How to reference this article

Boschetti, F., Del Grosso, A. M. (2020). L'annotazione di testi storico-letterari al tempo dei social media. *Italica Wratislaviensia*, 11(1), 65–99.
DOI: <http://dx.doi.org/10.15804/IW.2020.11.1.03>

Federico Boschetti

CNR-ILC of Pisa & VeDPH Ca' Foscari Venezia, Italia

federico.boschetti@ilc.cnr.it

ORCID: 0000-0002-7810-7735

Angelo Mario Del Grosso

CNR-ILC of Pisa, Italia

angelo.delgrosso@ilc.cnr.it

ORCID: 0000-0002-4867-6304

L'ANNOTAZIONE DI TESTI STORICO- LETTERARI AL TEMPO DEI SOCIAL MEDIA

THE ANNOTATION OF HISTORICAL AND LITERARY TEXTS IN THE AGE OF SOCIAL MEDIA

Abstract: The annotation of historical and literary texts is approached differently by traditional philologists and digital philologists. The former are concentrated on the detailed study of a given text (close reading) while the latter are focused on the study of large quantities of texts (distant reading). A structured and collaborative annotation makes it possible both to add information to particular passages of individual texts, as in a traditional linear comment, and to connect data from entire textual collections through rigorous protocols. However, the standards developed by digital philologists are not highly appreciated by traditional academics, since the effort necessary to apply the proposed technologies allegedly diverts researchers' attention from the object of study. As opposed to this objection, we intend to highlight that it is indeed possible to maintain the precision requisite for the application of computational tools to digital resources without renouncing the annotation practices established in traditional contexts. In support of the method, we report a number of case studies of digital scientific editions whose goals include both reconstructing respective texts and encouraging the dissemination of contents and public participation in the academic debate. In particular, we will discuss the following projects: a) the stylistic annotation of three different editions of Giacomo Leopardi's translation of the *Batracomiomachia*; b) the scientific edition of Bellini's letters; c) the multi-level annotated edition of Bassani; and d) the comparison of Umberto Eco's variants of his *Il nome della rosa*.

Keywords: digital philology, collaborative annotation, communities, digital scholarly edition, formalisation

1. INTRODUZIONE¹

La costituzione del testo e la sua interpretazione, che in potenza sono infinite², in atto trovano i propri limiti³ nell'insieme finito di (micro-)documenti che trattano dell'espressione e del contenuto⁴ del testo stesso, variamente costituito in molteplici edizioni critiche e poliedricamente interpretato in una moltitudine di monografie, articoli, commenti e annotazioni.

Ciò significa che *si potrebbe o si sarebbe potuto* far dire ad un testo qualsiasi cosa (cf. Eco, 2004), ma ciò che gli si è fatto *effettivamente* dire, in modo documentato e non come semplice frutto di fugaci intuizioni o di dotte chiacchierate, si trova in un numero limitato, e quindi quantificabile, di epitesti (ad es. commentari) o paratesti⁵ (ad es. note esegetiche a piè di pagina). Anche questi ultimi sono testi a pieno titolo, e possono quindi essere ricorsivamente oggetto di critica e d'interpretazione, dando origine a nuovi epitesti e paratesti, come avviene in ogni edizione più recente che si trovi a dialogare con tutte le precedenti⁶.

¹ Sebbene il contributo sia il frutto di un lavoro congiunto degli autori, sono da attribuire a F. Boschetti le sezioni 1, 4 (introduzione), 4.1, 5 (introduzione), 5.3, 5.4, 6 e ad A.M. Del Grosso le sezioni 2, 3, 4.2, 4.3, 5.1, 5.2.

² Così come per Schleiermacher (1974) l'ermeneutica (o, più precisamente, il comprendere, *Verstehen*, implicato da essa) è compito infinito (*unendliche Aufgabe*), anche l'ecdotica lo è, in quanto la costituzione del testo non si limita alla combinatoria effettuata dagli editori delle varianti storicamente attestate, che è finita, salvo il raro caso della scoperta di nuovi testimoni. L'ecdotica lascia spazio infatti alle infinite possibilità offerte dall'emendamento per congettura, che è ad un tempo opera interpretativa e creativa dell'editore del testo critico.

³ Cf. Rastier, 2009. Tali confini naturalmente si espandono ad ogni nuovo apporto critico o interpretativo documentato.

⁴ Qui i termini sono usati nell'accezione hjelmsleviana.

⁵ La terminologia è mutuata da Genette, 1992.

⁶ Si pensi ad esempio, fra gli innumerevoli casi, all'edizione Carbone (1886) del *Canzoniere* di Petrarca con le annotazioni di Leopardi e una scelta operata da Carbone stesso delle annotazioni di Ambrosoli non solo al *Canzoniere* ma anche alle interpretazioni di Leopardi sul testo di Petrarca.

L'attività di annotazione si è evoluta nel corso del tempo⁷ non soltanto grazie allo sviluppo dei paradigmi cognitivi, delle tecniche ecdottiche e dei metodi interpretativi, ma anche in base alle modalità di interazione fra annotatore, testo da annotare e fonti dell'annotazione stessa, offerte dai dispositivi di lettura⁸, scrittura e accesso alle informazioni (cf. Lamé *et al.*, 2015).

La separazione del paratesto dal testo, il riferimento puntuale al passo discusso, la citazione sempre più precisa delle fonti, sono il frutto di un'evoluzione millenaria della prospettiva storica sulla testualità, della disposizione del testo su diversi supporti fisici sempre più agili e diffusi (papiro, pergamena, cartaceo, elettro-magnetico), del trattamento del testo in modi sempre meno sequenziali.

Nell'era digitale (Roncaglia, 2010) ci troviamo dunque a dover scegliere se restare all'interno di una tappa evolutiva dove le pratiche di annotazione del testo elaborate nel passato si adattano al nuovo contesto, oppure se vogliamo agire in discontinuità col passato e introdurre pratiche aliene alla tradizione degli studi umanistici, creando di fatto una frattura fra la comunità di filologi e letterati tradizionali e la comunità degli umanisti digitali.

2. L'ANNOTAZIONE E IL SUO USO NEL PASSAGGIO ALL'ERA DIGITALE

Al fine di trattare in formato digitale le informazioni testuali, paratestuali ed extratestuali presenti nei documenti letterari è necessario implementare un sistematico processo di annotazione che ne espliciti la struttura e il contenuto. Nelle fasi di tale processo, attraverso opportune tecnologie, vengono gestiti i dati sia descrittivi che analitici prodotti dagli studiosi nel corso delle proprie attività di indagine all'interno di documenti che siano almeno *machine-readable*. Inoltre, le varie fasi prevedono l'uso di strumenti che rispettino le tradizionali pratiche della

⁷ Un'eccellente panoramica si trova in Frigerio, 2016.

⁸ Per alcune posizioni critiche sui nuovi dispositivi di lettura si vedano Casati, 2014 e Costa, 2017.

comunità di riferimento, ma che siano anche efficaci nel recuperare ed elaborare i dati rappresentati in documenti *machine-actionable* (Shillingsburg, 2015).

Per dirla con Buzzetti, un corretto modello di annotazione deve confrontarsi necessariamente con due requisiti di fondo: la rappresentazione esaustiva e la rappresentazione funzionale delle informazioni presenti nel testo stesso (Buzzetti, 2009). Il primo requisito, ossia l'esaustività, risponde al bisogno di resa fedele e affidabile dei dati originali (conservare l'integrità dell'espressione), mentre il secondo requisito, ossia la funzionalità, risponde al bisogno di efficienza nell'elaborazione e nell'interpretazione del contenuto informativo veicolato dal documento originale (conservare l'integrità del contenuto). Se, per esempio, si considera un'edizione facsimilare di un'opera, emerge la necessità di disporre di un meccanismo per descrivere la molteplicità di interpretazioni che insistono su di essa e, viceversa, se si considera una particolare interpretazione sorge la necessità di descrivere le molteplici espressioni e possibili forme di conservazione mantenendo lo stesso contenuto informativo.

Emerge, da quanto detto sopra, che la rappresentazione digitale dell'informazione comprende sia la descrizione dei dati testuali sia la definizione delle operazioni ammissibili su di essi, secondo i principi degli Abstract Data Types⁹ (ADTs). Ne consegue che la rappresentazione del testo digitale è condizionata dalle operazioni di accesso, di ricerca e di confronto previste nel modello originario e condiziona la possibilità di effettuare ulteriori operazioni nelle estensioni del modello e nelle sue implementazioni (Orlandi, 2010). Ad esempio, i tradizionali modelli per il testo digitale ben rappresentano la versificazione dei singoli testimoni, ma soddisfano solo parzialmente i requisiti di chi abbia la necessità

⁹ Gli ADT sono strutture dati definite indirettamente dalle operazioni eseguibili sui loro contenuti e dalle proprietà formali di tali operazioni (Liskov, 1974). Ad esempio il tipo *String* non definisce soltanto una sequenza di elementi *Character*, per il quale potrebbe essere sufficiente un vettore (array) di bytes, ma definisce anche le operazioni che si possono fare sulla stringa, come contare il numero di caratteri che la costituiscono, trasformarla da minuscolo in maiuscolo, etc. Ogni linguaggio di programmazione può implementare in modo diverso l'ADT *String*.

di comparare le differenti disposizioni colometriche¹⁰. Questo perché la versificazione è requisito universalmente accettato per la fruizione del testo poetico, mentre il confronto delle diverse colometrie di uno stesso testo non è ancora in testa alla lista dei requisiti funzionali necessari alla comunità dei filologi digitali.

