivate dal cosiddetto “incidente del corrupted blood”, avvenuto nel videogame “World of Warcraft”, si mirerà, dunque, alla rilevazione di schemi di comportamento da parte degli utenti ([18]) che possano risultare non solo ricorsivi e tipici delle dinamiche narrative partecipate proprie dei *MMORPG* ([22]), ma anche di determinati processi comunicativi tipici della contemporaneità digitale, con particolare riferimento alla proliferazione della disinformazione e delle fake-news ([17];[27];[12]).

PAROLE CHIAVE

Games studies, disinformazione, news media, role playing games, serialization.

INTERVENTO

1. INTRODUZIONE

Negli ultimi vent'anni, la narrazione mediatica dei contesti emergenziali ha subito una profonda mutazione in termini di prassi, strumenti e funzioni, tanto da rendere necessaria una riflessione sulla progressiva *ipermediatizzazione* del racconto della crisi, il quale, a sua volta, si nutre del materiale offerto dal reale per la strutturazione del “conflitto delle interpretazioni” ([28]): è così che, parafrasando Bourdieu ([7]), lo spazio digitale diviene il “luogo permanente per la definizione della realtà”. In altre parole, la produzione e la diffusione, su scala globale, delle diverse narrazioni mediatiche dell'emergenza finisce per assumere i tratti di una vera e propria contesa, che aspira a dotare di oggettività l'interpretazione del reale.

In tempi ancora più recenti, la pervasività dei *social network* (ivi incluse le piattaforme di messaggistica interpersonale) e la conseguente confluenza graduale fra *online* e *offline* – che, in una nota espressione di Floridi ([17]), è definita “onlife”; l'iper-frammentazione e la proliferazione delle fonti informative ([30]); la standardizzazione dei codici e, soprattutto, l'ibridazione fra corpo e *personal medium* (*smartphone*) come protesi di interazione tra il soggetto e la realtà hanno determinato fenomeni quali la *polarizzazione*, le *echo-chamber* ([27]) e la riontologizzazione della realtà ([17]) e un cambiamento nel ruolo degli algoritmi di gestione dell'agenda informativa, imponendo di fatto un ripensamento del mutato rapporto tra emergenza e media.

Il presente contributo intende riflettere sulla dimensione narrativa delle emergenze nella società dell'informazione, partendo dal presupposto che si tratti di un importante osservatorio di narrazioni parziali, ossia di un terreno dialettico che genera sempre nuovi discorsi in grado di incidere significativamente sugli assetti di potere del mondo contemporaneo.

Attraverso un approccio narratologico e critico-semiotico (con specifico riferimento al contributo apportato dalle teorie che analizzano la produzione ricorsiva industriale di schemi narrativi seriali tipici di contesti apparentemente remoti, come quelli dei *Massively Multiplayer Online Role Playing Game* e dei *serial*) ([29];[4]), si mirerà, dunque, alla rilevazione di anomalie comunicative che minano l'integrità, l'affidabilità e l'interoperabilità degli assetti informativi, generando dinamiche tipiche della contemporaneità (con particolare riferimento alla disinformazione e alle fake-news) ([17];[27];[8];[12]).

Va osservato, preliminarmente, che l'aspetto metodologico – improntato, in via prioritaria, all'osservazione delle strutture narrative – deve tenere necessariamente in conto la pluralità di forme differenti, ognuna riferita, potremmo dire, ai molteplici mondi possibili che ogni narrazione ludica costruisce per e attorno al giocatore.

Per tale ragione, riteniamo possa risultare coadiuvante un'analisi *user oriented*, attenta alle esigenze e alle motivazioni che si celano alla base dei comportamenti degli utenti, sulla scia, per esempio, della ricerca condotta da Dalisay, Kushin, Kim, Forbes, David, Somera (2021), atta all'individuazione di una possibile correlazione fra i diversi target di giocatori e le loro attitudini nei confronti dell'informazione e della politica.

Il contributo, in ultima analisi, quindi, intende proporre un possibile studio comparato tra le prassi di produzione di contenuti dei MMORPG, le strategie di contrasto adottate dai produttori di videogame per contrastare le manomissioni del gioco stesso da parte degli utenti e la generazione/propagazione di strutture e schemi narrativi non prevedibili, propri di determinate narrazioni mediatiche delle emergenze della contemporaneità, al fine di riflettere sui possibili risvolti del rapporto tra soggetto e costruzione sociale della realtà.

2. LE STRUTTURE NARRATIVE DI “WORLD OF WARCRAFT”

Negli ultimi venti anni, il panorama dell'intrattenimento virtuale si è caratterizzato per la proliferazione delle piattaforme ludiche, esaltato dalla possibilità di accesso multipiattaforma (computer, smartphone, console etc).

Il videogame *World of Warcraft*, prodotto dalla Blizzard Entertainment, è certamente uno dei MMORPG (Massive(ly) Multiplayer Online Role-Playing Game) ad ambientazione fantasy più giocati al mondo.

Sui cinque continenti del mondo virtuale di *Warcraft*, dal 2004, decine di milioni di utenti da tutto il mondo, attraverso i loro *alter ego* digitali, vivono quotidianamente una serie di avventure e partecipano alle storie create per loro da migliaia di storyteller e programmatori, tra i più capaci della scena internazionale ([18]).

Ogni giocatore, in base alle imprese compiute e alle alleanze stipulate nelle città e nei villaggi disseminati sul pianeta di Azeroth, è in grado di far progredire il proprio personaggio fino a divenire un vero e proprio *top player*, riconosciuto da tutta la comunità virtuale. Ma gli eroi del mondo di *Warcraft*, pari ai semidei della mitologia antica, sono i personaggi cosiddetti *non giocanti*, o PNG, guidati dai programmatori della Blizzard o dalle loro A.I. Sono questi che guidano le tante fazioni presenti nel gioco alla conquista del potere, conducono gli eserciti in battaglia e assoldano migliaia di truppe e avventurieri in imprese pericolose e avvincenti. L'entropia narrativa della storia viene alimentata da sempre più agguerriti antagonisti e intrecci narrativi, creati al fine di porre continuamente e *serialmente in gioco* l'interazione tra le avventure del mondo virtuale stesso e i giocatori, con un meccanismo molto simile a quello di serie tv di successo come *Il Trono di Spade* o *The Walking Dead*.

Gli abbonati della Blizzard *devono* poter giocare in virtù della sottoscrizione di un abbonamento; di conseguenza, la produzione assume dei narratori con l'incarico della creazione continua di *quest* (ovvero di *avventure*).

Tema predominante di molti di questi prodotti è una sorta di apocalisse imminente sui destini di tutto il mondo virtuale. L'intento di scongiurarla – o di facilitarla – compone la dicotomia manichea alla base del motore narrativo di tutto l'universo immaginario del gioco, fungendo da contesto anche – e, forse, soprattutto – all'orientamento politico ed etico dei personaggi creati dai singoli utenti.

3. NARRAZIONE MEDIATICA DELLE EMERGENZE E PROCESSI DI COSTRUZIONE DI QUEST: QUALI POSSIBILI ANALOGIE?

L'evento mediatico classico descritto nel 1992, all'apice dell'era dei mass media analogici, da Dayan e Katz ([14]) era costituito da uno *streaming* a bassa interazione, perlopiù passiva. L'evento mediatico avveniva nel *medium* e toccava l'individuo attraverso la sinapsi della *fredda* superficie di contatto (schermo della tv, radio, carta del giornale, etc), da cui poi veniva traslato nel mondo reale.

Nell'evento mediatico digitale, invece, il soggetto può interagire nel contesto dello stesso *medium*, mantenendo l'azione all'interno delle stesse *regole fisiche* dello spazio di appartenenza del mezzo.

Il concetto di *agitation*, coniato da Tolz e Teper (2018), si fonda su un assunto: la narrazione *mette in scena* i conflitti nei media, con un meccanismo di coinvolgimento attivo nei confronti del pubblico all'interno dell'arena virtuale stessa, su temi socialmente e politicamente sempre più rilevanti, in ragione del fatto che il fine della creazione artificiale del conflitto sia la proliferazione del conflitto stesso.

La notizia, inserita nel contesto dei social, diviene un'*avventura*, nella quale il soggetto, nei panni dell'*eroe*, intraprende il suo viaggio verso il ristabilimento dell'equilibrio. Ognuno può fare la sua parte, può condividere, commentare, ribattere. Più è alta l'entropia causata da tali interazioni, maggiori sono i volumi di traffico, economici, sociali e politici prodotti, più alto è il soddisfacimento da parte del pubblico nel momento dell'individuazione di rassicuranti schemi interpretativi

ricorsivi a cui affidarsi ([15]). Da questo punto di vista, quindi, risulta suggestivo l'accostamento tra la funzione delle *(fake)news* e degli algoritmi dei *social*, come veri e propri *generatori di entropia*, e la funzione tipica dello *storytelling* all'interno dei MMORPG, in quanto processo industriale per lo sviluppo di conflitti narrativi.

La produzione industriale di storie, o narrazioni, deve necessariamente, a sua volta, presentare una schematicità: ci si basa su strutture date e replicabili in infinite varianti, nelle quali, a grandi linee, un eroe intraprende un percorso di risoluzione dei conflitti, contrastato da vari antagonisti, in cerca di un *premio*.

In tal senso, gli algoritmi di produzione si pongono come strumenti fondamentali per coadiuvare i produttori di contenuti. Nell'osservazione della produzione e proliferazione di narrazioni e contro-narrazioni relative all'emergenza dovuta al Covid-19, ad esempio – e più in generale, nel caso di narrazioni mediatiche delle emergenze – è riscontrabile molto spesso e su differenti piattaforme, sia in Italia che all'estero, la produzione di racconti che rispettano e riproducono una logica strutturale di questo tipo. L'uso di termini generatori di contrapposizione, la schematizzazione degli attori, il diverso schema narrativo utilizzato a seconda dell'orientamento politico e a parità di dati reali – anche mediante operazioni di *debunking* – rivelano la creazione di uno schema narrativo teso alla formazione di un conflitto narrativo, che intrattiene con la realtà solo un legame di verosimiglianza ([12]).

D'altra parte, anche le recenti rivelazioni da parte dell'ex manager di Facebook, Frances Haugen, pubblicate sui media internazionali¹, rivelano l'attitudine degli algoritmi dei *social* a favorire il conflitto fra gli utenti, esattamente secondo la stessa logica del *gaming*.

È tuttavia possibile che, saltuariamente, all'interno degli schemi narrativi del gioco – o dell'opinione pubblica digitale – se ne creino alcuni non previsti, che vanno ad aggiungersi a quelli costruiti dagli *storyteller*. In alcuni casi, essi possono risultare anche in conflitto con gli altri. Il meccanismo di creazione di tali codici interpretativi inediti pare ricalcare quello di *ipercodifica* descritto da ([15]), per il quale è possibile che, in presenza di codici insufficienti a decodificare un testo complesso, il lettore proceda nella creazione di connessioni inedite e imprevedibili, come osservato a volte nei confronti dei *serial*.

4. CASO STUDIO: L'INCIDENTE DEL “CORRUPTED BLOOD”

È opportuno evidenziare, a questo punto, che tra le caratteristiche principali di un MMORPG vi è quella di creare o di stimolare la creazione di una *community* di giocatori che agisce su più livelli: da quello più strettamente associato alle dinamiche di gioco (*in game*), a quello metanarrativo della chat (*off game*), passando per la rete di relazioni *out of the game* che si svolgono attraverso i *social*, fino ad arrivare a un livello che, potremmo dire, si situa in *real life*. In questi due ultimi livelli, si concentrano tutta una serie di pratiche meta-ludiche che, nella maggior parte dei casi, sono costituite da discussioni e approfondimenti non solo e non tanto afferenti la narrazione stessa, quanto riguardanti l'individuazione di quelli che, gergalmente, vengono definiti *bug*, *crack*, *cheats* e *tips&tricks*, ovvero i metodi più o meno leciti (ossia non previsti dai programmatori e dai narratori) per *manomettere* il gioco stesso, a patto, naturalmente, che ogni progetto distruttivo da parte degli utenti si mantenga pur sempre all'interno dello stesso codice e, quindi, dello stesso *mondo possibile* del gioco. Possiamo quindi dire che, mentre al livello *in game*, i giocatori utilizzano gli avatar-eroi per sconfiggere l'antagonista all'interno di un codice narrativo non-lineare fornito dal narratore; a un livello *off game* e *out of the game* più giocatori assumono su di essi il ruolo di eroi e si coalizzano per sconfiggere, ponendolo sul piano dell'antagonista, il narratore – se non il gioco stesso.

Tale meta-scontro è così stringente e dannoso per la sopravvivenza del gioco stesso che il contrasto alle dinamiche distruttive da parte degli utenti rappresenta una delle attività più onerose e impegnative di ogni casa di produzione videoludica.

In tale dinamica di manomissione dello schema narrativo proposto – che genera, a sua volta, nuove soluzioni narrative imprevedute nelle quali, solitamente, il narratore assume il ruolo di antagonista – si può cogliere una rassomiglianza con il meccanismo di un particolare aspetto della disinformazione sui *social network*, in particolare in contesti emergenziali.

Mentre, infatti, per l'utente ideale della narrazione mediatica, eroi e antagonisti della storia narrata sono quelli dettati dal narratore, per alcuni utenti la narrazione *mainstream* diviene essa stessa l'antagonista da sconfiggere, mediante il ricorso a tattiche e risorse non previste dai narratori, generate dalla stessa *community*, mirate alla manomissione del *mondo possibile* ma che, tuttavia, sfruttano le stesse possibilità fornite dal codice stesso del *social* nel quale si propagano.

Un caso che potremmo analizzare per osservare tale fenomeno, in cui una *community* ha intrapreso una quest finalizzata alla distruzione del gioco stesso, è quella del cosiddetto evento del *corrupted blood*.

¹ <https://www.ilsole24ore.com/art/facebook-l-accusa-dell-ex-manager-profitti-piu-importanti-salute-utenti-AEky8Ln>

Nel settembre del 2005, i creatori di storie della Blizzard introdussero nel gioco una nuova *quest* dedicata ai giocatori di rango più elevato, al centro della quale era un nuovo PNG, il dio Hakkar, intenzionato a sterminare tutte le popolazioni del mondo grazie a un morbo debilitante.

Per sconfiggerlo, tutti i più forti giocatori di Warcraft furono chiamati a coalizzarsi, tentati soprattutto dai ricchi tesori promessi in caso di sconfitta del nemico.

L'avatar del dio Hakkar, tuttavia, aveva una peculiarità molto particolare: era in grado, cioè, di scagliare, su chiunque lo colpisse, un morbo – in linguaggio tecnico *debuff* – capace di debilitare l'avversario, gradualmente, fino alla morte.

Il morbo, definito dai narratori *corrupted blood*, poteva, inoltre, essere trasmesso a tutti i personaggi che si approssimassero al portatore infetto.

Apparentemente l'idea degli *storyteller* era quella di rendere inasprire la riuscita dell'impresa e di simulare una vera e propria malattia sui corpi immateriali dei giocatori.

Il problema, tuttavia, sorse quando fu evidente un *bug* del gioco, ovvero un gravissimo errore non previsto in fase di programmazione. Il *corrupted blood*, infatti, concepito per esaurirsi al momento dell'uccisione del personaggio colpito, non svaniva al momento del *respawn*, ovvero della rigenerazione automatica del personaggio dopo la morte. Ciò significava che il personaggio ucciso, una volta materializzatosi nuovamente in un'altra zona differente della mappa (quasi sempre il tempio o la piazza principale di una grande città), restava infetto e, soprattutto, altamente contagioso.

Il *corrupted blood*, così, da *debuff* limitato alla singola area di gioco del mostro Hakkar divenne, in breve tempo, un'epidemia. Dal momento che i suoi effetti erano stati tarati sui giocatori più forti e sui loro personaggi più anziani, esso risultava tremendamente letale per quelli più giovani.

Non ci volle molto perché le città di *World of Warcraft* si riempissero di cadaveri digitali che perivano anche dopo essere stati rigenerati.

Le reazioni da parte dei giocatori furono le più svariate. In un primo momento, alcuni personaggi guaritori si misero a disposizione di tutti gli utenti per limitare, tramite i loro incantesimi, gli effetti letali del virus. Nelle città, si crearono, così, spontaneamente dei veri e propri lazzaretti. La maggior parte dei giocatori, tuttavia, preferì la fuga dai grandi centri e il rifugio nelle sterminate lande disabitate virtuali di Azeroth.

Tuttavia, la necessità di recarsi nuovamente nelle città al fine di intraprendere le varie *quest* indispensabili per la progressione del gioco, portò inevitabilmente a una ripresa del contagio e alla stessa interruzione delle dinamiche narrative. I vertici della Blizzard, una volta individuato il problema, decisero allora di imporre una quarantena obbligatoria a tutti i giocatori, per guadagnare il tempo necessario a correggere il *bug*, senza essere costretti a resettare interamente tutti i personaggi. Molti aderirono. Tuttavia, vi furono nutrite frotte di giocatori che decisero, al contrario, di *parteggiare per il virus*, costituendo vere e proprie squadre di *untori*, circolando in gruppo o come solitari portatori del morbo, intenzionati, in ogni caso, a espandere l'epidemia e approfittare del *bug*, per arricchirsi attraverso l'appropriazione dei beni e degli averi dei personaggi infettati.

Di conseguenza, la Blizzard fu costretta a dichiarare *forfait* e a resettasse completamente tutti i *server* del gioco coinvolti, riportando ogni cosa allo status quo.

All'indomani dell'incidente, svariati epidemiologi e centri di ricerca in tutto il mondo iniziarono a interessarsi della vicenda, ipotizzando che l'incidente del *corrupted blood*, sebbene svoltosi in un contesto videoludico con *target* giovanile, potesse risultare una delle più accurate e documentabili simulazioni elettroniche massive di un'epidemia avvenute fino a quel momento ([23]).

Sebbene si trattasse di un'epidemia "immateriale", avvenuta in un mondo immaginario e virtuale, essa aveva, in effetti, stravolto le attitudini di diversi milioni di soggetti, che avevano agito in maniera assolutamente non predicibile, dando luogo a reazioni e interpretazioni differenti, circa la sua funzione, da parte degli utenti.

È proprio partendo da questo assunto circa la propria natura immateriale che la presente ricognizione, in ultima analisi, vuole suggerire un'analogia tra il contagio *virale virtuale* (nell'accezione specifica della conformazione narrativa che esso assume nel mondo seriale e del MMORPG) e il meccanismo di propagazione della disinformazione alla base di determinate narrazioni mediatiche della crisi e dell'emergenza: si tratta, infatti, di un contagio che si crea e si propaga nel *medium* e per mezzo del *medium*, secondo dinamiche imprevedute e, finora, imprevedibili, e che perlopiù non tiene in debita considerazione l'entità delle ripercussioni etiche che esso genera nella realtà.

BIBLIOGRAFIA

[1] Aarseth, Espen J. 1997. *Cybertext: Perspectives on Ergodic Literature*. Baltimore: Johns Hopkins University Press.

[2] ———., e Sebastian Möring. 2020. «The game itself?: Towards a Hermeneutics of Computer Games».

[3] Balkin, Jack M., e Simone N. Beth. 2006. *The State of Play: Law, Games, and Virtual Worlds*. New York University Press.

[4] Bandirali, Luca, e Enrico Terrone. 2009. *Il sistema sceneggiatura*. Lindau.

- [5] Bogost, Ian. 2006. *Unit Operations: an Approach to Videogame Criticism*. MIT Press.
- [6] Bolter, J. David, e Richard Grusin. 1999. *Remediation: Understanding New Media*. MIT Press.
- [7] Bourdieu, Pierre. 1988. *La parola e il potere. L'economia degli scambi linguistici*. Guida.
- [8] Caligiuri, Mario. 2019. *Come i Pesci nell'Acqua. Immersi nella disinformazione*. Rubbettino Editore.
- [9] Campbell, Joseph. 1949. *The Hero of a Thousand Faces*. Bollingen Foundation Inc.
- [10] Cinelli, Mattia, Andraž Pelicon, Igor Mozetič, e et al. 2021. «Dynamics of online hate and misinformation». *Scientific Reports* 11.
- [11] Ciracì, Fabio. 2020. «Ecco che cosa distingue una pandemia dalla guerra». *Gazzetta del Mezzogiorno*, 30 maggio 2020.
- [12] ———. 2021. «Per una teoria critica del digitale: fake-news e postverità alla luce della logica della verosimiglianza». In *Filosofia digitale*, Riccardo Fedriga, Fabio Ciracì, Cristina Marras, 87–112. Mimesis.
- [13] Consalvo, Mia, e Nathan Dutton. 2006. «Game analysis: Developing a methodological toolkit for the qualitative study of games». *Game Studies* 6.
- [14] Dayan, Daniel, e Elihu Katz. 1993. *Le Grandi Cerimonie dei Media*. Baskerville.
- [15] Eco, Umberto. 1979. *Lector in fabula*. Bompiani.
- [16] Ferraris, Maurizio. 2017. *Postverità e altri enigmi*. Il Mulino.
- [17] Floridi, Luciano. 2014. *The fourth revolution. How the infosphere is reshaping human reality*. Oxford University Press.
- [18] Garrelts, Nate. 2005. *Digital Gameplay: Essays on the Nexus of Game and Gamer*. Jefferson McFarland.
- [19] Griffiths, Mark. 1999. «Violent video games and aggression: A review of the literature». *Aggression and Violent Behavior* 4 (2): 203–12.
- [20] Harrigan, Pat, e Fruin-Noah Wardrip. 2007. «Second Person: Role-Playing and Story». In *Games and Playable Media*. MIT Press.
- [21] Iversen, Sara Mosberg. 2014. «Play and Productivity: The Constitution of Ageing Adults in Research on Digital Games». *Games and Culture* 11 (1–2): 7–27.
- [22] Juul, Jesper. 2005. *Half-Real: Video Games between Real Rules and Fictional Worlds*. MIT Press.
- [23] Lofgren, Eric T., e Nina H. Fefferman. 2007. «The untapped potential of virtual game worlds to shed light on real world epidemics». *The Lancet Infectious Diseases* 7 (9): 625–29.
- [24] Longo, Mariano, Luca Benvenga, e Matteo Zaterini. 2020. «Le emergenze tra nuovi e vecchi media». In *Politica dell'emergenza*, Mariano Longo, Gianpasquale Preite, Emiliano Bevilacqua, Vincenzo Lorubbio, 375–89. Tangram.
- [25] Marconi, Diego. 2019. «Fake news, the crisis of deference, and epistemic democracy». *Post-Truth, Philosophy and Law*, 86–92.
- [26] Mead, Corey. 2013. *War Play: Video Games and the Future of Armed Conflict*. Houghton Mifflin Harcourt.
- [27] Quattrociocchi, Walter, e Antonella Vicini. 2016. *Misinformation: Guida alla società dell'informazione e della credulità*. FrancoAngeli.
- [28] Ricoeur, Paul. 1969. *Le conflit des interprétations. Essais d'herméneutique* 1. Le Seuil.
- [29] Ryan, Marie-Laure. 2001. *Narrative as Virtual Reality*. Johns Hopkins University Press.
- [30] Tung, Wan Qing. 2016. «Have you get shocked today?! How content farms generate affective publics in cyberspace». *Cultural Studies@Lingnan* 51.
- [31] Vattimo, Gianni. 2009. *Addio alla verità*. Meltemi.
- [32] Williams, Bruce A., e Michael X. Delli Carpini. 2011. *After Broadcast News. Media Regimes, Democracy and the New Information Environment*. Cambridge University Press.

Narrazione e interazione

Matteo Jacopo Zaterini

Università del Salento, matteojacopo.zaterini@unisalento.it

ABSTRACT

Il rapporto tra narrazione e interazione all'interno del medium videoludico si esprime attraverso un ampio ventaglio di posizioni da parte di studiosi e accademici. Ludologi e narrativisti assumono punti di vista estremamente diversi per studiare le opere videoludiche: se i primi hanno tradizionalmente concentrato l'attenzione sulla struttura e le relazioni tra gli elementi compositivi riservando poco interesse per la narrazione, i secondi hanno invece investito risorse nello studio della componente narrativa e di come quest'ultima si sviluppa all'interno dell'opera a prescindere dai vincoli tecnologici e strutturali che la caratterizzano. In seguito, attraverso l'analisi dell'opera *Her Story* intendo esplicitare i punti di contatto tra narrazione e interazione all'interno del medium videoludico partendo da un approccio integrato.

PAROLE CHIAVE

Ludo-narrativa, interazione, game design.

INTERVENTO

Alla fine degli anni Novanta Murray ([7]) ha utilizzato un approccio narrativista per studiare le potenzialità espressive del videogame design: secondo Murray, il futuro dei videogiochi come mezzo di comunicazione risiede nel loro potenziale di raccontare storie e nelle nuove risorse creative che possono dare ai narratori del ventunesimo secolo. Nel suo *Hamlet on the Holodeck* troviamo il primo significativo dibattito teorico nel campo degli studi sui videogiochi. Da un vero e proprio rifiuto del potenziale narrativo dei videogiochi, la ludologia è diventata un movimento volto a portare la peculiarità del videogioco, ovvero il game design, come mezzo espressivo in primo piano nella formulazione di nuove domande di ricerca. Con il tempo la posizione dei ludologi è diventata gradualmente più moderata per quanto riguarda la narratività del videogioco ([5]). Ryan ([8]) ha suggerito di sviluppare approcci ludo-narrativi in cui si possano superare le posizioni polarizzate e si possano affrontare così le complesse relazioni tra narrazione e game design, una questione cruciale per comprendere la costruzione e la trasmissione del significato nei videogiochi.

Ma il game design trasmette davvero un significato? Per rispondere a questa domanda, Bogost e Frasca ([2]) hanno iniziato a sviluppare una branca della ludologia che sarebbe diventata nota come "retorica procedurale". Questa teoria postula che il nucleo procedurale dell'interazione del gioco (le regole e la meccanica dell'interazione) è il fulcro principale del significato del videogioco e diversi ricercatori di videogiochi interessati a questo argomento hanno occasionalmente utilizzato idee della semiotica: Frasca ([4]) ha applicato le teorie di Peirce ed Eco, e Maietti ([6]) ha utilizzato principalmente la semiotica francese e italiana (Greimas, Eco, Fabbri).

Negli ultimi anni, diversi lavori hanno fatto progredire in modo significativo lo studio del videogioco come mezzo espressivo. In *Beyond Choices: The Design of Ethical Gameplay*, Sicart ([9]) mette in evidenza la connessione tra il nucleo procedurale del gioco e il suo strato di rappresentazione audiovisiva, così come il ruolo attivo dei giocatori nel processo di costruzione del significato. Il suo modello analitico è parzialmente ispirato all'ingegneria semiotica dell'interazione uomo-macchina di De Souza ([3]); tuttavia, nel suo modello, la semiotica finisce per essere essenzialmente separata dalla struttura procedurale del gioco ("livello procedurale" contro "livello semiotico", [9]).

Piuttosto che vedere le narrazioni come esperienze altamente strutturate create da un autore per essere consumate da un pubblico, la ludo-narrativa emergente, quindi, studia la struttura, intesa come tutti gli elementi di game design specifico, all'interno della quale i singoli giocatori possono costruire le proprie narrazioni, o gruppi di giocatori possono impegnarsi nella costruzione sociale condivisa di narrazioni. Nasce quindi uno spazio di analisi creato dallo scollamento della figura del giocatore e della propria rappresentazione, uno spazio all'interno del quale il game design, le aspettative del giocatore, l'intenzionalità narrativa, il contesto socioculturale di fruizione dell'opera, si intrecciano. In seguito, attraverso un'opera che fa di questo spazio il fulcro principale su cui fa leva la narrazione, proverò a renderne esplicite le caratteristiche essenziali che permettono al racconto di sfruttare le peculiarità specifiche del medium.

Her Story è un titolo indipendente sviluppato da Sam Barlow e pubblicato nel 2015. All'interno del gioco impersoniamo un detective a cui è stato affidato il compito di recuperare delle prove da un pc danneggiato che consistono in videoregistrazioni di cinque diversi interrogatori di una donna coinvolta in un crimine. Il gioco all'avvio trasformerà il

nostro desktop in quello del pc del detective incaricato del caso. Essendo il gioco ambientato nei primi anni '90 anche il nostro desktop avrà le ridotte funzionalità di un pc dell'epoca. Una volta identificato l'hard disk in cui sono contenuti i file degli interrogatori scopriremo che i 5 file video sono danneggiati e trasformati in centinaia di clip di pochi secondi a cui potremo accedere solo attraverso il motore di ricerca della polizia presente sul pc: digitando una parola o una serie di parole il motore di ricerca ci restituirà esclusivamente le prime cinque clip all'interno delle quali le parole che abbiamo cercato risultano pronunciate dalla donna. Lo scopo del gioco è quello di ricostruire gli eventi che la coinvolgono.

Her story poggia su tre pilastri di game design:

- L'opera non evolve: *Her Story* non ha un "prima" e un "dopo". È un'opera che si mantiene costante nella struttura, nelle modalità narrative, nei suoi elementi costitutivi. L'opera non cambia "avanzando cronologicamente", a differenza della quasi totalità delle altre opere appartenenti al medium. Non si avanza come nella lettura di un libro o nella visione di un film. L'opera infatti sacrifica la propria crescita per promuovere lo sviluppo di una narrazione nel giocatore.
- La narrazione poggia sul sottotesto: in *Her story* il testo è solo il punto di partenza. Non esiste un terreno già battuto, una strada da percorrere accompagnati dalle parole e dalle immagini fornite dall'autore dell'opera. Il testo "deve" essere superato. C'è qualcosa "oltre" il testo, qualcosa che non è riconoscibile, qualcosa che non dipende dall'autore.
- L'opera non ha una forma predeterminata: in *Her Story* è l'autore che chiede al giocatore di delineare la presenza narrativa all'interno dell'opera.

L'aspetto Interattivo/narrativo: scorgiamo i legami e le connessioni tra i vari elementi dell'opera liberandone la struttura interna, interagendo con l'opera attraverso le parole. Non è un atto di forza quello del giocatore: l'opera viene interrogata, e le varie richieste che le vengono somministrate sono frutto dell'attività psichica/emotiva dello spettatore. In un medium nel quale gli input principali sono storicamente stati associati a dei tasti da premere, delle direzioni da seguire, quello di interagire con le parole frutto di un nostro atto creativo cambia totalmente la prospettiva interattiva.

L'aspetto Decostruttivo: la narrazione all'interno di *Her story* è frammentata in senso letterale. Le clip in cui sono stati trasformati i file video durano pochi secondi e solitamente trattano di aspetti di vita estranei al motivo "criminale" che porta la donna ad essere interrogata dalla polizia. La narrazione è decostruita, ridotta a brandelli che non hanno vita propria, affidata alla mente e di chi interrogando il database prova a dargli un senso compiuto.

L'aspetto Costruttivo: la natura frammentaria della narrazione ci costringe ad andare alla ricerca di un senso a cui le singole clip rimandano. Ma il senso non è presente all'interno dell'opera, non lo si può "raggiungere" in maniera tradizionale, per esempio "avanzando" nell'opera. L'oggetto che emerge dall'accostamento di questi segni è un oggetto "costruito" dal giocatore, dall'interpretante. È un oggetto unico, in quanto il percorso che ha portato alla sua formazione, che ne ha creato il perimetro è proprio di ognuno. L'opera diventa fruibile da ciascuno in maniera diversa, ognuno accosta i vari frammenti seguendo un proprio percorso, costruendo la narrazione a partire dai frammenti che ha individuato e da quelli che non ha individuato. La narrazione di *Her Story*, quindi, non prescinde dall'interattività del medium, ma si dà al fruitore proprio attraverso le infinite possibilità di interazione scaturite dalle modalità di interrogare l'opera attraverso il proprio linguaggio naturale. Quello di *Her story* è solo un esempio di come narrazione e game design interagiscono, permettendo all'autore di creare nuove tipologie di interazione tra utente e opera.

BIBLIOGRAFIA

- [1] Adams, Ernest. 1999. «Three problems for interactive storytellers». *Designer's Notebook Column - Gamasutra* 144.
- [2] Bogost, Ian, e Frasca Gonzalo. 2007. «Videogames Go to Washington: The Story Behind The Howard Dean for Iowa Game». In *Second Person. Role Playing and Story*, Pat Harrigan and Noah Wardrip-Fruin, 233–46. Games and Playable Media. Cambridge/MA.
- [3] De Souza, Clarisse Sieckenius, e Jenny Preece. 2004. «A framework for analyzing and understanding online communities». *Interacting with computers* 16.3: 579–610.
- [4] Frasca, Gonzalo. 2001. «Videogames of the oppressed: Videogames as a means for critical thinking and debate». School of Literature, communication, and culture. Georgia Institute of Technology.
- [5] Juul, Jesper. 2005. *Half-Real: Video Games between Real Rules and Fictional Worlds*. MIT Press.
- [6] Maietti, Massimo. 2004. *Semiotica dei videogiochi*. Edizioni Unicopli.
- [7] Murray, Janet Horowitz. 2017. *Hamlet on the holodeck: The future of narrative in cyberspace*. MIT press.
- [8] Ryan, Richard M., Przybylski Andrew, e Rigby C. Scott. 2006. «The motivational pull of video games: A self-determination theory approach». *Motivation and emotion* 30 (4): 344–60.

[9] Sicart, Miguel. 2013. *Beyond choices: The design of ethical gameplay*. MIT Press.

Sessione Testi 2

Ada Lovelace

Web e social media come nuove fonti per la storia

Chiara Aldini¹, Stefano Allegrezza², Tommaso Mazzoli³

¹Università degli Studi di Bologna, Italia, chiara.aldini4@unibo.it

²Università degli Studi di Bologna, Italia, stefano.allegrezza@unibo.it

³Università degli Studi di Udine, Italia, Tommaso.mazzoli@uniud.it

ABSTRACT

Il contributo intende mettere in evidenza come negli ultimi anni l'interesse verso i temi dell'archiviazione e conservazione del web e dei social media sia cresciuto enormemente, anche perché la consapevolezza dell'importanza di tali "risorse" come fonti privilegiate per ricostruire la storia della nostra epoca è ormai acquisita. Come faranno gli storici del futuro a ricostruire il periodo storico che stiamo vivendo se archivisti e bibliotecari non saranno capaci di archiviare e preservare i siti web e social media di istituzioni, enti pubblici, partiti, associazioni, organi di governo, personaggi politici, personaggi illustri in genere, dato che ormai tutto viene veicolato attraverso tali canali? La fragilità del web, poi, imporrebbe di agire subito ed avviare senza indugio iniziative di "web and social media archiving", pena la scomparsa di quanto è stato pubblicato e reso disponibile sul web negli ultimi anni, ma su questo punto la situazione in Italia – salvo poche eccezioni – appare molto in ritardo rispetto agli altri paesi europei ed enormemente in ritardo rispetto ai paesi dell'area anglosassone. Occorre avviare iniziative di sensibilizzazione su questi temi e di formazione delle competenze e delle professionalità necessarie per condurre progetti di archiviazione e conservazione del web e dei social media.

PAROLE CHIAVE

Web archiving; social media archiving; digital preservation; fonti archivistiche; storia.

INTERVENTO

1. INTRODUZIONE

L'interesse verso i temi della conservazione del web e dei social media è cresciuto notevolmente soprattutto negli ultimi anni, ovvero da quando è emersa sempre più distintamente la consapevolezza che essi costituiranno fonti insostituibili per la comprensione della storia e della civiltà contemporanee. Si pensi, ad esempio, alla pandemia da COVID-19 ancora in corso: le fonti web saranno fondamentali per ricostruire e raccontare gli avvenimenti di questo periodo ([14]) e senza di esse sarà molto difficile ricostruire questi due anni solo a partire dalle fonti tradizionali. È per questo motivo che negli ultimi vent'anni si è molto sviluppato il web archiving, cioè il processo finalizzato alla 'cattura' e conservazione sistematica di porzioni del web a cura di istituzioni della memoria, come archivi e biblioteche nazionali, istituzioni universitarie, fondazioni ([2]). Molti istituti di conservazione si sono attivati con iniziative e progetti di web archiving. Recentemente, si è cominciato anche a sviluppare un ulteriore settore, quello dell'archiviazione e conservazione dei social media, ritenuti anch'essi – non a torto – fonti insostituibili per ricostruire il periodo storico che stiamo vivendo.

2. STATO DELL'ARTE

Le prime riflessioni sul web archiving risalgono alla fine degli anni Novanta, periodo in cui inizia a porsi concretamente il problema della conservazione dei siti web. Nel 1996, sei anni dopo lo sviluppo del World Wide Web ad opera di Tim Berners Lee, prende avvio Internet Archive ([9]), organizzazione senza scopo di lucro con la *mission* di creare una digital library di siti internet e così salvaguardarne e garantirne l'accesso permanente. I fondatori, gli ingegneri statunitensi Brewster Kahle e Bruce Gilliat, 'catturarono' le prime istantanee di pagine web mediante l'utilizzo di un 'crawler', un'applicazione che riusciva a catturare una fotografia fedele e inalterata dei siti presenti sul web. Oggi Internet Archive vanta oltre 25 anni di cronologia web, per un totale di circa 70 petabyte di dati raccolti, resi accessibili tramite la Wayback Machine ([1]), ovvero l'interfaccia pubblica che consente di ricercare e visualizzare le versioni archiviate dei siti web. Inserendo l'URL (Uniform Resource Locator) che si vuole analizzare e selezionando un intervallo di date, è possibile navigare sulla versione archiviata di una pagina o un sito web.

Internet Archive mette a disposizione anche Archive-It, un servizio in abbonamento disponibile dal 2006 che consente alle istituzioni di costruire e preservare raccolte di contenuti nativi digitali. Attraverso l'applicazione web di facile utilizzo, i partner di Archive-It possono raccogliere, metadattare, gestire e generare una copia delle proprie raccolte digitali, le quali vengono archiviate ed ospitate nel data center di Internet Archive e rese accessibili al pubblico con ricerca full-text.

L'importanza di preservare le informazioni digitali è stata riconosciuta nel 2003 anche dall'Organizzazione delle Nazioni Unite per l'educazione, la scienza e la cultura (UNESCO), che nel "Charter on the Preservation of Digital Heritage" inserisce tra i materiali digitali che costituiscono il "digital heritage" anche le pagine web:

«The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are "born digital", there is no other format but the digital object. Digital materials include texts, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained. Many of these resources have lasting value and significance, and therefore constitute a heritage that should be protected and preserved for current and future generations. This ever-growing heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression» ([16]).

Nel 2003 viene fondato anche l'International Internet Preservation Consortium (IIPC), che riunisce alcune delle principali biblioteche nazionali e svolge attività di promozione e sviluppo di strumenti, tecniche e standard comuni per la creazione di archivi web internazionali. Attualmente partecipano all'IIPC organizzazioni di oltre 35 paesi, tra cui biblioteche e archivi nazionali, universitari e regionali ([8]).

Sebbene dal 2018 non sia più attiva, occorre citare anche la Internet Memory Foundation (fino al 2010 European Archive Foundation), fondata nel 2004 e coinvolta in progetti di ricerca, finanziati dalla Commissione Europea, volti a migliorare le tecnologie di web crawling, estrazione dati, text mining e conservazione degli archivi web delle istituzioni europee.

Oltre ad iniziative e strumenti per raccogliere e tenere traccia delle risorse sul web, all'interno della comunità internazionale è nata anche l'esigenza di uno specifico formato contenitore che consentisse di archiviare più risorse web in un unico file. Un grosso passo avanti in questa direzione è stata la pubblicazione, nel 2009, dello standard ISO 28500 (versione corrente aggiornata ISO 28500:2017) che ha definito il formato WARC (Web ARChive) – una revisione del formato ARC File Format usato inizialmente da Internet Archive per archiviare le catture del web – che oggi rappresenta il formato standard per gli archivi web ([10]), insieme al formato WACZ (Web Archive Collection Zipped) recentemente proposto ([5]). Nel 2013 viene pubblicato anche lo standard ISO/TR 14873:2013 che definisce principi, metodi e standard di qualità per le istituzioni culturali che si occupano di web archiving ([11]).

Dagli anni Novanta sono numerose le iniziative di archiviazione del web promosse in ambito internazionale. PANDORA, avviato nel 1996 ad opera della National Libraries of Australia, è stato il primo progetto di web archiving sviluppato da un'istituzione pubblica.

Molto attiva l'area del Nord Europa, con i progetti di Svezia (Kulturarw3, 1996), Norvegia (2001), Islanda (2004), Danimarca (Netarkivet, 2005). Il progetto americano della Library of Congress prende avvio nel 2000, quello della Bibliothèque Nationale de France nel 2006. Nel Regno Unito si occupano di web archiving sia i National Archives che le biblioteche incaricate del deposito legale, costituendo così un modello di riferimento internazionale. L'elenco più esaustivo ed aggiornato relativo alle iniziative internazionali di archiviazione del web è stato redatto sulla base dell'indagine condotta dal team di Arquivo.pt, l'archivio web del Portogallo. I risultati, che sono resi disponibili su Wikipedia nella pagina "List of web archiving initiatives" ([13]) evidenziano non solo i progetti ed i relativi paesi di appartenenza, ma anche le tecnologie di web archiving utilizzate ed il personale dedicato, distinguendo anche tra incarichi full-time o part-time.

Nel 2013 la Digital Preservation Coalition ha pubblicato un report specifico sul web archiving nella serie delle DPC Technology Watch Publications ([18]). Il rapporto discute le questioni chiave affrontate dalle organizzazioni impegnate in iniziative di archiviazione del web ed offre una panoramica dei principali software e strumenti attualmente disponibili.

In questo quadro di grandi sforzi a livello internazionale nel tentativo di trovare le strategie per preservare una risorsa che è di per sé estremamente effimera, l'Italia si distingue per l'enorme ritardo rispetto agli altri paesi, anche europei. L'unica iniziativa a livello nazionale meritevole di menzione è stata avviata nel 2018 dalla Biblioteca Nazionale Centrale di Firenze con il progetto di raccolta e archiviazione di siti web di 'interesse culturale' per la storia e la cultura italiana, secondo i principi della legge nazionale sul deposito legale (L. 106/2004 e suo Regolamento attuativo D.P.R. 252/2006) ([15]). La Biblioteca si fa carico, oltre alla raccolta, anche dell'organizzazione e della metadatozione "manuale" dei siti archiviati, avvalendosi per l'accesso e la conservazione della piattaforma Archive-it di Internet Archive. Il deposito legale dei documenti diffusi tramite rete informatica, tuttavia, non è ancora obbligatorio perché lo stesso D.P.R. 252/2006 all'art. 37 prevedeva che il deposito di tali documenti fosse subordinato alla redazione di uno specifico regolamento tecnico che non è ancora stato emanato. Pertanto, l'adesione al programma è su base volontaria da parte dei gestori dei siti, i quali possono

manifestare il proprio interesse compilando il form online disponibile sul sito della Biblioteca. La raccolta si presenta quindi a tutt'oggi molto parziale e frammentata rispetto alla produzione di interesse culturale presente sul web italiano. La scarsa sensibilità sul tema del web archiving e l'assenza di uno specifico quadro normativo rendono difficoltosa l'attuazione di strategie nazionali condivise, laddove un'azione sinergica sarebbe necessaria vista la rapida evoluzione del web e l'enorme quantità di risorse culturali che vi trovano sede e che rischiano di andare irrimediabilmente perdute, stante la rapida evoluzione del web e la sua 'fragilità' ([12]).

3. APPLICAZIONI E STRUMENTI

Per quanto auspicabile, l'individuazione di una strategia condivisa a livello nazionale riguardo la conservazione e la salvaguardia – anche parziale – dei contenuti del web è un traguardo ancora molto lontano da raggiungere.

Sembra quindi quanto mai opportuno che le singole istituzioni, gli enti di ricerca e forse anche i privati cittadini provino a dotarsi di sistemi autonomi di web archiving che possano sopperire a questa mancanza. A questo proposito occorre rilevare che il panorama delle applicazioni software che consentono di archiviare un sito internet è ancora molto frammentato. Vi sono soluzioni pensate per scaricare e consultare siti web con un limitato numero di pagine che possono essere utilizzate da singoli utenti o da gruppi di lavoro ristretti; altre che permettono processi di *harvesting* molto più strutturati e complessi ma che richiedono competenze e attrezzature informatiche decisamente più articolate.

Nel nostro caso, senza avere la pretesa di analizzare tutte le soluzioni presenti sul mercato – che sono numerose e in costante aumento – ci è parso importante segnalare due strumenti software particolarmente interessanti perché, dall'analisi condotta, si sono rivelati particolarmente efficaci pur rimando semplici da utilizzare, e, quindi, costituiscono un buon punto di partenza per un progetto di salvaguardia dei siti web e possono essere utilizzati anche da coloro che non hanno particolari competenze informatiche,

Il software più conosciuto è sicuramente HTTrack, un'applicazione *open source* per il mirroring di siti web e la loro navigazione offline ([7]). Il programma principale si esegue solo con istruzioni testuali dalla riga di comando dei principali sistemi operativi come Linux, Windows e Mac. Esiste però la possibilità di interagire con il software anche grazie ad un'interfaccia grafica predisposta per Windows (WinHTTrack) e per Linux (WebHTTrack) che ne rende sicuramente più semplice l'utilizzo. Questo programma permette di scaricare un intero sito internet sul proprio computer in una cartella a nostra scelta ricostruendone l'intera struttura. Vengono quindi memorizzati il codice, le immagini ed ogni altro tipo di file dal server al computer locale. È importante sottolineare che il software adattata la struttura originale dei link relativi al sito in modo da poterli comodamente navigare in off-line. Vengono supportati i protocolli HTTP e FTP ma non HTTPS. Dopo che il programma avrà eseguito la copia dall'indirizzo web selezionato sarà sufficiente aprire una qualsiasi pagina del sito nel proprio browser per poter navigare esattamente come se si fosse on-line. Il software permette inoltre di configurare numerose opzioni per limitare o estendere la raccolta di base e per controllare il tipo e le caratteristiche dei file da scaricare sul proprio computer. Per ora non è prevista la possibilità di salvare il sito in formato WARC anche se in rete esiste un tool di conversione, *httrack2warc* ([6]), dai risultati ancora non del tutto affidabili. HTTrack è scaricabile dal sito web ufficiale insieme ad una guida dettagliata al tipo di installazione che si intende eseguire e un manuale d'uso per configurare il lavoro di salvataggio del sito web. Purtroppo l'ultima versione del programma è stata rilasciata più di quattro anni fa e ciò fa presupporre che il progetto sia stato abbandonato e quindi non più supportato da nuovi aggiornamenti nel prossimo futuro. Il secondo software che vogliamo segnalare è Cyotek WebCopy ([4]) uno strumento gratuito ma non *open source* che consente di scaricare automaticamente il contenuto di un sito web sul proprio dispositivo locale. Come HTTrack anche WebCopy esegue la scansione del sito web specificato e ne scarica il contenuto: i collegamenti a risorse come fogli di stile, immagini e altre pagine del sito verranno automaticamente rimappati in modo da corrispondere al proprio percorso locale. Utilizzando il pannello di configurazione si potranno definire quali parti di un sito web verranno copiate, permettendo, ad esempio, di scaricare solo le immagini piuttosto che l'intero contenuto. L'interfaccia d'uso è molto completa e sicuramente più facile da utilizzare rispetto HTTrack. Tra le opzioni principali è presente il comando che consente di effettuare una scansione completa del sito prima di iniziare a scaricarlo. Ciò permette anche la creazione della mappa del sito stesso utile ad individuarne la struttura per selezionare zone di particolare interesse. WebCopy riesce anche a scaricare il contenuto di aree protette da password avendo cura di inserire le credenziali richieste in fase di avvio della raccolta. Anche in questo caso non abbiamo la possibilità di salvare il lavoro di copia nel formato WARC. Il software è disponibile sul sito del produttore ma esiste solo la versione per il sistema operativo Windows (dalla versione 7 in poi). Sullo stesso sito è inoltre presente una corposa documentazione di supporto sia per l'installazione che per il suo utilizzo con diversi esempi pratici. L'ultima versione stabile è stata rilasciata alla fine di marzo 2021 ma la presenza di altre versioni in fase di test fa pensare che la casa produttrice sia intenzionata a seguire e migliorare il suo prodotto anche in futuro.

I software qui brevemente presentati sono solo un esempio che dimostra come le pratiche di web archiving possano essere implementate anche all'interno di progetti d'archiviazione di portata limitata e, volendo, anche di tipo personale. Tuttavia, per progetti di più ampio respiro, come quelli portati avanti dalle biblioteche nazionali di alcuni Stati europei e dell'area anglosassone, questi strumenti non sono più sufficienti ed è necessario mettere in campo risorse – non solo tecniche ed economiche, ma anche umane – di dimensioni molto più ampie.

4. CONCLUSIONI

Se è vero che il tema dell'archiviazione e conservazione del web e dei social media sta acquisendo una rilevanza sempre maggiore, è altrettanto vero che ad oggi le figure professionali capaci di condurre progetti in questo ambito sono poche se non addirittura quasi del tutto assenti ([3]), salvo casi eccezionali e certamente meritevoli di segnalazione – come quello già citato della Biblioteca Nazionale Centrale di Firenze. Ciò è dovuto, da una parte, allo scarso interesse che fino ad oggi è stato riservato a questi temi e, dall'altra, alla mancanza di percorsi formativi che sarebbero invece estremamente importanti anche in considerazione delle difficoltà – non solo di tipo tecnico ma anche economico ed organizzativo – che devono essere superate per portare a termine progetti di questa natura. Per superare queste difficoltà, dal 6 al 10 settembre 2021 presso l'Università degli Studi di Bologna si è tenuta la prima edizione della Summer school in “Web and social media archiving and preservation”, con l'intento di:

«offrire una formazione di alto livello sui temi emergenti dell'archiviazione e conservazione dei siti web e dei social media, che rappresentano una nuova e diversificata tipologia di materiale la cui conservazione è imprescindibile per tutta una serie di ambiti scientifici (si pensi alla ricerca storica, sociologica, antropologica, etc.) ai fini della futura ricostruzione dell'attuale civiltà. La Summer school intende anche fornire le conoscenze e le competenze necessarie per favorire lo sviluppo di nuove professionalità ed avviare nuovi percorsi lavorativi da parte dei discenti interessati» ([17]).

La prima edizione della Summer school ha visto la partecipazione di quasi 40 partecipanti tra professionisti dei beni culturali (archivisti, bibliotecari, operatori museali), informatici, funzionari di enti pubblici ed aziende private, studenti/dottorandi in Library and information science e in Digital humanities oltre che persone interessate a vario titolo alle questioni legate all'archiviazione e alla conservazione dei siti web, dei blog e dei social media. Con questa iniziativa si è voluto, da una parte, avviare un percorso di sensibilizzazione sull'importanza della conservazione del web e dei social media, dall'altra, formare i professionisti che saranno in grado di gestire i processi di archiviazione e conservazione almeno di quella parte del web e dei social media che sarà insostituibile per la ricostruzione della nostra epoca e senza la quale la storia ne risulterebbe irrimediabilmente menomata. Si auspica che iniziative di questo genere vengano messe in campo anche da altre agenzie formative e dalle istituzioni di conservazione della memoria, al fine di garantire la conservazione per il futuro di queste nuove ed insostituibili fonti per la storia.

BIBLIOGRAFIA

- [1] «Archive-It». s.d. Archive-It. <https://archive-it.org>.
- [2] Bracciotti, Lorenzana. 2019. «Il Web Archiving. Conservazione e uso di una nuova fonte». *Officina della storia* (blog). 10 gennaio 2019. https://www.officinadellastoria.eu/it/2019/01/10/il-web-archiving-conservazione-e-uso-di-una-nuova-fonte/#_edn2.
- [3] Costa, Miguel, Daniel Gomes, e Silva. 2017. «The evolution of web archiving». *International Journal on Digital Libraries* 18 (settembre): 191–205.
- [4] «Cyotek WebCopy». s.d. <https://www.cyotek.com/cyotek-webcopy>.
- [5] «GitHub». s.d. GitHub. <https://github.com/webrecorder/wacz-spec>.
- [6] «GitHub - nla/htrack2warc». s.d. GitHub. <https://github.com/nla/htrack2warc>.
- [7] «HTTrack». s.d. HTTrack. <https://www.httrack.com>.
- [8] «International Internet Preservation Consortium (IIPC)». s.d. Netpreserve. <https://netpreserve.org/>.
- [9] «Internet Archive». s.d. <https://archive.org>.
- [10] «ISO 28500:2017 Information and documentation — WARC file format». 2017. ISO. <https://www.iso.org/obp/ui/#iso:std:iso:28500:ed-2:v1:en>.
- [11] «ISO/TR 14873:2013 Information and documentation — Statistics and quality issues for web archiving». 2013. ISO. <https://www.iso.org/obp/ui/#iso:std:iso:tr:14873:ed-1:v1:en>.
- [12] Landino, Costantino, e Lina Marzotti. «Perché dovremmo pensare al web archiving». *Forum PA - Cantieri PA* (blog). 20 marzo 2019. <https://www.forumpa.it/pa-digitale/gestione-documentale/perche-dovremmo-pensare-al-web-archiving>.
- [13] «List of Web archiving initiatives». s.d. Wikipedia. https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives#Archived_data.

- [14] Lorenzana, Bracciotti. 2020. «Pandemia e web archiving. Conservare le fonti online #igiornidellapandemia». *Il mondo degli archivi* (blog). 2 maggio 2020. <http://www.ilmondodegliarchivi.org/rubriche/archivi-digitali/815-pandemia-e-web-archiving-conservare-le-fonti-online-igiornidellapandemia>.
- [15] Storti, Chiara. 2019. «Web archiving, “sfida culturale”: il servizio della Biblioteca Nazionale Centrale di Firenze». *Forum PA - Cantieri PA* (blog). 12 giugno 2019.
- [16] Unesco. 2003. «Charter on the Preservation of Digital Heritage». *Unesco* (blog). 15 ottobre 2003. http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html.
- [17] «Web and social media archiving and preservation». s.d. <https://site.unibo.it/web-and-social-media-archiving-and-preservation/it>
- [18] «Web-Archiving: DPC Technology Watch Report». s.d. DPC online. <https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports>

Idee, persone, *realia*: un ambiente digitale per la Via della Seta

Emmanuela Carbé¹, Andrea Balbo², Chiara Ombretta Tommasi³, Francesco Stella⁴, Mario G.C.A. Cimino⁵,
Federico A. Galatolo⁶, Chiara Aiola⁷

¹Università di Siena, Italia – emmanuela.carbe@unisi.it

²Università di Torino, Italia – andrea.balbo@unito.it

³Università di Pisa, Italia – chiara.tommasi@unipi.it

⁴Università di Siena, Italia – francesco.stella@unisi.it

⁵Università di Pisa, Italia – mario.cimino@unipi.it

⁶Università di Pisa, Italia – federico.galatolo@ing.unipi.it

⁷Net7, Italia – aiola@netseven.it

ABSTRACT

Il contributo delinea la fase di progettazione di un ambiente digitale collaborativo multilingua dedicato a SERICA (Sino-European Religious Intersections in Central Asia. Interactive Texts and Intelligent Networks), un progetto che riunisce studiosi di diverse discipline per la realizzazione di una piattaforma integrata che raccolga documentazione testuale sulle connessioni tra mondo europeo e mondo orientale in un percorso diacronico molto ampio (dal II sec. a.C. all'età moderna). A partire da esperienze precedenti, e in particolare dal progetto ELA - Eurasian Latin Archive, è stata prodotta un'analisi dei requisiti e un'architettura della piattaforma per la realizzazione di un primo prototipo.

PAROLE CHIAVE

Digital Library, Multilingual Texts, NLP, Artificial Intelligence.

INTERVENTO

1. INTRODUZIONE

SERICA (Sino-European Religious Intersections in Central Asia. Interactive Texts and Intelligent Networks) è un progetto FISIR a cura dell'Università di Pisa e di Torino, in collaborazione con studiosi dell'Università di Ravenna, di Napoli Orientale e di Siena, e con alcune aziende, tra cui Net7. Il progetto, avviato a giugno 2021, si concentra sulle rotte dell'Asia Centrale tra Cina e Europa, con un'estensione diacronica dei documenti presi in esame che parte dal mondo ellenistico e tardoantico e arriva al XIX secolo, adottando dunque una prospettiva di ricerca globale e transdisciplinare in grado di mettere in luce le interconnessioni tra epoche diverse. Per questo ampio attraversamento sono stati individuati tre assi portanti, sintetizzabili nei concetti di 1. idee (scambi tra Oriente e Occidente in ambito filosofico, religioso, antropologico), 2. persone (storiografi, esploratori, missionari che contribuirono alla mediazione tra i due mondi), e 3. *realia* (documentazione su elementi materiali delle civiltà osservate, opere d'arte e oggetti d'uso quotidiano, strumenti scientifici e capi di abbigliamento).

L'obiettivo finale del progetto SERICA è la creazione di un ambiente digitale collaborativo e multilingua, basato anche su tecniche di NLP e di Intelligenza Artificiale, in grado di offrire agli utenti una piattaforma di Digital Library con strumenti di elaborazione, analisi avanzata e annotazione degli oggetti di ricerca, affiancati da percorsi di approfondimento e da strumenti specifici come mappe interattive.

2. BACKGROUND

Il gruppo di SERICA è partito da diversi case studies, prima di tutto dall'esperienza pregressa di ELA – Eurasian Latin Archive (ela.unisi.it, Stella 2021), un progetto biennale dell'Università di Siena cofinanziato da Regione Toscana tra il 2018 e il 2020, e realizzato con la collaborazione dell'azienda QuestIT. Il progetto ELA ha censito un gruppo di testi del XIII-XVIII secolo, digitalizzando un nucleo di documenti in lingua latina, contenenti tuttavia anche porzioni di testo in altre lingue ([6]). ELA ha realizzato un modello di codifica in XML TEI ponendo una particolare attenzione ad alcuni aspetti semantici dei testi, a partire dai nomi di luogo e nomi di persona, trattati da uno script (realizzato in Python) che in parte automatizza il processo di marcatura utilizzando le occorrenze già individuate in precedenza. La Digital Library di ELA dialoga con un framework messo a disposizione su GitHub (ELA Tools: [4]) e basato su CLTK ([3]) e NLTK ([2]): i testi codificati in TEI vengono dunque acquisiti da ELA Tool, in cui si attivano i processi di parsing, normalizzazione, tokenizzazione, individuazione delle parti del discorso, lemmatizzazione, estrazione di collocazioni, n-grammi, e di

coordinate geografiche dei luoghi citati nel testo attraverso i database di Pleiades (pleiades.stoa.org) e GeoNames (geonames.org). I dati così raccolti, insieme ad alcune elaborazioni statistiche, vengono restituiti in formato JSON alla Digital Library, che integra i risultati acquisiti nell'interfaccia utente ([5]).

Oltre a Eurasian Latin Archive, il progetto SERICA è partito da altri strumenti già esistenti, come la biblioteca digitale Sinica 2.0 dell'Università di Vienna, l'archivio ARSI di Roma (Archivum Romanum Societatis Iesu, in parte digitalizzato e disponibile in Internet Archive), il Chinese Christian Texts Database dell'Università di Lovanio e il progetto China Historical GIS (CHGIS) dell'Università di Harvard.

3. ANALISI DEI REQUISITI E REALIZZAZIONE DEL PROTOTIPO

Per il progetto di ambiente digitale è stato costituito un gruppo di lavoro specifico che ha collaborato alla stesura dell'analisi dei requisiti e alla realizzazione dei modelli di dati e dell'architettura della piattaforma. L'ambizione del progetto è quella di creare un ambiente modulare, in grado di unire armonicamente tecnologie, framework e modelli (IIIF per la gestione e l'annotazione delle immagini, codifica del testo in XML TEI, annotazione semantica in RDF, trattamento del linguaggio tramite strumenti NLP e di Intelligenza Artificiale) per rispondere alle esigenze dei gruppi di ricerca e alle maggiori criticità intrinseche al progetto: la prima di queste è l'estrema eterogeneità del *corpus*, costituito da lingue diverse di epoche diverse (testi in greco antico, latino, cinese, lingue persiane, sanscrito, italiano, francese, spagnolo, tedesco), lingue che dovranno essere riconosciute e processate nell'ambito degli strumenti NLP messi a disposizione dalla Digital Library.

Sono stati definiti, all'interno dei requisiti, i ruoli degli utenti (amministratore, editore, autore-revisore, lettore, visitatore) ed è stata data una particolare attenzione alla natura dei metadati da gestire.

Allo scopo di creare uno strumento flessibile e scalabile, il cuore dell'ambiente digitale è stato suddiviso in quattro moduli: 1. SERICA DL (Digital Library e relativo gestionale), che prevede più livelli di rappresentazione degli oggetti digitali: nel caso di documenti testuali la Digital Library include le immagini digitalizzate (se disponibili), il testo codificato, le possibili annotazioni semantiche, l'esito delle elaborazioni NLP; 2. SERICA MAP (Atlante per la navigazione interattiva) gestisce la realizzazione delle mappe, distinte per epoca e per tema e collegate a pagine informative, immagini, riferimenti ai testi della Digital Library, riferimenti ad altri progetti e basi di dati; 3. SERICA NLP (strumenti di analisi linguistica e semantica basati su tecniche di Natural Language Processing) è il modulo che attiva il livello NLP della Digital Library, estrae e gestisce i dati linguistici e i concetti chiave dei testi, con particolare attenzione alla risoluzione dei problemi legati alle diverse lingue presenti nei testi; 4. SERICA AI (strumenti di analisi e di guida alla navigazione basati sull'Intelligenza Artificiale) è il modulo per realizzare la ricerca di testi per similarità tramite algoritmi di apprendimento, e per creare una guida interattiva visuale in base ad algoritmi di raccomandazione.

I moduli interagiranno tra loro attraverso tecniche e formati standard: il singolo item, costituito da un documento XML in una codifica comune e documentata, dotato di identificatore univoco, sarà il nucleo su cui basare la comunicazione stessa – eventualmente anche da e verso agenti esterni – e la persistenza dei dati. Questa scelta può consentire da una parte il riutilizzo di eventuali dati già esistenti attraverso semplici trasformazioni automatiche (ad esempio parte dei documenti pubblicati in ELA), dall'altra la condivisione in evoluzioni o aggregazioni future, nella convinzione che l'obiettivo primario di ogni progetto digitale, ancora prima della costruzione di uno strumento gestionale, sia la preservazione e l'evoluzione del dato, nonché la sua documentazione.

In base all'analisi dei requisiti è in corso di realizzazione un prototipo operativo della piattaforma, realizzato da un gruppo di sviluppatori dell'azienda Net7, concentrato primariamente sul gestionale, sul quale i ricercatori sperimenteranno il processo di creazione di una risorsa all'interno della Digital Library. I quattro moduli saranno dunque accessibili attraverso un unico portale, che presenterà diversi livelli di operatività a seconda della tipologia di utente, mantenendo lo stesso linguaggio grafico e la stessa interfaccia.

4. RINGRAZIAMENTI

Il progetto è finanziato dal Ministero dell'Università e della Ricerca (MUR), nel quadro del programma FISR 2019 - Grant No. 03602.

BIBLIOGRAFIA

- [1] Balbo, Andrea, e Ahn Jaewon. 2019. *Confucius and Cicero: Old Ideas for a New World, New Ideas for an Old World*. Vol. 1. Roma: Sinica.
- [2] Bird, Steven, Klein Ewan, e Edward Loper. 2009. *Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit*. O'Reilly Media. www.nltk.org/book/.
- [3] Burns, Patrick J. 2019. «Building a Text Analysis Pipeline for Classical Languages». In *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*, 159–76. Boston: Berti, M., and De Grutyer, B.

- [4] Carbé, Emmanuela, e Francesco Garosi. 2019. «Ela Tools». GitHub. 2019. <https://github.com/EurasianLatinArchive/ELA>.
- [5] Carbè, Emmanuela, e Nicola Giannelli. 2020. «ELA: fasi del progetto, bilanci e prospettive». In *La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*, 61–66.
- [6] Carbé, Emmanuela, e Nicola Giannelli. 2021. «First steps in building the Eurasian Latin Archive». *QQLM Journal* 9: 577–89.
- [7] Stella, Francesco. 2020. «“Global Latin”: Testi latini d’america e d’Asia fra Medioevo e prima età moderna». In 3–298. Siena: SISMELE- Edizioni del Galluzzo.
- [8] ———. 2020. «The Eurasian Latin Archive and the ‘Long’ Latin Middle Ages». In *Digital Philology and Quantitative Criticism of Medieval Literature: Unconventional Approaches to Medieval Latin Literature*, II:219–35. Brepols Publishers.

Visualizzazione del cambiamento d'uso del maschile e femminile nei titoli occupazionali

Pierluigi Cassotti¹, Pierpaolo Basile¹, Marco de Gemmis¹, and Giovanni Semeraro¹

¹ Università di Bari, Italia – {nome}.{cognome}@uniba.it

ABSTRACT

In questo lavoro presentiamo uno strumento per la visualizzazione di statistiche riguardanti l'uso delle forme grammaticali maschile e femminile di titoli occupazionali e dei professionisti menzionati contestualmente ai titoli occupazionali in un corpus diacronico. Le statistiche sono state calcolate utilizzando un corpus diacronico di articoli estratti da quotidiani italiani, composto da 3.5 miliardi di tokens. Le occorrenze dei titoli occupazionali sono state filtrate per ridurre rumore introdotto dalla polisemia dei termini. L'interfaccia web permette un uso semplice ed intuitivo grazie all'utilizzo di tecnologie allo stato dell'arte per la visualizzazione di grafici. L'interfaccia offre la possibilità di visualizzare, confrontare e analizzare le serie temporali delle frequenze relative dei titoli occupazionali nella forma maschile e femminile e delle frequenze assolute delle occorrenze dei professionisti menzionati nel testo.

PAROLE CHIAVE

Diachronic, gender, occupational title, natural language processing.

INTERVENTO

1. INTRODUZIONE

La crescente disponibilità di corpus diacronici digitalizzati rende sempre più evidente la necessità di approcci computazionali che tengano conto della dimensione temporale nell'analisi dei documenti.

Sfruttando tecniche di Natural Language Processing è possibile condurre automaticamente un'analisi diacronica di ampie collezioni di documenti per investigare cambiamenti nella morfologia, nella grammatica e nella semantica del linguaggio. Un esempio di questo tipo di analisi è quella svolta in ([5]), che sfrutta Google Ngram ([4]), una collezione di n-grammi estratti da 3.5 milioni di libri pubblicati tra il 1520 e il 2008. Gli autori di questo lavoro sfruttano la grande quantità di informazioni contenute nel dataset Google Ngram per analizzare l'evoluzione del lessico nel tempo, tracciando ad esempio la diffusione dell'influenza attraverso lo studio della frequenza del termine *influenza* nel tempo.

In ([3]), invece, sfruttando metodi di semantica distribuzionale gli autori analizzano il cambiamento semantico del lessico nel corpus English Gigaword ([7]) per determinare le fasi dei conflitti mondiali.

In questo lavoro focalizziamo l'analisi sullo studio dell'uso del maschile e del femminile nei titoli occupazionali, in un corpus diacronico di articoli di quotidiani italiani che copre il periodo 1948-2005.

Fenomeni sociali e culturali possono portare a una trasformazione del lessico nel corso del tempo con l'introduzione di neologismi o cambi nella frequenza d'uso del maschile o del femminile di un particolare titolo occupazionale.

Questo tipo di cambiamenti diventa più evidente in lingue, come l'Italiano che presentano un genere grammaticale.

Il genere grammaticale può infatti influire sul modo in cui il genere semantico è percepito. Accade quindi che ad esempio sia preferito l'uso del maschile a quello del femminile di un titolo occupazionale o viceversa, per ragioni storiche o di uso del linguaggio, non tenendo conto del genere semantico dell'agente.

In questo lavoro presentiamo uno strumento per la visualizzazione delle occorrenze dei titoli occupazionali e delle entità collegate ai titoli occupazionali per permettere un'analisi quantitativa e qualitativa di dettaglio.

2. CORPUS

I dati sono stati estratti da due corpus diacronici. Il primo è il corpus diacronico de "l'Unità" ([1]), un corpus in lingua italiana ottenuto dall'elaborazione delle notizie estratte dall'omonimo giornale. Il corpus copre un esteso periodo temporale che va dal 1948 al 2014. Il corpus è stato elaborato riducendo errori dovuti all'errato riconoscimento dei caratteri da parte di sistemi di riconoscimento automatico dei caratteri e applicando sistemi di Natural Language Processing per l'annotazione automatica di PoS tags, Named Entities e relazioni di dipendenza.

Il secondo corpus è stato invece ottenuto utilizzando tecniche di crawling applicate all'archivio digitale pubblico della Stampa e copre il periodo 1945-2005. Il corpus è stato processato usando la stessa metodologia utilizzata per il corpus "l'Unità" descritta in ([1]).

I due corpus sono stati uniti, allineando i due diversi periodi temporali, ovvero considerando gli articoli dell'Unità la cui data di pubblicazione è antecedente al 2006.

Il corpus finale contiene 3,529,820,155 tokens e copre il periodo 1948-2005. Nonostante le elaborazioni mirate al filtraggio e all'attenuazione del rumore, permangono errori dovuti agli strumenti di riconoscimento automatico del carattere (OCR). Inoltre, il numero di articoli e quindi il numero di tokens non è uniformemente distribuito. Il numero di tokens, infatti, diminuisce notevolmente nel periodo iniziale.

3. ESTRAZIONE TITOLI OCCUPAZIONALI ED ENTITÀ

Per poter estrarre le occorrenze dei titoli occupazionali e delle entità collegate ad esse per prima cosa è necessaria una lista di titoli occupazionali e la relativa forma flessa maschile e femminile. Per la generazione di una lista di titoli occupazionali con la relativa forma flessa maschile e femminile abbiamo utilizzato WikiData.

WikiData è una base di conoscenza aperta e accessibile liberamente, parte della famiglia di progetti della fondazione Wikimedia. WikiData si presenta come la risorsa principale di dati strutturati contenuti in altri progetti come Wikipedia o Wiktionary. Per la generazione della lista abbiamo ritrovato tutte le entità WikiData di tipo *wd:Q28640* (professione) presenti in WikiData e dotate della proprietà *P2521*, che specifica la forma grammaticale femminile. La lingua italiana ammette quattro categorie di nomi, classificati sulla base delle regole morfologiche e grammaticali utilizzate per riferirsi al genere semantico:

- Nomi di genere comune: Il genere grammaticale è determinato dall'articolo che precede il nome es. il giudice/la giudice;
- Nomi di genere promiscuo: Il genere semantico può essere asserito solo aggiungendo descrittori es. la guardia;
- Nomi di genere mobile: Seguono le regole morfologiche della lingua italiana es. maestro/maestra;
- Nomi di genere fisso: Le forme grammaticali maschile e femminile non condividono la stessa radice es. genero/nuora.

Nella lista estratta abbiamo escluso i titoli occupazionali che rientravano nella categoria di nomi di genere promiscuo e di genere comune in quanto non distinguibili nel testo.

La lista è stata ulteriormente elaborata con lo scopo di rimuovere nomi che possono presentare ambiguità semantica e quindi alterare le statistiche. In particolare, abbiamo automaticamente filtrato quei titoli occupazionali nei quali la forma femminile corrisponde anche al nome di una disciplina, es. matematica, fisica. Per fare ciò, abbiamo utilizzato WordNet, un database lessicale che permette di ottenere a partire da un lemma i possibili *synset* (significati) associati a quel lemma. Un altro strumento utile al fine di diminuire l'ambiguità è quello del riconoscimento delle categorie grammaticali (PoS tagging). Nel calcolare le occorrenze dei titoli occupazionali, abbiamo infatti sfruttato le annotazioni fornite con il corpus per considerare nel conteggio solo quelle occorrenze dei titoli occupazionali annotate con la categoria grammaticale NOUN (*nome*). In questo modo abbiamo potuto escludere occorrenze che creavano rumore, come quelle dove *editrice* compare come aggettivo, es. "la casa editrice". Infine, abbiamo considerato solo le forme singolari dei titoli occupazionali, in quanto le forme plurali possono indurre ambiguità. Ad esempio, *infermiere* può riferirsi sia al maschile singolare che al femminile plurale.

Le entità presenti nel corpus (persone, luoghi, organizzazioni e date) sono state già etichettate nella fase di elaborazione del corpus. Per estrarre quelle che si riferiscono a particolari titoli occupazionali sfruttiamo la dipendenza sintattica presente tra il titolo occupazionale e l'entità. Un'entità viene associata ad un titolo occupazionale se tra loro intercorre una dipendenza diretta. Un esempio è illustrato in Figura 1.

Ludmilla Tcherina e il celebre ballerino Rudolf Nureyev sono a Montecarlo

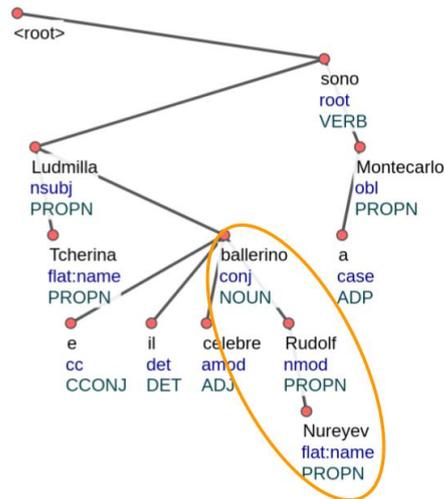


Figura 1. Estrazione occorrenza dell'entità "Rudolf Nureyev" sfruttando la dipendenza sintattica con il nome "ballerino".

4. ELABORAZIONE DEL DATO

Le occorrenze delle forme maschili e femminili dei titoli occupazionali sono state aggregate per anno calcolando la frequenza assoluta di ciascun titolo occupazionale in ogni anno.

La frequenza assoluta, tuttavia, può essere condizionata dalla distribuzione dei tokens, che, come detto nella Sezione 2, non è uniforme per tutti gli anni. Le frequenze assolute, pertanto, sono state normalizzate, dividendole per il numero di tokens presenti in ogni anno. Per ogni titolo occupazionale si ottengono così due serie temporale: quella delle frequenze relative del titolo al maschile e quella delle frequenze relative del titolo al femminile per il periodo 1948-2005.

Allo stesso modo le occorrenze di ciascuna entità sono state aggregate per anno. Per offrire una maggiore leggibilità e interpretabilità del dato nel caso delle entità abbiamo calcolato le frequenze assolute. Dalla frequenza assoluta di ciascuna entità è possibile determinare il periodo di attività, di maggiore fama o ad esempio correlare un incremento/decremento della frequenza a particolari eventi sociali/culturali. Un esempio di analisi che è possibile realizzare a partire dai dati estratti è descritta in ([2]).

Per indagare i motivi di un cambio di frequenza sarebbe necessario leggere gli articoli che parlano di quella particolare entità, operazione che può diventare onerosa in termini di tempo. Per superare questo ostacolo è possibile sfruttare i termini che più spesso compaiono con l'entità come indicatori della semantica della frase in cui l'entità compare. La maggior parte dei modelli di Natural Language Processing che trattano la semantica si basano sull'ipotesi distribuzionale, ovvero dal fatto che la semantica di una parola è determinata dai contesti in cui quella parola compare. Nei modelli di semantica distribuzionale (DSMs) le parole sono associate a vettori in spazi geometrici multidimensionali. I primi DSMs si basavano semplicemente sul contare quante volte una parola appare in un determinato contesto, frase, paragrafo o documento, a seconda della granularità desiderata.

Un approccio più recente e diffuso alla semantica distribuzionale è quello basato sulla predizione. I modelli basati sulla predizione creano una rappresentazione continua delle parole per predire la distribuzione di probabilità condizionata della parola dato il contesto. Un esempio di modelli basati su predizione è Word2Vec ([6]).

Sfruttando Word2Vec abbiamo costruito uno spazio vettoriale su ogni anno, utilizzando un token speciale per le entità. Per ogni entità e per ogni anno abbiamo calcolato le 10 parole più simili e la rispettiva similarità. Infine, abbiamo memorizzato la similarità dell'entità con le 10 parole più simili per ciascun anno.

5. VISUALIZZAZIONE

La visualizzazione dei dati estratti ed elaborati avviene per mezzo di un'interfaccia web. In particolare, l'interfaccia consiste di una pagina web il cui aspetto è definito con il linguaggio di markup HTML e la formattazione con i fogli di stile CSS. La parte di elaborazione e manipolazione del dato invece è gestita in Python.

L'interfaccia permette di selezionare un titolo occupazionale da un menu a tendina posto in alto nella pagina. Con la selezione verranno visualizzate le due serie temporali delle frequenze relative riferite rispettivamente al titolo occupazionale nella forma maschile e nella forma femminile.

La selezione del titolo occupazionale, inoltre, genera la lista delle entità estratte per ciascuna delle due forme grammaticali. Per ciascuna entità viene mostrato il grafico delle frequenze assolute, confrontabile con il grafico delle frequenze relative del titolo occupazionale di riferimento.

Se disponibile il vettore Word2Vec dell'entità viene visualizzato un secondo grafico. Il grafico rappresenta una matrice le cui righe sono le 10 parole più simili semanticamente all'entità e le colonne gli anni dal 1948 al 2005.

Ogni elemento della matrice, quindi, rappresenta la similarità tra il vettore dell'entità e una specifica parola in uno specifico anno. La scala dei colori va dal viola al giallo, dove il viola indica una similarità bassa e il giallo una similarità alta. Il sistema inoltre è in grado di interrogare WikiData per estrarre automaticamente informazioni aggiuntive riguardo l'entità come la foto o l'anno di nascita, utili per dare un contesto e facilitare l'analisi.

Un esempio è riportato in Figura 2. L'esempio mostra le serie temporali delle frequenze relative di ballerino e ballerina (in blu) e le serie temporali delle frequenze assolute delle entità *Roberto Bolle* e *Carla Fracci* (in verde).

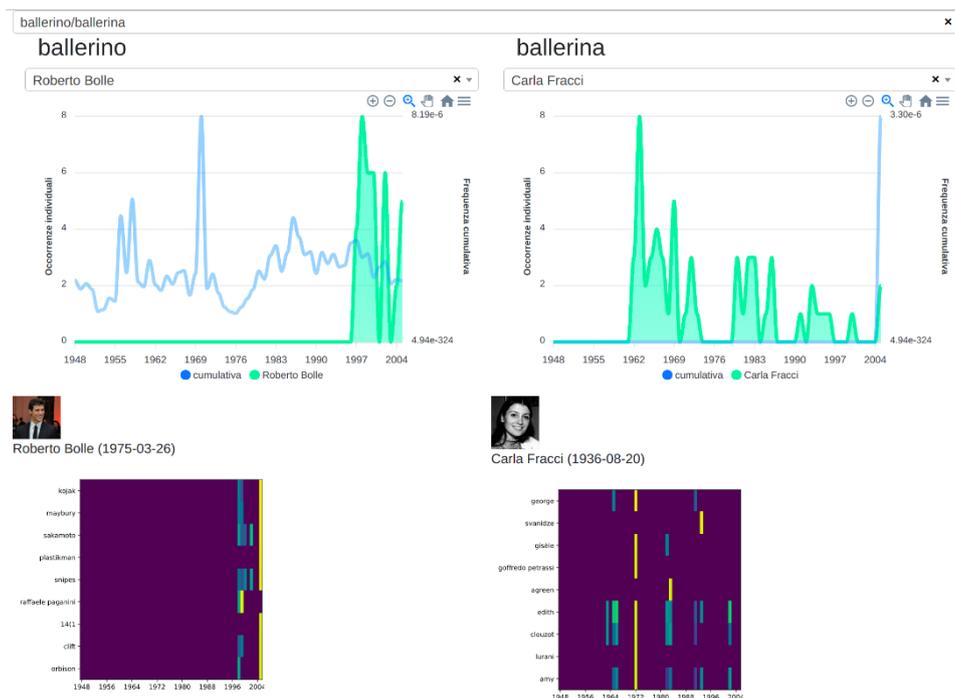


Figura 2. Interfaccia web.

6. CONCLUSIONI

In questo lavoro abbiamo presentato uno strumento per la visualizzazione della frequenza relativa di titoli occupazionali e delle frequenze assolute di entità collegate ai titoli occupazionali estratte da un corpus diacronico di articoli di quotidiani italiani. Lo strumento inoltre permette di visualizzare il cambiamento semantico dell'entità nel corso del tempo sfruttando tecniche di semantica distribuzionale. I dati estratti e lo strumento per la visualizzazione dei dati sono liberamente accessibili¹.

BIBLIOGRAFIA

- [1] Basile, Pierpaolo, Annalina Caputo, Tommaso Caselli, Pierluigi Cassotti, e Rossella Varvara. 2021. «A Diachronic Italian Corpus based on “L’Unità” 2020». In . Vol. Vol. 2769. Bologna, Italy: Accademia University Press.
- [2] Cassotti, Pierluigi, Andrea Iovine, Pierpaolo Basile, Marco De Gemmis, e Giovanni Semeraro. 2022. «Emerging trends in gender-specific occupational titles in Italian Newspapers». In . Milan, Italy.
- [3] Kutuzov, Andrey, Erik Velldal, e Lilja Øvrelid. 2017. «Tracing armed conflicts with diachronic word embedding models». In *Events and Stories in the News Workshop@ACL 2017*, 31–36. Vancouver, Canada.
- [4] Lin, Yuri, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman, e Slav Petrov. 2012. «Syntactic Annotations for the Google Books Ngram Corpus». In *Proceedings of the System Demonstrations*, 169–74. Jeju Island, Korea.
- [5] Michel, Jean-Baptiste, Yuan Shen, Aviva Aiden, Adrian Veres, Matthew Gray, Joseph Pickett, Dale Hoiberg, et al. s.d. «Quantitative Analysis of Culture Using Millions of Digitized Books». *Science* 331 (2011): 176–82.

¹ <https://github.com/pierluigic/igsot>

- [6] Mikolov, Tomás, Chen Kai, Corrado Greg, e Dean Jeffrey. 2013. «Efficient Estimation of Word Representations in Vector Space». In *Workshop Track Proceedings*. Scottsdale, Arizona, USA.
- [7] Parker, Robert, David Graff, Jumbo Kong, Ke Chen, e Kazuki Maeda. 2011. «Linguistic Data Consortium». Linguistic data consortium. 17 giugno 2011. <https://catalog ldc.upenn.edu/LDC2011T07>.

GenderedOntoComedy: Toward a Gendered Representation of Literary Characters in the Dante's Commedia

Selenia Anastasi¹, Andrea De Domenico², Marianna Nicolosi-Asmundo³

¹University of Genoa, Italy – 146253@studenti.unige.it

²Vrije Universiteit Amsterdam, Netherlands, - a.de.domenico@vu.nl

³University of Catania, Italy - marianna.nicolosiasmundo@unict.it

ABSTRACT

In this paper we describe GenderedOntoComedy, an OWL 2 ontology for the gendered description of Dante's female characters in the *Divina Commedia*. GenderedOntoComedy is the result of a transversal research activity integrating aspects from the fields of Gender Studies and Computer Science, focusing on an in-depth study of the representation of the female characters in Dante's *Commedia*. GenderedOntoComedy extends the ontology GenderedCHContents which in turn is an extension of the Europeana Data Model, designed to draw attention to the extensive presence of women in the fields of Cultural Heritage. Furthermore, for the description of *Divina Commedia*, we rely on OntoComedySources, an ontology modeling iconographic and literary sources of the *Divina Commedia*. GenderedOntoComedy is an attempt to investigate the Ontology of Gender, in consideration that literary characters, whether historical or not, can create a gendered conceptual map of women's representation through time.

KEYWORDS

Ontology of Gender, Semantic Web, Formal Ontology, Divina Commedia.

TALK

1. INTRODUCTION

In recent years, Gender Studies is attracting more and more interest between Humanities, Social Science and in the field of Computer Science, also due to the low representation of women in the so-called STEM disciplines. The emphasis attributed to the Web as an instrument of female liberation and resistance, which had characterized the creative period of Cyberfeminism during the past 1990s ([8];[16]), must reckon with an important gender gap that still needs to be filled. This lack has consequences on the level of the development of IT resources and proper tools for the analysis of gender contents, where, on the contrary, a greater demand and digital support is needed in the field of Literary Criticism.

Identifying gendered contents and symbols is a complex challenge, as for a long time women have been victims of prejudices too deep-rooted in common sense to attract curiosity. The uprising of the women's issue has a relatively recent history, thanks to the political struggles of feminism of the so-called first and second wave, following the critical articulation of oppression by intellectuals, writers and collectives of activists whose activities are hard to summarize in a single text. In the field of Semantic Web Technologies, still few studies have focused on the development of tools for the analysis of stereotypical gender representation within literary texts. As a specific representative example, the ontology GenderedCHContents ([12]) is a model for describing gendered aspects of Cultural Heritage, by focusing on the ontological category of gender itself.

In this contribution we present GenderedOntoComedy¹, an ontology for the gendered representation of Dante's female characters in the *Divina Commedia*. The choice to model Dante's female characters can be regarded as a complex challenge on two main fronts: the first one has to do with the difficulty of adapting feminist epistemology and Gender Theory – mostly hostile to essentializing labels – to the descriptive aims of formal ontologies, the second one with the scant presence of Semantic Web resources for the description of gendered traits. In order to develop new entities, properly designed to describe the difference between an historical and a fictional character, we follow the works of Fabio Ciotti ([3]) and Hastings and Schulz ([9]). To describe gendered contents, GenderedOntoComedy adopts the Constructionist approach to

¹ <https://github.com/SeleniaAnastasi/GeneredOntoComedy>.

gender ([13]). However, we felt the need of introducing new classes and properties in order to fully describe Dante’s particular female repertoire, such as that of *Gendered Stereotypes* and its subclass *Medieval Stereotypes*.

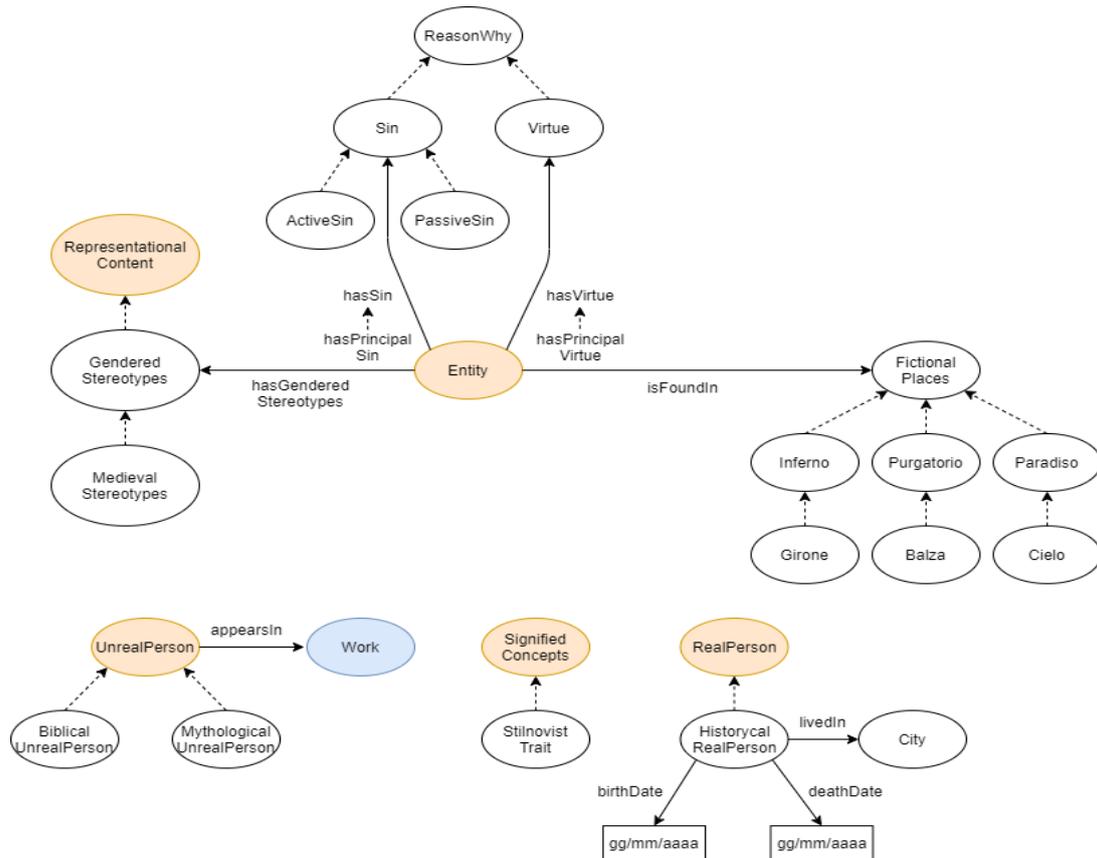


Figure 1: GenderedOntoComedy classes and properties related to the class *Entity* from the ontology *GenderedCHContents*.

At a first stage, we want to draw attention to the literary dominion, pointing out how women have always played an ancillary or inactive role within the narratives. As the pioneering works of Virginia Woolf underlined ([17];[18]), in the literary production women are mothers, wives, lovers, daughters or sisters, but they are rarely the heroines of their stories.

Dante’s *Commedia* is a representative case study of this trend in classic literary canon. In fact, in Dante, female characters are present in an extended way throughout the narration, according to well established stereotypes of the time (such as Beatrice the *donna angelicata*). Only in the *Inferno*, where Dante places what is blamable and distorted, he recognizes women a social status (as historical figures of power) and gives them a predominant role through dialogues. The *Commedia* therefore represents an interesting case study, considering also the importance of gendered features such as those typical of the *Dolce Stil Novo*.

In addition to the Gender Studies literature, we based our description of female characters on the study of Brooke L. Carey, “Le Donne in Dante: An Historical Study of Female Characters in The Divine Comedy” ([1]). Carey’s approach pays particular attention to the historical connotation of the female characters in Dante’s production, as well as to the aspects that we can consider remarkable from a literary as well as sociological point of view.

The expressive power of OWL 2 (Web Ontology Language), made possible to model relevant entities in a precise way, without reducing the complexity of the study in its theoretical implications: on the one hand, the literary criticism, on the other, the political needs of Gender Studies. Given the attention we paid to the description of women’s material conditions and historical connotation of each gendered connotation, we can consider our approach as epistemologically situated ([12]).

2. DESCRIPTION OF GENDEREDONTOCOMEDY

To define *GenderedOntoComedy* ontology, we chose to extend the classes *RepresentationalContent* and *SignifiedConcepts*, both subclasses of the class *Content*, main class of the ontology *GenderedCHContents*, which provides descriptions of the content of a Cultural Heritage artifact. The class *RepresentationalContent* is used to describe the content of the artifacts, and we also provide the possibility to express a *Non-representationalContent*, in case of the needing to describe artifacts that contain non-figurative art. Since we are dealing with narrative characters, and we are always facing

with figures with a clear describable appearance, such as human beings, animals, creatures, etc., it has been enough to extend the class *RepresentationalContent*. The *SignifiedConcepts* class is derived from the Semiotics of Saussurean matrix and indicates the meanings of certain entities (i.e., metaphorical, religious, mystical and gendered symbols). All the rest of the classes are subclasses of *RepresentationalContent*, which contains representative entities that convey gendered characteristics ([12]).