Di converso, la fruizione intelligente dei testi annotati deve tendere al miglior compromesso fra la rappresentazione esaustiva e la rappresentazione funzionale di dati rilevanti (precision) e completi (recall). Per chiarire, il facsimile della pagina di un manoscritto è una rappresentazione esaustiva (sotto certe condizioni) dell'espressione di un testo, mentre la trascrizione di tale pagina è una rappresentazione funzionale del contenuto testuale (isolato da altri contenuti, ad esempio iconografici). Se l'accesso al facsimile privo di metadati è oneroso in termini di individuazione dell'informazione rilevante, la ricerca sulla sola trascrizione non fornisce tutta l'informazione (testuale e non testuale) veicolata dal facsimile. Solo la combinazione delle due cose, tramite ad esempio la mappatura della trascrizione sul facsimile e la conseguente ricerca del testo rilevante sulla posizione precisa dello specchio di scrittura contornato da eventuali immagini, è ad oggi uno dei migliori compromessi fra integrità di espressione e integrità di contenuto, come spiegato sopra.

Inoltre, le annotazioni di diversa tipologia, quali quelle categoriali (ad esempio la lemmatizzazione), strutturali (ad esempio l'analisi metrica) e relazionali (ad esempio la corrispondenza di una unità testuale con la sua traduzione), oltre ad avere un valore autonomo, sono anche funzionali all'accesso e alla ricerca del testo, poiché le annotazioni permettono la creazione di indici multidimensionali, ripartiti per tipologia.

¹⁰ Negli studi filologici, la colometria antica, cioè la divisione originaria del verso in *cola* disposti su più linee, ha acquistato sempre maggiore attenzione (Battezzato, 2008), ma non sembra essere fra i temi di maggior rilievo per la rappresentazione digitale di edizioni critiche.

3. CONTESTO OPERATIVO E CENNI SULLO STATO DELL'ARTE

L'esperienza di ricerca condotta presso l'Istituto di Linguistica Computazionale "A. Zampolli" di Pisa ha permesso di sviluppare un circostanziato processo di indagine testuale in grado di rispondere alle esigenze tipiche degli studiosi di documenti storico-letterari (Montemagni, 2013). Progetti come BAMBI¹¹ e DIPHILOS¹² hanno posto le basi per una stazione di lavoro digitale e modulare dedicata agli studi filologici; DBT¹³ e Pinakes Text hanno coniugato gli aspetti di rappresentazione del testo propri della filologia digitale con gli aspetti della linguistica computazionale per il trattamento automatico del testo, sviluppando così metodi, modelli e strumenti di filologia computazionale; Greek into Arabic e Traduco si sono concentrati sulla creazione, la rappresentazione e l'analisi del paratesto in relazione a testi e relative traduzioni (per i riferimenti bibliografici dei singoli progetti e per ulteriori informazioni si veda complessivamente Bozzi *et al.*, 2019).

La linea di ricerca mira a sviluppare un ambiente che integri sia l'espressione sia il contenuto del testo sotto forma di un oggetto digitale complesso¹⁴. La proposta si concentra sulle esigenze operative del dominio di interesse, evitando di ingabbiare gli studiosi in ambienti nati per soddisfare esigenze aliene allo studio critico e all'analisi filologica del testo. Un sistema di annotazione, se adeguato, rende possibile sia una rappresentazione esaustiva – orientata cioè alla descrizione dei dati testuali – sia una rappresentazione funzionale – orientata cioè alla definizione delle operazioni ammissibili (Buzzetti, 2004; Renear, 2004; Burnard, 2014). In più, un sistema robusto¹⁵ è capace di operare su por-

¹¹ *Better Access to Manuscript and Browsing of Images* (BAMBI).

¹² *Digital PHILOlogy System* (DIPHILOS).

¹³ *Data Base Testuale* (DBT).

¹⁴ Il termine *oggetto digitale complesso* è usato nel senso indicato in Orlandi, 2010 e Ciotti & Tomasi, 2016.

¹⁵ Sistemi di annotazione con un'espressività limitata hanno difficoltà a gestire segmenti testuali discontinui o note che insistono su porzioni di testo sovrapposte (*overlapping*).

zioni di testo continue o discontinue e su blocchi di lunghezza variabile, e deve essere sufficientemente flessibile per rappresentare numerose tipologie di fenomeni quali pattern letterari, figure retoriche, interventi autoriali, interventi editoriali, ecc. Infine, un meccanismo di annotazione congruo deve essere in grado di gestire molteplici varianti interpretative che insistono su molteplici varianti testuali (Boschetti, 2013), garantendo un'opportuna formalizzazione di questo vincolo¹⁶.

Oggi giorno, le linee guida messe a punto dalla Text Encoding Initiative¹⁷ sono considerate lo standard de facto per una corretta annotazione digitale dei testi (Mancinelli & Pierazzo, 2020). Esse consistono in uno schema di codifica per linguaggi di markup¹⁸ volto alla costruzione e alla conservazione a lungo termine di archivi testuali. Il successo di XML-TEI è in buona parte attribuibile alla sua forte aderenza al modello di testo promosso dalla teoria OHCO (Ordered Hierarchy of Content Object, DeRose et al., 1990) – per il quale il testo è un oggetto complesso espresso sotto forma di una gerarchia ordinata di oggetti di contenuto, per quanto il modello sia stato messo fortemente in discussione e più volte perfezionato (Renear *et al.*, 1996 and Schmidt, 2010).

Data la centralità delle attività di annotazione (Del Gratta *et al.*, 2016), molti strumenti e applicazioni sono stati sviluppati e impiegati durante il lavoro di critica letteraria¹⁹. In seno alle tecnologie dei Linked Open Data²⁰ è stato sviluppato il modello Open Annotation Data Model (Sanderson *et al.*, 2013). Quest'ultima iniziativa, evolutasi nell'attua-

¹⁶ La formalizzazione si realizza ad esempio attraverso la stesura di Context-Free Grammar, come spiegato in Mugelli *et al.*, 2016.

¹⁷ Le linee guida della TEI sono pubblicate online all'indirizzo Web <https://tei-c.org/guidelines/>

¹⁸ Le tecnologie collegate all'ecosistema XML sono mantenute dal consorzio w3c, raggiungibile all'indirizzo Web <https://www.w3.org/standards/xml/>

¹⁹ Tra gli strumenti più apprezzati e documentati se ne possono citare alcuni, quali Pundit, Brat, GATE TeamWare, Hypothesis, Recogito.

²⁰ Il consorzio w3c mantiene le specifiche per la pubblicazione dei dati nel Semantic Web. Al seguente indirizzo è disponibile la descrizione della tecnologia LOD <https://www.w3.org/standards/semanticweb/data>

le specifica del Web Annotation Data Model²¹, è utilizzata per gestire annotazioni tra oggetti del Web Semantico (Pyysalo *et al.*, 2015). Sepur lodevole, nemmeno questa proposta è in grado di gestire le molteplici interpretazioni che possono insistere sulle molteplici versioni di una risorsa. A questo problema cerca di rispondere il progetto CIDOC-CRMInf²² (Van Ruymbeke *et al.*, 2017): esso implementa un'articolata ontologia²³ per descrivere modelli argomentativi.

Tuttavia, il quadro delle attuali tecnologie non risolve le necessità emerse durante le nostre attività di ricerca. Ci si riferisce, cioè alla 1) definizione formale dei casi d'uso coinvolgendo la comunità degli studiosi; 2) produzione di strumenti orientati all'esperienza-utente di specialisti; 3) costruzione di modelli che permettano di trattare molteplici varianti dell'espressione testuale accompagnate da molteplici interpretazioni.

Per i motivi suddetti, accanto all'impiego delle linee guida della TEI (sezione 5.1 e 5.2), il nostro lavoro mira alla sperimentazione di un processo di annotazione flessibile (sezioni 5.3 e 5.4) basato sullo sviluppo di linguaggi specifici di dominio (Fowler, 2010) che consentano di registrare informazioni testuali con politiche di *stand-off annotation*²⁴ e con meccanismi di *micro-tagging* molto noti nell'universo dei social-media (sezione 4.3).

L'attività di annotazione portata sul web è per sua natura distribuita e quindi intrinsecamente collaborativa, in forma sincrona, quando il gruppo di lavoro interagisce in tempo reale sul medesimo documento, oppure asincrona, quando uno o più annotatori lavorano autonomamente.

²¹ Le specifiche del modello sono disponibili all'indirizzo Web <https://www.w3.org/TR/annotation-model/>

²² Le specifiche CIDOC-CRMInf sono disponibili all'indirizzo Web <http://www.cidoc-crm.org/crminf/>

²³ L'ontologia, in questo contesto, è intesa nell'accezione di Gruber, 1993 e Borst, 1997; vale a dire "una specificazione formale ed esplicita di una concettualizzazione condivisa".

²⁴ La *stand-off annotation* è una tecnica per tenere distinti su documenti o parti di documento differenti il testo e il suo insieme di annotazioni, tale che il legame fra i due sia stabilito tramite riferimenti univoci dalle annotazioni ai passi in oggetto (per maggiori informazioni è possibile approfondire al seguente link: <https://bit.ly/2UvKPCd>).

4. METODI PER CREARE ANNOTAZIONI DIGITALI

Le annotazioni digitali possono essere effettuate a diversi livelli di granularità, di interdipendenza, di automatismo e di interoperabilità.

Dal punto di vista della granularità, le macro-annotazioni riguardano un intero corpus o un'intera opera; le meso-annotazioni si applicano ad ampie partizioni del testo, come ad esempio le battute di un'opera teatrale o i capitoli di un libro e le micro-annotazioni si associano alle unità testuali minime, siano esse dotate di significato (come le parole) o prive di significato proprio (come i singoli grafemi).