Fig.1 depicts the principal architecture of GenderedOntoComedy, where orange circles indicate classes from GenderedCHContents, blue circles classes from OntoComedySources and white circles new main classes from GenderedOntoComedy. These include gender characteristics, sins and virtues by which an entity is judged by Dante and fictional places in which the entity is subsequently placed. To better express these concepts, in particular the reference to stilnovistic poetic and medieval gender stereotypes, we extended the classes of GenderedCHContents by means of the class *GenderedStereotypes* and its subclass *MedievalStereotypes* and *StilnovistTrait*, both subclass of *SignifiedConcepts*. Furthermore, the connection with the ontology OntoComedySources is constituted by the *appearsIn* object property.

As stated in the paragraph above, one of the main aspects of GenderedOntoComedy is its situated approach. In fact, we considered the treatment that Dante reserves for character's bodies within the Commedia. Feminist Theory, in particular the French tradition (Christine Delphy, Monique Wittig and Colette Guillaumin), stressed the question of the classical representation of women as embodied entities, material subjects, therefore subjected to all the weaknesses that the Christianity, as social dispositive of oppression, has assigned to women's body – opposed to the rational and incorporeal virtue of the male sex, so much closer to the divine lightness. According to Carey, Dante represents this precise dichotomy in the Commedia through the gradual shift from the detailed physical descriptions of the tortures suffered by the penitents of the Inferno, to the lightness and candid innocence of the virtuous creatures of the Paradiso. We indicate this distinction in our ontology through the *DisembodiedAppearance* and *EmbodiedAppearance* classes. As an example of the use of this class we consider Francesca da Rimini, a main female character of the Inferno, represented as an embodied figure tormented by the opposite winds that characterize the circle of the lustful. Worthy of note is the ability of Francesca to persuade Dante of her innocence, thanks to the way she speaks about her love with Paolo – skill that is typical of the poetics of the Dolce Stil Novo. As Carey states, «Francesca speaks like a stilnovist whose gentle heart was taken by the love and overcome by it» ([1]). According to this interpretation, the piety aroused in Dante by Francesca's words reinforces the medieval belief that women possess an intrinsic seductive and deviant nature. Thus, the idea that women's speeches have a deviant and seductive nature for the male soul, is a trait shared by both the *StilnovistTrait* and the *MedievalStereotypes* classes, both subclasses of *GenderedStereotypes*. These classes are designed to carry the set of gender stereotypes, both male and female, providing that they change over time and can be roughly divided into historical periods. In addition to the concepts related to the sinful nature of the feminine, the Dolce Stil Novo affirms a new concept of love, as well as a new concept of woman: the platonic love and the *donna angelicata*. In the stilnovist tradition, these figures play the function of guiding the soul of a man towards noble gestures and thought, thanks to the expression of an absolute love, similar to that of God – and since similar to that of God, it cannot foresee sensual and bodily involvement. All these concepts related to the purifying or deviant function of women and love, are well expressed through the class *StilnovistTrait*.

Our approach to situated knowledge representation prompts us to take particular account of the fictional places where the characters appear and where the interactions between the pilgrim Dante and the damned take place. Carey once again underlines how in Dante the assignment of a specific place within the geography of the Divina Commedia is always justified, and often driven by gender prejudices, due to his historical time and his belonging to a certain poetic tradition. Taking once again as an example the emblematic case of Francesca da Rimini, the historian observes:

The setting in which Dante places Francesca and Paolo also provides a context for how the poet reinforces stereotypes about the inherent inferiority of women and superiority of men. [...] The fact that Dante renders Francesca the dominant figure in the canto illustrates the anomaly of the assertive woman and passive man. Everything in Hell, except for the justice that governs it, is perverse[...]. By switching feminine and masculine characteristics and situating the lovers in Hell, Dante perverts the lovers and makes them unnatural ([1]).

The importance of places for the understanding of the characters narrated in the Commedia, led us to consider the need to extend the *Entity* class with the subclass *FictionalPlaces*, dedicated to narrative places, which in turn contains the Inferno, Paradiso and Purgatorio classes and their respective geographical subdivision into Cerchi, Balze and Cieli. Thus, it will be possible to distinguish two cases: the one in which the character is associated with the place where she physically appears within the narrative (Francesca da Rimini in the II Girone of the Inferno), but also on the intradiegetic level, when Dante mentions characters he meets or who knows to be present in a specific place, neither they play an active role, nor does the pilgrim Dante interact directly with them.

To make this distinction clear we employ the *Line* class, belonging to *OntoComedySources*, to indicate the quotation of a character in the line, while the material and active presence of a character on the level of the plot is expressed with the subclasses of the class *FictionalPlaces*. To model such a distinction, we further introduce two object properties: *quotedIn* for the former case, having the class *Entity* as its domain and the class *Line* of *OntoComedySources* as its range, and the object property *isFoundIn* to express the level of the plot, with the class *Entity* as its domain and the class *FictionalPlaces* as its range.

As we already mentioned, Dante assigns each character a precise location in his geography, in consideration of the type of virtue or guilt he attributes to them in life. If, on the one hand, it was therefore necessary to describe these places through classes and object properties, in the same way it is essential to understand the reasons behind Dante’s choices. To this end, we introduce the *ReasonWhy* class. Here we instantiate the subclasses *Sin* and *Virtue*, where *Sin*, in turn, is divided into *ActiveSin* and *PassiveSin*, underlining a nuance of meaning recognized by Carey. She observes how often, in the *Commedia*, sins are committed by virtue of the power of external will, as is the case of Lust. According to Carey, this aspect has a gender connotation, as it implies the innocence of the man, driven to sin because of the pervert seduction operated by women.

3. REASONING WITH GENDEREDONTOCOMEDY

GenderedOntoComedy has been classified using the DL (Description Logics) reasoner Hermit ([7]). Along with standard inferences concerning characteristics of properties such as functionality and inverse properties, domains and ranges of properties, GenderedOntoComedy also includes SWRL rules in order to infer simple assertions concerning entities, in relation with the verse, the Canto, the Cantica, and in general the fragments of the *Commedia* where they are being quoted.

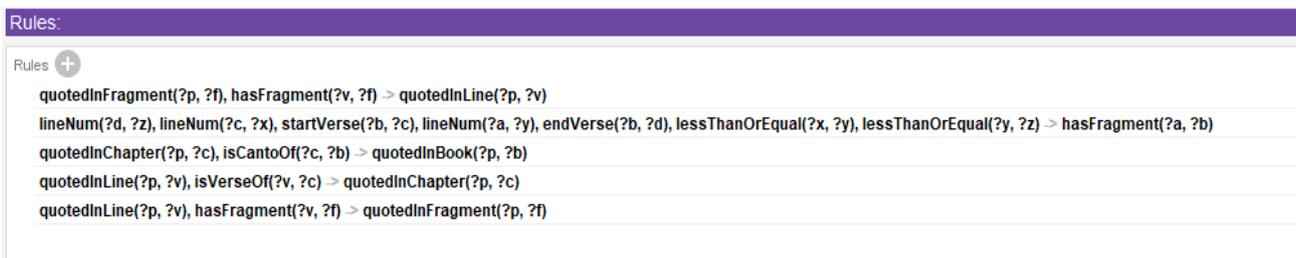


Figure 2: Some SWRL rules concerning entities and places in the text where they are quoted.

The rules are listed in Fig. 2. In particular, *hasFragment(a, b)* means that the verse *a* is contained in one of our fragments *b*, whereas *quotedInLine(p, v)* holds if an entity *p* is quoted in a certain verse *v*. If the same entity *p* is found in a Canto *c*, we say *quotedInChapter(p, c)*, and to express that *p* is being quoted in a Cantica *b*, we affirm *quotedInBook(p, b)*. Finally, *quotedInFragment(p, f)* means that the fragment *f* speaks of an entity *p*.

4. CONCLUSIONS

In this work we presented GenderedOntoComedy, an OWL 2 ontology for the representation of gendered traits of fictional characters in Dante’s *Commedia*. For the purpose of this paper, we focused exclusively on female characters, in order to emphasize Medieval gender stereotypes and several stylistic features in the poetics of *Dolce Stil Novo*. GenderedOntoComedy is modeled according to GenderedCHContents, an extension of the Europeana Data Model (EDM), in order to accurately represent gendered features in the Cultural Heritage domain. It is worthy of note that the design of GenderedOntoComedy follows a foundational approach: it not only aims to develop a resource to describe the ontological status of fictional characters, but most importantly, through the modeling of classes, it aims to question the Ontology of Gender. The emphasis placed on overcoming the nature-culture dichotomy, opens up interesting scenarios for investigating the relationship between power structures, society and gender identity, and therefore between the ontological and epistemological domains. Indeed, reflecting on the way in which gender is represented in the artifacts of Cultural Heritage made possible to bring to light, in the case of Dante, prejudices about women that are still present, such as the association between political and sexual power, the dichotomy of holiness-wickedness and the de-humanized sublimation of the beloved woman (as in the *Dolce Stil Novo*).

We plan to expand the ontology to represent additional gender traits, old and new cultural stereotypes, optimizing class management to include masculine gender characterizations in Dante’s *Commedia*. Given the scarcity of foundational

resources for describing gender characteristics, GenderedOntoComedy will be further extended to model other literary and iconographic works.

REFERENCES

- [1] Bartalesi, Valentina, and Carlo Meghini. 2017. "Using an Ontology for Representing the Knowledge on Literary Texts: The Dante Alighieri Case Study." *Semantic Web* 8 (3): 385–94. <https://doi.org/10.3233/SW-150198>.
- [2] Carey, Brook L. 2007. "Le Donne Di Dante: An Historical Study of Female Characters in The Divine Comedy." Syracuse University Honors Program Capstone Projects.
- [3] Ciotti, Fabio. 2017. "Toward a Formal Ontology for Narrative." *Centro de Literatura Portuguesa; Imprensa Da Universidade de Coimbra* 4(1).
- [4] Damiano, Rosanna, and Antonio Lieto. 2013. "Ontological Representation of Narratives: A Case Study on Stories and Actions." In *Proceedings of CMN13*. Hamburg.
- [5] Doerr, Martin. 2003. "The CIDOC-CRM An Ontological Approach to Semantic Interoperability of Metadata." *AI Magazine* 24 (3): 75–92.
- [6] Doerr, Martin, Stefan Gradmann, Steffen Henniecke, Antoine Isaac, Carlo Meghini, and Herbert Van De Sompel. 2010. "The Europeana Data Model (Edm)." In *World Library and Information Congress*, 10-15.
- [7] Glimm, Birte, Ian Horrocks, Boris Motik, Giorgos Stoilos, and Zhe Wang. 2014. "Hermit: An OWL2 Reasoner." *Journal of Automated Reasoning* 53: 245–69.
- [8] Haraway, Donna. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14(3): 575–99.
- [9] Hastings, Janna, and Stefan Schulz. 2019. "Representing Literary Characters and Their Attributes in an Ontology." *JOWO*.
- [10] Hofweber, Thomas. 2018. "Logic and Ontology." In *The Stanford Encyclopedia of Philosophy*, Summer 2018. Edward N. Zalta.
- [11] Kroon, Fred, and Alberto Voltolini. 2018. "Fictional Entities." Winter 2018. Edward N. Zalta.
- [12] Kyvernitou, Ioanna, and Antonis Bikakis. 2017. "An Ontology for Gendered Content Representation of Cultural Heritage Artefacts." *DHQ: Digital Humanities Quarterly* 11(3).
- [13] Mikkola, Mari. 2008. "Feminist Perspectives on Sex and Gender." In *The Stanford Encyclopedia of Philosophy*, Fall 2019. Edward N. Zalta.
- [14] Oberle, Daniel, Nicola Guarino, and Steffen Staab. 2009. "What Is an Ontology?" In *Handbook on Ontologies*. Springer.
- [15] O'Brien, Jodi. n.d. *Encyclopedia of Gender and Society*. Vol. 1. Seattle University, USA: SAGE.
- [16] Turkle, Sherry. 1998. *Sex/Machine: Readings in Culture, Gender, and Technology*. Patrick Indiana University: Patrick D. Hopkins.
- [17] Woolf, Virginia. 2001. *Three Guineas*. Penguin Books.
- [18] ———. 2004. *A Room of One's Own*. Penguin Books.

Sessione Filosofia 1
Marisa Bellisario

Gli indici della prima modernità come strumento storiografico: questioni preliminari metodologiche e pratiche

Enrico Pasini

(ILIESI/CNR) – enrico.pasini@cnr.it

ABSTRACT

Il contributo ha come finalità generale quella di tematizzare, a partire da un insieme di questioni metodologiche e pratiche, una possibile linea di indagine relativa all'euristica e al patrimonio terminologico contenuti negli indici (*indices rerum*) delle opere filosofiche (principalmente di medio e grande formato) della prima modernità, che si tratti di grandi trattati scolastici o di testi di filosofia naturale, di medicina teorica (non esistendo all'epoca la filosofia della medicina, ma esistendo una significativa concorrenza accademica tra filosofi e medici): un campo di ricerca che potrebbe essere estremamente fruttuoso ma che presenta significative difficoltà su entrambi i piani succitati. Il focus del contributo, che potrebbe unirsi ad altri di analoga impostazione in un panel, è sugli indici come chiavi di accesso dinamiche ai contenuti e come possibili 'infrastrutture concettuali' per l'interoperabilità. Lo studio degli indici è considerato come la produzione storiograficamente consapevole di una classe di *nouveaux observables*, secondo la formula di François Rastier ([9]).

L'obiettivo, comune ad altre proposte discusse nell'ambito della Conferenza AIUCD2022, è aprire un confronto con la comunità degli umanisti e dei filosofi "digitali" ([3]) sull'utilizzo e il trattamento degli indici dei testi filosofici come strumenti di navigazione e ricerca non gerarchica o lineare dei contenuti, e indagare e discutere in che misura e in che modo le entrate dell'indice vadano trattate da un lato, come metadati descrittivi e vocabolari controllati, dall'altro come strumenti dinamici e relazionali con un intrinseco potenziale euristico forte anche della polisemia e della stratificazione semantica che contraddistingue la terminologia filosofica ([12]).

PAROLE CHIAVE

Indices rerum, strumento storiografico, filosofia.

INTERVENTO

1. PROGETTO DI RICERCA E IPOTESI DI LAVORO

1.1 Di che si tratta

Quando Leibniz pubblicò la sua unica opera di vasta portata apparsa con lui vivente, si preoccupò di realizzarne personalmente la "Table des matières", che offre una visione privilegiata sullo scheletro terminologico e concettuale della *Teodicea*. Prodotti dagli autori o dai collaboratori dello stampatore, simili indici accompagnano molte opere filosofiche, mediche, scientifiche o di filosofia naturale, di teologia ecc., ove il formato dell'opera richieda una facilitazione della consultazione. La terminologia che si presenta in queste sezioni ha alcune caratteristiche di evidente interesse:

- rappresenta un vocabolario controllato nel senso stretto del termine (e quindi rappresentabile, in linea di principio, partendo da strutture dati e formalizzazioni già esistenti e sufficientemente standardizzate), con una chiara connotazione disciplinare e sub-disciplinare;
- ha sovente una struttura gerarchica, cioè risponde a un modello di organizzazione tradizionalmente usato nella rappresentazione di insiemi concettuali, offrendo delle istanziazioni storicamente radicate;
- rispecchia le trasformazioni linguistiche della produzione editoriale che accompagna, a partire dalla prevalenza iniziale del latino e la progressiva affermazione delle lingue che saranno dette nazionali;
- consente in importanti casi una comparazione plurilingue;
- è il frutto di scelte consapevoli e gode dunque di una rappresentatività particolare;
- è (da qui partono le nostre considerazioni) estremamente difficile da studiare.

Quest'ultimo punto è specialmente rilevante. Le ragioni della difficoltà di tale studio sono, prima facie, le seguenti:

- la sostanziale impossibilità di affrontare questo genere di paratesti con tecniche di close reading;
- la probabile scarsa utilità di affrontare questo genere di paratesti con tecniche semplificate di distant reading, quali quelle impiegate nei primi lavori di Moretti ([7]);

- la complessità della produzione di corpora adeguati allo studio mediante tecniche sofisticate di distant reading ([13]) e analisi testuale automatica.

La questione della costituzione di corpora specifici per la storia delle idee, o storia del pensiero, *Begriffsgeschichte*, *History of Knowledge*, etc. sta cominciando a porsi apertamente ([8];[1]). Ma, a differenza dei corpora di testi primari, su cui esistono alcune basi di ragionamento, la costituzione di corpora di paratesto, che sembrerebbe ovvia nell'indagine digitale della produzione culturale ([4]), è stata sinora affrontata in modo limitato e specialmente *non* nel caso degli indici. Alcune ragioni sono ovvie:

- la scarsa attrattività della materia (lo studio dei *carmina gratulatoria*, ad esempio, è intuitivamente meno noioso);
- la complessità dell'acquisizione dei testi, considerato che molto sovente la qualità tipografica di tali sezioni è non solo non superiore, ma perlopiù inferiore a quella del testo, di cui dunque condivide in misura incrementata i problemi di tipografia storica;
- il rilevante lavoro di correzione che ci si può dunque aspettare;
- la necessità di produrre un adeguato modello di rappresentazione (o più modelli) ([2];[11];[6]) che è evidentemente preliminare alla registrazione dei dati ottenuti dalla digitalizzazione in un formato utilizzabile per ogni successiva analisi;
- in fatto di analisi, l'indispensabilità di complicate analisi comparative la cui automatizzazione richiede una certa customizzazione di tools e procedure.

D'altra parte, gli indici rappresentano un materiale concettuale selezionato in tempo "reale" da operatori culturali competenti radicati nella produzione e nel contesto. Questo, banalmente, riduce di molto il peso dell'analisi automatica richiesta nello studio di corpora di testi di carattere argomentativo e concettuale (p. es. il problema non banale della "concept detection") e, anzi, potrebbe fornire un insieme di thesauri per facilitare significativamente la successiva analisi dei corpora testuali culturalmente corrispondenti, anche al di là delle opere da cui gli indici provengono; si noti che spesso gli indici del tempo contengono non soltanto singoli lessemi, ma spesso anche la descrizione del risultato ottenuto argomentativamente o espositivamente in un determinato luogo del testo.

1.2 Ragioni di interesse

Se consideriamo questi indici come percorsi di accesso facilitato, come accade nell'ordinaria fruizione da parte di chi studia la storia del pensiero – che ne fa uso per evitare di dover percorrere l'intera opera quando va in cerca di una sezione relativa a uno specifico tema di interesse – risulta chiaro che siffatti indici rappresentano una struttura di dati solo apparentemente ausiliaria, ma in realtà intrinseca ai contenuti di un testo. Tuttavia, se li si considera come un corpus interconnesso e rappresentativo della produzione complessiva, in un certo arco di tempo, di uno specifico genere paratestuale a sé stante, non solo essi consentono, secondo una prospettiva che è anche propria dell'odierna *information retrieval*, di recuperare più velocemente i dati evitando la lettura dell'intero contenuto ([5]), ma essi possono rappresentare un'interfaccia di rappresentazione, studio e analisi di un'intera epoca di produzione di terminologia intellettuale e dunque rappresentare una chiave di accesso non soltanto ai testi, ma al complesso della produzione culturale di cui tali testi fanno parte. Questo rappresenta il principale aspetto innovativo che offrirebbero la dimensione digitale della ricerca e la produzione di una massa critica di informazione.

- Nei colossali indici dei grandi volumi della prima modernità, grazie alla possibilità di utilizzare il latino come termine di riferimento, ed essendo evidente che gli indici stessi sono le fonti dirette di thesauri storici sia latini sia nelle lingue nazionali, è acquisibile uno strumento fondamentale di indagine del nascente plurilinguismo della cultura europea e della sua terminologia.
- Le trasformazioni nel tempo del lessico presente negli indici permette di studiarne le trasformazioni diacroniche in base a un materiale, come si è osservato, già selezionato, e in effetti selezionato secondo procedure intrinsecamente non anacronistiche, momento per momento.
- Un ulteriore aspetto interessante della produzione degli indici delle opere di medio e grande formato della prima modernità è inoltre che ogni indice ha non soltanto un momento, ma anche un luogo di produzione, permettendo così di costruire rappresentazioni:
 - di reti terminologiche,
 - distribuite diacronicamente,
 - collocate geograficamente.

Per ogni concetto, o insieme di concetti, questo permetterebbe di generare diverse mappe interattive in cui l'accesso alle informazioni sia riferito non soltanto ai testi, ma anche ai tempi e ai luoghi, questi ultimi sia in rapporto ai testi, sia in quanto tali. Il luogo infatti relaziona persone, temi, produzioni editoriali, accademie e università, controversie e dibattiti.

In questo senso, in una prospettiva comune anche ad altri contributi, la raccolta dei dati avrebbe valore non solo in termini estrattivi ma anche in termini generativi/euristici e la visualizzazione si configurerebbe come una modalità di accesso di cui la rappresentazione formalizzata degli indici storici costituirebbe i percorsi di accesso.

2. CONCLUSIONI

Nel contributo ci si è voluto concentrare, in accordo con una prospettiva di ricerca condivisa con altri relatori e colleghi dell'AIUCD, sulle potenzialità euristiche degli indici come chiavi d'accesso dinamiche ai contenuti di un insieme di testi filosofici di carattere storico ([10]), discutendo nello specifico la fecondità scientifica della formazione e studio analitico di una collezione in forma di corpus strutturato e annotato degli *indices rerum*, *tables des matières* ecc. delle opere filosofiche e di discipline affini pubblicate nella prima modernità. Da un lato questo offrirebbe la possibilità di specifiche modellizzazioni dei contenuti dell'insieme di testi senza richiedere l'estrazione dei concetti mediante tecniche di analisi testuale; d'altro canto permettere di disegnare delle mappe culturali dotate di dimensioni multiple, geografiche temporali linguistiche e plurilinguistiche, aprendo a nuovi percorsi di ricerca; dall'altro, come è facilmente visibile, consentirebbe di utilizzare i risultati come una rete concettuale autonoma, da una parte, e dall'altra come una chiave di accesso ai testi stessi. L'intento del contributo è, in sostanza, quello di suscitare una discussione preliminare alla definizione collaborativa di un concreto progetto partecipato promosso dall'Istituto di ricerca di cui fa parte l'estensore di questa nota.

BIBLIOGRAFIA

- [1] Betti, Arianna. 2021. «Comment to Pasini». In *DR2 Working Papers*. Torino: Accademia UP.
- [2] Betti, Arianna, e Hein Van den Berg. 2014. «Modelling the History of Ideas». *British Journal for the History of Philosophy* 22 (4): 812–35.
- [3] Ciraci, Fabio, Cristina Marras, e Riccardo Fedriga, a cura di. 2021. *Filosofia digitale*. Quaderni di «Filosofia». Milano: Mimesis.
- [4] Desrochers, Nadine, Daniel Apollon. 2014. *Examining Paratextual Theory and its Applications in Digital Culture*. Hershey, PA: IGI Globa.
- [5] Hjørland, B. 2011. «The Importance of Theories of Knowledge: Indexing and Information Retrieval as an Example». *Journal of the American Society for Information Science and Technology* 62 (1): 72–77.
- [6] McCusker, Jamie, Joanne Luciano, e Deborah L. Mcguinness. 2011. «Towards an Ontology for Conceptual Modeling». In *Proceedings of the 2nd International Conference on Biomedical Ontology*.
- [7] Moretti, Franco. 2009. «Style, Inc.: Reflections on 7,000 Titles (British Novels, 1740–1850)». *Critical Inquiry* 36 (1): 134–58.
- [8] Pasini, Enrico. 2021. «On Some Challenges Posed by Corpus-Based Research in the History of Ideas». In *DR2 Working Papers*. Torino: Accademia UP.
- [9] Rastier, François. 2011. *La mesure et le grain. Sémantique de corpus*. Paris: Champion.
- [10] Russo, Ada, e Michela Tardella. 2017. «TheofPhilo. Thesaurus of Philosophy». *ILIESI Digitale*, 1–29.
- [11] Sugumarán, Vijayan, e C. Veda Storey. 2002. «Ontologies for Conceptual Modeling: Their Creation, Use, and Management». *Data & Knowledge Engineering* 42 (3): 251–71.
- [12] Tiberi, M., e F. Mazzocchi. 2007. «La gestione della polisemia nei thesauri: il caso dei termini filosofici». *Bollettino AIB47* (1/2): 93–107.
- [13] Underwood, Ted. 2016. «Distant Reading and Recent Intellectual History». In *Debates in the Digital Humanities*, M. Gold, L. F. Klein. Minneapolis: University of Minnesota Press. <https://dhdebates.gc.cuny.edu/projects/debates-in-the-digital-humanities-2016>.

Indici e mappe digitali per l'iter italicum di G. W. Leibniz

Cristina Marras

Istituto Lessico Intellettuale Europeo e Storia delle Idee, Consiglio Nazionale delle Ricerche

IT – cristina.marras@cnr.it

ABSTRACT

Il contributo ha come finalità generale quella di tematizzare, a partire da un concreto caso studio, l'euristica contenuta nella relazione tra filosofia e pratiche computazionali nella ricerca. Presenta, in sintesi, il modello di lavoro e alcuni primi risultati del progetto *L'iter Italicum di G. W. Leibniz: coordinate per una mappa digitale* che ricostruisce, rappresenta e visualizza i diversi e complessi aspetti del viaggio italiano del filosofo tedesco riportati e descritti analiticamente nel volume dello studioso francese André Robinet, *G. W. Leibniz, Iter italicum (Mars 1689-Mars 1690). La dynamique de la République des Lettres*, pubblicato dalla casa editrice Olschki di Firenze nel 1988. Il focus del contributo è sugli indici come chiavi di accesso dinamiche ai contenuti e come possibili 'infrastrutture concettuali' per l'interoperabilità. L'obiettivo è aprire un confronto con la comunità degli umanisti e dei filosofi "digitali" sull'utilizzo e il trattamento degli indici dei testi filosofici come strumenti di navigazione e ricerca non gerarchica o lineare dei contenuti, e indagare e discutere in che misura e in che modo le entrate dell'indice vadano trattate, da un lato, come metadati descrittivi e vocabolari controllati per l'interoperabilità, dall'altro come strumenti dinamici e relazionali con un intrinseco potenziale euristico forte anche della polisemia e della stratificazione semantica che contraddistingue la terminologia filosofica.

PAROLE CHIAVE

Indici, mappe digitali, Leibniz.

INTERVENTO

1. INTRODUZIONE

Il contributo ha come finalità generale quella di tematizzare, a partire da un concreto caso studio, l'euristica contenuta nella relazione tra filosofia e pratiche computazionali nella ricerca ([1]). Presenta, in sintesi, il modello di lavoro e alcuni primi risultati del progetto *L'iter Italicum di G. W. Leibniz: coordinate per una mappa digitale* che ricostruisce, rappresenta e visualizza i diversi e complessi aspetti del viaggio italiano del filosofo tedesco riportati e descritti analiticamente nel volume dello studioso francese André Robinet, *G. W. Leibniz, Iter italicum (Mars 1689-Mars 1690). La dynamique de la République des Lettres*, pubblicato dalla casa editrice Olschki di Firenze nel 1988. Il focus del contributo è sugli indici come chiavi di accesso dinamiche ai contenuti e come possibili "infrastrutture concettuali" per l'interoperabilità. Il contributo è diviso in due parti: nella prima, dopo avere descritto sommariamente le tappe del viaggio in Italia di Leibniz, si descrivono i dettagli del progetto dando conto anche della tecnologia utilizzata e del metodo di lavoro; nella seconda, si presentano alcuni primi risultati mettendone a fuoco criticità e aspetti innovativi. L'obiettivo è aprire un confronto con la comunità di DH e dei filosofi "digitali" sull'utilizzo e il trattamento degli indici dei testi filosofici come strumenti di navigazione e ricerca non gerarchica o lineare dei contenuti, e indagare e discutere in che misura e in che modo le entrate dell'indice vadano trattate, da un lato, come metadati descrittivi e vocabolari controllati per l'interoperabilità, dall'altro come strumenti dinamici e relazionali con un intrinseco potenziale euristico forte anche della polisemia e della stratificazione semantica che contraddistingue la terminologia filosofica ([13]).

2. PROGETTO DI RICERCA E MODELLO DI LAVORO

Durante tutto il corso della sua vita Leibniz ha viaggiato instancabilmente per tutta l'Europa, intessendo rapporti con gli scienziati, i politici, gli intellettuali dell'epoca anche attraverso una fitta corrispondenza (c.a. 1.100 corrispondenti) di importanza fondamentale per la storia delle idee, tanto che il fondo leibniziano è stato incluso nel *Memory of the World Register* dell'UNESCO. La dimensione del viaggio è componente centrale e importante della vita del filosofo che conosce molto bene gli aspetti tecnici relativi ai mezzi di trasporto, alla viabilità dell'epoca. La via però non è solo uno strumento di collegamento, ma una possibilità per acquisire e sviluppare conoscenze, diffondere e far valere idee e scoperte, la scelta di una via è anche vincolata ai problemi e ai temi nei quali il filosofo è coinvolto e alla sua visione cooperativa dell'avanzamento della scienza ([5]). Il viaggio in Italia di Leibniz si svolge tra il marzo 1689 e il marzo 1690, partito da

Venezia raggiunge Napoli, soggiornando, all'andata e ritorno, nelle principali città italiane. Il filosofo viaggia nella *République des Lettres* italiana con il ruolo istituzionale di consigliere aulico, bibliotecario del duca di Brunswick-Luneburg, con l'obiettivo ufficiale di trovare documenti e informazioni riguardanti le origini della storia dei Guelfi legati alla famiglia Estense. Ma il viaggio fu anche, e soprattutto, mosso da motivazioni politiche e scientifiche.

Il testo di André Robinet è uno scrigno di informazioni complesse ormai datato, ma uno strumento importante per il lavoro degli studiosi leibniziani e sul quale vale la pena applicare metodologie innovative, capaci di valorizzare e rendere più fruibili i contenuti e i risultati preziosi della ricerca dello studioso francese. La ricchezza e la stratificazione delle informazioni nel testo rende infatti in qualche modo difficile la fruizione dei dati e la comprensione della fitta rete intellettuale e scientifica intessuta da Leibniz, non consente di avere una visione di insieme della complessità dei temi, non facilita la correlazione tra scritti e nuclei tematici.

Il progetto *L'iter Italicum di G. W. Leibniz: coordinate per una mappa digitale* anziché digitalizzare l'intero volume, si concentra sugli indici che vengono utilizzati come chiavi di accesso dinamiche ai contenuti, sulla mappa geografica del viaggio, costruisce una serie di mappe concettuali e grafi risultato dell'aggregazione dei dati, così da restituire la multilinearità e l'interdipendenza delle informazioni ([10]). Si tratta di una scelta principalmente metodologica: la ricerca storico filosofica è condotta a partire da un modello di lavoro (ecosistema) ([6]) che concettualizza i processi di interscambio tra filosofia e digitale, mette insieme le strategie di accesso ai dati, le modalità di rappresentazione della conoscenza, con particolare attenzione alla modellizzazione dei dati stessi, testa gli strumenti informatici per la ricerca a uso della comunità (open, standard etc...), ottimizza le risorse (approccio ecologico). Il lavoro ha tra gli altri anche l'obiettivo di: a. rendere i dati ottenuti F.A.I.R, dunque rintracciabili, accessibili, interoperabili e riutilizzabili, b. consentire agli studiosi l'accesso a delle informazioni talvolta poco fruibili, c. grazie agli strumenti digitali aprire possibili nuove interpretazioni e linee di indagine.

Il progetto di ricerca parte dagli indici del volume di André Robinet e non dalla digitalizzazione del testo integrale. Consideriamo gli indici come "cammini di accesso", alternativi alle vie di accesso diretto alle informazioni e ai contenuti del volume. L'utilizzo degli indici, considerati come indici dinamici, una struttura di dati apparentemente ausiliaria, ma sicuramente intrinseca ai contenuti di un testo che consente di recuperare più velocemente i dati (*information retrieval*) evitando la lettura dell'intero contenuto (*full table scan*), tramite una selezione più mirata ([4];[7]). Lo stesso Robinet struttura in qualche modo i contenuti tra indici e testo, organizzandoli in ordine alfabetico e non sequenziale, offrendoci un repertorio ben preciso di informazioni e selezionando le chiavi d'accesso alle diverse parti del suo complesso lavoro.

Il volume è corredato da sei indici, per ora ci siamo concentrati su 4 di questi: accademie, nomi, temi, indice del volume. Come prima azione abbiamo estrapolato e disaggregato i dati raccogliendoli e organizzandoli in una tabella inserita in un foglio di calcolo intervenendo, arricchendo e completando, là dove possibile, i dati a disposizione. Per esempio, i nominativi presenti nell'indice dei nomi sono stati normalizzati sulla base del VIAF (*Virtual International Authority File*), e collegati tramite un link alla biografia e alla iconografia su pagine di pubblico dominio; abbiamo inoltre corredato i nomi di persona con le date di nascita e morte e incrociato e correlato queste informazioni, là dove disponibili, con le date di incontro con Leibniz e/o con i dati sulla corrispondenza. Per quanto riguarda le città abbiamo registrato la data di arrivo e partenza di Leibniz così come riportata nel testo, abbiamo registrato le coordinate geografiche e le abbiamo localizzate nella mappa. Rispetto alle Accademie abbiamo registrato il luogo, le date di fondazione, la data della visita leibniziana, indicato il link a una scheda descrittiva su siti di pubblico dominio. Abbiamo poi correlato le informazioni sui luoghi, le persone e le accademie con i testi scritti in Italia così come indicato nell'indice di Robinet. Rimane da fare il complesso lavoro di rimando tra i riferimenti bibliografici nel testo e le fonti all'edizione dell'Accademia e, eventualmente, ad altre edizioni o alle traduzioni.

Dall'analisi dei contenuti abbiamo proceduto alla loro formalizzazione, passando dai dati semi strutturati ai dati formalizzati ([7];[9]). La scelta è caduta su strumenti semplici, diffusi, con linguaggi di programmazione standard e open source, in modo da creare un ambiente digitale aperto che fosse di facile mantenimento e consentisse l'economicizzazione delle risorse; strumenti semplici al fine soprattutto di valorizzare e facilitare lo scambio, il riuso e la collaborazione anche con altre importanti piattaforme e data-base leibniziani. Abbiamo usato l'XML per la descrizione dei dati, MySQL come database relazionale, PHP come linguaggio di scripting/programmazione. La relazione tra entità è descritta in RDF al fine anche di pubblicare i dati strutturati e collegati tra loro (*Linked Data*). Le fasi e lo schema di lavoro sono riassunti nella Fig. 1:

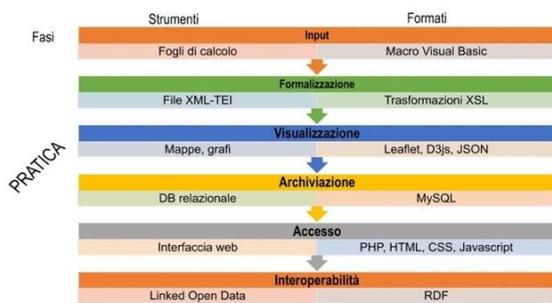


Fig. 1. Schema e fasi di lavoro

Si è voluto anche mettere in luce la multilinearità dei testi che nel loro formato digitale non necessariamente devono essere strutturati in sequenze fisse. Una scelta combinata dei nodi da parte dell'utente consente di superare il limite intrinseco degli elementi ordinati di un indice, riducendo i limiti della lettura di un testo lungo e complesso come quello di Robinet, paradossalmente disorientante. Questa "espansione digitale" delle possibilità del formato convenzionale dell'indice ordinato promette un miglioramento diffuso nel modo in cui le persone accedono a lunghi documenti digitali e apre a una serie di sfide chiave per la loro usabilità. Questa strada esplorata per la prima volta sui testi di Leibniz, a partire da un contributo della letteratura critica, ci sembra particolarmente promettente per un patrimonio così vasto, enciclopedico e complesso come quello del fondo leibniziano.

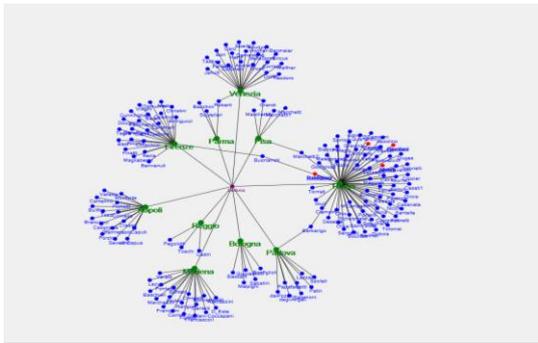
3. TEORIA CON PRATICA: MAPPE E GRAFI

Sulla mappa 'cartacea' di Mattheus Merian, *Nova Italiae Delineatio* (1640) utilizzata per la copertina del volume di Robinet, abbiamo costruito una interfaccia grafica e una mappa interattiva (Fig. 2) in cui l'accesso alle informazioni passa attraverso la rete dei luoghi (le città tappe del viaggio) che aggregano diversi contenuti e informazioni. Il luogo infatti è un nodo che relaziona persone, temi, accademie, riviste scientifiche, le società filosofiche, testi (tutte le informazioni che il testo presenta in modo lineare) ([8];[11]). Si è intesa dunque la raccolta dei dati non solo in termini estrattivi ma anche in termini generativi/euristici e la visualizzazione come una porta d'accesso di cui gli indici hanno costituito la chiave.

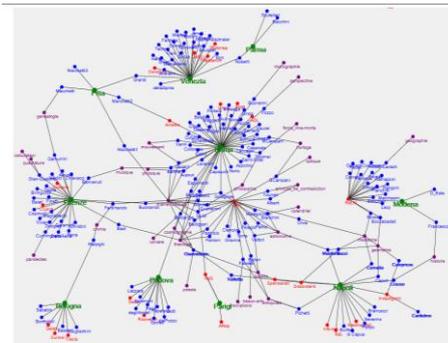


Fig. 2. Mappa interattiva

Tenuto conto che rappresentare e visualizzare attraverso grafi dinamici reticolari costituisce solo una delle possibili modalità di lavoro, ci è sembrato particolarmente fecondo preservare gli indici tradizionali, che offrono già di per sé delle informazioni testuali essenziali. Di questi abbiamo voluto valorizzare la possibilità che offrono di leggere i dati in modalità diverse, anche per esplorare nuovi approcci all'organizzazione dei dati stessi.



Grafo 1. Grafo relazionale Leibniz - Città – Persone



Grafo 2. Rete relazionale del viaggio

L'approccio è guidato da ciò che accomuna le diverse forme di *espressione* nella definizione data da Leibniz stesso. Per il filosofo non è necessario che ciò che si esprime sia simile alla cosa espressa se viene mantenuta anche solo una certa analogia tra le relazioni: qualcosa esprime (rappresenta, è un modello di) qualcos'altro, quando le relazioni tra gli elementi appartenenti al primo dominio corrispondono alle relazioni tra gli elementi appartenenti al secondo. Più l'analogia relazionale è ben strutturata, più cresce il valore euristico della rappresentazione ([3];[12]).

Uno degli aspetti ancora da affrontare è il trattamento linguistico degli indici stilati da Robinet. Si è optato, in questa prima fase, per il mantenimento della lingua originale, il francese, come lingua principale per l'accesso ai contenuti. L'intento è la realizzazione di un ecosistema plurilingue che permetta in primo luogo di rispecchiare la diversità linguistica della produzione scientifica ed epistolare leibniziana e, in secondo luogo, garantisca un accesso plurilingue ai dati e ai testi. Le successive fasi del lavoro a questo riguardo consisteranno dunque: 1. nella individuazione degli equivalenti interlinguistici delle entrate in italiano, tedesco e latino; 2. nel collegamento a ciascuna entrata dei contesti di occorrenza nei testi rilevanti.

4. CONCLUSIONI

Nel contributo ci si è voluti concentrare sulle potenzialità euristiche degli indici come chiavi d'accesso dinamiche ai contenuti di un testo filosofico. Tale tema ha messo a fuoco anche alcuni vantaggi strategici per la navigazione e accesso ai contenuti di un testo lungo e complesso: da un lato offre la possibilità di nuove modellizzazioni ([14]) dei contenuti e delle sue rappresentazioni ridisegnando le mappe del testo e aprendo a nuovi percorsi di navigazione e diverse prospettive di ricerca; dall'altro, come noto, consente di affiancare alla lettura e accesso lineare le potenzialità della lettura digitale e di un approccio al testo multilivello, valorizzandone l'uso ipertestuale, interattivo e multimodale ([2]) e attivando, fra le altre cose, la riscrivibilità dei contenuti del testo stesso.

BIBLIOGRAFIA

- [1] Ciraci, Fabio, Riccardo Fedriga, e Cristina Marras, a cura di. 2021. *Filosofia digitale*. Quaderni di «Filosofia». Milano: Mimesis.
- [2] Ciula, Arianna, Eide Oyvind, Cristina Marras, e Patrick Shale. 2018. «Models and Modelling between Digital and Humanities. A Multidisciplinary Perspective.» *Historical Social Research Supplement* 31.
- [3] Geissler, Nils, e Michela Tardella. 2018. «Observational Drawing. From Words to Diagrams». *Historical Social Research* 31: 209–25.
- [4] Hjørland, B. 2011. «The Importance of Theories of Knowledge: Indexing and Information Retrieval as an Example». *Journal of the American Society for Information Science and Technology* 62 (1): 72–77.
- [5] Marras, Cristina. 2017. *La métaphore chez Leibniz*. Paris: Lambert–Lucas.
- [6] ———. 2021. «Biodiversità ed ecosistema digitale Per una filosofia plurilingue e multiprospettica». In *Filosofia Digitale*, F. Ciraci, R. Fedriga, C. Marras, 17–36. Milano: Mimesis.
- [7] Meyns, Chris. s.d. «Data in the Royal Society's Philosophical Transactions, 1655-1886». *Notes and Records* 1.
- [8] Petrovich, Eugenio. 2020. «Drawing maps with R. A basic tutorial». *Blog DR2 Research Group* (blog). 23 marzo 2020. <https://dr2blog.hcommons.org/2020/03/23/drawing-maps-with-r-a-basic-tutorial/>.
- [9] Rosenberg, Daniel. 2013. «Data before the Fact». In *Raw Data' is an oxymoron*, Lisa Gitelman, 15–40. Cambridge: MIT press.
- [10] Ruecker, Stan, e et al. 2009. «The Table of Contents: A Dynamic Browsing Tool for Digitally Encoded Texts». In *The Charm of a List: From the Sumerians to Computerised Data Processing*, L. Dolezalova, 178–89. Cambridge: Cambridge Scholars Publishing.
- [11] Ruecker, Stan, Milena Radzikowska, e Stéfan Sinclair. 2011. *Visual Interface Design for Digital Cultural Heritage: A Guide to Rich-Prospect Browsing*. Farnham, Surrey: Ashgate Publishing.
- [12] Russo, Ada, e Michela Tardella. 2017. «TheofPhilo. Thesaurus of Philosophy». *ILIESI Digitale*, 1–29.

- [13] Tiberi, M., e F. Mazzocchi. 2007. «La gestione della polisemia nei thesauri: il caso dei termini filosofici». *Bollettino AIB47* (1/2): 93–107.
- [14] Underwood, Ted. 2016. «Distant Reading and Recent Intellectual History». In *Debates in the Digital Humanities*, M. Gold, L. F. Klein. Minneapolis: University of Minnesota Press. <https://dhdebates.gc.cuny.edu/projects/debates-in-the-digital-humanities-2016>.

Ermeneutica digitale del testo filosofico.

Problemi e opportunità

Roberto Evangelista¹, Leonardo Pica Ciamarra²

¹CNR-ISPF – ¹roberto.evangelista@ispf.cnr.it

²CNR-ISPF – ²leonardo.picaciamarra@ispf.cnr.it

PAROLE CHIAVE

Edizioni digitali; testo filosofico; scopo dell'elaborazione digitale.

INTERVENTO

Con questo contributo si intende avviare un ragionamento sul modo in cui stanno cambiando le scienze umane, e in particolare la filosofia, attraverso l'utilizzo sempre più pervasivo delle tecnologie digitali. In particolare, nel caso delle edizioni di opere filosofiche sono evidenti i molti vantaggi delle edizioni digitali e delle edizioni critiche digitali, ma nell'organizzazione e nella preparazione di esse emergono diversi problemi che mettono in discussione il ruolo dell'umanista.

Anche lo scarso finanziamento della ricerca umanistica di base rende attrattiva la conversione del lavoro del ricercatore verso l'utilizzo delle tecnologie digitali. Ma digitalizzare un testo, creare un'edizione critica digitale è un lavoro dispendioso e faticoso, e porta con sé diverse implicazioni. L'introduzione recente di nuovi strumenti adatti a velocizzare o rendere più agevole lo sviluppo delle Digital Scholarly Editions ([18];[20]), è certamente un vantaggio, perché teoricamente libera tempo dal lavoro tecnico, ma rischia lo stesso di imbrigliare il filologo e il filosofo in un lavoro meramente esecutivo di trascrizione. Nonostante i progressi, e nonostante la DSE sia vista come un'opportunità per allargare l'ermeneutica del testo attraverso la riproposizione del metatesto intellettuale di cui fa parte l'autore e la sua recezione, bisogna sempre tenere presente quello che suggeriva McLuhan quando affermava: “we shape our tools, and thereafter our tools shape us”.

Le edizioni digitali pongono alcuni problemi di ordine tecnico: i più comuni sono quelli relativi all'interoperabilità e all'invecchiamento precoce dei formati dei file. Abbiamo avuto modo di valutare questi rischi nel corso di due esperienze relative alla digitalizzazione di alcuni manoscritti inediti di Giambattista Vico e del testo del 1744 della *Scienza nuova*, entrambi materiali preparatori a un'edizione critica digitale delle tre versioni dell'*opus majus* vichiana. In questi casi, abbiamo operato scelte particolari, rinunciando a utilizzare gli strumenti TEI e preferendo concentrarci sulla possibilità di offrire materiali correttamente “interpretati” in un tempo relativamente breve, sfruttando così una ricca esperienza nel classico lavoro di ecdotica. Quello che vorremmo, però, rilevare in questa occasione non è tanto l'ordine delle valutazioni di scelte peculiari (in parte condivise con [8]), quanto il modo in cui cambia il lavoro dell'umanista e i problemi etico-sociali relativi all'utilizzo delle tecnologie.

La logica che permea la gestione della rete, spesso una logica per lo più volta al profitto ([13]), complica il lavoro dell'umanista, perché lo espone sempre al rischio di perdere il controllo del lavoro fatto, o peggio al rischio di trasformarsi in un mero esecutore per mancanza di strumenti liberi. Le enormi potenzialità del web e della tecnologia, in linea di principio, permetterebbero di inserire un'opera filosofica in una vera e propria “realtà aumentata” nella quale un intero contesto storico, culturale, intellettuale e anche paesaggistico e geografico possa essere restituito alla comunità. In questo modo sarebbe possibile non solo arricchire l'ermeneutica del testo filosofico, ma anche restituire all'umanista una funzione nella formazione degli individui e nello sviluppo di una coscienza civile critica. A questa prospettiva si oppone una apparente inversione tra mezzi e fini. A quale fine è inteso tutto il lavoro della digitalizzazione, dell'edizione digitale, dei testi filosofici? Per quale uso è sostenuto, promosso, finanziato? La conversione in *dati* del patrimonio di pensiero del passato sembra configurarsi come un fine di per sé, e non abbastanza si riflette su che cosa comporta – internamente al suo mestiere e in rapporto al suo ruolo sociale – la conversione del filosofo e dell'umanista in strumento per questo fine, né come, d'altro lato, queste trasformazioni impattino sulla stessa lettura del testo. C'è bisogno di un'indagine anche genealogica intorno alle trasformazioni dei modelli e delle politiche della conoscenza in corso nei nostri studi, a partire dalle implicazioni epistemologiche insite nell'acquisizione della stessa categoria delle “humanities”. La questione del digitale, anche proprio sul terreno di ciò che il digitale fa del testo filosofico della tradizione, è intimamente intrecciata con il destino degli studi umanistici.

BIBLIOGRAFIA

- [1] *Advances in Digital Scholarly Editing. Papers presented at the Dixit Conferences in The Hague. Cologne and Antwerp*. 2017. Peter Boot, Anna Cappelotto, Wout Dillen, Franz Fischer, Aodhán Kelly, Andreas Mertgens, Anna-Maria Sichani, Elena Spadini&Dirk van Hulle. Leiden: Sidestone.
- [2] Andrews, Tara. 2013. «The third way. Philology and critical edition in the digital age». *Variants* 10: 61–76.
- [3] Bryant, John. 2002. *The fluid text: A theory of revision and editing for book and screen*. Editorial Theory and Literary Criticism. Ann Arbor: University of Michigan Press.
- [4] Buard, Pierre-Yves. 2015. «Modélisation des sources anciennes et Édition Numérique». Université de Caen. <https://hal.archives-ouvertes.fr/tel-01279385/file/thesePierreYvesBuard.pdf>.
- [5] Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, e Jeffrey Schnapp. 2012. *Digital Humanities*. Cambridge, MA: MIT Press.
- [6] Causer, Tim, Justine Tonra, e Valerie Wallace. 2012. «Transcription maximized; expense minimized? Crowdsourcing and editing the collected works of Jeremy Bentham». *Literary and Linguistic Computing* 27: 19–137.
- [7] Del Vento, Christian, Monica Zanardo, e Thomas Lebarbé. 2016. «Re-thinking Leopardi: Towards the digital edition of the Canti». In *Edizioni Critiche Digitali – DigitalCritical Editions*, Italia Paola, Bonsi Claudia, 63–74. Roma: Sapienza Università Editrice.
- [8] Fiorimonte, Domenico, Teresa Numerico, e Francesca Tomasi. 2015. *The Digital Humanist. A Critical Enquiry*. New York: Punctum Books.
- [9] Guichard, Eric. 2019. «Les humanités numériques n'existent pas». 13 settembre 2019. <http://barthes.enssib.fr/articles/Guichard-Les-humanites-numeriques-n-existent-pas.pdf>.
- [10] «Guidelines for electronic text encoding and interchange». 2018. *TEI Consortium* (blog). 2018. <https://tei-c.org/guidelines/p5/>.
- [11] Hinrichs, Uta, Stefania Forlini, e Bridget Moynihan. 2018. «In defense of sandcastles: Research thinking through visualization in digital humanities». *Digital Scholarship in the Humanities*, 1–20.
- [12] Kirschenbaum, Matthew. 2015. *Track changes: A literary history of word processing*. Cambridge: Harvard University Press.
- [13] Mazzola, Roberto. 2019. «Note su internet e democrazia». *Laboratorio dell'ISPF* 16 (8). http://www.ispf-lab.cnr.it/article/2019_MZR_Abstract.
- [14] Moretti, Franco. 2013. *Distant Reading*. London: Verso.
- [15] Pierazzo, Elena. 2011a. «A Rationale of Digital Documentary Editions». *Literary and Linguistic Computing* 26 (4): 463–77. <https://doi.org/10.1093/llc/fqr033>.
- [16] ———. 2011b. «A rationale of digital documentary editions». *Literary and Linguistic Computing* 26 (4): 463–77.
- [17] ———. 2014. «Unpacking the draft page: A new framework for digital editions of draft manuscripts». *Variants* 11: 29–46.
- [18] ———. 2015. *Digital Scholarly Editing: Theories, Models and Methods*. Routledge. <https://www.routledge.com/Digital-Scholarly-Editing-Theories-Models-and-Methods/Pierazzo/p/book/9780367598600>.
- [19] Pierazzo, Elena, e Matthew James Driscoll, a c. di. 2016. *Digital Scholarly Editing: Theories and Practices*. Digital Humanities Series. Open Book Publishers. <https://doi.org/10.11647/OBP.0095>.
- [20] Pierazzo, Elena, e Elise Leclerc. s.d. «L'edizione scientifica al tempo dell'editoria digitale». *Ecdotica*, 12 12: 180–94.
- [21] Posner, Miriam. 2012. «Think Talk Make Do: Power and the Digital Humanities». *Journal of Digital Humanities* 1 (2).
- [22] Robinson, Peter. 2003. «Where we are with electronic scholarly editions, and where we want to be». *Jahrbuch für Computerphilologie* 5: 23–143.
- [23] ———. 2010. «Electronic editions for everyone». In *Text and genre in reconstruction. Effects of digitalization on ideas, Behaviours, products and institutions*, W. McCarty, 145–63. Cambridge: Open Book Publisher.
- [24] ———. 2016. «Project-based digital humanities and social, digital, and scholarly editions». *Digital Scholarship in the Humanities* 31 (4): 875–89.
- [25] Shillingsburg, Peter L. 2006. *From Gutenberg to Google: Electronic Representations of Literary Texts*. Cambridge: Cambridge University Press.
- [26] Stokes, Peter, e Elena Pierazzo. 2009. «Encoding the language of landscape: XML and databases at the service of Anglo-Saxon lexicography». In *Perspectives on lexicography in Italy and Europe*, S. Bruti, R. Cella, M. Foschi Albert, 203–38. Newcastle: Cambridge Scholars Publishing.
- [27] «Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities» 2020. ALLEA. <https://repository.dri.ie/catalog/tq582c863>.
- [28] Van Zundert, Boris, e Peter Boot. 2012. «The digital edition 2.0 and the digital library: Services, not resources». *Bibliothek and Wissenschaft* 44: 141–52.

Human Enhancement e soggetto Post-Umano alla prova delle DH: come le tecnologie digitali ci trasformano

Camilla Domenella

Università degli Studi di Macerata, Italia – c.domenella@unimc.it

ABSTRACT

Le posizioni intorno al tema dello *Human Enhancement* rintracciano nelle tecnologie GRIN (geno-, robo-, info-, nano-) le condizioni per un potenziamento umano desiderabile in termini di prestazioni e capacità fisiche e cognitive. Analizzando alcune delle prospettive affrontate dalle *Digital Humanities* e prendendo in considerazione le teorie sul Post-Umano, si intende presentare le tecnologie digitali come elementi fondanti di un potenziamento già in atto, capace di rendere conto delle caratteristiche di interrelazione, sincronicità, ibridizzazione dell'umano.

PAROLE CHIAVE

Human Enhancement, Digital Humanities, Post-Umano, tecnologie digitali.

INTERVENTO

1. INTRODUZIONE

Il dibattito intorno al potenziamento umano si concentra – pur senza esaurirsi – intorno alle migliori tecniche e tecnologiche utili alla *modificazione biologica dell'individuo*. Più in particolare, le teorie sullo *Human Enhancement* interpretano un potenziamento umano di ordine fisico e cognitivo, biologico e intellettuale, inglobando le riflessioni intorno ai progressi tecnici e tecnologici compiuti in particolare nei campi della medicina e della bioingegneria. Tali posizioni si concentrano sulla eventuale creazione di cyborg, sulla modificazione genetica, sulla “correzione” cognitiva degli esseri umani oppure sul progressivo allontanamento della vecchiaia e della morte.

Le riflessioni sullo *Human Enhancement* sembrano tuttavia aver trascurato quella forma di potenziamento umano attualmente in atto, realizzato per mezzo delle tecnologie digitali, e che rappresenta oggi un composto trasversale di informazioni, dati, identità, società, culture, capace di dischiudere l'orizzonte verso l'autentico soggetto Post-Umano. Attraverso un'analisi delle prospettive aperte dalle *Digital Humanities* e una disamina delle teorie del Postumanesimo, con la conseguente valorizzazione del *continuum* natura-cultura, si offre qui il principio di una tematizzazione di quel potenziamento (in) digitale cui assistiamo e del quale siamo oggetto ed artefici.

2. PROPOSTA

Le posizioni intorno allo *Human Enhancement* rintracciano nelle tecnologie GRIN (geno-, robo-, info-, nano-) le condizioni preliminari per un potenziamento umano desiderabile. Tuttavia, tali riflessioni non sembrano prendere in considerazione la dimensione digitale, promossa e attuata dalle tecnologie digitali intese come quel complesso di tecnologie dell'informazione, media digitali, strumenti ed espressioni abilitati digitalmente che hanno introdotto e introducono una trasformazione qualitativa in termini di codifica, registrazione, trasmissione, comunicazione di contenuti. Questa trasformazione qualitativa – colta e condotta dalle *Digital Humanities* – investe e ricopre di nuovi significati concetti quotidiani come quelli di intelligenza, socievolezza, informazione, privacy, welfare fino a toccare lo statuto antropologico dell'umano stesso.

Da un punto di vista teoretico, le posizioni intorno al tema dello *Human Enhancement* insistono su una rigida opposizione tra antropocentrismo e biocentrismo trovando la propria eco nelle riflessioni sul Post-Umanesimo.

Se il quadro contemporaneo può essere interpretato sulla base del rapporto tra natura e cultura, la presunta alternativa tra biocentrismo ed antropocentrismo implica gli estremi di una dialettica ulteriore, quella tra naturale e artificiale, fra dato biologico originario e prodotto e produzione umani. *L'homo technologicus* vive al centro di tale dialettica; il Postumanesimo si fa interprete di questo statuto antropologico.

La tesi di fondo della visione postumana insiste sulla relazione tra natura e cultura, affermando tale relazione come un *continuum*, sulla base di una visione monista che si concentra sulla forza autopoietica della materia vivente ([3]). A ciò si lega la dimensione “propria” dell'uomo, inteso come *ibridazione* tra umano e non-umano, frutto di una *partnership* storico-

evoluzionistica fra uomo e animale ([10]) e macchina. In questo senso, il progredire evolutivo dell'*homo technologicus* coincide con il progredire di una ibridazione tecnologica: sullo sfondo del *continuum* natura-cultura, del compenetrarsi fino a confondersi delle dimensioni naturale e artificiale, la questione dell'identità propriamente umana riaffiora, e con essa tutte le implicazioni che l'accompagnano.

Fra queste implicazioni, le teorie sullo *Human Enhancement* sembrano aver trascurato la peculiare forma di potenziamento umano offerta dal digitale e già in atto, organizzata intorno al presente e coerentemente sviluppata nel complesso delle tecnologie digitali, il cui impatto è ben delineato dalle *Digital Humanities*. Si tratta, in questo senso, di un potenziamento immateriale, tuttavia intenzionale, della condizione umana.

Le tecnologie digitali hanno anzitutto dischiuso una nuova *ecologia*, rappresentata dalla globalità dello spazio delle informazioni, in grado di riformulare lo statuto umano. Luciano Floridi ha approfondito questo aspetto definendo l'Infosfera come «il nuovo ambiente in cui gli esseri umani trascorrono gran parte della loro vita» ([7]: 84) ponendo l'accento sul processo di globalizzazione che ha condotto e tuttora pervade l'Infosfera. La contrazione dello spazio fisico e l'espansione dell'ambiente virtuale si toccano in una soglia porosa attraverso cui l'ibridizzazione fra reale e virtuale, fra umano e macchinico avviene e – letteralmente – ha luogo. In questo senso, la globalizzazione dell'Infosfera è una sincronizzazione sull'asse del tempo e una correlazione e interrelazione sull'asse sociale. Io sono ovunque ma non per questo non sono; al contrario, ci riconosciamo in questo ambiente anzitutto come fascio di informazioni.

L'Informatica Umanistica, con le sue declinazioni applicative, ci consente una riflessione più ampia e contemporaneamente più profonda. Le *Digital Humanities* ci fanno riflettere sul fatto che ogni contenuto digitale è anzitutto un contenuto codificato. Questo, per esempio, è particolarmente evidente nella linguistica computazionale, dove la digitalizzazione di un testo passa attraverso la marcatura XML e il *text encoding* e dove tali attività sono sviluppate per condurre ulteriori analisi testuali attraverso sistemi di *text mining* o di *information retrieval*. In altre parole e più in generale, un contenuto codificato diviene un insieme di dati processati, un fascio di informazioni col quale interagire. Ed è proprio la dimensione dell'interazione a rappresentare la cifra del potenziamento già in atto. Oggetti che utilizziamo quotidianamente come carte di credito, ATM o smart pen sono forme di deep text con cui interagiamo e attraverso cui il software diventa co-autore di contenuti. Queste “macchine” tracciano i nostri movimenti e le nostre attività nel tempo e nello spazio, costituendo la nostra stessa identità come consumatori (e non solo). L'interazione non fornisce semplicemente l'accesso al contenuto; essa configura un'azione, assume un significato pragmatico e quindi introduce una nuova dimensione di esistenza. Il ciclo produttivo di oggetti, eventi e azioni è così legato al digitale che la materialità, insieme alla nostra esperienza, non può essere interpretata o addirittura esistere senza esso. La produzione di contenuti per mezzo delle tecnologie digitali si lega dunque alla necessità di prendere consapevolezza di essere autori, di stare costituendo, scrivendola e riscrivendola, la propria identità.

In questo senso, la rappresentazione scritto-codificata dell'individuo, che il digitale ha posto in campo, va oltre i confini del sé materiale pur senza cancellarlo: il flusso di dati digitali ci precede, ci circonda, ci coinvolge, ci costituisce in nuove forme. In questa prospettiva, appare sensato chiedersi *chi sia l'individuo* e non *che cosa sia l'oggetto* digitale.

Alla luce di quanto espresso, le posizioni a sostegno di uno *Human Enhancement* radicale appaiono ridimensionate. Il progresso tecnologico, più che attestare la possibilità di un intervento sempre maggiore sulle qualità o capacità umane, si fa piuttosto testimone di «una coevoluzione tra automazione e ominazione», «dove la condizione umana non è [...] da presupporre come fundamenta, ma da valorizzare sempre come soglia» ([1]: 142), perciò porosa e in trasformazione. Altrettanto, le filosofie dello *Human Enhancement* sembrano non cogliere il potenziamento già attuato dalle tecnologie digitali. La condizione dell'*Onlife* – propria dell'Infosfera – ha introdotto un cambiamento dell'essere e dell'ambiente umano, ridefinendone lo statuto ontologico e la posizione “cosmologica”. Ubiquità, simultaneità, distribuzione frantumata e parallela sono la cifra di un potenziamento che ha traghettato l'umano oltre se stesso e il mondo oltre l'individuo. Si tratta, più strettamente, di imprimere una direzione a questo potenziamento.

Sotto questo punto di vista, la teoria Postumana si stacca dal vortice riflessivo dello *Human Enhancement* e realizza il decentramento dell'umano a favore di un soggetto postumano nomade interrelato, relazionale e polimorfo, capace di divenire – e non di subire – il criterio delle tecnologie che pone in atto.

BIBLIOGRAFIA

- [1] Accoto, Cosimo. 2019. *Il mondo ex machina. Cinque brevi lezioni di filosofia dell'automazione*. Milano: Egea.
- [2] Allegra, Antonio. 2017. *Visioni Transumane. Tecnica, salvezza, ideologie*. Napoli: Ortothes.
- [3] Braidotti, Rosi. 2014. *Il postumano. La vita oltre l'individuo, oltre la specie, oltre la morte*. Roma: DeriveApprodi.
- [4] Cudworth, Erika, e Stephen Hobden. 2021. «Posthuman International Relations: Complexity, Ecology and Global Politics». In *International Relations in the Anthropocene*, Müller F., Rothe D., Chandler D., 233–49. Cham: Palgrave Macmillan.
- [5] Dobson, James E. 2019. *Critical digital humanities: the search for a methodology*. University of Illinois Press.

- [6] Fiormonte, Domenico, Teresa Numerico, e Francesca Tomasi. 2015. *The Digital Humanist: A Critical Inquiry*. New York: Paperback.
- [7] Floridi, Luciano. 2009. *Infosfera. Etica e filosofia nell'età dell'informazione*. Torino: G. Giappichelli.
- [8] Harrasser, Karin. 2018. *Corpi 2.0. Sulla dilatabilità tecnica dell'uomo*. Alessandro Grassi e Tiziano Tanzini. Firenze: goWare.
- [9] Ireni-Saban, Liza, e Maya Sherman. 2021. «Cyborg ethics and regulation: ethical issues of human enhancement». *Science and Public Policy*. <https://doi.org/10.1093/scipol/scab058>.
- [10] Marchesini, Roberto. 2012. *Post-human: verso nuovi modelli di esistenza*. Torino: Bollati Boringhieri.
- [11] Margulies, Jared D., e Brock Bersaglio. 2018. «Furthering post-human political ecologies». *Geoforum* 94: 103–6. <https://doi.org/10.1016/j.geoforum.2018.03.017>.
- [12] Savulescu, Julian, e Nick Bostrom. 2009. *Human Enhancement*. New York: Oxford University Press Inc.

Sessione Testi 3
Grace Murray Hopper

Conservazione e fruizione di banche dati letterarie: l'archivio della poesia italiana dell'Otto/Novecento di Giuseppe Savoca

Antonio Di Silvestro¹, Christian D'Agata², Giuseppe Palazzolo³, Pietro Sichera⁴

¹Università degli Studi di Catania - silvanto [at] unict.it

²Università degli Studi di Catania - christian.dagata [at] studium.unict.it

³Università degli Studi di Catania - giuseppe.palazzolo [at] unict.it

⁴CNR-ISTC - pietero.sichera [at] istc.cnr.it

ABSTRACT

In un periodo nel quale si va sempre più verso l'interoperabilità delle banche dati, in particolare di quelle letterarie, si pone il problema di ovviare al rischio di obsolescenza degli archivi testuali dei decenni precedenti, assicurando non solo la corretta migrazione dei dati nelle nuove piattaforme ospitanti, ma anche la loro persistenza digitale. È una questione che interessa l'Archivio della poesia italiana dell'Otto/Novecento, ideato negli anni '80 da Giuseppe Savoca presso l'Università di Catania, a cui l'interazione tra gli elementi lemmatizzati e la loro codifica in XML-TEI offrirà continuità e nuove potenzialità di consultazione e interrogazione.

PAROLE CHIAVE

FAIR principles, lemmatizzazione concordanziale, vocabolario poetico italiano, data retrieval, archivi digitali.

INTERVENTO

1. INTRODUZIONE

In *Se una notte d'inverno un viaggiatore* ([4]) Lotaria dà una delle prime descrizioni letterarie dell'applicazione di processi elettronici alle pratiche di critica del testo. «La lettura elettronica mi fornisce una lista delle frequenze, che mi basta scorrere per farmi un'idea dei problemi che il libro propone al mio studio critico». Tra i modelli di letture e di misletture che l'autore si diverte a censire, quello presentato da Lotaria è la parodia dell'uso delle concordanze nella critica letteraria. Il romanzo di Calvino viene pubblicato nel 1979 e cita espressamente gli *Spogli elettronici dell'italiano letterario contemporaneo* ([2]). Negli stessi anni, a Catania, Giuseppe Savoca elabora programmi e metodi informatici di spoglio lessicale, con il supporto di un software proprietario. Lo studioso riesce a delegare a un sistema IBM/AS400 un'operazione altamente artigianale quale la realizzazione di una concordanza lemmatizzata, in un periodo nel quale, con l'eccezione di alcune pregiate concordanze cartacee (ad es. quella del *Canzoniere* petrarchesco realizzata dall'Accademia della Crusca ([11])) e dell'opera fondamentale di Padre Busa ([3]) a circolare erano soprattutto formari bruti, con pochissimi dati o inferenze di tipo statistico, di non secondaria importanza nell'interpretazione dei testi letterari.

Il processo computazionale che sta alla base delle concordanze lemmatizzate prodotte da Savoca e dalla sua équipe di ricerca ha prodotto finora degli output cartacei dotati di esaustività (le concordanze comprendono infatti anche i lemmi sinsemantici – articoli, preposizioni, congiunzioni) e di strumenti di studio statistico assai ricchi e diversificati¹.

L'esito più prestigioso di questa operazione editoriale è confluito nel *Vocabolario della poesia italiana del Novecento* [8]. Tuttavia, le stampe cartacee non riuscivano a tenere il passo con l'incremento quantitativo e qualitativo dell'archivio, dove intanto confluivano anche i poeti 'minori', nonché i corpora costituiti dalle traduzioni poetiche. Inoltre, l'ingresso di autori del Sette/Ottocento (Parini, Foscolo, Manzoni, il Leopardi 'puerile') e di una nuova concordanza del *Canzoniere* petrarchesco ([10]) rendevano l'archivio elettronico un paradigma esemplare della storia della poesia italiana da Petrarca al Novecento. Un simile patrimonio necessita di essere preservato e messo a disposizione della comunità scientifica nazionale e internazionale. Esigenze di calcolo, modalità di visualizzazione innovative e aggiornamento delle metodologie sia informatiche quanto linguistiche consentono, e in un certo senso obbligano, la migrazione di questi dati verso sistemi diversi, adottando anche un punto di vista computazionale moderno e portando il progetto verso la riusabilità e l'interoperabilità. È il nucleo dei principi FAIR ([13]): l'obiettivo non è più la semplice gestione corretta dei dati (da un

¹ Tutte le più significative concordanze dei poeti dell'Otto/Novecento (da Leopardi a Montale, da D'Annunzio a Pavese) sono state ospitate nella prestigiosa collana "Strumenti di Lessicografia Letteraria Italiana" dell'editore Olschki, che ha finora stampato quasi trenta volumi di concordanze e studi teorici ad esse collegati. Si ricordano in particolare la prima su Gozzano ([7]) e il volume teorico e metodologico *Lessicografia letteraria e metodo concordanziale* ([9]).

punto di vista tanto teorico quanto pratico, informatico quanto lessicografico), ma l'integrazione e il riutilizzo della conoscenza dopo la pubblicazione dei dati stessi. In tale contesto *Findability*, *Accessibility*, *Interoperability*, e *Reusability* diventano aspetti cardine nella gestione del processo di migrazione, la cui analisi e le cui ricadute scientifiche e tecnologiche sono oggetto di questo articolo.

2. LA CONCORDANZA NELL'ARCHIVIO DELLA POESIA ITALIANA DI G. SAVOCA: PREMESSE TEORICHE E FUNZIONAMENTO

Negli anni '80 i programmi esistenti in commercio si basavano essenzialmente, se non esclusivamente, sulla semplice messa in ordine alfabetico delle forme brute di un testo²; non erano in grado di eseguire una corretta lemmatizzazione, in quanto non distinguevano le forme omografe, né, tanto meno, prevedevano la classificazione grammaticale dei lemmi o la generazione automatica di indici statistici. Le concordanze realizzate secondo il modello Savoca, in quanto finalizzate alla lettura paradigmatica di un testo e/o allo studio del sistema linguistico di un autore, forniscono elementi oggettivi, formalizzati e quantificati, relativi a tutto il lessico, comprese le parole cosiddette funzionali – articoli, congiunzioni, preposizioni –, che occupano circa il 30% della lingua di qualunque testo.

I programmi di lemmatizzazione ideati da Savoca lavorano sulla piattaforma IBM AS/400³, uno dei pochi sistemi che negli anni '80 poteva essere in grado di elaborare la quantità di informazioni necessaria al lavoro di catalogazione. In *Figura 1* viene mostrata il menu di scelta programmi del sistema Savoca su AS/400.



Figura 1. Menù principale del software concordanziale su AS400

Secondo lo schema di Savoca, il programma importa il testo da analizzare attribuendo ad ogni sua riga un codice univoco a dimensione fissa. Prendendo come esempio la riga del “File versi” (riprodotto in *Figura 2*) “I corpi di Clitennestra e di Egisto”, si vede come alla riga vengono attribuiti un’etichetta “V” (indicante il tipo di riga, verso in questo caso), e un codice composto da “PA” (due lettere indicanti il codice dell’autore, Pasolini), “AA” (codice attribuito all’opera, *Pilade*) e un numero di otto cifre (la posizione del verso all’interno dell’opera).

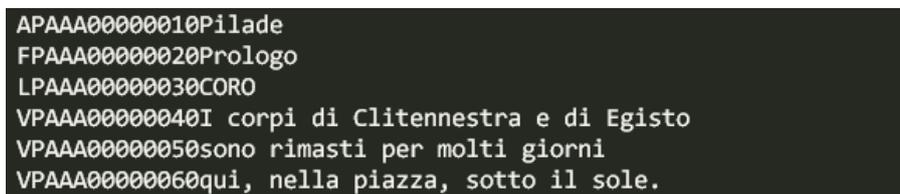


Figura 2. Esempio di “File versi” su AS400

Grazie a questa struttura tutte le righe da concordare risultano univocamente determinate.

A partire dal “File versi” viene generato un “File lemmi”, in ordine sequenziale o alfabetico. In *Figura 3* è possibile vedere il “File lemmi” relativo al “File versi” di *Figura 2*. Anche qui ogni riga risulta strutturata. Prendendo come esempio la riga relativa all’occorrenza “Clitennestra”, si vede come tale riga è composta da “V” (il codice indicante il tipo della riga nel “File versi”), “Clitennestra” (il lemma, l’entrata lessicografica da cui deriva l’occorrenza), “Clitennestra” (l’occorrenza,

² Si vedano a tal proposito, a titolo esemplificativo, gli output generati dal software “LIZ 3.0 - A cura di Pasquale Stoppelli ed Eugenio Picchi - 1998”

³ Con IBM AS/400 (Application Server/400) si intende quella famiglia di computer *midrange* progettati da IBM alla fine del secolo scorso che consentissero di ottenere alte prestazioni di calcolo a prezzi relativamente contenuti.

cioè la parola come è scritta nel testo), “np” (il codice della categoria grammaticale del lemma, nome proprio in questo caso), “PAAA0000040” (codice della riga come presente nel “File versi”), “000007” (contatore progressivo delle occorrenze nel file). Si noti quindi come sia il “File versi” che il “File lemmi” siano posizionali, impongano cioè l’uso di un certo numero di caratteri per ogni campo.

FPrologo	Prologo	smPAAA00000020000002
Lcoro	CORO	smPAAA00000030000003
Vi	i	arPAAA00000040000004
Vcorpo	corpi	smPAAA00000040000005
Vdi	di	prPAAA00000040000006
VClitennestra	Clitennestra	npPAAA00000040000007
Ve	e	coPAAA00000040000008
Vdi	di	prPAAA00000040000009
VEgisto	Egisto	npPAAA00000040000010
Vessere (aus.)	sono	vePAAA00000050000011
Vrimanere	rimasti	vePAAA00000050000012
Vper	per	prPAAA00000050000013
Vmolto	molti	inPAAA00000050000014
Vgiorno	giorni	smPAAA00000050000015
Vqui	qui	avPAAA00000060000016
Vin	nella	prPAAA00000060000017
Vpiazza	piazza	sfPAAA00000060000018

Figura 3. Esempio di “File lemmi” su AS/400

Il software dispone di un dizionario di macchina composto da circa 8.000 forme che gli consentono di avere una sensibilità del 70%, una precisione media del 90% e quindi un punteggio F1 di 0,78⁴. Si rende dunque necessario l’intervento umano per correggere e completare il processo concordanziale, operazione sicuramente non banale. Quantunque infatti la prima regola lessicografica del concordatore sia quella di adeguarsi al modello adottato dai dizionari (il verbo all’infinito, i sostantivi al singolare, il superlativo degli aggettivi ricondotti al grado positivo, ecc.), egli, tuttavia, si trova a dover elaborare un dizionario del tutto speciale perché limitato a un’opera (o a un corpus), con particolarità linguistiche, testuali e grafiche che non sempre possono trovare riscontro nell’uso medio attestato da un dizionario. Per di più, in un’opera poetica può esserci tutto un corredo di notizie, riferimenti, nomi propri, titoli, citazioni, ecc., che non rientrano fra i lemmi accolti nei dizionari, e che perciò pongono al concordatore problemi di lemmatizzazione a volte molto complessi.

Portata a termine la fase non automatizzata del processo, dall’unione del “File versi” e del “File lemmi” viene generata la concordanza vera propria.

Oltre alla realizzazione della concordanza, l’AS/400 consente la produzione di altri utili strumenti lessicografici di consultazione, come liste di frequenza (lemmi o forme in ordine alfabetico, lemmi o forme per numero di occorrenze in ordine crescente o decrescente, lemmi ordinati per categoria grammaticale, retrograde di forme e/o lemmi, utili ai fini dello studio delle rime), liste di lemmi specifici, liste di lemmi comuni a più opere e a più autori, quadri statistici (totale dei lemmi, delle forme e delle occorrenze del testo concordato, frequenza assoluta e frequenza relativa delle categorie grammaticali, ecc.).

Il numero di concordanze create con questo sistema dagli anni ’80 a oggi è notevole. L’archivio contiene oltre 1.400 file tra “File versi”, “File lemmi” e concordanze generate, per un totale di diversi milioni di record.

Si evince da questa descrizione che l’archivio dati descritto risulta di tipo *flat-file*, in cui la formattazione dei dati risulta essenziale, limitando però notevolmente l’interoperabilità e la riusabilità degli stessi. Inoltre la presenza sia nel “File versi” che nel “File lemmi” del campo “Tipo verso” contravviene alle regole sulla normalizzazione ([1]) dei database⁵. Diventa quindi essenziale ripensare questo sistema concordanziale sia come progettazione che come modalità di fruizione.

⁴ Con *precisione e sensibilità* (*precision e recall*) si intendono due indicatori per valutare la qualità di un modello decisionale. Nel modello preso in esame, la precisione indica il rapporto tra i casi corretti e quelli che il modello ha riconosciuto, la sensibilità indica il rapporto tra i casi corretti rispetto all’intero *corpus*. La media armonica di questi due indicatori ($2*((P*S)/(P+S))$) viene chiamata punteggio F1 ed è un metro ampiamente utilizzato per valutare complessivamente i sistemi di apprendimento automatico.

⁵ Nella teoria dei database le forme normali sono delle regole teoriche che garantiscono la corretta progettazione di una base di dati relazionale, evitando episodi di ridondanza e incoerenza.

3. LA CONCORDANZA SU WINDOWS, MACOSX, WEB: LIOTROCONCORD

Da alcuni anni il CINUM⁶ ha implementato un nuovo software per la generazione delle concordanze, chiamato LiotroConcord (versione 2). Il software, sviluppato a partire dal modello concordanziale di Savoca, è un database relazionale puro, creato utilizzando la piattaforma 4D.

4D⁷ è una piattaforma RAD⁸ per DBMS⁹ che integra al suo interno un motore per database SQL¹⁰, un linguaggio di programmazione proprietario e un server web: un modello LAMP¹¹ riunito in un unico strumento.

A queste caratteristiche se ne aggiungono varie altre come il supporto nativo SVG¹², i comandi DOM e SAX per l'analisi e la scrittura di file XML¹³, un interprete PHP¹⁴.

La creazione di un nuovo software si è resa necessaria per vari motivi. Ad esempio le esperienze maturate permettono di avere un progetto tecnicamente più avanzato, l'utilizzo di personal computer Windows o MacOSX permette di creare interfacce utente moderne e *user-friendly*, facilitando l'usabilità e l'interoperabilità con altri dati e altri software.

LiotroConcord è stato progettato seguendo la teoria delle forme normali ([12]), in modo da garantire correttezza di sviluppo e facilità di mantenimento. Il software è utilizzabile indifferentemente e contemporaneamente sia su sistemi Windows che MacOS, nonché via browser web.

Al centro di LiotroConcord viene posta la parola nelle sue varie declinazioni: occorrenza, forma e lemma. Tutte le concordanze generate dal sistema sono strettamente connesse, indipendentemente dall'autore o dal genere (prosa, poesia, teatro). La generazione della concordanza viene resa quanto più possibile automatica passando attraverso una metodologia ormai consolidata. L'utente, attraverso un editor di testi esterno, compone il corpus da importare nel sistema. Il file così generato (.txt o .docx) viene sottoposto da LiotroConcord ad un'analisi preventiva del file secondo la codifica UTF-8¹⁵. Preparato il testo, LiotroConcord lo analizza, lemmatizzando correttamente la quasi totalità del testo, grazie ad un dizionario di macchina di quasi 300.000 forme, disambiguando inoltre le forme omografe¹⁶. Il software utilizza un sistema di *machine learning* ad apprendimento con rinforzo: la correzione manuale della concordanza realizzata automaticamente fornisce dei feedback che gli consentono di imparare a correggere i propri errori. Il sistema di riconoscimento utilizza varie tecniche, come ad esempio la valutazione degli "intorni" (l'occorrenza o i gruppi di occorrenze che precedono/seguono la forma omografa) o la presenza di imitazioni, utilizzo cioè della stessa sequenza di occorrenze/lemmi/categorie grammaticali nelle opere dello stesso autore o nell'intero database. LiotroConcord acquisisce ed impara: attualmente presenta punte negli indici di precisione e sensibilità del 98% con un indice F1 che si attesta sempre a 0,97 (valori nettamente più elevati rispetto al sistema su AS/400).

Per il controllo e la correzione della concordanza, LiotroConcord offre un editor a interfaccia grafica appositamente progettato per rendere veloce ed accurato il processo di controllo, anche quando questo viene svolto contemporaneamente da più utenti.

⁶ Il CINUM (Centro di Informatica Umanistica) è un centro interdipartimentale istituito presso l'Università di Catania nel 2017. Riunisce studiosi di vari ambiti disciplinari (dalla linguistica alla filosofia del linguaggio, dalla filologia alla sociologia della comunicazione), e attualmente è impegnato, oltre che nell'Edizione Digitale dell'Opera Omnia di Luigi Pirandello (www.pirandellonazionale.it), in diversi progetti di ricerca su fondi di Ateneo riguardanti sia la messa in rete dell'archivio della poesia italiana dell'Otto/Novecento, realizzato da Giuseppe Savoca e dalla sua équipe di ricerca a partire dagli anni '80, sia la realizzazione di edizioni-archivio interrogabili della narrativa e poesia italiana del Novecento.

⁷ La piattaforma 4D è sviluppata e mantenuta dall'omonima società francese 4D SAS. Per un elenco di caratteristiche si veda <https://us.4d.com/new-4d-v19lts>.

⁸ RAD (Rapid Application Development) è una metodologia di sviluppo informatico basata sulla costruzione di prototipi, modelli di sviluppo iterativi e utilizzo di interfacce grafiche.

⁹ Con DBMS (Database Management System) si intende un software per la realizzazione di database. Infatti, alla realizzazione e interrogazione della base dati, aggiunge altre caratteristiche, come linguaggi di programmazione e gestori dell'interfaccia.

¹⁰ SQL (Structured Query Language) è un linguaggio per database inizialmente sviluppato da Donald D. Chamberlin e Raymond F. Boyce ([5]) basato sul modello relazionale introdotto da Edgar F. Codd.

¹¹ LAMP è l'acronimo di 'Linux, Apache, MySQL, PHP', uno dei tipici ambienti per lo sviluppo di applicazioni web, dove Linux è il sistema operativo, Apache è il software che funge da web server, MySQL è il motore del database e PHP il linguaggio di programmazione.

¹² SVG (Scalable Vector Graphics) è un formato che supporta oggetti di grafica vettoriale.

¹³ XML (eXtensible Markup Language) è un linguaggio di markup che consente la codifica di documenti in un formato leggibile sia dall'uomo che da una macchina. I documenti XML sono strutturati e DOM (Document Object Model) e SAX (Simple API for Xml) sono linguaggi usati per accedere a tale struttura.

¹⁴ PHP (PHP Hypertext preprocessor) è un linguaggio di scripting per pagine web dinamiche.

¹⁵ UTF-8 è un sistema di codifica per i caratteri Unicode. Mentre ASCII permette di codificare solo 128 caratteri e le codifiche ISO-8859 alcuni set di caratteri stampabili, UTF-8 amplia le possibilità di codifica facendo uso di codici fino a 4 byte. Lo standard è reperibile al link <https://datatracker.ietf.org/doc/html/rfc3629>

¹⁶ Forme omografe come ad esempio 'sale' (sostantivo maschile singolare, sostantivo femminile plurale o voce verbale) oppure 'che' (congiunzione o pronome relativo).

Il sistema è in grado di effettuare le analisi linguistiche e generare gli output già presenti nel sistema Savoca (concordanza, liste di frequenza, ecc.), nonché altre analisi come l'iperconcordanza di più corpora, l'intersezione lessicale (per lemma e per forma tra parti dello stesso corpus o tra corpora differenti), la valutazione delle co-occorrenze.

È stato sviluppato anche un sistema di interrogazione tramite pagine web¹⁷ attraverso cui è stato possibile pubblicare testi lemmatizzati prodotti dall'équipe di ricerca di Giuseppe Savoca e successivamente dall'équipe del CINUM. Essendo 4D, come detto in precedenza, un ambiente integrato, non è necessario trasferire i dati dall'ambiente di concordanza ad interfaccia utente all'ambiente web: i dati risultano sempre già pronti per essere pubblicati sul web e ogni modifica apportata attraverso l'interfaccia utente è immediatamente visibile nel sistema di interrogazione.

4. PROCESSO DI TRASFORMAZIONE

La migrazione delle concordanze generate su AS/400 verso il nuovo software LiotroConcord è stata svolta analizzando e risolvendo varie problematiche. Innanzitutto la codifica dei caratteri dei vecchi file risultava non standard. Per questo motivo si è dovuto ricorrere ad un'analisi puntuale dei file esportati dall'AS/400 correggendo le difformità che si evidenziavano, soprattutto per quanto riguardava i caratteri con codice ASCII superiore a 128 (le lettere accentate).

I "File lemmi" e i "File versi" necessitavano di una conversione dalla codifica proprietaria del sistema AS/400 a quella UTF-8. Per effettuare questo passo, si è utilizzato il vocabolario di macchina di LiotroConcord che, analizzando le parole sconosciute (ad esempio "perch?" al posto di "perché"), è riuscito ad effettuare in automatico l'aggiornamento della codifica.

Codificati i due file in maniera corretta, si è proceduto alla loro importazione combinata. Il "File versi" è stato importato tenendo traccia del codice attribuito al verso dal sistema AS/400. Di seguito sono state cercate all'interno del "File lemmi" le forme presenti in un certo verso; attraverso un algoritmo di *pattern matching*¹⁸ sono state confrontate le possibili occorrenze del verso importato con le occorrenze che il "File lemmi" attribuiva a quel verso. Il processo si è rivelato corretto in maniera automatica al 99,9%¹⁹.

È stato così possibile non solo attribuire le forme ai versi relativi, ma anche riscontrarne l'effettiva posizione, in modo da poter utilizzare le funzioni di LiotroConcord che analizzano le co-occorrenze. Il processo si è rivelato esaustivo, essendo infatti in grado di trasferire integralmente le concordanze dal vecchio al nuovo sistema.

5. RICADUTE SCIENTIFICHE. UN MODELLO PER IL FUTURO

L'obiettivo principale del processo di trasformazione è stato quello di incanalare le concordanze prodotte da Savoca all'interno dei principi FAIR: evitare l'obsolescenza digitale delle concordanze (dati reperibili e riutilizzabili) e garantirne l'accessibilità e l'interoperabilità. A tale scopo è attivo presso il CINUM il progetto ALPION²⁰, che si pone come obiettivo quello della trasmigrazione dei dati in un sistema interoperabile, permettendo un'interrogazione del corpus da parte degli studiosi. Il problema di fondo è tuttora al centro del dibattito nella comunità scientifica, ma certamente un modo per superarlo è quello di condividere i contenuti in formati aperti come XML-TEI²¹. La concordanza diventa così un prodotto che, come nella menzionata Edizione Digitale dell'Opera omnia di Pirandello²², è liberamente consultabile sia attraverso un form d'interrogazione sul web, sia attraverso un pdf scaricabile ([6]). La *Figura 4* mostra uno *specimen* dell'output in XML-TEI generato dalla ricerca del lemma «abate» all'interno dell'*Enrico IV* di Luigi Pirandello:

¹⁷ Il sistema di interrogazione è stato implementato per l'Edizione Nazionale dell'Opera omnia di Luigi Pirandello (<https://vocabolari.pirandellonazionale.it/>) e per il progetto ALPION (<https://alpion.unict.it/vocabolario/ricerca/>).

¹⁸ È stata implementata una variante del metodo ad indici che, basandosi sul *preprocessing* del testo, risulta molto performante. Dati m la lunghezza del pattern e Z il numero di occorrenze di un certo pattern, l'algoritmo impiega $O(m+Z)$ per trovare tutte le occorrenze.

¹⁹ Si è scelto volutamente di marcare come "Da controllare" le occorrenze omografe all'interno di un verso, quando tali occorrenze provenivano da lemmi differenti.

²⁰ <https://alpion.unict.it/>

²¹ TEI Consortium. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.3.0. Last updated on 31st August 2021. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>.

²² <https://www.pirandellonazionale.it/>

```

<entry n="3">
  <form>
    <orth>abate</orth>
  </form>
  <gramGrp>
    <pos>s</pos>
    <gen>f</gen>
  </gramGrp>
  <note>
    <num type="occorrenza_assoluta">6</num>
    <num type="occorrenza_relativa">0,030604</num>
  </note>
  <cit type="KWIC">
    <quote n="EN-AI-0089-7">ARIALDO Il signore vuol dire l'<term rend="bold">Abate</term>? Benissimo: Ugo di Cluny</quote>
    <quote n="EN-AI-0093-7">di Cluny che accompagna l'<term rend="bold">Abate</term>. </quote>
    <quote n="EN-AII-0106-8">duchessa Adelaide e dell'<term rend="bold">Abate</term> di Cluny. Trovandosi davanti</quote>
    <quote n="EN-AII-0114-7">Marchesa Matilde e dell'<term rend="bold">Abate</term> di Cluny. E io vi so dire, </quote>
    <quote n="EN-AII-0116-14">Dottore con la tonaca di <term rend="bold">Abate</term> di Cluny; Enrico IV è fra</quote>
    <quote n="EN-AII-0128-6">mascherata...E lui da padre <term rend="bold">abate</term>...-E mi portano con loro</quote>
  </cit>
</entry>

```

Figura 4. Risultato della ricerca del lemma «abate» all'interno dell'Enrico IV di Luigi Pirandello

In questo modo la concordanza diventa facilmente importabile e riutilizzabile in qualunque altro sistema per basi di dati. Avendo portato a termine la realizzazione del sistema di migrazione dei dati dal sistema AS/400 ed essendo già in fase avanzata l'importazione dei dati stessi verso il software LiotroConcord, uno degli obiettivi che ALPION si pone adesso è quindi quello di permettere il libero accesso agli studiosi attraverso un sistema di interrogazione dei dati su interfaccia web e l'esportazione dei risultati in file XML-TEI (elaborati direttamente da LiotroConcord) che rispettino le informazioni strutturate della concordanza.