Dal punto di vista dell'interdipendenza, le annotazioni sono in sé concluse oppure contengono riferimenti a fonti esterne o rimandi ad altre note.

Dal punto di vista dell'automatismo, le annotazioni sono completamente manuali, oppure completamente prodotte dall'applicazione al testo di algoritmi di analisi (linguistica, stilistica, etc.), oppure prodotte automaticamente e corrette o integrate manualmente.

Infine, dal punto di vista dell'interoperabilità, le annotazioni possono essere libere, rispondenti a formati proprietari oppure conformi a standard che garantiscano l'interoperabilità solo sintattica o anche semantica.

A seconda di come si combinano i parametri appena descritti, l'interazione uomo-macchina è favorita da diverse tipologie d'interfaccia utente per la creazione e la gestione di annotazioni, che possiamo distinguere a grandi linee in: interfacce con notevole impiego di grafica avanzata; interfacce con formattazione immediata del testo; interfacce focalizzate sull'inserimento di stringhe di caratteri (testo libero o codice conforme ad una grammatica generativa, nello specifico una *Context-Free Grammar*²⁵).

²⁵ Le *Context-Free Grammars* definiscono o generano linguaggi formali. La gerarchia di Chomsky (1959) prevede quattro livelli crescenti di espressività per le seguenti grammatiche: *Regular Grammars* (che generano *Regular Expressions*), *Context-Free Grammars*, *Context-Sensitive Grammars* e *Recursively Enumerable Grammars*, che hanno la piena espressività di una macchina di Turing.

4.1. Interfacce interattive complesse

Progetti che prevedono annotazioni fortemente strutturate, eseguite non solo da specialisti ma anche da studenti e volontari, con possibili collegamenti puntuali a fonti interne o esterne al corpus oggetto di studio, traggono giovamento da interfacce utente interattive complesse, che guidano passo passo l'annotatore, fornendo vincoli stringenti alla sequenza delle azioni da compiere e all'insieme dei valori selezionabili.

La possibilità d'errore formale (violazione del modello dei dati) o sostanziale (incoerenza dei riferimenti) si riduce al minimo già in fase d'inserimento dell'annotazione, facilitando le successive fasi di revisione, analisi e fruizione dei risultati.

Ad esempio, il progetto Cadmus (Fusi, 2011) fornisce una piattaforma web d'annotazione dove ciascuna unità testuale e le sue parti sono associate a maschere d'inserimento conformi al tipo di dato, con campi da riempire o menù a tendina con valori da selezionare.

Il progetto Perseids (<https://www.perseids.org>) è dotato di una raffinata interfaccia grafica per l'annotazione sintattica che permette la costruzione di grafi complessi per rappresentare le strutture ad albero conformi alle regole della grammatica di dipendenza.

Se si guadagna in precisione, si rischia di perdere in rapidità, automazione e concentrazione. L'interazione con interfacce grafiche elaborate spesso richiede (o almeno promuove) l'uso di più periferiche di input, come il mouse per accedere ai menu o spostare gli elementi grafici sullo schermo e la tastiera per inserire stringhe testuali. Inoltre, i vincoli stabiliti dall'interfaccia penalizzano o addirittura impediscono l'automazione di operazioni anche semplici, come il ritrovamento e la sostituzione di *patterns* identificabili tramite *regular expressions*. Infine, il continuo cambio di modalità mouse/elementi grafici vs tastiera/testo fa aumentare la perdita di concentrazione dall'obiettivo primario, ossia portare a termine una prestazione intellettuale e non meramente meccanica, come l'annotazione richiede.

4.2. Interfacce *What You See Is What You Get*

Tra le applicazioni maggiormente adottate dagli studiosi per trascrivere e annotare documenti letterari, si annoverano i sistemi di desktop publishing. Questo tipo di strumenti implementano complesse interfacce grafiche presentando a schermo il layout del documento in analogia con la stampa²⁶.

Se da un lato tali applicazioni, che nascono per una comunità molto diversa da quella dei critici letterari, sono molto flessibili e immediate per arricchire i documenti con informazioni di stile e formattazione, troppo spesso vengono impropriamente impiegate per registrare informazioni di tipo semantico, critico e filologico.

Per esempio, la figura 4.2a mostra l'annotazione di porzioni di testo per registrare interpretazioni editoriali. Nello specifico, si tratta della trascrizione e parziale annotazione di una lettera (LL1.4 del 1830) scritta dal compositore Vincenzo Bellini (si veda la sezione 5.2). In figura è evidente l'uso di un codice cromatico per determinare le diverse tipologie di dati.

Ulteriori strumenti *general purpose* molto utilizzati sono quelli che consentono la costruzione di fogli di calcolo²⁷. Tali applicazioni sono adottate durante la prima fase di raccolta delle informazioni poiché permettono un'agevole registrazione dei dati relativi agli oggetti di studio.

Infine, la figura 4.2b mostra un documento elettronico puntualmente formattato con strumenti tipografici specializzati e sistemi grafici avanzati²⁸. Questi ultimi permettono di gestire molteplici tipologie di note e di livelli di apparato adottando consolidati metodi di ecdotica (fonti, varianti, note critiche, ecc).

²⁶ Questi tipi di applicazioni sono classificate come applicazioni “WYSIWYG” – What You See Is What You Get. I più famosi sistemi di word processing sono Microsoft Word – della suite Microsoft Office, Apache Writer – della suite Apache OpenOffice, Google Document – della suite Google Drive e altri simili.

²⁷ Si annoverano in questa tipologia di applicazioni esempi come Microsoft Excel, Google Spreadsheet, Apache Calc e altri.

²⁸ Ambienti simili ad Adobe InDesign, oppure al Classical Text Editor.

1. Milano, [luglio 1830] LL1.4
 Vincenzo Bellini a **Vincenzo Ferlito**.¹ Lettera.
 AUT. I-CATm, Bellini lettere. Un foglio, una facciata più indirizzo nel verso (incompleta)
 ED. GIUSEPPE DELOGU, *La casa di Bellini (nel primo centenario della "Sonnambula")*, «Emporium. Rivista mensile illustrata d'arte e coltura», LXXXIII/435 (marzo 1931), pp. 164-175; 169; CAMBI 1943, pp. 252-253; Neri 2005, pp. 172-173.

[...] vi dissi che volea darvi un cenno sù la malattia sofferta ed ecco: La principale ragione fù l'aver scritto a Venezia i **Capuleti** in 26 giorni, ove m'applicava 10 ore di seguito nella mattina ed altre quattro nella sera. Ià qualche volta mi puzzava il fiato per le cattive digestioni: la stagione orrida si era unita pure a farmi soffrire: in Aprile ritornai in Milano e stiedi senza appetito sino che ai 21 di Maggio mi scoppiò una tremenda febbre infiammatoria gastrica biliosa che bisognò farmi un salasso e poi darmi l'emetico: il terzo giorno **Pollini** con l'intervento del dottore volle farmi trasportare in sua casa, perché la mia abitazione era composta di camere strette e basse di tetto, cosa che era pericolosa pel carattere della malattia che potea farla degenerare in putrida maligna: in casa Pollini, in una parola fui assistito con tanta premura ed affetto che non posso descriverlo. Si sono ancora dispendiati, perché io non ho pagato altro che il medico ed i medicamenti; perciò vedete quanti obblighi professo a questa buona **famiglia** che mi ama più che figlio.

La scrittura che forse combinerò con la **Scala**, forse sarà per due **opere** da scriverle una nell'autunno del 1831 ed una nel carnevale del **1832** e per queste l'**impressario** m'ha offerto **4000** ducati; ma io pretendo di più, ossia altri **600** ducati, che vengono ad essere **20000 franchi** cosa che si risolverà fra giorni. Intanto ricevetevi i miei **abbracci** e le tante cose per tutti i parenti

Vostro affezionatissimo nipote
 Vincenzo

FIG 4.2a. Documento Word annotato con strumenti di stile e formattazione

Una tamen reuexio, quae in ad angulum ex
 25 recto quinque octavas habentem, quoniam sus-
 cipit octogoni distinctionem et prope accedit
 A:13r ad dictam dimidii recti mediocritatem, | adep-
 ta est tantum virium ut posset secundariam
 iridem, licet languidiorem generare. 175 Quod
 30 autem talis reflexio prope accedat ad reflexio-
 nem dimidii recti, patet, quoniam excedit tale
 dimidium octava parte recti. Quod vero etiam
 talis reflexio suscipit divisionem octogoni, pa-
 tet, quoniam minimus numerus habens dimi-
 35 dium et octavam est 8. Et ideo oportebit sin-
 gulos quatuor rectos angulos, qui complent to-
 tum ambitum, in octo partes secari, et perinde
 totum circuitum in 32 partes distingui, qui nu-
 40 merus quater numeratur ab octonario. 176 Et
 ob id suscipit octogonalia intervalla, per quae
 successive repetitur repercussio ad exprimen-
 dos iridis secundariae colores. Qui quidem
 in hac secundaria iride sunt languidiores ob

2-3 octavamque post corr. A 25 quinque post corr. A $\diamond$ ante
 octavas del. decimas A 35 octavam post corr. A 37 octo ex
 duodeviginti A 38 32 ex 72 A 72 C 39 quater post corr. A
 $\diamond$ octonario post corr. A

175 in 25^o: Maur. Diaph. Theorema 25 $\diamond$ in additione 25^o:
 Maur. Diaph. Add. Theorema 25

COROLLARIUM

179 Constat ergo quod Sole horizontem occu-
 pante, sicut altitudo primariae iridis est gra-
 70 duum 45, ita tunc secundariae iridis fastigium
 habebit gradus $56 \frac{1}{4}$. Patet, nam gradus 45 sunt
 recti anguli dimidium, gradus vero $56 \frac{1}{4}$ sunt
 recti dimidium et pars octava. 180 Unde duarum
 iridum diametri sunt ad invicem sicut quinque
 75 ad quatuor. Item, Sole non minorem $56 \frac{1}{4}$ gra-
 dibus habente altitudinem, non posse apparere
 secundariam iridem, eiusque altitudinem cum
 Solis altitudine coniunctam efficere $56 \frac{1}{4}$ gradus.