Un ulteriore possibile output è quello della lemmatizzazione integrale del testo, laddove – è bene chiarirlo – siano già scaduti i diritti d'autore e sia possibile rendere disponibile alla comunità scientifica il testo per intero. Il file così condiviso sarebbe quello della lemmatizzazione, rendendola in tal modo integralmente interoperabile. A mo' di esempio mostriamo in Figura 5 il possibile output di un verso de *Le ceneri di Gramsci* di Pier Paolo Pasolini.

```

<>
  <w xml:id="CG-CG-156-1" pos="conj" lemma="ma">ma</w>
  <w xml:id="CG-CG-156-2" pos="prep" lemma="a">a</w>
  <w xml:id="CG-CG-156-3" pos="ie" lemma="che_2">che</w>
  <w xml:id="CG-CG-156-4" pos="v" lemma="servire" >serve</w>
  <w xml:id="CG-CG-156-5" pos="art" lemma="la">la</w>
  <w xml:id="CG-CG-156-6" pos="sf" lemma="luce">luce</w>
  <pc type="interrogative">?</pc>
</>

```

Figura 5. Un possibile output della concordanza de *Le ceneri di Gramsci* di Pier Paolo Pasolini in XML-TEI

In questo caso a essere reso disponibile non sarà tanto il prodotto finale, ma la lemmatizzazione *in fieri*, che può essere modificata, integrata, revisionata da qualunque studioso.

BIBLIOGRAFIA

- [1] Albano, Antonio, Giorgio Ghelli, e Renzo Orsini. 1997. *Basi di dati relazionali e a oggetti*. Bologna: Zanichelli.
- [2] Alinei, Mario. 1978. *Spogli elettronici dell'italiano letterario contemporaneo*. 3 voll. Bologna: Il Mulino.
- [3] Busa, Roberto. 1974. *Index Thomisticus Sancti Thomae Aquinatis Operum Omnium Indices et concordantiae*. Vol. 56. Stuttgart: Frommann-Holzboog.
- [4] Calvino, Italo. 1994. *Se una notte d'inverno un viaggiatore*. Vol. 2. Romanzi e racconti. Milano: Mondadori.
- [5] Chamberlin, Donald D., e Raymond F. Boyce. 1974. «SEQUEL: A Structured English Query Language». In *ACM SIGFIDET*.
- [6] Giuffrida, Milena, Christian D'Agata, Laura Giurdanella, e Pietro Sichera. 2021. «Pirandello Nazionale: per un nuovo modello di edizione digitale, collaborativa e integrata». In *Boschetti, Federico; Del Grosso, Angelo Mario; Salvatori, Enrica; (edd.); AIUCD 2021 - DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale. Raccolta degli abstract estesi della 10a conferenza nazionale*. Pisa.
- [7] Savoca, Giuseppe. 1984. *Concordanza di tutte le poesie di Guido Gozzano*. Firenze: Olschki.
- [8] ———. 1995. *Vocabolario della poesia italiana del Novecento*. Bologna: Zanichelli.
- [9] ———. 2000. *Lessicografia letteraria e metodo concordanziale*. Firenze: Olschki.

- [10] Savoca, Giuseppe, e Bartolo Calderone. 2011. *Concordanza del Canzoniere di Francesco Petrarca*. 2 voll. Firenze: Olschki.
- [11] Ufficio lessicografico dell'Opera del Vocabolario. 1971. *Concordanza del Canzoniere di Francesco Petrarca*. 2 voll. Firenze: Firenze: Accademia della crusca.
- [12] Ullman, Jeffrey D. 1988. *Principles of Database and Knowledge - Base System*. USA: Computer Science Press Inc.
- [13] Wilkinson, Mark D., Michel Dumontier, et al. 2016. «The FAIR Guiding Principles for scientific data management and stewardship». *Sci Data* 3 160018.

«Le varianti della rosa». Per un prototipo di edizione digitale del *Nome della rosa*: interpretazione, didattica, annotazione

Christian D'Agata

Università degli Studi di Catania - christian.dagata@gmail.com

ABSTRACT

Le varianti della rosa presenta un prototipo di edizione digitale per un pubblico non strettamente specialistico a cui offrire alcuni studi di filologia digitale sul *Nome della rosa* di Umberto Eco. Il contributo propone una riflessione sul paradigma 'IDEA' (Interpretazione, Didattica, Edizione, Annotazione) che fa da cornice allo sviluppo della presente edizione integrata ed estesa. La concreta realizzazione di questo modello si manifesta in un portale con l'edizione delle varianti tramite EVT2, con contenuti didattici proposti tramite TRAViz, Storymap JS e Timeline JS, e con un'annotazione sviluppata su Euporia. Il contributo presenta infine una *road map* dei contenuti futuri in un'ottica di sempre maggiore integrazione tra ricerca scientifica e dimensione *public*.

PAROLE CHIAVE

Digital Scholarly Edition, EVT e codifica dei testi, Euporia e annotazione DSL, Public Humanities, Digital Hermeneutics.

INTERVENTO

1. INTRODUZIONE

Nel 2012 Umberto Eco pubblicava un'edizione riveduta e corretta de *Il nome della rosa* ([9]), provocando un'aspra polemica alimentata da quotidiani nazionali e internazionali che avevano tacciato questa riscrittura di semplificazione per *minus habens*. *Le varianti della rosa* (www.variantidellarosa.it) nasceva nel 2019 come progetto di sintesi tra studi filologici tradizionali e *Digital humanities* per indagare le varianti d'autore del romanzo di Eco¹. La critica contemporanea si era infatti limitata a descrivere alcuni fenomeni di sinonimia, riduzione di latinismi, alleggerimento di elenchi e di riferimenti eruditi ([17]). In realtà, gli interventi correttori investivano circa 6500 parole e, pur nella loro asistematicità, sembravano rappresentare una spinta alla semplificazione, sebbene limitata ([7]). *Le varianti della rosa* offriva in origine un portale dove presentare agli studiosi i risultati *in itinere* della ricerca. La presente proposta intende invece andare oltre, ipotizzando un modello di edizione scientifica digitale per un'opera del tardo Novecento che faccia dell'integrazione tra annotazione, *distant reading*, concordanza, codifica XML-TEI ([21]), filologia digitale, e *data visualization* il suo paradigma, proponendo un modo integrato ([11]) di studiare e fruire l'opera letteraria, in particolare quelle ancora oggetto di diritto d'autore.

2. IL PARADIGMA 'IDEA'

Il Paradigma IDEA (Interpretazione, Didattica, Edizione, Annotazione) intende mettere in luce la centralità dell'Interpretazione come atto fondante di ogni operazione critica (sia tradizionale che digitale): dalle domande di ricerca che devono guidare il processo di elaborazione dell'edizione scientifica digitale alla costruzione di contenuti insieme scientifici e *public* per coinvolgere un pubblico ampio e variegato. D'altronde come dice Sahle: «Scholarly digital editions are scholarly editions that are guided by a digital paradigm in their theory, method and practice» ([19]). La specificità del paradigma IDEA non sta solo nell'elaborazione e fruizione di contenuti digitali, ma nella stessa progettazione che tiene conto delle specificità del digitale: dalla possibilità di presentare grandi quantità di dati, alla necessaria multimedialità, fino all'*user interaction* ([4]); specificità che investono le coordinate del tempo (testo diacronico), dello spazio (tutti i testi), della forma (il testo iconico) e del *social* (il testo wiki) ([3]).

¹ Il progetto è iniziato nel marzo 2019 all'interno del Master in *Digital humanities* della Ca' Foscari in collaborazione col CNR-ILC. Nel gennaio 2020 è stata pubblicata la prima versione del portale-edizione focalizzata sulla presentazione del progetto e su una selezione di varianti con EVT. Con la collaborazione del CINUM è stato sviluppato il paradigma IDEA e sono stati implementati contenuti di didattica digitale e di *public humanities* nel corso del 2021. Si prevede di pubblicare una seconda versione del portale alla fine del 2022 con l'edizione critica integrale delle varianti, un database delle varianti, l'annotazione e nuovi contenuti didattici e multimediali.

Le quattro diverse fasi del paradigma IDEA non devono però essere viste come consequenziali, ma in un'ottica rizomatica e, in un certo senso, ciclica, dove ogni fase presuppone le altre e le completa. Ad esempio, l'interpretazione fonda l'annotazione e poi i risultati di quest'ultima fanno emergere una nuova interpretazione, riprendendo in chiave digitale l'idea del circolo ermeneutico come atto fondamentale della comprensione ([10]). L'Edizione inoltre non deve essere vista soltanto come frutto di una fase specifica, prodotta a partire da una determinata codifica in XML-TEI, con una specifica visualizzazione del testo attraverso l'interfaccia scelta, ma è anche l'esito finale di tutto il paradigma, secondo un'ottica di edizione integrata ed estesa, dove per integrata ci si riferisce alla profonda connessione tra didattica, filologia, ermeneutica e DH, e con estesa si riprende un concetto della narratologia contemporanea, dove personaggi e storie eccedono il loro medium di riferimento ([16]). Così il paradigma IDEA mira all'allestimento di un portale-edizione dove far dialogare strumenti, contenuti, discipline diverse, aggiungendo informazioni critiche, arricchendo ed estendendo il testo cartaceo.

PARADIGMA 'IDEA' (EDIZIONE ESTESA)



Figura 4. Il paradigma 'IDEA' (Interpretazione-Didattica-Edizione-Annotazione)

3. INTERPRETAZIONE, DIDATTICA, EDIZIONE, ANNOTAZIONE

Per allestire un portale-edizione del *Nome della rosa*, col testo *in absentia* (perché riservato al cartaceo o all'e-book), bisogna costruire un discorso alternativo, che arricchisca ed estenda il testo originale, proponendo degli strumenti didattici, delle interpretazioni, degli approfondimenti critici. Il paradigma IDEA nel presente progetto prevede i seguenti passi:

- 1) Elaborazione della domanda di ricerca («*Il nome della rosa* è stato semplificato nel 2012?») e di un'ipotesi ermeneutica iniziale a partire da un'analisi quantitativa dei dati (CollateX per la collazione automatica e Voyant tools);
- 2) Costruzione di un modello di edizione in base alle specificità del testo e agli obiettivi scientifici e didattici;
- 3) Sviluppo dei primi contenuti didattici (es. traVIZ) e dello schema di codifica (finalizzato a una prima visualizzazione);
- 4) Progettazione dello schema di annotazione con un *Domain Specific Language* (DSL) per validare l'ipotesi ermeneutica²;
- 5) Allestimento del prototipo del portale e raccolta di *feedback* in funzione dell'*User-centered design*;
- 6) Nuovi cicli di revisione: riformulazione dell'ipotesi ermeneutica; rielaborazione del modello di edizione, dello schema di codifica TEI e dello schema di annotazione DSL (mettendoli in dialogo con l'obiettivo di integrare i due livelli diversi); sviluppo di nuovi contenuti didattici e di eventuali nuove funzionalità; rilascio di nuove versioni del portale;
- 7) Rilascio dell'edizione definitiva e di tutti i dati in *open access* al raggiungimento degli obiettivi prefissati e di un buon grado di esaustività dei risultati in grado di confermare o confutare l'ipotesi ermeneutica. In accordo coi principi FAIR tutti

² In questo contributo si usano i termini codifica e annotazione per indicare lo stesso processo di rappresentazione digitale del testo secondo due sfumature diverse: si usa codifica in riferimento alla rappresentazione degli aspetti essenziali del testo (tenendo conto delle divisioni strutturali, dei *tag* per l'apparato filologico, della marcatura delle entità nominate) mentre con annotazione ci si riferisce alla formulazione di annotazioni linguistiche e di postille critiche di vario genere. La scelta lessicale implica un preciso processo di differenziazione: laddove bisogna rappresentare fenomeni strutturali del testo già altamente formalizzati in TEI con l'obiettivo di condividerli con la comunità (e visualizzarli attraverso EVT) si è scelto di marcare tali aspetti in TEI sin dall'inizio; per rappresentare invece fenomeni meno formalizzati, *fuzzy*, talvolta persino tangenti al testo, ma utili al ricercatore nello sviluppo di un modello fedele alla propria domanda di ricerca e all'oggetto di studio, si è scelto di annotare costruendo un DSL per giungere allo standard TEI solo nel momento finale del paradigma con l'allestimento dell'edizione critica. Per una riflessione esaustiva sull'annotazione si rimanda a ([15]).

i contenuti devono essere rintracciabili, accessibili, interoperabili e riusabili e, in particolare, la codifica in TEI e l'annotazione DSL (esportata in TEI) vanno inserite in repository come TAPAS e Zenodo con licenza CC BY-NC-SA.

3.1 INTERPRETAZIONE, *DISTANT READING* E CONCORDANZA

L'interpretazione deve fondarsi innanzitutto su un'ipotesi di lettura del testo (desunta dalle interpretazioni storicamente determinate che sono state date nella letteratura secondaria) che può essere integrata grazie al *Distant reading* ([13]) e al metodo concordanziale ([20]) che mira a far emergere il significato del testo attraverso i rapporti tra lemma e contesto. L'interpretazione può dunque prescrivere delle linee di ricerca che possono essere realizzate nello sviluppo di unità didattiche, nell'allestimento di un'edizione secondo determinati criteri, fino all'annotazione di taluni aspetti considerati significativi. Nel nostro caso, l'ipotesi ermeneutica iniziale della semplificazione del *Nome della rosa* ha prodotto un'annotazione su Euporia con le marche d'uso di Tullio De Mauro ([8]) e attraverso una lettura quantitativa di questi dati è stato possibile dimostrare come vi sia una leggera semplificazione del lessico nelle correzioni echiane. Le due edizioni contano rispettivamente 174047 e 169661 parole, con una riduzione di 4386 parole tra la prima e la seconda edizione (il 2.5 % della prima) e una variazione che investe 6688 parole (tra cancellazioni, sostituzioni e aggiunte). Le parole marcate ad alto uso («FO», «CO», «AU», «AD») nell'edizione del 2012 sono in un rapporto di 1:5 rispetto al corpus totale per quella marca d'uso (4 parole su 5 fanno parte dell'edizione del 1980), rapporto che diventa di 1:10 per le parole appartenenti al lessico a basso uso («TS», «LE», «BU», «OB»)³, rappresentando così una forma limitata di semplificazione ([7]) che per essere valutata nel suo complesso deve essere messa in relazione al lessico dell'intera opera attraverso le concordanze⁴.

3.2 DIDATTICA, STORYTELLING E OTTICA *PUBLIC*

L'intenzione di produrre unità didattiche e approfondimenti sulla semplificazione del lessico echiano ha portato a costruire dei contenuti *ad hoc*. Si è pensato a una selezione di varianti significative perché più rappresentative di taluni fenomeni e si è data una visualizzazione tramite TRAViz⁵ ([12]).

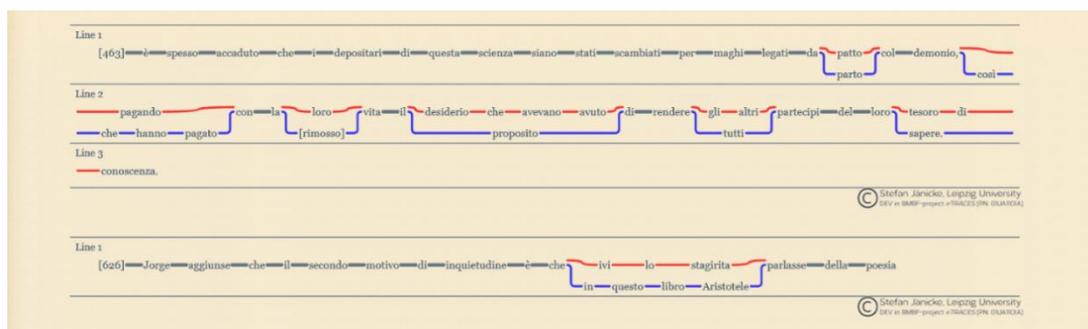


Figura 5. La top 10 delle varianti visualizzata tramite TRAViz

Un'altra unità didattica è stata pensata per approfondire la storia dell'eresia all'interno del romanzo (in quanto risulta essere il tema più variato dai dati desunti dall'annotazione tematica) con Storymap JS e Timeline JS ([5];[6]).

3.3 EDIZIONE, INTERFACCIA E VISUALIZZAZIONE

L'edizione, *stricto sensu*, non può essere del testo integrale (nel rispetto dei diritti d'autore) ma deve essere limitata all'edizione dell'apparato delle varianti, reso disponibile attraverso l'uso di EVT2 ([18]) che permette di visualizzare in modalità collazione i passi variati del testo, confrontando le lezioni differenti e facendo così emergere le varianti. Ad esempio, una codifica XML-TEI delle varianti ha la seguente struttura⁶:

³ Le categorie di De Mauro usate nell'annotazione sono: Fondamentale (FO), Comune (CO), Alto Uso (AU), Alta Disponibilità (AD), Termine specialistico (TS), Letterario (LE), Basso Uso (BU), Obsoleto (OB).

⁴ La lettura concordanziale delle varianti permette infatti di indagare in quali contesti viene variata una parola, se tutte le occorrenze di un determinato lemma vengono corrette, se la pratica correttoria è sistematica o rapsodica.

⁵ L'uso di TRAViz è un esperimento volto a rendere gli apparati di varianti maggiormente fruibili per i lettori non specialistici, sebbene la natura delle varianti (2 soli testimoni) ne limiti l'efficacia dal momento che TRAViz rende maggiormente l'idea con molti testimoni.

⁶ In una prima fase la codifica è stata circoscritta agli aspetti strutturali e filologici, demandando all'annotazione il commento puntuale. Nella successiva integrazione si prevede di inserire nel file TEI per ogni *rdg* gli attributi *type* e *cause* (secondo una tassonomia di valori definiti nell'annotazione DSL) e il tag *note* con il commento critico, in modo da implementare l'annotazione in EVT.

<app>
 <lem wit="#NR1980">i terribili avvenimenti che</lem>
 <rdg wit="#NR2012">quanto</rdg>
 </app>

Codifica del testo che poi in EVT2 viene visualizzata nel seguente modo, con la possibilità di cliccare sulla parola e avere una sorta di apparato a comparsa:

- [5] Forse, per comprendere meglio gli avvenimenti⁵ in cui mi trovai coinvolto, è bene che io ricordi quanto stava avvenendo in quello scorcio di secolo, così come lo compresi allora, vivendolo, e così come lo rammento ora, arricchito di altri racconti che ho udito dopo - se pure la mia memoria sarà in grado di riannodare le fila di tanti e confusissimi eventi⁶.
- [6] Accadde così che piegammo verso occidente mentre la nostra meta ultima stava⁷ a oriente, quasi seguendo la linea montana che da Pisa porta in direzione dei cammini di San Giacomo, soffermandoci in una terra che i terribili avvenimenti che⁸ poi vi vennero⁹ mi **sconsigliano**¹⁰ di identificare meglio, ma i cui signori erano fedeli all'impero e dove gli abati del nostro ordine di comune accordo si opponevano al papa eretico e corrotto.
- [7] La gioventù non vuole apprendere più nulla, la scienza è in decadenza, il mondo intero cammina sulla testa, dei ciechi conducono altri ciechi e li fan precipitare negli abissi, gli uccelli si lanciano prima di aver preso il volo, l'asino suona la lira, i buoi danzano, Maria non ama più la vita contemplativa e Marta non ama più la vita attiva, Lea è sterile, Rachele ha l'occhio carnale, Catone frequenta i lupanari, **Lucrezio diventa femmina**¹¹. Tutto è sviato dal proprio cammino.

• [6] Accadde così che piegammo verso occidente mentre avremmo dovuto andare¹ a oriente, quasi seguendo la linea montana che da Pisa porta in direzione dei cammini di San Giacomo, soffermandoci in una terra che quanto² poi vi accadde³ mi **sconsiglia**⁴

sconsigliano NR1980
 • GROUP 1: sconsigliano NR1980
 • GROUP 2: sconsiglia NR2012

Info aggiuntive XML

```
<app>
  <lem wit="#NR1980">sconsigliano</lem>
  <rdg wit="#NR2012">sconsiglia</rdg>
</app>
```

di identificare meglio, ma i cui signori erano fedeli all'impero e dove gli abati del nostro ordine di comune accordo si opponevano al papa eretico e corrotto.

Figura 6. EVT 2

L'edizione, però, *lato sensu*, è quella che emerge dall'integrazione delle varie sezioni del portale, quella che abbiamo definito 'edizione estesa' che nasce dal dialogo tra didattica e scientificità, tra interfaccia, *data visualization* e annotazione.

3.4 ANNOTAZIONE, EUPORIA E APPARATO

Last but not least, l'annotazione è il momento in cui emerge l'interpretazione dello studioso e sorregge l'intera impalcatura ermeneutica dell'edizione. Può essere una 'semplice' descrizione degli elementi formali e linguistici del testo oppure può essere il momento dell'interrogazione profonda dell'opera, seguendo la lezione 'analogica' di Barthes in *S/Z* ([2]). La nostra annotazione è stata sviluppata attraverso Euporia ([1];[14]), che permette di annotare attraverso un proprio DSL, in modo tale che l'interpretazione fosse lasciata il più possibile libera e scevra da precomprensione.

1109 Chiesi chi aveva messo in capo alla folla che bisognava attaccare gli ebrei. Salvatore non ricordava. Credo che quando si radunano tante folle troppe persone seguendo una promessa e chiedendo subito qualcosa, non si sappia mai chi parla tra di loro. Pensai che i loro capi si erano educati nei conventi e nelle scuole vescovili, e parlavano il linguaggio dei signori, anche se lo traducevano in termini comprensibili a ai pastori. E i pastori non sapevano dove stesse il papa, ma sapevano dove stavano gli ebrei. Insomma, **presero avevano preso** d'assedio un'alta e massiccia torre del re di Francia, dove gli ebrei spaventati erano corsi in massa a rifugiarsi. E gli ebrei usciti sotto le mura della torre si difendevano coraggiosamente e **spietatamente**, lanciando legna e pietre. Ma i pastorelli **applicarono avevano applicato** il fuoco alla porta della torre, **tormentando** e gli ebrei **asserragliati** coi **assedati dal fumo e coi dal fuoco** - **E gli ebrei non potendo salvarsi**, preferendo uccidersi piuttosto che morire per mano dei non circoncisi, **chiesero a uno di avevano chiesto** al più arditto tra loro **che sembrava il più coraggioso, di ucciderli trafiggerli** con la spada. **Egli acconsentì** - **E costui ne uccise aveva ammazzato** quasi cinquecento. Poi **uscì**, uscito dalla torre coi figli degli ebrei, **chiese aveva chiesto** ai pastorelli di essere battezzato. Ma i pastorelli gli **dissero avevano detto**: tu hai fatto una tale strage della tua gente e ora pretendi di sottrarti alla morte? e lo **fecero avevano fatto** a pezzi, risparmiando però i bambini, che poi fecero battezzare. Poi si diressero verso Carcassonne, compiendo molte sanguinose rapine durante il loro cammino. Allora il re di Francia avvertì che essi avevano passato il limite e ordinò che si opponesse loro resistenza in ogni città in cui passavano e si difendessero persino gli ebrei come fossero uomini del re...

```
[10] @narratore {storia;eresia $racconto/Salvatore
[11] [tante folle] : <trappe persona> = #errore : $sinonimo $alleggerimento || [tante] : <trappe> = #FO : #
[12] [a] : <ai> = #del : #varOrtografica
[13] [presero] : <avevano preso> = #indPassRemoto : #indTrappassProssimo
[14] [e spietatamente] : <$riduzione/min || $spietatamente> = #CO
[15] [applicarono] : <avevano applicato> = #indPassRemoto : #indTrappassProssimo
[16] [tormentando] : <e> = #del : $appiattimento || [tormentando] = #LE
[17] [asserragliati col] : <assedati dal> = #del : $semplificazione = || [asserragliati] : <assedati> = #
[18] [col] : <dal> = #del : #add
[19] (. E gli ebrei non potendo salvarsi) : <$riduzione/med || [ebrei] = #FO || [potendo] = #FO || [salvarsi]
[110] [chiesero a uno di] : <avevano chiesto al più arditto tra> = #del : #add $appesantimento || [chiesero]
[111] [, che sembrava il più coraggioso] : <$riduzione/med || [sembrava] = #FO || [coraggioso] = #AU
[112] [ucciderli] : <trafiggerli> = #del : $appesantimento || [ucciderli] : <trafiggerli> = #FO : #CO
[113] [Egli acconsentì, e] : <E costui> = #del : $riduzione/par || [Egli] : <costui> = #pronSogg : #CO || [
[114] [uccise] : <aveva ammazzato> = #indPassRemoto : #indTrappassProssimo || [uccise] : <aveva ammazzato> = .
[115] [uscì] : <, uscito> = #indPassRemoto : #partPerfetto
[116] [e chiese] : <aveva chiesto> = #indPassRemoto : #indTrappassProssimo
[117] [dissero] : <avevan detto> = #indPassRemoto : #indTrappassProssimo
[118] [una tale] : <$riduzione/min || [tale] = #FO
[119] [fecero] : <avevan fatto> = #indPassRemoto : #indTrappassProssimo
```

Sulla sinistra il testo con le varianti evidenziate attraverso l'algoritmo di collazione.

Sulla destra l'annotazione con un DSL in base alle domande di ricerca.

Figura 7. Una schermata dell'ambiente Euporia con l'annotazione in progress

Attraverso alcuni simboli predefiniti in fase di sviluppo del linguaggio di dominio (dal # usato per descrivere il fenomeno linguistico o la marca d'uso, § per i descrittori tematici e interpretativi, @ per indicare il personaggio, le parentesi graffe per indicare le rimozioni dall'edizione del 1980 e le parentesi uncinate invece le aggiunte nel 2012) si è così potuti arrivare a descrivere la pratica correttoria echiana per rispondere alla domanda di ricerca iniziale.

4. CONCLUSIONI. VERSO LA NUOVA EDIZIONE «RIVEDUTA E CORRETTA»

Il prototipo di edizione digitale qui proposto trova la sua ragion d'essere nell'integrazione tra le diverse modalità di approccio al testo, alcune delle quali solitamente marginali nello studio critico, ma centrali in un'ottica di diffusione degli studi specialistici anche a un pubblico non accademico. Se si pensa che la critica letteraria digitale non si debba limitare a riflettere soltanto su opere del lontano passato, ma debba interrogarsi e proporre contenuti sui testi significativi del presente, il paradigma IDEA dell'edizione estesa mira ad arricchire la conoscenza attraverso approfondimenti in rete. Ma la

potenzialità del digitale sta anche nel continuo ampliamento dei suoi confini, con nuove proposte e riflessioni. Diventa centrale quindi il *Versioning* e una *road map* dei contenuti futuri. Una versione 2.0 del portale-edizione⁷ è in sviluppo con:

- una nuova annotazione con il possibile sviluppo di un'ontologia secondo una tassonomia delle varianti ben definita e un motore di ricerca delle varianti per tema, personaggio, intervento correttivo⁸;
- integrazione tra la codifica TEI e l'annotazione DSL;
- altre unità didattiche sulle singole varianti e sulla top10 nell'ottica di uno *storytelling* transmediale delle varianti;
- alcuni esperimenti di *gamification* con dei quiz pensati per studenti di liceo e dell'università;
- rilascio dell'edizione scientifica digitale vera e propria (*open content*) nel rispetto dei principi FAIR.

Così come Umberto Eco, trentadue anni dopo la pubblicazione del suo capolavoro, decise di intervenire e correggere il proprio testo, con «un'edizione riveduta e corretta», così *Le varianti della rosa* nella sua stessa essenza si propone di essere in continuo aggiornamento, con la consapevolezza che «stat rosa pristina nomine, nomina nuda tenemus».

5. RINGRAZIAMENTI

Si ringrazia il CNR-ILC (e Federico Boschetti) per l'elaborazione dello schema di codifica su Euporia; Roberto Rosselli del Turco e gli sviluppatori di EVT; e infine il CINUM per la cornice interpretativa e l'utilizzo del software di concordanza.

BIBLIOGRAFIA

- [1] Bambaci, Luigi, Federico Boschetti, e Riccardo Del Gratta. 2019. «Qohelet Euporia: a Domain specific Language for the Encoding of the critical Apparatus». *International Journal of Information Science & Technology* 3 n.5: 26–37.
- [2] Barthers, Roland. 1970. *S/Z*. Du Seuil. Paris.
- [3] Bonsi, Claudia, Angelo Di Iorio, Paola Italia, e Fabio Vitali. 2015. «Manzoni's Electronic Interpretations». *The Mechanic Reader* LIII (febbraio): 91–99.
- [4] Buzzoni, Marina. 2016. «A Protocol for Scholarly Digital Editions? The Italian Point of View». In *Digital Scholarly Editing: Theories and Practices*, a cura di Matthew James Driscoll e Elena Pierazzo, 59–82. Open Book Publishers. <https://doi.org/10.11647/OBP.0095.04>.
- [5] Chicago-San Francisco: Knightlab. 2021a. «StorymapJS». StorymapJS. 2021. <https://storymap.knightlab.com/>.
- [6] ———. 2021b. «TimelineJS». TimelineJS. 2021. <http://timeline.knightlab.com/>.
- [7] D'Agata, Christian. 2021. «I Nomi della Rosa. Un'analisi testuale informatica delle varianti del Nome della Rosa tra Distant e Close reading». In *Letteratura e Scienze. Atti delle sessioni parallele del XXIII Congresso dell'ADI*. Roma: Adi editore.
- [8] De Mauro, Tullio. 1999. *Grande dizionario italiano dell'uso con CD-Rom*. Torino: Utet.
- [9] Eco, Umberto. 2012. *Il nome della rosa*. Milano: Bompiani.
- [10] Gadamer, Hans Georg. 2000. *Verità e metodo*. Trad. e Apparati di G. Vattimo. Milano: Bompiani.
- [11] Giuffrida, Milena, Christian D'Agata, Laura Giurdanella, e Pietro Sichera. 2021. «Pirandello Nazionale: per un nuovo modello di edizione digitale, collaborativa e integrata». In *Boschetti, Federico; Del Grosso, Angelo Mario; Salvatori, Enrica; (edd.); AIUCD 2021 - DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale. Raccolta degli abstract estesi della 10a conferenza nazionale*. Pisa.
- [12] Janicke, Stefan, Annette Geßner, Greta Franzini, et al. 2015. «TRAViz: A Visualization for Variant Graphs». *Digital Scholarship in the Humanities* 30 (dicembre). <http://www.traviz.vizcovery.org>.
- [13] Moretti, Franco. 2020. *A una certa distanza*. Roma: Carocci.
- [14] Mugelli, Gloria, Federico Boschetti, Riccardo Del Gratta, Del Grosso, Angelo Mario, Fahad Khan, e Andrea Taddei. 2016. «A User-Centred Design to Annotate Ritual Facts in Ancient Greek Tragedies». *Bulletin of the Institute of Classical Studies* 59 (2): 103–20. <https://doi.org/10.1111/j.2041-5370.2016.12041.x>.
- [15] Nantke, Julia, e Frederik Schlupkothen. 2020. *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization*. De Gruyter.
- [16] Pescatore, Guglielmo. 2018. *Ecosistemi narrativi*. Milano: Carocci.
- [17] Pischetta, Bruno. 2016. *Eco. Guida al Nome della rosa*. Milano: Carocci.
- [18] Rosselli Del Turco, Roberto, Chiara Di Pietro, e Chiara Martignano. 2019. «Progettazione e implementazione di nuove funzionalità per EVT 2: lo stato attuale dello sviluppo». *Umanistica Digitale*, No 7 (2019). <https://doi.org/10.6092/ISSN.2532-8816/9322>.
- [19] Sahle, Patrick. 2016. «What is a Scholarly Digital Edition». In *Digital Scholarly Editing. Theories and Practices*, by M. J. Driscoll and E. Pierazzo, 19–40. Cambridge, UK: Open Public Publishers.

⁷ Il prototipo del portale è stato sviluppato adattando un template CSS realizzato con il framework Bootstrap. La versione 2.0 si servirà invece di un Content Management System svincolando lo sviluppatore dalla gestione dell'intero sistema, avendo l'obiettivo di un allargamento progressivo dei contenuti e di un'apertura alla collaborazione con studenti e studiosi per la loro creazione e gestione.

⁸ In questo modo l'annotazione diventa un modo per esplicitare l'interpretazione dell'editore. L'utente può interagire con i dati, avere statistiche sulle annotazioni, ricercare, verificare ed eventualmente proporre interpretazioni alternative delle varianti, rendendo aperto e interattivo il lavoro del filologo, il quale può confrontarsi e ricevere *feedback* in ogni fase della propria ricerca.

- [20] Savoca, Giuseppe. 2000. *Lessicografia letteraria e metodo concordanziale*. Firenze: Olschki.
- [21] «TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.3.0». 2021. Tei Consortium. 2021.
<http://www.tei-c.org/Guidelines/P5/>.

Online lexical resources for translators: where do we stand?

A (possibly meaningful) case-study

Francesca Bianchi¹, Marco Zappatore²

¹University of Salento, Italy – francesca.bianchi@unisalento.it

²University of Salento, Italy – marcosalvatore.zappatore@unisalento.it

ABSTRACT

By means of a case-study of word ‘jab’ in its medical sense, this work aims to assess the effectiveness of a range of online lexical resources in providing word meanings. These include: monolingual and bilingual dictionaries; corpora; and machine translation services.

KEYWORDS

Word meaning; Machine Translation; Corpora and Corpus Analysis Software; Online Dictionaries; Digital WWW archive.

TALK

1. INTRODUCTION

The Web offers translators (and students) a wide range of lexical and textual resources to support them in their daily work. These include: online encyclopaedias (including wiki ones); online monolingual and bilingual dictionaries (from web versions of famous paper dictionaries to glossaries of various forms and origins); concordance tools that show translations in context by searching online parallel documents (e.g. *reverso context* and *linguee*, this latter being based on the DeepL machine translation service); tools to query existing corpora or create ad-hoc corpora and analyse them (e.g. The Sketch Engine); machine translation (MT) services. Corpus linguists claim that corpora are the best possible lexical resource ever. Students resort almost exclusively to bilingual dictionaries, *linguee* or *reverso context*, as well as MT services. Professional translators make reasoned and combined use of all the available resources.

The current paper aims to take stock of the current situation and assess the effectiveness of a range of different resources in helping users understand word meaning. Our case-study revolves around term ‘jab’, in its medical sense. This word has been chosen for the following reasons: it has a polysemic nature, being used in sports, general language (in both concrete and metaphorical meanings) and medicine, but also in a range of grammar forms and meanings within the medical area; it has been largely used in the medical field for several decades (see Preliminary results), a time span which is short enough for us to consider this term a fairly recent acquisition in the language of medicine, but long enough for resources to have included it; and has made the front pages in the media in the last 12 months because of the COVID-19 pandemic.

We expect corpora to confirm their role as the best lexical resource; monolingual dictionaries to be rather limited in the number of meanings provided, and bilingual dictionaries to be almost useless. Finally, we expect MT engines to offer rather inaccurate solutions for the translation of word ‘jab’ in medical contexts, but to improve their output fairly quickly in time.

2. ON THE ONLINE RESOURCES CONSIDERED IN THIS WORK

This work considers the following online resources and tools:

- Wikipedia English.
- Major online English dictionaries accessible through the Onelook¹ portal (e.g. Merriam Webster; Macmillan; Collins Cobuilt; Oxford English Dictionary; Cambridge Advanced Learners Dictionary; American Heritage Dictionary of the English language).

¹ <https://onelook.com/>.

- Online English-Italian bilingual dictionaries (Cambridge Bilingual²; Collins³; Sansoni⁴; Larousse⁵; Wordreference⁶).
- The Sketch Engine corpus tool⁷. Of all corpus tools, the Sketch Engine has been chosen for this analysis because it offers specific affordances which are unavailable elsewhere: word sketches, i.e. synthetic automatic outlines of a word's grammatical and collocational behaviour, several large corpora of English, and the possibility to quickly compile ad-hoc corpora from the Web.
- Four MT services: Google Translate⁸; Microsoft Bing Translator⁹; DeepL Translator¹⁰; and Systran Translate¹¹. These four MT services were specifically selected for a twofold reason. First, they are all accessible as free cloud services via dedicated Web pages where users can write or upload their source texts directly. Second, they take advantage of four different types of Neural MT (NMT), which is nowadays the best-performing approach to MT. Neural-based MT engines are grounded on artificial neural networks that guarantees better capturing of the context of full sentences before translating them, thus mimicking human cognitive processes and acquiring new knowledge over time. Therefore, NMT represents an improvement in comparison with previous phrase-based statistical MT approaches. More specifically, the selected MT services adopt the following approaches: ad-hoc Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks in the case of Google (also known as Google Neural Machine Translation or GNMT; ([8])); Marian NMT (a customised deep neural MT) in the case of Bing ([4]); Convolutional Neural Networks (CNNs) in the case of DeepL ([1]); and Pure Neural MT (PNMT) in the case of Systran ([2]).

It is interesting to observe that almost all of the resources listed above take advantage of electronic corpora in one form or another. Indeed, corpora have long proven to be an excellent means to identify and distinguish between word meanings. Even several famous dictionaries have long started to take advantage of corpora (e.g. Collins Cobuild; Macmillan). On the other hand, MT tools have passed from taking advantage of large parallel corpora to being based on complex self-learning algorithms – known as artificial neural-networks – which are trained on ever larger parallel corpora and beyond ([3]); some are even capable of supplying translations for zero-resource language pairs ([5]). Some MT tools, such as Google translate, have greatly improved the quality of translation into/from English, especially when the text is technical and words are provided in context (i.e. when full texts are fed into the system, rather than single words) ([6]).

3. MATERIALS AND METHODS

The resources listed above are analysed and assessed considering the quantity and quality of the lexical meanings they provide for the English word ‘jab’ in medical contexts. Furthermore, all the resources are analysed at various points in time, to see whether and how quickly the information they offer has changed. September 17th 2021 is our primary reference point in time (Time T₀), i.e. the benchmark day for our initial and subsequent comparisons. From September to the end of December 2021, monthly checks will be performed. This seems particularly necessary in the case of MT tools, for which rather quick improvements can be expected. Wherever possible, Wikipedia and dictionaries are also analysed backwards in time, to see when the word ‘jab’ first appeared in its medical meaning. To this aim, we used the Wayback Machine digital archive¹², which preserves browsable copies of Web pages cached starting from 1996. These archived pages are provided by Web crawling applications that periodically download all publicly accessible data from a given website into a so-called snapshot. Snapshots are captured with variable frequency which also changes per website.

To test the MT tools, five full texts are inputted, taken from online news. Taken together these five texts include an interesting range of meanings, collocations and grammar forms of word ‘jab’.

The Sketch Engine is used to analyse English corpora of various types and times. More specifically the following are considered: the British National Corpus (BNC), which is a general-purpose corpus manually created in the 1990s and containing about 100 million words; and three general corpora of the TenTen family (enTenTen12; enTenTen18; enTenTen20), which are several-million-word corpora automatically created from the Web. Furthermore, the Sketch

² <https://dictionary.cambridge.org/dictionary/english-italian/>.

³ <https://www.collinsdictionary.com/browse/english-italian/>.

⁴ https://dizionari.corriere.it/dizionario_inglese/.

⁵ <https://www.larousse.fr/dictionnaires/anglais-italien>.

⁶ <https://www.wordreference.com/>.

⁷ <https://www.sketchengine.eu/>.

⁸ <https://translate.google.com/>.

⁹ <https://www.bing.com/translator>.

¹⁰ <https://www.deepl.com/en/translator>.

¹¹ <https://translate.systran.net/>.

¹² <https://web.archive.org/>.

Engine is used to create a specialised corpus from the Web revolving around word ‘jab’ and filtering out non-medical fields; for the creation of this corpus – hereafter called ‘the Jab corpus’ – the following seeds (i.e. words inputted to the web crawling software) are used: jab; jabs; covid; vaccine; pandemic; booster; jabbed; jabbing.

4. PRELIMINARY RESULTS

A typical starting point to understand the meaning of a word is an encyclopaedia or a dictionary. Here we shall start with Wikipedia. When the user enters word ‘jab’ in the Wikipedia search box, a page about martial arts is displayed¹³. However, clicking on the disambiguation link, the user learns that “Jab [...] may also refer to” a list of things top of which is the following meaning: “Jab, an informal term for an injection”. At our Time T0, the Injection page¹⁴ clarifies that an injection is “often and usually referred to as a ‘shot’ in US English, a ‘jab’ in UK English, or a ‘jag’ in Scottish English and Scots”. Going back in time thanks to the Wayback Machine cached pages, it appears that ‘shot’ was added as a synonym on December 17th 2008, while ‘jab’ on March 23rd 2010. In Wikipedia, ‘jab’ clearly appears also with the meaning of ‘vaccine’ in the Influenza vaccine page¹⁵, where “Influenza vaccines, also known as flu shots or flu jabs” are explained. A Wayback Machine analysis of this page shows that the US synonym was entered on January 16th 2010, while the UK synonym (‘jab’) on January 23rd 2017.

If we look at dictionaries, according to the OED (paper version) word ‘jab’ “was born in 1825, and first appeared in glossaries of the Scots language, wherein it meant ‘a sharp prick’. That general sense of jab later became specific, gradually coming to also mean any hypodermic injection. [Furthermore] One 1914 vocabulary of criminal slang noted that a noun sense of jab was ‘common amongst morphine and cocaine fiends’ ([7]). Indeed, the Collins Cobuild – International Edition (paper version), issued in 1994, has an entry for ‘jab’ in the medical field, where it defines ‘jab’ as “an injection of a substance into your body to prevent illness; used in informal English”; the dictionary also offers ‘injection’ as a synonym. All the online dictionaries considered in this study illustrate the use of noun ‘jab’ with reference to the ‘injection’ meaning. The definitions range from minimal and rather vague (Collins English dictionary and Cambridge Advanced Learners: ‘an injection’), to minimal but more precise (American Heritage Dictionary: ‘a hypodermic injection’) to rather detailed (Merriam Webster: ‘an injection of something (such as medicine) into one’s body with a needle’; Oxford English dictionary: ‘A hypodermic injection, especially a vaccination’; Collins Cobuild: ‘an injection of something into your blood to prevent illness’). These last two definitions actually suggest the rather more specific meaning of ‘vaccination’ rather than the more general meaning of ‘injection’. Furthermore, with the only exception of the Oxford English dictionary, they all specify that it is used in informal contexts. Furthermore, all the online dictionaries considered except the American Heritage one clarify that it is chiefly (Merriam Webster only) or exclusively used in British English. The Wayback Machine digital archive goes as far back as 1996, but the actual initial date of cached pages largely differs depending on website. So, for instance, for the Cambridge English dictionary ‘jab’ page there are no cached pages before 2015. In any case, a backward analysis shows that for some dictionaries the entry of ‘jab’ as a noun meaning ‘injection’ is a relatively recent acquisition. This is the case of the Merriam Webster dictionary, where it is first attested on June 20th 2021. For the remaining dictionaries the noun entry quoted before is always present in all cached pages.

A few of the online dictionaries also report ‘jab’ to be used as verb form in the area of medicine. These are the Oxford dictionary (def.: Inject (someone) with a vaccine; vaccinate), the Collins English dictionary (def.: to administer an injection (to)), and the Cambridge Advanced Learners (def.: to use a needle and syringe (= small tube) to put a liquid such as a drug into a person’s body). Using the Wayback Machine digital archive it was possible to establish that the verb entry is a very recent addition, wherever present: the Oxford and the Collins dictionaries added it on August 25th and August 22nd 2021, respectively; while the Cambridge dictionary must have been added it even more recently since none of the cached pages include ‘jab’ as a verb in the medical field.

While encyclopaedias and dictionaries are the result of specific lexicographic efforts, corpora collect texts spontaneously produced by language users for their everyday living or working purposes, and offer a view of words in context. The BNC collects texts produced in 1990s and illustrates word usage around that time. In the BNC, ‘jab’ appears in medical contexts a few times, primarily used as a noun and modified by names of illnesses: tetanus/anti-tetanus/encephalitis/typhoid jab. (e.g.: *in three doses at two, three, and four months, at the same time as the triple diphtheria, whooping cough, and tetanus jab. Doctors hope vaccination will repeat the success of measles and whooping cough immunisation programmes*). The BNC also reports one sentence where jab is used as a verb, with the (semi-technical) meaning ‘to inject’ (e.g. *She exposed her wrist, and jabbed the painkiller into her bloodstream*).

¹³ <https://en.wikipedia.org/wiki/Jab>.

¹⁴ [https://en.wikipedia.org/wiki/Injection_\(medicine\)](https://en.wikipedia.org/wiki/Injection_(medicine)).

¹⁵ https://en.wikipedia.org/wiki/Influenza_vaccine.

The three TenTen corpora consulted, which are respectively based on the Web in 2012, in 2018 and in 2020, report similar uses of noun ‘jab’, though preceded by a different range of illnesses (clearly a consequence of different illnesses having different impacts on society at the various point in time): flu, avian flu (enTenTen12 only), swine flu, tetanus, rubella (enTenTen18 only) and MMR (measles-mumps-rubella). They also attest the use of ‘jab’ as a verb followed by ‘needle’ (e.g. *bone marrow biopsies - an awful, painful procedure that requires jabbing a thick needle into the hipbone*), which however illustrates the general meaning ‘to pierce’, rather than a specifically medical meaning.

The Jab corpus (616,069 words in all) includes almost exclusively instances of ‘jab’ related to its uses in the medical field, as expected. More specifically, it includes 537 hits of ‘jab’ as verb, and as many as 2099 hits of ‘jab’ as noun. Thanks to the Word Sketch feature it is rather easy to observe that, besides the well-established meaning of ‘injection’, as a noun ‘jab’ is also used to refer to the following:

- dose of vaccine (e.g. *The more jabs gotten, the greater the risk to health and a shortened lifespan*);
- vaccine (e.g. *Pfizer's jab has already been approved for 12- to 15-year-olds; Johnson & Johnson jabs were given so only one dose is needed; Millions of older Brits will miss out on third Covid-19 jab doses for now*);
- vaccination (e.g. *More than 700,000 Covid-19 jabs were booked in one day; Ursula von der Leyen announced the bloc will draw up plans for jabs certificates within weeks and aims to have them in place by June*).

As a verb, the corpus amply illustrates that ‘jab’ is used with the meaning of ‘vaccinate’: e.g. *Britain has been slipping down the rankings, not least because it was slow to jab children; those who have received two Covid jabs may no longer be considered "fully jabbed"*; *Government wants more young people jabbed soon*.

Let us now see what happens when we consider English-Italian bilingual dictionaries and MT engines.

Cambridge bilingual offers two translations for ‘jab’ as a noun: *puntura* (i.e. ‘injection’) and *vaccinazione* (i.e. ‘vaccination’). The Collins English-Italian dictionary and the Larousse dictionary offer only one translation: *puntura* (‘injection’). The Sansoni dictionary suggests three different translations: *iniezione* (‘injection’), *puntura* (‘injection’) and *vaccinazione* (‘vaccination’). Finally, Wordreference (particularly loved by students) offers *vaccinazione* (‘vaccination’) as the only possible translation. None of them, however consider ‘jab’ as a verb.

The picture with MT engines at Time T0 is rather varied: Google used the correct term 5 times out of 15, all of them in the noun occurrence; Bing used the correct term 6 times, all nouns; DeepL managed to provide the correct translation 8 times, including one of the only two verb occurrences; while Systran managed to correctly understand the meaning of ‘jab’ only once, as a noun.¹⁶ At Time T1, Google used the correct term 5 times (the same occurrences of Time T0); Bing used the correct term 6 times (the same occurrences of Time T0); DeepL managed to provide the correct translation 10 times (the previous ones plus two new instances); while Systran managed to correctly understand the meaning of ‘jab’ twice (the only previous one plus a second occurrence). It must also be observed that, regardless of the number of correct/incorrect ‘jab’ hits, in all systems the translations showed some kind of change between Time T0 and Time T1. In particular, the number of occurrences where ‘jab’ was translated as *jab* (i.e. no translation) passed from 4 to 2 in DeepL, and from 9 to 6 in Systran.

5. CONCLUDING REMARKS

This study has focused on a word with specific, though not uncommon, characteristics: beside having one or more concrete and metaphorical meanings in general language, and a technical meaning in a specialised area (sports), ‘jab’ is also an informal way to address specialised meanings in a different scientific area (medicine), where it is used both as noun and verb. The latter are the meanings we focused on this work.

Our preliminary analysis has shown that, when it comes to understanding the meanings of ‘jab’ in the area of medicine, a specially created corpus targeting the focus word is the best possible resource, thus confirming one of our hypothesis. Our Jab corpus immediately evidenced three different technical meanings for ‘jab’ as a noun, beside the well-established meaning of ‘injection’, and one technical meaning as a verb. The large general-purpose corpora used proved partially useful: they all clearly highlighted one technical meaning of the noun (‘vaccine’), besides the ‘injection’ one; furthermore, the BNC illustrated verb ‘to jab’ in the sense of ‘to inject a drug’, while the enTenTen corpora illustrated verb ‘to jab’ in the sense of ‘to pierce with a needle’.

On the other hand, monolingual dictionaries generally offer only the most general of all the technical meanings of the noun, i.e. ‘injection’, although a few of them, by enriching the definition with some level of detail, more or less explicitly suggest the more technical meaning of ‘vaccination’. The verb is not acknowledged by three of the six dictionaries considered, and is variably treated by the remaining three, with the Oxford dictionary ignoring the more general meaning of ‘administer an

¹⁶ In this preliminary analysis no distinction is made between incorrect translation (e.g. *pillola*, or *colpi*) and no translation (e.g. *i jab vengono offerti*). Furthermore, in this preliminary analysis attention is paid exclusively to the use of the correct Italian word, given its general context, regardless of morphological mistakes in the collocation or syntactic mistakes in the sentence.

injection’ in favour of the ‘vaccinate’ one. Bilingual dictionaries did not fare worse than monolingual ones, at least as far the noun form is concerned. Unfortunately, however, none of them considered ‘jab’ as a verb. Thus, also our second hypothesis can be considered confirmed.

Finally, as expected, MT engines are not yet able to correctly deal with such a polysemous word as ‘jab’, despite we feeded them with full texts that clearly revolve around medicine and vaccines, and despite the fact that the word, in its medical sense, has been around for decades. The analyses have also confirmed that MT engines are constantly ‘on the move’ and, if used at different times, they tend to produce ever different translations for the same sentences. For some of them (most evidently DeepL) such constant change has corresponded to actual improvements in the understanding and treating of our focus word.

The contrast between how fast some MT engines progress and how slowly dictionaries are updated with new meanings – as shown by our backwards analyses with the Wayback Machine – is striking and suggests that shortly MT engines might replace dictionaries, at least for translators and more generally bilingual users. For monolingual users, on the other and, an ad-hoc corpus remains the best solution. Corpora however have not yet achieved popularity outside specialised groups, such as corpus linguists and lexicographers. This is probably due to the fact that – as shown in this case – general corpora are too general to make a real difference compared to dictionaries, and ad-hoc corpora are thought to be difficult to create and analyse. However, the Sketch Engine makes corpus creation and lexical analysis really easy and the data could be interpreted with little training. If only the Sketch Engine, or a similar tool, were freely accessible on the Web, the chances of corpora replacing monolingual dictionaries as lexical tools in the hands of translators and students would dramatically increase.

REFERENCES

- [1] Coldewey, Devin, and Frederic Lardinois. 2017. “DeepL Schools Other Online Translators with Clever Machine Learning.” *TechCrunch* (blog). 2017. <https://techcrunch.com/2017/08/29/deepl-schools-other-online-translators-with-clever-machine-learning/>.
- [2] Crego, Joseph, et al. 2016. “SYSTRAN’s Pure Neural Machine Translation Systems.” *ArXiv* 1610.05540. <https://arxiv.org/abs/1610.05540>.
- [3] Hassan, Hany, et al. 2018. “Achieving Human Parity on Automatic Chinese to English News Translation.” *ArXiv*:1803.05567. <https://arxiv.org/abs/1803.05567>.
- [4] Microsoft. 2019. “Neural Machine Translation Enabling Human Parity Innovations in the Cloud.” *Microsoft Translator Blog* (blog). 2019. <https://www.microsoft.com/en-us/translator/blog/2019/06/17/neural-machine-translation-enabling-human-parity-innovations-in-the-cloud/>.
- [5] Schuster, Mike, Johnson Melvin, and Nikhil Thorat. 2016. “Zero-Shot Translation with Google’s Multilingual Neural Machine Translation System.” *Google AI Blog* (blog). 2016. <https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html>.
- [6] “The State of Machine Translation 2020 - Independent Multi-Domain Evaluation of Commercial Machine Translation Engines.” 2020. *Intento, Inc.*
- [7] Webb, Tiger. 2021. “The COVID-19 Pandemic Has Seen Few Words Loathed as Much as ‘Jab.’” *ABC NEWS* (blog). 2021. <https://www.abc.net.au/news/2021-04-19/jab-the-most-hated-word-coronavirus-pandemic-needle-phobia/100058980>.
- [8] Yonghui, Wu, et al. 2016. “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.” *ArXiv* 1609.08144. <https://arxiv.org/abs/1609.08144>.

Sessione Filosofia 3
Gilbert Simondon

Governare le piattaforme.

Cinque proposte su pluralismo e polarizzazione online¹

Gabriele Giacomini

Università degli Studi di Udine, Italia; Center for Advanced Studies Southeast Europe - gabrielegiacomini@hotmail.it

ABSTRACT

La filosofia politica ha spesso affrontato il problema del pluralismo. Autori come Rawls e Arendt, *mutatis mutandis*, teorizzano un sistema politico caratterizzato da un linguaggio pubblico grazie al quale comunicare, pur continuando a parlare linguaggi non pubblici divergenti e, a volte, incommensurabili. Tuttavia, questo approccio rischia di essere messo in crisi dall'Internet delle piattaforme. Infatti, tecnologie come la profilazione degli utenti e la personalizzazione dei contenuti rinforzano le dinamiche sociopsicologiche dell'omofilia (*echo chambers* e *filter bubbles*), radicalizzando le posizioni. Per questo motivo, si sostiene l'opportunità di promuovere alcune azioni per permettere alle piattaforme digitali, opportunamente orientate dalle istituzioni liberaldemocratiche, di mitigare il problema della polarizzazione online.

PAROLE CHIAVE

Polarizzazione, pluralismo, Rawls, Arendt, ICT, piattaforme, web.

INTERVENTO

1. INTRODUZIONE

Il pluralismo solleva diversi problemi per le istituzioni democratiche, che si trovano a dovere rispondere a complesse domande di giustificazione: come dovrebbe essere strutturata una società per assicurare che i diversi progetti di vita dei suoi membri siano perseguibili? Quali sono i limiti da imporre alle diversità affinché siano sostenibili in un ordine sociale? Quale risposta dovrebbe essere data innanzi a situazioni che mettono in crisi la stabilità della democrazia? Queste domande sono tanto più importanti quanto più il contesto è caratterizzato da un "iper-pluralismo" ([11]), da un "pluralismo polarizzato" ([20]), ovvero dalla presenza di idee e valori sempre più divergenti.

Negli ultimi decenni, recenti sfide alla sostenibilità politica del pluralismo sono state poste, ad esempio, dall'avanzamento della medicina e da dilemmi di tipo bioetico o dalla gestione dell'immigrazione. In questo contributo si intende segnalare la presenza di una nuova sfida al pluralismo in un contesto democratico liberale, che è posto dalle ICT. L'attuale Internet delle piattaforme, in particolare, a causa di strumenti tecnologici innovativi (IA, psicografia, big data eccetera) che permettono la profilazione degli utenti e la personalizzazione dei contenuti, sembra che aumenti la frammentazione e la polarizzazione della società. Queste dinamiche informative e comunicative hanno conseguenze sistemiche ed estremizzanti su tutte le altre².

È urgente, dunque, teorizzare alcune possibili azioni, ispirate agli ideali che dovrebbero informare la risposta del sistema liberaldemocratico e promosse dalle istituzioni attraverso le leggi e politiche pubbliche, affinché le ICT, in particolare le odierne piattaforme digitali, mitigino il problema della polarizzazione online.

2. II PLURALISMO SECONDO RAWLS E ARENDT

Nel corso del Novecento, fondamentali teorici della democrazia liberale, come John Rawls e Hannah Arendt, si sono misurati con il problema del pluralismo.

Il problema centrale posto da Rawls in "Liberalismo politico" ([19]), è la presenza nella società contemporanea di differenti dottrine "comprehensive", ovvero visioni del mondo che gli individui credono essere ispirate da verità fondamentali, tanto

¹ Una prima versione del contributo, riguardante tre delle cinque proposte qui avanzate, è Giacomini G. (2020), «Pluralismo liberale e polarizzazione digitale. Una diagnosi e tre possibili terapie», in Taddio L., Giacomini G. (a cura di), *Filosofia del digitale*, Mimesis, Milano, 261-282.

² Dal momento in cui l'Internet delle piattaforme tende a "chiudere le comunità", soprattutto quelle altamente attive ([9]), e a polarizzare le loro posizioni ([18];[20]), il rischio è che si aggravi il conflitto fra posizioni circa problemi bioetici (ad esempio, pro-vax e no-vax), oppure fra diverse identità culturali (ad esempio, comunità islamiche nelle società occidentali).

divergenti e radicate da essere, almeno in parte, inconciliabili. Rawls cerca di dare risposta al “problema del pluralismo” con il “consenso per intersezione”. Secondo questo modello, persone con differenti concezioni possono comunque convergere su un sotto-insieme di valori e continuare, così, nella cooperazione sociale. La convergenza su alcuni valori politici “minimi” consentirebbe, negli auspici di Rawls, di tenere in adeguata considerazione il fatto del pluralismo senza mettere a repentaglio la stabilità del sistema democratico.

Perché il sistema rawlsiano regga, è necessario che lo spettro delle concezioni del bene interessate non sia eccessivamente ampio ([5]). L’ampia parte delle persone è ragionevole e l’essere credenti o atei, cristiani o musulmani, pro-vax o no-vax non è – nella grande maggioranza dei casi – un elemento che impedisce la cooperazione continuativa nel tempo. Tuttavia, il liberalismo politico rawlsiano rischia di essere messo in difficoltà dal pluralismo radicale: è determinante che gli individui abbiano una seppur minima predisposizione all’intesa. Un aumento della polarizzazione può essere un problema e perciò, prima che raggiunga certi livelli non più sostenibili, potrebbe essere opportuno individuare meccanismi che la limitino ([12]).

Anche Arendt ha affrontato il tema politico delle relazioni fra diversi. Secondo Arendt è necessario riprendere il senso politico della “vita activa” tipica delle città greche: per difendere la libertà e, allo stesso tempo, ricreare le condizioni di una vita politica di confronto fra pensieri in movimento, aperti e non chiusi ([1]). Il concetto di “zoon politikon”, sostiene Arendt, è stato frainteso: non esiste nell’individuo qualcosa di politico che appartiene naturalmente. L’essere umano è, piuttosto, in prima battuta, un essere a-politico. La politica nasce solo in seconda istanza, nello spazio che si trova tra gli individui, dunque in qualcosa di esterno a loro. Nella riflessione arendtiana, nota Magni ([17]: 24), la politica nasce in quello spazio intermedio che resta tra i soggetti, al netto di appartenenze, affiliazioni, credenze, e si pone come un irrinunciabile “in-between” tra esseri differenti.

L’allegoria del tavolo è particolarmente eloquente. Con questa metafora, Arendt intende rappresentare l’equilibrio che dovrebbe realizzarsi fra distacco (inteso come la capacità di sottrarsi criticamente a condizionamenti o interferenze) e appartenenza (non solo alla comunità politica ma anche ad un mondo fattuale comune). Vivere insieme significa che esiste un mondo di cose tra coloro che lo hanno in comune, come un tavolo posto tra quelli che vi siedono intorno. Il mondo, come ogni “in-between”, mette in relazione e separa gli uomini nello stesso tempo. Si tratta, in altre parole, di preservare la comunanza nella separatezza, e viceversa: quando non c’è pluralità di punti di vista sul mondo comune non c’è spazio pubblico, ma quando il mondo comune viene meno e rimangono solo visioni diverse ed inconciliabili, svanisce la possibilità concreta di stabilire legami politici ([6]).

3. PROFILAZIONE, PERSONALIZZAZIONE, POLARIZZAZIONE

Secondo la letteratura specialistica, sembra che nel mondo online sia più facile esporsi soltanto alle informazioni, alle parole, ai messaggi che gradiamo. Pariser ([18]) parla di *filter bubble*: ognuno si ritrova in una bolla culturale o ideologica in cui riceve solo informazioni che confermano ciò che crede. Altri, fra cui Sunstein ([20]) parlano di *echo chambers*, indicando un ambiente chiuso che riflette sé stesso, una camera di risonanza in cui ognuno trova ciò che più gli piace e incontra le persone che hanno gli stessi suoi interessi, con conseguente effetto di rafforzamento.

Le piattaforme web sono propense a favorire comportamenti di esposizione selettiva a causa della quantità potenzialmente enorme di informazioni che possono essere presentate ad ogni utente ([14]). Inoltre, le piattaforme hanno fini commerciali: sono bastate su un meccanismo economico basato sulla pubblicità e la raccolta dei dati personali, a loro volta basati sulla gestione dell’attenzione dell’utente. Maggiore è il tempo impiegato dagli utenti sulla piattaforma, meglio è per quest’ultima. La piattaforma, quindi, deve profilare e personalizzare ([23]).

Uno degli studi sugli effetti della “cura algoritmica” dei *social network* ha monitorato se la tecnologia di personalizzazione aumenta i clic e il tempo finalizzati a leggere articoli di politica allineati all’opinione dell’utente (e diminuisce i clic e il tempo finalizzati a leggere articoli disallineati) ([10]). Lo studio conclude che la tecnologia di personalizzazione è particolarmente efficace nel ridurre la dissonanza cognitiva, evitando informazioni non allineate con gli interessi, i gusti e i valori degli individui e che la personalizzazione (ad esempio tramite *newsfeed* dell’algoritmo di Facebook) potenzia la “chiusura” in bolle. Secondo altri studi, sembra che su Facebook gli utenti di un determinato profilo ideologico (nel caso americano, *liberals* e *conservatives*) abbiano pochi amici di un profilo opposto e condividono pochi contenuti con loro, ma che soprattutto quasi il 71% delle nuove informazioni presentate all’utente dal sistema, nel *newsfeed*, mostrano opinioni che si allineano all’ideologia dell’utente stesso ([7];[2]).

Quella della polarizzazione, per il pluralismo, è una meta-questione che incide a cascata su molte altre questioni sociali. Ad esempio, alcuni studi suggeriscono che l’utilizzo di Internet e dei *social media* può favorire la radicalizzazione religiosa su soggetti che sono “chiusi” nei gruppi colpiti dalla propaganda, ad esempio da parte dell’ISIS ([21]). Simili studi sono stati condotti sulle (non) relazioni fra i gruppi di utenti favorevoli alla medicina ufficiale e quelli che credono alla medicina

alternativa (i no-vax): secondo questi studi, il fenomeno della chiusura “a riccio” delle comunità online sui social è evidente ([2];[9]).

Fu Arendt a sottolineare quanto è importante la libertà – d’azione e di pensiero – per uscire da sé stessi, e per alimentare la politica democratica. Invece, l’utilizzo dei cookie, il funzionamento degli algoritmi di Google che permettono ricerche personalizzate, i suggerimenti dei post sul *newsfeed* di Facebook e le liste su Twitter sono tutti elementi che vanno nella direzione di spingere le persone a vivere in un ambiente online sempre più “cucito addosso”, fornendo quello che l’utente desidera e riducendo la possibilità di incontri casuali. Questa “accelerazione” della polarizzazione nel mondo virtuale può essere un problema significativo per la democrazia liberale pensata da Rawls e Arendt: le possibilità di “minimo comune linguaggio” pubblico rischiano di essere erose.

4. CINQUE PROPOSTE POLITICHE PER LE PIATTAFORME DIGITALI

Prendere sul serio la sfida del pluralismo polarizzato nella società contemporanea, acuita dalle scelte algoritmiche delle grandi piattaforme, significa abbandonare un approccio di *laissez faire*, per adottare misure ispirate da un approccio più interventista, ispirato al principio della divisione e dell’equilibrio dei poteri, capace di valorizzare il ruolo del potere pubblico. Soltanto il potere pubblico, infatti, può tutelare in ultima istanza alcuni principi democratici fondamentali, se diventa evidente che la mano “invisibile del mercato” non riesce a rispettarli ([15]). I problemi sociali possono essere esacerbati dalle tecnologie dell’informazione, ma possono anche essere mitigate dalle tecnologie opportunamente orientate dalle istituzioni democratiche. In conclusione, si propongono alcune idee che possano contribuire a promuovere il valore del pluralismo “non polarizzato”, per come è stato inteso da Arendt e Rawls.

Una prima azione potrebbe consistere nel *potenziamento della normativa sulla protezione dei dati personali* (chiamata anche “privacy”). Se è vero che la profilazione e la personalizzazione che “nutre” le *filter bubbles* è basata sulla capacità di raccogliere una grande quantità di dati, allora porre dei limiti alla “pesca a strascico” dei dati allenterebbe la precisione algoritmica con cui si offrono contenuti su misura. Una normativa privacy più stringente, indirettamente, potrebbe rendere più “porosi” i confini delle bolle informative. In questa direzione sembra andare il recente GDPR europeo.

Una seconda azione riguarda la *normativa antitrust*. Oggi abbiamo grandi concentrazioni nel web: spesso un unico gruppo di aziende domina un determinato settore. Ad esempio, il gruppo Facebook-Instagram-Whatsapp gestisce gran parte del traffico nell’ambito dei social network (simile discorso circa il gruppo Alphabet-Google per quanto riguarda il settore motori di ricerca). Come spiega Wu ([22]), la concentrazione economica, soprattutto nel caso delle aziende digitali di comunicazione, è un fatto rilevante dal punto di vista democratico. In particolare, sappiamo che la polarizzazione è connessa alla gestione di grandi masse di dati. Indebolendo i grandi gruppi attraverso azioni di antitrust porterebbe a raggiungere un doppio obiettivo. In primo luogo, la profondità dell’analisi diminuirebbe con la “separazione” aziendale delle banche dati. In secondo luogo, avremmo un numero maggiore di intermediari nell’ambito della sfera pubblica digitale, maggiore sperimentazione per quanto riguarda il modello di business, quindi più metodi nel gestire l’informazione. Potrebbe così emergere un modello alternativo a quello basato alla personalizzazione (e polarizzazione).

Una terza azione, suggerita da Floridi ([13]), potrebbe essere porre dei *limiti alla pubblicità online*, attenuando le finalità di tipo commerciale. Stabilendo per legge una quota di pubblicità da non superare per ogni impresa, si limiterebbero gli incentivi di sistema che pungolano le aziende digitali a puntare con forza sullo studio dei gusti degli utenti e su un’offerta tanto personalizzata da produrre *echo chambers* e *filter bubbles*. Le piattaforme dovrebbero cercare altri mezzi per mantenersi e per sopravvivere, ad esempio con l’offerta di informazioni o servizi aggiuntivi, che verrebbero pagate dagli utenti. Gli utenti, invece di beneficiare dell’odierno “regalo” di servizi in cambio della loro attenzione e dei loro dati, dovrebbero pagare per i servizi che intendono usare, e si orienterebbero verso quelli che ritengono migliori, senza la pressione della profilazione e della personalizzazione. Man mano che la pressione della pubblicità si riduce, gli utenti potrebbero vivere in un sistema in cui è possibile uscire dalla propria bolla con maggiore facilità.

Una quarta azione potrebbe consistere nell’obbligare le piattaforme online ad *offrire agli utenti una quota minima di “incontri casuali”*, soprattutto per quanto riguarda le informazioni di rilevanza pubblica. Lo stato (ad esempio, attraverso un’autorità indipendente) potrebbe promuovere questo obbligo e monitorare l’adempimento. Introdurre notizie e informazioni casuali sarebbe un adattamento algoritmico semplice per le piattaforme dei social media e salutare per il dialogo pluralista. In fondo, il legislatore italiano, nell’ambito del medium televisivo, si è già preoccupato di garantire che i cittadini ricevano un’informazione legata a opinioni e valutazioni politiche il più possibile plurale ed equilibrata (con la cosiddetta legge sulla “par condicio”). Un principio simile potrebbe essere applicato, con i debiti distinguo, per quanto riguarda il medium digitale.

Una quinta linea di azione, in questo caso promossa direttamente dalle istituzioni democratiche, potrebbe essere la *promozione di esperienze pubbliche di comunicazione online* ispirate al principio del confronto fra diversi, come prescritto

dalle teorie di Rawls e Arendt. In ambito civico e comunale, sono già numerosi gli esempi di piattaforme digitali di informazione, partecipazione e collaborazione ([8]). Si tratta, in genere, di siti che mettono in comune dati, notizie, progetti su cui la comunità può dialogare e confrontarsi con l'obiettivo di maturare decisioni di interesse generale il più possibile inclusive. Le intenzioni dei decisori pubblici che promuovono queste iniziative sono di creare ecosistemi digitali ispirati ai principi democratici, assumendo che le preferenze degli attori, attraverso il confronto fra punti di vista contraddittori, possano trasformarsi nel corso dell'interazione ([4];[16]).

Con queste e simili iniziative, ispirate dagli ideali che dovrebbero informare la risposta delle istituzioni liberaldemocratiche al fatto del pluralismo, si potrebbero avvicinare due obiettivi. In primo luogo le ICT, in particolare l'Internet delle piattaforme, sarebbero maggiormente in grado di abilitare una sfera politica abitata da un insieme di soggetti in relazione fra loro. In secondo luogo si potrebbe disinnescare l'acuirsi di conflitti su faglie sociali già problematiche di per sé, come quelle che riguardano il rapporto fra religioni, visioni politiche, scelte alimentari o sulla vita, che su Internet, a causa del fenomeno della polarizzazione e delle *echo chambers*, potrebbero acuirsi. Favorire la ragionevolezza rawlsiana e il "in-between" arendtiano, anche a seguito della rivoluzione digitale e telematica, significa aiutare i cittadini ad entrare in rapporto dialogico tra loro, rendendo conciliabile la manifestazione della pluralità con la collaborazione.

BIBLIOGRAFIA

- [1] Arendt, Hannah. 1958. *The human condition*. University of Chicago Press.
- [2] Bakshy, Eytan, Solomon Messing, e Lada Adamic. 2015. «Exposure to Ideologically Diverse News and Opinion on Facebook». *Science* 348: 1130–32.
- [3] Bessi, Alessandro, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, e Walter Quattrociocchi. 2015. «Science vs conspiracy: Collective narratives in the age of misinformation». *PloS one* 10 (2).
- [4] Blondiaux, Loïc. 2008. *Le nouvel esprit de la démocratie*. La république des idées. Seuil.
- [5] Bohman, James. 1995. «Public reason and cultural pluralism». *Political theory* 23 (2): 253–79.
- [6] Canovan, Margaret, e Hanna Arendt. 1992. *A reinterpretation of her political thought*. Cambridge University Press.
- [7] Colleoni, Elanor, Alessandro Rozza, e Adam Arvidsson. 2014. «Echo Chamber or Public Sphere?» *Journal of Communication* 64 (2014): 317–32.
- [8] De Blasio, Emiliana. 2018. *Il governo online*. De Blasio.
- [9] Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, e Walter Quattrociocchi. 2016. «The spreading of misinformation online». *PNAS* 113 (3): 554–59.
- [10] Dolgov, Dylko, Igor Dolgov, William Hoffman, e et al. 2017. «The dark side of technology». *Computer in Human Behavior* 73: 181–90.
- [11] Ferrara, Alessandro. 2012. «Hyper-pluralism and the multivariate democratic polity». *Philosophy & Social Criticism* 38 (4–5): 435–44.
- [12] Gaus, Gerald. 1999. «Reasonable pluralism and the domain of the political». *Inquiry* 42 (2): 259–84.
- [13] Giacomini, Gabriele. 2018. *Potere digitale*. Meltemi.
- [14] Iyengar, Shanto, e Kyu Hahn. 2009. «Red media, blue media: evidence of ideological selectivity in media use». *Journal of Communication* 59 (1): 19–39.
- [15] Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money*. Palgrave Macmillan.
- [16] Loader, Brian, e Dan Mercea. 2011. «Networking democracy? Social media innovations and participatory politics». *Information, Communication & Society* 14 (6): 757–69.
- [17] Magni, Beatrice. 2018. «My thinking is my fighting». *Biblioteca della libertà* 53 (223): 3–47.
- [18] Parisier, Eli. 2011. *The Filter Bubble*. Penguin.
- [19] Rawls, John. 1993. *Political liberalism*. Columbia University Press.
- [20] Sunstein, Cass. 2017. *#Republic. Divided Democracy in the Age of Social Media*. Princeton University Press.
- [21] Wakeford, Laura, e Laura Smith. 2020. «Islamic State's Propaganda and Social Media». In *ISIS Propaganda: A Full-Spectrum Extremist Message*, Baele, S.J., Boyd, K.A., Coan, T.G. Oxford University Press.
- [22] Wu, Tim. 2020. *La maledizione dei giganti*. Il Mulino.
- [23] Zuboff, Shoshana. 2019. *The age of surveillance capitalism*. Profile Books.

A Taxonomy of Depictive Representations: From Paintings and Sculptures to Virtual Reality

Enrico Terrone

Università di Genova – enrico.terrone@unige.it

ABSTRACT

An account of depictive representations or depictions should satisfy two basic desiderata. First, to distinguish depictions from linguistic representations. Second, to figure out the differences among different kinds of depictive representations, thereby providing a compelling taxonomy. In §1 I show how some popular accounts of depiction satisfy the former desideratum, while in §2 I point out some difficulties they may face when it comes to satisfy the latter. In §3 I propose an account that has the potential to satisfy both the desiderata. In §4 I show how it does so.

KEYWORDS

Virtual reality, depictive representations, taxonomy.

TALK

1. DISTINGUISHING DEPICTIONS FROM LINGUISTIC REPRESENTATIONS

The term “representation” usually designates both representational states of the mind and representational artifacts. Here, I use the term “representation” only in the latter sense, conceiving of a representation as an artifact X that has the function of directing the user’s mind not only toward X itself but also toward some other thing Y. Since the user’s mind is directed towards two things, we might state that a representation is an artifact that has the function of putting the user in a state of double intentionality, that is, a complex mental state constituted by an intentional component CX directed toward X and an intentional component CY directed toward Y¹. I will elaborate on the features of CX and CY later. For the time being, I am just assuming that the suitable user of a representation is aware in some sense of both X and Y.

The divide between depictive and linguistic representations can be drawn in at least three different ways. First, by stressing the resemblance between X and Y in depictions as opposed to the conventionality of linguistic meaning ([9]). Second, by highlighting differences between representational structures ([4]). Third, by considering the different kinds of mental states that linguistic and depictive representations are meant to elicit from their users ([13];[10];[12]). This is the “user-oriented” approach to depiction which I will endorse in this paper.

Accounts such as Richard Wollheim’s and Kendall Walton’s are dubbed “experiential” since they cast depictions as eliciting peculiar *experiences* from users, while accounts such as Flint Schier’s are dubbed “recognitional” since they cast depictions as eliciting higher-level *states of recognition* from users. For Wollheim, a depictive representation X of Y is aimed to elicit a twofold experience which consists of two simultaneous perceptual components: CX and CY. For Walton, X is meant to elicit a complex imaginative experience in which the user imagines, of her perceptual experience CX, that this is another perceptual experience CY. For Schier, X is meant to trigger the user’s capacity of visually recognizing Y. Consider for example a painting of a mountain. For Wollheim, it makes one simultaneously perceive the colored surface and the mountain depicted. For Walton, it makes one imagine of one’s perception of the colored surface that this is a perception of the mountain. For Schier, it triggers one’s capacity of visually recognizing mountains.

All these accounts succeed in distinguishing depictions from linguistic representations. Wollheim casts depictions as involving a perceptual experience of both X and Y whereas linguistic representations are not meant to elicit a perceptual experience of Y. Walton casts depictions as involving imagining of one’s perception of X that this is a perception of Y whereas linguistic representations at most involve perceiving X and imagining Y. Schier also properly draws the distinction since linguistic representations do not trigger the user’s capacity of *visually* recognizing the thing represented.

¹ In principle, the complex mental state might reveal itself to be a pair of interconnected mental states ([5]), or even a pair of alternating mental states ([3]). Here, I am staying neutral on this issue.

2. DISTINGUISHING AMONG DEPICTIVE REPRESENTATIONS

An account of depiction should draw not only the distinction between depictive and linguistic representations but also illuminating distinctions among depictive representations. The latter desideratum has been quite underestimated so far.

By explaining depiction in terms of the user's capacity of visual recognition, accounts such as Schier's fall short of differentiating kinds of depictive representations since the triggered recognitional capacity remains the same despite changes in the triggering representation ([6]: 173). Experiential accounts, on the other hand, have more room for maneuver since they can differentiate kinds of depictive representations in terms of the different visual experiences that can be elicited. For instance, Walton ([12]: 227) draws a sensible distinction between two-dimensional pictures and sculptures in terms of the different imaginings they elicit.

Still, I contend, there is a kind of depictive representations that conceptions such as Wollheim's and Walton's find it hard to accommodate, namely, virtual reality. The reason is that both those conceptions presuppose that the user has a perceptual experience of *X*. For Wollheim, the specificity of depictions is the simultaneous perceptual experience of *X* and *Y*. For Walton, it is imagining of the perceptual experience of *X* that this is a perceptual experience of *Y*. If we assume that the specificity of virtual reality consists in providing the user with a perceptual experience of *Y* in the absence of a perceptual experience of *X*, neither Wollheim's account nor Walton's, as they stand, can accommodate this kind of depictive representations.

A supporter of Wollheim's or Walton's might reply that virtual reality scenarios are not depictive representations and thus an account of depiction is not meant to accommodate them. Wollheim ([13]: 32) seems to favor such a strategy when he states that trompe l'oeil paintings, which are somehow the ancestors of virtual reality, are not pictures. Still, at least some works of virtual reality are appreciated as depictions in our cultural practices. Alejandro González Iñárritu's *Flesh and Sand* (*Carne y Arena*), for instance, is a work of virtual reality which depicts immigrants at the border between Mexico and the United States. This classification helps to explain why *Flesh and Sand* premiered at the 2017 Cannes Film Festival as part of the official selection: it depicts immigrants just as a film might have done. If this is right, an experiential account of depiction which makes room for virtual reality is preferable to those that exclude it. In what follows, I will propose such an account, offering an abductive argument for it. I will show that the proposed account does not only draw the proper distinction between depictions and linguistic representations but also enables us to distinguish among kinds of depictive representations in an elegant and compelling manner.

3. THE "PERCEPTION ONLY" ACCOUNT

Experiential accounts such as Wollheim's and Walton's state that the appropriate response to a depictive representation *X* of *Y* is a perceptual experience of *Y* plus some further experiential component: for Wollheim, this is the simultaneous perceptual experience of *X* while, for Walton, the perceptual experience of *Y* is to be nested into the scope of an imaginative experience concerning the perceptual experience of *X*. Both these accounts can be dubbed "Perception Plus..." since they individuate the specificity of depictions in the perceptual experience of the thing represented plus some further experiential component involving the representation itself. A significant portion of the debate about depiction aims to establish which is the best way to characterize such further component. The account I am proposing, on the other hand, denies that any further component is required. Further experiential components can contribute to distinguish between kinds of depictions, but the perceptual experience of the thing represented is all we need to properly characterize depictive representations. That is why I dub this account "Perception Only".

The distinction between depictions and linguistic representations can be drawn by Perception Only in the same way as Wollheim's account does so, that is, by stressing that, in depiction, *CY* involves a perceptual experience of *Y*. The novelty of Perception Only is that there is no commitment to the nature of *CX*. Although many depictive representations elicit a state of double intentionality in which both *CY* and *CX* are perceptual experiences, there can be depictive representations such as works of virtual reality that elicit a state of double intentionality in which only *CY* is a perceptual experience while *CX* comes down to the merely cognitive awareness that there is an artifact *X* which is the source of *CY*. The basic form of this awareness consists in grasping a definite description such as "the technological device that caused this *CY* experience". More competent users might fill the description with details concerning the hardware and software components of the system producing the virtual reality experience. Yet, the mere grasping of the description is enough to constitute the *CX* component of the state of double intentionality whose *CY* component is the enjoyment of the virtual scenario.

All this shows that Perception Only can cast virtual reality as a depictive representation. In the remainder of the paper, I will deploy Perception Only to offer a taxonomy of depictive representations which ranges from paintings and sculptures to virtual reality.

4. THREE KINDS OF DEPICTIVE REPRESENTATIONS

While the perceptual nature of CY is essential to depictions, the nature of CX is a variable parameter that enables us to individuate three basic kinds of depictive representations, namely, K1, K2, and K3.

In K1, whose paradigm is the two-dimensional picture, CX is a perceptual experience of X in the viewer's egocentric space, that is, a framework "centering in the subject's body" and having as its coordinates «"up" and "down", "left" and "right", and "in front of" and "behind"» ([2]: 153). CY, instead, is a perceptual experience of Y in a pictorial space distinct from the viewer's egocentric space. The viewer of a picture experiences X as an object in front of both her sight and her body while she experiences Y as an object in front of her sight but not in front of her body. CY, just like CX, is perspectival in virtue of involving an unseen point, namely, the point of view, in relation to which everything seen is presented. Yet, CY, unlike CX, is not egocentric since that point does not match the viewer's body: the point remains the same even if the viewer moves.

Drawings, paintings, and photographs exemplify K1 as still depictions, while films exemplify it as temporally extended depictions. In looking at a painting, one has a perceptual experience of Y while being perceptually aware of X as the painting's surface in one's egocentric space. Likewise, in watching a film, one has a perceptual experience of Y while being perceptually aware of X as the enlightened screen in one's egocentric space ([7]).

Scholars have offered conflicting accounts of CX and CY for K1 depictions². I will not take stance on this issue. I limit myself to characterizing K1 as involving both a perceptual awareness of X in the viewer's egocentric space and a perceptual experience of Y in a distinct perspectival space, regardless of whether these two experiences are simultaneous or alternate. In K2, whose paradigm is the sculpture, both CX and CY are perceptual experiences in the viewer's egocentric space. While in K1 there is a sharp phenomenal contrast between CX and CY, in K2 they are almost indistinguishable at the perceptual level. In both CX and CY, the viewer perceives a three-dimensional object in her egocentric space. The difference between CX and CY lies in the concept the viewer applies to her perception, namely, the concept of an X in CX and the concept of an Y in CY. For example, in watching the statue of a lion, the viewer applies to her perception the concept "statue" in CX and the concept "lion" in CY. Alternating the application of the two concepts, thereby alternating CX and CY, can prevent our perceptual experience from inconsistently locating two distinct things, the statue and the lion, at the same place of our egocentric space³.

Although the viewer *perceives* both the statue and the lion as objects in her egocentric space, replacing the concept "statue" with "lion" may enable her to *imagine* that the lion has its place in a space different from hers ([12]: 227; [6]: 278). Still, this is just a higher-level imaginative supplementation. At the perceptual level, the sculpted lion, unlike the painted lion, is experienced as having its place in the viewer's egocentric space, as suggested by the viewer's sense of being able of making a tour of inspection of the sculpted lion but not of the painted lion. Interestingly, all this contributes to explain why statues (and more generally sculptures in the round), which do enable the tour of inspection, are central instances of sculpture while works of relief sculpture, which do not enable it, «seem to occupy a region between sculpture and painting, and that is how, in general, tradition has received them» ([6]: 287).

Theater is the temporally extended correlate of three-dimensional depictions such as sculptures ([13]: 151), just as film is the temporally extended correlate of two-dimensional depictions such as drawings, paintings, and photographs. Indeed, the proper correlate of sculpture is puppetry rather than live theater since both sculpture and puppetry involve the replacement of "sortal concepts" which enable us to single out the particulars they subsume ([11]: 168): just as the viewer of a sculpture can replace the concept "statue" in CX with "lion" in CY, the viewer of a puppet show can replace "puppet" in CX with "lion" in CY. Live theater, on the other hand, usually involves the application of the same sortal concept, namely "person", both in CX and CY. The relevant shift, here, concerns the application of an "individual concept" (or "mental dossier", see [2]: 276): the viewer of a play replaces the individual concept of a certain actor in CX with that of a certain character in CY. Nevertheless, live theater elicits experiences of same sort as those elicited by sculpture and puppetry, since both the actor in CX and the character in CY are perceived as having their place in the viewer's egocentric space. It is only at a higher imaginative level that the viewer casts the space of the stage as detached from her egocentric space. In film experience, instead, the space portrayed is immediately given in perceptual experience as detached from the viewer's

² As seen earlier, for Wollheim ([13]) CX and CY are meant to be simultaneous perceptual experiences while for Walton ([12]) they are meant to be in the scope of a complex imaginative experience. Ernst Gombrich's ([3]) account differs from Wollheim's in conceiving of CX and CY as alternate perceptual experiences instead of as simultaneous ones. Variants of these accounts have been more recently proposed (for a survey, see Hyman and Bantinaki [8]). Among those, Robert Briscoe's ([1]) is perhaps the closest to "Perception Only" since it does not presuppose that CX is a perceptual experience. Still, Briscoe characterizes CY as an experience of a "virtual model" rather than of the thing represented.

³ For Wollheim ([13]: 150-151) CX and CY are meant to be simultaneous in our experience of pictures, but they might alternate in our experience of sculptures.

egocentric space. That is why theater makes room for interactions (exploited especially by experimental theater) between characters and the audience which are not available to film.

In K3, whose paradigm is virtual reality, only CY is a perceptual experience while CX is a merely cognitive awareness. In CY, the viewer experiences Y in a space which involves not only perspective but also dependence of the point of view on the viewer's bodily movements in a way that matches ordinary perception. For example, if the viewer turns her head the perspective changes in the way it would change in ordinary perception⁴.

On the one hand, K3 resembles K2 and differs from K1, in which the viewer experiences Y in a perspectival space whose point of view does not depend on her bodily movements. On the other hand, K3 also differs from K2, in which both X and Y can be experienced as objects in the viewer's egocentric space. In K3, Y is the only object of perception. Just as virtual reality exemplifies K3 as a temporally extended representation, trompe l'oeil exemplifies it as a still representation.

Interestingly, these three kinds of depictive representations match three venerable conceptions of images in the philosophical tradition. K1 matches the Aristotelian conception of images as involving the experience of something absent: pictures such as paintings or films provide us with a perceptual experience of things that have their place in a space that we experience as detached from ours. K2 matches the Kantian conception of images as involving the application of concepts to the manifold of sensation: sculpture and theater encourage us to replace the concepts that *should* be applied to certain objects perceivable in our surroundings to properly categorize them with other concepts which *might* be applied to those objects to enjoy a rewarding experience of them. K3 matches the Platonic conception of images as essentially involving deception: trompe l'oeil and virtual reality make us experience things in our surroundings that are different from the things that actually surround us. The proposed taxonomy reveals that these three conceptions ultimately correspond to three basic kinds of depictive representations, which are all meant to elicit a perceptual experience of the things represented but differ as for how that experience might be related – or not related – to the experience of the representation itself⁵.

REFERENCES

- [1] Briscoe, Robert. 2016. "Depiction, Pictorial Experience, and Vision Science." *Philosophical Topics* 44 (2): 43–82.
- [2] Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Clarendon Press.
- [3] Gombrich, Ernst. 1960. *Art and Illusion*. London: Phaidon.
- [4] Goodman, Nelson. 1968. *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis: Bobbs-Merrill.
- [5] Hopkins, John. 2010. "Inflected Pictorial Experience: Its Treatment and Significance." In *Philosophical Perspectives on Depiction*, Abell C. and Bantinaki K., 151–80. Oxford: Oxford University Press.
- [6] Hopkins, Robert. 2003. "Sculpture and Space." In *Imagination, Philosophy and the Arts*, Lopes D. and Kieran M., 272–90. London: Routledge.
- [7] ———. 2009. "Depiction." In *The Routledge Companion to Philosophy and Film*, Livingston P. and Plantinga C., 64–74. London: Routledge.
- [8] Hyman, John, and Katerina Bantinaki. 2017. "Depiction." In *The Stanford Encyclopedia of Philosophy*, Zalta. <https://plato.stanford.edu/archives/sum2017/entries/depiction/>.
- [9] Pierce, Charles Sanders. 1931. *Collected Papers of Charles Sanders Pierce*. 1-6. Cambridge: Harvard University Press.
- [10] Schier, Flint. 1986. *Deeper into Pictures*. Cambridge: Cambridge University Press.
- [11] Strawson, Peter Frederick. 1959. *Individuals*. London: Methuen.
- [12] Walton, Kendall. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA: Harvard University Press.
- [13] Wollheim, Richard. 1980. *Art and Its Objects: An Introduction to Aesthetics*. 2nd ed. New York: Harper and Row.
- [14] ———. 1987. *Painting as an Art*. Princeton: Princeton University Press.

⁴ One might wonder whether the space experienced in virtual-reality's CY can be subsumed under the notion of egocentric space. My intuition goes in this direction, but here I do not need to argue for this. The dependence of the point of view on the viewer's bodily movements is enough to differentiate the CY of virtual reality from that of K1 depictions such as films.

⁵ I see this taxonomy as a framework favoring further research. For instance, one might investigate which specific sorts of aesthetic appreciation are afforded by each kind of depictive representations; or add further dimensions to the taxonomy, distinguishing between causal representations such as photographs and intentional representations such as paintings; or even extend the taxonomy beyond the visual domain, considering for instance auditory depictions. Still, all these lines of inquiry shall rely on the core taxonomy proposed in this paper.

Paesaggi dell'incontro mediale on-demand

Nicolò Villani

Università Telematica eCampus, Italia – nv.nico.villani@gmail.com

ABSTRACT

Il problema dell'incontro mediale all'interno del contesto delle piattaforme streaming on-demand pone al centro la questione della soggettività tanto dello spettatore/utente che sceglie e seleziona il prodotto da vedere all'interno del catalogo quanto della piattaforma stessa che riorganizza sé stessa e il proprio arredo per rinnovare continuamente il dialogo con il proprio utente. Una relazione reciproca, intersoggettiva, che genera una formazione di senso in continuo mutamento riscontrabile attraverso l'homepage stessa della piattaforma, vera e propria immagine dell'incontro avvenuto e continuamente sul punto di rinnovarsi, teso verso una pratica da perpetuare. In questa pratica entrano in gioco, al fianco della relazione intersoggettiva, due elementi strettamente fenomenologici: il tocco come veicolo preferenziale dell'incontro e il "paesaggio" come dispositivo teorico di comprensione dell'immagine generata dall'incontro stesso. Partendo dai risultati dei *Media Studies* in rapporto alle pratiche produttive, distributive e di fruizione che circondano il contesto delle piattaforme streaming e dalla comprensione che i *Platform Studies* permettono di avere riguardo le infrastrutture che sottendono queste pratiche, si applicano qui il metodo etnosemiotico e lo sguardo fenomenologico sui problemi teorici immanenti a queste pratiche, individuando e descrivendo il luogo teorico soggiacente all'efficacia dell'incontro mediale, per ottenere descrizioni operative del funzionamento delle piattaforme streaming.

PAROLE CHIAVE

Media environment, media studies, phenomenology, platform studies, subjectivity.