COROLLARIUM

181 Et manifestum fuit, huius secundariae iri-
 dis colores semper esse languidiores coloribus
 primariae, et cur ita sit assignatae sunt causae.]

48 octavam ex nonam A 56 32 ex 72 A 72 $56 \frac{1}{4}$ S ex 55 A
 $56 \frac{1}{4}$ C 73 $56 \frac{1}{4}$ S ex 55 A $56 \frac{1}{4}$ C 74 octava ex nonam A
 $\diamond$ Unde ~ $56 \frac{1}{4}$ gradus $\frac{1}{4}$ in marg. A 76 $56 \frac{1}{4}$ ex 55 A numero
 $56 \frac{1}{4}$ ($\frac{1}{4}$ ex $\frac{1}{4}$) C numero $56 \frac{1}{4}$ S 79 $56 \frac{1}{4}$ gradus $\frac{1}{4}$ ex 55 gradus
 A $56 \frac{1}{4}$ gradus et $\frac{1}{4}$ C S

178 ut praemissa: Maur. Diaph. Theorema 30

[A 12v-13r · C 45v-47r · S 58-60]

FIG. 4.2b. Esempio di documento elettronico formattato con strumenti tipografici avanzati

4.3. Interfacce basate su editor di solo testo

I linguaggi attuali per la rappresentazione dell'informazione semi-strutturata di tipo testuale sono – ancora oggi – riconducibili ai sistemi di annotazione dichiarativa basati su applicazioni XML. L'espressività di tali linguaggi, garantita dal formato testuale e da etichette metatestuali *human-readable*, risponde alla duplice necessità di rappresentare differenti tipologie di annotazioni (quali quelle categoriali, strutturali e relazionali) e di garantire interoperabilità a lungo termine tra infrastrutture e processi computazionali.

Per questo motivo XML si è imposto come tecnologia di riferimento per la codifica digitale di risorse testuali ed è l'unico formato di riferimento per implementare il modello di annotazione promosso dal consorzio TEI. l'XML, tuttavia, non è sempre la scelta ideale sia dal punto di vista pratico sia teorico (Renear, 2004; Ciotti, 2019). Sulla scia dei problemi citati, si registra oggi un rinnovato interesse nell'impiego di sistemi di annotazione orientati alla compilazione tipografica dei testi²⁹ (Fig. 4.3a).

<p>²⁵¹ Sambucam Troglodytae invenire, ut Solinus; ¹²⁸ Tubam aeneam Pisaeus Tyrrenus, ut Plinius, aut multo antea Moyses, ut Iosephus. ¹²¹ Arcades primos in Latium instrumenta musica tulisse, cum antea fistulis pastoralibus tantum uterentur: author est Dionysius. ¹²² Lacedaemones primum tibiis in praelio usos Thucydides; Cretenses in bellum egressos cithara praecinente Gellius; Parthos cum tympanis, Plutarchus et Appianus. ¹²³ Fistulae inventor Pan, teste Plinio et Virgilio; plectri Sappho, ut Suidas et Aelianus. ¹²⁴ Haec carptim hic posita, alibi latius tractantur. Exponetur calculus.</p> <p>28 Pisaeus: Picae 5</p> <p>¹²¹ Homerus est author in hymno ad Mercurium: <i>Hym. hymn. 4</i>. ¹²⁸ Alcorus: cf. <i>Plut. mus. 14 p. 1135f/1136a</i> + <i>Anticles et Inter contratenor: cf. Plut. mus. 14 p. 1136a</i>. ¹²² in sacris literis: <i>gen. 4.21</i> + <i>apud Iosephum: Is. Riv. antiq. 12.2</i> + <i>ut Solinus: cf. Pseudo-Vergilii, De inventibus rebus, 1.15</i> "Sambucum non multum instrumentum auctor Clemente (<i>Gen. Alex. Strom. 1.6.4</i>) Troglodytae Africae, ut Solinus (28.1-2, 31.2-3) plectri, populi invenirentur". ¹²³ ut Plinius: <i>Pin. nat. VI.201</i> + <i>ut Iosephus: Is. Riv. antiq. III.12.6</i>. ¹²⁴ author est Dionysius: <i>Dion. Halic. antiq. 1.33.4</i>. ¹²² Thucydides: <i>Thuc. V.76</i> cf. <i>Gell. I.1.1-1.5</i> + <i>Gellius: Gell. I.1.6</i> + <i>Plutarchus et Appianus: Plut. Cons. 2.2</i> + <i>Plinius: Pin. nat. VI.204</i> + <i>Virgilius: Verg. Ed. 2.32-33</i> + <i>Suidas: Suid. sigma 107 Ador</i> + <i>Aelianus: cf. Athen. XIV.36 p. 635b</i>.</p> <p>[199]</p>	<pre> \edtext{Pisaeus} {\footnotetext{Pisaeus\tag{:}\variantsep Piscus \tag{5}}}} Tyrrenus, \edtext{ut Plinius} {\footnotetext{\tag{Plin. \textit{nat.}} VII.201}}}, aut multo antea Moyses, \edtext{ut Iosephus} {\footnotetext{\tag{Ios. Flav. \textit{antiq.}} III.12.6}}}, \ledunit Arcades primos in Latium instrumenta musica tulisse, cum antea fistulis pastoralibus tantum uterentur: \edtext{author est Dionysius} {\footnotetext{\tag{Dion. Halic. \textit{antiq.}} I.33.4}}}, \ledunit Lacedaemones primum tibiis in praelio usos \edtext{Thucydides} {\footnotetext{\tag{Thuc. V.76; cf. Gell. I.1.1-1.5}}}, Cretenses in bellum egressos cithara praecinente \edtext{Gellius} {\footnotetext{\tag{Gell. I.1.6}}}; </pre>
--	--

FIG. 4.3a. Sistema di annotazione e resa grafica di un documento elettronico compilato con tecnologia LaTeX

Chi opera all'interno di un dominio circoscritto, come quello della lessicografia, della filologia o degli studi letterari, ha familiarità con

²⁹ Per esempio l'ambiente TeX.

l'organizzazione delle informazioni tramite linguaggi controllati e semi-strutturati, come il linguaggio degli apparati critici, dei repertori bibliografici, delle voci di dizionario etc.

Sono linguaggi concisi, ottimizzati per un singolo dominio di conoscenza. Bastano solo alcuni caratteri tipografici, come i due punti (:), la tilde (~), la *manicula* (☛), per esprimere la semantica delle parti di una voce di dizionario, distinguendo ad esempio il lemma dalle diverse accezioni e dagli impieghi del termine in contesto.

Se si eliminano le eccezioni sintattiche e le ambiguità semantiche, si può passare da un linguaggio informale o semiformale ad un linguaggio formale, rendendo l'annotazione sempre più *machine actionable* e non semplicemente *machine readable*.

Un Domain-Specific Language (DSL) (cf. Parr, 2014) è un linguaggio formale ottimizzato per esprimere, in maniera concisa ed efficace, entità³⁰, relazioni, proprietà e operazioni pertinenti ad un singolo dominio di conoscenza. Ad esempio, nell'ambito dei database, lo Structured Query Language (SQL) è un linguaggio specifico di dominio per interrogare una banca dati. Un DSL è descritto dalla propria Context-Free Grammar (CFG) che ne definisce il lessico e i costrutti. Nell'approccio che adottiamo, la CFG viene realizzata insieme agli studiosi o agli studenti che effettuano le annotazioni per rispondere a un'ipotesi di ricerca. Il DSL è un linguaggio familiare all'annotatore. Nei nostri progetti ciò significa, da un lato, essere vicini alle convenzioni tradizionali adottate nell'ambito degli studi umanistici (per apparati critici, etc.) e, dall'altro, essere vicini alle nuove convenzioni diffuse sui social media, come l'uso di *hashtag* per marcare parole chiave e trend topics.

Come dimostrato in Bambaci *et al.* (2019), se il DSL creato ha la medesima espressività di un sottoinsieme TEI, è sempre possibile attuare una traduzione di formato dal DSL a TEI, mediante serializzazione del codice DSL in XML. Infatti, l'uso del DSL non costringe ad

³⁰ Le ontologie modellano astrattamente i concetti di un dominio (entità), e vengono definite o popolate facendo ricorso a schemi (come RDF *schema*) di linguaggi *general-purpose* (come XML) oppure a DSL (come Turtle o altri linguaggi creati per esigenze specifiche).

abbandonare il modello testuale descritto dalle specifiche TEI. Se lo si adotta pedissequamente, la corrispondenza sarà biunivoca, mentre se lo si adotta solo parzialmente, perché si ha bisogno di un'espressività maggiore, sarà comunque possibile convertire in XML-TEI la parte di DSL che adotta il medesimo modello testuale. In questo modo si allontana il rischio di assenza di interoperabilità fra progetti di annotazione, ma nel caso in cui sia necessaria maggiore espressività, si può creare un DSL in grado di gestirla, delegando a un secondo momento la partecipazione ai tavoli di lavoro per l'estensione degli standard, ad esempio tramite i SIG della TEI.