INTERVENTO

1. INTRODUZIONE

Nel contesto postmediale ([15]) la fruizione audiovisiva – sempre più declinata attraverso la dinamica on-demand ([34]) – si ibrida e si fonde con lo spazio che circonda lo spettatore, arrivando a estendersi capillarmente ad ogni ambito del suo intorno ([11];[9]). In questo, le piattaforme streaming si pongono come porte d'accesso totalizzanti di luoghi che, attraverso le loro infrastrutture ([22]), puntano a farsi per l'utente paesaggi, trascendendo quella superficie di spazio virtuale che la loro interfaccia suggerisce ([12]) e formando con la loro partecipazione attiva all'esperienza mediale dell'utente la sua stessa identità come soggetto spettatoriale. Questa partecipazione si compone di pratiche che hanno nella dimensione tattile un canale preferenziale, attraverso un "tocco" che è tanto azione percettiva quanto strumento di modifica del mondo ([2]). Il soggetto spettatoriale si scopre quindi attore di un dialogo la cui formazione discorsiva è da indagare attraverso una metodologia che comprenda tanto la dimensione fenomenologica dell'evento mediale quanto l'esplorazione etnografica della pratica di visione ([24];[25]). Il risultato sarà la fotografia di un incontro ([19]) intersoggettivo in cui i due soggetti coinvolti – piattaforma e spettatore – confondono i loro ruoli nella generazione di un'immagine di senso condiviso, producendo nel loro incontrarsi continuamente rinnovato un paesaggio arredato e organizzato in armonia e in tensione con le pratiche che vi si consumano.

2. OGGETTI, PROBLEMI E METODOLOGIE

Questo intervento si vuole porre nel solco delle ricerche dei *Media Studies* tradizionali e dei *Platform Studies* cercando di dialogare con i risultati dei rispettivi dibattiti per fornire come contributo l'inizio di un'analisi teorica profonda che possa farsi descrizione delle pratiche soggiacenti alla fruizione mediale che queste discipline osservano dal punto di vista produttivo ([13];[23],[10]), distributivo ([31];[7];[30];[33]) e infrastrutturale ([29];[3];[28];[22];[8]). Partendo dalle ricerche di settore di stampo economico-etnografico già esistenti, che mappano e quantificano tanto le pratiche delle audience quanto le loro forme esperienziali rispetto a particolari prodotti mediali ([14];[30];[16]), si propone qui di introdurre un approccio etnosemiotico ([24]) in grado di restituire al dibattito la dimensione della soggettività relazionale ([21]) presente nel contesto della fruizione on-demand messa in forma dal funzionamento delle moderne piattaforme OTT; la metodologia etnosemiotica permette infatti di tradurre oggetti, pratiche e spazi in conformazioni testuali interpretabili,

in grado di restituire all'analisi elementi e strumenti di lettura altrimenti limitati al loro funzionamento tecnico puro¹. Facendo largo uso dei principi propri della fenomenologia moderna ([25];[2];[27]), la metodologia etnosemiotica, ricostruendo le pratiche, gli oggetti e gli ambienti mediali, consente di descrivere come la relazione intersoggettiva tra spettatore e piattaforma prenda forma, avvicinandosi a ciò che François Jullien ha descritto come "paesaggio" ([18]): è Jullien stesso a ricostruire un'attenta descrizione dell'incontro a partire dall'idea di "paesaggio" come luogo dell'intersoggettività in grado di produrre effetti di senso percepibili ([19]). Gli oggetti descritti e individuati da *Media e Platform Studies* – tra tutti, in particolare, gli algoritmi di raccomandazione e il loro funzionamento – diventano quindi qui elementi di una relazione intersoggettiva da indagare con gli strumenti di un'etnografia di stampo fenomenologico, capace di restituire le condizioni profonde dell'incontro mediale che avviene all'interno delle piattaforme streaming, basato sul costante bisogno del *rinnovo* – di visione, di permanenza, di abbonamento, di discorso – perpetuo e pervasivo, in modo da ottenerne descrizioni operative chiare, efficaci e comparabili tra i diversi soggetti del mercato mediale.

3. OBIETTIVI E PROSPETTIVE

Nel rendere conto, attraverso la nozione di "paesaggio mediale" ([5]) della relazione intersoggettiva che si sottende alle pratiche di visione on-demand, si vuole qui restituire un modello teorico e descrittivo generale dell'azione di riorganizzazione dell'home page della piattaforma come immagine significativa risultante dall'incontro avvenuto tra i due soggetti: piattaforma e spettatore. La speranza è quella di restituire la nozione di "soggetto" al cuore del dibattito dei *Media Studies*, a nostro avviso prettamente volti verso un determinismo intersoggettivo che dimentica la dimensione relazionale sottesa alle pratiche di scelta, visione, discorso e condivisione dei prodotti mediali. Questo tipo di ricerca si propone come modello da applicare alle specifiche piattaforme come casi di studio per declinare le diverse proporzioni della presenza e dell'efficacia relazionale al loro interno, ricostruendo metafore analitiche e descrizioni operative che si diano come lenti d'osservazione per comprendere gli effetti di senso del funzionamento tecnico di questi contesti mediali². In questo, si aprono due strade problematiche per l'indagine: da una parte, il ruolo cruciale dell'"inaudito" ([20]) in tutte le sue varie declinazioni come motore del rinnovo relazionale all'interno del paesaggio mediale e dall'altra gli aspetti gerarchici che incorrono tra le soggettività nel momento in cui si passa da un contesto di mercato concorrenziale puro a uno di servizio pubblico ([32];[36]).

4. RINGRAZIAMENTI

Si ringrazia Francesco Marsciani e il suo seminario sui fondamenti della teoria della significazione per avermi permesso più volte di confrontarmi sui problemi della soggettività ricavata dallo sguardo etnosemiotico; con lui François Jullien per la limpidezza di pensiero attraverso cui si scoprono nuovi lati delle proprie esperienze nel mondo.

BIBLIOGRAFIA

Ove presenti, si fa riferimento alle edizioni italiane.

- [1] Accardo, Lorenza, et al. s.d. *Via Mascarella. Declinazioni di uno spazio denso*. 2015: Esculapio.
- [2] Amoroso, Prisca. 2016. *Corpo, linguaggio e senso: tra semiotica e filosofia*. Esculapio.
- [3] Apperley, Thomas, e Jussi Parikka. 2018. «Platform Studies' Epistemic Threshold». *Games and Culture* 13 (4).
- [4] Avezzi, Giorgio. 2019. «I fantasmi nella macchina. Dati di comportamento e raccomandazioni personalizzate». *Fata Morgana* 36: 85–100.
- [5] Barberis, Mauro. 2021. *Ecologia della rete*. Mimesis.
- [6] Barra, Luca. 2015. *Palinsesto: storia e tecnica della programmazione televisiva*. Laterza.
- [7] Baschiera, Stefano, Francesco Di Chiara, e Valentina Re. 2018. *Re-intermediation: distribution, online access, and gatekeeping in the digital European market*. Mimesis.
- [8] Bloom, Andrew. 2021. *Tubi: viaggio al centro di internet*. Minimum Fax.
- [9] Bolter, J. David. 2020. *Plenitudine digitale: il declino della cultura d'élite e lo scenario contemporaneo dei media*. Minimum Fax.
- [10] Brembilla, Paola. 2018. *It's all connected: l'evoluzione delle serie TV statunitensi*. Franco Angeli.
- [11] Casetti, Francesco. 2015. *La galassia Lumière: Sette parole chiave per il cinema che viene*. Bompiani.
- [12] Coviello, Massimiliano, e Valentina Re. 2020. «"Continua a guardare". Strategie enunciative nelle piattaforme di video on demand». *E|C Rivista dell'Associazione Italiana di Studi Semiotici* 14 (30).

¹Per rendere conto del funzionamento e dell'efficacia del metodo etnosemiotico applicato, oltre al già citato ([24]), si vedano ([1];[26]).

²Alcuni risultati di questo approccio sono stati presentati all'Artes Spring Seminar (2021) dell'Universidade Católica Portuguesa con l'intervento *The Haunting of Netflix Home* e al convegno sulle forme della spazialità (2021) organizzato dall'Università di Bergen con l'intervento dal titolo *La Piattaforma delle meraviglie*.

- [13] Di Chio, Federico. 2017. *Mediamorfosi 2. Industrie e immaginari dell'audiovisivo digitale*. RTI.
- [14] Eugeni, Ruggero. 2010. *Semiotica dei media: le forme dell'esperienza*. Carocci.
- [15] ———. 2015. *La condizione postmediale: media, linguaggi e narrazioni*. La Scuola.
- [16] Hill, Annette. 2019. *Esperienze mediali: dalle serie tv al reality*. Minimum Fax.
- [17] Jenkins, Henry. 2014. *Cultura convergente*. Maggioli.
- [18] Jullien, François. 2017. *Vivere di paesaggio o l'impensato della ragione*. Mimesis.
- [19] ———. 2020. *L'apparizione dell'altro: lo scarto e l'incontro*. Feltrinelli.
- [20] ———. 2021. *L'inaudito. All'inizio della vita vera*. Feltrinelli.
- [21] Landowski, Eric. 2010. *Rischiare nelle interazioni*. Franco Angeli.
- [22] Lobato, Ramon. 2020. *Netflix nations: geografia della distribuzione digitale*. Minimum Fax,.
- [23] Lotz, Amanda D. 2017. *Post network: la rivoluzione della tv*. Minimum Fax.
- [24] Marschiani, Francesco. 2007. *Tracciati di etnosemiotica*. Franco Angeli.
- [25] ———. 2012. *Ricerche semiotiche I: Il tema trascendentale*. Esculapio.
- [26] ———. 2021. *Un etnosemiologo nel museo*. Esculapio.
- [27] Merleau-Ponty, Maurice. 2019. *Fenomenologia della percezione*. Bompiani.
- [28] Moschini, Ilaria. 2018. «Social semiotics and platform studies: an integrated perspective for the study of social media platforms». *Social Semiotics* 28 (5): 623–40.
- [29] Parks, Lisa, e Nicole Starosielski. 2015. *Signal traffic: critical studies of media infrastructures*. University of Illinois Press.
- [30] Pescatore, Guglielmo. 2018. *Ecosistemi narrativi: dal fumetto alle serie tv*. Carocci.
- [31] Re, Valentina. 2017. *Streaming media: distribuzione, circolazione, accesso*. Mimesis.
- [32] Scaglioni, Massimo. 2016. *Il servizio pubblico televisivo: morte o rinascita della RAI?* Vita e Pensiero.
- [33] Smith, Michael D., e Rahul Telang. 2019. *Streaming, sharing, stealing: i big data e il futuro dell'intrattenimento*. Minimum Fax.
- [34] Tryon, Chuck. 2017. *Cultura on demand: distribuzione digitale e futuro dei film*. Minimum Fax.
- [35] Vespignani, Alessandro, e Rosita Rijtano. 2019. *L'algoritmo e l'oracolo: come la scienza predice il futuro e ci aiuta a cambiarlo*. Il Saggiatore.
- [36] Zaccaria, Roberto. 2019. *RAI: il diritto e il rovescio: il servizio pubblico oggi*. Passigli.

Sessione Contenuti 2
Marshall McLuhan

Tra Public e Digital History: la soluzione ibrida dei registri parrocchiali di Monterosso on line

Chiara D'Anzi¹, Vittore Casarosa², Enrica Salvatori³

¹Università di Pisa, Italia – c.danzi@studenti.unipi.it;

²CNR-ISTI, Pisa, Italia – casarosa@isti.cnr.it;

³Università di Pisa, Italia – enrica.salvatori@unipi.it

ABSTRACT

I registri parrocchiali rappresentano la prima forma nota di registro anagrafico in Europa e, come tali, vanno a costituire un patrimonio di grande importanza per lo studio delle comunità di età moderna e contemporanea dal punto di vista storico, demografico e genealogico. I registri, infatti, sono libri manoscritti – o misti con parte di testo a stampa – che raccolgono i dati relativi a battesimi, matrimoni e funerali officiati in una parrocchia.

I registri parrocchiali di Monterosso al Mare (La Spezia) – battesimo, matrimonio e morte – vanno dal 1580 fino agli anni '80 del XX secolo, e sono stati fotografati e trascritti da un cittadino della comunità di Monterosso. Il signor Gustavo Moggia ha quindi reso disponibile un totale di circa 35.000 immagini (in cui sono comprese quelle con i singoli atti e quelle con *recto* e *verso* delle pagine dei registri) e circa 30.000 trascrizioni testuali. In questo articolo presentiamo la piattaforma web sviluppata per mettere questi dati a disposizione della comunità di Monterosso e di tutti gli studiosi, e una prima loro analisi.

PAROLE CHIAVE

Registri parrocchiali, crowdsourcing, public history, statistiche demografiche.

INTERVENTO

1. INTRODUZIONE

La piattaforma web “Registri parrocchiali di Monterosso” (<https://registrimonterosso.labcd.unipi.it/>), attiva dall'aprile 2021, rende possibile la ricerca di dati anagrafici entro queste fonti peculiari appartenenti alla parrocchia di Monterosso al Mare (SP), riguardanti un ampio periodo cronologico, dal 1580 agli anni '80 del XX secolo circa. I registri parrocchiali, com'è noto, raccolgono i dati relativi a battesimi, matrimoni e funerali officiati in una parrocchia e rappresentano la prima forma di registro anagrafico in Europa. Come tali, costituiscono un patrimonio immenso, di cui non ne è mai stata calcolata l'estensione¹, di enorme importanza per lo studio delle comunità di età moderna e contemporanea, dal punto di vista storico demografico, genealogico, sociale e culturale ([3];[2]).

Il loro trattamento digitale, tuttavia, è particolarmente problematico, sia per la quantità dei dati che contengono e sia per i formati e le caratteristiche diverse che presentano nel tempo: da testi inizialmente solo manoscritti e con formulario relativamente libero a registri prestampati con moduli, che il parroco doveva completare a mano (si veda la Fig. 1). Si tratta quindi di fonti che contengono dati semistrutturati, difficili da estrarre e passibili di trattamenti diversi. Nel caso che si presenta i registri sono stati fotografati e digitalizzati in anni recenti da un membro della comunità di Monterosso al Mare, il signor Gustavo Moggia, il quale ha anche trattato il materiale elaborando circa 35.000 immagini e operando 30.000 trascrizioni testuali. Questo materiale è andato a formare la struttura principale sulla quale è stata costruita la piattaforma, poi aperta alla collaborazione del pubblico.

¹ Rari i tentativi di censimento e sempre settoriali: si veda ad esempio ([5];[1]).

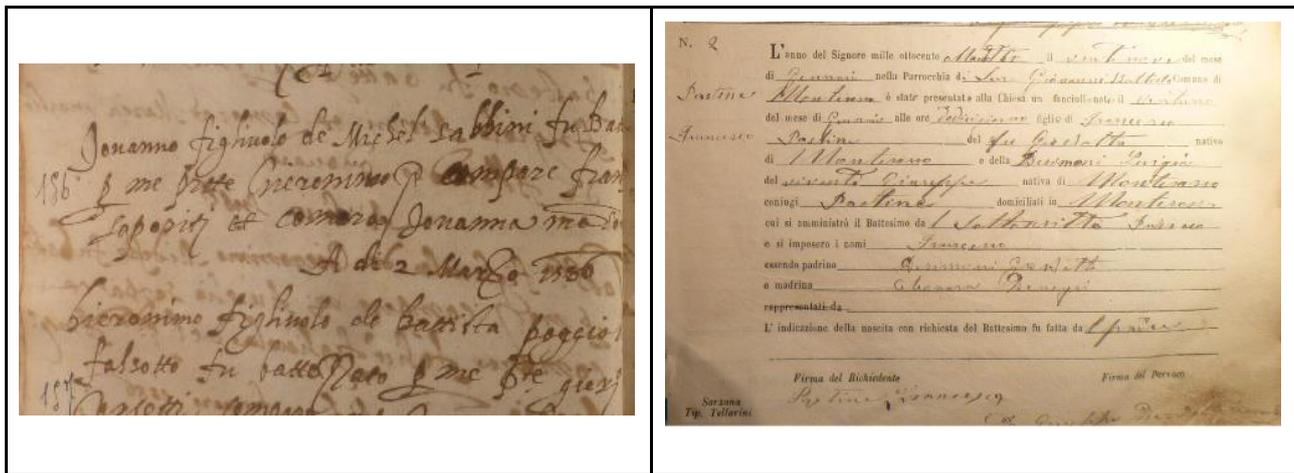


Fig. 1. Confronto tra una registrazione di battesimo del 1580 e una del 1899.

2. OPERARE SUL SEMILAVORATO

Il materiale che popola il sito dei registri è costituito sostanzialmente dalle immagini e dalle trascrizioni testuali fornite dal signor Moggia. Il trattamento delle immagini digitali che ci sono state consegnate non ha posto alcun problema di metodo: si trattava infatti di file in formato JPEG già denominati in modo corretto per corrispondere ai file di testo. Questi ultimi, al contrario, presentavano alcune criticità, in quanto la trascrizione non era stata restituita con criteri filologici. Nel database del sig. Moggia sono infatti presenti in maniera standardizzata solo il nome e il cognome delle persone a cui l'atto è riferito. Ad esempio, nel caso del battesimo si trovano l'anno, il nome del battezzato, il nome del padre e, se presente, il nome e il cognome della madre; non quelli di eventuali padrino e madrina. Le varianti grafiche dei nomi – nei primi registri in latino e in seguito in volgare – sono state tutte ricondotte alla forma italiana: ad esempio *Iohannes* – *Ioannes* è diventato Giovanni –, e *Iohannes Battista* con tutte le sue varianti grafiche, è stato sempre trasformato in GioBatta. Anche se tali scelte non rispondono ai correnti criteri di edizione scientifica delle fonti documentarie, nell'accettarla abbiamo tenuto conto dei seguenti fattori: in primo luogo la quantità dei dati effettivamente già disponibili (30.000 trascrizioni), che rendeva oggettivamente impossibile una loro nuova annotazione, soprattutto in assenza – a quanto ci risulta – di una soluzione agile per la fruizione di dati semi strutturati in fonti manoscritte o semi-manoscritte ([4]); in seconda istanza l'accuratezza del lavoro del signor Moggia che, benché amatoriale, era fondato su una profonda conoscenza delle famiglie del luogo: solo tale competenza poteva infatti consentire l'adeguata comprensione della grande mole di antroponimi presenti (nomi, cognomi e soprannomi). Abbiamo inoltre ritenuto che la possibilità di visionare contemporaneamente trascrizione e immagine consentisse agli utenti di verificare di persona la bontà della trascrizione, con la possibilità di comunicare agli amministratori l'esito di questa verifica. Facciamo infine notare che la valorizzazione – da parte di un ente di ricerca – di un lavoro prodotto da un membro della comunità avrebbe implicato – come poi effettivamente è successo – l'instaurarsi con la comunità stessa di una relazione diversa dal solito. In sostanza non si voleva semplicemente "offrire" un'edizione scientifica – la pubblicazione on line di documentazione inedita secondo i dettami della comunità accademica –, ma creare con le persone coinvolte un rapporto paritetico, in cui venivano riconosciute le competenze e professionalità di ognuno, al fine di ottenere il miglior risultato possibile date le forze in campo².

3. LA PIATTAFORMA WEB

Come detto in precedenza, il materiale iconografico di partenza era costituito da circa 35.000 immagini (ad alta risoluzione), divise in due categorie: le immagini di tutte le pagine dei registri (due pagine per immagine) e l'immagine di ogni singolo atto trascritto. Queste immagini erano già organizzate ed è stato relativamente semplice associarle ad ogni singola trascrizione e renderle fruibili tramite il software OpenSeadragon³. Le trascrizioni testuali erano distribuite in un numero molto elevato di cartelle e file di testo, per cui si è reso necessario raccoglierle in tre grandi fogli Excel (battesimi,

² Nello specifico il team del progetto ha coinvolto: una laureanda del corso di laurea di Informatica Umanistica (Chiara D'Anzi), un informatico (Vittore Casarosa), una digital public historian (Enrica Salvatori), un cittadino della comunità di Monterosso (Gustavo Moggia), il sindaco di Monterosso al Mare (Emanuele Moggia).

³ Una libreria Javascript che permette di visualizzare immagini e di compiere azioni su di esse, come zoom e pan (<https://openseadragon.github.io/>). OpenSeadragon è una delle migliori librerie per questo tipo di funzionalità e viene utilizzata in numerosi progetti online come, per citare solo un esempio, quelli della Biblioteca Nazionale Centrale di Roma (<http://digitale.bnc.roma.sbn.it/libri>).

matrimoni, morte), per facilitare la ricerca di corrispondenze tra i vari registri e l'immissione dei dati sulla piattaforma web. Nei fogli Excel di battesimo e di morte ogni riga corrisponde a un atto (e quindi a una persona) a cui è stato assegnato un codice identificativo unico (ID). Anche nel foglio Excel dei matrimoni ogni riga corrisponde a un atto (con il proprio ID), ma di fatto l'atto "corrisponde" a due persone.

A partire da questi dati sono state ricercate (tramite un programma scritto in Python) le corrispondenze per mettere in relazione gli atti di battesimo, matrimonio e morte riconducibili alla stessa persona. Il programma confronta i dati delle tabelle Excel relativi ad ogni persona (nome, cognome, nome del padre, cognome e nome della madre, data di nascita e di morte) e restituisce le possibili corrispondenze. Il confronto è stato effettuato ricercando ogni persona della tabella dei battesimi all'interno delle tabelle di matrimonio e morte, con opportuni filtri temporali. Al fine di non tralasciare nessuna possibile corrispondenza tra gli atti, lo stesso processo è stato ripetuto per ogni persona nella tabella dei morti ricercandola nella tabella dei battesimi. I risultati di questo processo sono stati memorizzati in due nuove tabelle Excel contenenti in ogni riga il valore identificativo di una persona, associato ai valori identificativi delle persone (possibili corrispondenze) trovati dal programma Python.

Queste cinque tabelle Excel sono state poi convertite in CSV e importate all'interno della piattaforma in modo da creare il database relazionale MySQL alla base della navigazione.

La piattaforma web permette di navigare le immagini e i dati, relativi a ogni registro e a ogni persona, tramite due modi: la ricerca per registro e la ricerca per nome della persona. Nella ricerca "per registro", si specifica subito se la ricerca viene fatta nei registri di battesimo, di matrimonio o di morte; si seleziona poi il registro relativo all'anno di interesse, e infine un atto (corrispondente a una persona) di quel registro. Anche nella ricerca "per nome" si specifica subito se la ricerca deve avvenire nei registri di battesimo, di matrimonio o di morte, con la differenza che in questo caso non viene selezionato un registro, ma si ricerca direttamente la persona interessata attraverso una lista di cognomi, associati agli anni in cui questi cognomi compaiono.

In entrambi i casi il punto di arrivo, all'interno della piattaforma, è rappresentato dalla "scheda anagrafica" di ogni persona, la quale contiene i dati relativi a una persona e le immagini degli atti in cui si ritrova questa persona (si veda la Fig. 2); i dati mostrati per una scheda vengono generati dinamicamente grazie a query SQL fatte sulle tabelle del database utilizzando procedure in PHP.

Anagrafica

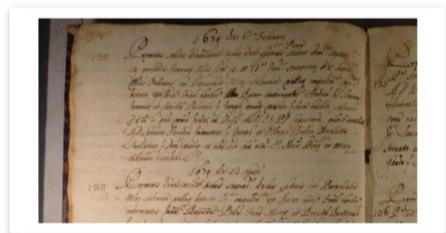
Anno di nascita: 1641
 Nome: Andrea
 Cognome: Desimoni
 Sesso: M
 Padre: Desimoni Giovanni
 Madre: (Il cognome manca) Elisabetta
Possibile coniuge:
 • Pallerano Camilla, atto di matrimonio del 1679

Anno di morte: 1721 (80 anni)

[Suggerisci altre informazioni relative a questo atto](#)



Atto di battesimo del 1641



Atto di matrimonio del 1679



Atto di morte del 1721

Fig. 2. La scheda anagrafica di una persona

Durante la ricerca delle corrispondenze è apparso chiaro come sia difficile trovare automaticamente tutte le relazioni e le corrispondenze tra gli atti di battesimo, di matrimonio e di morte per almeno quattro ordini di motivi: la presenza di molti dati mancanti o illeggibili; la presenza di molti omonimi; la possibilità che la persona descritta, ad esempio, in un atto di battesimo non sia presente nei registri di matrimonio e/o di morte perché magari non si è sposata o è "emigrata" da Monterosso; infine, dato che l'uso dei soprannomi nel passato era molto frequente e, soprattutto, molto incisivo nella vita quotidiana delle persone, è possibile trovare persone che nei registri di matrimonio e morte compaiono con un altro nome, anche molto differente da quello di battesimo.

Per questi motivi nella scheda anagrafica sono evidenziati in arancione i dati incerti o mancanti, e viene data la possibilità agli utenti (registrati) di fornire correzioni o informazioni aggiuntive che possano servire ad aggiornare e correggere il

database. Per fare questo è stato creato un form (costituito dal box in cui scrivere il testo del messaggio e un tasto di invio) e una tabella del database volta a conservare i messaggi ricevuti. Cliccando sul pulsante di invio, infatti, si inserisce un'istanza alla tabella tramite l'istruzione SQL di inserimento.

I messaggi vengono visualizzati dagli amministratori all'interno del pannello di controllo personale, tramite il quale è possibile inserire, modificare ed eliminare dati nelle principali tabelle del database senza avere necessariamente competenze tecniche sulla formulazione di comandi MySql. I messaggi, quindi, vengono visualizzati in questo modo (si veda la Fig. 3).

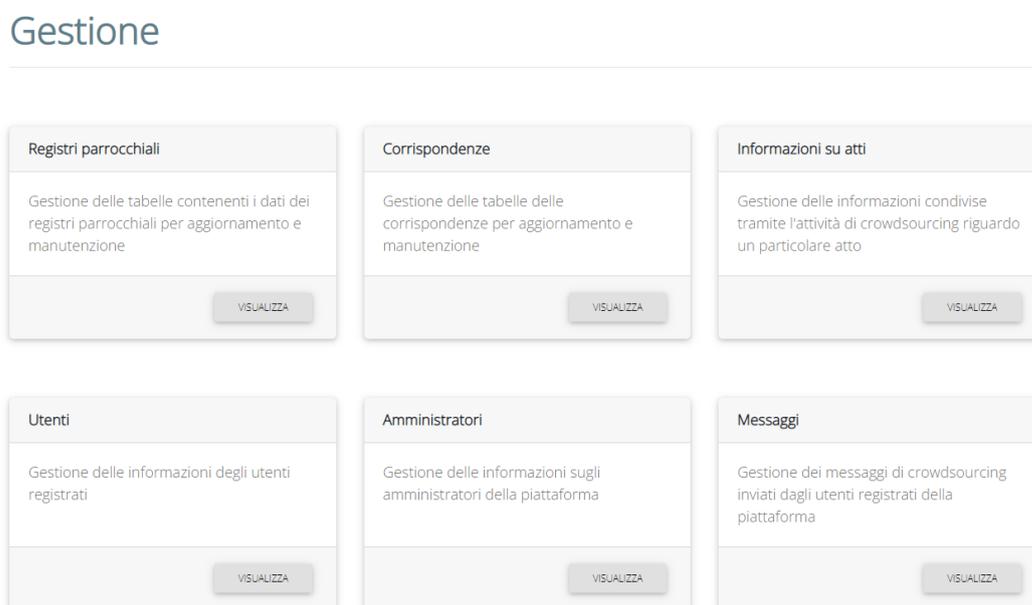


Fig. 3. Pannello di gestione della dashboard di amministrazione

4. STATISTICHE

Utilizzando le trascrizioni testuali è stato inoltre possibile fare uno studio storico-demografico sui dati, pubblicato nella sezione "Dati e statistiche" della piattaforma. In particolare, gli aspetti approfonditi da questo punto di vista riguardano prevalentemente la natalità e la mortalità della popolazione dalla fine del XVI secolo fino a più della metà del XX secolo, e la distribuzione dei cognomi.

Per fare un esempio, nel grafico che segue (si veda la Fig. 4) è possibile osservare come la quantità di nati è generalmente superiore a quella delle morti. Questa tendenza si inverte in pochi casi: il più evidente è tra il 1940 e il 1950, cioè durante e subito dopo la Seconda Guerra Mondiale. I dati "anomali" visibili negli estremi cronologici (1580/1610 e 1960/1984) dipendono invece dal fatto che la copertura dei registri non è omogenea.

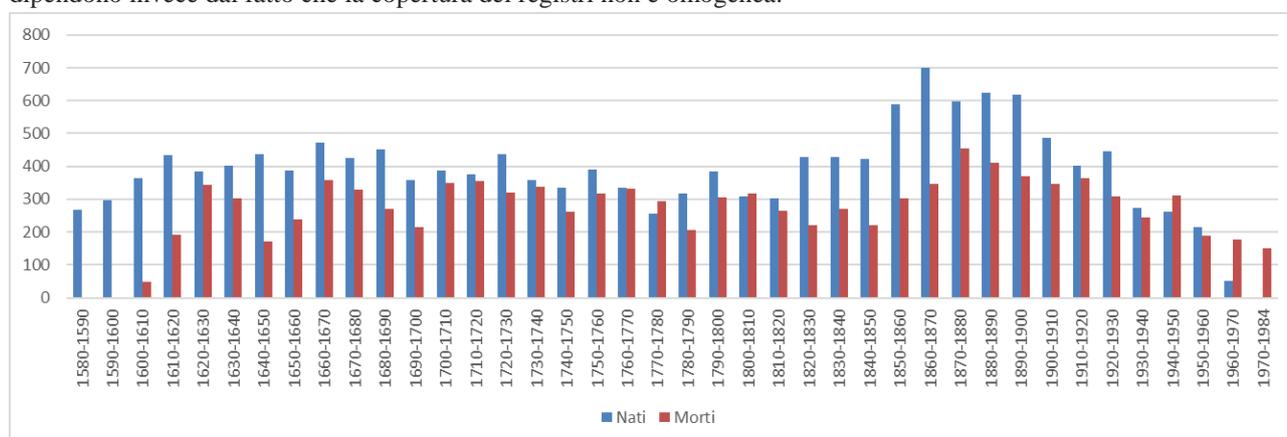


Fig. 4. Quantità di nati e di morti per decennio (1580-1984)

Come altro esempio abbiamo la distribuzione dei cognomi, che costituisce un elemento interessante per individuare le famiglie autoctone. Nel caso di Monterosso sono stati individuati i cognomi presenti all'interno dei registri e contate le

occorrenze di ogni cognome all'interno dello stesso registro. La quantità di cognomi diversi all'interno di ogni registro è la seguente:

Battesimo: 621 (di cui 331 compaiono solo una volta)
 Morte: 630 (di cui 392 compaiono solo una volta)
 Matrimonio Sposo: 480 (di cui 319 compaiono solo una volta)
 Matrimonio Sposa: 247 (di cui 121 compaiono solo una volta)

Nelle tabelle qui sotto (si vedano le Fig. 5 e 6) vengono riportati i 10 cognomi più frequenti per registro, che rappresentano da soli una percentuale piuttosto consistente sul totale dei cognomi all'interno dei registri.

Cognomi	Numero occorrenze	%
Moggia	830	5,46%
Rossignoli	779	5,12%
Benvenuto	760	5,00%
Poggi	733	4,82%
Cavallo	627	4,12%
Desimoni	574	3,78%
Gando	509	3,35%
Alzari	452	2,97%
Morando	404	2,66%
Contardi	383	2,52%
Totali	6051	39,80%

Battesimo

Cognomi	Numero occorrenze	%
Rossignoli	584	5,47%
Poggi	582	5,45%
Benvenuto	545	5,10%
Moggia	483	4,52%
Desimoni	422	3,95%
Gando	390	3,65%
Cavallo	367	3,44%
Alzari	341	3,19%
Morando	326	3,05%
Contardi	272	2,55%
Totali	4312	40,37%

Morte

Fig. 5. I 10 cognomi più frequenti nei registri di Battesimo e Morte

Cognomi	Numero occorrenze	%
Benvenuto	141	4,21%
Poggi	135	4,17%
Moggia	131	3,82%
Rossignoli	130	3,69%
Cavallo	119	2,86%
Desimoni	115	2,57%
Gando	89	2,31%
Contardi	80	2,05%
Alzari	72	2,02%
Martelli	64	1,77%
Totali	1076	29,47%

Sposo

Cognomi	Numero occorrenze	%
Moggia	169	4,58%
Benvenuto	164	4,23%
Cavallo	143	4,13%
Rossignoli	142	3,68%
Desimoni	131	2,61%
Poggi	128	2,55%
Gando	114	2,36%
Alzari	81	2,26%
Contardi	79	2,23%
Durante	73	2,23%
Totali	1224	30,86%

Sposa

Fig. 6. I 10 cognomi più frequenti nei registri di Matrimonio

Queste e altre statistiche visibili sulla piattaforma sono state ricavate dai dati consolidati raccolti nei tre fogli Excel menzionati prima. Un utente registrato può scaricare questi dati come tre file CSV (battesimi, matrimoni, morte) a scopi di studio per ulteriori ricerche e statistiche.

5. APRIRE ALLA COMUNITÀ

Sia per risolvere questi problemi, sia per far sì che il lavoro fatto avesse una sua ricaduta nella comunità, abbiamo deciso di aprire la piattaforma alla collaborazione del pubblico. Registrandosi nella piattaforma è possibile contribuire in due modi:

- 1) segnalare correzioni per eventuali errori di trascrizione, riempire lacune o correggere errori nel riconoscimento delle parentele e delle corrispondenze di una persona tra un registro e l'altro, segnalare nuove corrispondenze;
- 2) inviare materiali agli amministratori per pubblicare, in una sezione apposita ("Ricerche sulle famiglie"), gli studi genealogici specifici condotti dagli utenti sui dati resi disponibili dalla piattaforma.

L'apertura al pubblico, presentata nel luglio 2021 e periodicamente rilanciata sui social network (gruppi locali), ha già raccolto alcune manifestazioni di interesse e la pubblicazione di diversi studi specifici, tra cui si segnalano quelli di Gustavo Moggia e l'albero genealogico ricostruito da Danilo Francescano, ancora da pubblicare nella sua interezza (si veda la Fig. 7).

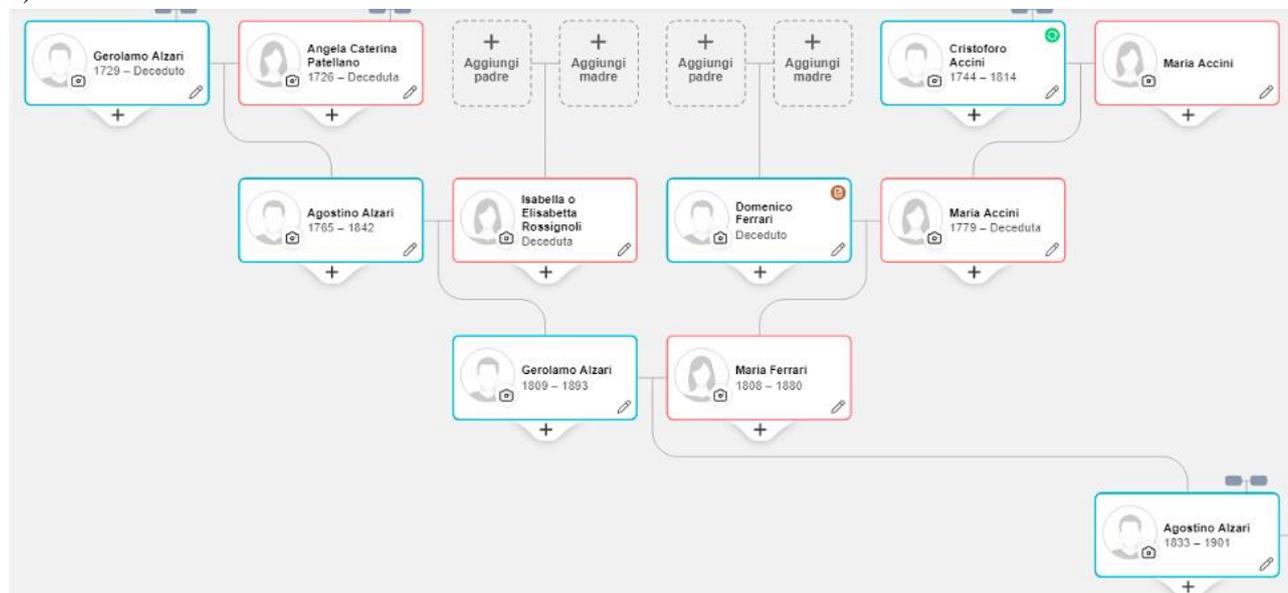


Fig. 7 Porzione di albero genealogico degli antenati di Danilo Francescano costruito su MyHeritage con i dati recuperati dai Registri Parrocchiali di Monterosso al Mare.

6. CONCLUSIONI

La piattaforma "Registri parrocchiali di Monterosso" rende fruibili a tutti (e in particolare alla comunità di Monterosso) i dati sui registri parrocchiali del luogo in una piattaforma semplice e intuitiva nella navigazione. Un solido database relazionale garantisce la consultazione contemporanea dei dati e delle immagini degli atti, consentendo ricerche anche sulle singole persone in tutte e tre le categorie di registro (battesimo, matrimonio, morte) e permettendo, contemporaneamente, agli utenti di contribuire all'arricchimento di quanto messo a disposizione. Come servizio aggiuntivo, il sito presenta inoltre anche statistiche storico-demografiche estratte dal database e approfondimenti sulle singole famiglie, prodotti, questi ultimi, sempre dagli utenti.

Si tratta di una realizzazione per ora unica in Italia. Esiste – a nostra conoscenza – solo un'altra iniziativa che dichiara di perseguire i medesimi scopi: "La Memoria dei Sacramenti" (<http://registriparrocchiali.weebly.com/>) col fine di mettere on line «oltre 5.000 registri parrocchiali, elenchi nominativi di oltre 250.000 atti di battesimo, matrimonio e sepoltura». Il sito ha un'impostazione simile a quella offerta dal nostro progetto, ma in molti casi non si visualizzano le immagini dei registri e non c'è la possibilità di ricercare singole persone⁴.

L'ostacolo oggettivamente più difficile da superare nella creazione di una banca dati consultabile *on line* con dati provenienti da un certo numero di registri parrocchiali deriva probabilmente dalle caratteristiche ibride e complesse di questa tipologia di fonti che abbiamo chiarito nell'introduzione. I registri infatti contengono una enorme quantità di dati semistrutturati, in parte manoscritti e in parte a stampa, dalla forma e contenuto variabile, che è "facile" fotografare, ma assai più complesso estrarre e correlare in modo da poterli utilmente interrogare.

Da questo punto di vista il caso di Monterosso al Mare non può costituire un modello "di per sé", in quanto per ottenere un simile risultato, ogni paese/parrocchia dovrebbe avere un suo "Gustavo Moggia", ossia una persona che per passione e dedizione si metta a schedare (non propriamente trascrivere) il contenuto dei registri. Tuttavia, proprio l'apertura al commento e alla collaborazione della gente che abbiamo implementato, riteniamo apra una possibile soluzione generalizzabile: si tratterebbe infatti di promuovere, per i registri parrocchiali, una via appunto ibrida, in cui le immagini dei registri, suddivise nelle singole dichiarazioni e accessibili tramite un database relazionale, siano visualizzabili e

⁴ Dal 2014 ad oggi il sito non è cambiato e non vi si leggono annunci o nuove pagine che facciano pensare al proseguimento della raccolta.

permettano l'interazione da parte dell'utenza. L'utenza dovrebbe coincidere di fatto con la comunità locale e dovrebbe essere sollecitata, con pratiche di crowdsourcing, non tanto a operare la trascrizione completa delle registrazioni, quanto appunto a schedare i dati essenziali, utili alla ricerca genealogico-demografica.

Ulteriori sviluppi potrebbero inoltre riguardare la possibilità per gli utenti di crearsi in autonomia il proprio albero genealogico a partire dai dati inseriti e validati, mutuando tecniche già operative e disponibili in altri servizi commerciali *on line* come ad esempio MyHeritage (<https://www.myheritage.it/>), Familysearch (<https://www.familysearch.org>) o Ancestry (<https://www.ancestry.com>)⁶.

BIBLIOGRAFIA

- [1] Ciacci, Francesca. 2003. «Umbria. Registri parrocchiali conservati negli archivi storici comunali. Censimento e inventari». SOPRINTENDENZA ARCHIVISTICA PER L'UMBRIA, ARCHIVIO DI STATO DI PERUGIA, ARCHIVIO DI STATO DI TERNI.
- [2] Coppola, Gauro, e Casimira Grandi. 1989. *La conta delle anime: popolazioni e registri parrocchiali: questioni di metodo ed esperienze*. Bologna: il Mulino.
- [3] Jedin, Hubert. 1944. *Le origini dei registri parrocchiali e il Concilio di Trento*.
- [4] Rosselli Del Turco, Roberto, Enrica Salvatori, Andrea Nanetti, Marco Giacchetto, Vera Isabell Schwarz-Ricci, e Antonella Ambrosio. 2021. «Introduzione: "Fonti archivistiche medievali nel digitale. La sfida di trattare e visualizzare dati semi-strutturati"». *Umanistica Digitale*, settembre 2021.
- [5] Sala, P. 1985. *I registri parrocchiali asburgici nella Milano del XIX secolo. Analisi di una fonte per la demografia storica e la storia sociale*. Milano: Franco Angeli.

⁵ MyHeritage è un social network che permette la creazione del proprio sito di famiglia, con foto, eventi e alberi genealogici; per questi ultimi si basa sul software Family Tree Builder. FamilySearch è il celebre servizio di accesso a una vasta banca dati genealogica offerto dalla Chiesa di Gesù Cristo dei santi degli ultimi giorni. Ancestry è una rete internazionale di siti genealogici con 9 miliardi di documenti di storia della famiglia ricercabili (dati del 2012).

⁶ MyHeritage è un social network che permette la creazione del proprio sito di famiglia, con foto, eventi e alberi genealogici; per questi ultimi si basa sul software Family Tree Builder. FamilySearch è il celebre servizio di accesso a una vasta banca dati genealogica offerto dalla Chiesa di Gesù Cristo dei santi degli ultimi giorni. Ancestry è una rete internazionale di siti genealogici con 9 miliardi di documenti di storia della famiglia ricercabili (dati del 2012).

Una nuova mappatura digitale per i borghi delle aree interne

Vincenzo Colaprice¹, Sergio Chiaffarata Iannelli², Nicola Barbuti³

¹ Università degli Studi di Bari, Italia – vincenzo.colaprice@uniba.it

² D.A.Bi.Mus. S.r.l., Spin-off dell'Università degli Studi di Bari, Italia - sergkj@gmail.com

³ Università degli Studi di Bari, Italia – nicola.barbuti@uniba.it

ABSTRACT

La rigenerazione dei borghi delle aree interne è uno degli obiettivi prioritari delle strategie delineate nel PNRR. È indispensabile ripensare e reinterpretare questi ecosistemi diffusi, per favorirne il rilancio riattivando l'interesse da parte sia delle comunità residenti, sia dei potenziali utenti. Produrre una nuova mappatura digitale è essenziale per comprendere le diverse esigenze e caratteristiche dei borghi a partire dalle relazioni con i contesti e con i territori in cui insistono.

PAROLE CHIAVE

Mappatura digitale, borghi, aree interne.

INTERVENTO

1. INTRODUZIONE

Lo scenario dei borghi italiani oscilla tra due estremi opposti: luoghi abbandonati o in progressivo spopolamento, o mete di devastanti turismi di massa. Nel 2016 l'indice di internazionalizzazione dei borghi ha raggiunto il 30%, pari a 15 milioni di turisti con un indotto vicino al miliardo di euro ([1]). Eppure, il 72% dei borghi ha registrato nell'ultimo decennio un calo medio del numero degli abitanti del 3%.

In seguito all'istituzione della *Strategia Nazionale per le Aree Interne* (SNAI)¹, è sorto un ampio confronto sui temi dello spopolamento e del recupero e rigenerazione delle aree interne. Il punto di svolta è rappresentato dal lancio del *Manifesto per Riabitare l'Italia* ([2]), che emancipa i borghi dal mero ruolo di mete turistiche ([3]: X-XI). Ripopolamento, identità abitativa, marginalità, patrimonio paesaggistico e culturale sono le direttrici su cui si sta articolando la strategia sui borghi. In questo scenario, è fondamentale comprendere le peculiarità non solo geografiche, ma soprattutto identitarie e di contesto che contraddistinguono i territori ([3]: 52). Procedere con una mappatura che non sia solo posizionamento geografico, ma dia conto delle relazioni dei borghi con le aree di contesto diventa quindi indispensabile per comprendere le peculiarità territoriali e orientare le *policies* utili alla rigenerazione.

Questa esigenza è ben nota: l'ANCI (Associazione Nazionale Comuni Italiani) da diversi anni realizza un *Atlante dei Piccoli Comuni*², pubblicato sul web nel 2018. Altre associazioni ed enti hanno sviluppato carte tematiche e rappresentative dei comuni affiliati. Al giorno d'oggi, tuttavia, manca un lavoro di mappatura dei 5.521 borghi italiani che raccolga indicatori e informazioni utili a valutarli in relazione ai contesti territoriali, culturali ed economici in cui insistono.

Al fine di affrontare questa criticità, è stata avviata una mappatura multilivellare digitale, attualmente in corso, nella quale si prendono in considerazione e si integrano analiticamente i diversi indicatori che illustrano le peculiarità dei borghi in relazione ai territori di contesto.

2. PERCHÉ MAPPARE I BORGHI

Nel 2017, il MiBACT ha definito i borghi come «*i comuni italiani con al massimo 5.000 abitanti caratterizzati da un prezioso patrimonio culturale*»³. I borghi rappresentano il 69,8% dei 7.904 comuni italiani, occupando il 54% della penisola. Ospitano il 17% della popolazione italiana, circa 10 milioni di abitanti. Il 72% dei borghi si trova nelle aree interne, seguendo la dorsale appenninica e l'arco alpino ([5]). Dal punto di vista culturale e produttivo, i borghi includono il 31,1% di luoghi e beni culturali, il 32,8% dei musei e danno origine al 90% dei prodotti DOP e IGP ([6]: 12-15).

Nonostante la loro rilevanza, i borghi presentano contesti condizionati da carenze infrastrutturali e logistiche, scarsa valorizzazione del patrimonio culturale, difficoltà di accessibilità ai servizi. Un quadro lontano dall'immaginario collettivo

¹ Vedi <https://www.agenziacoesione.gov.it/strategia-nazionale-aree-interne/>.

² Vedi <https://www.anci.it/atlane-dei-piccoli-comuni/>.

³ Direttiva n. 555 del Ministro dei Beni e delle Attività Culturali e del Turismo, 2 dicembre 2016, recante norme relative all'indizione dell'Anno dei Borghi Italiani.

che, ancora oggi, condiziona non poche scelte politiche, identificando il “borgo” nel piccolo centro storico ben conservato, incastonato in un patrimonio naturalistico e architettonico attraente ([4]).

La mappatura digitale multilivello dei borghi può restituire una fotografia della reale condizione in cui versa ciascun borgo, consentendo di progettare investimenti mirati, finalizzati a rigenerare i borghi quali *attuatori d'interesse*. L'obiettivo è fornire un'analisi che incroci i dati numerici della popolazione con quelli relativi ai contesti territoriali di riferimento.

3. LA MAPPATURA

Il lavoro di mappatura è in corso. Come primo approccio metodologico, si è reso necessario procedere con un censimento dei borghi aggiornato al 2021. L'ultimo censimento è rappresentato dall'*Atlante interattivo dei Piccoli Comuni*, prodotto dall'ANCI sulla base di dati ISTAT al 31/12/2017⁴. L'ANCI ha suddiviso i borghi quantitativamente in base a tre fasce di residenti: da 0 a 1.000 abitanti, da 1.001 a 3.000 abitanti e da 3.001 a 5.000 abitanti.

Per la mappatura, quindi, sono stati raccolti i dati ISTAT al 01/01/2021, con conseguente rideterminazione dei dati demografici ANCI: la fascia da 0 a 1.000 residenti è stata mantenuta, mentre il segmento intermedio è stato ridotto da 1.001 a 2.500 e la terza fascia è stata portata da 2.501 a 5.000. La ridefinizione risponde alla necessità di avere segmenti realisticamente rappresentativi dei contesti territoriali in cui i borghi insistono. Infatti, è emerso che, in diverse regioni, buona parte dei borghi con popolazione residente superiore alle 2.500 unità presenta condizioni di stabilità e sostenibilità del tutto assimilabili a quelle dei comuni al di sopra dei 3.000. Inoltre, i comuni compresi dall'ANCI nella fascia tra i 1.001 e 3.000 residenti sono 2.224, ma, tra questi, 1.810 non superano i 2.500 abitanti. Quindi, portare il segmento intermedio a 2.500 residenti, con uno scarto di soli 414 comuni rispetto all'ANCI, non è significativo in termini quantitativi, mentre in termini qualitativi l'analisi di una serie di indicatori consente di delineare uno scenario sia di dettaglio che d'insieme realistico, chiaro ed esaustivo.

Stabiliti questi criteri metodologici preliminari, si è provveduto alla creazione di mappe digitali regionali e multilivello utilizzando come base Google Maps, al fine di favorire una consultazione agevole e intuitiva. Le mappe sono accessibili da un sito web dedicato, di imminente pubblicazione online.

Un primo set di mappe rappresenta i dati demografici e la diffusione dei borghi. Ogni mappa regionale contiene tre livelli (Figura 1): confini amministrativi di tutti i comuni; puntinato dei borghi; poligoni della superficie dei borghi. Il puntinato permette di geolocalizzare i borghi nella regione. I segnaposto, una volta cliccati, offrono dati relativi alla demografia del borgo. L'uso dei poligoni consente di comprendere la dimensione demografica dei borghi in relazione al contesto spaziale e altimetrico. Le tre fasce di residenti hanno colori diversi: rosso da 0 a 1.000 abitanti, giallo da 1.001 a 2.500, blu da 2.501 a 5.000.



Figura 8: Livelli di mappatura dei borghi del Friuli-Venezia Giulia.

Questo set di venti mappe regionali permette di avere un quadro aggiornato della contestualizzazione territoriale dei borghi. L'attività di mappatura, tuttavia, non può tenere considerazione del solo dato demografico. La ricerca vuole mettere in evidenza le potenzialità di ogni borgo attraverso indicatori che tengano conto di dati qualitativi. È in corso di lavorazione il set riguardante il patrimonio culturale e paesaggistico delle regioni.

⁴ ANCI, Atlante Interattivo dei Piccoli Comuni, <http://www.anci.it/atlane-dei-piccoli-comuni/>, ultimo accesso 28 luglio 2021.



Figura 9: Mappatura del patrimonio culturale e paesaggistico della Puglia

Sono tre i livelli previsti: puntinato dei borghi con numero di beni culturali censiti da ISTAT, prossimità di beni UNESCO, eventi e tradizioni riconosciute da progetti ICCD, parchi e aree protette riconosciute da Stato e Regione (Figure 2-3).

Infine, si prospetta la creazione di due ulteriori set di mappe:

- Accessibilità e logistica. Livelli: infrastrutture e collegamenti, integrità e conservazione del patrimonio architettonico, rischio del territorio.
- Servizi ed economia. Livelli: sistemi locali del lavoro, produzioni agroalimentari e artigianali, ricettività, presenze turistiche.

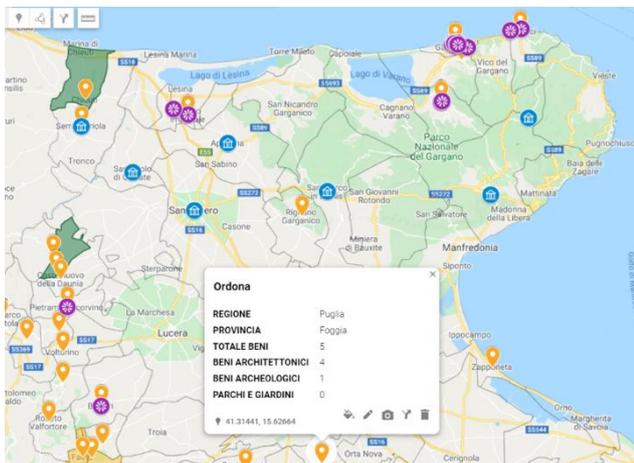


Figura 10: Dettaglio della mappatura del foggiano.

BIBLIOGRAFIA

- [1] Arduini, Clelia. 2017. «Il 2017 è l'Anno dei Borghi». *Touring Club Italiano* (blog). 17 febbraio 2017. <https://www.touringclub.it/notizie-di-viaggio/il-2017-e-lanno-dei-borghi>.
- [2] Cersosimo, Domenico, e Carmine Donzelli. 2018. *Manifesto per Riabitare l'Italia*. Roma: Donzelli Editore.
- [3] De Rossi, Antonio. 2020. *Riabitare l'Italia*. Roma: Donzelli Editore.
- [4] Gugliotta, Federica. 2010. «Di cosa parliamo quando parliamo di borghi». *Repubblica*, 4 aprile 2010. <http://temi.repubblica.it/casa-made-2010/2010/02/04/di-cosa-parliamo-quando-parliamo-dei-borghi/>.
- [5] «Identikit dei piccoli Comuni, Unioni di Comuni e Fusioni di Comuni». 2019. Area Studi e Ricerche ANCI. <https://www.anci.it/atlane-dei-piccoli-comuni/>.
- [6] *Piccoli comuni e cammini d'Italia*. 2020. Roma: Fondazione Symbola. <https://www.symbola.net/ricerca/piccoli-comuni-cammini-italia/>.

Intelligenza artificiale e archivi audiovisivi: potenzialità e sfide del progetto “PH-Remix”

Giovanni Grasso¹, Chiara Mannari², Davide Serramazza³

¹Università di Pisa, Italia - giovanni.grasso@fileli.unipi.it

²Università di Pisa, Italia - chiara.mannari@fileli.unipi.it

³Università di Pisa, Italia - davide.serramazza@fileli.unipi.it

ABSTRACT

PH-Remix è un prototipo di piattaforma web basata sull'intelligenza artificiale che consente il caricamento, la catalogazione, la ricerca, la consultazione, l'estrazione e il remix di contenuti multimediali.

La piattaforma nasce con l'obiettivo di creare uno strumento da integrare alle metodologie definibili classiche di ricerca e studio delle fonti prevedendo, di fatto, l'accesso agevolato e il riutilizzo di una grande quantità di fonti filmiche primarie. Ponendo particolare attenzione alle sfide ed alle criticità che possono nascere durante lo sviluppo di uno strumento di questo tipo, il paper in questione intende presentare:

- Il prototipo di remix e le possibili applicazioni pratiche nel contesto della *public history* scaturite dalla possibilità di creare nuovi contributi;
- Le tecniche di *machine learning* utilizzate per l'estrazione delle informazioni dalle opere audiovisive.

PAROLE CHIAVE

Cinema e storia, remix culture, intelligenza artificiale, archivi audiovisivi, digital public history.

INTERVENTO

1. INTRODUZIONE

La piattaforma PH-Remix (Public History Remix)¹ è sviluppata nell'ambito dell'omonimo progetto di ricerca biennale coordinato dal Laboratorio di Cultura Digitale dell'Università di Pisa in collaborazione con il Festival dei Popoli e la Fondazione Sistema Toscana.

Il materiale filmico dell'archivio del Festival dei Popoli rappresenta il caso di studio per la realizzazione del progetto e per lo sviluppo del prototipo: i dati dell'archivio sono gestiti nella piattaforma sia attraverso la metadattazione tradizionale, importando le schede catalografiche realizzate nell'ambito del parallelo progetto di digitalizzazione dell'archivio, sia tramite l'uso di tecniche di intelligenza artificiale che permettono l'estrazione automatica di numerosi segmenti video significativi di durata variabile (clip). A tale fine sono stati utilizzati diversi algoritmi di *machine learning* specializzati nell'estrazione di informazioni da contenuti audiovisivi quali: il riconoscimento di oggetti, l'estrazione di sottotitoli mediante OCR, la segmentazione in inquadrature etichettate in base al colore dominante.

In un'ottica metodologica tipica della *digital public history*, il progetto intende sperimentare innovative modalità collaborative per fare la storia con il pubblico con finalità di favorire la valorizzazione di possibili fondi filmici depositati presso archivi ed istituzioni che intendano potenziare la catalogazione e la conseguente promozione dell'archivio attraverso lo studio della storia.

2. LA PIATTAFORMA “PH-REMIX”

La piattaforma, tuttora in corso di sviluppo, è caratterizzata da un'architettura software basata su microservizi che dialogano tra loro attraverso API (*application programming interface*) [Fig. 1].

¹ <http://www.labcd.unipi.it/ph-remix>.

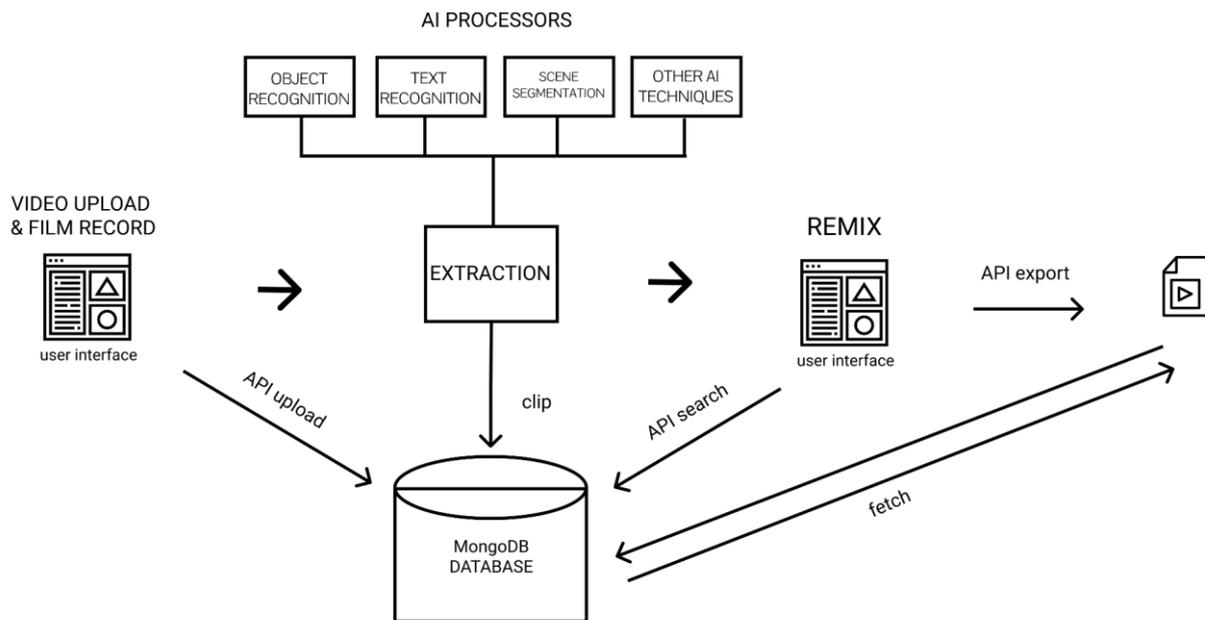


Figura 1. Diagramma della piattaforma PH-Remix

I film vengono caricati sul server deputato all'estrazione automatica attraverso un'interfaccia web ad accesso riservato. Nella fase di estrazione, ciascun film viene analizzato da diversi processi basati su algoritmi di intelligenza artificiale che restituiscono come risultato numerose clip di durata variabile. Parallelamente viene creato un record relativo al film attraverso l'importazione delle schede catalografiche impostate seguendo gli standard proposti dall'International Federation of Film Archives (FIAF) ([3]). Le clip estratte sono infine pubblicate nell'ambiente di remix, un'applicazione JavaScript accessibile agli utenti finali che consente la ricerca delle clip, la visualizzazione e il video editing [Fig. 2].

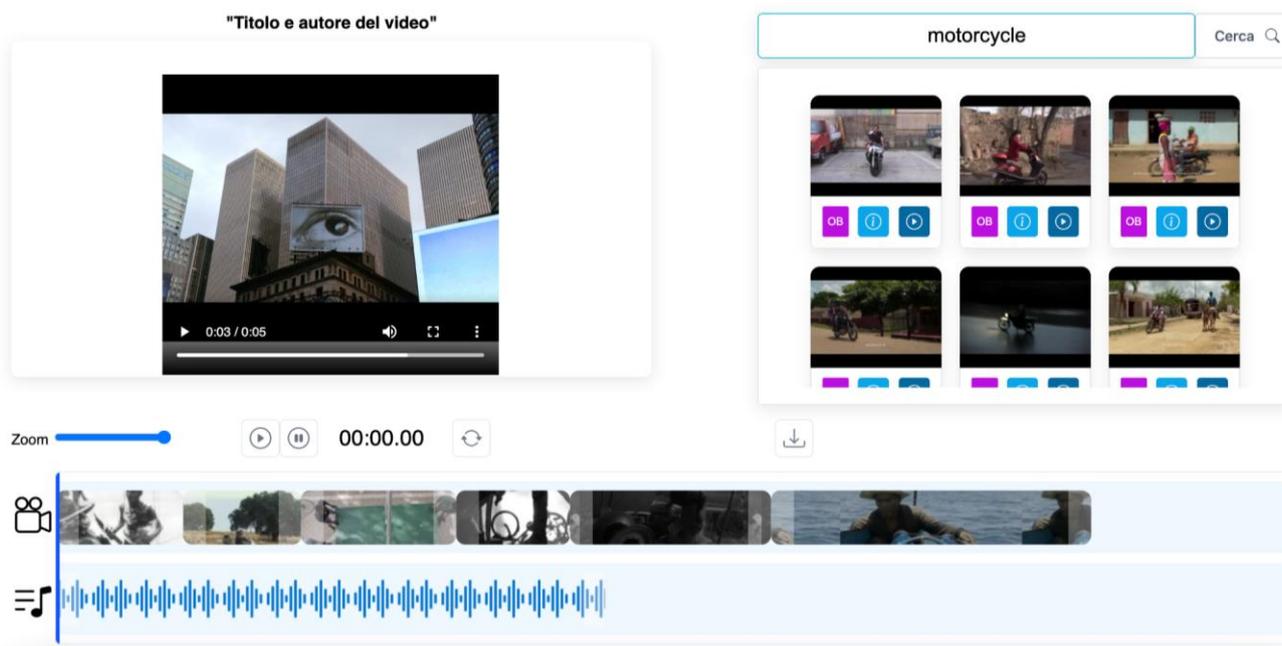


Figura 2. Prototipo dell'interfaccia di remix

L'operazione di montaggio video consente di mixare clip, effettuare ritagli ed aggiungere tracce audio. In qualsiasi momento è possibile visualizzare l'anteprima del video e, una volta completato il processo, lanciare la funzione di esportazione che restituisce un file in formato mp4 corredato da un documento con l'elenco dei film di provenienza dei diversi frammenti utilizzati. In successive fasi di sviluppo si potrà prevedere l'aggiunta di funzionalità di composizione

video avanzate e l'evoluzione della funzionalità di ricerca delle clip (al momento basata sui metadati estratti dagli algoritmi di *machine learning*) attraverso l'applicazione di modelli di *information retrieval*.

3. IL CONTRIBUTO DELL'INTELLIGENZA ARTIFICIALE

Una parte profondamente innovativa del progetto è costituita dall'adozione di tecniche di *machine learning* per l'analisi del contenuto audiovisivo allo scopo di indicizzare porzioni di video o clip.

In letteratura scientifica sono presenti diversi modelli che, grazie alla loro capacità di identificare specifiche caratteristiche all'interno del contenuto visuale, possono essere utilizzati per dividere i video sottoposti alla piattaforma in clip: ad esse inoltre verrà assegnata una label necessaria per l'indicizzazione. Chiameremo tali algoritmi *processor* da qui in avanti.

Per entrare nello specifico di tale processo descriviamo come alcuni modelli già presenti in letteratura siano stati sfruttati per il nostro progetto. Il primo *processor* utilizzato è TransNet V2 ([12]), un modello per *shot transaction* che consiste nell'identificazione delle transizioni fra un'inquadratura e la successiva. Con tale modello vengono estratte dal film tutte le clip corrispondenti ai cambi di inquadratura assegnando come label il colore dominante all'interno della scena.

Gli altri due *processor* sin qui adoperati utilizzano, invece, modelli ideati per lavorare su immagini piuttosto che su materiale video; per adoperarli è risultato necessario dunque apportare delle modifiche per cambiarne il dominio di utilizzo. Il primo è RetinaNet ([7]: 2980-2988), un modello per *object detection*: la versione da noi utilizzata fa affidamento su ResNet152 ([13]) per l'estrazione delle *features* dalle immagini ed è stato allenato sul *dataset* Open Image ([5]: 1956–1981) rilasciato da Google. Tale modello fornisce come output una serie di predizioni riguardanti varie sottosezioni delle immagini con la relativa predizione (oggetto rilevato) e *confidence* (probabilità assegnata alla predizione effettuata). Per un singolo film, tale algoritmo viene applicato su tutti i frame che lo compongono scartando tutte le predizioni con un valore di *confidence* al di sotto di una determinata soglia. Da quest'ultima lista vengono estratte le clip in cui lo stesso oggetto è presente in più frame successivi; le label di tali clip saranno ovviamente gli oggetti riconosciuti. [Fig. 3]



Figura 3. Esempio visivo di informazione estratta da RetinaNet

L'ultimo *processor* da menzionare è utilizzato per estrarre il contenuto dei sottotitoli presenti nei documentari effettuando la concatenazione di due diversi modelli: il primo è EAST ([14]: 5551–5560), il cui scopo è quello di rilevare all'interno di un'immagine le parti in cui è presente del testo; il secondo è un OCR implementato dalla popolare libreria tesseract che analizza tali porzioni ricavandone il testo contenuto. Tale pipeline viene applicata a tutti i frame che compongono un video. La label di tali clip sarà il transcript estratto; le varie clip estratte utilizzando questi tre algoritmi andranno a costituire l'indice delle clip utilizzabili dalla piattaforma di remix. La label assegnata dall'algoritmo corrisponderà alla parola chiave usata per ricercare tale clip.

4. CONCLUSIONI

La piattaforma di remix si presenta come uno strumento complesso ed altamente innovativo per il video remix inteso come il riutilizzo di frammenti video precedentemente pubblicati. L'azione di copiare, trasformare, rielaborare² è legittimata dalle pratiche di produzione tipiche del web che si fondano sulla condivisione di contenuti ormai parte della cultura di massa. La piattaforma PH-Remix nasce in seno ad un progetto di *public history* il cui obiettivo è far uso delle modalità di interazione diffuse coinvolgendo gli utenti finali nell'elaborazione di contenuti di argomento storico mediante l'utilizzo del cinema documentario.

Con il test del prototipo presso un ampio pubblico, caratterizzato da studiosi per la produzione di nuovi materiali didattici e studenti di diversi ordini scolastici per la fruizione e la rielaborazione di contenuti multimediali di carattere storico, si intende realizzare prodotti di remix volti all'individuazione di un modello di *public history remix*³ e fare una prima riflessione sulle sfide e le potenzialità portate da un simile strumento. In primo luogo, la sfida di introdurre in ambito accademico uno strumento di ricerca in grado di produrre e restituire notevoli quantità di documenti multimediali frammentati da algoritmi di intelligenza artificiale e successivamente l'effettiva possibilità di messa online dei materiali audiovisivi d'archivio nel rispetto delle vigenti leggi sul diritto d'autore. Riguardo invece alle potenzialità, oltre ai già citati vantaggi per gli utenti finali, si intende evidenziare quanto la disponibilità di una piattaforma online che permette la visualizzazione e il riutilizzo di clip estratte da migliaia di film digitalizzati, schedati ed indicizzati in maniera sistematica grazie al contributo dell'intelligenza artificiale, inneschi un meccanismo di rilancio e di rivalorizzazione dei fondi filmici contribuendo alla diffusione della conoscenza dell'archivio.

BIBLIOGRAFIA

- [1] Bell, Desmond. 2011. «Documentary film and the poetics of history». *Journal of Media Practice* 12: 3–25.
- [2] Burrough, Xtine, Owen Gallagher, e Eduardo Navas. 2021. «Forking Paths in New Media Art Practices: Investigating Remix». *Media-N, Journal of the New Media Caucus*, 17 (1).
- [3] Fairbairn, Natasha, Maria Assunta Pimpinelli, e Thelma Ross. 2016. *The FIAF Moving Image Cataloguing Manual*. Indiana University press.
- [4] Gallagher, Owen. 2018. *Reclaiming critical remix: the role of sampling in transformative works*. Routledge.
- [5] Kuznetsova, Alina, et al. 2020. «The open images dataset v4». *International Journal of Computer Vision* 128: 1956–81.
- [6] Lagny, Michèle. 2001. «Il cinema come fonte di storia». In *Storia del cinema mondiale*, Brunetta Gian Piero. Vol. 5. Teorie, strumenti, memorie. Einaudi.
- [7] Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, e Piotr Dollár. 2017. «Focal loss for dense object detection». In *Proceedings of the IEEE international conference on computer vision*, 2980–88.
- [8] Navas, Eduardo. 2012. *Remix theory. The aesthetics of sampling*. Springer.
- [9] Pang, Natalie, Kai Khiun Liew, e Brenda Chan. 2014. «Participatory archives in a world of ubiquitous media». *Archives and Manuscripts* 42: 1–4.
- [10] Rosenstone, Robert. 1995. *Revisioning History. Film and the Construction of a New Past*. Princeton University Press.
- [11] Sorlin, Pierre. 1999. *L'immagine e l'evento. L'uso storico delle fonti audiovisive*. Paravia.
- [12] Souček, Tomáš, Jaroslav Moravec, e Jakub Lokoč. 2020. «Transnet v2: An effective deep network architecture for fast shot transition detection». *arXiv* 2008.04838.
- [13] Targ, Sasha, Diogo Almeida, e Kevin Lyman. 2016. «Resnet in resnet: Generalizing residual architectures». *arXiv* 1603.08029.
- [14] Zhou, Xinyu, Cong Yao, He Wen, Yuzhi Wang, He Shuchang, e Liang Weiran Jiajun. 2017. «East: an efficient and accurate scene text detector». In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 5551–60.

² Cfr. Kirby Ferguson “Everything is a Remix” <https://www.everythingsaremix.info/watch-the-series/>.

³ Un primo esperimento è stato realizzato in occasione del seminario di Cultura Digitale dal titolo “Remix Culture” svolto il 12 maggio 2021 presso l'Università di Pisa. La registrazione del seminario è disponibile al link <http://www.labcd.unipi.it/seminari/chiara-mannari-remix-culture/> e la presentazione del public history remix parte da [1:17:00](#).

Sessione Intelligenza 2
John von Neumann

Un nuovo approccio per la descrizione e gestione del patrimonio culturale digitale relativo a MAB

Nicola Barbuti¹, Stefano Ferilli²

¹ Università degli Studi di Bari Aldo Moro, Italia – nicola.barbuti@uniba.it

² Università degli Studi di Bari Aldo Moro, Italia – stefano.ferilli@uniba.it

ABSTRACT

Il patrimonio culturale di musei, archivi e biblioteche (MAB) è oggetto di rinnovato e crescente interesse. A fronte di una massa enorme e sempre crescente di dati digitali relativi al patrimonio, però, troppo spesso si confonde il DCH con semplici *digital twins* di beni analogici, continuando a trascurare il valore culturale dei dati stessi. Recenti progetti che cercano di superare questa limitazione raccolgono informazioni comunque strettamente legate agli originali analogici rappresentati negli oggetti digitali.

Si presenta qui la ricerca, in corso, su un modello di rappresentazione degli oggetti digitali relativi al patrimonio MAB che supporta questa necessità di ampliamento delle informazioni, includendo il fra l'altro il contesto del bene ed il ciclo di vita della risorsa digitale. La proposta si basa su tecniche di Intelligenza Artificiale (IA) che supportano non solo l'organizzazione dei dati, ma anche modalità innovative di interazione proponendo soluzioni personalizzate.

PAROLE CHIAVE

Patrimonio culturale digitale, MAB, basi di conoscenza, ontologie concettuali, intelligenza artificiale.

INTERVENTO

1. STATO DELL'ARTE

La digitalizzazione dei patrimoni di musei, archivi e biblioteche (MAB) è oggetto di rinnovato e crescente interesse. A riguardo, anche il recente PNRR prevede cospicui investimenti, nella prospettiva di rilanciare le relazioni e le interazioni delle comunità con il patrimonio culturale pubblicando online raccolte e collezioni digitalizzate in varia maniera.

Tuttavia, già oggi in rete fluttua una mole ingente di dati relativi ad artefatti analogici, considerati del tutto impropriamente *Digital Cultural Heritage*. Infatti, la situazione è ben differente da quanto comunemente si rappresenta. I metodi e le pratiche di indicizzazione in uso per il patrimonio culturale¹ – soprattutto, ma non solo MAB – sono a oggi ancorati alle norme catalografiche di ciascuna tipologia di beni e si focalizzano solo sugli artefatti originali rappresentati nei layout degli oggetti digitali, trascurando le informazioni relative ai dati. Ne consegue che, a nostro parere, poco o nulla di quanto prodotto ha effettiva valenza culturale, a meno che non si intenda considerare *patrimonio digitale* le entità i cui contenuti visuali e le relative descrizioni materializzano *digital twins* di beni analogici. Questo incide direttamente sulle pratiche di conservazione e preservazione dei dati, focalizzate sulla sostenibilità nel tempo dei layout, piuttosto che delle informazioni sul ciclo di vita delle risorse digitali ([1];[4]).

In questo scenario, alcuni recenti progetti si segnalano come interessanti tentativi di evolvere verso modelli di valorizzazione degli oggetti digitali, in quanto l'indicizzazione delle risorse si basa su descrizioni reticolari tramite LOD² che recepiscono il modello RDF³. Ancora una volta, però, si descrivono prevalentemente le caratteristiche degli originali rappresentati nei layout, prestando poca attenzione a interi insiemi correlati sia agli oggetti digitali, sia agli stessi artefatti analogici. Di fatto, si reitera il modello "tradizionale" di descrizione che standardizza e rende i database ancora oggi statici e poco interattivi.

¹ Ci stiamo riferendo all'italiano MAG e agli internazionali Dublin Core, METS, MODS, xDams.

² Si veda il recente Catalogo Generale dei Beni Culturali curato dall'ICCD <<https://catalogo.beniculturali.it/>> (ultima consultazione: 7 ottobre 2021). Sebbene l'utilizzo ben calibrato dell'ontologia di LOD ArCO segni un'evoluzione senza dubbio di valore e significativa rispetto alla maggioranza delle collezioni presenti in rete, l'indicizzazione delle risorse esposte si focalizza esclusivamente sulle immagini dei layout.

³ <https://it.wikipedia.org/wiki/Resource_Description_Framework> (ultima consultazione: 7 ottobre 2021).

2. MOTIVAZIONI

Da questi recenti tentativi genera la ricerca in corso. Si sta studiando e sperimentando un modello di indicizzazione di oggetti digitali relativi al patrimonio MAB basato su ontologie formali che estendano le tipologie di classi, relazioni ed attributi tipicamente usate in tale ambito, sia ampliando le informazioni di contesto relative all'artefatto analogico tramite relazioni concettuali che lo connettono ad altre risorse, sia raccogliendo e rendendo disponibili informazioni sulla *provenance* rappresentativa del ciclo di vita dei dati ([8]).

L'obiettivo è delineare strutture di metadati che sostanzino culturalmente gli artefatti digitali, nella prospettiva di evolvere le attuali *basi di dati* in *basi di conoscenza*, nelle quali le descrizioni degli artefatti originali si espandono grazie all'uso sapiente delle relazioni concettuali e, nel contempo, sono correlate alle descrizioni fisiche e di contesto dei dati.

Si sta anche sperimentando l'applicazione di tecniche di Intelligenza Artificiale (IA) capaci di gestire sia gli elementi legati all'oggetto culturale rappresentato, che quelli fisici e di contesto relativi alla risorsa digitale, e di supportare ed espandere le possibili interazioni tra le *basi di conoscenza* e molteplici differenti comunità di utenti, proponendo a ciascun fruitore soluzioni personalizzate grazie all'utilizzo di tecniche avanzate di profilazione.

Rendendo disponibili nel lungo termine questo ecosistema informativo, le risorse e le collezioni digitali assumono così la forma e la funzione di *record* coerenti con i principi *FAIR* ([3];[2]) e acquisiscono valore culturale nella prospettiva di generare nuovo DCH.

3. METODOLOGIA

La transizione tecnologica dalle *basi di dati* alle *basi di conoscenza* si concretizza passando da un'organizzazione delle informazioni basata su schede a una reticolare, che spinge sulle inter-relazioni fra informazioni per creare un valore aggiunto significativo. La tipologia di informazioni di una base di conoscenza e la loro organizzazione sono definite nelle ontologie studiate dalla branca dell'IA nota come *Rappresentazione della Conoscenza e Ragionamento Automatico*.

L'unione delle ontologie con le relative istanze viene spesso definita *Grafo di Conoscenza*, tradizionalmente rappresentato con formalismi che sono riconducibili a insiemi di triple <Soggetto,Predicato,Oggetto>. Tali formalismi sono usati da programmi che realizzano forme di ragionamento automatico in grado di fare inferenze sulla conoscenza a disposizione, mettendo però in secondo piano l'efficienza tipica delle basi di dati.

Unire le due tecnologie valorizza l'efficienza dalle basi di dati e la flessibilità e la potenza espressiva dalle ontologie. I DB a grafo (a es., Neo4j) sono particolarmente adatti alla rappresentazione reticolare delle informazioni, ma afferiscono al modello dei grafi *Labeled Property Graphs* (LPG) che è parzialmente incompatibile con quello tradizionalmente usato in ambito ontologico. A tal fine, è stato definito un formalismo ontologico intermediario che consenta di esprimere il modello dei DB a grafo, ma possa essere tradotto nei formalismi ontologici classici (attualmente, OWL) così da consentire l'applicazione di tecniche di ragionamento automatico.

Questa visione è stata implementata nel framework Gr@phBRAIN ([7]). Caratteristica interessante del framework è che può gestire più ontologie sullo stesso grafo, il che consente di creare delle "viste" sulle informazioni che esprimono particolari prospettive o domini. Tuttavia, essendo il grafo sottostante unico, le informazioni presenti in ciascun dominio si collegheranno con quelle di altri domini, arricchendosi a vicenda. Attualmente il sistema è in via di ulteriore sviluppo e di espansione, e contiene informazioni relative ai domini MAB, turistico e della storia dell'Informatica ([5];[6])⁴. Sebbene sia ancora in costruzione, attualmente la nostra ontologia include 61 classi e 161 relazioni. La consistenza complessiva è di 336483 istanze di classe, descritte da un totale di 1875571 valori di attributi, e di 496564 istanze di relazione, descritte da un totale di 41301 valori di attributi. La Figura 1 mostra una sezione del grafo che connette istanze di entità tramite istanze di relazioni. Colori diversi dei nodi rappresentano classi diverse.

⁴ Un'applicazione Web dimostrativa che consente di esplorare e interagire con la base di conoscenza è disponibile all'indirizzo <http://193.204.187.73:8088/GraphBRAIN/>

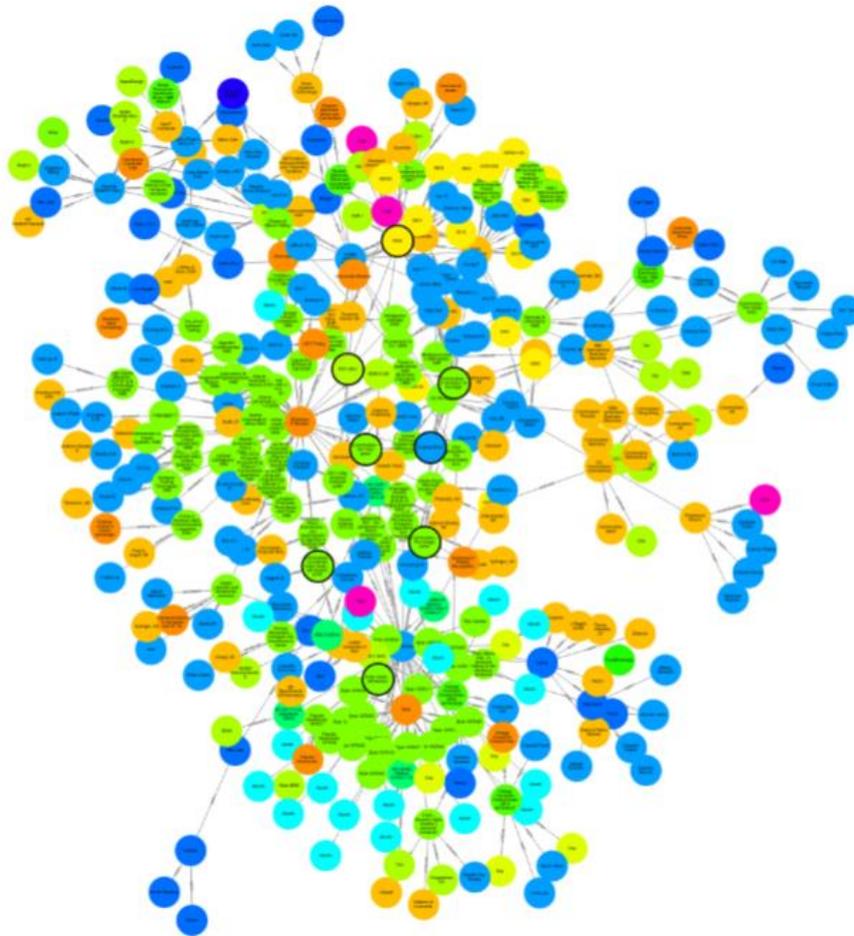


Figura 1. Sezione del grafo

L'ontologia che si sta definendo per la descrizione dei beni MAB e dei relativi oggetti digitali prevede un insieme di concetti e di relazioni che superano quanto prospettato finora, sia nella descrizione degli artefatti analogici, sia in ambito LOD. Una parte è finalizzata a descrivere il ciclo di vita degli oggetti digitali, incluse le attività e gli attori che ne fanno parte a vario titolo. Altri elementi ampliano la gamma delle informazioni di contesto oggi utilizzate. Sono previsti elementi per descrivere e gestire non solo i beni MAB, ma anche gli utenti e le loro caratteristiche, così da adattare il comportamento del sistema alle specifiche esigenze e finalità di ciascuno.

Usando varie tecniche dell'IA⁵, la conoscenza espressa nel sistema attiva un'interazione avanzata degli utenti con le informazioni, favorendo, a esempio:

- *ritrovamento di informazioni*⁶;
- *question answering*⁷;
- *clustering*⁸;
- *raccomandazione*⁹;

⁵ A es., Ragionamento automatico, Data Mining, Analisi di reti sociali, Modellazione di utenti, Raccomandazione, ecc.

⁶ Una medesima interrogazione produce risultati diversi, o organizzati diversamente, per diversi utenti, evidenziando per ciascuno i contenuti più rilevanti. A es., cercando ulteriori opere d'arte simili a una di riferimento, per alcuni la similarità potrà tenere in maggior conto i materiali realizzativi, per altri lo stile.

⁷ Gli utenti possono porre domande relative ai beni culturali e ottenere risposte "su misura". A es., chiedendo del Colosseo, un utente più interessato alla storia potrà ricevere informazioni sugli usi che esso ha avuto nel tempo, mentre un altro più interessato all'architettura potrà ricevere informazioni sullo stile e le caratteristiche costruttive.

⁸ È possibile individuare gruppi di artefatti culturali omogenei basandosi non solo sui loro metadati formali, ma anche su informazioni di contesto. A es., si possono ottenere informazioni su tutte le opere di scrittori che nella vita hanno avuto relazioni con Lecce.

⁹ Ciascun utente può ricevere suggerimenti su opere, autori o altri elementi verosimilmente di suo interesse, in relazione a formazione, preferenze, obiettivi, interessi, etc.

- supporto ai ricercatori e ai fruitori¹⁰.

BIBLIOGRAFIA

- [1] Bailey, Lila. 2015. «Digital Orphans: The Massive Cultural Black Hole on Our Horizon». Tech dirt. 13 ottobre 2015. <https://www.techdirt.com/articles/20151009/17031332490/digitalorphans-massive-cultural-blackhole-our-horizon.shtml>.
- [2] Barbuti, Nicola. 2020a. «Creating Digital Cultural Heritage with Open Data: From FAIR to FAIR5 Principles». *Digital Libraries: The Era of Big Data and Data Science. Communications in Computer and Information Science* 1177: 1–9.
- [3] ———. 2020b. «Ripensare i dati come risorse digitali: un processo difficile?» In *Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*, 19–20. Milano: Università Cattolica del Sacro Cuore.
- [4] Duranti, Luciana, e Elizabeth Shaffer. 2012. «The memory of the world in the digital age: digitization and preservation». In *UNESCO Conference Proceedings*,. http://ciscra.org/docs/UNESCO_MOW2012_Proceedings_FINAL_ENG_Compressed.pdf.
- [5] Ferilli, Stefano. 2021. «Integration Strategy and Tool between Formal Ontology and Graph Database Technology». *Electronics* 10 2616.
- [6] Ferilli, Stefano, e Domenico Redavid. 2020a. «An ontology and knowledge graph infrastructure for digital library knowledge representation». *Digital Libraries: The Era of Big Data and Data Science. Communications in Computer and Information Science* 1177: 47–61.
- [7] ———. 2020b. «The GraphBRAIN System for Knowledge Graph Management and Advanced Fruition». *Foundations of Intelligent Systems*. 12117: 308–17.
- [8] Tomasi, Francesca. 2017. «La preservazione del contenuto degli oggetti digitali: formalizzare la provenance». *Bibliothecae.it* 6: 17–40.

¹⁰ Il sistema può scoprire relazioni non evidenti ed utilizzarle per proporre nuove ipotesi di ricerca o di approfondimento; questo sarà possibile grazie alla capacità del sistema automatico di gestire ed incrociare quantità di dati molto maggiori di quanto possano fare gli esseri umani. A es., se due autori in anni vicini hanno prodotto opere sullo stesso argomento, e il sistema scoprisse che anni prima entrambi hanno lavorato a un progetto su di esso, potrebbe proporre di approfondire le eventuali influenze reciproche fra tutti questi elementi, e fra questi ed altri autori.

Sulla funzionalità di un'ontologia della filosofia alto medievale.

Il caso dei «Moralia in Iob» di Gregorio Magno

Armando Bisogno¹, Gianmarco Bisogno², Raffaella D'Urso³, Renato de Filippis⁴, Giuseppe Palermo⁵,

Timothy Tambassi⁶

¹Università degli Studi di Salerno, Italia – arbisogno@unisa.it

²Università degli Studi di Salerno, Italia – gibisogno@unisa.it

³Università degli Studi di Salerno, Italia – rduroso@unisa.it

⁴Università degli Studi di Salerno, Italia – rendefil@yahoo.it

⁵Università degli Studi di Salerno, Italia – gipalermo@unisa.it

⁶Università degli Studi di Salerno, Italia – ttambassi@unisa.it

ABSTRACT

Obiettivo del progetto OPhEMA (Ontology for the Philosophy of Early Middle Ages) è contribuire all'ampliamento dell'ontologia della filosofia in OWL di Barry Smith e Pierre Grenon, esplorando la possibilità che in essa possano rientrare autori e testi latini, collocati tra il secondo e il dodicesimo secolo d.C., attualmente non inclusi. La scelta dell'arco cronologico di riferimento è legata all'individuazione, in esso, di un contesto problematico: se nel mondo greco (e fino all'epoca tardo-antica) il termine “filosofia” ha indicato, per esempio, una specifica professione o un particolare stile di vita, la ricezione (avvenuta tramite Cicerone) del termine nel periodo patristico e altomedievale non preserva, tuttavia, il medesimo significato, in quanto non è più possibile parlare di una disciplina istituzionalizzata nelle forme e specificatamente legata a una produzione scritta. Ciononostante, gli studiosi successivi hanno associato a tale arco cronologico un'attività speculativa definita “filosofica”, lasciando così intuire l'esistenza di criteri secondo i quali testi non dichiaratamente filosofici possano essere comunque giudicati tali. Fine ultimo del contributo è perciò mostrare i diversi passaggi che, a partire dalla peculiarità del contesto d'applicazione e dal caso specifico dei «Moralia in Iob» di Gregorio Magno (540 d.C.-604 d.C.), conducono alla costruzione di un'ontologia della filosofia altomedievale concepita come modulo dell'ontologia della filosofia di Grenon e Smith, con particolare attenzione al suo essere un potenziale strumento tanto di ri-valorizzazione di singoli testi (resa possibile dall'utilizzo di strumenti digitali) quanto di un significativo ampliamento dell'ontologia già esistente.

PAROLE CHIAVE

Filosofia, medioevo, Gregorio Magno, ontologia.

INTERVENTO

1. INTRODUZIONE

Nel mondo greco e fino all'epoca tardo-antica, il termine “filosofia” ha indicato una specifica professione o un particolare stile di vita. Complici la rara conoscenza del greco, l'assenza dei dialoghi platonici e delle opere non logiche di Aristotele, la scarsa circolazione dei testi neoplatonici coevi e, in generale, una conoscenza frammentaria delle dottrine delle filosofie ellenistiche, la ricezione (avvenuta tramite Cicerone) del termine nel periodo patristico e altomedievale non preserva, tuttavia, nell'Occidente latino il medesimo significato. “Filosofia” non indica né una dottrina specifica, né un *corpus* di testi, né una professione, bensì una non ben definita tradizione del passato, che è di conseguenza dagli autori attivi tra il secondo e il dodicesimo secolo d.C. recepita in modi talvolta diametralmente opposti. Solo nel tredicesimo secolo la filosofia riacquista, con la riscoperta nel mondo latino dell'intero *corpus* aristotelico, una definizione ben precisa, vale a dire quella di dottrina specifica insegnata nelle neonate università, restando tale fino ai giorni nostri. Gli studiosi successivi hanno tuttavia associato al periodo patristico e altomedievale un'attività speculativa definita “filosofica”, un'attribuzione *ex post* che prescinde dal fatto che nell'arco cronologico considerato gli autori non si autodefiniscano “filosofi” (o le cui opere non sono dalla tradizione sempre state giudicate “filosofiche”).