5. CASI DI STUDIO

In questa sezione presentiamo quattro casi di studio per esemplificare come la varietà di parametri e metodi discussa all'inizio della sezione 4, congiunta a vincoli di gestione dei singoli progetti, porti a scelte tecnico-implementative che ci hanno permesso di ottimizzare tempo, risorse di calcolo e risorse umane.

Per ciascun caso di studio vengono descritte le ipotesi di ricerca e gli obiettivi (scientifici, didattici o divulgativi) dei progetti, le tecnologie utilizzate, le analisi fatte sui dati, le modalità di fruizione, le risorse operative offerte dalle istituzioni coinvolte e i profili degli annotatori.

Il primo progetto, *L'edizione digitale delle postille di Bassani*, è stato condotto in collaborazione con la dottoranda Angela Siciliano, sotto la direzione scientifica di Marina Riccucci (Università di Pisa), Enzo Neppi (Université de Grenoble-Alpes) per gli aspetti filologici e la direzione tecnico-scientifica di A.M. Del Grosso per gli aspetti di codifica.

Il secondo, *BellinInRete Digital Correspondence*, supportato dal Fondo Europeo Sviluppo e Coesione 2014–2020: Piano per il Mezzogiorno, è stato condotto in collaborazione con Graziella Seminara e Erica Capizzi (Università di Catania), Salvatore Cristofaro (CNR-ISTC di Catania), Emiliano Giovannetti e Simone Marchi (CNR-ILC di Pisa), sotto la direzione tecnico-scientifica di A.M. Del Grosso e Daria Spampinato (CNR-ISTI di Catania) per gli aspetti di codifica e di fruizione.

I materiali sono stati forniti dal Museo Civico Belliniano di Catania, per la valorizzazione del patrimonio posseduto.

Il terzo, *Studi linguistici sulle traduzioni leopardiane della Batracomiomachia*, è stato condotto in collaborazione con l'allora laureando Matteo Cazzato (Università di Pavia), sotto la guida scientifica di Federico Boschetti per gli aspetti linguistici, filologici e di codifica.

Infine il quarto, *I Nomi della Rosa*, riguardante le semplificazioni d'autore di Umberto Eco alla seconda edizione del 2010 del suo romanzo, è stato condotto in collaborazione con l'allora studente di Master in Digital Humanities Christian D'Agata (Università Ca' Foscari Venezia) sotto la direzione scientifica di Marina Buzzoni (Università Ca' Foscari Venezia) per gli aspetti di codifica del testo delle edizioni, la direzione scientifica di Daniele Silvi (Università di Tor Vergata) per gli aspetti di distant reading e di Federico Boschetti per gli aspetti di progettazione della piattaforma e l'elaborazione del DSL d'annotazione.

5.1. L'edizione digitale delle postille di Bassani

Il progetto di edizione digitale delle postille di Giorgio Bassani³¹ si inserisce in un'azione più ampia (Siciliano, 2020) dedicata allo studio della biblioteca del letterato bolognese³². In particolare, l'iniziativa si prefigge di fornire un testo critico di tutte le annotazioni appuntate dallo scrittore sui suoi volumi (Siciliano, 2018; Siciliano, 2019). Le postille consentono di ricostruire la formazione giovanile dell'autore, il debito nei confronti delle fonti e le modalità della loro rielaborazione. Accanto alle annotazioni testuali, le postille non verbali (o mute) rappresentano

³¹ Giorgio Bassani (Bologna 1916–Roma 2000) è stato una figura di primo piano della cultura italiana del Novecento. Accanto alla produzione narrativa di alto spessore, che gli valse prestigiosi riconoscimenti (Premio Strega nel 1956 con le *Cinque storie ferraresi*; Premio Viareggio nel 1962 con la sua opera più nota, *Il giardino dei Finzi-Contini*; Premio Campiello nel 1969 con *l'Airone*), ricordiamo l'attività nel campo editoriale (nelle riviste 'Botteghe Oscure' e 'Paragone', e per la casa editrice Feltrinelli) e l'impegno per la causa ambientale e culturale, con la fondazione e la direzione di 'Italia Nostra'.

³² La biblioteca è conservata presso i locali della Fondazione "Giorgio Bassani" di Ferrara.

una modalità di evidenziazione importante: sottolineature interlineari, barre laterali, asterischi (Fig. 5.1a).

FIG. 5.1a. Riproduzione meccanografica per lo sviluppo dell'edizione facsimilare e ultra-diplomatica

La necessità di descrivere le postille ha sollevato da subito il problema di come strutturare, conservare, elaborare e pubblicare correttamente tali annotazioni (cf. §2 e §4). Ne consegue la scelta quasi obbligata di adottare metodi e tecniche digitali³³. Infatti, come anticipato in §2 e §3, un'edizione digitale è una soluzione efficace per rappresentare i fenomeni testuali oggetto di indagine (esaustività) e per soddisfare tutti i requisiti editoriali di visualizzazione e fruizione del testo (funzionalità). In primo luogo attraverso la riproduzione facsimilare della pagina (cf. §2 e §3) si consente la visualizzazione immediata delle postille e delle relative sezioni di testo (Fig. 5.1a).

Alcune postille rimandano ad altre pagine dello stesso volume o ad altri volumi dello stesso autore oppure contengono citazioni di brani, versi o concetti presenti in altre opere. Questo tipo di connessioni

³³ L'edizione digitale delle postille di Giorgio Bassani è per il momento limitata a un solo volume (Guido Calogero, *La scuola dell'uomo*, Firenze, Sansoni, 1939).

in un'edizione cartacea sarebbero espresse in modo non adeguato col rischio di perdere importanti informazioni (cf. §1 e §4).

Ulteriore vantaggio è la possibilità di svolgere interrogazioni che permettano di ricostruire la scansione cronologica delle note, la loro stratificazione nel volume e la metodologia postillatoria.

Dal punto di vista ermeneutico, sia le annotazioni – espressione analitica e critica dello studioso – sia la conseguente elaborazione automatica – finalizzata alla fruizione avanzata dei contenuti digitalizzati – sono istanze del modello discusso nelle sezioni precedenti (cf. §2, §3 e §4) e strumenti collaborativi di indagine. Esse permettono, infatti, un'efficace combinazione tra metodi di *close reading* (analisi puntuale di piccole porzioni di testo registrate sotto forma di annotazioni inline o in stand-off) e metodi di *distant reading* (analisi esplorativa di grandi quantità di dati) e, infine, ulteriori metodi di *close reading* alla luce dei dati elaborati (condotti generalmente con tecniche di stand-off annotation per aumentare la significatività dei dati senza dover sincronizzarne le dipendenze). In tal modo, l'accesso funzionale al testo per recuperare efficientemente informazioni quantitative, quali il numero di postille vergate a matita rossa, quante a lapis, quante volte ricorre una barra laterale doppia, quante il nome di uno studioso (Tab. 5.1), è la chiave per un'ulteriore investigazione puntuale che può essere condotta dallo stesso studioso oppure da altri colleghi della comunità.

TAB. 5.1. Statistiche portate sui primi dati analizzati nell'edizione digitale delle postille di Bassani

<i>Elementi</i>	<i>Quantità</i>
pagine	120
postille	305
categorie in tassonomia	64
entità nominate	13
occorrenze annotate di entità	16

Il modello di annotazione così implementato permette, quindi, di studiare la metodologia di lavoro di Bassani sia in fase di codifica sia

poi in fase di interrogazione: il fine ultimo del lavoro è l'interpretazione dei segni usati da Bassani per postillare i testi letti, alla luce di evidenze quantitative (indici per i conteggi) e qualitative (concordanze per la valutazione di termini in contesto).

Durante la prima fase di lavoro, il processo di annotazione ha risposto alle necessità editoriali più significative seguendo le linee guida della *Text Encoding Initiative*. Tra i requisiti editoriali individuati, sia per l'edizione digitale sia per l'edizione a stampa si citano (Fig. 5.1b):

- Rispettare le abitudini grafiche dell'autore, con criterio conservativo.
- Emendare gli errori non corretti da Bassani; sciogliere le abbreviazioni; integrare le espressioni rilevate talvolta omesse.
- Riprodurre la sottolineatura d'autore e la barra laterale nonché l'a capo.
- Segnalare la posizione della postilla rispetto al testo.
- Indicare i dati tipografici, gli inchiostri e gli strumenti utilizzati.

Dai primi capitoli di questo libro dovrebb'essere risultato chiaro che di spirito, di pensiero, di soggettività, di coscienza o non posso più seriamente parlare, quando voglia intendere lo sviluppo del pensiero moderno e favorirne l'ulteriore avanzamento, se non dando a quei termini ► il significato di *spirito mio*, *pensiero mio*, soggettività mia, coscienza mia, mia e di tutti coloro che io ponga in me come me, nell'attuazione della mia esperienza morale. ◀ [p. 108, rr. 15-24] (L.) ¶

¶

*stupendo | lo Spirito, si attua | non per una sorta | di trascendenza, | *ma per consapevole* | *atto morale** ¶

FIG. 5.1b. Testo digitale per l'edizione a stampa delle postille di Bassani, con le specificità tipografiche per accomodare le esigenze editoriali evidenziate

La particolarità della proposta risiede nell'approccio di codifica multi-dimensionale. Nello specifico, è stato adottato il metodo *embedded transcription*³⁴ per la rappresentazione del documento – il quale associa la riproduzione facsimilare alla corrispondente trascrizione della postilla (unendo l'integrità dell'espressione con quella di contenuto, cf. §2, §3 e §4) cf. circoscrivendo, in più, la pericope a cui la postilla è riferita (Fig. 5.1d).