Le diverse letture storiografiche concernenti i «Moralia in Iob» di Gregorio Magno, opera redatta verso la fine del sesto secolo, riflettono efficacemente la complessità evidenziata a proposito del rapporto tra filosofia e Alto Medioevo: se da un lato, infatti, l'opera è raramente inclusa – insieme al suo autore – nei manuali di storia di filosofia o, in generale, in studi che la considerino dal punto di vista filosofico, dall'altro vi sono casi in cui la letteratura secondaria si riferisce alla

presenza, nei «Moralia in Iob», dell'elemento della tragicità umana ([1]) e di una certa considerazione del problema del male ([4]), nonché al suo essere, tra gli altri, luogo di espressione di un vero e proprio «pensiero gregoriano» ([2]). Tali aspetti rientrano certamente nell'odierna idea di filosofia, che tuttavia non coincide con quanto espresso sull'argomento da Gregorio Magno: nell'opera, infatti, le occorrenze del termine «philosophia» e dei suoi derivati rimandano a una considerazione negativa dei filosofi, giudicati portatori di una dottrina sempre inadeguata se reputata valida indipendentemente dalla rivelazione cristiana e, come sempre accade in ambito altomedievale, in ogni caso a essa inferiore; l'unica valutazione positiva della filosofia coincide con la descrizione della nozione di «vera philosophia», considerata un lodevole modello di comportamento seguito dal protagonista del testo biblico ma che, in ogni caso, differisce da quanto (non) individuato dagli studiosi successivi che (non) ne parlano da una prospettiva filosofica.

D'altra parte, la complessità della struttura e il gran numero di temi trattati che caratterizzano il micro-universo dei «Moralia in Iob» li rendono, insieme alla già citata individuazione in essi di un qualche tipo di riflessione filosofica, al contempo il luogo di partenza ideale per la costruzione di un'ontologia della filosofia altomedievale, che può in tale contesto costituire, per mezzo dei passaggi di seguito riportati, uno strumento utile allo sviluppo di una lettura complessivamente inedita dell'opera (resa possibile, come si vedrà, dall'utilizzo di strumenti digitali) e, insieme, all'arricchimento di un modulo potenzialmente applicabile anche ad altri testi e autori.

Il lavoro sui «Moralia in Iob» si inserisce nel più ampio contesto del progetto OPhEMA (Ontology for the Philosophy of Early Middle Ages), promosso dal centro FiTMU e dal dottorato RAMUS dell'Università degli Studi di Salerno. Il progetto, infatti, mira a indagare la complessità dei significati assunti dal termine latino «philosophia» e dai suoi derivati nella letteratura latina prodotta tra il secondo e il dodicesimo secolo d.C., così da avere una prima immagine dei diversi modi in cui essa veniva concepita nel periodo di riferimento. A tale passaggio segue lo sviluppo di un'ontologia della filosofia altomedievale che, intesa come modulo dell'ontologia della filosofia di Smith e Grenon, comprenda le caratteristiche in base alle quali autori mai proclamatisi filosofi sono stati (o possano eventualmente esserlo in futuro) ritenuti tali dagli studi storico-filosofici.

Questo l'iter che ha condotto alla costruzione di una forma embrionale dell'ontologia:

a) **Individuazione di temi, metodi, termini e concetti genericamente considerati 'filosofici'** (es. libertà, linguaggio, conoscenza). Essi hanno rappresentato i primi elementi della classe (già prevista nell'ontologia di Smith e Grenon) *Concetti filosofici* (figura 1), senza tuttavia esaurirne immediatamente il contenuto ulteriormente ampliato, insieme a quello delle classi (anch'esse già previste da Smith e Grenon) *Metodi filosofici* e *Persone*, in seguito alla (b) lettura del testo latino dei «Moralia in Iob».

b) **Creazione di tag, semantizzazione del testo, verifica dell'affidabilità dei tag e integrazioni.** Tutti gli elementi raccolti (a) sono stati convertiti in *tag* (intesi come 'etichette' applicabili a un testo in formato .html) tramite il plug-in *Hypothes.is* e associati, per mezzo dello stesso, a porzioni di testo in cui l'argomento indicato dal *tag* è effettivamente trattato. L'operazione ha previsto, in modo da stabilire una corrispondenza che fosse quanto più possibile accurata, la sostituzione di alcune diciture (es. *libero arbitrio* anziché *libertà*), l'introduzione di divisioni interne (es. quella tra *linguaggio interiore* e *linguaggio esteriore* o tra *conoscenza corrotta* e *conoscenza mancante*) o, ancora, l'aggiunta di altri elementi alle classi di cui sopra, avvenuta sulla base di quanto suggerito dal testo stesso. Tra i *tag* integrati vi sono, per esempio: *ordo*, *Dio*, *creazione* e *ragione* per la classe *Concetti filosofici*; *etimologia*, *esegesi biblica* e *analogia* per la classe *Metodi filosofici*; *Girolamo di Stridone* per la classe *Persone*.

c) **Creazione dell'ontologia.** L'insieme completo, verificato e revisionato dei *tag*, in parte già reso graficamente tramite il software *XMind* (figura 2) e all'interno del quale sono state, sempre sul modello proposto da Grenon e Smith (figura 3), stabilite precise relazioni, costituirà a fine lettura dei «Moralia in Iob» la forma embrionale di un'ontologia della filosofia altomedievale, che sarà successivamente integrata con i risultati ottenuti dall'applicazione del medesimo processo a testi latini che presentano le medesime criticità riscontrate, tanto a livello testuale quanto storiografico, nel testo di Gregorio Magno.

evidenziati, un contributo significativo allo studio e alla rivalutazione di una delle opere più tramandate nel Medioevo, fornendo gli strumenti necessari a esplorarne ulteriormente la ricchezza e a integrare i risultati raggiunti nell'ambito della medievistica tradizionale.

BIBLIOGRAFIA

- [1] D'Onofrio, Giulio. 2013. *Storia del pensiero medievale*. Città Nuova.
- [2] Evans, Gillian Rosemary. 1986. *The thought of Gregory the Great*. Cambridge University Press.
- [3] Grenon, Pierre, e Barry Smith. 2011. «Foundations of an Ontology of Philosophy». *Synthese* 182 (2): 185–204.
- [4] Stump, Eleonore. 2014. *The problem of evil*. Robert Pasnau. Vol. 2. Cambridge University Press.
- [5] Tambassi, Timothy. 2021. *On the Content of Information-system Ontologies*. Acta Analytica.
- [6] ———. s.d. «Intrecci meta-ontologici tra filosofia e informatica». *Giornale di metafisica* in printing.

La Visualizzazione Grafica di Sensi e Relazioni Semantiche di un Lessico Computazionale della Lingua Italiana

¹Marianna Colombo, ²Emiliano Giovannetti

^{1,2}Istituto di Linguistica Computazionale “A. Zampolli” - CNR, Italia,

¹marianna.colombo@ilc.cnr.it

²emiliano.giovannetti@ilc.cnr.it

ABSTRACT

La visualizzazione, intesa come ausilio alla consultazione e alla ricerca, può giocare un ruolo fondamentale nella fruizione di dati linguistici e di conoscenza. Attraverso la rappresentazione di una risorsa lessicale o di una ontologia sotto forma di grafo, ad esempio, uno studioso ha la possibilità di farsi un'idea immediata delle relazioni che intercorrono tra i vari elementi di un lessico (di lingua o specialistico) o dei concetti che formalizzano un determinato dominio. In questo contributo presentiamo un'applicazione *web based* per la visualizzazione grafica della componente semantica di un lessico computazionale dell'italiano. Il grafo è stato sviluppato con Cytoscape.js, una libreria Javascript per la creazione di grafi. L'interfaccia al grafo, sviluppata come applicazione Angular, offre un primo insieme di funzionalità di interazione per la navigazione e la manipolazione della risorsa lessico-semantica.

PAROLE CHIAVE

Visualizzazione grafica di risorse linguistiche, rappresentazione della conoscenza, human-computer interaction, lessico computazionale, grafi.

INTERVENTO

1. INTRODUZIONE

Al giorno d'oggi l'impiego sempre maggiore di soluzioni grafiche per presentare dati e informazioni ha ottenuto un responso decisamente positivo in funzione della capacità che offre di stimolare l'attenzione dell'utente migliorandone notevolmente la *user experience* ([2]). Nel campo delle ricerche che coinvolgono grandi quantità di dati linguistici e ontologici il supporto visivo si sta affermando come elemento strategico e di grande impatto ([3];[6]). Nell'ambito della visualizzazione di dati linguistici, già da molti anni sono disponibili strumenti per la rappresentazione di strutture sintattiche sotto forma di alberi ([9]). Tuttavia, l'adozione della *graph visualization* in ambito linguistico è una tecnica abbastanza recente: si menzionano di seguito alcuni progetti nei quali viene sottolineata l'utilità di questo tipo di visualizzazione. Sia nei progetti di seguito illustrati che nel presente lavoro l'approccio grafico ha una natura “esplorativa” (da *exploratory graphics* [5]): lo scopo della visualizzazione, infatti, non è quello di presentare l'informazione in modo statico e schematico (come avviene nella *presentation graphics*) ma di permettere a un utente di effettuare attivamente ricerche tra i dati disponibili e analizzarne la natura. La creazione di un'interfaccia esplorabile dall'utente trova una corrispondenza in ([11]) nel quale i dati lessicali vengono presentati tramite una GUI (*Graphic User Interface*). A differenza del nostro progetto, in cui viene analizzata una porzione di lessico dell'italiano, questa interfaccia permette di esplorare i dati relativi a DiCoInfo, un dizionario online che descrive termini nei campi dell'informatica e di Internet in francese, inglese e spagnolo. All'autore preme sottolineare come l'impiego dei grafi per mostrare le relazioni tra termini sia finalizzato ad una migliore comprensione della risorsa da parte dell'utente¹.

Nel caso di dataset ampi la possibilità di poter zoomare su un particolare gruppo di elementi consente di avere una visione più chiara dei collegamenti che sussistono tra un numero circoscritto di nodi escludendo la parte di grafo che non interessa. In ([1]), per esempio, gli autori descrivono l'applicazione WordNet Atlas concepita per visualizzare il database WordNet in una mappa zoomabile. Più in generale, le *Zooming User Interfaces* (ZUI) costituiscono un tipo di interfaccia grafica che offre all'utente l'abilità di zoomare e spostare l'ambiente (*pan*) a seconda delle proprie necessità. I dati sono così rappresentati con diversi livelli di dettaglio secondo il livello di ingrandimento scelto.

¹ “We assumed that relationships between terms (perhaps not all, but a large part of them) were likely to be better understood by end users if they were first shown graphically rather than simply listed in tables with textual explanations.” ([10]: 244).

In ([10]) le azioni di *panning* e *zooming* consentono di esplorare le associazioni lessicali presentate in un grafo diretto flessibile in cui i concetti sono visualizzati come nodi e le *crosslinguistic colexifications* come archi.

Gli studiosi del progetto esposto in ([4]) e ([3]) attribuiscono la medesima importanza all'aspetto estetico e funzionale della *graph visualization* esplicitata attraverso tre criteri da seguire per una rappresentazione funzionale e visivamente appetibile ([4]: 468): i) l'efficacia nella visualizzazione (es. garantire che la rappresentazione grafica comprenda tutto il set di dati disponibile); ii) l'efficienza nella visualizzazione (es. evitare informazioni ridondanti e la proliferazione di elementi grafici inutili); iii) la chiarezza della visualizzazione (es. scelta dei colori, livello di luminosità e contrasto adeguato).

Questi principi si ritrovano nei progetti descritti in ([13]) e ([8]) in cui i modelli grafici sono stati sviluppati al fine di aiutare l'utente a comprendere meglio la struttura delle risorse terminologiche e delle relative relazioni semantiche.

Quanto finora citato dimostra come le tecnologie di visualizzazione dei dati possano essere utilmente integrate nell'ambito della linguistica e quali benefici possano portare a livello di studio e di apprendimento.

Il lavoro qui presentato si concentra sull'applicazione di tecniche di visualizzazione per la consultazione del *layer* semantico di un lessico computazionale dell'italiano, la cui struttura generale (seppur non ancora trattata in questo primo lavoro nella sua interezza) risulta molto più articolata della risorsa WordNet utilizzata in molti dei lavori citati, soffermandosi sull'aspetto estetico e funzionale. Le tecnologie impiegate nello sviluppo del modello e dell'interfaccia, nativamente concepite per fornire all'utente un alto livello di usabilità, sono state utilizzate per rendere il più possibile chiara e comprensibile la rete di relazioni che lega i vari elementi semantici considerati. La relativa applicazione, seppure ad uno stato di sviluppo ancora prototipale, può essere liberamente consultata online².

2. LA TECNOLOGIA UTILIZZATA

Nella teoria dei grafi, un grafo è definito come una struttura relazionale formata da un numero finito di vertici (o nodi) e un numero finito di segmenti (archi o spigoli) che collegano ogni nodo agli altri³.

Visivamente, i grafi sono rappresentati da un punto (o cerchio) per ogni vertice, e una linea (o arco) per ogni segmento che connette due nodi. Nel caso di un grafo diretto, la direzione del segmento è indicata da una freccia.

Lo sviluppo di un'interfaccia per la visualizzazione di grafi ha richiesto una fase preliminare di studio delle tecnologie disponibili, nella quale sono state analizzate le tecnologie *open* più recenti che potessero essere agevolmente integrate in un ambiente di sviluppo Angular⁴, uno dei framework *open source* più utilizzati per lo sviluppo front-end. La scelta è ricaduta su Cytoscape.js⁵, una libreria *open-source* in Javascript con licenza MIT per la visualizzazione e analisi di grafi ([7]). Mediante questa libreria è possibile visualizzare e manipolare svariate tipologie di grafi, tra cui grafi diretti (*directed graphs*), grafi non diretti (*undirected graphs*), grafi misti (*mixed graphs*), loops, multigrafi (*multi graphs*) e grafi composti (*compound graphs*). Le interazioni disponibili, implementabili dal programmatore, agevolano la *user-experience* permettendo all'utente varie azioni quali: *grab and drag*, zoom in e zoom out, selezione con il *tap*, ecc. Gli elementi che costituiscono il grafo vengono forniti con un file JSON e organizzati in un *array* di nodi e archi.

Il posizionamento dei nodi dipende dal tipo di layout impiegato: Cytoscape.js offre diversi modelli di layout ognuno dei quali è integrato con un proprio specifico algoritmo che influenza la disposizione dei nodi e la lunghezza degli archi, modificando perciò la forma finale del grafo.

3. IL LESSICO COMPUTAZIONALE

La risorsa linguistica che è stata presa in esame per questo lavoro è Parole-Simple-Clips (PSC), un lessico computazionale dell'italiano contemporaneo ([12]) disponibile sotto forma di database MySQL su CLARIN⁶. A differenza di WordNet, già oggetto di esperimenti di visualizzazione come mostrato nell'introduzione, PSC si presenta come una risorsa particolarmente ricca di informazione linguistica articolata su quattro livelli: fonologico, morfologico, sintattico e semantico. Ognuno di questi livelli è descritto attraverso unità informative specifiche (unità fonologiche, unità morfologiche, e così via) corredate di numerose proprietà e correlate tra loro. In questo contesto sono state considerate le unità semantiche che popolano il *layer* semantico, visualizzate nel grafo come nodi e un primo set di relazioni semantiche (iponimia, meronimia e sinonimia), visualizzate come archi tra tali nodi. Questo insieme di dati iniziale è stato estratto dal database e serializzato in JSON come descritto nella sezione seguente.

² <https://klab.ilc.cnr.it/demoLexiconGraph/> (si consiglia l'utilizzo del browser Chrome).

³ <https://www.andreaminini.com/ricerca-operativa/teoria-dei-grafi/grafico>.

⁴ <https://angular.io/>.

⁵ <https://js.cytoscape.org/>.

⁶ <https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-88>.

4. LE UNITÀ SEMANTICHE SOTTO FORMA DI GRAFO

Nel file JSON che rappresenta la serializzazione del grafo ogni elemento (nodo o arco) è stato descritto mediante alcune proprietà alle quali è stato possibile collegarsi programmaticamente nella fase di implementazione.

Le proprietà che descrivono un nodo sono le seguenti:

- id*: l'identificatore univoco dell'unità semantica così come appare nella risorsa PSC (es: "USemD2389fiore");
- label*: la parola associata all'unità semantica in questione (es: "fiore");
- definition*: la definizione dell'unità semantica;
- pos*: la parte del discorso (*part of speech*).

Le proprietà di un arco sono le seguenti:

- id*: l'identificatore univoco dell'arco (es. "USem70823anemone isA USemD2389fiore");
- source*: l'id del nodo di partenza (es. "USem70823anemone");
- target*: l'id del nodo di arrivo (es. "USemD2389fiore");
- relationType*: il tipo di relazione (es. "hyponymy");
- label source/target*: le parole associate ai sensi relativi ai nodi source e target (es. "anemone", "fiore").

Il criterio per la disposizione delle unità semantiche all'interno del grafo segue un'idea precisa: i nodi uscenti (gli iperonimi nella relazione di iponimia, gli olonimi nella relazione di meronimia) vengono posizionati a un livello più alto rispetto all'unità di interesse che si vuole analizzare, mentre i nodi entranti (iponimi nella relazione di iponimia e meronimi nella relazione di meronimia) sono posizionati a un livello inferiore. Il posizionamento dei nodi cambia a seconda del numero di elementi associati all'unità analizzata:

- nel caso di un numero di archi (entranti o uscenti) inferiori a 20, i nodi vengono posizionati a un'equa distanza sull'asse delle ordinate tenendo al centro il nodo analizzato (figura 1); l'algoritmo per il calcolo della disposizione dei nodi sfrutta la funzione *nodes.positions()*⁷ della libreria Cytoscape.js che permette di recuperare le posizioni sull'asse *x* e *y* di un nodo;

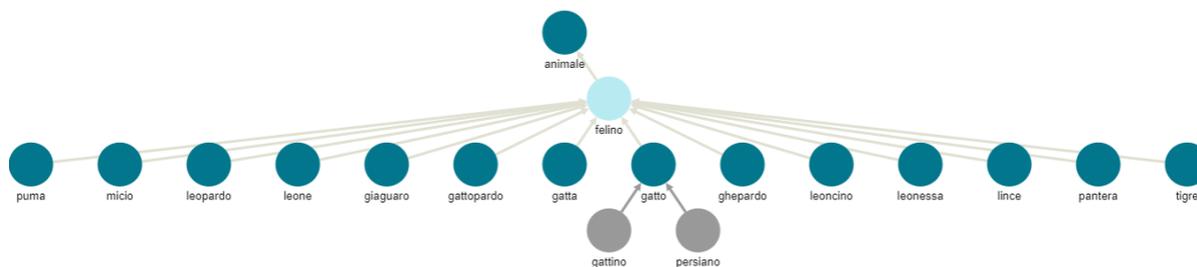


Figura 1. Disposizione di nodi in numero inferiore a 20

- nel caso di un numero di archi maggiore di 20, si è optato per il layout a griglia (*grid layout*⁸) presente nella libreria che, come suggerisce il nome, dispone gli elementi in una griglia ordinata; nel nostro modello sono state aggiunte delle proprietà per avere un controllo maggiore sull'effetto visivo, come la funzione *avoid overlap* al fine di prevenire la sovrapposizione dei nodi (figura 2).

⁷ <https://js.cytoscape.org/#nodes.positions>.

⁸ <https://js.cytoscape.org/#layouts/grid>.

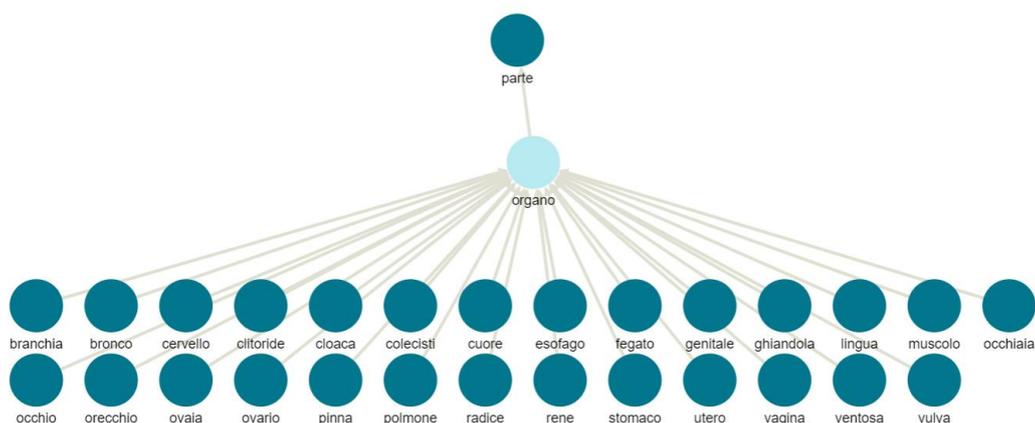


Figura 2. Disposizione di nodi in numero superiore a 20

Nella creazione del grafo sono stati adottati alcuni accorgimenti estetici al fine di agevolare l'utente nella navigazione della risorsa.

Prendendo come riferimento quanto espresso in ([4]) e ([3]), per rendere i vari elementi della risorsa più facilmente riconoscibili sono state innanzitutto effettuate delle scelte sui colori. Le unità semantiche, infatti, assumono un colore specifico in funzione delle diverse relazioni nelle quali sono coinvolte. L'unità di interesse, ossia quella che viene selezionata mediante doppio clic, è sempre colorata di una tonalità più chiara rispetto alle altre per sottolinearne l'importanza (figura 3). Nella scelta dei colori, inoltre, si è tenuto conto dell'accessibilità da parte di utenti daltonici scegliendo tonalità che non risultassero, dal punto di vista della luminosità, troppo simili tra loro.

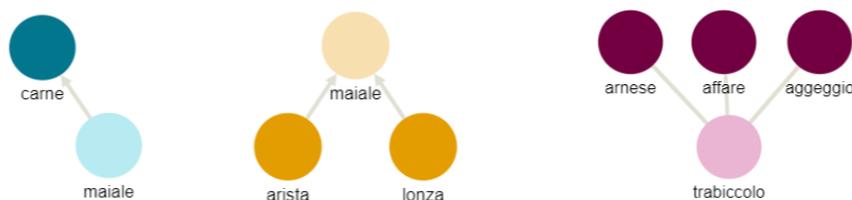


Figura 3. I colori utilizzati per rappresentare le relazioni di iponimia, meronimia e sinonimia

Nella navigazione del grafo, i nodi già espansi che non sono più direttamente collegati con il nodo di interesse (i.e. nodi posti a distanza due o più) si colorano di grigio per far risaltare all'utente le relazioni dirette (si vedano i nodi in grigio della figura 1). Nel caso in cui i nodi grigi aperti superino il numero di 20, vengono nascosti così da non appesantire troppo il *viewport* e confondere l'utente.

5. L'INTERFACCIA UTENTE

La struttura dell'applicazione di interfaccia al grafo si compone di due colonne laterali, un *header* orizzontale dotato di bottoni di funzionalità e una sezione centrale per la visualizzazione del grafo.

L'*header* è provvisto di una bottoniera associata a diverse funzionalità di navigazione del grafo, per: i) nascondere i nodi, ii) centrare il grafo nel *viewport*, iii) effettuare *zoom in / zoom out*. Più a destra sono collocati tre *radiobutton* che permettono di selezionare uno dei tre tipi di relazione semantica disponibili. In base al tipo di relazione selezionata vengono visualizzati solamente i nodi delle risorse associate e gli archi con la proprietà *relationType* corrispondenti alla relazione specificata.

La colonna a sinistra mostra la lista delle unità semantiche disponibili. Esse sono filtrabili dall'utente tramite una apposita casella di testo: alla selezione (clic) di una unità viene visualizzato il nodo corrispondente nella sezione centrale. L'utente può analizzare le singole unità semantiche e le varie relazioni visualizzando le proprietà ad esse associate cliccando sui rispettivi nodi e archi: il clic singolo su un elemento apre un pannello nella colonna di destra con l'elenco delle proprietà e i relativi valori.

La navigazione della risorsa può avvenire attraverso due modalità principali. Data una unità semantica visualizzata nella sezione centrale:

1. con *doppio clic*: vengono visualizzati i nodi relativi alle istanze di relazione semantica (sulla base della relazione preselezionata nell'*header*) che la collegano ad altre unità;

2. con il *clic del tasto destro*: si apre un *context-menu* a fianco dell'unità selezionata che offre le seguenti funzioni:
- *apri/chiedi nodi entranti*: nel caso della relazione di iponimia vengono visualizzati/nascosti gli iponimi del senso selezionato, per la relazione di meronimia i meronimi corrispondenti e per la sinonimia i sinonimi;
 - *apri/chiedi nodi uscenti*: per la relazione di iponimia vengono visualizzati/nascosti gli iperonimi, per la relazione di meronimia gli olonimi e per la sinonimia i sinonimi;
 - *apri/chiedi tutti i nodi*: vengono visualizzati/nascosti iponimi e iperonimi nella relazione di iponimia, meronimi e olonimi nella relazione di meronimia e i sinonimi per la sinonimia;
 - *rimuovi nodo*: il nodo viene rimosso dal grafo; questa funzione facilita la navigazione del grafo nel caso in cui vi sia la necessità di rimuovere nodi senza perdere traccia della ricerca effettuata fino a quel momento.

La figura 4 offre una panoramica generale delle funzionalità che l'interfaccia espone come descritte in questo paragrafo.

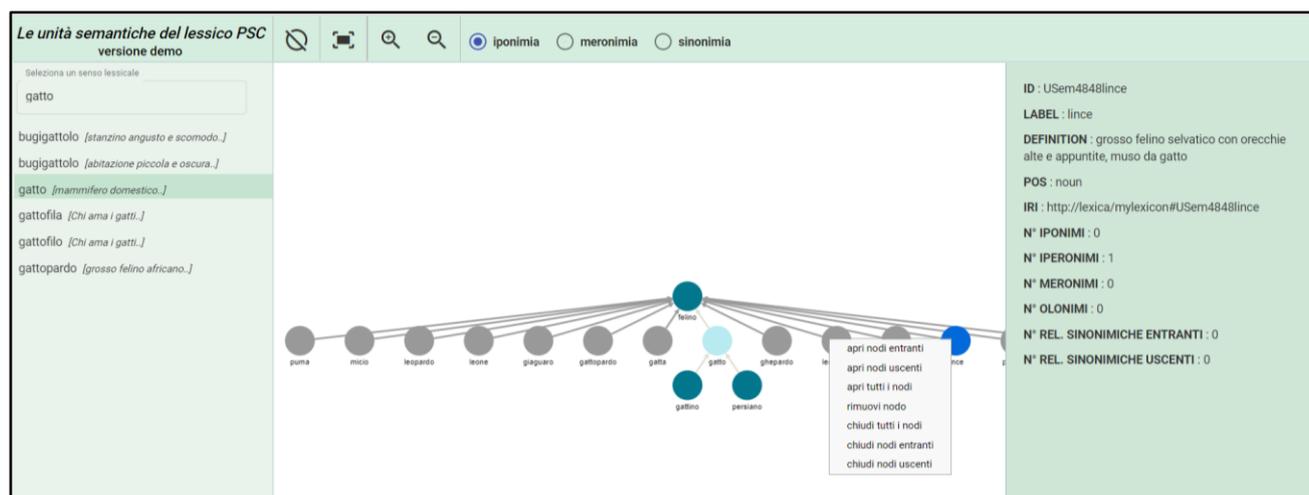


Figura 4. Panoramica dell'interfaccia utente

Come esposto in ([1]) le *ZUI* consentono di controllare il livello di dettaglio aumentando o diminuendo lo zoom e spostando l'ambiente (*pan*) a secondo le necessità. Nell'applicazione qui illustrata l'utente può modificare il grado di zoom attraverso la rotellina del mouse o tramite i bottoni di funzionalità presenti nell'*header*. L'azione di *pan* si attiva mantenendo premuto il tasto sinistro del mouse e spostando il puntatore nella direzione voluta. Infine, è sempre possibile, tramite il pulsante "centra", impostare automaticamente *zoom* e *pan* in modo tale che il grafo appaia contenuto e adattato alla finestra.

6. DISCUSSIONE E LAVORI FUTURI

Attraverso la visualizzazione grafica di una risorsa lessicale è possibile esplicitare in modo semplice e immediato gli elementi linguistici che la costituiscono e le relazioni che intercorrono tra di essi. In questo modo si intende agevolare la consultazione del lessico fornendo una vista del dato alternativa rispetto alla rappresentazione classica tabellare o *XML-like*. Per quanto riguarda i possibili profili utente ai quali una visualizzazione di questo tipo potrebbe essere utile, citiamo innanzitutto il lessicografo, che potrà esplorare graficamente il lessico seguendo i percorsi che conducono da un elemento linguistico agli altri elementi al fine di analizzarne la struttura e apportare gli eventuali aggiornamenti. Un altro profilo di riferimento potrebbe essere l'apprendente della lingua italiana L2, che avrà la possibilità di esplorare, visivamente, la rete di relazioni (per il momento semantiche, ma in seguito anche di altra natura) che intercorrono tra i vari elementi linguistici (fonologici, morfologici, sintattici e semantici) del lessico.

Il lavoro descritto nel presente articolo rappresenta la fase preliminare di una attività finalizzata a sviluppare una interfaccia liberamente accessibile per la navigazione grafica e la manipolazione dell'intero lessico computazionale PSC della lingua italiana. Affinché ciò diventi possibile si prevede di articolare il prosieguo dell'attività di ricerca e sviluppo nei seguenti passi: i) integrazione di tutte le relazioni semantiche disponibili (la risorsa ne annovera 138), ii) aggiunta, progressiva, degli altri livelli linguistici, a partire da quello morfologico (che consentirà innanzitutto di ancorare ogni senso al rispettivo lemma) e delle relative relazioni, iii) adeguamento dell'interfaccia dalla sola modalità di visualizzazione a quella di editing, mediante la quale sarà possibile correggere eventuali errori presenti nella risorsa oppure di estenderla con nuovi elementi, fonologici, morfologici, sintattici e semantici.

7. RICONOSCIMENTI

Il presente lavoro è stato condotto nel contesto del Progetto TALMUD e nell'ambito della cooperazione scientifica tra S.c.ar.l. PTTB e ILC-CNR.

BIBLIOGRAFIA

- [1] Abrate, Matteo, Clara Bacciu, Andrea Marchetti, e Maurizio Tesconi. 2012. «WordNet Atlas: a web application for visualizing WordNet as a zoomable map». In *Proceedings of the 6th International Global WordNet Conference*, 23–29. Matsue, Japan.
- [2] Bahador, Saket, Alex Endert, e John Stasko. 2016. «Beyond Usability and Performance: A Review of User Experience-focused Evaluations in Visualization». In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 133–42. New York - USA.
- [3] Caldarola, Enrico Giacinto, Antonio Picariello, e Antonio Maria Rinaldi. 2016. «Experiences in WordNet Visualization with Labeled Graph Databases». *Communications in Computer and Information Science* 631: 80–99.
- [4] Caldarola, Enrico Giacinto, Antonio Picariello, Antonio Maria Rinaldi, e Marco Sacco. 2016. «Enhancing the WordNet Exploration and Visualization in Neo4J with a Tag Cloud Based Approach». *International Journal On Advances in Intelligent Systems* 9: 462–71.
- [5] Chen, Chun-hou, Wolfgang Kar Härdle, e Antony Unwin. 2007. *Handbook of data visualization*. Springer Science & Business Media.
- [6] Dudáš, Marek, Steffen Lohmann, Vojtěch Svátek, e Dmitry Pavlov. 2018. «Ontology visualization methods and tools: A survey of the state of the art». *The Knowledge Engineering Review* 33 (E10).
- [7] Franz, Max, Christian T. Lopes, Gerardo Huck, e et al. 2016. «Cytoscape.js: a graph theory library for visualisation and analysis». *Bioinformatic* 32 (2): 309–11.
- [8] Katricheva, Nadezda, Alyaxey Yaskovich, Anastasiya Lisitsina, Tamara Zhordaniya, Andrey Kutuzov, e Elizaveta Kuzmenko. 2020. «Vec2graph: A python library for visualizing word embeddings as graphs». *Communications in Computer and Information Science* 1086: 190–98.
- [9] Little, Alexa N., e Stephen Tratz. 2016. «Easytree: A graphical tool for dependency tree annotation». In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 2343–47. Portorož, Slovenia.
- [10] Mayer, Thomas, Johann-Mattis List, Anselm Terhalle, e Matthias Urban. 2014. «An interactive visualization of cross-linguistic colexification patterns». In *Proceedings of VisLR: Visualization as Added Value in the Development, Use and Evaluation of Language Resources*. Reykjavik, Iceland.
- [11] Robichaud, Benoît. 2011. «A graph visualization tool for terminology discovery and assessment». In *Proceedings of the Fifth International Conference on Meaning-Text Theory*, 243–52. Barcelona, Spain.
- [12] Ruimy, Nilda, Monica Monachini, Raffaella Distanti, e et al. 2002. «Clips, a multi-level italian computational lexicon: A glimpse to data». In *Proceedings of the Third International Conference on Language Resources and Evaluation*. Las Palmas de Gran Canaria.
- [13] Suchowolec, Karolina, Piotr Bański, e Andreas Witt. 2018. «Bridging standards development and infrastructure usage by means of concept graphs: the liaison of CLARIN and ISO TC37SC4 in practice». In *Proceedings of VisLR III: Visualization as Added Value in the Development, Use and Evaluation of Language Resources*, 1–8. Miyazaki, Japan.

Sessione Testi 4

Hedy Lamarr

Dalla codifica alla fruizione: l'edizione digitale Bellini Digital Correspondence

Santa Pellino¹, Pietro Sichera², Angelo Mario Del Grosso³, Daria Spampinato⁴

¹ Università di Pisa, Italia - s.pellino1@studenti.unipi.it

² Istituto di Scienze e Tecnologie della Cognizione - CNR, Italia - pietro.sichera@istc.cnr.it

³ Istituto di Linguistica Computazionale "A. Zampolli" - CNR, Italia - angelo.delgrosso@ilc.cnr.it

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

ABSTRACT

L'articolo illustra le attività svolte per la realizzazione dell'edizione scientifica digitale delle lettere autografe belliniane (Bellini Digital Correspondence). Il progetto si inserisce in un contesto didattico finalizzato alla rappresentazione e fruizione della corrispondenza del maestro catanese. In seno alle attività è stato sviluppato un tool di gestione delle regioni d'interesse e la personalizzazione del software EVT.

PAROLE CHIAVE

Digital scholarly edition, correspondence, software design, XSLT, Vincenzo Bellini.

INTERVENTO

1. INTRODUZIONE

L'edizione scientifica digitale delle lettere autografe belliniane "Bellini Digital Correspondence" (BDC) si prefigge tanto la rappresentazione del corpus conservato presso il Museo civico Belliniano di Catania, attraverso un processo di codifica che segue il vocabolario e le linee guida della Text Encoding Initiative (TEI) ([6];[3]), quanto la conseguente fruizione via web mediante l'utilizzo del software EVT (Edition Visualization Technology)¹ per offrire un'edizione image-based ([5]). Altre fasi del progetto sono descritte in ([4]).

Il corpus consta di 40 unità testuali che sono state riprodotte in formato fototipico. Alcune lettere sono scritte in bifogli, altre in più fogli, per un totale di 35 unità codicologiche; infatti, alcune minute coesistono nello stesso folio. Le immagini riproducono le singole carte (recto, verso), per un totale di 111 riproduzioni, di cui 23 carte bianche. Al fine di garantire la correttezza scientifica delle trascrizioni, si è fatto riferimento all'edizione critica a stampa a cura Seminara ([1]), che raccoglie tutta la corrispondenza belliniana costituita, alla data di edizione, da 517 documenti. All'interno del carteggio belliniano (segnato con LL), la segnatura LL1 corrisponde alle missive scritte da Vincenzo Bellini. La lettera n-esima in ordine cronologico si indica con LL1.n. Il numero I o II, in coda alla segnatura (quando presente), specifica la coesistenza di più missive nella stessa carta (unità testuale).

Il progetto BDC si inserisce in un contesto didattico che prevede la codifica di buona parte del corpus da parte di studenti, tirocinanti e tesisti del corso di codifica dei testi dell'Università di Pisa, sulla base del modello descritto in [7]. Per questo motivo la codifica dei documenti del corpus non è uniforme e ha necessitato di un attento processo di armonizzazione per la realizzazione dell'intera edizione.

Nel suo insieme, il progetto ha quindi previsto diverse fasi redazionali per la produzione, analisi, elaborazione, presentazione e fruizione del repertorio belliniano oggetto d'indagine. Il modello proposto in questo contributo è del tutto generale e quindi riusabile in altri contesti simili.

2. METODOLOGIA

L'edizione del corpus belliniano si sviluppa in un articolato processo redazionale che ha coinvolto vari attori e prodotto molteplici risorse.

Per la realizzazione del testo delle missive in formato XML-TEI [2], al fine di coadiuvare al meglio il lavoro di codifica svolto dagli studenti, sono stati messi a disposizione materiali di tipo archivistico e bibliografico: un modello di codifica per ogni lettera, i file XML-TEI con le liste quasi complete delle persone, luoghi, opere, termini, organizzazioni e

¹ BDC è accessibile a <http://bellinicornespondence.cnr.it> (ospitato già sui server CNR all'URL <http://licodemo.ilc.cnr.it/bellini-in-rete>).

bibliografia, la trascrizione dei testi e le immagini. Un dettagliato elenco di elementi, attributi e relativi valori sono stati sistematizzati in documenti di lavoro e pubblicati sul repository *GitHub* di progetto, così da rendere il più efficiente possibile le attività del gruppo di ricerca.

Il processo di annotazione ha coinvolto diverse decine di studenti che hanno realizzato la codifica e la visualizzazione di singole lettere in modo autonomo con risultati molto diversi tra loro. Infatti, gli studenti, nel corso del proprio lavoro, hanno realizzato proposte di edizione digitale selezionando, studiando e codificando parte delle lettere belliniane.

Alla prima fase di selezione dei materiali e alla successiva codifica delle lettere, si sono avvicinate varie altre fasi che hanno previsto la realizzazione di nuove risorse per la: 1) validazione formale e scientifica dei documenti codificati; 2) armonizzazione delle diverse codifiche; 3) creazione del corpus belliniano; 4) presentazione e fruizione dell'edizione.

L'attività di verifica scientifica e di armonizzazione delle lettere ha previsto la revisione e/o la registrazione di fenomeni testuali nel documento digitale tra cui:

- L'inserimento e la revisione della codifica dell'intestazione di ogni lettera, della punteggiatura, delle note, della bibliografia, dell'interruzione di linea, delle cancellature e delle annotazioni di altre mani.
- Lo scioglimento delle abbreviazioni.
- La codifica dei collegamenti e dei riferimenti alle entità esterne.
- Il completamento delle liste – sia di entità nominate sia di altri dati notevoli – quali persone, luoghi, organizzazioni, termini, opere, bibliografia.
- La codifica e la revisione delle regioni d'interesse a partire dalle fonti facsimilari.
- Il completamento e la revisione delle descrizioni del supporto scrittoria con eventuale codifica di timbri, segni, danni, strappi, ed altri fenomeni materiali.
- La codifica accurata dei dati relativi alla corrispondenza, la componente degli indirizzi e i dati riguardanti l'imbustamento.

Al fine di supportare al meglio le attività elencate, è stato sviluppato anche uno strumento di valenza più generale rispetto al progetto specifico, quale il tool di verifica e correzione delle aree disegnate sulle immagini ZoneRW.

Il repertorio prodotto consta di 40 documenti XML-TEI (quante sono le unità testuali) più 6 documenti XML-TEI a corredo, contenenti le liste delle entità e dei dati rilevanti. In ragione del fatto che il software di visualizzazione scelto (EVT) necessita di alcune strutture specifiche definite per un singolo file, si è reso necessario realizzare l'intero corpus mediante un meccanismo di inclusione dinamica dei singoli documenti, basato sulla tecnologia XSLT, che garantisce al tempo stesso modularità, manutenibilità e flessibilità. In più, EVT richiede una struttura del documento TEI differente dal modello di codifica adottato per le lettere belliniane, quali, ad esempio, le informazioni bibliografiche, la gestione dei metadati, le descrizioni codicologiche, la gestione dei dati facsimilari, le note, la terminologia. Perciò sono state implementate le opportune trasformazioni strutturali dei documenti XML-TEI del repertorio per la pubblicazione web dell'edizione scientifica digitale.

3. GESTIONE DELLE IMMAGINI

La fruizione web dell'edizione è stata notevolmente migliorata da una fase di ottimizzazione della gestione delle immagini. In particolare, le attività hanno previsto sia lo sviluppo del tool ZoneRW (Zone Read Write)² per la verifica delle coordinate delle regioni d'interesse, sia la realizzazione delle versioni in formato piramidale delle immagini originali e sia la creazione delle miniature per ciascuna singola lettera.

Per quanto riguarda l'individuazione delle zone da segnalare per la corrispondenza tra immagine e relativo contenuto testuale del file TEI, siano esse linee di testo oppure hotspot, è ampiamente utilizzato lo strumento TEI Zoner³. Sfortunatamente, in caso di errori anche minimi, è necessario ripetere tutto il processo, lavoro insostenibile per il progetto BDC poiché gli elementi *zone* superano le 1200 unità. Inoltre, il tool non fornisce la visualizzazione a posteriori delle aree tracciate a partire dalle coordinate. Di conseguenza, per accelerare sia la visualizzazione che la correzione delle zone è stata sviluppata l'applicazione ZoneRW.

ZoneRW (Fig. 1a) è un tool WYSIWYG che, a partire dalla dichiarazione delle *surface* con i relativi elementi *graphic* e *zone*, mostra le regioni codificate nel file TEI con i relativi attributi *@xml:id* e *@rendition*, consentendone la contestuale modifica (aggiornamento delle coordinate, aggiunta o rimozione di zone) e generando il corrispondente codice

² Il tool è liberamente disponibile su *GitHub* al seguente indirizzo: <https://github.com/pierpaolosichera/ZoneRW>. ZoneRW è stato sviluppato utilizzando 4th Dimension, comunemente chiamato 4D, (<https://www.4d.com/>) ambiente RAD per database relazionali prodotto e mantenuto dalla società francese 4D SAS.

³ <http://teicat.huma-num.fr/zoner.php>.

XML-TEI. Il risultato presente a video può essere esportato anche in formato immagine (ad esempio JPG) ottenendo così una vista delle aree che sia indipendente dal tool di creazione. In questo modo, i dati associati alle differenti zone delle immagini vengono correttamente estratti dal software EVT⁴ al fine di evidenziare le regioni d'interesse sovrapposte alle immagini facsimilari. Infatti, il *viewer* fornisce le funzionalità per la vista parallela testo-immagine a granularità di riga (attivabile tramite il pulsante con l'icona *link*) nonché la visualizzazione degli *hotspot* mediante il relativo pulsante posto nella barra degli strumenti al di sopra dell'immagine (Fig. 1b).

Da un punto di vista tecnico, l'attuale versione 2beta di EVT è sviluppata con il *framework* javascript AngularJS⁵, assicurando al *viewer* un approccio client-side. D'altro canto le immagini ad alta risoluzione sono acquisite attraverso apparecchiature specifiche, generando file di notevoli dimensioni⁶. Tuttavia, la visualizzazione di immagini in formati web standard come JPG o PNG richiede che il file venga interamente caricato dal browser, provocando lunghi tempi di attesa per immagini di grandi dimensioni.

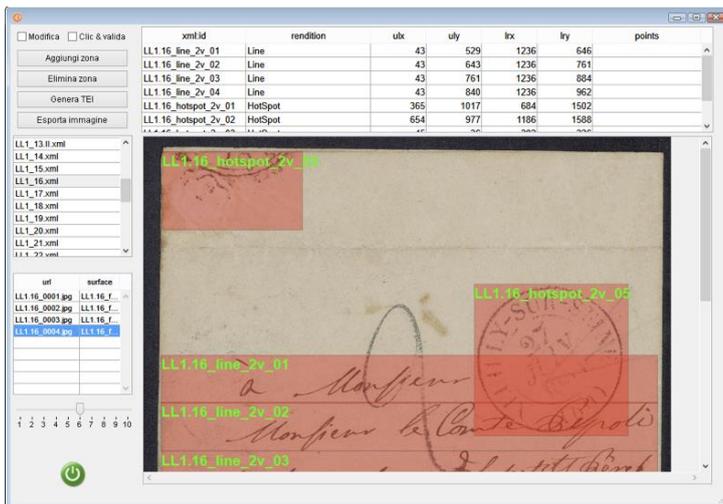


Figura 14a. Schermata principale di ZoneRW per la visualizzazione delle regioni di interesse codificate mediante l'elemento zone della TEI.

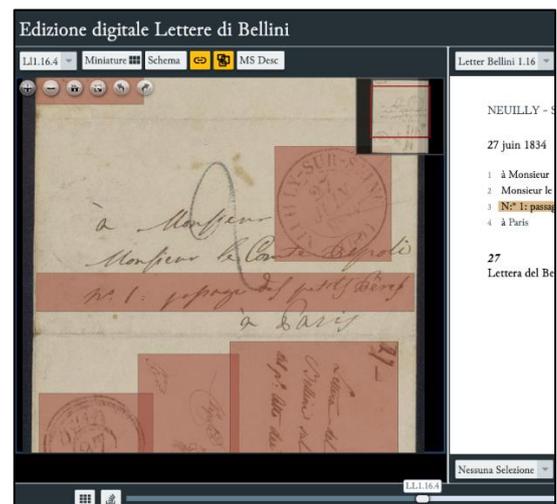


Figura 15b. Schermata di EVT viewer con evidenziate le zone di interesse (linee e hotspot).

EVT integra il *viewer* OpenSeadragon⁷, che, oltre ai formati standard, consente di visualizzare immagini in formato piramidale, quale DZI (Deep Zoom Image), che suddivide l'immagine originale in *tiles*, porzioni cioè della stessa immagine che vengono caricate in funzione dell'attuale *viewport*⁸. Utilizzando questa caratteristica, le immagini originali ad alta risoluzione sono state trasformate in DZI, ottenendo così un cospicuo incremento nelle prestazioni senza rinunciare alla qualità delle scansioni. La conversione delle immagini JPG in *tiles* DZI è stata eseguita attraverso uno script⁹ che utilizza il tool a riga di comando PyramidIO¹⁰. Inoltre, sfruttando la caratteristica di EVT che consente di avere le immagini ad alta risoluzione e le miniature in cartelle diverse, si è scelto di utilizzare immagini di ridotte dimensioni e a bassa risoluzione per le anteprime delle oltre 100 carte¹¹.

⁴ Disponibile su <http://evt.labcd.unipi.it/>.

⁵ <https://angularjs.org/>.

⁶ Nel caso belliniano i file delle immagini in formato JPG hanno una dimensione di circa 4MB.

⁷ <https://openseadragon.github.io/>.

⁸ Con il termine *viewport* ci si riferisce all'area effettiva di visualizzazione dell'immagine.

⁹ Il comando da utilizzare per la conversione delle immagini in formato piramidale è: `java -jar pyramidio-cli-1.1.0.jar -i ./nomefileinput.jpg -o ./nomecartelladestinazione`. È importante notare che `nomefileinput` non deve contenere il carattere "." se non prima dell'estensione.

¹⁰ <https://github.com/usnistgov/pyramidio>.

¹¹ Le miniature sono state create col software ImageMagick (<https://github.com/ImageMagick/ImageMagick>) coll'istruzione da riga di comando: `convert input.jpg -format jpg -define jpeg:size=192x292 -auto-orient -thumbnail 96x146 -units pixelsperinch -density 72x72 output.jpg`. Il parametro `size` impone un downsampling all'immagine prima che ne venga generata la thumbnail; `density` porta la densità dei pixel dall'alta risoluzione a 72 dpi.

4. ADATTAMENTO DEL MODELLO DI CODIFICA

Il repertorio codificato secondo il modello di progetto ha implicato un'ulteriore fase di adattamento al fine di ottimizzarne l'esperienza utente in ambiente web¹². Nello specifico, sono stati ristrutturati alcuni frammenti XML-TEI riguardanti:

- 1) la registrazione delle coordinate;
- 2) la disposizione dei metadati codicologici e la corrispondenza nonché la struttura dei metadati del corpus nella sua interezza;
- 3) le liste di entità, di dati notevoli e le note.

Data l'arbitrarietà delle dimensioni delle carte e delle immagini nel corpus belliniano, si è resa necessaria un'attività di normalizzazione delle coordinate registrate nelle zone di interesse. In tal modo è stato possibile ottenere una corretta funzione di allineamento testo-immagine con OpenSeadragon in EVT. Tale normalizzazione è stata eseguita mediante lo sviluppo di un foglio di stile XSLT¹³.

I metadati codicologici, della corrispondenza e del corpus codificati in ciascun documento XML sono stati opportunamente elaborati per renderli visibili dal software EVT aggiungendo specifiche regole al foglio di stile XSLT.

Grazie a queste trasformazioni, in visualizzazione (come si vede in Fig.2), la sezione accessibile tramite il pulsante *Info* riporta, per ogni lettera, le informazioni riguardanti il titolo, la lingua, la collocazione, il tipo di supporto usato, le sue condizioni fisiche, la presenza di timbri, di filigrana, di sigilli, di piegature, di annotazioni di altre mani e le informazioni sulla corrispondenza (mittente, destinatario, luogo di invio, eventuale luogo di ricezione e datazione).

Mentre, nella sezione relativa alla visualizzazione dei dati del manoscritto (pulsante *Ms Desc* dell'interfaccia web) possiamo notare le informazioni riguardanti il titolo, la bibliografia di riferimento, l'incipit e l'explicit di ogni lettera.

In dettaglio, per ottenere la ridisposizione degli elementi, si applicano alcune regole di trasformazione (riportate in Fig.3) che ricodificano gli opportuni dati dell'intestazione (teiHeader) da ogni singolo documento di input al contenuto del blocco text/front del documento di output, come si vede in Fig.4.

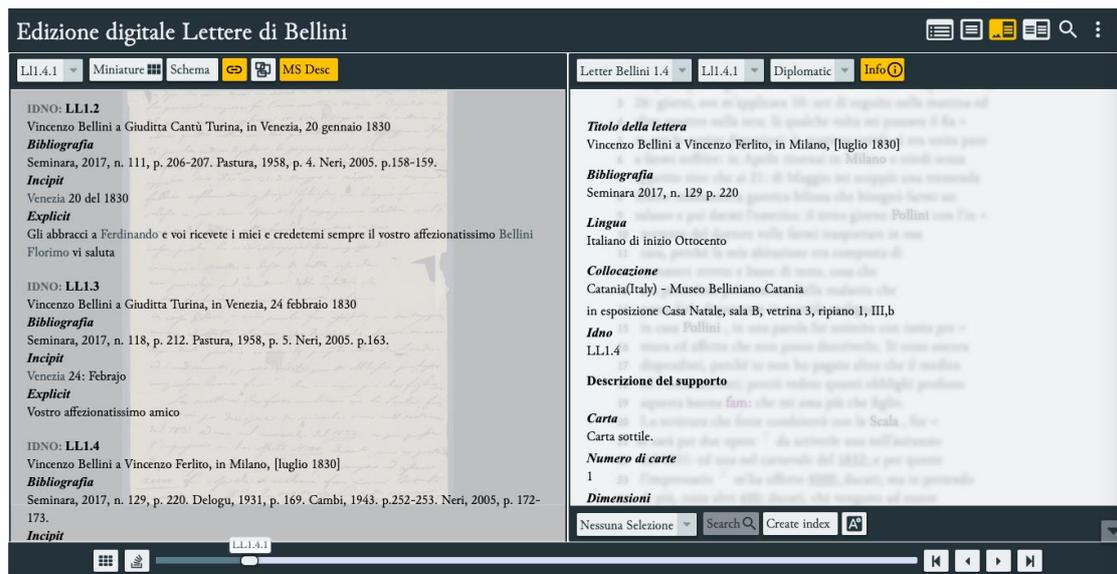


Figura 16. Risultato finale su EVT delle sezioni Info e MsDesc.

¹² La fase di adattamento è stata realizzata mediante specifiche regole definite in fogli di stile XSLT. Le trasformazioni fanno uso del processore saxonb-xslt, che naviga il documento di input tramite XPath ed esegue le regole definite nel file XSLT, generando così il documento XML-TEI di output compatibile con lo strumento di visualizzazione web EVT.

¹³ La normalizzazione è calcolata in funzione del rapporto tra la dimensione di ogni singola immagine e un coefficiente costante, successivamente il valore di ogni coordinata si divide per il rapporto precedentemente calcolato.

```

<xsl:template match="tei:front">
  <xsl:element name="front" namespace="http://www.tei-c.org/ns/1.0">
    <xsl:element name="div" namespace="http://www.tei-c.org/ns/1.0">
      <xsl:attribute name="type">Title</xsl:attribute>
      <xsl:element name="fw" namespace="http://www.tei-c.org/ns/1.0">
        Titolo della lettera
      </xsl:element>
      <xsl:element name="p" namespace="http://www.tei-c.org/ns/1.0">
        <xsl:copy-of select="//tei:titleStmt/tei:title/node()"/>
      </xsl:element>
    </xsl:element>
  </xsl:template>

```

Figura 17. Regole di trasformazione XSLT per la sezione 'info'.

Le informazioni relative alla descrizione del manoscritto sono state estratte a partire dall'elemento `msDesc`. In particolare, sfruttando il content model dell'elemento `msContent`, è stato possibile ricollocare la descrizione di ogni unità testuale nel documento di output con l'elemento `msItem`. All'interno di quest'ultimo sono registrate le informazioni che descrivono il contenuto del manoscritto, quali il titolo della lettera, la bibliografia e a seguire l'incipit e l'explicit di ciascuna lettera. L'edizione è ricca di entità nominate e fenomeni notevoli che vengono presentate in EVT in varie modalità grafiche (liste, sezioni in overlay, pop-up, etc). Ad esempio, per accogliere pienamente le informazioni relative alle note, alle opere citate e alla terminologia, ulteriori regole XSLT favoriscono la visualizzazione di aree in modalità pop-up.

In particolare, per ogni opera citata nel testo cliccando accanto ad essa si apre un pop-up nel quale possiamo visualizzare le informazioni riguardanti il titolo, il compositore, il librettista, la prima rappresentazione e le note che spiegano e danno approfondimenti sull'opera stessa.

```

<text type="letter" n="Bellini 1.4" xml:id="LL1.4">
  <front>
    <div type="Title"><fw>Titolo della lettera</fw><p>Vincenzo Bellini a Vincenzo Ferlito, in Milano, [luglio 1830]</p></div>
    <div type="Bibliography"><fw>Bibliografia</fw><p>Seminarà 2017, n. 129 p. 220</p></div>
    <div type="Language"><fw>Lingua</fw><p>Italiano di inizio Ottocento</p></div>
    <div type="Collocation">
      <p><fw>Collocazione</fw>Catania(Italy) - Museo Belliniano Catania</p>
      <p><fw>in esposizione Casa Natale, sala B, vetrina 3, ripiano 1, III,b</p>
      <p><fw>Idno</fw>LL1.4</p>
    </div>
    <div type="desc">
      <head>Descrizione del supporto</head>
      <p><fw>Carta</fw>Carta sottile.</p>
      <p><fw>Numero di carte</fw>1</p>
      <p><fw>Dimensioni</fw>269x195<unit> mm</unit></p>
      <p><fw>Watermark</fw>La carta non presenta filigrana</p>
      <p><fw>Timbri</fw>Sono presenti nel folio retro centrale un timbro postale di partenza e un frammento di sigillo di cerallacca rossa.</p>
      <p>Il folio presenta segni di piegature asimmetriche che lo suddividono in nove sezioni. Così il folio stesso funge anche da busta con l'indirizzo n
    </div>

```

Figura 18. Esempio di output della lettera LL1.4 dopo aver applicato la trasformazione per la sezione 'info'.

Infine, sono state condotte anche varie modifiche al codice del software EVT che hanno permesso la personalizzazione della resa grafica del pop-up soprattutto per le note di approfondimento terminologico e bibliografico¹⁴ (Fig. 5). Queste sono molto corpose tanto da dover cambiare dinamicamente la dimensione dei pop-up in modo che il browser attivi, in automatico ove necessario, le barre di scorrimento laterali.

¹⁴ La modifica ha riguardato tre file di EVT: `index.html`, `evtviewer.6cd6ca9b.js` e `main.ffb09194.css`. È stato inoltre aggiunto un file CSS per le regole di stile personalizzate.

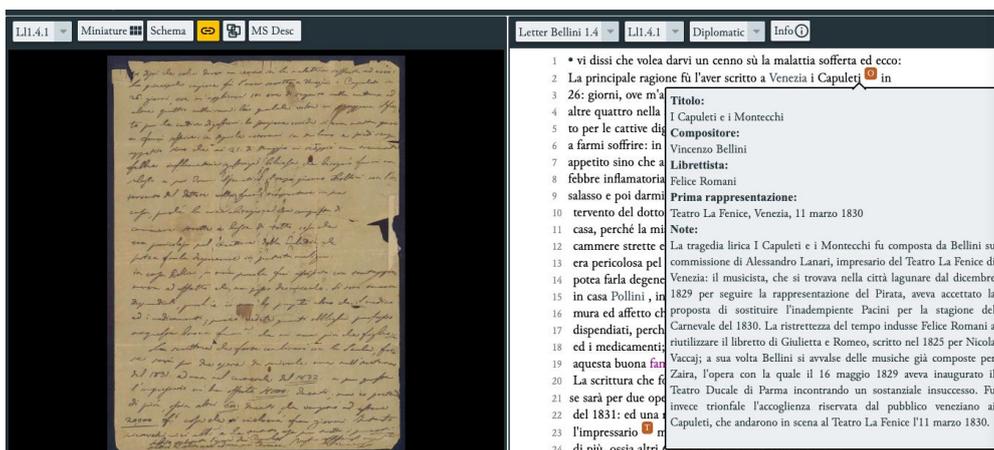


Figura 19. Risultato finale su EVT per la visualizzazione dei pop-up delle opere citate.

5. CONCLUSIONI E SVILUPPI FUTURI

Il contributo ha presentato lo stato attuale dell'edizione digitale delle lettere belliniane denominata Bellini Digital Correspondence. In seno al progetto sono state realizzate numerose attività a partire dalla codifica effettuata da studenti di informatica umanistica, fino alla personalizzazione del software di visualizzazione EVT, passando per l'armonizzazione di ogni singolo documento tramite la definizione di regole XSLT. Tra i prodotti della ricerca è stato anche sviluppato uno strumento per la verifica, la correzione e la gestione delle regioni di interesse (il tool ZoneRW). Le funzionalità sviluppate possono essere la base di una estensione futura di EVT. Intenzione del gruppo di lavoro è quello di migliorare ulteriormente i risultati del progetto, quali l'aggiunta sistematica di entrate del Lessico della letteratura musicale italiana (LESMU) da apportare alla descrizione terminologica. In più, il tool ZoneRW permetterà anche la modifica delle coordinate tramite puntatori grafici.

Tra le attività in corso menzioniamo la realizzazione di alcune pagine di presentazione all'interno del sito web di progetto in cui saranno pubblicate le scelte editoriali, di codifica, i progetti più validi di codifica realizzati dagli studenti, alcune pagine di presentazione di dati estratti dal corpus codificato, quali indici di luoghi, destinatari, mittenti, cronologici. La funzione di ricerca nei testi, prevista nella nuova versione ufficiale di EVT, sarà integrata nell'edizione digitale delle lettere di Bellini. Infine, è stata avviata una fase interlocutoria per la creazione di una scheda di catalogo e per il deposito del repertorio all'interno dell'infrastruttura CLARIN-IT.

6. RINGRAZIAMENTI

Il progetto BDC è stato in parte finanziato dal Patto per Catania a valere sul Fondo Sviluppo e Coesione 2014-2020. Gli autori, inoltre, ringraziano Carla Congiu e Laura Mazzagufò per il lavoro svolto nell'ambito del progetto.

BIBLIOGRAFIA

- [1] Bellini, Vincenzo. 2017. *Carteggi*. Graziella Seminara. Vol. 131. *Historiae Musicae Cultores*. Firenze: Olschki.
- [2] Burnard, Lou. 2014. *What Is the Text Encoding Initiative?: How to Add Intelligent Markup to Digital Resources*. Marseille: OpenEdition Press. <http://books.openedition.org/oep/426>.
- [3] Del Grosso, Angelo Mario, Daria Spampinato, Erica Capizzi, Salvatore Cristofaro, e Graziella Seminara. 2019. «Promoting Bellini's legacy and the Italian opera by scholarly digital editing his own correspondence». In *What is Text, really? TEI and beyond*. Graz, Austria.
- [4] Del Grosso, Mario Angelo, Erica Capizzi, Salvatore Cristofaro, Maria Rosa De Luca, Emiliano Giovannetti, Simone Marchi, Graziella Seminara, e Daria Spampinato. 2019. «Bellini's Correspondence: a Digital Scholarly Edition for a Multimedia Museum». *Umanistica Digital* 3 7: 23–47.
- [5] Di Pietro, Chiara, e Roberto Rosselli Del Turco. 2018. «Between Innovation and Conservation: The Narrow Path of UI Design for the DSE». In *Digital Scholarly Editions as Interfaces*, a cura di Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, e Gerlinde Schneider. Vol. 12. *Schriften Des Instituts Für Dokumentologie Und Editorik*. Norderstedt: Books on Demand.
- [6] Pierazzo, Elena. 2015. *Digital Scholarly Editing: Theories, Models and Methods*. Digital Research in the Arts and Humanities. Farnham Surrey: Ashgate.
- [7] Salvatori, Enrica, Federica Boschetti, e Angelo Mario Del Grosso. 2019. «From Collaborative Transcription to Interdisciplinary Education: The Postcards of the Great War Case». In *Teaching and research in Digital Humanities' era*. Udine, Italy.

Dante e Petrarca allo (stesso) scrittoio. Per lo sviluppo di un'ontologia di IDP a partire dall'istanza manoscritti di Itinera

Fara Autiero¹, Gennaro Ferrante², Sandra Gorla³, Serena Picarelli⁴

¹Università degli Studi di Napoli "Federico II", Italia – fara.autiero@unina.it

²Università degli Studi di Napoli "Federico II", Italia – gennaro.ferrante@unina.it

³Università degli Studi di Napoli "Federico II", Italia – sandra.gorla@unina.it

⁴Scuola Superiore Meridionale, Italia – serena.picarelli@unina.it

ABSTRACT

Illuminated Dante Project, iniziativa dipartimentale dell'Università degli Studi di Napoli "Federico II", si occupa di censire e catalogare i manoscritti miniati tre-quattrocenteschi della *Commedia* di Dante le cui immagini recino anche il più piccolo riferimento significativo al testo del poema. Negli ultimi anni IDP si è posto l'obiettivo di sviluppare un'ontologia per gestire i diversi formati di *database* utilizzati nei suoi sottoinsiemi locali. In questo modo, la grande mole di dati inserita in IDP permetterebbe agli utenti di poter effettuare ricerche per artista, data, luogo, autore o titolo dell'opera e perfezionare la ricerca tramite una serie di filtri specificamente pensati per la particolare tipologia del *database*. Il confronto con i membri del progetto PRIN Itinera (*Italian Trecento Intellectual Network and European Renaissance Advent*) ha portato a valutare il *framework* Cadmus – e in particolare l'*item Manuscript* progettato da Itinera – come promettente possibilità per il passaggio del *database* relazionale di IDP al web semantico.

PAROLE CHIAVE

Database, iconografia, semantic web, catalogazione.

INTERVENTO

Nel 2015, in seguito al programma STAR promosso dalla Compagnia San Paolo, è stato ottenuto un finanziamento per il lancio di un progetto start-up denominato *Illuminated Dante Project* (IDP). L'obiettivo del progetto era quello di fornire un'indagine sistematica e una descrizione accurata dei primi manoscritti miniati della *Commedia* di Dante recanti anche il più piccolo riferimento significativo al testo del poema ([2];[4]). Una volta definito il *corpus* il team di IDP ha portato avanti tre azioni principali: 1. Digitalizzazione dei manoscritti; 2. Strutturazione delle descrizioni codicologiche e iconografiche; 3. Configurazione di un archivio di immagini interoperabile.

In merito al *primo punto*, grazie ad un accordo tra IDP e il Ministero per i beni e le attività culturali e per il turismo, IDP ha ottenuto il *copyright* dei codici conservato presso le biblioteche statali e ha stretto collaborazioni con biblioteche non statali e internazionali arrivando a costruire il più grande archivio di immagini dantesche attualmente esistente. Insieme alle digitalizzazioni, è stata avviata un'indagine sperimentale di "diagnostica" utilizzando la tecnologia RTI (*Reflectance Transformation Imaging*). A questo scopo, IDP ha iniziato una collaborazione con il gruppo *Illuminare* della KU di Leuven che ha analizzato alcuni codici danteschi utilizzando il Microdome© (il database sarà presto implementato con il viewer Pixel+ per la visualizzazione dei file zun).

Per quanto riguarda il *secondo punto*, IDP ha stipulato un accordo con l'Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane (ICCU) al fine di condividere le descrizioni codicologiche dei codici del *corpus* con il Catalogo Nazionale dei Manoscritti online (MOL) tramite lo standard di markup XML TEI-P5. A sua volta, IDP ha strutturato un *database* MySQL per l'iconografia dantesca legata al supporto manoscritto ospitato su www.dante.unina.it; in tale *database* la sezione 19 del modulo di descrizione di MOL (corrispondente al modulo <decoDesc> TEI-P5) è stata profondamente rielaborata grazie all'introduzione di categorie indicizzabili come "Attribuzione stilistica", "Tipologia decorativa" e "Iconografia". Quest'ultima sezione è stata suddivisa in due parti: a. *Immagine/Testo* per la classificazione delle relazioni di ogni immagine con il testo del poema, il paratesto (rubriche, glosse e commenti), il contesto interno (topografia della *Commedia*) e il contesto esterno (tradizione iconografica dantesca e non) e per determinare il livello diegetico dell'illustrazione (narrazione di primo livello, ad esempio "Dante incontra Virgilio", o narrazione di secondo livello, es. "Il Conte Ugolino nella torre"); b. *Tematizzazione*, articolata in *Soggetto*, ovvero l'esecuzione dell'immagine come una sorta di "figurativa" *parole* Saussuriana ("Dante spaventato dalle fiere"); *Macrosoggetto*, ovvero la sovrastruttura dell'immagine considerata secondo l'elaborazione Saussuriana della *langue* ("Uscita dall'inferno") e *Parole chiave*.

Infine, in merito al *terzo punto*, per rendere l'archivio conforme agli attuali standard di interoperabilità delle immagini sul web, IDP ha iniziato una collaborazione con i laboratori di ingegneria informatica dell'Università "Federico II" di Napoli (CSI) al fine di testare una configurazione del web server conforme agli standard IIIF. I file immagini multi-pagina (JPEG2000) sono attualmente disponibili sul sito del progetto attraverso il viewer Mirador 3; ogni manoscritto è stato corredato del proprio *manifest* che ne permette l'importazione e la fruibilità in tutti i *repositories* che supportano la IIIF.

Nel corso degli anni, IDP si è posto l'obiettivo di creare un'ontologia basata sul web semantico per gestire i diversi formati di *database* utilizzati nei suoi sottoinsiemi locali. In questo modo, la grande mole di dati inserita in IDP permetterebbe agli utenti di poter effettuare ricerche per artista, data, luogo, autore o titolo dell'opera e perfezionare la ricerca tramite una serie di filtri specificamente pensati per la particolare tipologia del *database*. Il fruttuoso confronto con i membri del progetto Itinera (*Italian Trecento Intellectual Network and European Renaissance Advent*) ha portato a valutare il *framework* Cadmus come possibilità molto promettente per la trasformazione del *database* relazionale di IDP in un'ontologia.

Il progetto PRIN Itinera (diretto da Natascia Tonelli e che coinvolge le Università di Siena, Roma Tre, Stranieri di Perugia e "Federico II" di Napoli) è dedicato al censimento e allo studio delle corrispondenze di Petrarca e indaga al contempo la tradizione manoscritta dei *Rerum vulgarium fragmenta* ([1]). Uno degli obiettivi principali del progetto è la costruzione di una banca dati digitale consultabile online in *open access* che raccolga e renda disponibili i risultati delle ricerche delle varie unità. La creazione del *database* è stata affidata a Daniele Fusi (Università Ca' Foscari) ed è basata sul *framework* di *editing open source* Cadmus da lui sviluppato nell'ambito di una *research fellowship* presso il *Venice Centre for Digital and Public Humanities* (<https://github.com/vedph>). Cadmus permette di costruire sistemi per la creazione *web-based* di contenuti e consente di organizzare i dati in set indipendenti e componibili in strutture più complesse, dinamiche e modulari: le unità di grado superiore dell'architettura, gli *item*, sono infatti costituite da più *part*, insiemi di dati coerenti e autonomi, a loro volta assemblate attraverso la combinazione di diverse *properties* e *classes*. Gli *item* così modellati sono in grado di rappresentare diverse tipologie di oggetti, e permettono una strutturazione dei dati molto specifica. Nella creazione del *database* del progetto petrarchesco sono stati elaborati modelli con un alto livello di granularità descrittiva, ma che aspirano al contempo a risultare ampiamente generalizzabili. Tra gli *items* di Itinera, quello dedicato alla descrizione dei manoscritti si è rivelato un interessante spunto di riflessione per la ristrutturazione del *database* di IDP. L'*item Manuscript* è stato elaborato con l'obiettivo di conciliare il rispetto della prassi della descrizione codicologica tradizionale con una modellizzazione il più possibile parcellizzata dei dati, tentando in parallelo di produrre una struttura descrittiva potenzialmente funzionale per qualsiasi progetto che abbia tra i suoi scopi la descrizione – e lo studio – di manoscritti.

Ragionando sui limiti dell'XML-TEI e sulle potenzialità offerte da Cadmus (cfr. anche [3]), la ristrutturazione di IDP risulterà estremamente utile per alcune sezioni particolari. Il modulo "Scrittura e mani", attualmente dipendente dai metadati inseriti in MOL, permetterà di individuare ogni copista con uno specifico *ID* collegato al Virtual International Authority File (VIAF <http://www.viaf.org>), consentendo di mettere in comunicazione non solo le descrizioni dei codici copiati da uno stesso scriba, ma anche le immagini di tali manoscritti presenti in IDP o in altri *repositories*, con un'attenzione particolare al portale Biblissima (<https://iiif.biblissima.fr/collections/>) che raccoglie tutti i *repositories* che utilizzano la IIIF.

Una prima idea della riformulazione di questa sezione attraverso l'utilizzo di Cadmus è illustrata nella seguente struttura schematica:



La *part* che nell'*item Manuscript* di IDP sarà dedicata all'immissione dei dati relativi agli atti scrittori di ogni singolo copista su un dato manoscritto sarà costituita da sei campi principali:

-Id: un identificativo univoco *human-readable* (con relativo *authority file* potenzialmente comunicante con VIAF) che, sfruttando le potenzialità del web semantico, sia collegato ad un *item Copyst* (che raccolga le informazioni riguardanti il copista) all'interno del quale si darà conto della patina linguistica dello scriba e che sarà a sua volta associato attraverso

triple ad altri eventuali codici da lui copiati (sia interni, e quindi corrispondenti a un *item Manuscript* del *database*, sia esterni, con rimando ad altri *repositories*);

- Tipologia grafica: definita attraverso un *thesaurus* precedentemente stabilito;
- Inchiostro: i cui colori sono indicati attraverso un *thesaurus* precedentemente stabilito;
- Testo copiato: un *thesaurus* dedicato alle possibili tipologie e porzioni di testi copiati (ad es. *Commedia*, singole cantiche o parti di esse, commenti, chiose irrelate) con collegamento al *range* di carte;
- Progetto originario: un *flag booleano* che indichi se la copia sia contestuale al confezionamento del manoscritto (caso marcato);
- Datazione: attraverso l'utilizzo del modello *historical date* già presente nel *framework* Cadmus (cfr. [3]: 264-266).

Il risultato della riformulazione di questa parte punta alla creazione del più grande repertorio comparatistico di copisti danteschi.

Di particolare interesse sarà la sezione dedicata alle “Descrizioni interne”. Attualmente i dati inseriti in questa parte sono frantumati tra MOL (*incipit* ed *explicit*, titoli delle partizioni testuali) e IDP (intitolazioni). La rielaborazione di questo modulo permetterà di collegare l'attività dei copisti alle diverse porzioni testuali e prevederà un campo interamente dedicato alle lacune e alle interpolazioni. Questi ultimi due dati, importantissimi per la tradizione della *Commedia*, non disponendo di un *tag* specifico, sono oggi slegati da qualsiasi possibilità di ricerca e non sono restituiti all'utente finale.

Parimenti, la sezione “Decorazione” beneficerà del passaggio al web semantico. Questo varrà in particolare per le due sottosezioni “Iconografia” ed “Evidenze materiali”. Per quanto riguarda la prima, la riformulazione di IDP consentirà di comparare i risultati dell'analisi del singolo manoscritto con altri *items* seguendo diversi criteri (ad es. operato di una stessa bottega, operato di diverse botteghe rispetto agli stessi passaggi testuali), permettendo in ultima battuta di ottenere una panoramica esaustiva della nascita e dello sviluppo di formule iconografiche legate alla prima tradizione miniata del poema. Inoltre, sarà possibile confrontare i soggetti iconografici di IDP con quelli indicizzati nei più importanti *database* iconografici, come ICONCLASS (www.iconclass.org/help/outline) e Index of Christian Art (<https://library.princeton.edu/resource/title/index-christian-art>).

La sezione “Evidenze materiali”, vera e propria novità di IDP, è stata strutturata tenendo conto delle “Istruzioni verbali” (descrizioni della scena da figurare) e delle “Istruzioni abbreviate” (rimandi a istruzioni verbali o a modelli visivi). Anche in questo caso, la granularità del dato permetterà di scomporre l'informazione in più punti che tra loro potranno essere messi in dialogo. Sarà possibile confrontare le istruzioni al miniatore, da un lato, di codici della *Commedia* presenti in IDP arrivando a delineare un quadro complessivo della progettazione dei corredi iconografici dei primi manoscritti illustrati del poema e, dall'altro, di altri manoscritti di diverse tradizioni figurative descritti in *database* come Iniziale (<http://initiale.irht.cnrs.fr/>). Inoltre, tale sezione dell'*item* potrà dialogare con quella deputata all'analisi del rapporto testo-immagine, permettendo di indagare anche la relazione che le istruzioni per il miniatore instaurano con il testo, i paratesti (attraverso una prospettiva collaborativa automatica, sarà auspicabile un'interazione con CoDa, *database* in corso di costruzione da parte dell'OVI che raccoglierà tutti i commenti danteschi) e le illustrazioni del manufatto.

BIBLIOGRAFIA

- [1] De Nisco, Nicola, Sandra Gorla, e Alessia Valenti. 2021. «Una banca dati per Petrarca e il suo tempo: criteri, modelli e obiettivi». *DigItalia 2*.
- [2] Ferrante, Gennaro. 2019. «Illuminated Dante Project. Per un archivio digitale delle più antiche illustrazioni della ‘Commedia’. I. Un case study quattrocentesco (mss. Italien 74, Riccardiano 1004 e Guarneriano 200)». In *Dante visualizzato. Carte ridenti II: XV secolo*, M. Ciccuto e L. M. G. Livraghi, 229–55. Firenze: Franco Cesati Editore.
- [3] Fusi, Daniele. 2018. «Sailing for a Second Navigation: Paradigms in Producing Digital Content». », *Seminari romani di cultura greca 7*: 213–76.
- [4] Perna, Ciro. 2020. «IDP - Illuminated Dante Project: un archivio e database per la più antica iconografia dantesca (secc. XIV-XV)». *DigItalia 2*: 150–58.
- [5] «VeDPH». s.d. Github. Consultato 11 novembre 2021. <https://github.com/vedph>.

Il progetto ‘epistolarITA’ e una proposta di applicazione di algoritmi di prossimità testuale su documenti epistolari italiani (XV-XVII s.)

Gianluca Valenti¹, Alexander Zeisberg Militerni²

¹Université de Liège, Belgio - gianluca.valenti@uliege.be

²Université de Liège, Belgio - alexander.zeisberg85@gmail.com

ABSTRACT

Negli ultimi decenni si è assistito a un incremento significativo di ricerche digitali su testi epistolari. Se molte di esse si concentrano su edizioni di lettere e su schedature di metadati, meno esteso è il numero di progetti al servizio dell’analisi – sia essa storica, linguistica o letteraria. Un algoritmo sviluppato di recente all’Università di Liegi permette di inserire un qualsiasi testo epistolare all’interno di un database collegato con le principali edizioni – digitali, ma non solo – di lettere italiane scritte tra il XV e il XVII secolo, per ottenere in ritorno le lettere ad esso semanticamente più prossime, facendo emergere connessioni inaspettate all’interno di quella *respublica litteraria* composta da scriventi di testi epistolari della prima modernità.

PAROLE CHIAVE

TF-IDF, Named-Entity Recognition, sistema di raccomandazione, prossimità testuale, documenti epistolari.

INTERVENTO

1. TESTI EPISTOLARI E PROGETTI DIGITALI

Negli ultimi decenni si è assistito a un incremento significativo di progetti incentrati sul reperimento, sull’edizione e sull’analisi di testi epistolari della ‘prima modernità’ (sintagma volutamente fluido con cui intendiamo, *grosso modo*, i secoli XV–XVII). Grazie in particolare alla ISCH COST ACTION IS1310 «Reassembling the Republic of Letters» e al progetto «Cultures of Knowledge», entrambi portati avanti, all’università di Oxford, sotto la supervisione di Howard Hotson, si è recentemente consolidato un fecondo dialogo tra gruppi di ricerca europei che lavorano su *corpora* epistolari. Senza pretese di essere esaustivi qui si segnalano il database «EMLO – Early Modern Letters Online» e il progetto «Mapping the Republic of Letters». Potenzialmente di estrema importanza, benché al momento contenga un numero relativamente limitato di informazioni, è il motore di ricerca «correspSearch», che permette di effettuare ricerche tra i metadati di edizioni di lettere moderne. In tempi ancora più recenti, e sempre all’interno del summenzionato network europeo, hanno pure visto la luce, tra gli altri, i progetti «Skillnet – Sharing Knowledge in Learned and Literary Networks» e «LetterSampo».

Nel medesimo arco cronologico nella Penisola da un lato si registra un’attenzione senza precedenti indirizzata verso il preziosissimo patrimonio epistolare italiano – fin troppo a lungo sacrificato, dagli studiosi, a vantaggio di altri generi letterari – ma, in parallelo, bisogna riconoscere che l’impatto di tecnologie e metodi digitali applicati a questo specifico campo di ricerca è ancora lungi dall’aver esaurito tutte le sue potenzialità, soprattutto se paragonato ai progetti europei – tra cui quelli precedentemente citati – che hanno per oggetto la medesima tipologia testuale. Un forte slancio a iniziative di questo tipo è stato dato dal progetto «Archilet – Archivio delle corrispondenze letterarie italiane di età moderna (secoli XVI-XVII)» il quale, «nato timidamente» ([3]: 340a) nel biennio 2011-2012, ha visto progressivamente aumentare la propria notorietà ed è oggi un polo d’eccellenza, catalizzatore di iniziative scientifiche del massimo rilievo, tra cui si segnala perlomeno il PRIN 2016-2020 «Repertorio Epistolare del Cinquecento. Teorie, lingua, pratiche di un genere», coordinato da Paolo Procaccioli e articolato su sette unità di ricerca¹.

¹ Non avendo spazio sufficiente per fornire una sintesi ragionata sulla vasta bibliografia attualmente disponibile sulla materia epistolare in ambito italiano, si rimanda – per un primissimo approccio alla materia – alle pubblicazioni *Open Access* di Archilet (<http://www.archilet.it/Pubblicazioni.aspx>, ultima visualizzazione: 07.09.2021) e della collana BITeS (per la quale sono uscite, solo nell’ultimo anno, le edizioni di testi epistolari a cura di ([2];[4];[10]).

Altri progetti digitali, pur non ruotando interamente intorno all'oggetto-lettera, ad esso ritornano con insistenza, facendone di fatto il nucleo fondante dei rispettivi *corpora* di riferimento. Pensiamo in particolare al «Medici Archive Project» e al suo database «BIA», all'interno del quale si possono visualizzare i metadati e le parziali trascrizioni di circa 25.000 fonti mediche, di cui più della metà sono, appunto, lettere. Varie edizioni di documenti – in buona parte epistolari – legati alla famiglia Gonzaga sono state pubblicate sul sito delle «Banche dati Gonzaga», mentre è basato su testi relativi a Isabella d'Este il progetto «IDEA», che si compone anch'esso di migliaia di lettere, le cui trascrizioni dovrebbero essere divulgate online nei prossimi anni. Tra le numerose altre edizioni digitali relative a documenti epistolari composti tra Quattro- e Seicento segnaliamo l'edizione delle lettere contenute nell'«Archivio Datini», nonché quelle della corrispondenza di Michelangelo, Vasari e Filippo Baldinucci che sono state rese disponibili sul sito della «Fondazione Memofonte». Nato al di fuori dei confini nazionali ma direttamente collegato alla materia epistolare in lingua italiana è il progetto «EpistolART», promosso da Paola Moreno all'Università di Liegi, grazie al quale è stata diffusa online (affiancata da riproduzioni fotografiche) una nuova edizione della monumentale antologia di corrispondenza di artisti rinascimentali pubblicata originariamente da Gaye ([5]).

Tutti i progetti appena citati possono essere suddivisi in due macrocategorie, motori di ricerca ed edizioni digitali, strumenti cioè che mirano a fornire all'utente le informazioni primarie – i metadati e le trascrizioni – concernenti le fonti epistolari. Qualora però si passi a menzionare progetti direttamente legati allo studio di questi stessi testi, il panorama è meno ricco. Sono pochi infatti quei progetti in cui le metodologie digitali sono messe al servizio dell'*analisi* – sia essa storica, linguistica o letteraria – di testi epistolari, sebbene la loro sempre più impellente necessità sia ormai pienamente riconosciuta: «lo studio della lettera non è mai stato così vitale come nella stagione presente. Campagne di digitalizzazione sempre più estese e sofisticate hanno garantito una disponibilità inedita dei materiali epistolari. [...] Nessuno studioso del passato ha avuto a disposizione tanti materiali e tanto facilmente accessibili. Una vera montagna di dati. Sta a noi impegnarci nella sua scalata o lasciare che quella montagna imploda e il materiale ci travolga» ([7]: 9). È dunque proprio come supporto alla «scalata» di questa «montagna di dati» che nasce il progetto «epistolarITA», di cui si vogliono qui presentare i fondamenti teorici e la metodologia utilizzata.

2. LA 'REPUBBLICA DELLE LETTERE'

Molti dei già citati progetti internazionali insistono fin dal titolo sull'idea di *respublica litteraria* (o 'republic of letters', nella sua versione anglosassone). Benché già nel Rinascimento l'associazione tra 'lettere' ed 'epistole' fosse sicuramente attiva, in realtà il significato del sintagma andava molto più lontano: «to foreground correspondence in our conception of the republic of letters is not entirely unwarranted; but to define the *respublica litteraria* primarily as a republic of epistolary communication would be a grave mistake» ([11]: 26). Per capire le sfumature di tale concetto si preferisce oggi immaginare questa repubblica come un sodalizio di dotti (*sodalitas doctorum*), un mondo di eruditi (*orbis eruditorum*), una schiera di letterati (*omnis litteratorum cohors*) (*Ibid.*).

Nella sua pretesa di esaustività questo approccio cela tuttavia un'altra insidia, e cioè quella di limitare lo sguardo ai soli intellettuali. Benché infatti, nei secoli in esame – e proprio in virtù di un uso più intensivo dello strumento-lettera – l'*intelligenza* europea si trovò a comunicare per iscritto in maniera crescente², gli eruditi che ne facevano parte non rappresentavano che una porzione circoscritta di un insieme molto più esteso e trasversale. Se lo sviluppo della rete postale europea fu certamente una causa importante della riduzione della distanza che separava ad esempio un umanista fiammingo da uno portoghese, non si può limitare la portata degli scambi epistolari di età moderna alle sole interazioni tra intellettuali: scrivevano lettere anche i mercanti e i militari, i medici e i diplomatici, i nunzi e gli artisti, di ogni ceto e condizione socio-culturale. Possediamo epistole scritte da Leonardo da Vinci, autodefinitosi «omo senza lettere» (*Codice Atlantico*, c. 327v), di Michelangelo Buonarroti, «scorrecto in gramatica» ([1]: IV, 177), nonché di una vasta rappresentanza di semicolti che, lungi dall'aver ambizioni letterarie, proprio nel *medium* epistolare trovarono il loro canale privilegiato di comunicazione. In questa *respublica* il sodalizio – se di sodalizio si può ancora parlare – di certo non si limitava agli eruditi.

Non per questo però il concetto dovrà essere abbandonato, anzi: l'intuizione a nostro avviso più interessante dietro l'idea di una *respublica litteraria* è proprio quella, originaria, di un'unione di scriventi, non necessariamente letterati ma uniti dalle lettere, dove 'lettere' questa volta significherà, *strictu sensu*, 'epistole'. Questa comunità si fondava su un'interrelazione che ovviamente aveva numerosi vicoli ciechi – tranne rare eccezioni, l'umanista non corrispondeva con il fabbro, né l'artigiano scriveva al nunzio apostolico – ma che oggi, grazie alle tecnologie digitali, possiamo analizzare come un unico, grande documento. Una «montagna di dati» che possiedono più connessioni di quante immaginiamo, e che sta a noi riuscire a mettere in evidenza.

² Nel contributo appena citato Van Miert, Hoston e Wallnig forniscono vari esempi di umanisti, eruditi e scienziati che, tra Cinque- e Ottocento, rifletterono sulla questione (cfr. in particolare le pagine 33-40).

3. ALLA RICERCA DI SIMILARITÀ SIGNIFICATIVE

Posta la questione dell'interconnessione dei dati epistolari, la tematica che andrà affrontata è come rendere manifesti i collegamenti esistenti all'interno del corpus di lettere italiane della prima modernità. Esistono vari modi per fornire una risposta a tale problematica, a seconda del taglio che le si vuole dare. Con il progetto «epistolarITA» abbiamo voluto creare un algoritmo che portasse alla luce similarità significative – torneremo in seguito sul sintagma – le quali permettano di identificare gruppi di lettere che altrimenti sarebbe difficile, se non quasi impossibile, mettere in relazione. Alcuni esempi aiuteranno a capire meglio l'idea.

Uno storico che voglia indagare i rapporti tra Michelangelo Buonarroti e Leone X si orienterà probabilmente, presto o tardi, verso le lettere del primo, alla ricerca di brani testuali in cui l'artista descrive ai suoi corrispondenti alcuni episodi che hanno il papa come protagonista. Il suo carteggio offre numerosi passaggi di questo tipo: «La spesa della facciata, nel modo che io intendo di farla e metterla in opera, fra ogni cosa, *che 'l Papa non s'abbi a impacciare più di niente*, non può esser mancho [...] che di trenta cinque mila ducati d'oro» (02.05.1517, a Domenico Buoninsegni); «Ora, *se 'l Papa fa bole da potere rubare*, io priegho Vostra Signioria eminentissima ne facci fare una anchora a me, perché n'ò più bisogno di loro» (07.15.1518, a Giulio de' Medici)³. Tuttavia sarebbe inevitabilmente utopistico pensare che il testo di una lettera rifletta la reale percezione della realtà da parte del mittente, senza alcun condizionamento di fattori quali, ad esempio, la presenza in filigrana del destinatario. A seconda dell'interlocutore Michelangelo avrà dunque calibrato le sue rimostranze contro il papa in modi sempre diversi, a volte più sfumati, a volte più diretti; sarà dunque solo dalla lettura simultanea di tutte le fonti che trattano di questo medesimo argomento che ci si potrà avvicinare il più possibile all'opinione che lo scultore aveva di Leone⁴.

L'algoritmo che stiamo sviluppando permette di inserire un qualsiasi testo epistolare all'interno di un database collegato con le principali edizioni – digitali, ma non solo – di lettere italiane della prima modernità⁵, per ottenere come output i testi ad esso più 'simili' (anche questo è un aggettivo il cui significato verrà precisato oltre). Per proseguire con l'esempio fittizio appena riportato si potrebbe ipotizzare che Michelangelo abbia discusso di Leone X anche a voce con suo fratello Buonarroti, e che questi abbia riferito parti di conversazione a sua moglie Cassandra Puccini, la quale potrebbe avere accennato alla questione in una lettera inviata a suo padre Bernardo. Così, se tale lettera fosse presente nel nostro database, inserendo l'estratto della lettera di Michelangelo a Domenico Buoninsegni citata sopra, l'algoritmo sarebbe in grado di metterla in corrispondenza con questa lettera inviata da Cassandra a Bernardo. Nel segnalarla all'utente, l'algoritmo gli permetterebbe di venire a conoscenza di una fonte documentaria supplementare, a cui egli difficilmente sarebbe arrivato attraverso ricerche tradizionali, e che fornirebbe una nuova prospettiva sull'argomento da lui indagato.

Un'idea simile era già stata proposta, dieci anni fa, dai membri del progetto «ePistolarium», il cui algoritmo è stato aggiornato per l'ultima volta – secondo quanto si evince dal loro sito – nel 2013⁶. Benché l'approccio alla base di «ePistolarium» converga con il nostro negli obiettivi perseguiti, la metodologia da noi utilizzata – complice anche una drastica evoluzione nell'ultimo decennio di alcune specifiche tecniche di analisi – è alquanto differente.

4. COME FUNZIONA L'ALGORITMO DI «EPISTOLARITA»

La messa online della versione *alfa* dell'algoritmo di «epistolarITA» è prevista per la fine del 2021. Sul modello di quanto avviene con i «Voyant Tools», l'utente del database «epistolarITA» sarà invitato a introdurre nel sito un testo epistolare a sua scelta per il quale ha interesse a ottenere ulteriori informazioni. L'algoritmo soggiacente al database è un sistema di raccomandazione content-based il quale, dopo avere messo in relazione tra loro i testi epistolari italiani della prima modernità, si propone di fare emergere le loro prossimità lessicali e semantiche. L'algoritmo si basa sulla combinazione di un sistema di word embedding che sfrutta l'utilizzo congiunto delle tecniche W2Vec e TF-IDF e di un sistema di named entity recognition che consente di associare i medesimi luoghi, date e persone citate nei testi. L'integrazione dei risultati di queste operazioni permette di ottenere un valore numerico compreso tra 0 e 1 indicante la distanza che separa il testo

³ Le citazioni sono tratte da <http://www.memofonte.it/ricerche/michelangelo-buonarroti/> (ultima visualizzazione: 07.09.2021); i corsivi sono nostri.

⁴ Un caso simile è descritto in ([9]), dove si osserva come Alessandro Farnese descriva la carestia che nell'inverno 1586 colpì l'esercito del generale Charles de Mansfeld con modalità molto diverse a seconda del destinatario delle sue lettere e, di conseguenza, dei suoi scopi comunicativi.

⁵ Al momento hanno aderito a «epistolarITA» vari progetti di ricerca nazionali e internazionali, più – a titolo personale – numerosi altri editori. Per una panoramica completa e aggiornata delle collaborazioni si veda la pagina <http://web.philo.ulg.ac.be/epistolarita/le-collaborazioni/> (ultima visualizzazione: 07.09.2021).

⁶ Cfr. <http://ckcc.huylgens.knaw.nl/> (ultima visualizzazione: 07.09.2021). Sui presupposti teorici alla base dell'algoritmo di *topic modeling* utilizzato nel quadro di «ePistolarium» si veda ([12]).

dell'utente da ogni altro testo facente parte del corpus a cui «epistolarITA» ha accesso⁷; più il valore tende a 1, più la coppia di lettere a cui si esso riferisce può essere considerata 'simile'.

I documenti epistolari di cui si compone il corpus sono stati inizialmente sottoposti a fasi standard di pre-processing quali la tokenizzazione, la rimozione del rumore e delle stopword, e procedimenti semiautomatici di lemmatizzazione e di suddivisione in paragrafi⁸. In seguito all'introduzione di una lettera da parte dell'utente, l'algoritmo calcolerà il TF-IDF effettuando un test per ottimizzare il processo di parameter tuning (max features, min/max df); dopo questa operazione, tramite l'applicazione di FastText, per ogni vocabolo ricavato dal passaggio precedente saranno acquisite le 300 caratteristiche ('features') che costituiscono il word embedding di W2Vec⁹. Ogni documento sarà descritto tramite un vettore composto dalla media delle 300 caratteristiche di ogni termine, ciascuna di esse moltiplicata per il valore del TF-IDF. Questa matrice verrà poi sottoposta a tecniche di riduzione della dimensionalità, quali la PCA; successivamente sarà calcolata, per ogni coppia di documenti epistolari in nostro possesso, la distanza del coseno ([8]).

Il valore così conseguito sarà in seguito combinato con il valore complessivo di entità nominali uguali rinvenute, con l'utilizzo della libreria Spacy, per ogni coppia di documenti. Il valore di ogni coppia di entità nominali sarà stabilito sulla base sul TF-IDF dell'entità in questione, moltiplicato per due iperparametri: il primo, *c*, è una variabile dummy che varia in funzione dell'importanza da noi accordata alla categoria (antroponimo, toponimo, data etc.) di cui l'entità fa parte; il secondo parametro, *k*, avrà un valore fisso – ancora da stabilire – compreso tra 0 e 1, a seconda del peso che reputeremo necessario dare alla named entity recognition.

In seguito alla normalizzazione dei valori, dopo aver combinato le due parti dell'algoritmo, sarà possibile determinare il ranking finale di similarità testuale tra i documenti facenti parte del corpus. La creazione dei vettori che rappresentano i singoli paragrafi delle nostre lettere sarà calcolata in batch al fine di rendere più efficiente il software in produzione poiché, in seguito all'introduzione di un nuovo documento da analizzare, la macchina applicherà gli algoritmi con i parametri appresi nell'addestramento precedente solo su tale nuovo documento, inserito dall'utente a run time.

Attualmente l'algoritmo è in fase di ultimazione, e – come si è accennato sopra – la sua versione *alfa* sarà disponibile online entro la fine del 2021. Il convegno AIUCD sarà dunque un'eccellente occasione per presentare per la prima volta a un pubblico di specialisti il funzionamento del database, e per discutere di aspetti tecnici e teorici legati alle analisi di similarità testuali portate avanti nell'ambito del progetto «epistolarITA».

BIBLIOGRAFIA

- [1] Barocchi, Paola, e Renzo Ristori. 1965. *Il Carteggio diretto di Michelangelo*. Edizione postuma di G. Poggi. 5 voll. Firenze.
- [2] Bramanti, Vanni. 2020. *Lettere*. Luigi Alamanni. Roma.
- [3] Carminati, Clizia. 2020. «Le corrispondenze letterarie del Cinquecento e del Seicento: metodi e iniziative di studio». *Mélanges de l'École française de Rome - Italie et Méditerranée modernes et contemporaines* 132: 339–53.
- [4] Comelli, Michele. 2020. *Corrispondenza con Alessandro Farnese*. Giovanni Della Casa. Vol. 1. Roma.
- [5] Gaye, Johannes. 1839. *Carteggio inedito d'artisti dei secoli XIV, XV, XVI*. 3 voll. Firenze.
- [6] Migliorini, Bruno. 1960. *Storia della lingua italiana*. Firenze.
- [7] Procaccioli, Paolo. 2019. «Epistolografia tra pratica e teoria». In *L'epistolografia di Antico Regime*, Procaccioli, Paolo, 9–33. Sarnico.
- [8] Raunak, Vikas, Vivek Gupta, e Florian Metze. 2019. «Effective Dimensionality Reduction for Word Embeddings». In *Augenstein, Isabelle et al. Proceedings of the 4th Workshop on Representation Learning for NLP*, 235–43. Firenze.
- [9] Rocco, André, e Gianluca Valenti. in corso di stampa. *Les correspondances des expatriés italiens aux anciens Pays-Bas: réflexions pragmatiques et historiographiques à partir de quelques lettres d'Alexandre Farnèse*.

⁷ L'elenco dei testi presenti nel database è disponibile qui: <http://web.philo.ulg.ac.be/epistolarita/il-corpus/> (ultima visualizzazione: 10.11.2021).

⁸ Come punto di partenza per la lemmatizzazione dei testi abbiamo preso le forme e i lemmi catalogati nei database TLAVI (<http://www.tlavi.it/tesoro/>) e AGLIO (<http://aglio.ovi.cnr.it/>) (ultima visualizzazione: 07.09.2021). Questi repertori sono stati ampliati con l'aggiunta di nuove forme – per le quali si è provveduto a una lemmatizzazione manuale – riscontrate in un campione di lettere facenti parte del nostro corpus epistolare. I limiti di un'operazione di questo tipo condotta su un corpus non omogeneo, composto da un'estesa gamma di scriventi e che – per sua stessa natura – presenta un'ampia variazione sociolinguistica, costituiscono un problema teorico di vasta portata, non esauribile in questa sede (sulla questione della norma linguistica nell'Italia della prima modernità si rimanda all'ancora attualissima sintesi di Migliorini ([6]), capitoli VII-IX). Tuttavia si anticipa che i primi risultati sono incoraggianti, e che il repertorio di forme e lemmi così ottenuto permette al momento di lemmatizzare automaticamente l'85-90% delle parole presenti nelle nostre lettere. È previsto un periodico aggiornamento del repertorio sulla base delle nuove forme che entreranno a far parte del corpus con il progressivo inserimento di ulteriori testi.

⁹ I vettori attuali sono ricavati da modelli preesistenti, allenati sull'italiano contemporaneo. Questa scelta si è resa inevitabile poiché la quantità di documenti che abbiamo attualmente a disposizione non è tale da permetterci di addestrare un modello direttamente sui testi che fanno parte del nostro corpus; un addestramento *ad hoc* potrà essere approntato solo in una fase successiva del progetto, nel momento in cui avremo integrato nel database una massa di lettere sufficientemente elevata.

- [10] Russo, Emilio, a cura di. 2020. *Torquato Tasso. Lettere (1587-1589)*. Roma.
- [11] Van Miert, Dirk, Howard Hotson, e Thomas Wallnig. 2019. «What Was the Republic of Letters?» In *Reassembling the Republic of Letters in the Digital Age*, Hotson, Howard&Wallnig, Thomas, 23–40. Göttingen: Standards, Systems, Scholarship.
- [12] Wittek, Peter, e Walter Ravenek. 2011. «Supporting the Exploration of a Corpus of 17th-Century Scholarly Correspondences by Topic Modeling». In *Supporting Digital Humanities 2011: Answering the unaskable*, Maegaard, Bente. Copenhagen. <https://www.diva-portal.org/smash/get/diva2:887360/FULLTEXT01.pdf>.

Sessione Testi 5

Hélèn Metzger

Visualizing the genetic process of literary works

Valentina Pallacci¹, Andrea Benedetti², Tommaso Elli³, Elena Spadini⁴,

Michele Mauri⁵, Daniel Maggetti⁶

¹Politecnico di Milano, Italy - valentina.pallacci@mail.polimi.it

²Politecnico di Milano, Italy - andrea.benedetti@polimi.it

³Politecnico di Milano, Italy - tommaso.elli@polimi.it

⁴University of Lausanne, Switzerland - elena.spadini@unil.ch

⁵Politecnico di Milano, Italy - michele.mauri@polimi.it

⁶University of Lausanne, Switzerland - daniel.maggetti@unil.ch

ABSTRACT

The paper presents the preliminary results of a research aimed at designing a visual model to represent the genetic process of literary works, namely the network of texts (such as drafts, diary notes, and previously published materials) used by an author to develop a new work. In the paper, the visual model is described highlighting the main design choices, the design process, and its assessment with a domain expert. Both the underlying data model and the visual model aim to be general enough to offer possibilities for application to other cases of scholarly research.

KEYWORDS

Data visualization, Visual model, Genetic criticism, Filologia d'autore.

TALK

1. INTRODUCTION

In this paper we introduce the preliminary results of a research aimed at designing a visual model for studying the genetic process of literary works. The focus of the paper is the design process carried on in collaboration between communication designers specialized in data visualization, and scholars of the project “Gustave Roud. *Œuvres complètes*”¹.

Gustave Roud (1897-1976) was a Swiss poet, photographer, and translator, also active in arts and literary criticism. The genesis of Roud’s works is grounded in his diary and is characterized by the reuse of already published materials. For the study of genetic processes, scholars consider multiple sources and examine the relationships between them. A particularly relevant concept in the fields of genetic criticism and *filologia d'autore* is the one of genetic dossier: a group of plans, sketches, drafts, and clear copies that testify to the project of a literary work ([7];[6]); they are identified by scholars interpreting the documents, and can be represented as a node in the network that stands between documents and publications.

Before the beginning of the collaboration with communication designers, scholars of the project “Gustave Roud. *Œuvres complètes*” developed a data model for literary genesis in the form of an OWL 2 ontology². The data model is aimed at storing and representing the literary genesis in the form of a conceptual network of documents and publications ([5]). Presenting the data model is beyond the scope of this paper, which instead focus on the visual model.

In what follows, we pursue the design process (section 2), briefly introduce the assessment activity (section 3) and present future works in the conclusions (section 4).

2. DESIGN PROCESS

The aim of the collaboration between the two partners is the creation of a model that enables the visual reading of the genesis of Roud’s works. The visual model is built on top of the aforementioned data model and could be adapted to the study of other genetic processes. The design process was divided in three key moments: (section 2.1) definition of design requirements, (section 2.2) visualization process, and (section 3) validation of a preliminary outcomes with a domain expert.

¹ University of Lausanne, 2017-2021 (<https://p3.snf.ch/Project-157970>).

² The project ontology is available at. The project makes use of the framework DSP (DaSCH Service Platform, <https://dsp.dasch.swiss/>): DSP provides a general ontology, on top of which each project defines its own ontology. The part of the ontology devoted to literary genesis is documented at <https://gen-o.github.io/> and at <https://github.com/gustaveroudproject/geneticNetworksDataViz>.

2.1 DESIGN REQUIREMENTS

Designers used structured interviews ([9]) to inform their design process. The method is commonly used to gather a rich and holistic understanding about the topic of a work of visualization ([4]). The interviews focused on Roud’s production and on the work of the domain experts, that is the data modeling process and preliminary data representations (sketches and hand-made network visualizations). In addition to interviews, designers autonomously explored the data using an open-source tool aimed at the visual exploration of network graphs ([1]). The most important finding of these research activities regards the nature of the genetic dossiers: they support the understanding of the genesis of a work, but, at the same time, they are interpretations of scholars, which could potentially evolve with the sedimentation of their knowledge about Gustave Roud and with the growth of the author’s archive.

The interviews and the exploration of the data lead to the identification of four design requirements for the visual model, which should:

- R1.** be designed for an audience of domain experts;
- R2.** explicit the role of genetic dossiers as an interpretive layer;
- R3.** differentiate between typologies of objects (diary entries, drafts, books and periodicals), clusters (genetic dossiers, and works separated into parts) and relationships (used in, part of, rewritten in);
- R4.** mediate between the complexity of the data model and the legibility of the networks.

2.2 VISUALIZATION PROCESS

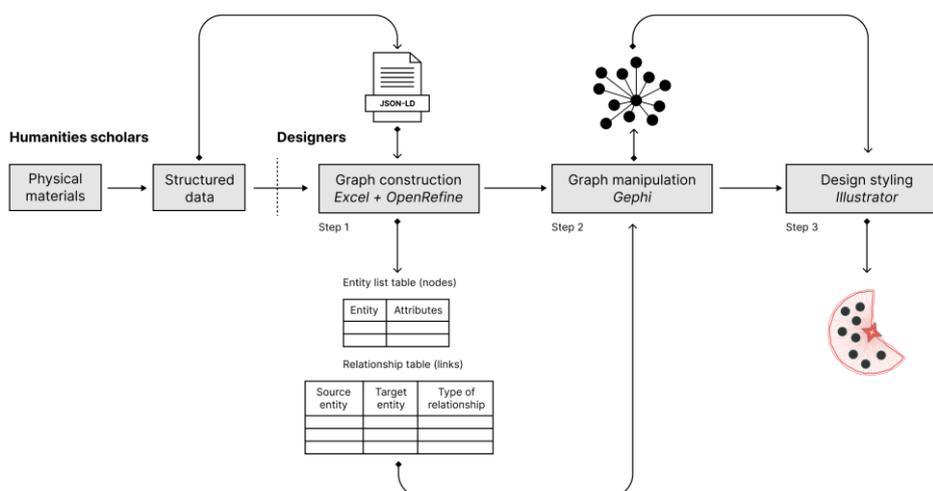


Figure 1. Schematization of the visualization process in relationship to the work done by humanities scholars. The process unfolds in three major steps, each with its own output.

The visualization process unfolds in three steps (fig. 1). Firstly, it was necessary to convert the data into a tabular format (step 1) to fit the selected tool for network visualization ([1]). Every genetic network constitutes of a JSON-LD file produced by scholars³, and designers defined a process to convert them into two tables: one for nodes (publications, manuscripts, periodicals and their parts) and one for links (relationships among nodes, like rewriting and reuse). By taking into account the outcomes of preliminary explorations, designers decided to consider genetic dossiers not as nodes, but as sets of nodes, the center of which would be the resulting publication. Having a separate way of defining the genetic dossiers was crucial in simplifying the reading of the network (R4); in addition, it helped readers in understanding the existence of two layers: a factual one with publications and manuscripts, and an interpretive one with genetic dossiers (R2, fig. 2). This operation was implemented by bypassing the genetic dossiers nodes and directly connecting used materials to the publication that resulted from them.

³ The data in JSON-LD are available at <https://github.com/gustaveroudproject/geneticNetworksDataViz>.

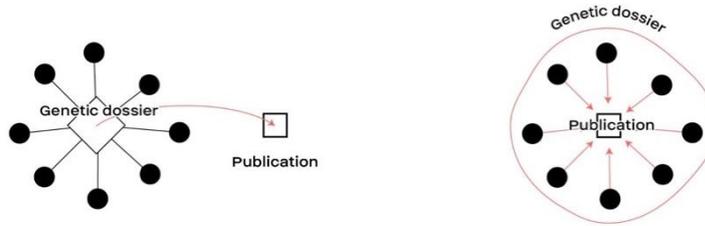


Figure 2. The shift from a node to an enclosure, to represent genetic dossiers.

Successively, designers spatialized the genetic networks using a force directed layout algorithm (step 2), which simulates a physical system in order to calculate the position of the elements of the networks ([8]). In the simulation, nodes repulse or attract each other according to the links that connect them. In addition, designers set a stronger pull between the nodes belonging to the same genetic dossier; they achieved this result by incrementing the weight value of the corresponding links. This kind of direct manipulation resulted in a compact positioning of the elements of the genetic dossiers, that are visually separated from nodes connected through different kinds of connections. This spatialization allowed researchers to perform a visual network analysis ([10]) and to identify clusters and outliers (R3), and observing the overall shape of the network (fig. 3).

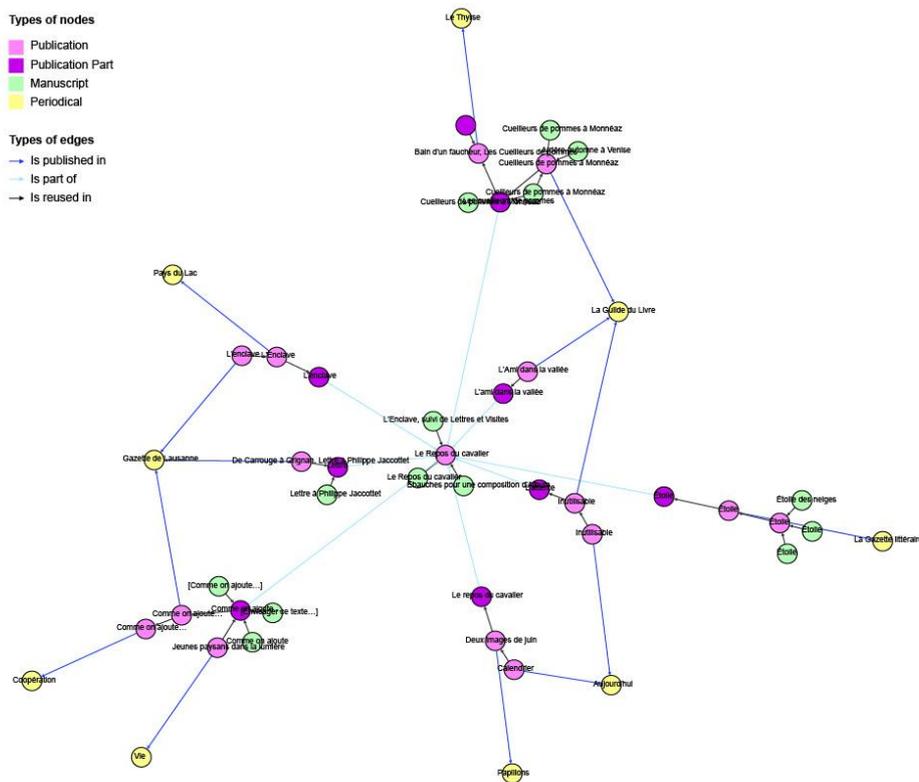


Figure 3. The result of the spatialization using *Force Atlas 2* (in this case the genesis of Roud's work *Le Repos du Cavalier*) is a visualization in which the quality of linkages determines the position of nodes. Different colored nodes identify different kinds of entities (books, articles, etc...) and different colored links identify different kinds of relationships (rewriting, reuse, publications...).

Lastly, designers reworked the network visualizations using a vector-based editing software (step 3). During this activity they applied a visual embellishment that employs a visual metaphor, following a consolidated strategy for improving comprehension and memorability of visualizations ([2]). Gustave Roud was often inspired by nature, countryside, or stars and, after having analyzed three different options, astral maps emerged as a suitable metaphor for supporting the creation of a visual language: they represent objects of different nature (stars), and they embed the human interpretations of stars arrangements (constellations). The metaphor inspired the form of nodes and connections (fig. 4) and allowed for the

integration of a layer of domain experts' interpretations (fig. 5). The genetic dossiers, indeed, were treated as an additional layer of the visualization, one that groups elements.

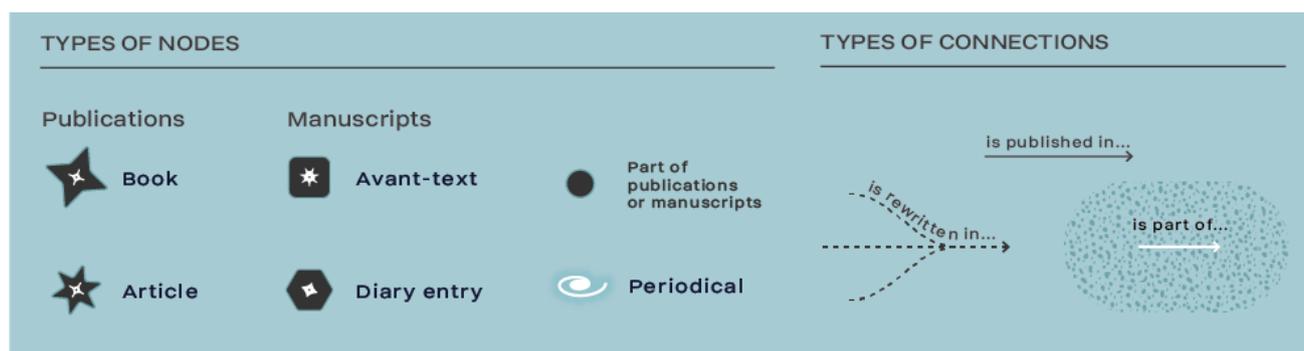


Figure 4. The legend of the visualization, divided between nodes and connections, with resemblances with stars, nebulae, and galaxies. Each typology of entity is represented by a different glyph, and each typology of connection is stylized in a different manner.



Figure 5. Representations of genetic dossiers. In the visualization, genetic dossiers are treated as an interpretive layer: they surround documents that converge in a publication.

3. PRELIMINARY ASSESSMENT

Before starting the semi-automatic production of a larger number of visualizations, designers and scholars deemed important to assess the functioning of the visual model. The assessment activity is intended to inform another design iteration and to identify meaningful aspects on which to further work.

3.1 METHODS

The design outcomes were assessed with one of the scholars, creator of the data model and knowledgeable about the overall goals of the project, but who was never involved in any visualization activity. The scholar will be referred to as domain expert. The assessment uses three genetic networks⁴ (one of them in fig. 6) and is based on a talk-aloud protocol ([3]) and on a task-oriented approach. The domain expert was asked to read aloud the network and enunciate aloud any comment that might occur. The domain expert was then asked to address specific tasks: (1) identify the central work of the visualization, (2) count the total amount of genetic dossiers, (3) identify the documents included in each dossier and the resulting publication, and (4) identify the periodicals in which the poet published.

3.2 OUTCOMES

The domain expert was able to read the visualizations almost entirely (R1, R3 and R4) and managed to distinguish genetic dossiers from other elements (R2). The result suggests that the design needs refinements but is overall solid. By looking at the networks, the domain expert was able to formulate hypothesis about the works genesis, especially on the role of periodicals in the selection of already published materials to reuse. The domain expert also found stimulating the possibility to move from the observation of the genesis of a single work, to the observation of general trends in Roud's way of working. Possible improvements that emerged are to: (1) **improve the structure of the genetic dossiers**, making it easier to identify its center and the elements belonging to them; (2) **differentiate between hierarchies of genetic dossiers**, since the expert identified genetic dossiers related to the main work and those related to further reuses; (3) **use chronology to clarify the reading order of the network**.

⁴ All design outcomes can be consulted at: <https://bit.ly/roud-aiucd22>

- [5] Carpendale, Sheelagh. 2008. "Evaluating Information Visualizations." In *Information Visualization*, A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, 4950:19–45. Berlin, Heidelberg: Springer Berlin Heidelberg.
- [6] Christen, Alessio, and Elena Spadini. 2019. "Modeling Genetic Networks. Gustave Roud's Œuvre, from Diary to Poetry Collections." *Umanistica Digitale*, 7 (December).
- [7] De Biasi, A. 2004. "Toward a Science of Literature: Manuscript Analysis and the Genesis of the Work." In *Genetic Criticism: Texts and Avant-Textes*, J. Deppman, D. Ferrer, and M. Groden, 26–68. Philadelphia: University of Pennsylvania Press.
- [8] Grésillon, Almuth. 1994. *Éléments de Critique Génétique: Lire Les Manuscrits Modernes*. Paris: Presses universitaires de France,.
- [9] Jacomy, Venturini, Heymann, and Bastian. 2014. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software." *PLoS ONE* 9 (6): e98679.
- [10] Seidman, Irving. 2006. *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences*. 3rd ed. New York: Teachers College Press.
- [11] Venturini, Tommaso, Mathieu Jacomy, and Pablo Jensen. 2021. "What Do We See When We Look at Networks: Visual Network Analysis, Relational Ambiguity, and Force-Directed Layouts" *Big Data & Society*, 8 (1).

Analisi linguistica e pseudonimizzazione: strumenti e paradigmi

Laura Clemenzi¹, Francesca Fusco², Daniele Fusi³, Giulia Lombardi⁴

¹Università degli Studi della Tuscia, Italia - laura.clemenzi@unitus.it

²Università del Salento, Italia - francesca.fusco@unisalento.it

³Bamberg University, Germania - daniele.fusi@unive.it

⁴Università di Genova, Italia - giulia.lombardi@edu.unige.it

ABSTRACT

In questo contributo si presenta la procedura innovativa messa a punto nell'ambito del progetto PRIN "La chiarezza degli atti del processo (AttiChiari): una base di dati inedita per lo studioso e il cittadino" per il trattamento dei testi giuridici, funzionale sia alla pseudonimizzazione dei dati sensibili, sia all'analisi linguistica. Si introduce inoltre il motore di ricerca che consentirà di esplorare il *corpus* in fase di costruzione.

PAROLE CHIAVE

Analisi linguistica, *corpora*, marcatura, motore di ricerca, pseudonimizzazione.

INTERVENTO

1. IL PROGETTO, GLI OBIETTIVI, LE QUESTIONI¹

Il PRIN 2017 "La chiarezza degli atti del processo (AttiChiari): una base di dati inedita per lo studioso e il cittadino" – progetto a cui collaborano linguisti e giuristi degli atenei di Genova, Firenze, Lecce e Viterbo – si prefigge di creare una nuova risorsa per una scrittura efficace degli atti processuali². In particolare, in una prima fase l'obiettivo è allestire, per fini di studio linguistico, un *corpus* sincronico di atti di parte di circa tre milioni di parole rappresentativo, per tipologie testuali e provenienza geografica, delle diverse prassi di scrittura degli avvocati. Successivamente, con i testi raccolti, si intende realizzare una base dati interrogabile che in una specifica sezione includa esempi di scrittura chiara ed efficace, utili per il giurista e anche per il cittadino³.

La peculiarità dei testi che compongono il *corpus* è la presenza al loro interno di dati sensibili, la cui diffusione violerebbe il diritto alla riservatezza delle parti, di eventuali terzi coinvolti e dei procuratori costituiti. È dunque necessaria, propedeuticamente a qualsiasi tipo di studio, e come requisito stesso per ottenere l'accesso agli atti, un'attività di anonimizzazione dei documenti che renda irricognoscibili le vicende e i soggetti.

Le prassi di anonimizzazione usate tradizionalmente in Italia per riprodurre e diffondere testi giuridici che contengono dati sensibili, come ad esempio i provvedimenti giudiziari, consistono nella mera eliminazione di tali dati tramite l'omissione o la cancellatura con tratti neri, oppure nella loro sostituzione con asterischi, *omissis*, lettere, o altri segni grafici⁴: tutte prassi non compatibili con le esigenze del linguista, che necessita di testi massimamente leggibili e quanto più possibile completi per poter analizzare appieno le strategie usate dagli avvocati nel riferirsi alla parte assistita, alla controparte e agli altri soggetti del processo, sia all'interno di uno stesso atto, sia, in un'ottica di studio di tipo "verticale" e intertestuale, negli altri atti relativi allo stesso giudizio. Oscurando nomi, toponimi, date e ogni altro dato sensibile, difatti verrebbe meno la possibilità di individuare e distinguere le parti processuali e di ricostruire le vicende narrate: sarebbe quindi impossibile dipanare l'intreccio delle voci scriventi (cfr. [12]: 30).

Con questo contributo intendiamo proporre un modello possibile di trattamento semiautomatico degli atti di parte italiani come operazione propedeutica e funzionale sia alla tutela dei dati sensibili, sia all'analisi linguistica e contenutistica dei

¹ Il testo è stato concordato e rivisto da tutti gli autori; tuttavia, ai fini dell'attribuzione della paternità delle singole parti di cui si compone, vanno attribuiti a Laura Clemenzi il paragrafo 1, a Francesca Fusco i paragrafi 2-3, a Daniele Fusi i paragrafi 6-7, a Giulia Lombardi i paragrafi 4-5.

² Per alcuni primi studi sulla lingua degli atti di parte, cfr. ([22];[18];[19];[3];[5];[16];[8];[2]).

³ Per maggiori dettagli sugli obiettivi del progetto e sulle procedure adottate, si rinvia agli interventi raccolti nel volume curato da Gualdo e Clemenzi ([15]); per alcuni esempi di fenomeni linguistici ricercabili nella base dati Atti Chiari, cfr. in particolare ([4]).

⁴ Per alcuni esempi di prassi di anonimizzazione tradizionali, cfr. ([1]); in questo testo, si veda più avanti la Figura 4. Segnaliamo di passaggio, come eccezione, il caso del Consiglio di Giustizia Amministrativa della Regione Sicilia, che con la sentenza n. 1134/2020 ha deciso di sostituire gli *omissis* con nomi di fantasia.

testi. Il modello sviluppato deve ancora essere testato nella sua interezza, ma sono stati già condotti, con esito positivo, alcuni test pilota sugli atti a disposizione del gruppo di ricerca del PRIN Atti Chiari.

2. I REQUISITI DEL PROGRAMMA

Sulla base degli obiettivi e delle esigenze del gruppo di ricerca descritte poco sopra, si rende necessario sostituire i dati sensibili contenuti negli atti con dati fittizi della stessa categoria, secondo una procedura di “pseudonimizzazione”, definita dal *Regolamento generale sulla protezione dei dati* (Reg. U.E. n. 2016/679), art. 4, c. 5, come «il trattamento dei dati personali in modo tale che i dati personali non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e soggette a misure tecniche e organizzative intese a garantire che tali dati personali non siano attribuiti a una persona fisica identificata o identificabile»⁵.

In particolare, ai fini dell'analisi linguistica è fondamentale mantenere la coerenza concettuale-semantica tra i dati originali e quelli fittizi e la coerenza morfosintattica dei dati fittizi con il contesto (è importante, dunque, che il dato nuovo corrisponda in maniera univoca all'originale in tutte le occorrenze del testo e che ne conservi il genere, per non alterare la morfosintassi della frase in cui è inserito) (cfr. [12]: 30-34). A tale fine, una sostituzione automatica dei dati sensibili, tramite un programma che attinga a liste predefinite per i nomi e che modifichi sequenze numeriche alfanumeriche (quali date, targhe, fax, numeri di telefono, ecc.) si rivela la soluzione più efficace, in grado di ridurre il rischio di errore e di garantire un risultato uniforme.

Inoltre, dal momento che lo studio che si intende condurre sui testi è non solo di tipo linguistico, ma anche giuridico, è opportuno prevedere insieme di metadati diversi a seconda degli scopi: se, ad esempio, l'analisi linguistica presuppone l'inserimento di metadati relativi al paratesto, quella giuridica richiede che la sostituzione delle date non pregiudichi la ricostruzione cronologica dei fatti.

3. LA MARCATURA E LA PSEUDONIMIZZAZIONE

Ai fini del progetto è stato ideato un nuovo metodo per il mascheramento dei dati sensibili funzionale alle analisi da condurre. Si tratta di un metodo di annotazione a due fasi, ispirato ai modelli di Douglass et al. ([9]), Noumeir ([20]), Elger ([10]) e Dalianis ([7]) per la pseudonimizzazione delle cartelle cliniche, e di Oksanen ([21]) per la pseudonimizzazione degli atti giudiziari finlandesi.

Inizialmente si interviene sul testo con una leggera marcatura manuale, che, invece di togliere, aggiunge informazioni: l'operatore annota il testo di partenza direttamente in un applicativo di videoscrittura, secondo una sintassi concordata, che segnala sia la categoria del dato sensibile, sia il genere⁶. Posto che, come si è detto, il trattamento dei testi non è funzionale solo alla pseudonimizzazione in senso stretto, ma anche a uno studio di tipo linguistico, vi sono poi altri marcatori che non comportano la sostituzione della porzione di testo marcata, bensì sono propedeutici solo alla successiva analisi linguistica: ad esempio si è deciso di marcare i forestierismi, per i quali si usano i codici ISO 639, preceduti da *f-* (*foreign*).

Alla marcatura manuale segue la pseudonimizzazione automatica: per sostituire i dati identificativi il programma attinge a repertori di prenomi maschili e femminili, cognomi e toponimi⁷ (nel caso di nomi iniziati per vocale e preceduti da un *d* eufonica, il programma attinge a repertori di soli nomi iniziati per vocale). La coerenza concettuale-semantica all'interno del documento – o dei documenti, nel caso di più atti afferenti allo stesso giudizio – è garantita dal fatto che uno stesso dato sensibile, quando preceduto dallo stesso marcatore, è sostituito dal programma con il medesimo dato fittizio in tutte le sue occorrenze (cfr. [12]: 33-34). Cifre e sequenze alfanumeriche (come targhe, fax, numeri di telefono, ecc.) sono invece sostituite dal programma con stringhe di numeri e lettere casuali di pari estensione; un'attenzione particolare meritano le date, visto che per uno studio di tipo giuridico degli atti è necessario mantenere la coerenza dei riferimenti cronologici delle vicende fattuali e processuali in essi narrate: a tale scopo, il programma, per impostazione predefinita, lascia intatti mese e

⁵ La tecnica della pseudonimizzazione è richiamata in più parti del Reg. U.E. n. 2016/679 proprio come misura di «garanzia adeguata» della riservatezza dei dati: cfr. gli artt. 6, c. 4, 25, c. 1, 32, c. 1, 40, c. 2, 89, c. 1 (oltre ai *considerando* 26, 28, 29, 75, 78, 85, 156). Cfr. anche ([12]: 30-31).

⁶ I marcatori finora usati per identificare i dati sensibili sono: a-f-f (*anthroponym, female, first*) per gli antroponimi femminili; a-m-f (*anthroponym, male, first*) per gli antroponimi maschili; a-l (*anthroponym, last*) per i cognomi; j-f (*juridic person, female*) per i nomi propri di persone giuridiche di genere grammaticale femminile; j-m (*juridic person, male*) per i nomi propri di persone giuridiche di genere grammaticale maschile; t (*toponym*) per i toponimi; ad (*address*) per gli indirizzi; m (*e-mail*) per gli indirizzi di posta elettronica; d (*date*) per le date; n (*number*) per le cifre (es. numeri di telefono, importi in denaro, particelle catastali, ecc.); u per le stringhe alfanumeriche (es. codici fiscali, sigle delle province, targhe, ecc.); x per i dati da oscurare che non rientrano in nessuna delle precedenti categorie (sostituiti con ###).

⁷ Per gli scopi previsti dal progetto, non si è reso necessario distinguere ulteriori sottocategorie di toponimi, quali ad esempio città, paesi, Stati.

giorno, sottraendo all'anno un valore compreso fra un minimo (ad es. 5) e un massimo (ad es. 15), uguale in tutta la sessione di analisi (anche se resta comunque possibile optare per una sostituzione randomica delle date).

I documenti che si ottengono non contengono dati sensibili, ma restano perfettamente leggibili (e quindi ben si prestano ad analisi sia linguistiche, sia giuridiche): per vedere concretamente il funzionamento del procedimento di marcatura e pseudonimizzazione adottato nell'ambito del PRIN Atti Chiari e i suoi vantaggi in termini di leggibilità rispetto alle prassi anonimizzatorie tradizionali, si riporta di seguito l'*incipit* di un facsimile di atto di citazione in opposizione a decreto ingiuntivo nelle versioni originale, marcata e pseudonimizzata, cui viene affiancata, per confronto, la versione del medesimo documento anonimizzata mediante mero oscuramento dei dati⁸.

GIUDICE DI PACE DI TERMOLI
ATTO DI CITAZIONE IN OPPOSIZIONE
A DECRETO INGIUNTIVO EX ART. 645 C.P.C.

E CONTESTUALE ISTANZA DI SOSPENSIONE DELLA PROVVISORIA ESECUTIVITA'

La sottoscritta Avv. Gianna Barbieri del foro di Termoli, C.F.: BRBGNN87S46G045T, che rappresenta e difende ad ogni effetto di legge, in virtù di delega posta in calce al presente atto la **SIG.RA BELLINI GIANCARLA** nata a Vicenza il 24.11.1972 e residente in Termoli, Via Garibaldi, n. 4, C.F.: BLL GNC 72P52 R557X, ed elettivamente domiciliata presso lo studio del predetto avvocato in Termoli, Via XX Settembre, n. 56, con indicazione del n. fax al 0435/4530202, PEC: barbieri@pec.lambfa.it,

PREMESSO CHE

-in data 19.11. 2015 il Giudice di Pace di Termoli emetteva a favore di **Beta NPL S.p.a** decreto ingiuntivo provvisoriamente esecutivo n. 1234/2015 per la somma di € 3.400,00 oltre interessi di mora e spese della procedura;
-tale decreto, munito di formula esecutiva in data 03.12.2015, veniva notificato in data 12.12.2015 alla Sig.ra Giancarla Bellini;
- il predetto decreto è ingiusto ed illegittimo e avverso lo stesso si propone formale opposizione per i seguenti

MOTIVI

Figura 1 - Facsimile atto di citazione (versione originale)

ATTO DI CITAZIONE IN OPPOSIZIONE
A DECRETO INGIUNTIVO EX ART. 645 C.P.C.

E CONTESTUALE ISTANZA DI SOSPENSIONE DELLA PROVVISORIA ESECUTIVITA'

La sottoscritta Avv. Gradita Bertanzetti del foro di Tarcento, C.F.: TUWZWL08K38R553U, che rappresenta e difende ad ogni effetto di legge, in virtù di delega posta in calce al presente atto la **SIG.RA BANDINO GOFFREDA** nata a Verrès il 24/11/1965 e residente in Tarcento, Via Khepri Santori, 84, C.F.: IEE QGR 53E69 Z7781, ed elettivamente domiciliata presso lo studio del predetto avvocato in Tarcento, Via Giosia Lovine, 51, con indicazione del n. fax al 6259/8989756, PEC: bk5172@tiscali.it,

PREMESSO CHE

-in data 19/11/2008 il Giudice di Pace di Tarcento emetteva a favore di **Brillante S.p.a** decreto ingiuntivo provvisoriamente esecutivo n. 3079/4358 per la somma di € 4.287,51 oltre interessi di mora e spese della procedura;
-tale decreto, munito di formula esecutiva in data 3/12/2008, veniva notificato in data 12/12/2008 alla Sig.ra Goffreda Bandino;
- il predetto decreto è ingiusto ed illegittimo e avverso lo stesso si propone formale opposizione per i seguenti

MOTIVI

Figura 3 - Facsimile atto di citazione (versione pseudonimizzata)

GIUDICE DI PACE DI {t:TERMOLI}
ATTO DI CITAZIONE IN OPPOSIZIONE

A DECRETO INGIUNTIVO {f-lat:EX} ART. 645 C.P.C.

E CONTESTUALE ISTANZA DI SOSPENSIONE DELLA PROVVISORIA ESECUTIVITA'

La sottoscritta Avv. {a-f-f:Gianna} {a-l:Barbieri} del foro di {t:Termoli}, C.F.: {u:BRBGNN87S46G045T}, che rappresenta e difende ad ogni effetto di legge, in virtù di delega posta in calce al presente atto la **SIG.RA {a-l:BELLINI} {a-f-f:GIANCARLA}** nata a {t:Vicenza} il {d:24.11.1972} e residente in {t:Termoli}, Via {ad:Garibaldi, n. 4}, C.F.: {u:BLL GNC 72P52 R557X}, ed elettivamente domiciliata presso lo studio del predetto avvocato in {t:Termoli}, Via {ad:XX Settembre, n. 56}, con indicazione del n. fax al {n:0435/4530202}, PEC: {m:barbieri@pec.lambfa.it},

PREMESSO CHE

-in data {d:19.11. 2015} il Giudice di Pace di {t:Termoli} emetteva a favore di {j-f:Beta NPL} S.p.a decreto ingiuntivo provvisoriamente esecutivo n. {n:1234/2015} per la somma di € {n:3.400,00} oltre interessi di mora e spese della procedura;
-tale decreto, munito di formula esecutiva in data {d:03.12.2015}, veniva notificato in data {d:12.12.2015} alla Sig.ra {a-f-f:Giancarla} {a-l:Bellini};
- il predetto decreto è ingiusto ed illegittimo e avverso lo stesso si propone formale opposizione per i seguenti

MOTIVI

Figura 2 - Facsimile atto di citazione (versione marcata)

GIUDICE DI PACE DI ██████████

ATTO DI CITAZIONE IN OPPOSIZIONE

A DECRETO INGIUNTIVO EX ART. 645 C.P.C.

E CONTESTUALE ISTANZA DI SOSPENSIONE DELLA PROVVISORIA ESECUTIVITA'

La sottoscritta Avv. ██████████ del foro di ██████████, C.F.: ██████████, che rappresenta e difende ad ogni effetto di legge, in virtù di delega posta in calce al presente atto la **SIG.RA ██████████** nata a ██████████ il ██████████ e residente in ██████████, Via ██████████, C.F.: ██████████, ed elettivamente domiciliata presso lo studio del predetto avvocato in ██████████, Via ██████████, con indicazione del n. fax al ██████████, PEC: ██████████,

PREMESSO CHE

-in data ██████████ il Giudice di Pace di ██████████ emetteva a favore di ██████████ S.p.a decreto ingiuntivo provvisoriamente esecutivo n. ██████████ per la somma di € ██████████ oltre interessi di mora e spese della procedura;
-tale decreto, munito di formula esecutiva in data ██████████, veniva notificato in data ██████████ alla Sig.ra ██████████;
- il predetto decreto è ingiusto ed illegittimo e avverso lo stesso si propone formale opposizione per i seguenti

MOTIVI

Figura 4 - Facsimile atto di citazione (versione anonimizzata)

Sempre per gli scopi del progetto e per la natura dei testi, non si è reso necessario prevedere un sistema di depseudonimizzazione (cfr. [10];[20]), ovvero il processo inverso alla pseudonimizzazione che permette di recuperare in maniera inequivocabile i dati personali univocamente associati ai dati fittizi.

4. IL FLUSSO DEI DATI

Il processo descritto, inoltre, è guidato da un insieme variabile di regole configurabili a seconda degli obiettivi: infatti,

⁸ La Figura 1 riproduce il testo e la formattazione dell'originale ma contiene dati già fittizi. I metadati relativi alle informazioni paratestuali vengono conservati dal programma e dunque lo stile dell'originale (grassetto, corsivi, ecc.) è riprodotto anche nella versione pseudonimizzata. Per altri esempi di atti pseudonimizzati, cfr. ([12]: 31-39).

come abbiamo già sottolineato, nel nostro caso il trattamento dei testi non è funzionale solo alla pseudonimizzazione in senso stretto, ma anche all'analisi linguistica; alcuni accorgimenti come la marcatura dei forestierismi e l'attenzione posta al rispetto dei fenomeni fonosintattici (ad esempio la *d* eufonica) anche in fase di pseudonimizzazione, contribuiscono alla raccolta dei metadati. Come anticipato sopra e come si dirà meglio più avanti, altre fonti di metadati sono lo stesso formato digitale in *rich text*, che consente di recuperare aspetti tipografici, e altri strumenti esterni come i *POS taggers*⁹.

Un ulteriore beneficio offerto da questo apparentemente paradossale approccio, che aggiunge informazione solo per poterla togliere, è inoltre costituito dal fatto che il sistema di pseudonimizzazione diviene in grado di rimodellare il documento di partenza, dalla struttura puramente tipografica, in un documento semanticamente strutturato. In effetti, avvalendosi delle diverse fonti di metadati incluse nell'*input* il sistema ha la capacità di aggiungere al processo di pseudonimizzazione anche quello di conversione del documento, che dal formato di videoscrittura viene convertito in un vero e proprio documento TEI. In tal modo, nel processo completo vengono accodate le fasi di decodifica del formato originale, di pseudonimizzazione secondo un insieme variabile di regole, e di generazione di un documento TEI, corredato da eventuali rese tipografiche in HTML (v. Figura 3), sì da fornire agli operatori un immediato riscontro del loro operato (v. Figura 5)¹⁰. In effetti, le fonti dei metadati di un documento sono molteplici. Anzitutto, la leggera marcatura applicata, destinata sia all'offuscamento delle informazioni sensibili, sia all'annotazione di aspetti utili solo in funzione dell'analisi linguistica. In secondo luogo, un'ulteriore fonte è costituita dal formato di videoscrittura (DOCX nello specifico) nel quale vengono raccolti la pressoché totalità degli atti. Da esso naturalmente interessa estrarre solo un minimo sottoinsieme di informazioni tipografiche ritenute utili in fase di analisi degli aspetti paratestuali. Fra questi, un sottoinsieme della formattazione del testo viene direttamente estratto dal formato Office Open XML (ISO/IEC 29500). Infine, l'utilizzo di sistemi di *POS tagging* consente di ottenere con una buona approssimazione ulteriori metadati relativi alla lemmatizzazione e alla classificazione morfologica di ogni parola. Tutti questi metadati devono poi trovare posto nell'indice che nutrirà la base del motore di ricerca.

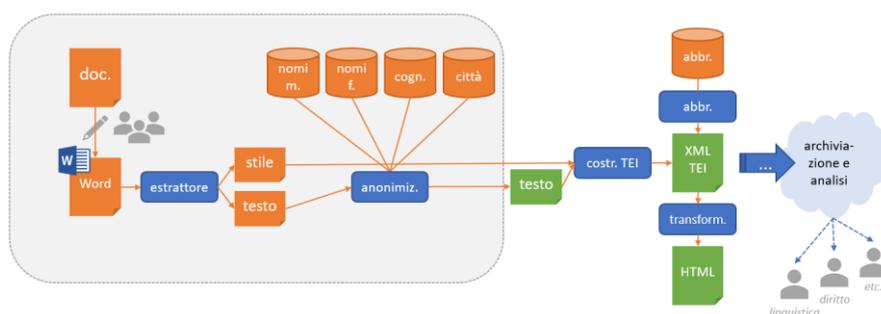


Figura 5 - Prima parte del flusso generale dei dati: il riquadro in grigio delimita l'area protetta, dalla quale nessun dato personale può uscire

Si riproducono di seguito alcuni estratti dei file intermedi del processo di trasformazione illustrato nella Figura 5, a partire dallo stesso atto usato per le esemplificazioni riportate nelle Figure 1-4. Per il riconoscimento delle abbreviazioni cui si fa riferimento nelle didascalie delle Figure 8 e 9, si veda più avanti il par. 6.

```
<pick>
<par nr="5" fmtId="4">
<run nr="1" fmtId="4">La sottoscritta Avv. {a-f-f:Gianna} {a-l:Barbieri} del foro di {t:Termoli}, C.F.: {u:BRBGNN87S46G045T}, che rappresenta e difende ad ogni effetto di legge, in virtù di delega posta in calce al presente atto la </run>
<run nr="13" fmtId="5">SIG.RA {a-l:BELLINI} {a-f-f:GIANCARLA} </run>
<run nr="19" fmtId="4"> nata a {t:Vicenza} il {d:24.11.1972} e residente in {t:Termoli}, Via {ad:Garibaldi n. 4}, C.F.: {u:BLL GNC 72P52 R557X}, ed elettivamente domiciliata presso lo studio del predetto avvocato in {t:Termoli}, Via {ad:XX Settembre n. 56}, con indicazione del n. fax al {n:0435/4530202}, PEC: {m:barbieri@pec.lambfa.it},</run>
</par>
</pick>
```

Figura 6 - Estratto del documento DOCX in un formato intermedio XML con i dati originali (gli attributi *fmtId* rimandano a insiemi di caratteristiche tipografiche ricavate dall'originale, e sono sciolti in un'apposita sezione)

⁹ Sui diversi livelli di annotazione dei *corpora*, tra cui il *POS (part of speech) tagging*, cioè l'attribuzione delle categorie grammaticali, si vedano almeno ([11]: 18-25) e ([6]: 84-94).

¹⁰ La Figura 5 è tratta da ([14]: 69); sul funzionamento e sui vantaggi del programma, cfr. ancora ([14];[15]).

```

<pick>
<par nr="5" fmtId="4">
<run nr="1" fmtId="4">La sottoscritta Avv. {a-f-f:Gradita} {a-l:Bertanzetti} del foro di {t:Tarcento}, C.F.: {u:TUWZWL08K38R553U}, che rappresenta e difende ad ogni effetto di legge, in virtù di delega posta in calce al presente atto la </run>
<run nr="13" fmtId="5">SIG.RA {a-l:BANDINO} {a-f-f:GOFFREDA} </run>
<run nr="19" fmtId="4"> nata a {t:Verrès} il {d:24/11/1965} e residente in {t:Tarcento}, Via {ad:Khepri Santori, 84}, C.F.: {u:IEE QGR 53E69 Z778I}, ed elettivamente domiciliata presso lo studio del predetto avvocato in {t:Tarcento}, Via {ad:Giosia Iovine, 51}, con indicazione del n. fax al {n:6259/8989756}, PEC: {m:fk5172@tiscali.it},</run>
</par>
</pick>

```

Figura 7 - Estratto del documento DOCX in un formato intermedio XML con i dati pseudonimizzati

```

<p rend="j">La sottoscritta Avv. <persName type="fn">Gradita</persName> <persName type="s">Bertanzetti</persName> del foro di <placeName>Tarcento</placeName>, C.F.: <num>TUWZWL08K38R553U</num>, che rappresenta e difende ad ogni effetto di legge, in virtù di delega posta in calce al presente atto la <hi rend="b">SIG.RA <persName type="s">BANDINO</persName> <persName type="fn">GOFFREDA</persName> </hi> nata a <placeName>Verrès</placeName> il <date>24/11/1965</date> e residente in <placeName>Tarcento</placeName>, Via <address><addrLine>Khepri Santori, 84</addrLine></address>, C.F.: <num>IEE QGR 53E69 Z778I</num>, ed elettivamente domiciliata presso lo studio del predetto avvocato in <placeName>Tarcento</placeName>, Via <address><addrLine>Giosia Iovine, 51</addrLine></address>, con indicazione del n. fax al <num>6259/8989756</num>, PEC: <email>fk5172@tiscali.it</email>,</p>

```

Figura 8 - Estratto del documento TEI senza il riconoscimento delle abbreviazioni

```

<p rend="j">La sottoscritta <choice><abbr>Avv.</abbr><expan xml:lang="ita">avvocato</expan></choice> <persName type="fn">Gradita</persName> <persName type="s">Bertanzetti</persName> del foro di <placeName>Tarcento</placeName>, <choice><abbr>C.F.</abbr></choice> <expan xml:lang="ita">codice fiscale</expan></choice>: <num>TUWZWL08K38R553U</num>, che rappresenta e difende ad ogni effetto di legge, in virtù di delega posta in calce al presente atto la <hi rend="b"><choice><abbr>SIG.RA</abbr><expan xml:lang="ita">signora</expan></choice> <persName type="s">BANDINO</persName> <persName type="fn">GOFFREDA</persName> </hi>nata a <placeName>Verrès</placeName> il <date>24/11/1965</date> e residente in <placeName>Tarcento</placeName>, Via <address><addrLine>Khepri Santori, 84</addrLine></address>, <choice><abbr>C.F.</abbr><expan xml:lang="ita">codice fiscale</expan></choice>: <num>IEE QGR 53E69 Z778I</num>, ed elettivamente domiciliata presso lo studio del predetto avvocato in <placeName>Tarcento</placeName>, Via <address><addrLine>Giosia Iovine, 51</addrLine></address>, con indicazione del <choice><abbr>n.</abbr><expan xml:lang="ita">numero</expan></choice> fax al <num>6259/8989756</num>, <choice><abbr>PEC</abbr><expan xml:lang="ita">posta elettronica certificata</expan></choice>: <email>fk5172@tiscali.it</email>,</p>

```

Figura 9 - Estratto del documento TEI con il riconoscimento delle abbreviazioni

5. I REQUISITI DEL MOTORE DI RICERCA

Per soddisfare una serie di requisiti del progetto si è introdotto un particolare motore di ricerca (*Pythia*) nel flusso di lavoro che conduce dai documenti Word ai loro *output* pseudonimizzati e trasformati in TEI. Anche se l'obiettivo del contributo non è quello di illustrare in dettaglio *Pythia*, trattandosi di un prototipo ancora in via di sperimentazione, ci limitiamo qui ad accennare alla sua impostazione generale in funzione del progetto di ricerca qui trattato e rimandiamo alla bibliografia ([13]) e alla documentazione che accompagna il suo codice *open source* (github.com/vedph/pythia) per ulteriori approfondimenti.

I requisiti del motore destinato ad accogliere indici e metadati sono infatti piuttosto complessi: il primo è che si tratti di un motore capace di fornire concordanze, un attributo non scontato se confrontato al panorama tecnologico dei motori di ricerca testuale più diffusi in ambito informatico e nati con scopi diversi da quelli del progetto Atti Chiari (per esempio, individuare un documento in un *corpus*, oppure localizzare con precisione le occorrenze di ogni parola nel loro contesto). Il secondo requisito è quello di incorporare i metadati descritti nel contesto di un livello di astrazione più elevato, che consenta di trattare in modo omogeneo non solo le parole e i loro metadati, ma anche le strutture testuali più estese (come frasi, versi e strofe) con i loro eventuali metadati.

Simili strutture, naturalmente, molto spesso non sono affatto sovrapponibili, in quanto giacciono su livelli di analisi linguistica completamente distinti: per esempio, non sempre l'albero sintattico di un testo trova corrispondenze puntuali con la sua organizzazione metrica in versi o strofe, o con la sua disposizione colometrica a livello grafico.

6. IL MOTORE DI RICERCA

La necessità di delimitare alcune essenziali strutture (come la frase) determina ulteriormente l'evoluzione del sistema di pseudonimizzazione qui illustrato. La possibilità di incorporare i confini di frase in un indice, per quanto approssimativamente determinabili in base all'interpunzione, consente infatti ricerche contestuali più precise all'interno di un contesto sintatticamente definito, piuttosto che affidato al solo computo della distanza relativa.

L'individuazione dei confini di frase viene effettuata da uno dei numerosi filtri inseriti nella *pipeline* del sistema di indicizzazione, che opera per moduli. Nel caso specifico, trattandosi di input in formato TEI, su un generico algoritmo di *sentence splitting* viene innestato un approccio configurabile che considera anche la natura di determinati marcatori: ad esempio, un marcatore come *head*, associato all'intestazione, viene considerato come corrispondente a una frase, anche se il suo testo manca dei consueti indicatori come la punteggiatura. Il sistema può così disporre di un modulo di *sentence splitting* che si avvale di ulteriori informazioni fornite dalla marcatura XML (TEI o meno, dato che è parametrizzabile), accanto a uno che prende in considerazione solo il testo, adatto ad esempio a input *plain text*; la *pipeline* di indicizzazione viene poi configurata, come per ogni altro suo aspetto, inserendo l'uno o l'altro modulo a seconda dei documenti trattati. Questo approccio modulare è proprio dell'intero sistema di indicizzazione, utilizzando una serie di componenti destinati a

estrarre i testi da una fonte (che non necessariamente è un *file system*), filtrarli in vario modo per prepararli all'analisi, estrarne metadati, calcolarne, secondo vari, algoritmi data e chiave di ordinamento, *tokenizzare* e filtrare i *token*, e individuare una serie di strutture testuali (frasi, versi, strofe, ecc.), in qualsiasi numero e di qualsiasi genere, anche quando esse si sovrappongano.

Ulteriori componenti configurabili riguardano poi la mappatura dell'articolazione interna del testo (ad esempio divisioni in sezioni, paragrafi, ecc.), in modo tale da fornire una mappa di navigazione del testo interattiva nel *frontend* del sistema e l'estrazione di porzioni di testo da presentare come contesto semanticamente congruo (basato su questa medesima mappa), e la trasformazione del formato originale del testo in un formato destinato alla sua presentazione, tipicamente HTML e CSS. Nel caso dei testi TEI qui trattati, il modulo di trasformazione utilizza semplicemente uno script XSLT fornitogli tra i suoi parametri operativi.

In questo ambito, la peculiare natura dei testi trattati ha determinato un'ulteriore evoluzione del sistema di pseudonimizzazione destinata a individuare in modo automatico (sulla base di un elenco e su una rosa di variazioni formali trattate in modo algoritmico) le numerosissime abbreviazioni, che non sarebbe economico affidare alla marcatura manuale. Infatti, poiché l'individuazione delle strutture relative alle frasi si basa essenzialmente sull'interpunzione (anche se non esclusivamente, come nel caso dei documenti dotati di una marcatura in grado di implicare i confini sintattici), la massiccia presenza di abbreviazioni contenenti punti costituirebbe una rilevante fonte di errore. In considerazione di ciò, oltre che a vantaggio della chiarezza del testo per un pubblico non necessariamente specialista in ambito giuridico, si è allora scelto di affidare al sistema di pseudonimizzazione anche il compito di marcare automaticamente le abbreviazioni in una fase distinta e successiva del suo operato. Si tratta quindi di un ulteriore esempio di come la natura stessa di questo sistema sia modellata sulle esigenze del suo uso in sede di analisi, anzitutto linguistica ma anche di altra specie.

Si è infatti visto che un ruolo essenziale in questa analisi linguistica rivestono i metadati e spesso anche le strutture testuali: metadati relativi a informazioni linguistiche (ad es. un latinismo, una classificazione morfologica, un'abbreviazione, un antroponimo, un toponimo, un numero, ecc.), paratestuali (ad es. una parola in grassetto o in corsivo), e sintattiche (qui essenzialmente le strutture rappresentate dalle frasi). In questo ambito, il motore deve poter ricercare allo stesso modo qualsiasi entità estratta dal testo con i suoi metadati, sicché l'approccio adottato consiste nell'elevare il livello di astrazione: un testo non viene più trattato come una sequenza di caratteri all'interno dei quali individuare delle sequenze (*token*) variamente filtrate e indicizzate per essere ricercabili; piuttosto, tale sequenza viene in certo modo smaterializzata per produrre un semplice insieme di oggetti. Ogni oggetto dell'insieme può essere dotato di un qualsiasi numero di metadati appartenenti a un elenco aperto, fra cui anche la posizione nel documento di origine. A questo punto, la ricerca consiste solo nell'individuare gli oggetti di proprio interesse attraverso questi metadati, per poi presentarli nel loro contesto originale. Tali oggetti non sono quindi più solo 'parole', ma anche un qualsiasi tipo di struttura testuale estratta dal testo, la cui posizione viene definita con due punti (primo e ultimo *token*) anziché uno solo (come nel caso di una singola 'parola'). Inoltre, una serie di operatori consente non solo di operare un confronto molto articolato fra il valore ricercato e quello indicizzato, ma anche di rappresentare indicazioni posizionali. Assimilando un oggetto con una singola posizione a un punto (per esempio una 'parola'), e uno con due posizioni a un segmento, questi operatori consentono di trovare un elemento dentro l'altro, o parzialmente sovrapposto a un altro, o alla testa o alla coda di un altro, e così via: è il caso della ricerca di una parola a inizio di frase o a fine di verso, o a fine di frase e di verso, o di una frase parzialmente sovrapposta a un verso o strofe, ecc. Qualsiasi tipo di elemento, derivi esso da una parola o da una struttura, non è che un oggetto con dei metadati: il motore interroga i metadati per giungere agli oggetti e ai loro rapporti, poi li localizza nel testo di origine, e li presenta opportunamente trasformati all'utente finale.

In questo ambito, il sistema offre l'ulteriore vantaggio di fornire un ambiente di lettura dei testi completo, sia in funzione del testo trovato sia in base alle esigenze dell'utente, che dispone anche di una mappa navigabile automaticamente generata per ogni documento. Tutto questo inoltre opera all'interno di un insieme di tecnologie standard e di uso universale: l'indice non è che un database relazionale, facilmente integrabile in qualsiasi progetto e consultabile in vario modo anche al di là del motore di ricerca; inoltre, tutto il processo che conduce dal documento nel suo formato di *input*, quale esso sia (TEI nel nostro esempio), e ovunque sia contenuto (*file system*, *cloud storage*, *web*, *database*, ecc.), è configurato in una *pipeline* componibile, dove ogni stadio viene configurato da una serie di parametri, all'interno di un semplice file JSON di configurazione. Il sistema può dunque arricchirsi di nuove funzionalità semplicemente introducendo nuovi moduli in questa *pipeline*: ad esempio, per introdurre il dettagliato esito di analisi fonologiche o metriche automatiche in seno all'indice, o quello dell'analisi prodotta da sistemi esterni di *POS tagging*, ecc. Nel caso qui esemplificato dunque, l'uso di questo motore può risultare particolarmente vantaggioso proprio in ragione delle peculiarità dei testi trattati e delle soluzioni adottate, a cominciare dal sistema di pseudonimizzazione da cui questo intervento ha avuto principio, modellandosi in funzione dei suoi obiettivi.

7. CONCLUSIONI

Il processo di pseudonimizzazione adottato nel progetto PRIN Atti Chiari e illustrato in questo contributo coniuga l'esigenza di tutelare la riservatezza e la necessità di disporre di testi formalmente completi che possano consentire l'analisi linguistica e l'individuazione di esempi di scrittura forense chiara ed efficace.

L'approccio qui adottato è funzionale, in primo luogo, ad assicurare una completa e non reversibile anonimizzazione dei dati, che non è solo un ovvio requisito legale, ma rappresenta un aspetto fondamentale per ottenere la fiducia di chi contribuisce alla costituzione del *corpus* di atti. In secondo luogo, esso serve a operare sul testo uno o più tipi di trasformazioni, in rapporto agli scopi della procedura, che pur garantendo questo primo requisito preservino la leggibilità del testo e la sua usabilità per analisi di ampio spettro. In queste convergono non solo le annotazioni e i metadati di ogni documento inseriti dagli anonimizzatori, ma anche ulteriori informazioni provenienti dalla conversione del formato del testo da DOCX a TEI (come, ad esempio, gli stili tipografici), o aggiunte da processi supplementari (come lo scioglimento delle abbreviazioni o il *tagger* di terza parte). A sua volta, questo richiede un sistema di ricerca capace di ingerire un insieme aperto di annotazioni, estese sulle parole come su altre strutture linguistiche (ad es. la frase), e di fornire una ricerca per concordanze che integri sul medesimo livello tutte queste fonti di dato, finendo così per mettere in campo un insieme di strumenti il cui valore pratico e metodologico può superare i confini del singolo progetto di ricerca.

BIBLIOGRAFIA

- [1] Candrilli, Fernanda. 2021. «Il progetto di archiviazione e anonimizzazione». In *Atti Chiari. Chiarezza e concisione nella scrittura forense*, a cura di Riccardo Gualdo e Laura Clemenzi, 19–29. Viterbo: Sette Città.
- [2] Caponi, Remo. 2014. «Il processo civile telematico tra scrittura e oralità.» In *Lingua e processo. Le parole del diritto di fronte al giudice, Atti del Convegno*, 176–86. Firenze: Firenze: Accademia della Crusca.
- [3] Cavallone, Bruno. 2010. «Un idioma coriaceo: l'italiano del processo civile.» In *L'italiano giuridico che cambia, Atti del Convegno*, 85–95. Firenze: Firenze: Accademia della Crusca.
- [4] Clemenzi, Laura. 2021. «L'interrogazione della base dati Atti Chiari.» In *Atti Chiari. Chiarezza e concisione nella scrittura forense*, a cura di Riccardo Gualdo e Laura Clemenzi, 41–52. Viterbo: Sette Città.
- [5] Conte, Giuseppe. 2013. «Il linguaggio della difesa civile.» In *Lingua e diritto. Scritto e parlato nelle professioni legali*, Alarico Mariani Marini e Federigo Bambi, 35–67. Pisa: Pisa University Press.
- [6] Cresti, Emanuela, e Alessandro Panunzi. 2013. *Introduzione ai corpora dell'italiano*. Bologna: Il Mulino.
- [7] Dalianis, Hercules. 2019. «Pseudonymisation of Swedish Electronic Patient Records Using a Rule-Based Approach». In *Proceedings of the Workshop on NLP and Pseudonymisation, a cura di Lars Ahrenberg e Beáta Megyesi*, 16–23. Turku: Linköping Electronic Press.
- [8] Dell'Anna, Maria Vittoria. 2014. «Fra attori e convenuti. Lingua dell'avvocato e lingua del giudice nel processo civile.» In *Lingua e processo. Le parole del diritto di fronte al giudice, Atti del Convegno a cura di Federigo Bambi*, 83–101. Firenze: Accademia della Crusca.
- [9] Douglass, Margaret, et al. 2004. «Computer-Assisted De-Identification of Free Text in the MIMIC II Database». *Computers in Cardiology* 31: 341–44.
- [10] Elger, Bernice S., e et al. 2010. «Strategies for health data exchange for secondary, cross-institutional clinical research». *Computer Methods and Programs in Biomedicine* 99 (3): 230–51.
- [11] Freddi, Maria. 2019. *Linguistica dei corpora*. Roma: Carocci.
- [12] Fusco, Francesca. 2021. «Marcatura linguistica e tutela della riservatezza nello studio di un corpus di scritture forensi». In *Atti Chiari. Chiarezza e concisione nella scrittura forense*, a cura di Riccardo Gualdo e Laura Clemenzi, 29–40. Viterbo: Sette Città.
- [13] Fusi, Daniele. 2020. «Text Searching Beyond the Text: a Case Study». *Rationes Rerum* 15: 199–230.
- [14] ———. 2021. «Digitalizzazione e marcatura XML degli atti». In *Atti Chiari. Chiarezza e concisione nella scrittura forense*, a cura di Riccardo Gualdo e Laura Clemenzi, 59–73. Viterbo: Sette Città.
- [15] Gualdo, Riccardo e Laura Clemenzi (a cura di). *Atti Chiari. Chiarezza e concisione nella scrittura forense*. Viterbo: Sette Città, 2021.
- [16] Gualdo, Riccardo, e Maria Vittoria Dell'Anna. 2014. «Per prove e per indizi (testuali). La prosa forense dell'avvocato e il linguaggio giuridico.» In *La lingua variabile nei testi letterari, artistici e funzionali contemporanei. Analisi, interpretazione, traduzione, Atti del XIII Congresso SILFI. A cura di Giovanni Ruffino e Marina Castiglione*, 623–35. Firenze: Cesati.
- [17] Lombardi, Giulia. 2021. «I vantaggi del programma an-tool.» In *Atti Chiari. Chiarezza e concisione nella scrittura forense*, a cura di Riccardo Gualdo e Laura Clemenzi, 29–40. Viterbo: Sette Città.
- [18] Mortara Garavelli, Bice. 2003a. «L'oratoria forense: tradizione e regole.» In *L'avvocato e il processo. Le tecniche della difesa*, a cura di Alarico Mariani Marini e Maurizio Paganelli, 66–91. Milano: Giuffrè.
- [19] ———. 2003b. «Strutture testuali e stereotipi nel linguaggio forense.» In *La lingua, la legge, la professione forense*, a cura di Alarico Mariani Marini, 3–19. Milano: Giuffrè.

- [20] Noumeir, Rita. 2007. «Pseudonymization of Radiology Data for Research Purposes». *Journal of Digital Imaging* 20 (3): 284–95.
- [21] Oksanen, Arttu, et al. 2019. «A Pseudonymization Service for Finnish Court Documents». In *Legal Knowledge and Information Systems. JURIX 2019: The Thirty-second Annual Conference, a cura di Michał Araszkiewicz e Víctor Rodríguez-Doncel*, 251–54. Amsterdam: IOS Press.
- [22] Sabatini, Francesco. 2015. «Dalla lingua comune al linguaggio del legislatore e dell'avvocato». In *L'avvocato e il processo. Le tecniche della difesa, a cura di Alarico Mariani Marini e Maurizio Paganelli*, 3–14. Milano: Giuffrè.

RePIM in LOD: semantic technologies to preserve knowledge about Italian secular music and lyric poetry from the 16th-17th centuries

Paolo Bonora¹ and Angelo Pompilio²

¹Department of Classical Philology and Italian Studies, University of Bologna, Italy – paolo.bonora@unibo.it

²Department of Cultural Heritage, University of Bologna, Italy – angelo.pompilio@unibo.it

ABSTRACT

The *RePIM in LOD* project aimed to publish the “*Repertorio della Poesia Italiana in Musica, 1500-1700*” (RePIM) as Linked Open Data (LOD) dataset. For the extent and detail of its contents, RePIM is a reference archive for research on Italian secular music from the 16th-17th centuries. In recent years, scholars have been able to access it through a public web-based application. Due to the obsolescence of its information technology platform, the RePIM repository was set to be taken offline. To preserve this precious source, the project migrated its contents into a knowledge base (KB) adopting semantic technologies and designed an up-to-date end-user application. The paper illustrates this process as a use case of digital knowledge preservation of bibliographic and philological information in the field of Italian secular music and lyric poetry of the 16th-17th centuries.

KEYWORDS

Linked Open Data, Semantic Web, CIDOC-CRM, FRBRoo, digital knowledge preservation, Italian poetry.

TALK

1. INTRODUCTION

The “*Repertorio della Poesia Italiana in Musica, 1500-1700*” (RePIM), is the digital version of the incipit index of Italian poetry in music, a project conceived at the end of the 70s by Lorenzo Bianconi in collaboration with Angelo Pompilio and Antonio Vassalli. Since 1990, materials collected on paper cards during the 70s-80s, as well as bibliographic descriptions of poetic and musical sources, have been revised and widely integrated. As a rule, musical sources of secular music do not declare authorship of the intoned literary text. To identify the poet, it is necessary to collate the musical sources with coeval poetry books. The RePIM incipit index was designed to respond to this basic scientific need: to identify the greatest possible number of authors of poetic texts set to music in secular and spiritual compositions of the 16th-17th centuries through the systematic examination of poetry books. A minor subset of the archive was published online during in 2006 but it has never been fully published in any other analogic or digital form. Unfortunately, the technological platform has become obsolete since then and now must be discontinued. The *RePIM in LOD* project identified the publication as a Linked Open Data (LOD) dataset as the way to preserve it as a free and unlimited resource for the community of present and future scholars. Contents have been migrated in Resource Description Format (RDF) adopting two reference ontologies such as FRBR Object Oriented (now Library Reference Model) and CIDOC Conceptual Reference Model. The result is a KB relying on a set of open and well-established formal ontologies and then fully interoperable within the Linked Data ecosystem. Finally, to develop a new web-based application, the ResearchSpace platform ([5]) has been selected as an open source, cultural heritage oriented and scholar friendly tool which also aims to address digital knowledge preservation and maintainability over time.

2. FALLING IN LOD

RePIM starts from the consideration that, in the secular vocal music of the 16th-17th centuries, for the purposes of historical and critical understanding of the work, the lyric component has a structural and artistic equal to the musical one. Thanks to the systematic collection of textual incipits from both musical and poetry books from Italian and international libraries, about 30% of the musical compositions of the collected works have been identified.

The conceptual model of the repository has been designed adopting the Functional Requirements for Bibliographic Records (FRBR) guidelines. The consequent distinction between Work and Manifestation levels enables bibliographic,

musicological and literary analytical search. This gives scholars the opportunity to explore literary and musical production from a single perspective.

The repository is structured into three main sections: the repertoire of works, the bibliographic sources and the incipit index. The repertoire consists of the authority file of names, literary and musical works and the links identified between them. Musical and literary sources are described in an analytical form both in terms of bibliographic description and contents. Textual incipit, complete text (diplomatic and modern edition), title, dedication, caption, voices and articulation in parts or sections are available. The sources are also linked to their reprints or newer editions. In addition, bibliographic references used during the study of sources are recorded in the KB together with a complete authority file of owning libraries and a repository of digital copies, if available.

The current consistency is: a bibliography of literary sources of 1.500 titles; the incipit of over 66.000 musical compositions; a complete bibliography of musical books of secular and spiritual music from the 16th-17th centuries (ca. 3.500 items), references to bibliographic musical repertoires (RISM A/I e B/I; Vogel-Einstein 1962 and Nuovo Vogel by Lesure and Sartori, 1977); an authority file with ca. 1.400 musicians and 3.200 poets; digital reproduction of approx. half of the recorded musical sources and literary books and full text transcriptions of more than 5.000 poetry works are also available.

For literary sources, the record contains: the complete transcription of the title page; the essential data about the dedicatory letter; the list of authors; the incipit (first two lines, or more) of the composition; any further information of a musical and literary nature such as: the complete transcription, incipits of texts marked as “*per musica*”, references and citations relating to musical occasions, etc.

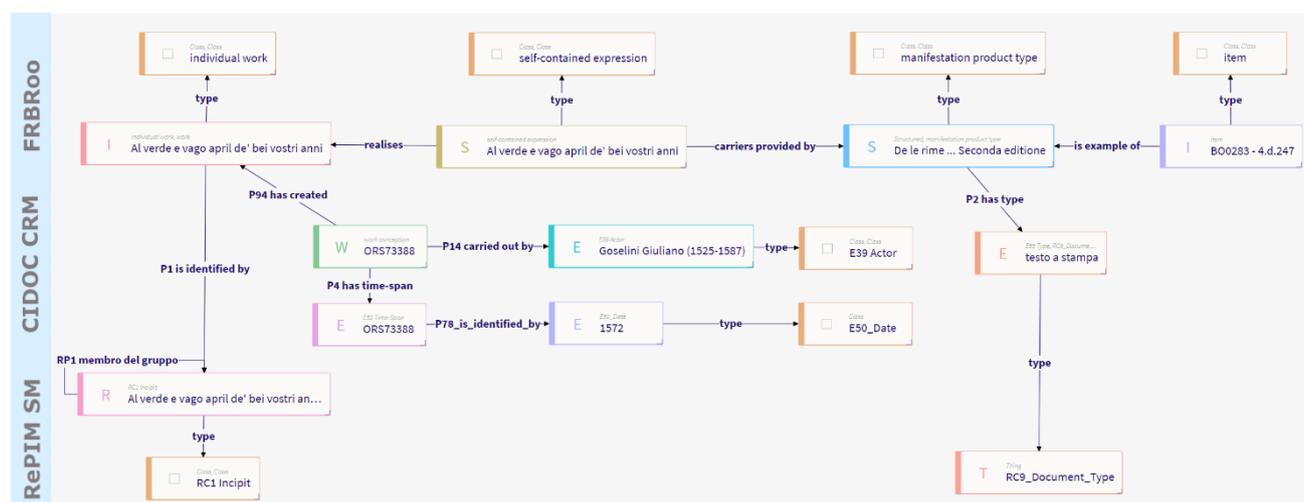


Figure 1. Part of the graph representing the work “Al verde e vago April de’bei vostri anni” in “Rime” by Paolo Gottardo Pontio, 1572.

In order to represent this structured knowledge, we adopted a set of formal ontologies integrated within the RePIM Semantic Model (RePIM SM)¹. This introduces some domain specific concepts (such as the “incipit” as the key to identify works) and several specializations of properties from those defined by the two reference ontologies where a more specific semantics was required. Thus, the role of the RePIM SM is to extend the CIDOC CRM and FRBRoo to better represent domain-specific concepts while maintaining full alignment with them. This makes the dataset fully interoperable as data can be accessed through the standard CIDOC CRM or through the specialization proposed by the RePIM SM only when required. To express information about the document segmentation and bibliographic references, we adopted some other specialized ontologies such as DoCO² and BiRO³. The model relies on FRBRoo’s concepts of work, expression and manifestation to describe musical and poetry works, their contents and books (fig.1). Starting from this core assertions, the other ontologies assert information about authors, primary and secondary sources, references, libraries and some domain specific topics as genres, voices, settings, etc.

The resulting KB’s conceptual model consists of 30 classes (12 from CIDOC CRM, 11 from FRBRoo and 4 from DoCO, 2 from Repim SM and 1 from BiRO) and 49 properties (17 from CIDOC CRM, 5 from FRBRoo, 19 from Repim SM and others from DoCO, BiRO, DCTerms and RDFs).

¹ The RePIM SM is maintained at: <https://github.com/paolobonora/RePIM-LOD/>.

² DoCO: Document Components Ontology ([3]), <http://www.sparontologies.net/ontologies/doco>.

³ BiRO: Bibliographic Reference Ontology ([4]), <http://www.sparontologies.net/ontologies/biro>.

Data were extracted from the legacy database management system with a Relational-to-RDF mapping tool following the semantic alignment between the FRBR-ER inspired proprietary relational model and the RePIM SM ([1]). The resulting knowledge graph contains 778.699 class instances and 5.324.973 property assertions (more than 5.5 million triples in total). We expect that the publication of RePIM's contents as a LOD resource through Open Science platforms as Zenodo⁴ would lower the barriers of data access, facilitating both the reuse and extension of the KB by a wider community. We also expect that this should increase the resilience against the technological unavoidable obsolescence of any information system leveraging dissemination of knowledge among different projects and players.

To reach the wider community of users not directly interested in raw data reuse, we also developed a new web-based application that exploits potential of the RDF dataset. The ResearchSpace⁵ platform has been selected to implement basic functionalities plus a set of requirements expressed over the years by the community of users but never fully implemented. The new application leverages both semantic and NLP technologies and introduces textual search and indexing features such as: full-text search, PoS and lemma based searching capabilities; graph-based browsing; NER (Named Entity Recognition) of names, toponyms, etc.; faceted browsing of topics⁶.

The design of the web application followed a strictly functional approach trying to achieve an effective representation of knowledge to the end-user. This means reducing the complexity of the conceptual model with views that increase data readability. Although the ResearchSpace's templating mechanism allows to arrange an effective representation of resources as user's views, it does not support their explicit formalization within the semantic model as the introduction of a dedicated semantic layer would allow to ([2]). Besides the framework supports basic mechanism for expert-led refinements and expansions of the knowledge graph. This will be leveraged to introduce edit and annotation features into the web application.

3. FURTHER DEVELOPMENTS

The current implementation aims to support basic search and browsing functionalities of the current knowledge expressed within the KB. The next phase should aim to let user annotate further findings achieved through the exploitation of already available sources or newly added ones through the contribution by the community of users. The ResearchSpace platform supports the implementation of data editing procedures as well as creation of semantic narratives and visual representation of resources from the knowledge graph. This would lead to a collaborative approach to knowledge extension and maintenance. This will require a fine-grained information provenance tracing and a sharp profiling of contributors. Besides, the outcome would be a shared effort of knowledge preservation and extension among the wider community of users through the years to come. We believe that publishing the RePIM archive as LOD will preserve it and offers, both to music scholars and those interested in Renaissance and Baroque poetry, a useful tool and a valuable source of information, resulting from over 40 years of research, which otherwise would be lost in the digital mist.

4. ACKNOWLEDGEMENTS

The *RePIM in LOD* project was founded by I Tatti – The Harvard University Center For Italian Renaissance Studies with a Digital Humanities Fellowship grant.

REFERENCES

- [1] Bizer, Christian, and Andy Seaborne. 2004. 'D2RQ-Treating Non-RDF Databases as Virtual RDF Graphs'. In *Proceedings of the 3rd International Semantic Web Conference (ISWC2004)*.
- [2] Bonora, Paolo, and Angelo Pompilio. 2021. 'Corago in LOD. The Debut of an Opera Repository into the Linked Data Arena'. *JLIS.It* 12 (2): 54–72.
- [3] Constantin, Alexandru, Silvio Peroni, Steve Pettifer, and et al. 2016. 'The Document Components Ontology (DoCO)'. *Semantic Web* 7 (2): 167–81.
- [4] Di Iorio, Angelo, Andrea Giovanni Nuzzolese, Silvio Peroni, and et al. 2014. 'Describing Bibliographic References in RDF'. In *CEUR Workshop Proceedings*, 1155.
- [5] Oldman, Dominic, and Tanase Diana. 2018. 'Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace'. In *The Semantic Web – ISWC 2018, a Cura Di Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, e Elena Simperl*, 325–40. Cham: Springer International Publishing.

⁴ The dataset is available through the DOI: 10.5281/zenodo.5692109.

⁵ <https://researchspace.org/>.

⁶ The provisional URL for the web application is: <https://repim.itatti.harvard.edu/>.

Sessione Filosofia 2
Giulio Cesare Vanini

Computare o comporre?

Riflessioni sul rapporto tra poesia e digitalità alla luce di alcune considerazioni bachelardiane

Ivana Brigida D'Avanzo

Università degli Studi di Perugia, Italia – ivanabdavanzo@gmail.com

ABSTRACT

Il contributo propone una discussione circa il rapporto tra poesia e digitale con l'intento di mostrare come l'incalzante avanzamento tecnologico ci ponga di fronte a un nuovo tipo di fatto estetico, spingendoci a riconsiderare anche le più tradizionali categorie del pensiero storico-filosofico. Attraverso il pensiero estetico del filosofo francese Gaston Bachelard è possibile individuare un plesso di confronto tra computazione e composizione che può essere funzionale al dialogo tra due universi apparentemente non comunicanti.

PAROLE CHIAVE

Poesia, digitale, Bachelard.

INTERVENTO

Il presente contributo intende proporre alcune considerazioni, a partire dalle quali scaturiscono o possono scaturire non pochi interrogativi, circa il rapporto che lega la cultura umanistica, in particolare nella sua declinazione poetica, e il digitale come dimensione ormai e sempre più avvolgente e coinvolgente ogni aspetto delle nostre quotidianità.

I

L'espedito scelto come punto d'avvio di tali considerazioni è il seguente esperimento mentale: dati due componimenti poetici, null'altro che due testi in versi, sprovvisti di qualsivoglia descrizione o specifica, è possibile riconoscere quale di essi sia stato composto da un uomo e quale invece da un software?

Eccone un esempio:

A home transformed by the lightning
the balanced alcoves smother
this insatiable earth of a planet, Earth.
They attacked it with mechanical horns
because they love you, love, in fire and wind.
You say, what is the time waiting for in its spring?
I tell you it is waiting for your branch that flows,
because you are a sweet-smelling diamond architecture
that does not know why it grows ([13]: 30-31)¹.

O cercle magique: oeil de tuot être!
Oeil de volcan injecté des sangs malsains
Oeil de ce lotus noir
Surgi des calmas du songe ([12] : 45).

Sospendendo momentaneamente il giudizio su quelle più immediate considerazioni di carattere tecnico, come la forma e la metrica, o estetico, o ancora inerenti al messaggio che si è inteso trasmettere, la domanda che più ci urge porre è se, a partire da una lettura alla cieca, una sorta di *blind review*, di due poesie come le suddette, sia plausibile congetturare che si tratti di una composizione artificiale, e se sì, quali siano le prove a carico di tale congettura. In altri termini, dove

¹ Disponibile sul sito https://issuu.com/dukeupb/docs/thearchive_fall2011.

concretamente andremmo a cercare le differenze tra una poesia umana ed una artificiale, sempre posto che tali differenze esistano? In definitiva, come ci aspettiamo che sia una poesia scritta da un computer?

Queste domande, ad oggi, non dovrebbero sembrarci affatto insensate. La digitalità ha ormai penetrato a fondo ogni parte della nostra vita, a quanto pare persino quegli aspetti – come la poesia appunto – che ci siamo a lungo illusi di poter preservare dalla contaminazione digitale, come le ultime roccaforti dell'umanità e dell'umanità soltanto. A riprova di ciò, vale la pena menzionare il seguente episodio.

Nel 2011 – e dieci anni sono un tempo estremamente lungo nell'era digitale e dei suoi velocissimi progressi – uno studente della Duke University, Zackary Scholl, ha partecipato ad una iniziativa della prestigiosa rivista della sua università, «The Archive», tesa ad incoraggiare la scrittura di poesia. Il componimento proposto e accettato dagli editori della rivista è esattamente la prima delle due poesie riportate qui sopra, *For the Bristlecone Snag*. Quattro anni dopo Scholl scopre le sue carte e scrive sul suo blog: “my poetry generator passed the Turing Test [...]. The real Turing Test of this AI was to get it accepted to a literary journal, which was accomplished – this poetry was successfully accepted into a literary journal at a prestigious university” ([14])². Scholl continua spiegando di aver ideato un programma che utilizzava un sistema grammaticale privo di contesto per generare poesie: l'operatore suggerisce al programma il tema da svolgere, una sorta di key-word, in associazione tematica con la quale il programma sceglie parole affini distribuite in varie liste di verbi, aggettivi e sostantivi, che andranno a comporre il corpo del testo. Le parole, classificate come positive, negative o neutre, sono combinate casualmente ma seguendo regole prefissate di sintassi e di distribuzione di elementi positivi e negativi. Anche la metrica e la suddivisione in strofe sono casuali. Aggiungiamo che il caso di Scholl non è isolato; nell'oramai incalcolabile catalogo dell'AI ci sono ormai innumerevoli esempi di programmi generatori di poesie a campionatura casuale sempre più avanzati e di sofisticatissimi algoritmi di *deep learnig*³.

Ma sorge spontanea la domanda: è così che ci immaginiamo un poeta al lavoro? La commissione della rivista incaricata di selezionare le poesie inviate avrà pur individuato qualcosa di dignitosamente poetico, per scegliere questa a discapito di altre, eppure non possiamo fare a meno di chiederci se il verdetto fosse rimasto invariato qualora l'origine della poesia fosse stata nota in fase di valutazione. Facciamo forse *ancora* un po' di difficoltà – o almeno così supponiamo – ad immaginare un poeta che, a mo' di generatore di poesie di Scholl, compone versi consultando liste di parole combinate in modo random, tenendo presente tuttalpiù un po' di sintassi. Un software non distingue un *fare* artistico da una qualunque altra forma di attività umana, *fa* arte così come *fa* medicina: calcolando, processando dati, organizzando output. Sia una combinazione di verbi e aggettivi in un generatore di poesie o una combinazione di sintomi in un generatore di diagnosi, non fa differenza. Ma noi ci aspettiamo che un software segua una logica tanto quanto che un poeta la disattenda⁴. Che ne sarebbe altrimenti delle grandi intuizioni estetiche e delle divine ispirazioni artistiche?

II

Non vi è alcun dubbio che una differenza tra un poeta che “crea arte” e un software che genera un codice costituito da una sequenza di 0 e di 1, non solo esiste ma è radicale. E sebbene si possa parlare a lungo di tale differenza, è un altro il nodo problematico che si intende qui mettere in discussione. Per tentare non dico di scioglierlo, ma per lo meno di individuarlo, mi servirò della particolare interpretazione che Gaston Bachelard fa della “lettura dei poeti [che] è essenzialmente *rêverie*” ([1]: 45), in quanto mi sembra che apra ad una dimensione problematica di non poco conto, soprattutto in relazione allo scarto tra poesia umana e poesia artificiale, e indirettamente e più in generale, alle implicazioni antropologiche legate alla relazione tra digitale e poesia.

Lo spunto viene, ancora una volta, dall'episodio di Scholl e della sua poesia artificiale; in particolare vorrei portare l'attenzione sull'inganno inconsapevolmente vissuto dagli editor o da quanti, leggendo quella poesia sulla rivista, *hanno creduto di vivere un'esperienza estetica*. Lo è stata *davvero*? È questa la domanda che ci preme discutere. Chiunque abbia letto quella poesia e, beatamente inconsapevole della sua genesi artificiale, abbia goduto sinceramente di quelle immagini, si può dire abbia vissuto un'esperienza estetica diversa o inferiore di chi abbia ad esempio letto, provando il medesimo pathos, un sonetto di Shakespeare? Se accogliamo l'uso bachelardiano di leggere poesie al fine di trovare felicità, quiete e *bonheur* in quelle immagini poetiche, al fine di ripristinare l'unione tra uomo e cosmo, in una *felice appartenenza* in cui “una sola immagine invade tutto l'universo, alimentando l'immensa felicità che proviamo all'idea di abitare il mondo di questa immagine” ([4]: 179); se accogliamo tutto questo, mentre lo accogliamo, in che misura ha incidenza interrogarsi

² <https://zackaryscholl.com/other/poetry/>.

³ Per una ancora valida analisi delle potenzialità (e i rischi) dell'intelligenza artificiale, si veda ([7]: 29).

⁴ È la tesi che vede nel messaggio estetico un meccanismo teso a disattendere il sistema di aspettative condivise dalle culture e sottoculture sociali. Su ciò, cfr. ([11: 61-82]).

sulla composizione e sulla genesi di queste immagini poetiche? Prima di tentare una risposta a queste domande, o meglio, prima di continuare a ragionarci su, vorrei approfondire la questione della lettura bachelardiana per giustificarne l'impiego. Si assuma come punto di partenza del discorso la tesi, portata avanti dal Bachelard a partire dalla trilogia degli elementi fino – e sempre più – alle poetiche della maturità, secondo cui la felicità sta nelle immagini⁵. Ma in quali immagini? Certamente le immagini che prendono forma davanti al focolare o alla fiamma di una candela, le immagini acquatiche dei ruscelli delle valli della Champagne tanto care al giovane Bachelard. Ma non solo. In quali immagini trova felicità il Bachelard professore alla Sorbonne che non ha più boschi, campagne e fiumi lungo i quali sognare? E dove possiamo trovarla noi, ancora non del tutto liberi dalle costrizioni di questa disastrosa crisi pandemica?

In fondo “noi non siamo che lettori” ([2]: 11), ci rassicura Bachelard, che ci suggerisce dunque di trovare nelle immagini dei libri, della letteratura, della poesia, la nostra felicità. È evidente l'accento che Bachelard pone sulla scrittura come dimensione propria delle sue *rêverie*. “La *rêverie* che ci proponiamo di studiare – scriverà Bachelard nel '60 – è la *rêverie* poetica [...], una *rêverie* che si scrive, o almeno ci si ripromette di scrivere, facendola emergere da quel grande universo che è la pagina bianca” ([4]: 12). E poco dopo, “a differenza di un sogno, una *rêverie* non si racconta. Per comunicarla, bisogna *scriverla*, trasmettendo emozioni, rivivendola nel momento in cui la si trascrive” ([4]: 14).

“Le felicità consiste allora – come spiega chiaramente Sertoli – nel riattivare, sulle pagine di un libro, quella *rêverie* naturale che in esse si è cristallizzata. Felicità è *leggere* le immagini scritte” ([15]: 291). Ma affinché tale felice meccanismo di riattivazione delle immagini si verifichi si rende necessario un particolare tipo di lettura, promosso da Bachelard come esemplare: un tipo di lettura che, lungi dal razionalizzare con i concetti le immagini, *violentandole* e riducendole a segni di qualcos'altro, renda ragione del fatto che non c'è forma di «comprensione» possibile dell'immagine se non attraverso l'atto immaginativo stesso. Un'immagine va *semplicemente* immaginata, non capita – d'altronde “in poesia, il non-sapere è una condizione primaria” ([1]: 23) – essa va vivificata attraverso la partecipazione e l'identificazione, non ridotta a «concetto poetico».