³⁴ Il modello di codifica “embedded transcription” è descritto al seguente indirizzo Web, <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHZLAB>

Accanto alla dimensione documentale, l'edizione digitale presenta, quindi, anche la ricostruzione editoriale del contenuto, determinando così la relativa dimensione testuale. Infine, la catalogazione di ciascuna postilla è stata implementata con gli opportuni elementi del modulo per la descrizione del manoscritto (Fig. 5.1c).

```
<msItem n="113" xml:id="p113" class="#postilla_filosofica_di_commento #sottolineatura_interlineare"
  facs="#pag108.p4">
  <msItem n="113.1" class="#postilla_filosofica_di_commento">
    <locus target="#t108.4" facs="#pag108.p4det"> p. 108 </locus>
    <textLang> italiano </textLang>
    <note resp="#AS"> postilla verbale a lapis, sul margine sinistro </note>
  </msItem>
  <msItem n="113.2" class="#sottolineatura_interlineare">
    <locus facs="#pag108.p4det1"> p. 108 </locus>
    <note resp="#AS"> postilla muta, costituita da sottolineatura interlineare; a lapis </note>
  </msItem>
</msItem>
<msItem n="114" xml:id="p114" class="#postilla_filosofica_di_commento
#barra_laterale_ondulata_e_sottolineatura_interlineare" facs="#pag108.p5">
  <msItem n="114.1" class="#postilla_filosofica_di_commento">
    <locus target="#t108.5" facs="#pag108.p5det"> p. 108 </locus>
    <textLang> italiano </textLang>
    <note resp="#AS"> postilla verbale a lapis, sul margine sinistro; presenta alcune parti
    sottolineate; è preceduta e seguita da asterisco </note>
  </msItem>
  <msItem n="114.2" class="#sottolineatura_interlineare">
    <locus facs="#pag108.p5det1m1"> p. 108 </locus>
    <note resp="#AS"> postilla muta, costituita da sottolineatura interlineare; a lapis </note>
  </msItem>
  <msItem n="114.3" class="#postilla_muta">
    <locus facs="#pag108.p5det1m2"> p. 108 </locus>
    <note resp="#AS"> postilla muta, costituita da doppia barra laterale, sul margine destro; a
    lapis </note>
  </msItem>
  <msItem n="114.4" class="#barra_laterale_ondulata">
    <locus facs="#pag108.p5det1m3"> p. 108 </locus>
    <note resp="#AS"> postilla muta, costituita da barra laterale ondulata, sul margine sinistro; a
    lapis </note>
  </msItem>
</msItem>
```

FIG. 5.1c. Frammento di documento XML-TEI per la catalogazione delle postille impiegando il modulo msDesc

Per il prosieguo del progetto è prevista l'integrazione di due ulteriori tipologie di annotazione: da un lato la marcatura TEI per tutto ciò che pertiene al farsi del testo (cancellature, correzioni, trasposizioni, sottolineature, ecc.) e dall'altro, l'adozione delle tecnologie del Web Annotation Data Model al fine di aggiungere, attraverso un sistema citazionale

globale³⁵, informazioni sulle entità nominate e su connessioni di varia natura.

La postilla in figura 5.1d si riferisce a quanto annotato dallo stesso Bassani alla pagina 108 del volume selezionato.

```
<zone xml:id="pag108.p5" ulx="132" uly="175" lrx="165" lry="198">
  <zone xml:id="pag108.p5det" ulx="22" uly="65" lrx="76" lry="34" hand="#m1">
    <line>
      <metamark function="highlighting" rend="asterisk" place="at_the_beginning_of_the_note" />
      <hi rend="italic"> stupendo </hi>
    </line>
    <line> lo Spirito, si attua </line>
    <line> non per una sorta </line>
    <line> di trascendenza, </line>
    <line>
      <hi rend="italic"> ma per consapevole </hi>
    </line>
    <line>
      <hi rend="italic"> atto morale </hi>
    </line>
    <metamark function="highlighting" rend="asterisk" place="at_the_end_of_the_note" />
  </zone>
  <zone xml:id="pag108.p5det1" ulx="122" uly="144" lrx="123" lry="166">
    <zone xml:id="pag108.p5det1m1" ulx="24" uly="52" lrx="78" lry="84" hand="#m1">
      <metamark function="highlighting" rend="underlining" place="inline" target="#pag108.p5det1m1" />
    </zone>
    <zone xml:id="pag108.p5det1m2" ulx="26" uly="85" lrx="96" lry="134" hand="#m1">
      <metamark function="highlighting" rend="double_straight" place="vertical_line" target="#pag108.p5det1 #pag108.p5det" />
    </zone>
    <zone xml:id="pag108.p5det1m3" ulx="24" uly="52" lrx="78" lry="84" hand="#m1">
      <metamark function="highlighting" rend="wavy" place="vertical_line" target="#pag108.p5det1 #pag108.p5det" />
    </zone>
  </zone>
</zone>
```

FIG. 5.1d. Codifica delle postille con metodo embedded transcription: si nota l'elemento metamark, specificandone le caratteristiche con l'attributo function, e lo si collega a postilla marcata

³⁵ L'architettura CITE-CTS è tra le tecnologie impiegabili a tale scopo. La tecnologia è descritta all'indirizzo Web <http://cite-architecture.org>

5.2. L'edizione digitale delle lettere di Bellini

L'edizione “*BellinInRete Digital Correspondence*”, sviluppata nel contesto del progetto “Museo Virtuale della Musica BellinInRete”³⁶, ha l'obiettivo di rendere fruibile un corpus di 41 lettere scritte dal compositore Vincenzo Bellini (1801–1835).

BellinInRete³⁷ è un progetto fortemente interdisciplinare che ha visto impegnate diverse figure professionali con competenze complementari e l'implementazione di un processo di annotazione collaborativo, coinvolgendo anche studenti di vari corsi di laurea in Digital Humanities.

Il progetto segue prassi consolidate nell'ambito del *digital scholarly editing* (prassi, metodi e strumenti descritti in §2, §3 e §4) avvalendosi perlopiù del vocabolario XML-TEI per la codifica della corrispondenza (Fig. 5.2a).

```
<profileDesc>
  <correspDesc>
    <correspAction type="sent">
      <persName ref="TEI-ListPerson.xml#VB" role="composer">Vincenzo Bellini</persName>
      <placeName ref="TEI-ListPlace.xml#NSS" type="postmark">Neuilly-sur-Seine</placeName>
      <date when="1834-06-27" type="postmark">27 giugno 1834</date>
    </correspAction>
    <correspAction type="received">
      <persName ref="TEI-ListPerson.xml#CP" role="librettist">Carlo Pepoli</persName>
      <placeName ref="TEI-ListPlace.xml#Paris" type="writing">Paris</placeName>
      <date when="1834-06" type="postmark">giugno 1834</date>
    </correspAction>
    <correspContext>
      <p>non è presente un contesto di riferimento</p>
    </correspContext>
  </correspDesc>
  <textClass>...
</textClass>
  <langUsage>...
</langUsage>
</profileDesc>
```

FIG. 5.2a. Frammento XML-TEI per la codifica della corrispondenza usando il tagset `correspDesc`

³⁶ L'obiettivo dichiarato del progetto è quello di creare un “cambiamento duraturo” per la valorizzazione del Museo Civico Belliniano di Catania, grazie all'impiego di tecnologie multimediali digitali.

³⁷ Sia per la trascrizione sia per le annotazioni, il progetto usa come fonte principale la recente edizione critica Bellini, 2017 a cura di Seminara.

In particolare, sono stati impiegati i tagset *correspDesc*³⁸ e *msDesc*³⁹ (Del Grosso *et al.*, 2019). Il primo, contenuto nel descrittore TEI *profileDesc*, è adatto alla registrazione della corrispondenza, mentre il secondo, presente nel modulo TEI “*manuscript description*”, è usato per la descrizione fisica delle lettere originali.

<i>Entità</i>	<i>Numero di entrate</i>
persone	35
luoghi	14
opere	7
termini	16
organizzazioni	8
bibliografia	10
TOTALE	90

TAB. 5.2. Quadro sinottico delle entità descritte ed annotate nel carteggio belliniano

Come mostra la tabella 5.2, notevole attenzione è stata posta per la costruzione di numerose liste di entità nominate (luoghi, persone, organizzazioni, eventi) e non nominate (opere, termini, note critiche), che andranno a costituire parte degli indici per la fruizione avanzata dell'edizione (cf. §2). In più, per cercare di raggiungere sia l'eshaustività sia la massima funzionalità dei dati, sono stati curati gli aspetti metatestuali relativi alla conformazione fisica delle lettere, come la modalità di piegatura e le condizioni di conservazione, i timbri postali, la presenza del sigillo in ceralacca (Fig. 5.2b).

³⁸ <http://www.tei-c.org/release/doc/tei-p5-doc/it/html/ref-correspDesc.html>

³⁹ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-msDesc.html>

FIG. 5.2B. Immagine con timbro in ceralacca e indicazione del destinatario

In tale scenario anche le buone pratiche del Semantic Web⁴⁰ sono state impiegate per rispondere ai requisiti di annotazione di carattere funzionale emersi nel progetto, con particolare attenzione ai dati di dominio musicale. Infatti, la realizzazione dell'edizione delle lettere di Bellini si è articolata su due fronti: da un lato, la codifica TEI dei fenomeni testuali e dei metadati e, dall'altro, la descrizione delle entità rilevanti per mezzo di formalismi e vocabolari definiti in ambiti propri del Semantic Web.

Inoltre, il processo di annotazione collaborativa sviluppato nel corso del progetto consente di registrare a diversi livelli di granularità i collegamenti testo-immagine mediante tecnologie utili all'identificazione univoca e globale delle regioni d'interesse (cf. §2, §3 e §4). A tale scopo sono stati scelti i protocolli CTS per gli aspetti citazionali del testo, e del IIIF per gli aspetti di recupero delle immagini.