Si tratta, in buona sostanza, di mettere in pratica quella lettura fenomenologica – definitivamente esplicitata a partire da *La poétique de l'espace* – che si rivolge all'immagine colta nella sua presenza immediata, all'immagine per l'immagine, senza avanzare alcuna pretesa interpretativa o archeologica. “L'immagine – infatti – nella sua semplicità, non ha bisogno di un sapere, essa è la ricchezza di una coscienza ingenua”, ([1]: 9), ricchezza che appartiene tanto all'autore quanto al lettore, al punto tale – si ponga attenzione su questo particolare aspetto – da non poter quasi distinguere dove finisca l'uno e dove cominci l'altro. La dinamica poetica autore-lettore supposta da Bachelard dovrebbe funzionare più o meno in questo modo: un'immagine poetica, prodotto di un atto di coscienza dell'autore, si offre al lettore che, per poterne cogliere appieno il valore, per poter godere della felicità insita in quell'immagine, deve trasformarla a sua volta in un proprio atto di coscienza, diventando lui stesso autore di quell'immagine. Così il lettore trasforma quell'immagine in qualcosa d'altro, in qualcosa di *proprio*, e più vi partecipa più diviene l'autore di quell'immagine, di un'immagine che è insieme uguale e diversa, vecchia e nuova⁶. D'altronde “un'immagine letteraria dice quello che non sarà mai immaginato due volte” ([1]: 9). In tal senso, leggere poesie e immagini, non significa affatto limitarsi a subire l'arte di qualcun altro, sperimentare un meccanismo passivo e immobile, la cui controparte attiva si identifichi invece con la composizione e la scrittura di tali immagini. Il modello bachelardiano di lettura, ponte tra la *conscience rêvante* e la *rêverie* stessa, assume la natura del proprio oggetto e diviene essa stessa poesia. “La lettura dei poeti è essenzialmente *rêverie*” ([1]: 9), Bachelard non ha dubbi, anzi sostiene che “la lettura dei poeti dovrebbe essere un'attività telepoetica” ([5]: 127) per il suo indurre, attraverso la lettura di poesie, alla creazione di nuove poesie. D'altronde fa questo un vero poeta: “trasmette il germe del sogno senza fine” ([3]: 149). E persino il più apparentemente meccanico degli approcci alla poesia, la trascrizione di poesie altrui, è essa stessa poesia, *rêverie*. “Come il Pierre Menard di Borges – rileva con sagacia Sertoli – che volendo riscrivere parola per parola il Don Chisciotte, pur riuscendovi scrisse qualcosa di «quasi infinitamente più ricco» così [...] il Bachelard che ricopia le parole-immagini dei suoi poeti, è un lettore che sta diventando, *lui*, in quel momento, poeta, che al di là della pagina altrui sta iniziando a scrivere la *propria* pagina” ([15]: 313)⁷.

⁵ In questa sede assumiamo la posizione di un fine interprete italiano di Bachelard, Giuseppe Sertoli, che nella sua monografia *Le immagini e le realtà. Saggio su Gaston Bachelard*, 317-319 si fa sostenitore di una fluida continuità tra il Bachelard della dottrina degli elementi e quello delle sue ultime poetiche, in particolare per i temi che si sta per affrontare circa il rapporto tra immagini e *repos*, e *bonheur*.

⁶ “L'immagine che la lettura del poema ci offre, eccola diventare veramente nostra: essa si radica in noi stessi e, sebbene noi non abbiamo fatto che accoglierla, abbiamo l'impressione che avremmo potuto crearla noi, che avremmo dovuto crearla noi”. Cfr. ([1]: 13).

⁷ E tuttavia vale la pena ricordare che nel processo di immedesimazione proposto dal Borges del *Pierre Menard*, non c'è mai un'identità senza residui tra lo scrittore Menard e lo scrittore Cervantes. Si tratta piuttosto, a mio modo di vedere, del tentativo ‘matto e disperatissimo’ di una finzione letteraria di rendersi reale attraverso la duplicazione di un'altra finzione. Solo un genio poetico come Borges poteva immaginare una siffatta ‘realtà’. Cfr. ([6]: 649-658).

L'approccio bachelardiano al fatto estetico implica un netto taglio metodologico che lascia fuori tutto ciò che riguarda la composizione dell'opera d'arte, e più nello specifico della poesia, sacrificando perciò ogni forma di discussione circa la metrica, le strutture linguistiche, le forme retoriche, tutto a favore delle immagini da cogliere nella loro *purezza*. Questo taglio, dichiarato da Bachelard in virtù di una opportuna «modestia fenomenologica» ([1]: 15)⁸, viene decisamente incontro all'esperimento che qui si sta tentando di portare avanti: leggere una poesia, partecipare *felicemente* delle sue immagini, sospendendo ogni forma di giudizio sulla sua *sostanza* – e dunque della sua composizione – a favore della sua *funzione*, gesto che finisce col mettere in discussione quella stessa sostanza (poetica)⁹.

III

Ora, alla luce di queste poche considerazioni bachelardiane, vorrei riproporre, con maggiore consapevolezza, il seguente quesito: se adottassimo la metodologia bachelardiana di lettura – che è *già* poesia – dei testi poetici, rivolgendo la nostra attenzione alle immagini e alle immagini soltanto, a nient'altro che alla forza immaginativa attraverso cui diamo, come *nuovi* poeti, *nuova* vita a quelle immagini, per godere infine della felicità che riposa sul fondo di esse; se lo facessimo a partire dalla lettura della poesia di Scholl, nella beata ignoranza della sua genesi, sperimenteremmo il *bonheur* e la felice quiete tanto cara a Bachelard? Se questo è il fine, che si tratti di una poesia, opera genuina di un artista, di una falsa appropriazione di un imbroglione, o del prodotto di un mero calcolo probabilistico, che differenza fa? Non diciamo che non ci sia – beninteso – e la domanda è posta volontariamente in chiave provocatoria, ma non si può non riconoscere che la sempre più rapida e totale rivoluzione digitale ci costringe ad aggiungere un altro problematico tassello al già complesso puzzle dal titolo “che cos'è l'arte?” e, volendo aggiungere persino un sottotitolo, “che cosa il digitale *fa* all'arte poetica?” – sempre ammesso che la dimensione poetica possa essere isolata e considerata come un aspetto singolare e opzionale dalla natura umana, e non, piuttosto, la sua dimensione originaria, imprescindibile. L'incalzante avanzamento tecnologico ci lancia una nuova sfida e ci pone di fronte a un nuovo tipo di fatto estetico, di cui non avremmo potuto fare esperienza fino ad alcuni anni fa, di cui, ad esempio non ha potuto fare esperienza Bachelard. Cosa avrebbe pensato se si fosse trovato di fronte alla poesia di School, o al posto dell'inconsapevole editore di *The Archive*? E una volta scoperto l'inganno, come avrebbe valutato quel prodotto artificiale? Non ci è dato conoscere le risposte ma credo valga la pena quanto meno porsi le domande. Prendere atto della prepotente “invasione” digitale, persino in quell'attività *umana, troppo umana* che è la poesia, significa riconoscere la necessità di aprire il discorso filosofico sull'uomo a sempre nuove questioni solo apparentemente marginali: agli occhi di ogni buon filosofo, mediamente *engagée* coi e dai ‘segni del presente’¹⁰, che uno studente universitario inganni una prestigiosa rivista elaborando un generatore di poesie, non dovrebbe rappresentare un mero episodio isolato di una furba bravata, ma piuttosto il segnale di una necessaria messa in discussione continua e costante di cosa sia l'*umano*, in ogni sua forma e prassi.

BIBLIOGRAFIA

- [1] Bachelard, Gaston. 1975. *La poetica dello spazio*. Bari: Dedalo.
- [2] ———. 1989. *La terra e le forze. Le immagini della volontà*. Como: Red.
- [3] ———. 2008a. *Il diritto di sognare*. Bari: Dedalo.
- [4] ———. 2008b. *La poetica della rêverie*. Bari: Dedalo.
- [5] ———. 2017. *Psicanalisi dell'aria. L'ascesa e la caduta*. Como: Red.
- [6] Borges, Jorge Luis. 1996. «Pierre Ménard, autore del “Chisciotte”». In *Tutte le opere*, 1:649–58. Milano: Mondadori.
- [7] Bostrom, Nick. 2018. *Superintelligenza. Tendenze, pericoli, strategie*. Torino: Bollati Boringhieri.
- [8] Cassirer, Ernst. 1999. *Sostanza e funzione. Ricerche sui problemi fondamentali della critica della conoscenza*. M. Ferrari. Firenze: La Nuova Italia.
- [9] Christodifès, Constantine G. 1962. «Bachelard's Aesthetics». *The Journal of Aesthetics and Art Criticism* 20 (3): 263–71.
- [10] De Biase, Riccardo. 2019. «Segni del presente. Filosofia della cultura e culture del digitale». *EDA – Esempi di Architettura*, n. Special Issue.
- [11] Eco, Umberto. 1968. *La struttura assente*. Milano: Bompiani.
- [12] Goll, Yvan. 1951. *Les cercles magiques*. Paris: Falaize.

⁸ In quello stesso anno – quello della pubblicazione de *La poétique de l'espace* – nel corso di un'intervista con il professor A. Aspel, Bachelard ammise “as for the problems of the composition of a poem, I understand very well that in forming a general poetic theory this should be done, but I don't know, I don't know – I am not qualified to do this research. on the contrary, I would like to devote all my remaining strength to continuing what I have done ... A poem without unity, but with ten beautiful images, well, that suits me. You see that my role is very modest”. Intervista riporta in ([9]: 268).

⁹ Gesto, quello del passaggio dalla sostanza alla funzione che, com'è noto è stato messo brillantemente in luce, tra i primi, dal lavoro Cassirer, *Sostanza e funzione* ([8]).

¹⁰ Rimandiamo al volume collettaneo del 2019 che ha cercato di indagare proprio la centralità del problema rappresentato dalle culture digitali ai nostri giorni. Cfr. ([10]).

- [13] Scholl, Zackary. 2011. «For the Bristlecone Snag». *The Archive*, 30–31.
- [14] ———. s.d. «My poetry generator passed the Turing Test». Zackaryscholl. <https://zackaryscholl.com/other/poetry/>.
- [15] Sertoli, Giuseppe. 1972. *Le immagini e le realtà. Saggio su Gaston Bachelard*. Firenze: La Nuova Italia.

Schemi, ipotesi e algoritmi

Approcci kantiani alla filosofia delle tecnologie digitali

Riccardo De Biase

Università degli Studi di Napoli Federico II, Italia – rdebiase@unina.it

ABSTRACT

Il contributo intende confrontare lessici e concettualità di due ambiti apparentemente assai distanti: l'universo del criticismo kantiano con l'orizzonte delle tecnologie digitali. La proposta teorica è che attraverso un'indagine delle linee-guida della dottrina kantiana dello schematismo trascendentale, sia possibile cogliere l'elemento determinante del processo algoritmico: la sua funzione di ipotesi euristica e probabilistica.

PAROLE CHIAVE

Criticismo, Algoritmo, Schema, Agente bayesiano.

INTERVENTO

Il mio contributo consisterà essenzialmente in due momenti e alcune conclusioni di carattere propositivo. Nel primo momento, attraverso l'analisi di alcuni tratti ben delimitati del criticismo kantiano, mostrerò la direzione precipuamente *pragmatica* (o addirittura *pragmatista*) del trascendentale, un modo di leggere la filosofia di Kant funzionale ai miei scopi, e che ha a mio avviso avuto i più gravidi sviluppi sulla più "adulta" riflessione filosofica circa l'epistemologia contemporanea. Nel secondo momento misurerò, appunto, questo modello interpretativo nel confronto con quelli che io chiamerò, sulla falsariga di Kant, i "fenomeni digitali", ossia quegli "ipereventi tachicritici" che caratterizzano – almeno secondo me – le questioni più interessanti delle tecnologie digitali per un filosofo. Le conclusioni finali riguarderanno alcune riflessioni generali sul rapporto tra *Kulturwissenschaften* e *Naturwissenschaften* e la necessità, assolutamente non aggirabile, di un confronto serrato ma orizzontale e condiviso.

I

Nel 1798 compare nel panorama editoriale della Germania di fine secolo l'*Antropologia da un punto di vista pragmatico*, l'ultimo libro composto e pubblicato da Kant in vita e in grado di intendere e di volere ([7]). Dal punto di vista del Kant "canonico", questo libro può apparire – e ad alcuni è senz'altro apparso – un fuor d'opera, un'impresa connotata da elementi di discontinuità con il rigore teorico delle tre *Critiche*. E per certi aspetti lo è davvero. Frutto della redazione e dell'ordinamento di un quasi trentennale impegno didattico, e organizzata nell'usuale modo architettonicamente sistematico (almeno nelle intenzioni del suo autore), l'*Antropologia* si apre con una serie di "dichiarazioni" che non mi pare inadatto definire "rivoluzionarie", almeno se confrontate con la cautela con la quale Kant solitamente si esprime: a partire dalla considerazione circa la priorità e i benefici dello studio dell'uomo «nella cultura civile», Kant distingue un'antropologia «da un punto di vista *fisiologico*» da quella del «punto di vista *pragmatico*» ([7]: 99), in base al fatto che la seconda «mira a indagare ciò che *egli* [l'uomo], in quanto essere che agisce liberamente, fa ovvero può e deve fare di se stesso» ([7]: 99). La centralità dell'analisi delle strutture dell'agire umano – perché di questo tratta fundamentalmente il libro del '98 – non ha una funzione meramente didascalica o dizionariale. Non si tratta, ossia, di dotare il panorama culturale della sua epoca di una precettistica comportamentale (anche se tracce di ciò sono evidenti), di un manuale di "buone maniere", o di una rassegna analitica delle facoltà implicate nella suprema rappresentazione, «quella dell'Io», ciò che dell'uomo «fa una *persona* e, in virtù dell'unità della coscienza in tutte le alterazioni che possono toccarlo, una medesima e sola persona, cioè un essere del tutto differente per rango e dignità dalle *cose*» ([7]: 109). È, oltre a tutto ciò, anche e forse soprattutto la rivendicazione dell'unicità e irriducibilità dell'umano, di gusto illuministico, certamente, ma che ridimensiona l'intangibilità di una Ragione astratta e autoreferenziale a vantaggio di una ragionevolezza agente e relazionale. Ma non c'è solo questo. È nella "natura" stessa dell'indagine, e non solo sui suoi oggetti, che si rivela l'autentica funzione di un'antropologia *pragmatica*, quella che ha di mira e che «contiene la conoscenza dell'uomo inteso come *cittadino del mondo*» ([7]: 100); e non basta "studiare" l'uomo come cittadino del mondo in una dimensione asettica di siderea lontananza del suo oggetto (che appunto non è l'uomo isolato ma l'uomo-nel-mondo), perché espressioni come

«conoscere il mondo» e «aver pratica del mondo» sono nel loro significato piuttosto lontane l'una dall'altra; infatti, nel primo caso ci si limita a *comprendere il gioco* cui si è assistito, mentre nell'altro vi si è *preso parte* ([7]: 100).

È difficile non assaporare in queste righe un appello agli intellettuali (utile anche a quelli dei tempi nostri, mi spingo a dire) a scendere dall'empireo della teoresi e confrontarsi coi segni del presente, del suo, di Kant, come del nostro. Ed è complicato sottrarsi alla suggestione di vedere in questo appello un potente stimolo nei confronti degli “umanisti” – e in particolari dei filosofi – troppo spesso incagliati in dibattiti autoreferenziali e poco, appunto, pragmaticamente rivolti ai “segni del presente” ([2]). Ma su ciò tornerò al termine del contributo.

Ora però è opportuno fare un passo indietro di una decina d'anni o giù di lì, e rileggere alcuni brevi passaggi della *Critica della ragion pura*. Ciò a motivo di far penetrare gradualmente, e più specificamente sul piano concettuale, nelle motivazioni che mi spingono a vedere nel trascendentale una “via regia” per l'approccio a una filosofia delle tecnologie digitali. Se ammettiamo che queste ultime fanno parte della realtà, *sono* un pezzo preponderante della *nostra* realtà, interessiamoci a cosa Kant dica di quest'ultima: «realtà, nel concetto puro dell'intelletto, è ciò che corrisponde a una sensazione in generale e quindi ciò il cui concetto significa in se stesso un essere (nel tempo); la negazione è ciò il cui concetto rappresenta un non essere (nel tempo)» ([5]: 193), Sono le pagine di straordinaria centralità (ed anche difficoltà) per l'equilibrio dei pesi teorici complessivi della prima *Critica*, riguardanti la discussione sugli schemi trascendentali dell'intelletto puro, in particolare lo schema del concetto puro di realtà. Kant prosegue: «la contrapposizione di realtà e negazione ha dunque luogo nel differenziarsi, nel medesimo tempo, di tempo pieno e tempo vuoto» ([5]: 193). A mio avviso – e mi si perdoni l'eccesso di sinteticità – qui Kant sta dicendo che lo stare-nella-realtà è il risultare di un rapporto di contiguità tra noi e gli oggetti in quanto fenomeni, rapporto che, intaccandoci perché esseri senzienti, procura in noi delle modificazioni. Ora, come enti dotati di sensibilità, quel che in noi corrisponde agli oggetti in quanto fenomeni, Kant lo chiama «materia trascendentale di tutti gli oggetti quali cose in sé (la cosità, la realtà)» ([5]: 193-194), ma un adeguamento siffatto (“materia” senziente umana-materia degli oggetti reali), a causa del costante alternarsi di “vuoti” e “pieni” di cui è costitutivamente intessuta la realtà, non può mai essere un perfetto e simmetrico aderire punto a punto. Siccome, argomenta Kant, «ogni sensazione ha un grado o quantità, in cui può riempire in misura maggiore o minore lo stesso tempo», non si può non dedurre che «ciò che fa sì che ogni realtà sia rappresentata come un *quantum*, è un rapporto e una connessione, o piuttosto un passaggio dalla realtà alla negazione» ([5]: 194). Kant sta descrivendo lo schema della realtà (come tutti gli altri schemi dei concetti) come un processo, una serie mobile (perché intimamente temporale) di “fatti” che, nella loro globalità, sono rappresentabili come un continuo, ma che intesi come pacchetti di quantità sono discreti. L'alternarsi di vuoti e pieni, di spazio e materia, di sentire (umano) passivo e re-agire (umano) attivo, genera un'ondulazione di “pezzettini” spaziomateriali, una dinamica che si nutre delle istruzioni che lo schema della realtà impone a chi si trovi “nella” realtà e voglia comprenderla. Non solo: la dottrina dello schematismo, e in particolare lo schema del concetto di realtà, è pensato da Kant come quel dispositivo che non si limita a rappresentare lo status quo di ciò a cui si applica, ma che in un qualche modo pure lo determina e lo “crea”. Mi attengo su ciò, non potendo per ovvi motivi entrare in dettaglio, a quanto ha sapientemente scritto Umberto Eco qualche anno fa, proprio riguardo a ciò: «introducendo lo schematismo nella prima versione del sistema (...), Kant si trova tra le mani un concetto esplosivo» ([3]: 72). Facendo una distinzione tra un “acerbo” schematismo (quello della prima *Critica*) e quello della terza *Critica*, della *Logica* e dell'*Opus postumum*, lo studioso italiano avanza l'ipotesi che con quest'ultima versione dello schematismo, si «cerca di costruire il concetto di albero (lo si assume) *come se* gli alberi fossero quali noi possiamo pensarli» ([3]: 73). Ciò ha delle conseguenze incalcolabili, e non soltanto per la dottrina specificamente kantiana, ma per l'intero sviluppo del pragmatismo peirceano e del fallibilismo popperiano ([3]: 80-87; 113; 394). Fatto sta, che «con questo tardo schematismo l'intelletto non costruisce la semplice determinazione di un oggetto possibile, ma *fa* l'oggetto, lo *costruisce*, e in questa attività (di per sé problematica) procede per tentativi» ([3]: 78). È fin troppo evidente, qui, che “procedere per tentativi” significa fare delle ipotesi che possono e debbono essere poste al vaglio dell'esperienza. Ossia – ma su ciò ritornerò più avanti, significa *imparare dall'esperienza*. Schema come processo, come ciò che è a un tempo repertorio di “fatti” e dispositivo di istruzioni di ordinamento dei “fatti” stessi: non si sente un'aria di famiglia per orecchie sensibili alle (poche) certezze dell'epistemologia contemporanea? Non si avverte che si sta parlando di noi, del nostro modo di avvicinarsi al reale, anche e soprattutto a quello mediale-digitale?

II

Derrick De Kerckhove ha diffuso qualche tempo fa nel dibattito italiano il termine “brainframes”, dove con questo semineologismo si è oramai abituati a intendere «qualcosa di diverso da un atteggiamento, da una mentalità (...) dato che non è mai localizzato nella struttura superficiale della coscienza, ma nella sua struttura profonda» ([9]: 69), e che «la prima dimensione della nostra esperienza che viene modificata dai *brainframes* è l'aspetto temporale» ([9]: 68). Sia chiaro: quel che si adombra in questa espressione è qualcosa di diverso dalle linee portanti della funzione schematizzatrice presentataci

da Kant, pur con tutta la complessità e problematicità ivi connesse. Non lo è perché, nel bene e nel male – più nel male, dal mio punto di vista – il *brainframes* resta qualcosa che “sta nella testa” delle persone, e non una pura funzione del processo di mediazione tra esterno e interno o, per dirla kantianamente, tra sensibilità e intelletto. Eppure, questa nozione ha avuto e ha indubbiamente un suo valore euristico per i filosofi del digitale; ad esempio – e qui invece il discorso si fa molto più affine a Kant (e a Cassirer) – Luciano Floridi ha individuato nel «problema del *frame* e della fondazione simbolica» ([4]: 154) un punto nevralgico della riflessione sulla “natura” delle ITC. L’argomentazione di Floridi nasce in fondo dalla domanda che non solo noi “attuali” ci facciamo continuamente, ma che è stata fatta sin dalla preistoria dell’era informatica: saremo mai in grado di costruire macchine “intelligenti” a tal punto da rivelarsi indistinguibili, nei loro outputs, dai comportamenti umani? E se sì, quali potranno essere le conseguenze sui diversi piani della vita sociale? La lettura che il filosofo italiano è che «se sia veramente possibile costruirlo [un computer intelligente come l’uomo] è tuttora una congettura, se non l’atto di fede di qualcuno» ([4]: 154). Ciò perché due ordini di problemi, peraltro intimamente collegati, si oppongono – e per Floridi probabilmente si opporranno per sempre – alla realizzazione concreta del progetto in questione. Il primo, come già detto, è quello del *frame*, ossia quello «relativo al modo in cui un agente situato in un contesto può rappresentarsi un ambiente che cambia e interagire con esso *nel tempo* in maniera efficiente» ([4]: 155)¹. Avere a che fare, infatti, con «l’imprevedibilità del mondo là fuori» fa sì, per Floridi, che oggi «nessuno ha un’idea particolarmente chiara del modo in cui l’intelligenza artificiale possa risolvere il problema del *frame*» ([4]: 155). La questione della fondazione simbolica trova la sua ragion d’essere (insormontabile, per Floridi) nel fatto della radicale sintatticità del linguaggio delle macchine, visto che «possono invero maneggiare distinzioni elementari. Possono identificare identità in quanto eguaglianze (...) e differenze in quanto semplici mancanze di identità (...). Ma non sono in grado di apprezzare le caratteristiche semantiche degli enti coinvolti e delle loro relazioni» ([4]: 156). I computer, i robot, gli algoritmi – anche quelli genetici di ultima generazione – possono sì, secondo Floridi, compiere «atti vagamente protosemantiche (...). A meno che (...) non adattiamo l’ambiente o il problema all’ambiente al computer» ([4]: 157). Ma a questo impensato e impensabile adattare l’ambiente alle macchine dovrebbe corrispondere un «cruciale passo avanti nella soluzione del problema del *frame*», ossia l’articolazione dotata di senso alla domanda «in che modo i dati acquistano il loro significato?» ([4]: 157). Floridi conclude – e io con lui questo momento – che esiste «una soglia semantica tra noi e le nostre macchine e non sappiamo come fare in modo che queste ultime la valichino» ([4]: 158).

Non entro neppure per un attimo circa la questione se Floridi abbia ragione oppure no, se sia troppo pessimista (o ottimista?) o vi siano buoni motivi per dubitare delle sue posizioni. Vorrei solo concentrare la mia attenzione sul linguaggio – tanto di Floridi quanto di De Kerckhove. *Frame* (ma anche il *pattern* spessissimo usato dagli esperti informatici), simbolo, semantica: non mi sembra di deviare troppo dall’alveo principale della proposta, se ora mostro quanto di queste suggestioni entrino a far parte dell’universo semantico del “digitale”. Ma che non si tratti solo di affinità nominalistiche, proverò a dirlo in maniera solo sintetica, riproponendomi di farlo in altri luoghi e in altre circostanze con maggior respiro.

Cos’è un algoritmo? Ma soprattutto, *che fa un algoritmo*? Già solo la differenza ponderale tra queste due domande – tutta a favore della seconda, a mio parere – può dare un indizio sulle intenzioni della mia proposta. E per farlo, comincio allora da quello che oggi è forse tra i più ascoltati tra i filosofi delle tecnologie informatiche, Nick Bostrom, che discute in uno dei suoi ultimi libri quasi esclusivamente di quegli oggetti noti come «algoritmi genetici» ([1]: 17), che assieme alle reti neurali sono stati e sono il presupposto per l’attuale stato della ricerca, stato che riassuntivamente si può definire con la formula del *Machine Learning*, ossia del modello del flusso di dati indirizzati a far apprendere la macchina ad apprendere. Questa classe di algoritmi, derivata da quelli di «propagazione all’indietro» e associata alle reti neurali, mostrano di essere capaci di «imparare dall’esperienza» nel loro «generalizzare a partire dagli esempi forniti e individuando configurazioni statistiche nascoste nei dati di ingresso» ([1]: 31). Generalizzare, qui, può essere inteso in stretta analogia con quella «facoltà conoscitiva così particolare» ([6]: 97) che è il giudizio, ossia quella disposizione a «sussumere particolari leggi date sotto leggi più generali», ma che quando la legge non è data lavora appunto per ipotesi. Se gli algoritmi sono e funzionano come un dispositivo logico-matematico programmato per ricevere dati “esterni”, processarli attraverso una computazione, e generare un “risultato” che abbia i caratteri della stabilità e della efficienza, intesa quest’ultima – è stato detto con acume – come «potere realizzante dell’enumerazione» ([11]: 7), si può inferire senza troppo eccedere in voli teorici, che essi si comportino come il kantiano “fantasma” di Robinson Crusoe della famosa pagina della terza *Critica* ([6]: 341), oppure come un agente bayesiano: «questo uso più generale della probabilità per quantificare credenze è noto come il punto di vista *bayesiano*, ed è anche conosciuto come interpretazione *soggettiva* della probabilità, poiché le probabilità dipendono da ipotesi. I sostenitori di un approccio bayesiano alla modellazione dei dati e al riconoscimento dei modelli non considerano questa soggettività un difetto, poiché a loro avviso *non si possono fare inferenze senza supposizioni*» ([8]: 26). Ma se questo è plausibile, va tenuto conto del fatto che ora, nel mondo delle macchine, che si

¹ Corsivo mio.

presenta e si vuole meccanico, logico, lucido, è penetrata la caratteristica più propria dell'esperienza, la sua irrisolutezza, la sua enigmaticità, la sua parzialità, la sua ipoteticità: «nella logica classica aristotelica, il *principio del medio escluso* asserisce che ogni proposizione è ogni volta o vera o falsa, senza alcun ombra di sfumatura. Ma nel mondo reale, così com'esso è percepito dagli umani, ciò che è piuttosto onnipresente è la *parzialità* e non la *categoricità*. Generalmente, noi possediamo conoscenze parziali, certezze parziali, credenze parziali, parziale comprensione, trattandola con causalità e verità parziali. Il concetto di parzialità, e in special modo quello di parzialità della verità, gioca un ruolo centrale nella logica fuzzy» ([10]: 2).

Quel che propongo, allora, e nel concludere, è questo modello di lettura delle ITC e in particolare dell'AI: un varco, una via da percorrere in cooperazione tra filosofi e ingegneri, tra umanisti e informatici. L'apparente estraneità e incomunicabilità dei due "regni" del sapere viene a mio modo di supporre di molto ridimensionata quando si trovi un mezzo comune di dialogo. L'idea di "usare" Kant e alcune delle sue più geniali intuizioni – lo schema in analogia all'algoritmo, entrambi come contesto regolativo di istruzioni atte a testare ipotesi e a produrre il "nuovo" – può forse rappresentare un primo spunto per risillabare insieme il gran problema della conoscenza.

BIBLIOGRAFIA

- [1] Bostrom, Nick. 2018. *Superintelligenza. Tendenze, pericoli, strategie*. Torino: Bollati Boringhieri.
- [2] De Biase, Riccardo. 2019. «Segni del presente. Filosofia della cultura e culture del digitale». *EDA – Esempi di Architettura*, n. Special Issue.
- [3] Eco, Umberto. 1997. *Kant e l'ornitorinco*. Milano: Bompiani.
- [4] Floridi, Luciano. 2017. *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*. Milano: Cortina.
- [5] Kant, Immanuel. 1967. *Critica della ragion pura*. P. Chiodi. Torino: Utet.
- [6] ———. 1993. *Critica del giudizio*. A. Bosi. Torino: Utet.
- [7] ———. 2010. *Antropologia da un punto di vista pragmatico*. M. Foucault. Torino: Einaudi.
- [8] MacKay, D. J. C. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
- [9] Riva, Giuseppe. 2014. *Nativi digitali. Crescere e apprendere nel mondo dei nuovi media*. Bologna: Il mulino.
- [10] Zadeh, Lofti Aliasker. 1997. *Some Reflections on the Relationship Between AI and Fuzzy Logic (FL). A Heretical View, in Fuzzy Logic in Artificial Intelligence*. A. L. Ralescu, J. G. Shanahan. Heidelberg, Berlin: Springer.
- [11] Zellini, Paolo. 2018. *La dittatura del calcolo*. Milano: Adelphi.

Tra chair e empiétement lo spazio topologico: contributo merleau-pontyano ai sistemi informatici

Daniela De Leo

Università del Salento – daniela.deleo@unisalento.it

ABSTRACT

Partendo dagli studi condotti nell'ambito delle applicazioni e delle metodologie informatiche e computazionali in cui è dimostrato che il dualismo cartesiano di mente e corpo non è più sostenibile, il presente lavoro, attraverso la prospettiva fenomenologica di Maurice Merleau-Ponty in cui è teorizzata la natura intrinsecamente "incarnata" della cognizione, ha come obiettivo quello di avvallare l'unità mente e corpo.

Le direttive della riflessione fenomenologica potranno fornire un valido contributo per ripensare la cognizione e restituire la valenza esistenziale della *Leibzentrierung* nei sistemi informatici, per intenderli come sistemi incarnati.

PAROLE CHIAVE

Fenomenologia, corpo, *chair*, *empiétement*, sistemi informatici.

INTERVENTO

Hubert Dreyfus ([3]) sosteneva che era impossibile che il comportamento intelligente, negli esseri umani, fosse generato dall'elaborazione computazionale di informazioni e rappresentazioni percettive. Identificando tre caratteristiche del sistema nervoso che facilitano il verificarsi di tale attività: l'orizzonte interno di prontezze e aspettative sviluppate dall'esperienza che anticipano e, di fatto, modellano le nostre interazioni in corso; il carattere globale del processo che può interpretare i dettagli di una particolare interazione all'interno di un contesto più generale; e la possibilità di trasferire le esperienze in un organo o modalità di senso ad altri. Più praticamente, la ricerca è stata intrapresa al MIT da Rodney Brooks ([1]) e Daniel Dennett ([2]) per sviluppare robot che non lavorano nel modo rappresentazionale tradizionale ma usano interazioni fisiche come base per sviluppare comportamenti cognitivi di livello superiore. Questi lavori mostrano, tuttavia, che il progetto è ancora ampiamente computazionale e lontano dalle condizioni delineate sopra. Brooks stesso riconosce questi limiti asserendo che dovrebbe essere individuata una nuova prospettiva da cui guardare i sistemi informatici, un punto zero dell'orientazione che possa illuminare la correlazione tra percezione e azione.

Nel presente lavoro si è inteso direzionare l'obiettivo sull'approccio fenomenologico per individuare in esso una nuova prospettiva di indagine.

La percezione, in questo approccio fenomenologico, viene intesa come un'originaria apertura che esplose verso due direttrici: il mondo e il vissuto. Da un lato essa è un modo di accesso alla realtà quale è in se stessa, dall'altro è sensibile, vale a dire propria di un soggetto.

La percezione, in questa impostazione, appare non caratterizzata soltanto da un processo di tipo concettuale, ma da un sapere intuitivo e soggettivo, è sia un'attività mentale, sia il prodotto di questa attività, che rinvia contemporaneamente ad una posizione di pensiero e ad una posizione di realtà.

La nostra percezione mette capo a oggetti e, una volta costituita, l'oggetto appare come la ragione di tutte le esperienze che di esso abbiamo avuto o potremmo avere. È infatti da intendere come la traduzione dell'informazione acquisita dai nostri sensi in un'esperienza significativa. Può essere definita come un processo continuo di aggiornamento e mantenimento del modello interno dell'ambiente in cui viviamo. È proprio questo modello che ci permette di adattarci all'ambiente in modo funzionale e sicuro.

Fondamentale in questa relazione percettiva è il soggetto situato, il soggetto incarnato in un corpo.

Nella speculazione fenomenologica, e in particolare in quella merleau-pontyana, viene interrogato questo soggetto, quale struttura di significati, e viene definita la sua localizzazione nella relazione di mondo, non come *Ichzentrierung* ma come *Leibzentrierung*.

Nel lavoro, pertanto, si intende presentare la riflessione merleau-pontyana intorno alla tematica della concezione del corpo come: *abituale* e *attuale*. Il primo è impersonale e generale, ovvero il nostro abitare un corpo sviluppato attraverso intenzioni sedimentate, di cui le cose non sono altro che il naturale prolungamento. Il secondo, il *corpo attuale*, ricopre il

contesto singolare e particolare del soggetto, questo corpo “proprio” è il terzo termine, sempre sottinteso, della struttura figura sfondo, l'*entriade* corpo-oggetto-mondo; insieme di organi sistematicamente coerente nell'unità o totalità dei sensi, il corpo è “un *système de systèmes voué à l'inspection d'un monde*” ([12]); struttura, implicita in tutte le analisi del mondo, struttura originaria che sola rende possibile il senso e i significati; cornice da cui tutte le esperienze derivano, sempre già presente e presupposta. «Le corps propre est dans le monde comme le *coeur* dans l'organisme: il maintient continuellement en vie le spectacle visible, il l'anime, il le nourrit intérieurement, il forme avec lui un système» ([10]: 235); organismo vivente (*Leib*) ancora prima di essere un qualcosa di materiale (*Körper*) è un fluire continuo, luogo *inoltrepassabile*, unico luogo da cui si può dispiegare una dimensione spazio/temporale. Il corpo vivente esprime innanzitutto la modalità del nostro essere al mondo, il punto di ogni orientazione di mondi possibili. Il corpo «cet étrange objet qui utilise ses propres parties comme symbolique générale du monde et par lequel en conséquence nous pouvons “fréquenter” ce monde, le “comprendre” et lui trouver une signification» ([10]: 274).

Nello specifico si proporrà, con un metodo storico-critico in prospettiva sincronica e diacronica, la comparazione degli scritti e dei manoscritti inediti merleau-pontyani per scoprire i prodromi di una filosofia della *chair*, in cui il filosofo francese tiene uniti tre livelli: la dimensione dell'io, lo sconfinamento del mondo estesiologico, e il pensiero con le sue fenomenizzazioni. In essi, contro l'ontologia cartesiana dell'oggetto, Merleau-Ponty riabilita l'ontologia dell'esperienza sensibile per riaffermare la sua collocazione – *io sono il mio corpo* – ponendo così le basi per una filosofia della carne. Una filosofia che non è un'opposizione proiettiva tra esterno ed interno, ma una scrittura dello sconfinamento, *empiétement* in cui prende posto l'intenzionalità e il desiderio. L'infrastruttura esistenzialista della concezione della *chair* si presenta, così, costruita su tale *empiétement*. Lontana da ogni forma di umanesimo esplicativo, come da ogni teologia positiva la filosofia della *chair* si intreccia con il movimento dell'*empiétement*.

Il lavoro di scavo archeologico degli scritti merleau-pontyani porterà in luce le sfumature semantiche del termine *chair*, che ribalteranno le interpretazioni svianti che lo vedono come mera trascrizione del termine *Leib*.

Tale procedere contribuirà a chiarificare equivoci interpretativi, dovuti al fatto di inglobare il *depositum* merleau-pontyano in pensieri egemoni, come quello di Edmund Husserl o di Martin Heidegger: ad esempio come il leggere *Leib* (il corpo vissuto contrapposto a *Körper* corpo proprio husserliano) là dove Merleau-Ponty scrive *chair*, o l'intendere i costanti riferimenti all'*Être* un calco del *Sein* heideggeriano.

Da questa ermeneutica testuale si approderà alla teorizzazione merleau-pontyana della spazialità pre-proiettiva e pre-metrica capace di sostenere l'ontologia della *chair*, in opposizione all'ontologia cartesiana. Merleau-Ponty amplia le strutture topologiche per ritrovare una spazio-temporalità naturale «topologica» e «ontogenetica», affrancata dagli schemi che sostengono l'estensione cartesiana. L'interazione pre-oggettiva dello spazio e del tempo è rinvenuta nella caratteristica delle infrastrutture più profonde dello schema corporeo: l'interrogazione non è più direzionata ad uno spazio e ad un tempo oggettivi, ma interroga questo spazio e questo tempo che “nous sommes” ([11]). Questa interrogazione sottende ad ogni conoscenza e accompagna la percezione di mondo.

Pertanto, con un approccio genetico-teoretico si propone, nella combinatoria delle opere edite e degli studi sui manoscritti, una ermeneutica filologico-testuale che metta in luce la composita eredità concettuale merleau-pontyana e la sua intrinseca disposizione ad essere incessantemente interrogata, secondo multiple direzioni, come modello generativo di pratiche sperimentali nel campo dei sistemi informatici.

Dunque, portare in discussione il contributo della fenomenologia merleau-pontyana condurrà ad uno spostamento dal regno cartesiano del pensiero puro verso il coinvolgimento *chiasmatico* della cognizione incarnata e il riconoscimento indiscutibile dell'unità corpo-mente.

La teoresi fenomenologica costituirà, così, un valido fondamento per tutti gli studi che inquadrano i sistemi informatici come sistemi incarnati.

Con l'ausilio dell'informatica le direttive della riflessione fenomenologica sulla corporeità potrebbero essere *documentate*. Quindi parallelamente alla questione teoretica, sarà utile procedere ad una documentalità. Nello specifico, si propone di fornire specifici contenuti per la realizzazione di un ipertesto multimediale:

- trascrizione dei testi merleau-pontyani sulla corporeità, anche degli inediti ([11]);
- comparazione con altri testi (tra gli altri quelli di Dreyfus ([1]), Brooks ([2]) e Dennett ([3])).

I contenuti saranno rappresentati, nell'ordine, da un *Lessico*, un *Glossario*, una *Tassonomia* e una *Mappa Semantica Navigabile* e risulteranno fruibili come oggetti singoli all'interno di uno strumento unico, tramite portale web. Il portale fornirà anche un motore di ricerca testuale per navigare all'interno dei contenuti dei testi. L'archiviazione dei dati verrà effettuata secondo *modelli a grafo*. Nello specifico, si farà ricorso alle seguenti sotto-attività:

- a) definizione del *modello dati concettuale*: per consentire un'adeguata elaborazione e gestione del contenuto informativo dei testi, verrà definito un modello dati esteso, funzionale alle scelte di archiviazione su database a grafo, che consenta agli utenti di effettuare ricerche;
- b) creazione di un *Lessico* di riferimento: a partire da un set di documenti di riferimento;
- c) creazione di un *Glossario*: verrà definita ed implementata un'interfaccia utente che consenta di associare i termini appartenenti al *Lessico* ottenuto al punto b);
- d) creazione di *Tassonomie*: per ottenere una struttura gerarchica multi-criterio da applicare ai termini del *Lessico* e alle relative definizioni, verrà adottato un sistema basato su *tag* (etichette di categoria predefinite e applicabili dall'utente esperto di dominio ai termini del *Lessico*);
- e) creazione di una *Mappa Semantica*: la *Tassonomia* ottenuta al punto d) verrà arricchita da ulteriori relazioni semantiche intercorrenti fra i suoi elementi, al fine di agevolare un'analisi multi-disciplinare del contenuto informativo dei testi. Le relazioni semantiche verranno definite secondo le specifiche della teoria dei database a grafo.

BIBLIOGRAFIA

- [1] Brooks, Rodney A. 1991. «Intelligence without representation». *Artificial Intelligence* 47: 139–59.
- [2] Dennett, Daniel C. 1994. «The practical requirements for making a conscious robot». *Philosophical Transactions of the Royal Society of London A* (349): 133-146.
- [3] Dreyfus, Hubert L. 1992. *What computers can't do: a critique of artificial reason*. Cambridge: MIT Press.
- [4] Flores, Fernando, Michael Graves, Brad Hartfield, e Terry Winograd. 1988. «Computer systems and the design of organizational interaction». *ACM Trans. Office Information Systems* 6 (2): 153–72.
- [5] Husserl, Edmund. 1977. *Cartesian meditations*. The Hague: Martinus Nijhoff.
- [6] Hutchins, Edwin. 1995. *Cognition in the wild*. Cambridge: MA: MIT Press.
- [7] Johnson, Mark L. 1995. «Incarnate minds. Minds and Machines». *Minds and Machine* 5: 533–45.
- [8] Madison, Gary B. 1981. *The phenomenology of Merleau-Ponty*. Athens: Ohio University Press.
- [9] Maturana, H. R. 1983. «What is it to see?» *Archives of Biol. Med. Exp.* 16: 255–69.
- [10] Merleau-Ponty, Maurice. 1945. *Phénoménologie de la perception*. Paris: Gallimard.
- [11] ———. 1958. «Projets de livre». Vol. VI, ff.386.
- [12] ———. s.d. «La Prose du monde». Vol. III, ff. 263.
- [13] Turner, Bryan S. 1984. *The body and society*. Oxford: Blackwell, 1984.
- [14] Varela, Francisco, Evan Thompson, e Eleanor Rosch. 1991. *The embodied mind*. Cambridge: MIT Press.

Sessione Testi 6
Katherine Johnson

There and back again: what to expect in the next EVT version

Giulia Cacioli¹, Giacomo Cerretini², Chiara Di Pietro³, Sara Maenza⁴, Roberto Rosselli Del Turco⁵, Simone Zenzaro⁶

¹ M.E.T.A. S.r.l., Italia - giulia.cacioli@gmail.com

² Università di Pisa, Italia - cerre.giacomo93@gmail.com

³ M.E.T.A. S.r.l., Italia - dipi.chiara@gmail.com

⁴ Università di Pisa, Italia - s.maenza1@studenti.unipi.it

⁵ Università di Torino, Italia - roberto.rosselidelturco@unito.it

⁶ Istituto di Linguistica Computazionale “A. Zampolli” - CNR, Italia - simone.zenzaro@ilc.cnr.it

ABSTRACT

Developing software as complex as EVT requires a significant amount of time and resources. As a result, the release frequency of new versions has never been particularly high, especially for major versions. The first release of EVT 1 dates back to 2014, in 2016 we published the first alpha version of EVT 2, after which more complete versions followed: a beta1 in 2017, and a beta2 in 2020. Not only did it take three years between the two beta versions of EVT 2, but we still don't see the light for a 1.0 version. What has happened in the meantime? How is it possible that a stable version for EVT 1 was achieved relatively quickly, and development slowed down significantly thereafter? This talk aims at updating the Italian DH community about the current status of EVT and its future prospects with regard to the next version (EVT 3).

PAROLE CHIAVE

Digital philology, digital scholarly edition, TEI/XML, web publishing, EVT.

TALK

1. INTRODUCTION

Edition Visualization Technology (EVT)¹ is an open-source tool to publish Digital Scholarly Editions starting from TEI/XML-encoded documents. Although it was born to serve the goals of a single project, the Digital Vercelli Book, it has been subsequently developed in such a way as to become an easy to use and flexible tool, which offers all the features that are to be expected in a modern digital edition. As of today, there are several DSE projects using both EVT 1 (v. 1.3), the original version still very effective for publishing digital facsimiles together with diplomatic transcriptions, and EVT 2, currently available as a second beta version which adds support for diplomatic editions in addition to critical edition support².

The migration of the source code base from EVT 1 to EVT 2 was due to several reasons, in particular to the search for greater flexibility in the management of the many, and sometimes very complex, functionalities necessary for the proper functioning of EVT, and to shortcomings inherent in the development framework chosen. Both of these causes occurred again at some point in the development of EVT 2, but in addition a flaw in the development method was also highlighted, a problem that was addressed in the new code migration to EVT 3.

2. A NEW CODE MIGRATION

As a consequence of the increasing adoption of EVT as a platform to visualise and publish editions across different projects with very different needs, we realised that the software has the potential to be extended in several directions leading to new features that would have to be integrated into the software. EVT has already gone through an overall rewrite of its code base, switching from an XSLT-based platform to a more flexible JavaScript framework coupled with established software design patterns (such as the Model-View-Controller one) in order to provide the support for complex features such as critical edition support, image-text linking in the digital facsimile, named entities support, and more³.

The framework we chose is the well-known and widely adopted AngularJS⁴ that served well the purposes of EVT 2. Unfortunately, on July 1, 2018 AngularJS entered a 3-year Long Term Support (LTS) period (extended to December 31,

¹ Home page of the project: <http://evt.labcd.unipi.it/>.

² For more information about EVT and its history see ([5];[3]).

³ See Table 1 for a list of all existing EVT versions.

⁴ <https://angularjs.org/>.

2021). During this time the framework will not be improved and only critical functional or security flaws will be addressed. When this LTS period will come to an end, any further support will cease. Since the EVT time frame is longer than the LTS period we started debating the adoption of a different framework.

At the same time, integrating the multiple new functionalities – both to reach feature parity with the previous version and to include features developed for specific projects, often by creating forks of the main project – added great complexity to the existing code base. As a result, the source code has bloated making its management increasingly harder. The previously clean architecture of the software started to become clunky and difficult to explain for new contributors. Moreover, the integration of new features was centralized toward a single person who retained the overall picture of the project.

Meanwhile, Google replaced AngularJs with a new version of the framework that is just named Angular⁵. This entirely new framework incorporates most of the good software design patterns and encourages good practices with regard to web application development, bringing the updated standard technologies to the table (such as the Typescript language, a well-integrated testing framework, container deployment through Dockerization, etc.).

For these reasons, we decided that porting EVT to the new Angular framework would benefit the future of the project by granting the time to re-establish the software foundations on an architecture that will be able to address the flexibility required for the implementation of new features. A codebase restart also allowed us to streamline the contributions to the software, making the process more distributed and subject to the scrutiny of more people, so that a larger group of people could understand and approve the changes. More in detail, we adopted a Continuous Integration/Continuous Delivery (CI/CD) pipeline that avoids the integration of faulty code, along with the git-flow⁶ model coupled with code reviews to encourage good practices and to improve the quality of the codebase in anticipation of future requirements. Finally, the new codebase management requires to trace every change to the code, thus allowing the traceability of the features.

The new version of EVT is called EVT 3 and is currently in active development. Our first goal for EVT 3 is to reach feature parity with EVT 2. One notable consequence of adopting the new Angular framework is reflected by the new XML/TEI parser that has been implemented in EVT 3. The parser is in charge of transforming one or more valid XML/TEI documents into a Typescript Object Model that represents the actual subset of encoded information EVT is able to compute and visualise (at the moment it mostly corresponds to a subset of the TEI schemas).

The parser itself has been designed to be independent from the actual viewer in order to achieve two important goals: (1) independence from the XML/TEI encoding, which opens the possibility of future encodings to be accepted as input data (e.g. new versions of the XML/TEI, TAGML, or others); (2) independence from the edition type and/or level (i.e. diplomatic/interpretative editions, critical editions, integrated editions etc.).

The EVT 3 architecture is also designed to be as modular as possible in order to allow the arrangement of smaller building blocks that cooperate to the construction of new views that, in turn, will manage the visualisation of different levels of edition.

Version	Release date	Technology used	GitHub repository
EVT 1	2014	XSLT	https://github.com/evt-project/evt-builder
EVT 2	2016/20	AngularJS	https://github.com/evt-project/evt-viewer
EVT 3	2022	Angular	https://github.com/evt-project/evt-viewer-angular

Table 1: EVT versions at a glance.

3. NEW FEATURES IN EVT 3

Distributed facsimiles

With respect to the previous version, EVT 3 implements a component that is able to display digital facsimiles described in an IIIF⁷ manifest.json file. The manifest can be local to the edition visualised, or it can be remotely available as a URI. Usually using an IIIF manifest means that the actual images are stored remotely and published by an IIIF server. The image viewer is based on the well-known OpenSeaDragon⁸ viewer. This component needs to be extended in order to reintroduce the features already available in EVT 2, such as direct loading of local images ([7]).

⁵ <https://angular.io/>.

⁶ <https://nvie.com/posts/a-successful-git-branching-model/>.

⁷ <https://iiif.io/>.

⁸ <https://openseadragon.github.io/>.

Multiple manuscript descriptions and a new image-image view

A new feature already implemented in EVT 3 is the support for multiple TEI <msDesc> elements, allowing the user to view and compare the descriptions of multiple manuscripts. This functionality will be complemented by a new image-image view useful for comparing facsimiles of different manuscripts and/or different versions of the same image.

There are several use cases where support for multiple <msDesc> proves to be useful. It is possible, in fact, that an XML document would include texts from different manuscripts and detailed descriptions for each of them, therefore the reader might want to choose which one to view. Another situation where such an implementation could be useful is the need to examine different manuscripts with different transcripts of the same work, browsing <msDesc> elements related to different witnesses.

It will be equally useful to let the user have an image-image view: to compare different versions of the same image, f.i. images that testify the before and after the restoration of the manuscript, multispectral scans or with a particular lighting; or digital facsimiles of the various witnesses of the same work, or even of completely unrelated works, f.i. to study a specific scribal hand.

Currently, the tool for multiple manuscript description is accessible through a selector placed inside the header of the frame containing the text or the facsimile of the document, so that the user can choose the description to be displayed. By clicking on the drop-down menu, a list of the <msDesc> elements existing within the encoded document will be proposed, each represented by its own identifier (Figure 1). Once opened, it will be possible to close the view using the appropriate "x" button in the selector. Within the manuscript description view, each section is highlighted by means of an underlined title, followed by various labels with their associated information.

The image-image view will be implemented as two image frames presented side by side on the screen, each one enriched by a selector that allows the choice of the image to be displayed in order to facilitate comparison.

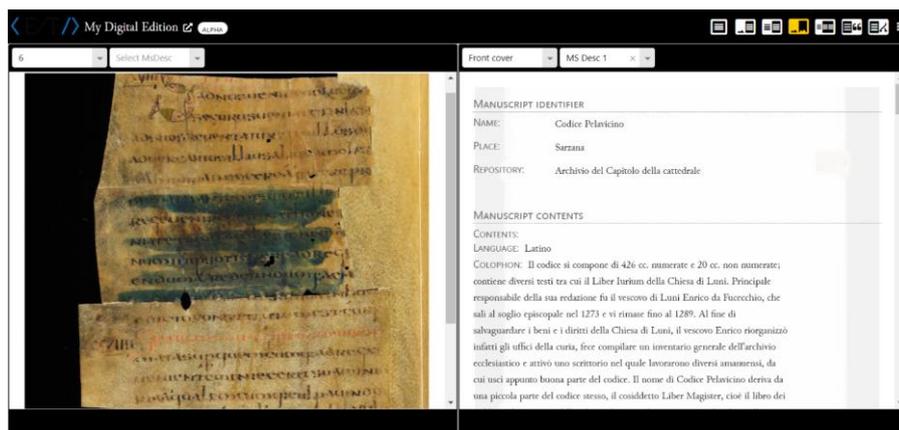


Figure 2. Multiple manuscript descriptions available thanks to the new selector.

User annotations

One of the new features developed for version 3 of EVT is the creation of an annotation system for text and images found within the digital edition. The main focus of this feature is to recreate, within EVT, the same actions performed by a reader when creating a note that relates to the displayed text, thus adding an additional layer of information to the edition data.

The approach used for the creation of a structure suitable for the definition of an annotation refers to the W3C standards, the Web Annotation Data Model. These models use a data structure based on the JSON-LD format, very useful in our case because it allows us to serialize the data in a traditional JSON model using a structure common to other EVT 3 functionalities, in this way every kind of resource is uniform and easy to manage.

Once the data model was defined, we moved on to the feature development phase, which was initially thought of as internal to EVT. With the progress of research, however, it was decided to develop a tool that could match the use of this feature both internally and externally to the context of EVT, opting for a Typescript library.

We can then divide and define the final development of the library in three macro areas of work, the first based on data storage, the second on the development of the library and finally the graphic part related to the internal GUI in EVT.

Regarding the part related to data storage, since EVT is based on a client-only model without using a server architecture, the notes are saved locally using the IndexedDB technology, that is a browser NoSQL database capable of managing structured data in JSON format.

The development of the library (EVT - Text Annotator), which to date only provides the creation of text-related annotations, exposes two APIs, *createAnnotation* for the creation of notes and *getAnnotation* for finding notes.

The GUI part instead has been developed within EVT, so it is not part of the library, and includes the appearance of a menu near the area selected by the user, giving him the possibility to choose whether to create a note or a simple text highlighting (Figure 2).

In the next step of development we will implement the second part of the library, extending it to the annotation of images and paving the way for new interesting features such as the choice of color of the highlights and the integration of custom tags for notes (already partially implemented), the option to export the entire database of created notes and to re-import it later (possibly on a different computer), and the ability to compare notes (even of different users) and to comment on them. This last feature requires some kind of support at server level in order to make notes persistent across browsers and users, but, since EVT is born as a client-only application, this kind of interaction will be implemented when the EVT 3 platform will provide server side capabilities.

Lunensis episcopus, predecessor suus, cum operariis omnibus de curte Sarzanei ¹⁰. Uterque en
instrumento legebatur, condonavit et remisit per se et suos successores omnibus suprascriptis o
nullus eorum cogatur ad canevas recipiendam **nec ad castaldiam** neque ad iscariam, nisi per vol
anno debet unus eorum esse iscarius, si fuerit v
amasiamenta omnia retinuerunt in se et success
suprascripti operarii omni anno dare viginti mo
congia vini ad palmentum et deferre usque in c
portare usque in canevas. Et episcopus debet e
ad socerandum, non debent dare nisi terciam pa
prato debent portare ad fenile et debent bis in c
coadunare fenum in prato; et alios attractus de
Soleria, de curte de Bolano, de curte de Carrari
facto pape vel imperatoris vel cardinalis vel cano
pape vel imperatoris vel guerre alii homines cor
predicto instrumento continerentur:

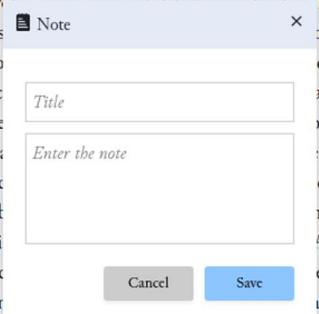


Figure 3. Highlighting the text to add a personal note

Semi-structured data processing

One of the last steps that will distinguish EVT 3 from its predecessor is the search for alternative methods to relational databases to process textual data in semi-structured format within a digital edition, and to provide at least part of the processing useful to historians and other scholars⁹.

Thanks to a series of complex queries, this new functionality will allow the user to consult a digital edition in a more effective and informative way, by cross-referencing the information already present in the XML/TEI encoding of the document.

For the initial development of this functionality, it has been essential to define a data structure that would be able to contain all the useful information contained in the XML/TEI document. This was possible through the use of JSON, a text-based data-interchange format. This choice was made mainly because of three reasons: the first one concerns the loss of information, which after the conversion from XML to JSON is often minimal or null; the second one is related to the actual development of the functionality, in fact this object notation is very suitable to being interrogated and to submit queries; finally, because the JSON notation is already being used in EVT since it is very suitable to being interrogated and to submit queries.

The beginning of the research focused its attention on the in-depth study of the documents encoded in XML/TEI which have been found to contain a lot of underutilized information. We realized that, once related to each other, this information could lead to a more precise and granular type of research. On the basis of this aspect we have therefore outlined a series of sample interrogations, which could reflect the possible queries carried out by the user within a digital edition. In this way, what emerged were four types of basic queries: (1) `<char>` queries, i.e. queries based on the encoding of characters, (2) `<table>` queries, which aims to perform calculations within tables, (3) entity queries, a type of query that works on (named) entities considering the tags *listPerson/person*, *listPlace/place*, *listEvent/event*, and (4) ontology queries, which

⁹ This new research avenue is the result of a workshop held at the University of Pisa in June 2020: “Medieval Archival Sources into the Digital. The Challenge of Processing and Visualising Semi-structured Data” (URL with full programme and video recordings: <http://www.labcd.unipi.it/fonti-archivistiche-medievali-nel-digitale/>). A thematic dossier based on the workshop papers has been published in the *Umanistica Digitale* journal (see [8];[4];[9]) for a different take on the ‘assertive edition’).

will serve to define the relationships between different entities by means of RDF-like triples. This last category is still in the development phase because it requires further study of the encoding method to be adopted.

As for the annotator, the need to develop a tool that could be used both inside and outside EVT emerged, so the direction we decided to follow is towards a Typescript library that can expose some of the APIs related to the query types. To date, the API is in a state of refinement and not yet published.

During the next development phase we will introduce a new graphical component, this time exclusively internal to EVT 3, to implement an advanced search tool, so that it will be possible to create a search query in a simple and intuitive way.

The next steps for this functionality will be to conceive a standard encoding system for ontologies and therefore to give a solution to cross-queries of XML/TEI documents which include RDF-like triples. A fundamental point on which we are working is to make the results of the queries totally processable in EVT, this would in fact allow us to have a unique connection between the visualization of the digital edition, the hyperlinks generated within it and the results of the queries.

Integrated edition

The last new feature we are developing for EVT 3 is the so-called *integrated edition* (see [6] for a description of this feature), a critical edition in which one or more witnesses are available as a separate diplomatic transcription, but can also be browsed on their own together with the corresponding manuscript images (if available). While maintaining the automatic generation of witnesses, EVT 3 will also make use of separate transcripts, when available, to connect them with the critical text.

The main goal is to make it possible for the final user to check the variant readings in their original context, including the original document, and to move from the critical text to the witnesses transcriptions and digital facsimiles (and back) in a seamless way, offering a wide range of new opportunities for research and study.

This will introduce an important change in the navigation system because, in addition to seeing the witness in the collation view, you will be able to navigate to the corresponding diplomatic edition and vice versa. Moreover, from the collation view it will be possible to reach the diplomatic edition/digital facsimile view (*image-text view*), where you will be able to view the text of the witness side by side with its transcript and related images. Thanks to this view it will be possible to make a full comparison between the two versions and to examine specific readings in their textual and documental context (Figure 3).

Two interesting navigation options will be provided. The first one is a link from the variant in the critical edition to the corresponding reading within the diplomatic transcription. The second one involves a text-image link, connecting the lesson of the witness to the corresponding manuscript area and vice versa.

From the theoretical point of view, this will be a very important result because it allows the editor(s) to overcome the dichotomy between the stemmatic method (Lachmannism and neo-Lachmannism) and the new philology approach (see [2]).

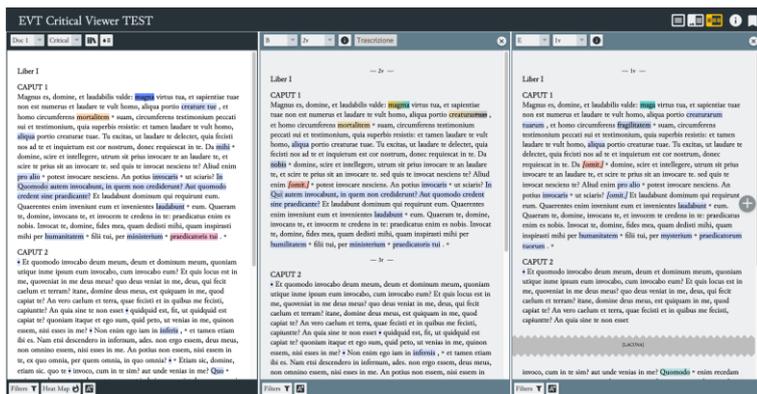


Figure 4. The new "Transcription" button leading to a separate diplomatic edition of the selected witness.

4. CONCLUSION

The development of complex software usually goes through multiple reengineering phases in order to tackle changes of technology, better understanding of the underlying architectural structure, and the emergence of novel features that could not be easily integrated in the already existing software. EVT has not been immune to this fate.

In this paper we presented the reasons behind the choice to embark on yet another new version of the software, and we also disclosed the features that are being developed in order to provide a better experience both for the technical contributors to EVT and for its users, be they digital editors or final users.

REFERENCES

- [1] Leff, Avraham, and James T. Rayfield. 2001. 'Web-Application Development Using the Model/View/Controller Design Pattern'. In *Proceedings Fifth IEEE International Enterprise Distributed Object Computing Conference*.
- [2] Monella, Paolo. 2019. 'L'edizione Scientifica Digitale: La Critica Del Testo Nella Storia Della Tradizione'. In *Textual Philology Facing 'Liquid Modernity': Identifying Objects, Evaluating Methods, Exploiting Media. Storie e Linguaggi. Rivista Di Studi Umanistici*. libreriauniversitaria.it edizioni.
- [3] Rosselli Del Turco, Roberto. 2019. 'Designing an Advanced Software Tool for Digital Scholarly Editions': *Textual Cultures* 12 (2): 91–111. <https://doi.org/10.14434/textual.v12i2.27690>.
- [4] ———. 2021. 'Elaborazione Di Dati Semi-Strutturati: Ipotesi Implementative e Casi d'uso Tratti Da Testi in Inglese Antico'. *Umanistica Digitale* 10: 387–407.
- [5] Rosselli Del Turco, Roberto, Giancarlo Buomprisco, Chiara Di Pietro, Julia Kenny, Raffaele Masotti, and Jacopo Pugliese. 2015. 'Edition Visualization Technology: A Simple Tool to Visualize TEI-Based Digital Editions'. *Journal of the Text Encoding Initiative*, no. Issue 8: 1–21. <https://doi.org/10.4000/jtei.1077>.
- [6] Rosselli Del Turco, Roberto, Chiara Di Pietro, and Chiara Martignano. 2019. 'Progettazione e implementazione di nuove funzionalità per EVT 2: lo stato attuale dello sviluppo'. *Umanistica Digitale*, No 7 (2019). <https://doi.org/10.6092/ISSN.2532-8816/9322>.
- [7] Rosselli Del Turco, Roberto, and Paolo Monella. 2020. 'Extending the DSE: LOD Support and TEI/IIIF Integration in EVT'. In *Atti Del IX Convegno Annuale AIUCD. La Svolta Inevitabile: Sfide e Prospettive per l'Informatica Umanistica*, edited by Cristina Marras, Marco Passarotti, Greta Franzini, and Eleonora Litta, 148–55. Bologna: Quaderni di Umanistica Digitale. <https://doi.org/10.6092/UNIBO/AMSACTA/6316>.
- [8] Rosselli Del Turco, Roberto, Enrica Salvatori, Andrea Nanetti, Marco Giacchetto, Vera Isabell Schwarz-Ricci, and Antonella Ambrosio. 2021. 'Introduzione: "Fonti Archivistiche Medievali Nel Digitale. La Sfida Di Trattare e Visualizzare Dati Semi-Strutturati"'. *Umanistica Digitale*, September 2021.
- [9] Vogeler, Georg. 2019. "'The "Assertive Edition": On the Consequences of Digital Methods in Scholarly Editing for Historians". *International Journal of Digital Humanities* 1 (2): 309–22.

XML-TEI: Un modello per la filologia d'autore

Giulia Tancredi¹, Cristina Fenu²

¹Università di Siena, Italia, gtancredi94@gmail.com

²Biblioteca civica "Attilio Hortis" di Trieste, Italia, cristina.fenu@gmail.com

ABSTRACT

Il paper propone uno standard di codifica XML-TEI per le varianti d'autore, integrando l'approccio cronologico-descrittivo dell'edizione documentaria digitale con il modulo *Critical Apparatus*: il metodo di codifica è la *parallel transcription*; la segmentazione del testo marcato, invece, riprende l'apparato sistemico usato in filologia d'autore.

PAROLE CHIAVE

XML-TEI, filologia d'autore, mark-up, varianti, Saba.

INTERVENTO

1. INTRODUZIONE

Il presente prototipo di mark-up per la filologia d'autore è stato elaborato all'interno del Progetto Saba 2021¹, iniziativa che si propone di pubblicare in *open access*², in occasione del centenario del *Canzoniere* di Saba, un'edizione digitale del manoscritto del *Canzoniere* datato 1919-20 (R.P.Ms I-18, in seguito *C19*) e conservato presso la Biblioteca civica Attilio Hortis di Trieste.

La proposta di un mark-up standard per i manoscritti moderni risponde all'esigenza di uniformare la codifica XML-TEI per la filologia d'autore così da evitare l'attuale eterogeneità di modellizzazione (cfr. [4]: 68; 179), nonostante le possibilità fornite allo studio della materia dalla sperimentazione di nuove strategie³.

Il modello qui esposto prende avvio dalla ricerca di Pierazzo sull'inserimento della variabile tempo nella trascrizione di un manoscritto⁴, fino all'introduzione nella *release* TEI P5 version 2.0 (12/2012) di alcuni elementi di codifica messi a punto dal gruppo TEI SIG MS⁵. L'approccio diacronico dell'edizione documentaria digitale⁶ è stato integrato nel presente metodo con il cap. 12 di TEI ([13]): l'elemento `<app>` e i corrispettivi `<rdg>` e `<lem>`, oltre alle varianti a stampa, indicheranno le varianti d'autore.

La scelta della *parallel transcription* è stata dettata dalla relativa semplicità del testo sabiano, oltre perché più intuitiva e facilmente gestibile dai software di visualizzazione⁷. Inoltre la gerarchia "orizzontale" della *parallel transcription* rappresenta più coerentemente un apparato di varianti, intese nella loro interdipendenza e stratificazione, e fornisce uno strumento utile soprattutto per indagare i rapporti paradigmatici delle correzioni.

Un modello così costituito ha il vantaggio di proporre una descrizione diplomatica immediatamente agganciata a un'interpretazione diacronica e "sistemica", cioè per fasi correttive⁸. Da un'unica trascrizione si potranno ottenere un'edizione diplomatica e una critica, usando due diverse *view* in un unico software⁹, oppure un'edizione arricchita da diversi strumenti-guida per l'utente. A parte questa scelta di non creare due edizioni distinte, ma "fondere" le trascrizioni e lasciare alle *view* la loro distinzione e interoperabilità, l'elaborazione del modello non si è già proiettata in una particolare resa grafica o compatibilità con un qualsiasi software di visualizzazione: quanto più il modello risulterà scollegato da

¹ Il Progetto Saba 2021, ideato dalla Biblioteca civica di Trieste, è promosso in collaborazione con l'Università Ca' Foscari e coinvolge studiosi e studenti delle università di Trieste, Bologna, Pisa, Torino e del Boston College (Mass.).

² Per la visualizzazione, verrà usato EVT, software *open source* per le edizioni digitali: cfr. ([10]).

³ Nel progetto VaSto, ad esempio, sono state collazionate con l'elemento `<witness>` le due volontà (d'autore e di "editor") conniventi nel manoscritto della *Storia fiorentina*: cfr. ([1]:155).

⁴ Cfr. ([6];[8]). La soluzione di Pierazzo è combinare un mark-up personalizzato di tipo *embedded* con JavaScript e XSLT, così da ricreare la diacronia della codifica durante la visualizzazione.

⁵ Il gruppo ha ampliato gli elementi di trascrizione, spostando il focus del mark-up dal testo al documento e proponendo una sintassi più attenta alla fenomenologia delle correzioni, oltre che alla topografia: cfr. ([12]).

⁶ Per la validità critica ed "ergodica" del concetto di "edizione documentaria digitale" cfr. ([7];[9]:56-9) e ([17]:196).

⁷ Al momento, nonostante la maggiore potenza del metodo *double-end-point-attached*, non esiste ancora uno strumento che consenta di preparare e visualizzare un'edizione digitale che lo usi: cfr. ([10]:157-8).

⁸ Per l'apparato sistemico o per fasi usato in filologia d'autore cfr. ([5]:59).

⁹ Sull'importanza di questa strategia cfr. ([10]:150).

programmi esistenti per la pubblicazione e la consultazione di edizioni digitali, tanto più potrà essere scalare, duraturo e valido per edizioni eterogenee.

2. IL MANOSCRITTO SABIANO

C19 è un quadernetto di complessive 214 pagine che riporta 186 liriche suddivise in sezioni e sottosezioni. Si tratta di una copia in bella predisposta da Saba per la stampa, come si evince dalla accurata messa in pagina e dalle frequenti notazioni autografe di carattere editoriale che si sommano a cinque fasi correttorie stratificate sul manoscritto, redatte ciascuna con una penna diversa¹⁰. Tra gli interventi autoriali si notano anche numerosi cartigli adesi alle pagine del quaderno, ritagli provenienti da altri taccuini, oltre che da pagine dello stesso quaderno: il *collage* è stato senz'altro adottato dal poeta per comporre il menabò in economia di tempo. Le varianti d'autore immediate e/o riconducibili alle diverse campagne correttorie interessano il 67,20% delle liriche, i cartigli il 31,72% del *corpus*. Il manoscritto offre dunque una ricca campionatura di interventi autoriali testuali, intertestuali e metatestuali e ben si presta, perciò, come caso di studio per la definizione di un modello generalizzato per la codifica di un'edizione genetica.

3. IL MARK-UP

Riprendendo la stratigrafia "testuale" del manoscritto d'autore (cfr. [3]), il presente mark-up si definisce principalmente intorno ai 4 livelli fondamentali del testo critico:

- il *livello testuale*, cioè la lezione che il filologo decide di mettere a testo;
- il *livello genetico*, cioè le correzioni cronologicamente successive al testo base;
- il *livello intratestuale*, cioè le varianti alternative;
- il *livello metatestuale*, cioè le postille d'autore.

Il livello testuale

Il presente metodo prevede che <lem> venga usato per la lezione da mettere a testo: nell'edizione di un unico manoscritto, marcherà l'ultima delle varianti, qualora il filologo non ritenga che l'ultima volontà dell'autore sia da ricercarsi in casi intermedi o nel testo base¹¹; in un'edizione multitestimoniale, invece, <lem> indicherà l'edizione a stampa o il testimone più attendibile. Il vantaggio, in entrambi i casi, è che si potrà offrire un testo "in pulito" grazie all'estrazione delle lezioni in <lem>¹².

Il livello genetico

La codifica delle correzioni

Una prima classificazione spazio-temporale del manoscritto riguarda gli strati scrittori, ovvero le correzioni appartenenti a un'unica campagna revisionale. Nel <teiHeader> un elemento <creation> riporterà un elenco di <change>, ognuno per strato scrittoria, ordinati cronologicamente; qualora avessimo nell'edizione più manoscritti con varianti d'autore si definiranno più <creation> e <change>. Il link a @change all'interno del testo è sempre in <mod>, elemento aggiunto da TEI SIG MS. <mod> indica una qualsiasi modifica dell'autore e contiene tutte le microcorrezioni a essa connesse, per le quali si segue il cap. 11 della TEI [13]: , <add>, <transpose>, ecc. <mod> non verrà usato qualora la trascrizione riguardi lo stato del documento oppure se la correzione appartiene ad un copista (indicato da @hand o <handShift/>).

La codifica delle varianti genetiche

All'interno di <app>, le varianti genetiche sono marcate con <rdg> e numerate con @varSeq, usato anche per <lem>. Il testo segmentato da <rdg> o <lem> corrisponde non all'evento singolo del *ductus* scrittoria, quanto alla fase individuata dal filologo.

Consideriamo l'esempio sabiano (fig. 1). Qui l'autore ha corretto 1) "dura" con "assidua" con la stessa penna del testo base, ma sicuramente in una variante tardiva; 2) ha cancellato con penna blu e ha aggiunto in linea "dura pena", con prima la variante immediata "lun<ga>"; 3) ha definitivamente scritto "avversa pena,".

¹⁰ *C19* è stato descritto e studiato da Giordano Castellani per l'edizione princeps del Canzoniere 1921 ([5]) e più di recente da Silvia Vodopivec, ai cui studi filologici sulla cronologia degli strati correttori il Progetto Saba 2021 si riferisce ([16]).

¹¹ Cfr. [3]:210, dove per l'edizione digitale di *Eros e Priapo* è stata marcata con <lem> la lezione base del manoscritto A con le revisioni antecedenti alla campagna correttoria del '60.

¹² EVT 2 ha già questa funzionalità.

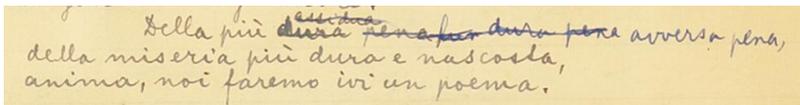


Figura 1. C19, p. 102, da *Verso casa*.

In un apparato a stampa le fasi sono:

^{1A}dura pena, > ^{2A}assidua pena, > ^{3B}dura (*prima lun<ga>*) pena > ^{4B}avversa pena

In grassetto è indicata l'ultima lezione, i numeri indicano la sequenza delle varianti, le lettere distinguono le penne (A testo base, B penna blu). Il mark-up sarà:

```
<l n="20">Della più <app>
  <lem varSeq="3">
    <mod change="#strato-2">
      <add>avversa pena,</add>
    </mod>
  </lem>
  <rdg varSeq="2">
    <mod change="#strato-2">
      <del>
        <subst>
          <del instant="true">lun<supplied>ga</supplied></del>
          <add instant="true">dura</add></subst> pena
        </del>
      </mod>
    </rdg>
    <rdg varSeq="1">
      <mod change="#strato-2">
        <del>
          <mod change="#strato-0">
            <subst>
              <del seq="1">dura</del>
              <add seq="2">assidua</add>
            </subst></mod> pena
          </del>
        </mod>
      </rdg>
    </app></l>
```

Figura 2.

Se confrontiamo l'apparato a stampa con il mark-up, noteremo pressoché la stessa segmentazione di testo coinvolto nelle fasi: solamente in `<rdg varSeq="1">` sono marcate insieme le due fasi “dura pena” e “assidua pena”. Il mark-up ha il vantaggio però di restituire la corrispondenza paradigmatica degli aggettivi “dura” e “assidua” oltre che il riuso di “pena,”, fenomeni andati persi nell'apparato a stampa (`<mod>` non marca “pena,” ma soltanto “dura” e “assidua”).

Nella figura 2 `<lem>` riporta l'ultima lezione. E se volessimo pubblicare *C19* all'interno di un'edizione multitestimoniale che ha a testo la *princeps* del Canzoniere (*C21*)? In tal caso, “assidua pena” diventerà un `<rdg>` e `<lem>` riporterà la variante di *C21* (“lunga pena”), non marcata da `@varSeq`, ma da `@wit` che punterà al relativo `<witness>` (fig. 3).

```
<l n="20">Della più <app>
  <lem wit="#P_C21">lunga pena</lem>
  <rdg wit="#MS_C19" varSeq="3">
    <mod change="#strato-2">
      <add>avversa pena,</add>
    </mod>
  </rdg>
  <rdg wit="#MS_C19" varSeq="2">
    <mod change="#strato-2">
      <del>
        <subst>
          <del instant="true">lun<supplied>ga</supplied></del>
          <add instant="true">dura</add></subst> pena
        </del>
      </mod>
    </rdg>
    <rdg wit="#MS_C19" varSeq="1">
      <mod change="#strato-2">
        <del>
          <mod change="#strato-0">
            <subst>
              <del seq="1">dura</del>
              <add seq="2">assidua</add>
            </subst></mod> pena
          </del>
        </mod>
      </rdg>
    </app></l>
```

Figura 3.

Tutte le varianti possono, inoltre, essere classificate attraverso un @type all'interno di <rdg> e <lem>: si potranno quindi definire varianti grafiche ("orthographic"), interpuntive ("punctuational"), lessicali ("substantive") o comunque personalizzate secondo gli interessi del curatore.

Le varianti non paradigmatiche

Se il modulo 12 funziona perfettamente per le correzioni paradigmatiche¹³, si adatta meno ad altri fenomeni correttori, quali aggiunte, cancellature, trasposizioni e inserimenti complessi, che dialogano con più luoghi del testo: queste varianti, che chiameremo non paradigmatiche, hanno bisogno di una codifica a parte, anche perché spesso sono le stesse che vanno incontro a problemi di *overlapping*¹⁴. Dovremmo includere nel nostro modello una marcatura che funziona attraverso il solo <mod> e sottoelementi e, in caso di *overlapping*, con il supporto di elementi vuoti (<addSpan/> al posto di <add>, ecc.): in caso di correzioni sintagmatiche, in luoghi diversi ma dialoganti tra loro, si potrà scegliere un metodo di collegamento delle modifiche (@xml:ide @corresp).

Le varianti immediate

TEI SIG MS ha definito il mark-up per le varianti immediate¹⁵, fornendo a un instant="true". Nel presente sistema, sarà introdotto comunque da <mod>; instant="false" non verrà usato per le varianti tardive, considerate tali di default. Si veda la trascrizione della variante immediata dell'esempio precedente:

```
<rdg varSeq="2">
  <mod change="#strato-2">
    <del>
      <subst>
        <del instant="true">lun<supplied>ga</supplied</del>
        <add instant="true">dura</add</subst> pena
      </del>
    </mod>
  </rdg>
```

Figura 4.

Come si nota, "lun<ga>" è messa in relazione attraverso <subst> con "dura", marcata a sua volta da instant="true". <subst> si usa solo per varianti immediate in rapporto biunivoco con la continuazione del *ductus*, come in questo caso (cambio di aggettivo); quelle abbandonate dall'autore non avranno nessun <subst> o <add>.

Il livello intratestuale

Le varianti alternative sono anch'esse marcate con <rdg>, ma un mark-up specifico le distingue dalle altre, ana="#altVariant", di rimando ad un <interp> nel <teiHeader>. Anche le varianti alternative avranno al loro interno una parte trascrizionale, come finora nel modello: <mod> con @change per lo strato, <add> con @place per la topografia e @seq per la cronologia.

Il livello metatestuale

Le postille d'autore verranno segnate con <note> e connesse all'autore con @resp, così da distinguerle dalle note del curatore¹⁶. <note> avrà al suo interno anche la parte trascrizionale (<mod> e sottoelementi), presentando così un vantaggio sia per il filologo che per l'utente: poiché la descrizione topografica è affidata al blocco <mod>, <note> si troverà nel luogo del testo con cui idealmente dialoga, così da agevolare la trascrizione e la visualizzazione nell'interfaccia.

¹³ I sottoelementi di <app> sono tra loro in rapporto di autoesclusione e dunque rappresentano perfettamente il movimento del testo sull'asse paradigmatico (tutte le sostituzioni in un puntuale luogo del testo).

¹⁴ Si pensi ad esempio all'aggiunta di un verso tra altri due oppure a una cancellatura che interessa contemporaneamente il secondo emistichio di un verso e il primo del successivo. La soluzione di Fiormonte ([2]), cioè marcare il verso con elementi vuoti e non con <1>, sarebbe utile per evitare l'*overlapping* tra marcatura di variante e marcatura di verso, ma non risolverebbe il problema delle varianti non paradigmatiche, come anche delle correzioni che investono il testo base su un asse più sintagmatico che paradigmatico.

¹⁵ Per le varianti immediate e tardive cfr. ([2]:54).

¹⁶ I commenti del curatore, non essendo parte del manoscritto, non hanno nessuna indicazione topografica.

Infine, tutte le note metatestuali potranno essere classificate da @type in autocommenti, indicazioni di struttura, citazioni ("metatextual", "autocomment", "quotation").

4. CONCLUSIONI

Considerando la natura *in itinere* del progetto, sarà proprio la realizzazione dell'edizione digitale di *C19* a perfezionare la codifica, implementando l'uso di <app> per la variantistica d'autore. La presentazione del modello in una sede collegiale come AIUCD vorrebbe anche alimentare la discussione intorno alla standardizzazione della codifica, fondamentale per lo sviluppo di software di visualizzazione scalabili e non circoscritti solamente all'edizione per cui sono stati progettati.

BIBLIOGRAFIA

- [1] Brancato, Dario, Milena Corbellini, Paola Italia, Valentina Pasqual, e Roberta Priore. 2021. «VaSto: un'edizione digitale interdisciplinare». *magazén* 1: 139–69.
- [2] Fiormente, Domenico. 2001. «La representación digital de la génesis del texto. Un caso de estudio». In *el taller del escritor: génesis textual y edición de textos*, a cura di A. Arcocha-Scarcia, J. Lluch-Prats e M.J. Olaziregui, Servicio Editorial del País Vasco, 147–76.
- [3] Italia, Paola. 2019. «Filologia d'autore digitale». *Ecdotica* 1: 203–16.
- [4] ———. 2020. *Editing Duemila*. Salerno.
- [5] Italia, Paola, e Giulia Raboni. 2010. *Che cos'è la filologia d'autore*. Carocci.
- [6] Pierazzo, Elena. 2009. «Digital Genetic Edition». In *Text Editing, Print and the Digital World*, a cura di M. Deegan e K. Sutherland, 169–86. Ashgate.
- [7] ———. 2019. «Edizione documentaria digitale: rinuncia intellettuale o opportunità scientifica?» *Ecdotica* 1: 174–85.
- [8] Pierazzo, Elena, e Julie André. s.d. «Proust Prototype». http://peterstokes.org/elena/proust_prototype/.
- [9] Pierazzo, Elena, e Tiziana Mancinelli. 2020. *Che cos'è un'edizione scientifica digitale*. Carocci.
- [10] Rosselli Del Turco, Roberto, e Chiara Di Pietro. 2019. «La visualizzazione di edizioni digitali con EVT: una soluzione per edizioni diplomatiche e critiche». *Ecdotica* 1: 148–73.
- [11] Saba, Umberto. 1981. *Il Canzoniere 1921*. Edizione critica a cura di Giordano Castellani. Milano.
- [12] TEI Manuscripts Special Interest Group (TEI SIG MS). s.d. «An Encoding Model for Genetic Editions». <https://tei-c.org/Vault/TC/tcw19.html>.
- [13] Text Encoding Initiative (TEI). s.d. «Critical Apparatus». <https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>.
- [14] ———. s.d. «Representation of Primary Sources». Text Encoding Initiative (TEI). <https://tei-c.org/release/doc/tei-p5-doc/en/html/PH.html>.
- [15] Università di Bologna. s.d. «Progetto VaSto». <https://dharc-org.github.io/progetto-vasto>.
- [16] Vodopivec, Silvia. 2016. «Le penne e le matite di Saba. Tracce di volontà autoriale perduta nel Canzoniere (R.P. Ms. 1-18, Biblioteca Civica "A. Hortis" di Trieste).» Trieste, Italy: Università degli Studi di Trieste.
- [17] Zaccarello, Michelangelo. 2019. «Testo, teoria, edizione. Come cambia la filologia nel contesto digitale». *Ecdotica* 1: 186–201.

La svolta empirico-computazionale negli studi culturali e letterari: una nuova scienza della cultura

Fabio Ciotti

Università di Roma "Tor Vergata", Italia – fabio.ciotti@uniroma2.it

ABSTRACT

L'espansione degli approcci di matrice computazionale negli studi letterari configura una vera e propria svolta paradigmatica negli studi letterari e culturali. In questo intervento intendiamo individuare i tratti fondanti di questa svolta, che si articola prioritariamente sul piano metodologico e delineare un quadro degli orizzonti teorici sulla cultura e sulla letteratura che meglio si prestano a fornire il contesto teorico di riferimento di una nuova scienza empirica della cultura.

PAROLE CHIAVE

Cultural analytics, distant reading, biopoetica, poetica cognitiva, studi letterari computazionali, studi quantitativi della cultura, studi culturali.

INTERVENTO

1. INTRODUZIONE

La metafora della svolta è un tratto ricorrente nella retorica degli studi letterari e culturali degli ultimi decenni, e testimonia come questa area di studi sia stata soggetta a numerose fasi di innovazione metodologica e teorica, sulla spinta di pressioni esterne: il rapporto/contatto/ibridazione con altre aree del sapere, più o meno contigue, quali le scienze sociali, le teorie cognitive ed evolucionistiche, le scienze del territorio e dell'ambiente; e interne: il processo ciclico di avvicinamento degli atteggiamenti generali verso i fatti letterari e culturali, come quello tra formalismo e contenutismo, o tra storicismo/contestualismo e autonomia della letteratura.

Tra queste innovazioni di paradigma, quella che ho chiamato la svolta empirico-computazionale, tuttavia, potrebbe avere conseguenze profonde e radicali, poiché essa investe e trasforma il campo in tutti i suoi livelli di articolazione: il piano del dominio di riferimento; il piano del metodo; il piano della teoria; e il piano della sociologia della ricerca.

2. IL CONTESTO TEORICO DISCIPLINARE DELLA SVOLTA

Le scelte terminologiche che denotano le fasi evolutive delle discipline scientifiche e dei campi di ricerca sono sempre parziali e rischiano di ridurre fenomeni complessi e plurali a una etichetta monodimensionale. Per limitare almeno in parte questo schiacciamento iniziamo con una mappatura del nostro campo discorsivo. Ovviamente l'aggettivo computazionale definisce la dimensione principale di tale campo, poiché esso si riferisce al fatto che i metodi di rappresentazione, modellazione e analisi dei testi sono metodi computazionali, intesi come processi di natura formale e algoritmica implementati in forma di programmi e dati. Volutamente abbiamo evitato il termine digitale, che pure appare da ormai venti anni come parte determinate dell'etichetta disciplinare di Digital Humanities ([13]), con cui si denomina il campo esteso dei vari approcci alle scienze umanistiche che in qualche modo hanno a che fare con l'informatica. Digitale infatti pertiene alla sfera della rappresentazione, che è fatto contingente, poiché come noto l'applicabilità dei processi computazionali è invariante rispetto all'insieme dei simboli usati nella computazione, purché questo sia discreto e finito. Ma questa sottigliezza teorica potrebbe essere tralasciata – lo stesso Turing ha usato 'digital' in un senso coestensivo a 'computazionale' in alcuni suoi scritti ([15]) – se non fosse che la parola 'digitale' è ormai inflazionata e che nel campo delle Digital Humanities sono oggi inclusi approcci e studi che in fondo di intrinsecamente e strettamente computazionale hanno poco. Poco male, si intenda, la grande tenda delle DH è stata giustamente accogliente per motivi tattici, potremmo dire, ma questa fluidità ai limiti dell'indeterminazione, come già osservava ([4]), ha ormai raggiunto i limiti della sua spinta espansiva, ed è oggi opportuno avviare una fase di ridefinizione del campo umanistico digitale, o per continuare a usare la metafora della tenda, iniziare a fare qualche compartimento al suo interno, per consolidare il tetto comune.

Tuttavia, per quanto attiene il tema di questo intervento, la scelta della determinazione di computazionale non esaurisce l'argomentazione. La svolta, infatti, ha ulteriori caratterizzazioni teoriche e metodologiche: la predilezione per i metodi quantitativi; l'applicazione di tecniche di elaborazione statistico/probabilistica; la considerazione dei fatti culturali e letterari come insiemi di dati estesi, diversificati e complessi, sia in termini sincronici sia in termini diacronici. La

digitalizzazione su vasta scala del patrimonio culturale, insomma, contribuisce alla costituzione dei cosiddetti *Big data*, il prodotto della transizione digitale in tutte le sfere dell'agire umano degli ultimi decenni. Una quantità di dati digitali enorme, che rispecchia i processi e i fenomeni naturali e sociali che li hanno originati, e che possono essere studiati solo su vasta scala. In questo senso, la nostra svolta si colloca in un contesto scientifico e culturale più vasto ed è sospinta da indirizzi di ricerca paralleli negli studi tecnico/scientifici: ci riferiamo alla recente emergenza di quella che stata definita la *data science*, un campo di studi che a sua volta si sostanzia sia di metodi e tecniche analitiche, come il *data mining* e il *machine learning*, sia di un assunto epistemologico: l'idea che la produzione di conoscenza possa fondarsi prioritariamente sull'analisi di grandi collezioni di dati, che alcuni spingono fino al limite estremo – e non condivisibile a nostro parere – di negare ogni ruolo alla teoria e alla modellizzazione ([1]).

La traslazione di queste tendenze nella sfera degli studi culturali e letterari ha favorito l'emergenza di proposte teoriche metodologiche e disciplinari come quella della *Cultural Analytics* ([9]) e del *Distant reading* ([10];[18]). Esse identificano una costellazione di teorie, metodologie, pratiche analitiche e sperimentazioni convergenti, che sostanziano una svolta profonda nel modo di capire i fatti culturali in generale e quelli letterari in senso stretto.

Un ultimo chiarimento è opportuno sull'attributo "empirico" che ho adottato nel titolo. Sebbene l'approccio ermeneutico e idiografico sia stato di gran lunga prevalente, nella storia degli studi letterari non sono mancati fasi e correnti che hanno rivendicato l'importanza di approcci empirici nella ricerca. Basti ricordare gli studi di impianto sociologico quantitativo nella tradizione della sociologia della letteratura e degli studi sul libro; oppure alle indagini di impianto psicologico nella tradizione degli *empirical literary studies* ([8];[11]); per non dimenticare alcune tendenze dei Cultural Studies della scuola di Birmingham ([16]). E d'altra parte, non si può certo dire che alcuni ambiti degli studi letterari non abbiano solide basi su evidenze materiali, si pensi alla critica del testo. Da questo punto di vista l'enfasi che segnala la scelta del termine "svolta" nel titolo di questo paper potrebbe essere considerata eccessiva se non abusiva. Tuttavia, a parte le ovvia difesa secondo cui a cercare bene si trovano precedenti e antesignani per ogni concetto e artefatto umano, vorrei osservare che in gran parte di queste precedenti correnti 'empiriche' ciò che alla fine rimaneva fuori dallo studio era proprio il testo o meglio i testi *in sé*, intesi nella loro materialità di oggetti linguistici. A titolo di esempio, negli *empirical literary studies* la maggior parte delle analisi vertono sull'adozione dei metodi della psicologia sperimentale o delle scienze cognitive all'analisi della lettura/ricezione del testo e dei suoi effetti nei lettori empirici. Ora, è indubbio che il paradigma che propongo in questa sede erediti, includa e valorizzi queste tradizioni di studi, finora liminari rispetto al *mainstream* degli studi letterari. Ma mi pare di poter dire che per la prima volta, grazie alla convergenza tra digitalizzazione di massa dei prodotti culturali (*legacy* e no) e *data analytics/machine learning*, siamo in grado di considerare i testi, la letteratura nel suo insieme, e la cultura come un fenomeno empirico e dunque indagabile con approcci quantitativi ed empirici.