⁴⁰ Sono stati utilizzati repertori pubblici e autorevoli quali VIAF (Virtual International Authority File, <https://viaf.org>), Geonames (<http://www.geonames.org>), CIDOC-CRM (<http://www.cidoc-crm.org>) e RISM (Répertoire International des Sources Musicales, <http://www.rism.info>).

Contestualmente, è stato realizzato un primo ambiente Web di fruizione⁴¹ (Fig. 5.2c) personalizzando due strumenti software open source: *Edition Visualization Technology*⁴² e *Omega Scholarly platform*⁴³.

FIG. 5.2c. Interfaccia di fruizione Web per il progetto di edizione digitale delle lettere di Bellini

⁴¹ Il prototipo (<http://licodemo.ilc.cnr.it/bellini>) rappresenta un efficiente visualizzatore di immagini adottando tecnologie Web ampiamente utilizzate, come OpenSeadragon (<https://openseadragon.github.io>) per la visualizzazione di immagini IIIF. Una tra le attività in corso più interessanti è la definizione di una ontologia dei contenuti del patrimonio belliniano (Cristofaro & Spampinato, 2020), la quale delinea formalmente i concetti notevoli presenti nelle lettere, ricorrendo a schemi di metadati standard per la modellazione di oggetti culturali, come RDA (Resource Description and Access, <http://www.rda-rsc.org>), FRBROO (Functional Requirements for Bibliographic Records object oriented, <http://www.cidoc-crm.org/frbroo/home-0>) e FOAF (Friend Of A Friend, <http://xmlns.com/foaf/spec>).

⁴² EVT (Edition Visualization Technology), un progetto open source, ideato per la navigazione Web di edizioni digitali di manoscritti.

⁴³ Omega Digital Platform è un progetto di ricerca e sviluppo in corso di realizzazione presso l'Istituto di Linguistica Computazionale, che mira ad implementare un ambiente digitale flessibile e modulare per lo studio scientifico dei testi. Il codice sorgente del progetto è disponibile online: <https://github.com/literarycomputinglab/OmegaProject>; E' possibile approfondire gli aspetti principali dell'iniziativa nei seguenti contributi: Del Grosso *et al.*, 2016; Del Grosso *et al.*, 2017; Del Grosso *et al.*, 2017b; Del Grosso *et al.*, 2018.

La Fig. 5.2c mostra l'edizione digitale con la riproduzione facsimilare di ciascuna lettera autografa accompagnata dalla sua trascrizione (cf. §2). Attraverso le diverse modalità di accesso e di ricerca, è possibile navigare il carteggio in modo puntuale ed esaustivo, presentando all'utente finale le informazioni più rilevanti (cf. §2 e §3).

5.3. Studi linguistici sulle tre redazioni leopardiane della traduzione della *Batracomiomachia*

Giacomo Leopardi, prima di dedicarsi al poemetto eroicomico in ottave dei *Paralipomeni alla Batracomiomachia*, studia a lungo il testo greco da cui l'idea trae origine, tanto da produrre tre redazioni della traduzione, in sestine, della *Batracomiomachia* antica.

A differenza dei precedenti progetti (§5.1 e §5.2), in cui i materiali sono stati codificati direttamente in XML-TEI, si è scelto di definire un linguaggio specifico di dominio (DSL: cf. §4), per un approccio più adatto al coinvolgimento di studenti e studiosi privi di familiarità con le pratiche diffuse nella comunità dei filologi digitali (cf. §4.3).

Matteo Cazzato, durante un periodo di tirocinio presso il CNR-ILC di Pisa, ha elaborato dunque un DSL per annotare le differenze stilistiche e linguistiche fra le diverse redazioni.

Il linguaggio specifico risultante combina le informazioni di un apparato critico positivo, non troppo dissimile da come sarebbe disposto in un'edizione a stampa, con le informazioni derivanti dall'analisi linguistica delle varianti d'autore. L'efficienza del DSL è data dalla sua elevata espressività ecdotica (rappresentazione delle varianti testuali) ed ermeneutica (interpretazione delle varianti diastratiche e diafasiche, sostanziate in variazioni di registro linguistico). L'efficacia del linguaggio è data dalla sua concisione (soprattutto rispetto al vocabolario XML-TEI) e dalla sua comprensibilità per chi sia familiare con il dominio filologico-linguistico.

Come si vede in Fig. 5.3a, la pagina è suddivisa in sette blocchi: il primo contiene il testo greco, il secondo e il terzo rispettivamente le differenze fra la terza e la seconda redazione e fra la seconda e la prima redazione; i tre blocchi successivi contengono il testo di ciascuna redazione e l'ultimo permette l'inserimento delle annotazioni.

Ogni annotazione è costituita da un riferimento ai luoghi di variazione delle diverse traduzioni, messi in relazione fra di loro dai due punti, e dall'elenco di marcatori linguistici e stilistici creati dallo studente. Il marcatore, prefissato da cancelletto (#), è costituito da due parti, separate da trattino basso (_): la prima parte indica le redazioni (ad es. *a*, oppure *bc*) e la seconda indica una valutazione sul registro linguistico della variante o su altri fattori linguistici e/o stilistici.

Anche se il lavoro è stato interrotto alla conclusione del tirocinio e presta il fianco a critiche di eccessiva soggettività nei criteri di classificazione, ci ha permesso di elaborare una notazione quanto più possibile vicina a quella familiare ad un filologo, almeno per la parte di riferimento alle varianti.

È importante sottolineare che il DSL creato ha la stessa espressività dei sottoinsiemi di specifiche TEI dedicati rispettivamente alla codifica delle varianti (*critical apparatus*: modulo 12 delle TEI Guidelines⁴⁴) e alla codifica di annotazioni strutturate (*feature structures*: modulo 18 delle TEI Guidelines⁴⁵), come discusso in §4.3.

Hom. 9 Μῦθος ποτε θεῖατος γαῖης κίνδυνον ἀνείρεν,
Hom. 10 πῶταρον ἐν λίανι ἄλκων προορῶντες νέεον,
Hom. 11 ὕπαι γερμένους μέληδ' ἰὼν δὲ κατέϊδε
Hom. 12 λιγυῶρας πολέμοιο, ἔπος δ' ἐθέλγιστο τοῖον

1826 - 1822
Un topo *un*, di *e* fra' *topi* il più *leggiadro*,
Venne d'un lago in *asu* la *fangosa* sponda: un *epigondarno*.
Campato *all'opcor d'un* *gatto* anzi *eracuto* e da un *legadro*.
Aschela *Ch* *insaguito timor* *con* *la* *freseh* l'aves per quel *dirindarno*:
Medianevo, *l'ha* *ovewisi* *a* *un* *ber*, *guarvendo* un *ranocchio*,
D *l'ai* *piaggando* *el* *stagno* *a* *lui* *i* *vicin*, *gli* *vi* *pose* l'occhio.

1822 - 1815
Un topo *un* di, fra' *topi* il più *ben* *leggiadro*,
Venne d'un lago al la *fangosa* sponda:
Se *Campato* *egali* *elora* *all'ard* un *ega* *un* *to* *ben* *astuto* e *gladiro*,
E *Accaimelava* il *timor* con la *freseh* *onda*:
Mentre *beveva*, un *garulo* *ranocchio*
Dalla *pigro* *palude* *stagno* a lui *rivolse* l'occhio.

1826
C.1.4.1 Un topo, de le membra il più ben fatto,
C.1.4.2 Venne d'un lago in su la sponda un giorno.
C.1.4.3 Campato poco innanzi era da un gatto
C.1.4.4 Ch'insaguito l'aves per quel dirindarno:
C.1.4.5 Stanco, fessasi a ber, quando un ranocchio,
C.1.4.6 Passando da vicin, gli pose l'occhio.

1822
B.1.4.1 Un topo un di, fra' topi il più leggiadro,
B.1.4.2 Venne d'un lago a la fangosa sponda:
B.1.4.3 Campato all'or d'un gatto astuto e ladro,
B.1.4.4 Achela Ch'insaguito il timor con la fresch'onda:
B.1.4.5 Mentre beveva, un garulo ranocchio
B.1.4.6 Dal pigro stagno a lui rivolse l'occhio.

1815
A.1.4.1 Un topo un di, fra' topi il più ben fatto,
A.1.4.2 Venne d'un lago alla fangosa sponda:
A.1.4.3 Scampato egli era allor da un tristo gatto,
A.1.4.4 E calmava il timor colla fresch'onda:
A.1.4.5 Mentre beveva, un garulo ranocchio
A.1.4.6 Dalla palude a lui rivolse l'occhio.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

FIG. 5.3a. Annotazione tramite DSL della *Batracomiomachia* su Euporia

⁴⁴ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>

⁴⁵ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html>

5.4. I nomi della rosa

Umberto Eco nel 2012 pubblica una nuova edizione del suo best seller *Il nome della rosa*, riveduta e corretta rispetto alla prima edizione del 1980. Se l'autore minimizza la portata dei suoi interventi fra un'edizione e l'altra, la dimensione e il valore di tale operazione possono essere in realtà messe in luce soltanto dal critico letterario che procede allo studio analitico delle varianti d'autore, forte dei metodi della filologia (digitale) e della linguistica (computazionale).

Christian D'Agata⁴⁶, per un saggio finale del Master in Digital Humanities dell'Università Ca' Foscari Venezia, ha svolto un accurato lavoro d'analisi volto ad acquisire le prove di una tendenza generale alla semplificazione del testo. Le varianti sono state individuate automaticamente tramite la procedura di allineamento fra versioni diverse di uno stesso testo messa a disposizione nella libreria di componenti software di CollateX⁴⁷.

Il nuovo DSL estende l'espressività del linguaggio descritto in §5.3 organizzando l'analisi delle varianti su due piani: l'uno resta aperto all'interpretazione soggettiva, tramite osservazioni stilistiche e storico-letterarie, l'altro invece è basato su rigorose evidenze linguistiche, tramite l'assegnazione a ciascuna parola oggetto di variazione della sua marca d'uso, così come è registrata nel lavoro di De Mauro (cf. De Mauro, 2000).

Per l'annotazione, è stato ideato un DSL in grado di descrivere il luogo e la natura della variazione (inserzione, sostituzione, trasposizione o cancellazione di testo) e di associare alla variante marcatori filologici e linguistici, soprattutto per indicare interventi testuali (ad es. la cancellazione) e la frequenza d'uso delle componenti lessicali (ad es. #FO: lessico fondamentale), nonché stilistici e storico-letterari, per indicare prevalentemente figure retoriche (ad es. antonomasia) e modificazioni del registro linguistico (ad es. semplificazione).

⁴⁶ Maggiori informazioni si trovano in <https://www.variantidellarosa.it> e in D'Agata, 2020.

⁴⁷ <https://collatex.net>

Nell'esempio riportato in Fig. 5.4a si vede come la variante del 2012, “in questo libro Aristotele”, sostituisca la lezione del 1980. I marcatori stilistici e storico-letterari, prefissati dal segno di sezione (§), indicano la semplificazione, il mutamento degli orizzonti culturali del lettore modello, l'esplicitazione e la soppressione di una figura retorica; dopo la tilde (~) si trovano le singole parole delle due varianti con le marche d'uso (#LE: letterario, #FO: fondamentale, #TS/stor: termine specialistico di ambito storico) e categorie grammaticali (#nomeProprio/autore).

La parte precedente il segno di uguale (=) di ciascuna annotazione è stata prodotta automaticamente⁴⁸ durante la fase di identificazione e marcatura delle varianti sul testo. L'annotatore interviene manualmente al fine di aggiungere le analisi.

FIG. 5.4a. Annotazione tramite DSL de *Il nome della rosa* su Euphoria

Attraverso un interprete⁴⁹ del DSL, è facile tradurre il linguaggio d'annotazione compatto, ma non standard, usato in questo progetto in codice XML con uno schema proprietario che, grazie a semplici fogli di trasformazione XSL, diventa conforme alle linee guida della TEI P5, secondo i principi discussi in §4.3.

Il lavoro di Christian D'Agata è riuscito a dimostrare, con metodi quantitativi applicati alle marcature linguistiche e stilistiche adottate, che l'edizione del 2012 segue una chiara tendenza alla semplificazione.

⁴⁸ Soltanto per la trasposizione l'annotatore deve intervenire manualmente per segnalare il legame fra le due operazioni riconosciute in automatico come distinte: la cancellazione di testo in un luogo e l'aggiunta di testo (il medesimo) in un altro luogo.

⁴⁹ L'interprete è realizzato con ANTLR 4.0, il cui funzionamento è illustrato in Parr, 2013.

L'uso del DSL ha permesso allo studioso di focalizzare la sua attenzione soltanto sull'oggetto dell'indagine, senza dover gestire la complessità tecnica dell'annotazione XML. Ne è risultato un repertorio di annotazioni compatte e computazionalmente efficienti.

6. CONCLUSIONE

Come abbiamo discusso sopra, l'attività di annotazione di testi storico-letterari, in questa epoca di passaggio dalla stampa al digitale, trae giovamento non solo dalle nuove pratiche dei filologi digitali, ma anche dalle pratiche consolidate dei filologi tradizionali.

Questi ultimi, sostenitori a ragione del metodo storico, hanno elaborato nel corso dei secoli metodi e dispositivi (come apparati critici e commentari lineari strutturati) per ottimizzare, anche nella *mise en page* delle annotazioni, lo studio puntuale di singoli testi canonici, linea per linea e parola per parola, vagliando criticamente varianti, congetture e interpretazioni offerte dalla tradizione testuale ed ermeneutica di quell'opera (*close reading*).

I filologi digitali invece, fautori del metodo scientifico-sperimentale applicato anche agli studi umanistici, stanno mutuando da domini diversi dal proprio (ad esempio l'informatica), strumenti e metodi (come la marcatura XML) al fine di operare analisi quantitative su imponenti basi di dati testuali per identificare pattern linguistici e stilistici (*distant reading*).

In questo articolo abbiamo cercato di far vedere come i Domain-Specific Languages nel dominio dell'annotazione filologico-letteraria facciano convergere i due approcci, fornendo leggibilità e compattezza per il *close reading* insieme a *machine actionability* ed efficienza per il *distant reading*. I vari casi di studio presentati hanno mostrato i primi risultati dell'applicazione di questi principi.

RINGRAZIAMENTI

Si ringraziano Matteo Cazzato, Università di Pavia, tirocinante presso l'Istituto di Linguistica Computazionale "A. Zampolli" di Pisa; Chri-

stian D'Agata, Università di Catania, tirocinante presso il Master in Digital Humanities dell'Università Ca' Foscari Venezia; Angela Siciliano, dottoranda presso l'Università di Pisa; Daria Spampinato, CNR-ISTC.

BIBLIOGRAFIA

- Agosti, M., & Ferro, N. (2007). A Formal Model of Annotations of Digital Content. *ACM-TOIS*, 26(1), 1–57.
- Bambaci, L., Boschetti, F., & Del Gratta, R. (2019). Qohelet Euphoria: a Domain specific Language for the Encoding of the critical Apparatus. *International Journal of Information Science and Technology*, 3(5), 26–37.
- Battezzato, L. (2008). *Quaderni Urbinati di Cultura Classica*, 90(3), 137–158.
- Bellini, V. (2017). *Carteggio* (edited by G. Seminara). Firenze: Leo S. Olschki.
- Borst, W.N., & Borst, W.N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse* [PhD Thesis]. University of Twente.
- Boschetti, F. (2013). Annotations in collaborative environments. *Studia graeco-arabica*, 3, 185–194.
- Bozzi, A., Corradini, M.S., & Ferrari, G. (2019). *Percorsi di linguistica e di filologia computazionali*. Pisa: ETS.
- Burnard, L. (2014). *What is the Text Encoding Initiative?: How to add intelligent markup to digital resources*. Marseille: OpenEdition Press.
- Buzzetti, D. (2004). Diacritical Ambiguity and Markup. In D. Buzzetti, G. Pancaldi, & H. Short (Eds), *Augmenting Comprehension: Digital Tools and the History of Ideas. 2000* (pp. 175–188). London: Office for Humanities Communication.
- Buzzetti, D. (2009). Digital editions and text processing. In K.S. Marilyn Deegan (Ed.), *Text editing, print and the digital world* (pp. 45–62). Farnham Surrey. Ashgate.
- Casati, R. (2014). *Contro il colonialismo digitale: Istruzioni per continuare a leggere*. Bari–Roma: Laterza.
- Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, 2(2), 137–167.
- Ciotti, F., & Tomasi, F. (2016). Formal Ontologies, Linked Data, and TEI Semantics. *JTEI*, 9, 1–22.
- Ciotti, F. (2019). A realistic theory of textuality and its consequences on digital text representation. In G. Vogeler (Ed.), *TEI2019 – Book of Abstracts*

- (pp. 12–13). Graz: Centre for Information-Modelling – Austrian Centre for Digital Humanities.
- Costa, P. (2017). *#letturasenzafine: Il futuro del testo nell'era social*. Milano: EGEA.
- Cristofaro, S., & Spampinato, D. (2020). Una proposta di ontologia basata su RDA per il patrimonio culturale di Vincenzo Bellini. In C. Marras, M.C. Passarotti, G. Franzini & E. Litta (Eds), *Proceedings of AIUCD2020* (pp. 82–88). Milano: Università Cattolica del Sacro Cuore.
- D'Agata, C. (2020, to appear). «I Nomi della Rosa». Un'analisi testuale informatica delle varianti del Nome della Rosa di U. Eco tra Distant e Close reading. in *Atti delle sessioni parallele del XXIII congresso dell'ADI*.
- Del Gratta, R., Boschetti, F., Del Grosso, A.M., Khan, F., & Monachini, M. (2016). Cooperative Philology on the Way to Web Services: The Case of the CoPhiWordNet Platform. *WLSI*, 9442, 173–187.
- Del Grosso, A. M., Albanesi D., Giovannetti E., & Marchi, S. (2016). Defining the Core Entities of an Environment for Textual Processing in Literary Computing. In M. Eder & J. Rybicki (Eds), *DH2016 Abstracts* (pp. 771–75). Kraków: ADHO.
- Del Grosso, A.M., Giovannetti, E. & Marchi, S. (2017). Il modello a micro-kernel di Omega nello sviluppo di strumenti per lo studio dei testi: dagli ADT alle API. In F. Ciotti & G. Crupi (Eds), *AIUCD2017 – Book of Abstracts* (pp. 199–205). Firenze: AIUCD.
- Del Grosso, A.M., Giovannetti E., & Marchi, S. (2017b). The Importance of Being... Object-Oriented: Old Means for New Perspectives in Digital Textual Scholarship. *Advances in Digital Scholarly Editing*, 1, 269–274.
- Del Grosso, A.M., Bellandi, A., Giovannetti, E., Marchi, S., & Nahli, O. (2018). Scanning is Just the Beginning: Exploiting Text and Language Technologies to Enhance the Value of Historical Manuscripts. In M. El Mohajir, M. Al Achhab, B. E. El