3. UN NUOVO PARADIGMA METODOLOGICO

La mappatura del contesto appena delineata ci permette di enucleare i tratti peculiari della svolta computazionale, che non consiste solo nell'uso del computer, nell'adozione di metodi statistici, o di uno sguardo empirico sui fatti culturali. Presi singolarmente, infatti, ciascuna di queste pratiche di ricerca ha una storia lunga, con precedenti e precursori brillanti ([7]), ([17]). Ma dalla loro convergenza emerge a mio avviso un paradigma scientifico che ha tratti fortemente innovativi rispetto alla tradizione degli studi letterari, caratterizzato dai seguenti elementi concettuali:

- l'adozione di un approccio empirico verso i fatti culturali/letterari, basato su pratiche osservative e (latamente) sperimentali;
- l'adozione di un approccio quantitativo nello studio dei fenomeni culturali;
- il cambiamento di scala nell'osservazione e analisi dei fenomeni: dal singolo testo e dall'autore individuale, la scala molecolare, si passa al macrofenomeno, la scala molare;
- lo spostamento sul piano del metodo dall'interpretazione alla spiegazione causale basata sulla evidenza quantitativa e l'analisi statistico probabilistica.

Ciascuno di questi aspetti necessita di una trattazione approfondita e critica. In questa sede ci limitiamo a due osservazioni. In primo luogo, l'adozione di questo paradigma non comporta una rinuncia alla prospettiva storica, quanto piuttosto un cambiamento della scala temporale e la necessità di proiettare la dimensione storica sul piano della completezza e della rappresentatività dei dati, operazione, questa, non priva di difficoltà come ha osservato ([3]). In secondo luogo, occorre intendersi sul significato della parola interpretazione. In senso lato, anche i dati (e il prodotto della loro elaborazione) sono 'interpretati', così come è vero che la critica e la storiografia letteraria propongono (anche) spiegazioni. Tuttavia, quando parliamo di interpretazione del testo, e ancor più del testo letterario, ci riferiamo a un processo che è sempre contingente e situato, determinato dalla soggettività dell'interprete e dalla storicità della comunità interpretante, che ha lo scopo di arrivare alla comprensione ([12]). La spiegazione invece è basata su dati che sono esterni al soggetto, caratteristiche

misurabili i cui valori e gradienti sono indipendenti dall'osservatore (il fatto che un osservatore definisca a priori come procedere alla "datificazione" è qui irrilevante, poiché tali scelte, una volta fatte, sono sempre potenzialmente pubbliche e ripetibili) e ha lo scopo di identificare relazioni causali e di produrre generalizzazioni.

4. CONVERGENZE TEORICHE

Quanto abbiamo detto delinea i razionali metodologici di una nuova scienza dalla cultura. È opportuno chiarire che questa scienza della cultura non è sostituiva degli studi culturali e letterari tradizionali poiché si pone a un diverso livello di astrazione, esattamente come la biologia molecolare non ha soppiantato l'anatomia comparata o la fisiologia. Certo, resta da approfondire teoricamente come e se il livello delle spiegazioni macro possa e debba interagire con quello delle interpretazioni e analisi ermeneutiche del livello micro. Questione che è in relazione con un altro problema metateorico: quali framework teorici sono adeguati agli approcci empirico computazionali? La mia idea è che il nuovo paradigma metodologico non possa recuperare e includere tutto il patrimonio teorico della tradizione (già di per sé assai variegato). Tale patrimonio teorico in gran parte si fonda sul *close reading* e sul metodo ermeneutico, che sono atti soggettivi e fortemente connessi con il giudizio estetico.

Su quali basi teoriche, allora, possiamo costruire una scienza della cultura e della letteratura empirica e computazionale? Una possibile direzione da esplorare, come suggerisce Underwood ([6]) è che essa rientri all'interno del più ampio campo delle scienze sociali, una opzione per cui ci sono molte buone ragioni. Ma penso che siano altrettanto rilevanti gli orizzonti teorici forniti dagli approcci cognitivi e bio-evoluzionistici alla letteratura e degli studi sull'evoluzione culturale.

La poetica/narratologia cognitiva ([2]) e gli studi letterari bio-evoluzionistici ([5]) sono stati due delle più interessanti correnti teoriche in campo letterario degli ultimi decenni e sono ormai campi di indagine consolidati. Con diverse gradazioni, a seconda degli autori, hanno sostenuto l'introduzione di una metodologia scientifica nello studio della letteratura, cercando approfondimenti metodologici e teorici nelle scienze cognitive e nella psicologia evolutiva. Non possiamo addentrarci nei dettagli delle varie aree di ricerca che hanno caratterizzato questo campo, e dei diversi approcci proposti, e vedere come i tipi specifici di problemi studiati nella poetica cognitiva e nella narratologia potrebbero essere analizzati attraverso metodi computazionali. Ciò che è più interessante dal punto di vista della mia tesi è che il dibattito sulla legittimità e accettabilità degli approcci cognitivi negli studi letterari ha determinato una discussione sul ruolo dell'interpretazione che ha molte similitudini con gli argomenti che ho proposto in questo intervento.

L'altro campo scientifico in cui gli studi letterari di matrice computazionale possono trovare un quadro teorico è quello dell'evoluzione culturale. Questo campo di studi mira a fornire una spiegazione naturalista ed empirica della natura e dell'evoluzione della cultura, adottando ampiamente la modellazione matematico/statistica. Uno dei fondamenti teorici dell'evoluzione culturale è l'adozione del *population thinking*, tratto dalla biologia evolutiva (secondo l'interpretazione di Ernest Mayr della teoria di Darwin) e dalla genetica delle popolazioni, e la sua applicazione ai fenomeni culturali ([14]). La letteratura fa parte della sfera culturale, quindi può essere considerata una popolazione di oggetti individuali (i testi) la cui descrizione un dato stato (sincronica) e la sua evoluzione (diacronica) è possibile attraverso l'analisi statistica.

Nello spazio delle teorie, in conclusione, la teoria bio-cognitiva della letteratura e dell'evoluzione culturale possono offrire la migliore opzione per usufruire dei vantaggi dei metodi computazionali e per comprendere quei fenomeni culturali che sono fuori dalla portata dell'approccio ermeneutico.

BIBLIOGRAFIA

- [1] Anderson, Chris. 2008. «The End of Theory: The Data Deluge Makes the Scientific Method Obsolete». *Wired* (blog). 23 giugno 2008. <https://www.wired.com/2008/06/pb-theory/>.
- [2] Bernini, Marco, e Marco Caracciolo. 2013. *Letteratura e scienze cognitive*. 1a edizione. Roma: Carocci Editore.
- [3] Bode, Kat. 2018. *A world of fiction: digital collections and the future of literary history*. Ann Arbor, MI: University of Michigan Press.
- [4] Ciotti, Fabio. 2019. «Oltre la galassia delle Digital Humanities: per la costituzione di una disciplina di Informatica Umanistica». In *Didattica e ricerca al tempo delle Digital Humanities. Book of Abstracts*, 67–72.
- [5] Cometa, Michele. 2018. *Letteratura e darwinismo: introduzione alla biopoetica*. 1a edizione. Roma: Carocci editore.
- [6] English, James F., e Ted Underwood. 2016. «Shifting Scales: Between Literature and Social Science». *Modern Language Quarterly* 77 (3): 277–95. <https://doi.org/10.1215/00267929-3570612>.
- [7] Hoover, David L. 2013. «Textual Analysis». In *Literary Studies in the Digital Age*, K. M. Price e R. Siemens. i Modern Language Association of America.
- [8] Kuiken, Don, e Arthur M. Jacobs. 2021. *Handbook of Empirical Literary Studies*. De Gruyter.
- [9] Manovich, Lev. 2020. *Cultural analytics*. Cambridge, Massachusetts: The MIT Press.
- [10] Moretti, Franco. 2013. *Distant Reading*. London: Verso.

- [11] Nemesio, Aldo. 2014. «Le ragioni della ricerca empirica sul testo». In *Dalla parte dell'uomo. L'economia nella letteratura e nelle scienze umane*. CoSMo Comp. Stud. Mod.
- [12] Ricoeur, Paul. 1976. *Interpretation Theory: Discourse and the Surplus of Meaning*. Texas Christian University Press.
- [13] Schreibman, Susan, Ray Siemens, e John Unsworth. 2004. *A companion to digital humanities*. Malden, Mass.: Blackwell Pub.
- [14] Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach*. Oxford: Basil Blackwell.
- [15] Turing, Alan. 1950. «Computing machinery and intelligence». *Mind* 59 (236): 433–60.
- [16] Turner, Graeme. 2003. *British cultural studies: an introduction*. 3rd ed. London; New York: Routledge.
- [17] Underwood, Ted. 2017. «A Genealogy of Distant Reading». *Digital Humanities Quarterly* 11 (2).
- [18] ———. 2019. *Distant horizons: digital evidence and literary change*. Chicago: The University of Chicago Press.

Poster

Wordforms and Meanings: an Updated Report on the LiLa Project

Marco Passarotti, Flavio Massimiliano Cecchini, Eleonora Litta, Francesco Mambrini, Giovanni Moretti,
Giulia Pedonese, Matteo Pellegrini, Paolo Ruffolo, Rachele Sprugnoli, Marinella Testori
Università Cattolica del Sacro Cuore di Milano, Italy - {nome.cognome}@unicatt.it

ABSTRACT

This contribution presents the current status of the ERC project “LiLa: Linking Latin”, the main objective of which is to connect and exploit the wealth of existing linguistic resources for Latin by making them interoperable, through the creation of a Knowledge Base following Linked Data standards. We describe the textual and lexical resources linked to the Knowledge Base and the ways in which it is possible to query and explore them.

KEYWORDS

Linguistic resources, Latin, Semantic Web.

POSTER

1. INTRODUCTION

Linguistic resources are machine-readable collections of language data and descriptions. Thanks to international efforts, several resources as well as Natural Language Processing (NLP) tools are currently available for ancient languages, including Latin. Linguistic resources are usually classified in two main categories depending on the kind of content they contain: (a) textual resources, such as written corpora, featuring either partial or full texts which may differ in genre, author or time period and (b) lexical resources like lexica, dictionaries and terminological databases providing information on lexical items for one or more languages including definitions, translations and morphological properties.

However, despite the increase in their quantity and coverage, linguistic data and metadata today are scattered in isolated resources, preventing users (in particular those from the humanities, such as historians, philologists, archaeologists and literary scholars) from honing both their individual and joint potential across platforms.

A current approach to making linguistic resources interact takes up Linked Data principles ([2];[3]), according to which data in the Semantic Web ([1]) are interlinked through connections that can be semantically queried so that the structure of web data can better answer to the needs of users.

With this in mind, the “LiLa: Linking Latin project” (2018-2023: <https://lila-erc.eu>) was awarded funding from the European Research Council (ERC) to build a Knowledge Base (KB) of linguistic resources for Latin following the Linked Data paradigm: the KB is a collection of diverse, interlinked data sets described with the same vocabulary of knowledge description that uses common data categories and ontologies ([10]). Given the presence and role played by lemmatization in various linguistic resources and the good accuracy rates achieved by state-of-the-art lemmatizers for Latin (up to 95.30% ([7]))¹, LiLa uses the lemma as the most productive interface between lexical resources, annotated corpora and NLP tools. Accordingly, the LiLa KB is highly lexically based, grounding on the simple postulation that strikes a good balance between feasibility and granularity: textual resources are made of (occurrences of) words, lexical resources describe properties of words, and NLP tools process words. This granted, the heart of the LiLa KB consists of a large collection of Latin lemmas called Lemma Bank, currently comprising of more than 130,000 canonical forms: interoperability is attained by linking all those entries in lexical resources and tokens in corpora that point to the same lemma. The linguistic properties of the Latin lemmas in LiLa are expressed as RDF triples using the LiLa ontology semantics.

¹ Such high rates of automatic lemmatization of Latin should be taken with a grain of salt. Indeed, performances of stochastic NLP tools heavily depend on the training set on which their models are built, and so decrease when they are applied to out-of-domain texts. This problem is particularly challenging for Latin owing to its wide diachrony (spanning two millennia), genre diversity (ranging from literary to philosophical, historical and documentary texts) and diatopy (Europe and beyond). For the state of the art in automatic lemmatization and PoS tagging for Latin, see the results of the first edition of EvaLatin, a campaign devoted to the evaluation of NLP tools for Latin ([12]).

This abstract introduces the current status of the LiLa KB, focussing on the textual and lexical resources that were interlinked so far thanks to their association to the collection of lemmas of LiLa².

2. RESOURCES

In this section we provide a brief description of the resources linked so far via the LiLa KB covering different linguistic aspects (from morphology to syntax and semantics) and different time periods (from Late Antiquity to the Middle Ages) of Latin linguistic material. More specifically, the textual resources currently available are the Index Thomisticus Treebank (ITTB) containing the works by Thomas Aquinas, the corpus of Latin texts by, or disputedly attributed to, Dante Alighieri (UDante), the text of the comedy “Querolus sive Aulularia” and the eighth chapter of the “Liber Abaci”, a mathematical treatise by Fibonacci. All these corpora are annotated following the Universal Dependencies framework ([4]): the last two resources are annotated with Part-of-Speech tags and lemmas whereas ITTB and UDante also contain syntactic information.

For what lexical resources are concerned, the LiLa KB currently contains: a collection of Proto-Italic and Proto-Indo-European reconstructed forms taken from the “Etymological Dictionary of Latin and the other Italic Languages” ([5]), the LatinAffectus sentiment lexicon, a collection of Ancient Greek loanwords in the Latin language extracted from the “Index Graecorum vocabulorum in linguam Latinam translatorum quaestiunculis auctus” ([11]), around 1800 manually checked entries of the Latin WordNet mapped onto Princeton WordNet 3.0, a valency lexicon for Latin and a derivational morphology lexicon. In order to achieve interoperability, all these resources are modeled and described using ontologies such as Ontolex ([9]) and encoded in a graph-based data structure in RDF.

3. QUERYING THE KNOWLEDGE BASE

At the time of writing, there are two ways for querying the LiLa KB: through the Query Interface (<https://lila-erc.eu/query/>) or using the SPARQL endpoint. The Query Interface is a user-friendly graphical web application for searching the lemmas in the Lemma Bank, suitable for those unfamiliar with SPARQL. Users can search for a specific lemma or part of it or compose their own query by dragging and dropping any combination of query modules: each query module allows to filter the results with respect to a grammatical or morphological feature (such as gender, PoS, presence of a suffix) by choosing an option from a drop-down menu. Results can be saved as a CSV file. Alternatively, it is possible to copy the underlying SPARQL query and view the complete lemma description or the corresponding graph representation. Figure 1 shows a query retrieving all common nouns with masculine gender having the suffix -(t)or: this query has 1,528 results and the first three lemmas in alphabetical order are *abactor* “a cattle-stealer” and *abbreviator* “epitomist”.

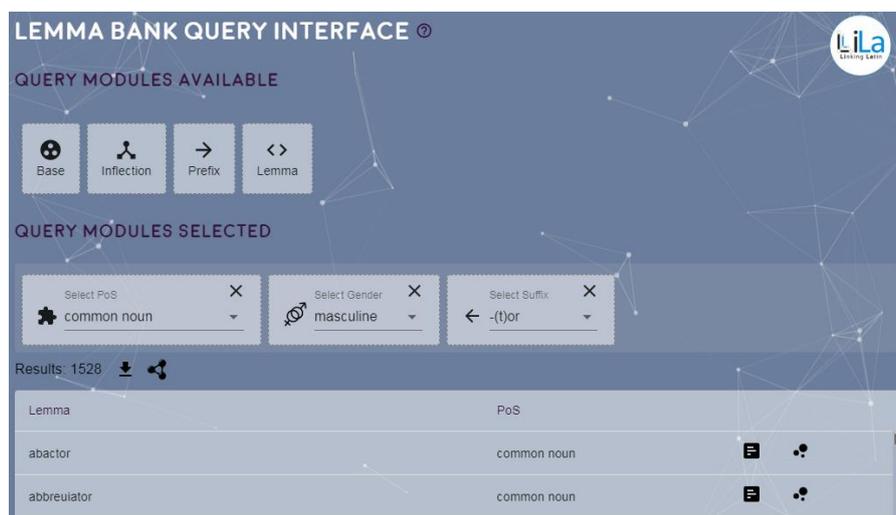


Figure 1. Screenshot of the Lemma Bank Query Interface.

Via the SPARQL endpoint (<https://lila-erc.eu/sparql/>) it is instead possible to access the ever-growing collection of connected resources beyond the Lemma Bank and perform more complex searches. We release and constantly update a set of queries in a dedicated GitHub repository to facilitate the use of the endpoint: <https://github.com/CIRCSE/SPARQL->

² Both the collection of lemmas and the source data of the resources linked to LiLa (together with their TTL files, which provide the RDF triples) are freely available from the GitHub page of the host institution’s CIRCSE research center: <https://github.com/CIRCSE>.

[queries](#). For example, the query UDante-sentiment.rq in the repository works on 3 different interlinked resources, i.e., LatinAffectus, the Lemma Bank and UDante to retrieve all lemmas in UDante that appears in the sentiment lexicon with a negative polarity and count the total number of occurrences per lemma. This query results in the following top 5 lemmas with a negative sentiment: *peccatum* “sin” (17 occurrences), *litigium* “quarrel” (16), *mors* “death” (15), *malus* “bad” (12), *iniura* “injurious” (11).

4. UPCOMING RESOURCES

We are currently working on modelling and linking the two following resources:

1. the bilingual “Latin Dictionary” curated by Ch. T. Lewis and Ch. Short and published by Harper and Oxford University Press in 1879 ([8]).
2. the LASLA corpus developed by the homonymous laboratory in Liège, Belgium, which currently includes more than 150 texts from around 20 authors for a total of approximately 1,700,000 words ([6]).

REFERENCES

- [1] Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. “The Semantic Web.” *Scientific American* 284 (5): 34–43.
- [2] Chiarcos, Christian, Philipp Cimiano, Thierry Declerck, and John. P. McCrae. 2013. “Linguistic Linked Open Data (Llod). Introduction and Overview.” In *Proceedings of the 2nd Workshop on Linked Data in Linguistics: Representing and Linking Lexicons, Terminologies and Other Language Data*.
- [3] Chiarcos, Christian, Sebastian Nordhoff, and Sebastian Hellmann. 2012. *Linked Data in Linguistics*. Heidelberg: Springer.
- [4] De Marneffe, Marie-Catherine, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. “Universal Dependencies.” *Computational Linguistics* 47 (2): 255–308.
- [5] De Vaan, Michiel. 2008. *Etymological Dictionary of Latin and the Other Italic Languages*. Vol. 7. Boston: Brill, Leiden.
- [6] Denooz, Joseph. 2007. “Opera Latina: Le Nouveau Site Internet Du Lasla.” *Journal of Latin Linguistics* 9 (3): 21–34.
- [7] Eger, Steffen, Tim Vor der Brück, and Alexander Mehler. 2015. “Lexicon-Assisted Tagging and Lemmatization in Latin: A Comparison of Six Taggers and Two Lemmatization Methods.” In *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*.
- [8] Lewis, Charlton Thomas. 1884. *Harpers’ Latin Dictionary: A New Latin Dictionary Founded on the Translation of Freund’s Latin-German Lexicon*. Harper & brothers.
- [9] McCrae, John. P., Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. “The Ontolex-Lemon Model: Development and Applications.” In *Proceedings of ELex 2017 Conference*.
- [10] Passarotti, Marco, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo, and Rachele Sprugnoli. 2020. “Interlinking through Lemmas. the Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin.” *Studi e Saggi Linguistici* 58 (1): 177–212.
- [11] Saalfeld, Alexander. 1874. *Index Graecorum Vocabulorum in Linguam Latinam Translatorum Quaestiunculis Auctus*. F. Berggold.
- [12] Sprugnoli, Rachele, Marco Passarotti, Flavio Massimiliano Cecchini, and Matteo Pellegrini. 2020. “Overview of the Evalatin 2020 Evaluation Campaign.” In *Proceedings of LT4HALA 2020-1st Workshop on Language Technologies for Historical and Ancient Languages*.

From Close to Distant Reading. Towards the Computational Analysis of “Liber Abbaci”

Letizia Ricci¹, Francesco Grotto², Margherita Fantoli³, Rachele Sprugnoli⁴, Marco Passarotti⁴,
Enrica Salvatori¹, Maria Simi¹

¹Università degli Studi di Pisa, Italy, l.ricci29@studenti.unipi.it - {maria.simi,enrica.salvatori}@unipi.it

²Scuola Normale Superiore di Pisa, Italy, francesco.grotto1@sns.it

³University of Leuven, Belgium, margherita.fantoli@kuleuven.be

⁴Università Cattolica del Sacro Cuore di Milano, Italy, {rachele.sprugnoli,marco.passarotti}@unicatt.it

ABSTRACT

This contribution presents the first steps towards the analysis of Leonardo Fibonacci's *Liber Abbaci* using computational linguistics methods. The work is currently carried out in the context of a joint research project between the Tuscany Region and the University of Pisa with the help of an interdisciplinary team.

KEYWORDS

Computational Linguistics, annotation, text encoding, Latin.

POSTER

1. INTRODUCTION

Leonardo Fibonacci's *Liber Abbaci* is a weighty medieval treatise on arithmetic and algebra that had a decisive influence in the development of Western mathematics. Traditional reading of the text has never been easy in the past and it is not easy now either. The characteristics of the work have, in fact, delayed its critical print edition until very recently. And, looking at the new editorial format and its price - 17 x 24 cm, cxviii-824 pp. 22 plates f.t. color pp., slipcase, Indian paper, silk binding gilded impressions, € 300 - the volume edited by Enrico Giusti ([2]) certainly is not "manageable" and is clearly addressed to an extremely small niche market. The work itself is full-bodied, complex, and presents some problems in the correct understanding and contextualization of terms related to the world of medieval Mediterranean trade, used by the author in illustrating mathematical problems. For this reason, *Liber Abbaci* has been studied more by mathematicians and historians of Science than by other types of humanists ([5];[7]). In 2018 a joint research project between the Tuscany Region and the University of Pisa (p.i. Pier Daniele Napolitani, University of Pisa) has started with the aim of transforming the critical print edition into a completely searchable digital edition, in order to recover the treasure of linguistic, mathematical and historical information that the work contains and therefore to facilitate the access to its content by different users. Within this overall project, an in-depth study was undertaken on the application of computational linguistics methodologies to the *Liber Abbaci* with the aim of developing systems for the automatic extraction of morphosyntactic and semantic information. Due to the linguistic peculiarities of the text, off-the-shelf tools cannot be used without a considerable loss in the performance (see Sec. 2.1). Thus, manually created high-quality annotated data are needed. In other words, in our work we start from the linguistic annotation of a chapter of the *Liber Abbaci*, relying on digital tools for encoding the results of a critical close reading of the text. As future work, we will develop models based on these data to facilitate a comprehensive approach of such a large-scale masterpiece, that will be interrogated with distant reading methods for the first time.

2. COMPUTATIONAL ANALYSIS

The *Liber Abbaci* is a complex work containing a large variety of topics. While the first 7 chapters discuss mathematical operations, chapters 8-11 have an empirical approach, describing commercial practices and monetary topics. The final chapters (12-15) go back to more abstract mathematical problems. The information contained in the chapters on commercial practices is valuable not only for those interested in the contribution of the work to the history of science, but also as a testimony of the history of economics and trade practices. For this reason, we decided to start our work from chapter VIII using that text for our pilot annotations at both the morphosyntactic and the semantic level. Chapter VIII is made up of 29,858 tokens, corresponding to about 10% of the total length of the book.

2.1 MORPHO-SYNTACTIC ANALYSIS

Beside its scientific interest, *Liber Abbaci* features a very peculiar lexicon, not often represented in the currently available linguistically annotated corpora for Latin. In order to fill this gap, we manually performed tokenization, sentence splitting, Part-of-Speech (PoS) tagging and lemmatization of chapter VIII following the Universal Dependencies framework ([9]). During the annotation, we had to deal with several complex linguistic peculiarities of the text that are typical of Medieval Latin such as monophthongization, the presence of analytical verb forms and a very limited use of enclitics. The greatest difficulties, however, concerned the annotation of units of measurement, names of coins, toponyms and arabisms often not even lemmatized in Medieval Latin dictionaries. The annotation was performed by a master’s degree student in Classical languages supported by experts in Latin linguistics and computational linguistics. The Inter-Annotator Agreement was calculated on 30 sentences (1,010 tokens), with the participation of a second scholar, and we registered an almost perfect agreement with a Cohen’s kappa of 0.97 for lemmatization and 0.94 for PoS tagging. The resulting dataset is freely available online¹ and has been used to evaluate current available automatic models for the processing of Latin. More specifically, we tested the accuracy of five UDPipe ([12]) models with respect to our gold standard: 1) EvaLatin2020, trained on classical texts in prose released for the EvaLatin evaluation campaign ([11]); 2) ITTB, trained on medieval texts of Thomas Aquinas ([4]); 3) LLCT, trained on Early Medieval charters written in Tuscany ([3]); 4) Proiel, trained on selections classical texts plus the Vulgate New Testament translation ([6]); 5) Perseus, trained on classical texts in prose and poetry ([1]). The scores, reported in Tab. 1, clearly show that current models are not good enough to process the Latin of Fibonacci: indeed, the best participating system at the EvaLatin 2020 achieved an accuracy of 96,2% for lemmatization and 96,7% for PoS tagging on the corresponding test set. The specific domain of the text has a negative impact on both lemmatization and PoS tagging: for example, chapter VIII contains a high frequency of lemmas not present in the training data of the model (>50%). Moreover, not all training data follow the latest version of the Universal Dependencies guidelines (v 2.8) causing some inconsistency of the annotations.

Model	Lemma	PoS
LLCT	68.8	82.8
EvaLatin2020	63.6	81.9
Perseus	67.5	78.4
ITTB	65.6	77.1
Proiel	60.2	51.6

Table 1. Accuracy of UDPipe models tested on chapter VIII.

The lemmatized text of chapter VIII has been linked to the Knowledge Base of interoperable linguistic resources for Latin developed by the ERC project “Lila: Linking Latin” ([8]). Thanks to the linking, our dataset becomes part of an interoperable ecosystem made of resources of different kinds² that can be queried using the SPARQL endpoint of LiLa³.

2.2 SEMANTIC ANALYSIS

We performed a lexical-semantic analysis of chapter VIII in order to identify and classify single terms and multi-token expressions specific to the domain of trade and commerce, so as to facilitate the search within the text. To do so, we have adopted the UCREL Semantic Analysis tagset⁴ ([10]), which provides a set of hierarchical semantic tags. Among those tags, we have decided to select only those relevant to the research objective of the project. From the 21 major discourse fields identified by the original UCREL set, we considered 7 of them, in particular I: money and commerce in industry, M: movement, location, travel and transport, N: numbers and measurement, O: substances, materials, objects and equipment, S: social actions, states and processes, T: time, Z: names and grammar. Each field has specific tags and each tag has an

¹ <http://dialogo.di.unipi.it/LiberAbaci/>.

² <https://lila-erc.eu/data/corpora/CorpusFibonacci/id/corpus/Liber%20Abbaci>.

³ <https://lila-erc.eu/sparql/>.

⁴ <http://ucrel.lancs.ac.uk/usas/>.

identification code with a short definition; for example for coins we have the generic tag I1 Money generally which is divided into I1.1 Money: Affluence, I1.2 Money: Debts, I1.3 Money: Price.

As an annotation tool, we have chosen Catma (see Fig. 1), a flexible and user-friendly online application. Catma allows to work on shared projects and to create a tagset with a hierarchical set of labels. It also provides tools for searching and analyzing the annotated text. In our case we started with the preliminary annotation of chapter VIII of *Liber Abbaci* using the tags briefly mentioned above. The most used tags are: I1 Money generally (frequency 872), with which terms of various types of coins are annotated; N3 Measurement (972), which annotates various units of measurement, the most common being the units of weight; I2 Business (291), which indicates terms referring to commerce; Z2 Geographical names (281) usually occurring with units of measurement or coins, for example *rotuli gerovi* referring to the unit of weight *rotoli* with the value used in Genoa. The text does not contain some tags, such as I3 Work and employment, S2 People, T1 Time, Z1 Personal names.

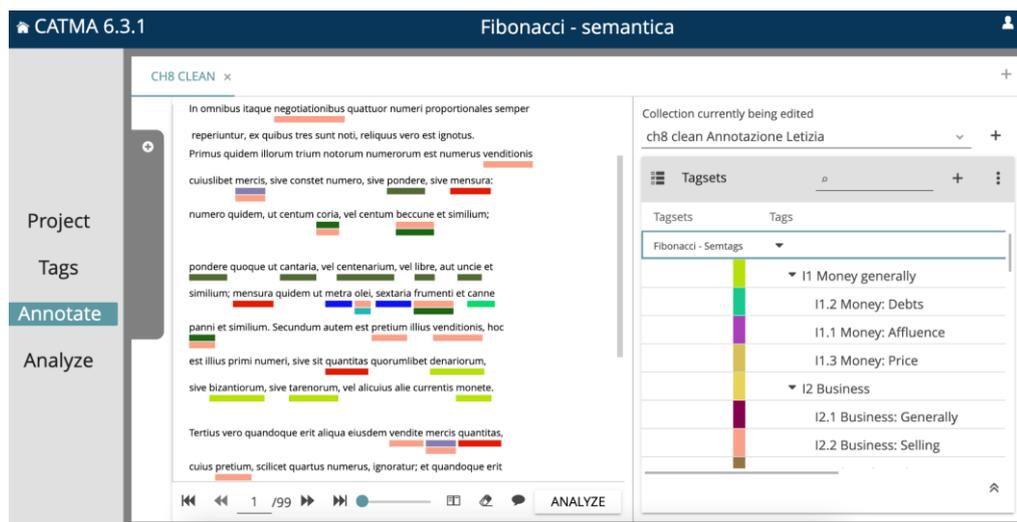


Fig.1 - Catma interface with annotation of chapter VIII.

REFERENCES

- [1] Bamman, David, and Gregory Crane. 2011. "The Ancient Greek and Latin Dependency Treebanks." In *Language Technology for Cultural Heritage*, Caroline Sporleder, Antal van den Bosch, Kalliopi Zervanou, 79–98.
- [2] Bigolli Pisani, Leonardo vulgo Fibonacci. 2020. *Liber Abbaci*. Edited by Enrico Giusti and Paolo D'Alessandro. Firenze: Olschki.
- [3] Cecchini, Flavio Massimiliano, Timo Korhakangas, and Marco Carlo Passarotti. 2020. "A New Latin Treebank for Universal Dependencies: Charters between Ancient Latin and Romance Languages." In *Proceedings of the Twelfth International Conference on Language Resources and Evaluation*. Marseille, France: European Language Resources Association (ELRA).
- [4] Cecchini, Flavio Massimiliano, Marco Passarotti, Paola Marongiu, and Daniel Zeman. 2018. "Challenges in Converting the Index Thomisticus Treebank into Universal Dependencies." In *Proceedings of the Second Workshop on Universal Dependencies*, 27–36.
- [5] Ciocci, Argante, and Enrico Giusti. 2018. "The Twelfth Chapter of Fibonacci's Liber Abaci in Its 1202 Version, Bollettino Di Storia Delle Scienze Matematiche." *Nuncius* 1 (33): 137–39.
- [6] Dag, Trygve, Truslew Haug, and Marius L. Jøhndal. 2008. "Creating a Parallel Treebank of the Old Indo-European Bible Translations." In *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data*, edited by Caroline Sporleder and Kiril Ribarov, 27–34.
- [7] Franci, Raffaella. 2002. "Il Liber Abaci Di Leonardo Fibonacci 1202-2002." *Bollettino Dell'Unione Matematica Italiana* 5 (A.2): 293–328.
- [8] Passarotti, Marco, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo, and Rachele Sprugnoli. 2020. "Interlinking through Lemmas. the Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin." *Studi e Saggi Linguistici* 58 (1): 177–212.
- [9] Petrov, Slav, Dipanjan Das, and Ryan McDonald. 2012. "Universal Part-of-Speech Tagset." In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, 2089–96. Istanbul, Turkey.
- [10] Piao, Scott, Dawn Archer, Olga Mudraya, Paul Rayson, Roger Garside, Tony McEnery, and Andrew Wilson. 2005. "A Large Semantic Lexicon for Corpus Annotation." In *Proceedings from the Corpus Linguistics Conference Series On-Line e-Journal*. Vol. 1. Birmingham, UK.

- [11] Sprugnoli, Rachele, Marco Passarotti, Flavio Massimiliano Cecchini, and Matteo Pellegrini. 2020. “Overview of the EvaLatin 2020 Evaluation Campaign.” In *Proceedings of LT4HALA 2020-1st Workshop on Language Technologies for Historical and Ancient Languages*.
- [12] Straka, Milan, and Jana Straková. 2017. “Tokenizing, POS Tagging, Lemmatizing and Parsing Ud 2.0 with Udpipes.” In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 88–99. Vancouver, Canada.

Citizen Humanities in Tyrol:

A case study on historical newspapers

Greta Franzini¹, Egon W. Stemle¹, Verena Lyding¹, Andrea Abel¹

Johannes Andresen², Karin Pircher²

Silvia Gstrein³, Barbara Laner³, Johanna Walcher³, Maritta Horwath³, Christian Koessler³

¹Eurac Research, Italy, {name.surname}@eurac.edu;

²Landesbibliothek Dr. Friedrich Teßmann, Italy, {name.surname}@tessmann.it;

³Universitäts- und Landesbibliothek Tirol, Austria, {name.surname}@uibk.ac.at.

ABSTRACT

This paper outlines a citizen humanities activity developed for a transnational project devoted to the digitisation and promotion of newspapers from the historical region of Tyrol.

KEYWORDS

Digital humanities, citizen humanities, Tyrol, historical newspapers, digital libraries, urban history, cultural history, community history.

POSTER

1. INTRODUCTION

The *Zeit.shift* project is an ongoing digital humanities cooperation between Eurac Research, the Landesbibliothek Dr. Friedrich Teßmann and the Universitäts- und Landesbibliothek Tirol, which seeks to contribute to the preservation of the memory and cultural heritage of the historical region of Tyrol. The project focusses on historical newspapers written in German and mostly blackletter script, which are currently scattered across North, East and South Tyrol and are only partially digitised. The objective of the project is twofold: increase access to these historical collections by digitising some 500,000 pages of Tyrolean newspapers published between 1850 and 1950; and investigate the potential of citizen engagement as a means of making these cultural assets more widely known while harvesting data for research.

Here, we describe one citizen humanities activity devised for the project. As launch is planned for October 11th, at present we can only report on expected results.

2. STATE OF THE ART

Recent studies have shown that the Social Sciences and the Humanities account for only 11% of citizen humanities practice. Within this small pool of projects, few tackle digitised historical sources and typically adopt participatory approaches in three areas: transcription, georeferencing and annotation ([3]). Participation is often remote and is geared towards the enhancement and acceleration of research (ibid.).

The contributory initiatives¹ that most closely resemble the present project are *Altes-Leipzig*, a reconstruction of historical Leipzig made possible thanks to digitised archival documents and genealogical information provided by citizens. Other analogous initiatives focussing on historical newspapers are²; the National Library of Australia's ongoing *Trove* effort, which allows members of the public to correct the OCR'd text of newspaper articles³; and the now completed *DigitalKoot* task commissioned by the National Library of Finland to correct noisy OCR generated from historical Finnish newspapers⁴.

3. METHODOLOGY

The two libraries digitise their Tyrolean newspaper holdings while Eurac Research is charged with the development of citizen humanities activities. Here, we describe the first activity to have been developed, which asks citizens to geolocate and semantically tag newspaper advertisements to help recreate the economic landscape of Tyrol from roughly 100 years

¹ For more information about this typology of project participation, see ([2]).

² <https://www.altes-leipzig.de/>.

³ <https://trove.nla.gov.au/newspaper/>.

⁴ <https://scistarter.org/digitalkoot>.

ago. We believe historical advertisements lend themselves well to citizen engagement not only because of their concise, visual and often amusing essence, but for the captivating sense of nostalgia they evoke in reminding us of products, traditions, people or businesses from years passed. The project seeks to leverage this powerful sense of familiarity to crowdsource citizen knowledge for the purposes of both cultural dissemination and research. Indeed, although the primary goal of the activity is to increase citizens' awareness of these digitised transregional newspaper collections, proper names tagged by participants will be used towards the improvement of Named Entity Recognition of newspaper texts written in the Tyrolean variant of the German language, as well as Named Entity search functionality for the two libraries.



Fig. 1. In this humorous 1923 advertisement from *Tiroler Grenzboten* (Kufstein, North Tyrol, Austria), the “Russolin” pesticide is described as a “unique mass exterminator of Russians and Swabians”. In the past, cockroaches were often named after political opponents, hence the colloquial use here of “Russen” and “Schwaben” to identify these insects.

The activity makes use of the existing third-party platform *Historypin*⁵, used by many GLAM institutions globally to showcase and open up their image collections to the public. The platform was chosen for its long-term data storage strategy, its intuitive interface, as well as its range of capabilities, which for users include, among others, the option of pinning images on a map, adding comments, creating virtual tours, sharing content and saving copies of the images in personal collections; for content providers, the platform supports bulk-upload/download of data and further automation through its API.

For a balanced representation of all Tyrolean communities, we select newspapers from as many cities as possible. The data chosen includes both display and classified adverts, and the imaging and OCR quality vary greatly across newspapers. We run a custom script⁶ on the TIFF and ALTO XML files to automatically extract the adverts and then manually filter out the false positives before bulk-uploading the extractions to the platform. In the interest of time, our opportunistic data preprocessing workflow does not weed out unreadable or duplicate adverts.

While there are no rules to tagging, participants are asked to use tags to, as a minimum, flag duplicate and illegible adverts, as well as adverts that do not provide location information. Both long and short instructions are provided to help participants complete the task. Usability tests were conducted with both German and non-German speakers to optimise the activity, which is open to all but is primarily targeted at the former and scholarly communities especially (e.g. historians). Interested citizens will be invited to complete the activity from home and to join the online workshops planned for the entire duration of the project. The activity will be advertised in research, museum and library networks alike by way of social media, blog posts, flyers, newsletters, as well as presentations at conferences and at other relevant venues.

To date, the platform hosts over 3,000 adverts from two different newspapers, with more to come.

4. DISCUSSION AND EXPECTED RESULTS

As a “distributed intelligence” citizen humanities project, *Zeit.shift* does not include input from citizens at each step of the research but relies on their cognitive and observation abilities to enhance data, granting them the possibility of influencing

⁵ For which the present project has also created a German language interface at: <https://www.historypin.org/de/>.

⁶ Adapted from <https://github.com/cneud/alto-tools>.

changes in methodology, objectives, development, results and dissemination. Unlike most citizen humanities projects working with digitised historical newspapers, Zeit.shift endeavours to move beyond mere OCR correction and data annotation, looking to, instead, lead participants on a trail of serendipitous discovery of the past (e.g. learning more about local ononymy or about professions that no longer exist).

Among the major concerns of citizen science and humanities projects are the quality and the sustainability of user contributions. Data quality in this activity (i.e. tagging consistency) is verified by project staff manually and through API calls, while spam protection is managed by the hosting platform. As previously mentioned, this first activity is not immediately focussed on data quality so much as community building. With regard to sustainability, while the project has signed a two-year agreement with *Historypin* for maintenance support, all content contributed to the platform will remain online indefinitely thereafter.

The project has set out to achieve 10,000 actions between the first and second citizen humanities activities by autumn 2022. We expect the majority of these to come from our upcoming *microtask* as opposed to the *macrotask* described here owing to the gamified nature of the former ([1]). Moreover, in line with most such projects, we expect an 80-20 pattern of participation, that is, that most contributions will come from a small percentage of committed users (“Pareto principle”). The success and long-term feasibility of the *Historypin* task will be measured on the number of citizens involved and contributions or actions completed, on the amount of community and data management required of project staff and on the quantity of additional traffic the activity will direct to the partner libraries.

5. ACKNOWLEDGEMENTS

Zeit.shift is funded by the European Regional Development Fund and Interreg V-A Italia - Austria 2014-2020 (ITAT 3030). Special thanks go to our contributing citizens.

REFERENCES

- [1] Ridge, Mia. 2020. *Crowdsourcing in Cultural Heritage*. Routledge Handbooks Online.
- [2] Shirk, Jennifer, Heidi Ballard, Candie Wilderman, Tina Phillips, Andrea Wiggins, Rebecca Jordan, Ellen McCallie, et al. 2012. “Public Participation in Scientific Research: A Framework for Deliberate Design.” *Ecology and Society* 17 (2). <https://doi.org/10.5751/ES-04705-170229>.
- [3] Tauginienė, Loreta, Eglė Butkevičienė, Katrin Vohland, Barbara Heinisch, Maria Daskolia, Monika Suškevičs, et al. 2020. “Citizen Science in the Social Sciences and Humanities: The Power of Interdisciplinarity.” *Palgrave Communications* 6 (1): 1–11. <https://doi.org/10.1057/s41599-020-0471-y>.

Un esperimento di visualizzazione grafica della terminologia del Talmud babilonese

Simone Marchi¹, Marianna Colombo¹, David Dattilo², Emiliano Giovannetti¹

¹Istituto di Linguistica Computazionale “A. Zampolli”, CNR, Italia - nome.cognome@ilc.cnr.it

²PTTB S.c.a r.l., Italia - david.dattilo@talmud.it

ABSTRACT

L’impiego di tecnologie di *information visualization* nel settore delle *digital humanities* può aprire nuove frontiere di ricerca. Le informazioni veicolate attraverso modalità grafiche, infatti, possono apparire agli studiosi più immediatamente comprensibili e le interfacce grafiche realizzate fornire inediti paradigmi di studio e di manipolazione dei dati analizzati. Il caso d’uso sperimentale illustrato in questo contributo è stato concepito per fornire allo studioso una modalità visiva, immediata, per l’analisi comparativa del contenuto terminologico di un corpus testuale.

PAROLE CHIAVE

Visualizzazione grafica di risorse testuali, terminologia, linguistica computazionale, tf-idf, grafi.

POSTER

1. INTRODUZIONE

Nel presente contributo si illustra un’applicazione web sperimentale per la visualizzazione grafica dei termini salienti presenti in un corpus di testi. Mediante un’interfaccia grafica specifica, descritta più avanti, l’applicazione consente di visualizzare sotto forma di grafo i testi che costituiscono il corpus e i termini contenuti in essi, entrambi espressi come nodi, e le relazioni di appartenenza di ogni termine al testo (o ai testi), rappresentati da archi tra i nodi. Attraverso questa tecnica si intende sperimentare i vantaggi di una esplorazione grafica della componente terminologica di un corpus testuale nella quale i termini peculiari di un testo sono più facilmente distinguibili da quelli condivisi tra più testi. Una modalità di analisi di questo tipo potrebbe, ad esempio, aiutare nella comprensione del contenuto dei testi considerati, così come esso traspare attraverso la terminologia che li contraddistingue; analogamente, poter visualizzare graficamente quali sono i termini condivisi tra due o più testi potrebbe fornire una prima indicazione di similarità argomentale.

Il corpus trattato in questo esperimento è composto da otto trattati del Talmud tradotti in italiano in seno al Progetto di Traduzione del Talmud babilonese¹. La varietà di temi affrontati nel Talmud, composto da trentasette trattati raggruppati in sei ordini, rende il corpus scelto particolarmente adatto a questa prima sperimentazione. Nonostante ogni trattato (come il titolo stesso suggerisce) verta su un tema specifico (“le benedizioni”, “il digiuno”, “lo Shabbat”, ecc.), ogni volume tratta anche di altri temi, anche non direttamente legati a quello principale. Come è possibile vedere, anche a colpo d’occhio, dal grafo prodotto mediante la metodologia descritta nella prossima sezione e manipolabile attraverso l’interfaccia descritta nella sezione 3, il contenuto argomentale di ogni trattato può essere in parte già evinto dai termini salienti che lo contraddistinguono (Fig. 1). Sebbene altri lavori abbiano trattato la navigazione grafica di terminologie (si vedano, a titolo di esempio, ([5];[6]), ma si veda anche ([4]) per una rassegna più generale delle tecniche di visualizzazione di dati testuali), non ci risultano lavori sulla visualizzazione congiunta di testi e terminologia con i quali confrontare il presente contributo.

2. METODOLOGIA

Il grafo è stato ottenuto attraverso una metodologia riassumibile nei seguenti passaggi: i) estrazione dei termini dal corpus e relativa indicizzazione; ii) conversione dell’indicizzazione ottenuta in un grafo serializzato in JSON al quale sono stati aggiunti, come nodi specifici, gli otto testi che costituiscono il corpus; iii) visualizzazione del grafo tramite un’applicazione basata sul framework Angular² e la libreria Cytoscape.js ([2]). Per l’estrazione della terminologia è stato utilizzato T2K² ([1]) un sistema per l’estrazione terminologica da testi composto da una serie di strumenti per il trattamento automatico della lingua italiana. Una delle funzioni principali di T2K² è l’estrazione di termini da una collezione di testi basata su regole linguistiche combinate a una serie di filtri che fanno uso di misure statistiche. T2K² prevede la possibilità di

¹ <https://www.talmud.it/>.

² <https://angular.io/>.

configurare l’algoritmo di estrazione terminologica tramite la scelta sia di pattern di estrazione di sintagmi nominali sia di soglie di frequenza dei sintagmi stessi. Per questo esperimento le soglie sono state stabilite empiricamente al fine di ottenere, nei passaggi successivi, un grafo che fosse, allo stesso tempo, significativamente ricco di informazione ma non troppo esteso e, quindi, difficile da visualizzare e manipolare. Tra gli output del sistema T2K² vi è l’indicizzazione terminologica dove ad ogni sintagma estratto sono associate informazioni di varia natura, tra cui la collocazione all’interno del corpus e una misura di rilevanza calcolata con la *tf-idf* (term frequency–inverse document frequency)³. Questa indicizzazione ha costituito il punto di partenza per il successivo passo di selezione e trasformazione dell’informazione in un grafo in formato JSON compatibile con la libreria Cytoscape.js adottata per la visualizzazione del grafo. Dall’indicizzazione sono state selezionate tre tipologie di termini: i) peculiari (i.e. i termini che contraddistinguono un trattato in termini di rilevanza), ii) condivisi ad alta rilevanza e iii) condivisi ad alta frequenza. I tipi i) e ii) sono stati selezionati come i primi della lista dei termini ordinati in modo decrescente di *tf-idf*, mentre il tipo iii) selezionando i termini a più alta frequenza e *tf-idf* nulla. Seppure in un contesto diverso, la terminologia del Talmud babilonese è già stata oggetto di analisi in ([3]), dove gli autori, tuttavia, non hanno affrontato il tema della visualizzazione.

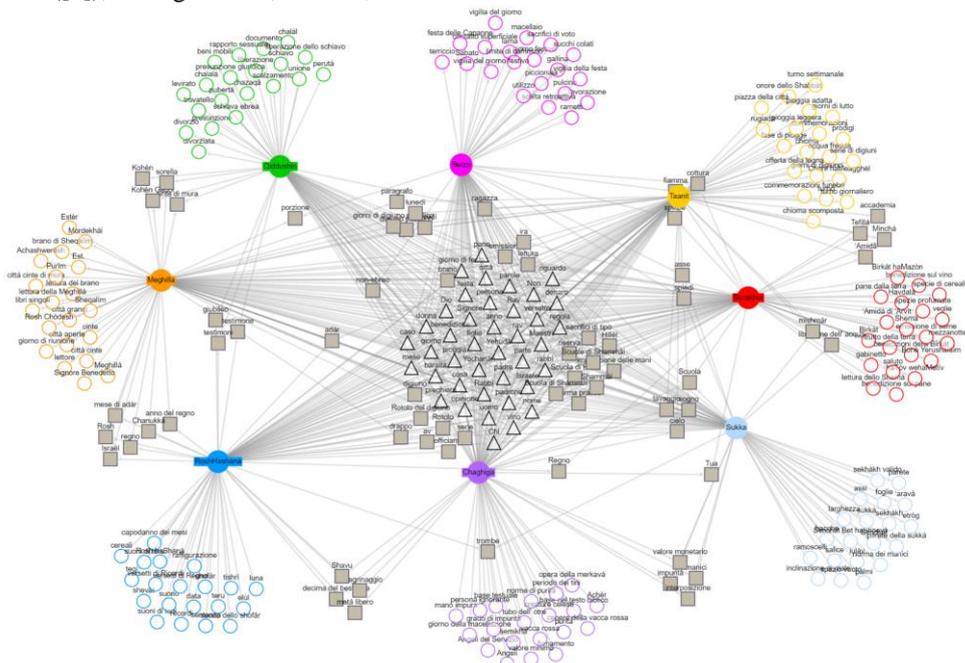


Figura 1. Il grafo degli otto trattati del Talmud selezionati, e dei relativi termini, visualizzato nella sua interezza

3. L’INTERFACCIA GRAFICA

L’interfaccia, liberamente accessibile nella sua prima versione prototipica all’indirizzo in nota⁴, è organizzata in tre zone. In alto è presente una bottoniera che agisce sull’intero grafo e che offre, da sinistra a destra, le seguenti funzionalità: i) mostra grafo; ii) mostra/nascondi termini peculiari; iii) mostra/nascondi termini condivisi (tra i trattati visualizzati) ad alta rilevanza; iv) mostra/nascondi termini condivisi (tra i trattati visualizzati) ad alta frequenza; v) nascondi grafo; vi) adatta (il grafo all’area di visualizzazione); vii) *zoom in / zoom out*. A sinistra vi è una colonna con la lista dei trattati, la lista dei termini e, in alto, una casella di testo che consente di filtrare trattati e termini per agevolarne la selezione. Al centro, infine, è presente l’area di visualizzazione del grafo. L’utente può visualizzare nodi specifici del grafo cliccando sui relativi termini o trattati presenti nella lista di sinistra e, successivamente, visualizzare gli altri nodi ad essi collegati mediante l’uso del mouse, come più avanti descritto. Il clic singolo con il tasto sinistro del mouse su un nodo del grafo apre la colonna di destra a comparsa con informazioni relative al trattato o al termine selezionato. Nel caso di un trattato vengono mostrate le seguenti informazioni: i) il nome del trattato selezionato; ii) l’elenco di tutti i termini che compaiono in quel trattato (corredati del relativo numero di occorrenze) ordinati in modo decrescente per rilevanza. Nel caso in cui, invece, sia selezionato un termine, i dettagli mostrati sono: i) il termine stesso; ii) il nome del trattato (o trattati) in cui esso compare corredato dal numero di occorrenze. La distinzione tra elementi testuali e terminologici di diversa natura all’interno del

³ La *tf-idf* misura la rilevanza di un termine rispetto a uno specifico documento di un corpus; un alto valore di *tf-idf* indica che il termine appare frequentemente in pochi documenti (e quindi è peculiare di quei documenti) mentre un basso valore di *tf-idf* indica che il termine si distribuisce in molti documenti diversi.

⁴ <https://klab.ilc.cnr.it/demoTermGraph/> (si consiglia l’utilizzo di un PC con browser Chrome).

grafo è stata agevolata attraverso l'uso di colori e di forme diverse: i trattati sono rappresentati da cerchi pieni di colore differente, i termini peculiari da cerchi trasparenti con contorno colorato della stessa tonalità del trattato di appartenenza, i termini condivisi ad alta rilevanza da quadrati e, infine, i termini condivisi ad alta frequenza da triangoli. Nella colonna di sinistra, invece, i termini condivisi ad alta rilevanza sono in verde, i termini ad alta frequenza in blu e i termini peculiari appaiono in rosso. In aggiunta alle funzionalità accessibili attraverso la bottoniera in alto, concepita per agire a livello globale, il grafo è esplorabile a livello di singoli nodi mediante due modalità di interazione: i) il doppio clic con il tasto sinistro del mouse sul nodo di interesse e ii) il menù contestuale. Nel primo caso, il primo doppio clic su un nodo trattato apre tutti i nodi termine ad esso collegati. I successivi doppi clic aprono e chiudono i termini peculiari del trattato, lasciando sempre visibili i termini condivisi. Il doppio clic su un nodo termine, invece, fa apparire i nodi trattato (o trattati) a cui esso è collegato. Il menù contestuale, che appare al clic destro del mouse su un nodo, offre un insieme di azioni distinte in base al nodo cliccato, a seconda che rappresenti un trattato o un termine. Nel caso di un trattato, l'utente ha la possibilità di: i) visualizzare tutti i termini ad esso collegati; ii) visualizzare solo i termini peculiari; iii) visualizzare solo i termini condivisi (ad alta rilevanza o frequenza); iv) nascondere tutti i termini; v) nascondere i termini condivisi (ad alta rilevanza o frequenza); vi) nascondere i termini peculiari; vii) chiudere il trattato. Nel caso di un termine le azioni possibili sono solo due: i) mostrare il trattato (o i trattati) in cui il termine compare; ii) nascondere il termine.

4. CONCLUSIONI

Nel presente contributo si è illustrato un esperimento di visualizzazione grafica di testi e terminologia, concepito primariamente per indagare modalità grafiche innovative per lo studio comparato del contenuto terminologico (e quindi, di conseguenza, argomentale) di testi appartenenti ad un corpus. Pur non portando, in questo primo contesto sperimentale, evidenze empiriche circa i possibili vantaggi di tale approccio, riteniamo sia già possibile, analizzando il grafo, individuare in esso alcuni casi d'uso interessanti. Innanzitutto, la visualizzazione del grafo nella sua interezza consente, in prima battuta, di capire quali possano essere gli argomenti principali affrontati nei vari testi e, subito dopo (attraverso i termini condivisi ad alta rilevanza), quali siano i possibili argomenti condivisi tra due o più testi. Un esempio interessante è dato dai trattati Berakhòt e Ta'anit: visualizzandoli entrambi e mostrando i termini tra loro condivisi ad alta rilevanza (con l'apposito pulsante in alto) è possibile individuare il termine "Tefillà" (preghiera ebraica), ad indicare che la Tefillà è, quindi, un argomento trattato in entrambi. Inoltre, selezionando il nodo relativo al termine, è possibile notare una distribuzione diversa del termine sui due trattati: 25 occorrenze in Berakhòt (che, infatti, tratta profusamente di benedizioni e di preghiere) e 7 occorrenze in Ta'anit. Anche i termini condivisi ad alta frequenza, visualizzati nella parte centrale del grafo come nodi di forma triangolare, forniscono un dato molto interessante: di fatto, essi costituiscono i "termini talmudici" (come "Dio", "rabbi", "regola", ecc.) che appaiono trasversalmente in tutti i trattati considerati indipendentemente dai loro argomenti specifici.

5. RICONOSCIMENTI

Il presente lavoro è stato condotto nel contesto del Progetto TALMUD e nell'ambito della cooperazione scientifica tra S.c.ar.l. PTTB e ILC-CNR.

BIBLIOGRAFIA

- [1] Dell'Orletta, Felice, Giulia Venturi, Andrea Cimino, e Simonetta Montemagni. 2014. «T2K²: a System for Automatically Extracting and Organizing Knowledge from Texts». In *Proceedings of 9th Edition of International Conference on Language Resources and Evaluation*, a cura di N. Calzolari et al. Reykjavik.
- [2] Franz, Max, Christian T. Lopes, Gerardo Huck, Yue Dong, Onur Sumer, e Gary D. Bader. 2016. «Cytoscape.js: a graph theory library for visualisation and analysis». *Bioinformatics* 32 (2): 309–11.
- [3] Giovannetti, Emiliano, Andrea Bellandi, David Dattilo, Mario Del Grosso, Simone Marchi, Alessandra Pecchioli, e Silvia Piccini. 2020. «The Terminology of the Babylonian Talmud: Extraction, Representation and Use in the Context of Computational Linguistics». *Materia Giudaica* 25: 61–74.
- [4] Kucher, Kostiantyn, e Andreas Kerren. 2015. «Text visualization techniques: Taxonomy, visual survey, and community insights». *IEEE Pacific Visualization Symposium (PacificVis)*, 117–21.
- [5] Miljkovic, Dragana, Jan Kralj, Uroš Stepišnik, e Senja Pollak. 2019. «Communities of Related Terms in a Karst Terminology Co-occurrence Network». In *Proceedings of eLex*. Sintra.
- [6] Robichaud, Benoît. 2011. «A graph visualization tool for terminology discovery and assessment». In *Proceedings of the Fifth International Conference on Meaning-Text Theory*, 243–52. Barcelona, Spain.

Una edizione critica digitale per la cristianistica dell'antichità

Luca Avellis

Università degli Studi di Bari Aldo Moro, Italia, luca.avellis@uniba.it

ABSTRACT

Alcuni testi per loro natura suscettibile a cambiamenti, come calendari e opere computistiche, non hanno avuto una edizione critica in senso proprio. Tali opere di fondamentale interesse sono caratterizzate da un gran numero di testimoni e di varianti. L'edizione critica digitale è forse la soluzione a questa esigenza peculiare della letteratura cristiana antica.

PAROLE CHIAVE

Edizioni critiche digitali, textual data, *Martyrologium Hieronymianum*.

POSTER

1. INTRODUZIONE

Una descrizione dei rapporti tra cristianistica e strumenti digitali è stata offerta recentemente da Galavotti ([3]). Questi pone come esperienza fondativa la schedatura integrale del *corpus* di Tommaso d'Aquino di padre Busa, che giunse nel 1949 a una prima formalizzazione stabile attraverso schede perforate IBM. Ma il primo tentativo di automazione dei processi, non solo di una schedatura bensì di un metodo ecdotico, ebbe un precursore nel domenicano Henri Quentin agli inizi del '900.

2. UNA EDIZIONE CRITICA PER LA CRISTIANISTICA DELL'ANTICHITÀ

Secondo Milanese ([6]) il metodo quentiniano basato su matrici e colonne anticipa concettualmente la struttura di un dataset. Orlandi limita questa ipotesi ([7]). L'approccio quentiniano, prima applicazione del sistema cladistico ([2]), fu affiancato presto da quello di altri studiosi come Greg, primo ad applicare il calcolo delle probabilità alla critica del testo, seguito da Dearing, ed ebbe in Froger un momento di incontro delle due idee. Queste videro la prima applicazione pratica nel software Quentin/80 solo nel 1980 ([6]). I problemi posti da questi tentativi erano di due nature: trovare un metodo oggettivo (sono diversi tra loro tanto quelli 'puri' di Lachmann, Bédier, Quentin, Greg, Maas, quanto quelli 'misti' di Clark, Collomp, Froger, etc.) e gestire il rapporto tra testimoni e varianti. Quentin creò la *méthode* (per la *Vulgata* geronimiana) per rispondere a due esigenze: un'oggettività scevra dal giudizio personale e la gestione di molti testimoni. Poco noto è che il metodo fu applicato da Quentin all'edizione del *Martyrologium Hieronymianum* (1931), caratterizzato da un alto numero di testimoni e un più alto numero di varianti e interpolazioni: l'esito fu parzialmente positivo. Una soluzione a questo problema potrebbe essere ora l'edizione critica digitale ([5]). The Versioning Machine con il sistema di affiancamento non gerarchico delle varianti potrebbe risultare il sistema più indicato in questo caso, a meno di non concepirne uno proprietario, ma si ritiene, per semplicità d'uso, collegamento delle immagini e flessibilità nella programmazione che sia EVT quello che offre le migliori possibilità. Un ulteriore aiuto, sulla scorta di Kinzig ([4]) potrebbe derivare dall'uso del coefficiente di correlazione di Spearman ([1]).

BIBLIOGRAFIA

- [1] Arsov, Nino, Milan Dukovskiy, Blagoja Evkoskiz, e Stefan Cvetkovskix. 2019. «A Measure of Similarity of Textual Data Using Spearman's Rank Correlation Coefficient». *arXiv* 1911.11750. <https://arxiv.org/abs/1911.11750>.
- [2] De Pinna, Mário, Fábio A. Bockmann, e René Zaragueta i Bagils. 2016. «Unrooted trees discovered independently in philology and phylogenetics: a remarkable case of methodological convergence». *Systematics and Biodiversity* 14 (4): 317–26.
- [3] Galavotti, Enrico. 2017. «La storia dei cristiani nell'era digitale». *Cristianesimo nella storia* 2: 357–82.
- [4] Kinzig, Wolfram. 1990. *In Search of Asterius. Studies on the Authorship of the Homilies on the Psalms*. Vandenhoeck & Ruprecht.
- [5] Michelone, Francesca. 2021. «L'edizione critica tra digitale e stampa: riflessioni metodologiche». *Umanistica Digitale* 10: 25–48.
- [6] Milanese, Guido. 2021. *Filologia, Letteratura, Computer. Idee e strumenti per l'informatica umanistica*. Vita e Pensiero.
- [7] Orlandi, Tito. 2010. *Informatica testuale. Teoria e prassi*. Editori Laterza.

Ritmi postumani: produzione poetica e machine learning

Lorenzo Demma¹, Daniele Silvi²

¹Università di Bologna, Italia – lorenzo.demma@studio.unibo.it

²Università di Roma “Tor Vergata”, Italia – silvi@lettere.uniroma2.it

ABSTRACT

Questo poster vuole offrire gli strumenti per una riflessione filosofica sull'attività simbolica del pensiero umano e la capacità del computer di creare connessioni, per capire se è possibile che l'intelligenza artificiale possa avere anche capacità creative ed eventualmente in che modo queste possano esplicitarsi in una poetica postumana. Inoltre questo poster si compone anche di una parte interattiva per stimolare la discussione con il pubblico che prevedrà l'uso di un test di Turing inverso per testare le capacità di diversi programmi di generazione di testi poetici, i cui risultati saranno ulteriormente discussi.

PAROLE CHIAVE

Post-umanesimo, intelligenza artificiale, Alan Turing, machine learning.

POSTER

Nel 1950 Alan Turing scrive un articolo in cui si interroga sull'intelligenza artificiale. Questo dibattito è ancora in corso, in parte per la difficoltà di definire il concetto di “intelligenza umana” e in parte, e di conseguenza, per la simile difficoltà nel definire quello di “intelligenza artificiale”. Partendo dall'affermazione di Turing: «Propongo di considerare la domanda: “Le macchine possono pensare?” Questo dovrebbe iniziare con le definizioni del significato dei termini macchina e pensare» ([6]: 1).

In questo poster vogliamo offrire gli strumenti per una riflessione filosofica sull'attività simbolica del pensiero umano e sulla sua capacità di creare connessioni tra simboli, confrontandola con la stessa funzione di un computer, per capire se è possibile che l'intelligenza artificiale possa avere anche capacità creative. Inoltre discuteremo, alla luce della critica filosofica e letteraria contemporanea, se ciò sia eticamente accettabile. Per perseguire questi due obiettivi, il nostro poster offrirà sia una parte informativa sulle tesi e sugli esperimenti fatti in merito all'argomento trattato, sia una parte interattiva per stimolare la discussione con il pubblico.

La nostra posizione è orientata a non separare la creatività dall'intelligenza ma ad ammettere che una macchina possa scrivere “intelligentemente” una poesia o un romanzo in modo creativo, senza quella capacità di creare una connessione interna di simboli che è propria dell'uomo e che gli permette di cogliere affinità e connessioni dove normalmente non ci sarebbero. Ancora di più, questa facoltà appartiene a poeti e scrittori, che fanno da ponte tra un mondo di idee e un mondo sensibile, attraverso strumenti di ricodifica e associazioni simboliche che non sono comuni. Chiameremo questo atteggiamento “capacità di rompere gli schemi”, in opposizione alla natura algoritmica di qualsiasi forma di intelligenza artificiale finora ipotizzata, una sorta di nuovo clinamen epicureo. La nostra intenzione è quella di stimolare i partecipanti su un tipo di test di Turing inverso, cioè verificare attraverso una serie di domande ed esperienze, se testare la macchina equivalga – e in che misura – a testare l'interrogante. In altre parole, se tramite il Test di Turing, la ricerca delle evidenze riguardo all' “*inner information processes*” ([4]) della macchina non sia un test dei processi intellettivi e immaginativi dell'uomo stesso fino a giungere a conclusioni come quella di Brooks: “*intelligence is in the eye of the observer*” ([1]).

Nel poster discuteremo la nostra posizione in merito alla questione se abbia senso parlare di creatività artificiale e se una poesia scritta da un agente artificiale abbia piena dignità alla luce della storia dell'intelligenza artificiale e della cibernetica. Esporremo sinotticamente le tesi esistenti sull'argomento, partendo dalle Macy Conferences ed arrivando ai giorni nostri, cercando di far emergere luci ed ombre del lungo dibattito.

Nella parte di coinvolgimento interattivo, proporremo a chi interverrà una discussione per sondare il rapporto – se esiste – tra intelligenza e creatività: alimentando alcune domande, alla luce dei materiali che forniremo e presentando anche un test di Turing inverso i cui risultati saranno ulteriormente discussi al termine della conferenza stessa. Alcuni degli interrogativi su cui ci preme stimolare le reazioni altrui sono: È possibile associare una macchina che gioca a scacchi (già esistente) a una macchina che scrive la poesia (ipotizzabile)? Le due cose sono divise o no (creatività ed intelligenza)?

Il nostro intento è quello di ipotizzare, insieme ai partecipanti, un nuovo (inverso) Test di Turing per circoscrivere le caratteristiche dell'intelligenza umana, che rimane il problema principale. Cercheremo di far emergere i processi che ci

sono dietro il meccanismo del riconoscimento della macchina come tale, dal momento che già nel test di Turing la risposta di chi interrogava la macchina era cruciale ([5]).

La capacità di creazione poetica pertiene ad un processo emozionale, piuttosto che di intelligenza matematica ([6])? Se le cose stanno così, per capire la natura e i confini del Postumano dobbiamo ancora ripensare la natura dell'umano e metterci di fronte alla macchina come se fossimo davanti ad uno specchio ([2];[3])?

Il consesso della conferenza ci appare il luogo naturale per proporre una simile attività, proprio come nel corso delle Macy Conferences, ed esattamente come allora con il contributo di studiosi appartenenti ai più svariati campi disciplinari.

BIBLIOGRAFIA

- [1] Brooks, Rodney A. 2018. *The Artificial Life Route to Artificial Intelligence*, Londra. Londra: Routledge.
- [2] Hayles, N. Katherine. 1999a. *How We Became Posthuman. Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: The University of Chicago Press.
- [3] ———. 1999b. "Simulating Narratives: What Virtual Creatures Can Teach Us." *Critical Inquiry* 26 (1): 1–26.
- [4] ———. 2010. "How We Became Posthuman: Ten Years On An Interview with N. Katherine Hayles." In *Psychoanalysis and the Posthuman*, 33:318–30. Edinburgh University Press.
- [5] Proudfoot, Diane. 2013. "Rethinking Turing's Test." *The Journal of Philosophy* 110 (7): 391–411.
- [6] Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind* LIX (236): 433–60.

Argument-Checking: A Critical Pedagogy Approach to Digital Literacy

Ruben Brave¹, Federica Russo², Jean Wagemans³

¹ CEO Entelligence and co-founder MMGA, The Netherlands – brave@entelligence.nl

² University of Amsterdam, The Netherlands – f.russo@uva.nl

³ University of Amsterdam, The Netherlands – j.h.m.wagemans@uva.nl

ABSTRACT

The digital revolution brought about unprecedented changes in people's daily lives as well as in techno-scientific contexts. In this paper, we address the problem of information overload people experience in online media, news outlets, and social media. The problem is well-known for its negative influence on the quality of online information, with abundant discussion on the promise of fact-checking and the potential role of censorship and moderation by social media. We instead discuss the issue from the perspective of digital literacy; specifically, we advance the view that our procedure of *argument-checking* can enhance such literacy, as a form of critical pedagogy, thereby contributing to improving the quality of online information.

KEYWORDS

Argument-checking, critical pedagogy, critical thinking, digital literacy, information overload.

POSTER

1. INFORMATION OVERLOAD IN THE DIGITAL ERA

The digital revolution has brought about profound changes. These changes do not only concern the technologies and artefacts developed for interacting with reality but have rendered the dichotomous distinction between online and offline obsolete – what we rather experience is *onlife* ([8]). We have entered the so-called “zettabyte era”, with an incredible amount of information being shared, and at an incredible speed. The implications thereof reach far beyond amount and speed: digital technologies, and more specifically *information and communication technologies* (ICTs), are changing how we form and perceive ourselves as well as our relationships with others and the surrounding world ([7];[21]).

We wish to focus here on ICTs, and in particular on one consequence of their use: people are confronted with an overload of information that is difficult to assess ([4];[17];[22]). These difficulties concern both the processing of the information as well as putting it to good use. This, in turn, may influence the way in which people see themselves (identity), their relationship to knowledge (manipulation), as well as their perception of the reliability of the source (authority).

In the context of online information, two different aspects of “information overload” stand out. First of all, the *quantity* of information: there is simply too much, and this imposes limits on our time and capacity for selecting and assessing it. Second, the *quality* of information: it is unclear what is true and who is a reliable source. While these are well-known and studied problems ([3];[5]), in this paper we explore the prospects of enhancing people's digital literacy, and especially by means of a semi-automated approach to “argument-checking”.

2. “ARGUMENT-CHECKING” AS A FORM OF DIGITAL LITERACY

From the perspective of communication science, information overload is tackled by studying phenomena such as polarization ([10]). Our approach is different in that it focuses on *how* the communication flow contributes to the information overload, namely how *arguments* are used in online settings. In providing tools for analyzing and evaluating these arguments, we also go beyond fact-checking, which we consider a valuable but limited way of fighting mis-, dis-, and mal-information ([2]). While adherence to reality – the facts – is clearly vital for these purposes, there is more than just facts in online information: facts are often embedded in argumentative discourse, true facts can be used to support bad conclusions, many actions are not based on facts only, but also on values, and how we present facts is as important as the facts themselves ([20]).

The limited scope of fact-checking is one motivation for our shift from facts to arguments, but there is another important point to note: more often than not, in online media, people engage with one another in an *antagonistic* and *aggressive* way. Communication seems to be aimed at showing off their individual qualities, promoting individual goals, or winning

arguments by “knocking down” the other party. This, we submit, hinders or even obstructs the efforts of finding a solution for the problems of online communication. We instead advocate “argument-checking” as a way of promoting goals such as collective problem solving, building a shared knowledge-base, reaching consensus about theoretical issues (what to believe) as well as practical ones (what to do). These values call for a more diverse, inclusive, and empowering approach to argumentation, as described in work on virtue argumentation and the ethics of communication ([1];[6];[14]).

Many tools for analyzing and evaluating arguments are developed from a (formal) logical perspective. Our approach differs from this perspective in two ways. First, it can not only be employed by experts in formal logic but by individuals of various educational levels. Second, by closely connecting to the way in which arguments are expressed in natural language, the procedure of “argument-checking” enables the analyst to deal with natural arguments, i.e., arguments as encountered in their everyday lives, including online. The procedure itself consists of a limited number of steps that enable the analyst to find the relevant elements of the argumentation, to reconstruct these elements, and to assess them in a systematic way ([11];[20];[24];[25]).

While existing methods for argument analysis rely on the skills of the analyst in matching the characteristics of predefined argument types with the argument under scrutiny, we take a procedural approach that provides the analyst with a theoretically informed and justifiable analysis and evaluation of the argumentation as it is found “in the wild” ([15]). Because argument-checking aims to promote values of inclusiveness and diversity, and to empower people in the context of online media, we take our approach to be in line with critical pedagogy (see next section).

3. A CRITICAL PEDAGOGY APPROACH TO ARGUMENT LITERACY

We take inspiration and guidance here from the field of critical pedagogy ([9];[16]). Critical pedagogy promotes a specific approach to education, and notably one in which we strive to *empower* students, citizens and, in our case, users and producers of online contents. We aim to empower users and producers of online information by awakening their critical consciousness, and also by providing them with tools that they can put to use: argument-checking as a form of digital literacy.

We would like to emphasize here that our approach does not merely focus on sharpening rhetorical strategies for the eristic purpose of winning discussions or persuading audiences to believe or do something, but considers ways to establish the common aim of advancing knowledge in various communicative contexts. Our procedure for argument-checking is designed to promote the creation of shared knowledge and the improvement of the quality of online information. However, by putting these values first, we don’t aim to take a “moralistic” approach, which would mainly consist in calling out offenders for having committed fallacies and addressing imperatives to people that are never going to change their behaviour because it is simply not in their interest to do so. Such a response runs the risk of being counterproductive.

Our approach is a critical one in that we aim to create awareness about these problems and to empower people to do something about it. For this purpose, we provide them with tools for assessing the arguments put forward in a variety of online pieces, and teach them how to use these tools to pursue a diverse and inclusive online space. This empowers people in taking the responsibility to engage with online discussions and, at the same time, gives them the opportunity to master skills that shield them against manipulative persuasion while contributing to a shared knowledge-base or consensus. Specifically, by teaching argument-checking in online contexts, we aim to:

- I. Increase the literacy of individuals (as online *users*) to defend themselves against the negative effects of dis- and mis-information;
- II. Empower individuals (as online *agents*) to intervene and block in appropriate ways episodes of dis- and misinformation, of trolling, or other;
- III. Teach individuals (as online content *producers*) to share and disseminate information online that is of high enough quality.

4. FROM THEORY TO PRACTICE

The analysis and considerations of the previous sections lead us to delineate the main aspects of a long-term project in which researchers, social entrepreneurs, software engineers, and citizens can take part in. We want to build a sustainable community of people that can impact the overall quality of online information and communication. For this purpose, we will develop “argument-checking” as an offspring of a research area known as the philosophy of argument ([26]), capable of offering individuals (users, agents, producers of online contents) tools that help them assess the quality of information and engage with one another in a more constructive and fruitful way ([19];[23];[13]). We explicitly inject values in the design of these tools, e.g., creating a safe learning environment by role-playing and gamification (cf. debate contests). These tools, in practice, can be used to manually annotate online content, as is currently done in MMGA ([18];[12]).

MMGA is a blockchain-based annotation platform (with hundreds of registrants) in which screened and trained expert and/or critical thinking readers can annotate high-impact news sites such as NU.nl and AD.nl, two of the “Big Four” Dutch online news platforms. These tools are also currently explored for their prospects to design [KRINO](#), a glass-box AI engine that can assist humans in a semi-automated process of argument evaluation. KRINO is not a fully-automated engine, but rather aids human agents in analyzing written text and disentangling critical aspects of the underlying argument structure. Our main goal in engaging with this venture is the possibility of positively contributing to improving on the critical thinking and argument-checking skills of users, agents, and producers of online contents.

5. CONCLUSION

We are *onlife*. One consequence of this new dimension, brought about by the digital revolution, is that we need to learn strategies to cope with too much and too fast information. We can’t handle quantity, speed, and quality at the same time, and individually. We can’t simply rely on an army of fact-checkers – there will never be enough, and possibly they won’t cover the topics that interest *us*. A venue worth exploring to improve the quality of online information is to adopt a critical pedagogy approach, to empower individuals – users, agents, and producers of online contents – to critically assess that information. This is the goal of “argument-checking”, an approach based on insights from the philosophy of argument made applicable for use in online contexts, that can enhance digital literacy. We believe in the potential of combining human values with the use of digital technologies, and in the possibility that human users, agents, and producers of online contents can make a change in the infosphere, to deliver it to future generations in a better state than it is now.

6. ACKNOWLEDGEMENTS

We are very grateful to Ondrej Uzovic for the numerous conversations about KRINO, argument-checking, and human annotations. Our collaboration on this ongoing project is fruitful and so enriching. We also would like to thank Federico Gobbo, whose input and participation in the earlier stages of the KRINO project have been essential.

BIBLIOGRAPHY

- [1] Aberdein, Andrew, and Daniel H. Cohen. 2016. “Introduction: Virtues and Arguments.” *Topoi* 35 (2): 339–43.
- [2] Andersen, Jack, and Sille Obelitz S e. 2020. “Communicative Actions We Live by: The Problem with Fact-Checking, Tagging or Flagging Fake News – the Case of Facebook.” *European Journal of Communication* 35 (2): 126–39.
- [3] Borg, Stanley. 2019. “‘We Are Edging to a World Where Reality Is a Matter of Personal Opinion’ Academic and Strategist Alex Grech on Fake News and What to Do about It.” *Times of Malta*, July 30, 2019. <https://timesofmalta.com/articles/view/we-are-edging-to-a-world-where-reality-is-a-matter-of-personal-opinion.725056>.
- [4] Brave, Ruben. 2021. “Public Rebuttal, Reflection and Responsibility. Or an Inconvenient Answer to Fake News.” In *Media, Technology and Education in a Post-Truth Society (Digital Activism and Society: Politics, Economy And Culture In Network Communication)*, A. Grech, 145–54. Emerald Publishing Limited, Bingley.
- [5] ———. n.d. “Post-Truth Conference Malta 2019.” Talk on Media, Journalism & Fake News. <https://open.spotify.com/episode/3WzhTSRe1TSxnZQKz6e7iN>.
- [6] Dalgleish, Adam, Patrick Girard, and Maree Davies. 2017. “Critical Thinking, Bias and Feminist Philosophy: Building a Better Framework through Collaboration.” *Informal Logic* 37 (4): 351–69.
- [7] Floridi, Luciano. 2014. *The Fourth Revolution. How the Infosphere Is Reshaping Human Reality*. Oxford University Press.
- [8] ———. 2015. *The Onlife Manifesto: Being Human in a Hyperconnected Era*. 1st Edition. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-04093-6>.
- [9] Freire, Paulo, Myra Bergman Ramos, and Donald P. Macedo. 2014. *Pedagogy of the Oppressed*. Bloomsbury Publishing, 2014.
- [10] Hameleers, Michael, and Toni G. L. A Van der Meer. 2020. “Misinformation and Polarization in a High-Choice Media Environment: How Effective Are Political Fact-Checkers?” *Communication Research* 47 (2): 227–50.
- [11] Hinton, Martin, and Jean H. M. Wagemans. September. “Evaluating Reasoning in Natural Arguments: A Procedural Approach.” *Argumentation*. <https://doi.org/10.1007/s10503-021-09555-1>.
- [12] Internet Society Chapitre Du Benin. 2020. “Retour Sur Le Webinaire: R le Des Journalistes Dans l’utilisation de l’Internet Comme Instrument / Support de Communication En Temps de Crise.” 2020. <https://isoc.bj/webinaire-journaliste/#.YWN1eC8RppR>.
- [13] Internet Society Netherlands Chapter. n.d. “From Fact-Checking to Argument-Checking as Award Nominated Researchers of University of Amsterdam Join MMGA with Human-AI Framework.” Internet Society Netherlands Chapter. <https://isoc.nl/nieuws/whoever-does-not-study-rhetoric-will-be-a-victim-of-it/>.
- [14] Kidd, Ian James. 2016. “Intellectual Humility, Confidence, and Argumentation.” *Topoi* 35 (2): 395–402.
- [15] King, Colin Guthrie, and Jean H. M. Wagemans. forthcoming. *Argumentation in the Wild: An Introduction to Critical Thinking*. MIT Press.

- [16] Knight, Janine, Melinda Dooly, and Elena Barberà. November. “Getting Smart: Towards Critical Digital Literacy Pedagogies.” *Social Semiotics*, 1–24.
- [17] Lock, Irina, and Ramona Ludolph. 2020. “Organizational Propaganda on the Internet: A Systematic Review.” *Public Relations Inquiry* 9 (1): 103–27.
- [18] Make Media Great Again. 2019. “Introducing ‘Public Annotations’ in Journalism.” *MMGA.Io — Make Media Great Again* (blog). March 13, 2019. <https://medium.com/@MakeMediaGreatAgain/introducing-public-annotations-in-journalism-e688b04be903>.
- [19] Make Media Great Again News. 2020. “Argument-Checking AI Shields against Fake News and Misinformation.” Make Media Great Again News. December 22, 2020. <https://mmga.pr.co/191859-argument-checking-ai-shields-against-fake-news-and-misinformation>.
- [20] Plug, H. José, and Jean H. M. Wagemans. 2020. “From Fact-Checking to Rhetoric-Checking: Extending Methods for Evaluating Populist Discourse.” In *Vox Populi: Populism as a Rhetorical and Democratic Challenge*, Ingeborg van der Geest, Henrike Jansen, and Bart van Klink, 236–52. Edward Elgar Publishing. <https://doi.org/10.4337/9781789901412.00023>.
- [21] Russo, Federica. 2018. “Digital Technologies, Ethical Questions, and the Need of an Informational Framework.” *Philosophy & Technology* 31 (4): 655–67.
- [22] “Transforming Skills to Meet Innovation Challenges. EuroDIG 2019.” 2019. *EuroDIG* (blog). June 20, 2019. https://eurodigwiki.org/wiki/Transforming_skills_to_meet_innovation_challenges_-_WS_05_2019.
- [23] Van Gestel, Maarten. 2020. “Deze Filosofen Maken Een Argumentatiemachine Die, Hopen Zij, Drogredenen Uit Toespraken Kan Vissen.” *De Volkskrant* (blog). February 2020. <https://volkskrant.nl/ts-b77a1a75>.
- [24] Wagemans, Jean H. M. 2020a. “Argument Type Identification Procedure (ATIP) – Version 3.” *Periodic Table of Arguments* (blog). 2020. www.periodic-table-of-arguments.org/argument-type-identification-procedure.
- [25] ———. 2020b. “Why Missing Premises Can Be Missed: Evaluating Arguments by Determining Their Lever.” In *Proceedings of OSSA 12: Evidence, Persuasion & Diversity*. <https://scholar.uwindsor.ca/ossaarchive/OSSA12/Saturday/1>.
- [26] ———. 2021. “The Philosophy of Argument.” In *The Cambridge Handbook of the Philosophy of Language*, 1st edition, 571–89. Cambridge University Press. <https://doi.org/10.1017/9781108698283.032>.

“Nostra Signora Experience”: il Placetelling® in Ambiente Digitale

Isabella Hernandez

Università del Salento, lauraysabella.hernandezgarcia@studenti.unisalento.it

ABSTRACT

Questo poster si occuperà di “Nostra Signora Experience”¹ (@nostrasignoraexp, Instagram 2019), un progetto di placetelling che applica gli strumenti della gamification per creare uno storyworld in forma di esperienza immersiva multimediale. Coniato dalla scuola geografica salentina, il placetelling è una declinazione specifica dello storytelling che interessa diversi settori di ricerca. Si tratta di un concetto programmatico mediante il quale le storie possono essere studiate come racconto dei luoghi e i luoghi, a loro volta, come narrazioni ([7];[2]). Il placetelling è anche un’importante occasione di intersezione tra diversi ambiti scientifici, ed è per questo motivo che in seno all’Università del Salento nasce la “Scuola di Placetelling”, punto di convergenza dove si instaura un dialogo tra addetti ai lavori, accademici e studenti.

PAROLE CHIAVE

Placetelling, gamification, remediation, storyworld.

POSTER

Nel suo narrare, il cinema è costretto dal proprio statuto ontologico a mostrare gli ambienti in cui le storie si svolgono, in più, secondo Pollice e Bandirali, gli ambienti “non si limitano ad accogliere le storie: le generano” ([2]). L’edizione più recente della Scuola di Placetelling (2019) è stata dedicata alla profonda relazione che lega il placetelling al racconto audiovisivo. In particolare, le sessioni metodologiche e applicative si sono interessate allo studio della produzione audiovisiva in Puglia, luogo in cui lo strumento delle Film Commission ha generato una significativa economia ed estetica attraverso un’intensa azione di branding territoriale ([1]).

È nel contesto della Scuola di Placetelling dell’Università del Salento che nasce “Nostra Signora Experience”, un progetto di placetelling realizzato a partire dell’opera filmica “Nostra Signora dei Turchi” (1968) di Carmelo Bene. Il film di Bene, oltre ad essere stato interamente girato in Salento, è una compiuta espressione della matrice identitaria locale: paesaggi naturali e architetture assumono un ruolo autobiografico², così come leggende e miti locali sono accolti all’interno della struttura narrativa. Ora ci chiediamo, come far allestire quel particolare storyworld in un contesto di cultura digitale e rimediazione ([5])?

Per mettere in atto la rimediazione dello storyworld nel medium digitale, in “Nostra Signora Experience” si è fatto ricorso agli strumenti della gamification, applicandoli ai social media. La piattaforma social Instagram è particolarmente adatta al racconto interattivo dell’ontologia primaria e secondaria di questo mondo narrativo³. Seguendo la disposizione a tre colonne del *feed* (Fig. 1), sono stati individuati tre tipi di contenuti da condividere: fotografie dei luoghi delle riprese (Fig. 2), riproposizioni amatoriali in formato foto o video di scene del film (Fig. 3) e, infine, fotogrammi del film stesso, che ristabiliscono il legame con lo storyworld di riferimento⁴ (Fig. 4). Gli elementi di gioco sono introdotti da un regolamento⁵ che, mediante strumenti specifici del medium, come gli user-generated content o la geolocalizzazione, offre agli utenti

¹ Realizzato in collaborazione con Roberto Greco, Carola Gatto ed Emanuele Gatto (Università del Salento).

² Nella sua estesa opera sull’immagine-tempo, Gilles Deleuze riconosce che in *Nostra Signora* “la storia del palazzo diventa autobiografia del protagonista” ([6]: 222).

³ “Occorre dunque distinguere due ontologie del cinema: l’ontologia primaria (il mondo dove si crea il film: il reale di cui la pellicola serba traccia) e l’ontologia secondaria (il mondo creato dal film: ciò che è reale nella storia narrata).” *Nell’occhio e nel cielo: teoria e storia del cinema di fantascienza* ([3]: 16).

⁴ I contenuti che vengono condivisi nei social media sono caratterizzati da una sovrapposizione indifferenziata di diversi stili e forme mediali, un aspetto dei media digitali già individuato da Bolter e Grusin nel 1999: “Yet these same old and new media often refuse to leave us alone. Many web sites are riots of diverse media forms—graphics, digitized photographs, animation, and video—all set up in pages whose graphic design principles recall the psychedelic 1960s or dada in the 1910s and 1920s.” ([4]: 6).

⁵ “Regolamento: 1) Segui la griglia! Ogni riga contiene un indizio, 2) Recati al posto, esplora l’universo di NST, 3) Inviaci una foto e condividi la tua esperienza con noi. Sarai ricompensato con i segreti del film. Ricorda di taggarci e di usare i nostri hashtag: #nostrasignoraexp #carmelobene #NSTexp. Condividi con noi la tua creatività” (Disponibile sulla raccolta storie di @nostrasignoraexp, Instagram 2019).

l'esperienza di immergersi in questo storyworld così particolare, incentivando la loro creatività attraverso la condivisione partecipativa. In tal senso, il processo restituisce una rinnovata centralità ai luoghi del territorio salentino.

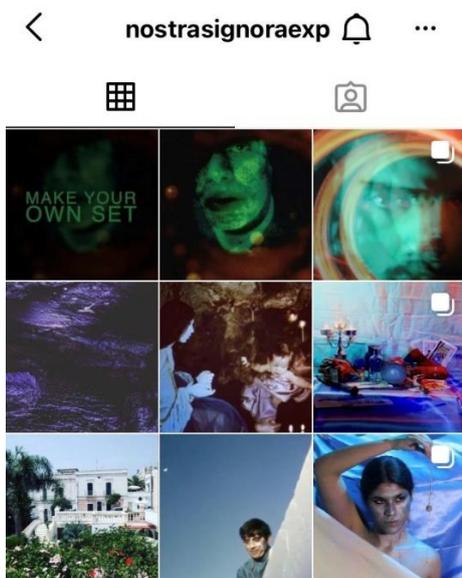


Figura 1. Feed, profilo Instagram di “Nostra Signora Experience” (@nostrasignoraexp, Instagram 2019).

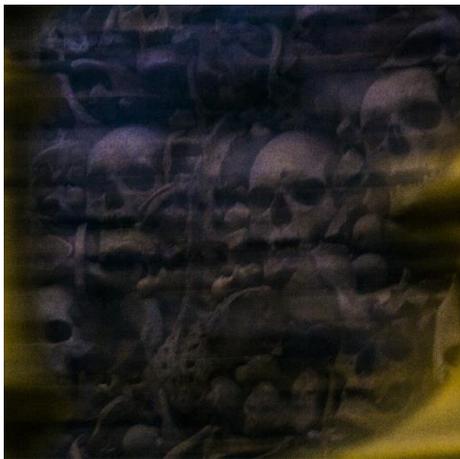


Figura 2. Foto della Cappella dei Martiri, Cattedrale di Santa Maria Annunziata, Otranto (LE). Autoproduzione, “Nostra Signora Experience” (@nostrasignoraexp, Instagram 2019).

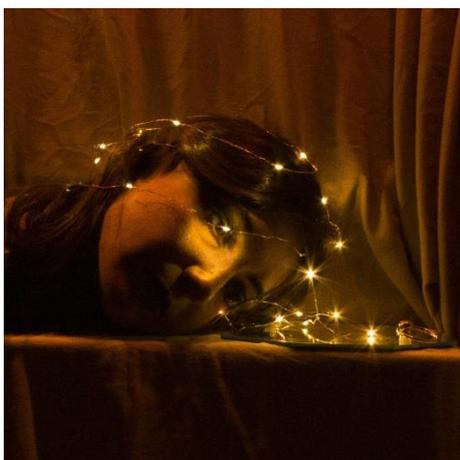


Figura 3. Riproposizione amatoriale di una scena del film. Autoproduzione, “Nostra Signora Experience” (@nostrasignoraexp, Instagram 2019).

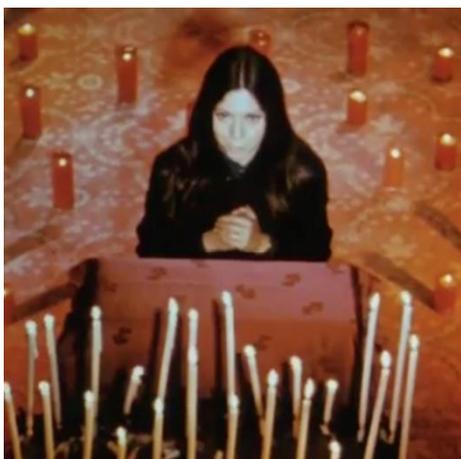


Figura 4. Fotogramma del film (@nostrasignoraexp, Instagram 2019). Fonte: “Nostra Signora dei Turchi” (1968) di Carmelo Bene.

BIBLIOGRAFIA

- [1] Bandirali, Luca. 2019. «Il Salento si alza. Come una regione di confine nel bacino del Mediterraneo è diventata un’area di interesse nell’ambito del cinema europeo contemporaneo». In *Cinema e identità italiana*, a cura di Parigi, Stefania, Christian Uva, Vito Zagarrò, 4:581–89. Roma: TrE-Press.
- [2] Bandirali, Luca, e Fabio Pollice. 2018. «Il cinema e il racconto dei luoghi». *Segnocinema*, 2018.
- [3] Bandirali, Luca, e Enrico Terrone. 2008. *Nell’occhio e nel cielo: teoria e storia del cinema di fantascienza*. Lindau.
- [4] Bolter, J. David, e Richard Grusin. 1999. *Remediation: Understanding New Media*. MIT Press.
- [5] Bolter, Jay David. 2019. *The Digital Plenitude: The Decline of Elite Culture and the Rise of New Media*. MIT Press.
- [6] Deleuze, Gilles. 2017. *L’immagine-tempo*. Cinema 2. Einaudi.
- [7] Pollice, Fabio. 2017. «Placetelling® per uno sviluppo della coscienza dei luoghi e dei loro patrimoni». *Territori della Cultura*, 2017.

AIUCD 2022



Consorzio Regionale
per le Arti e la Cultura



Con il patrocinio di:

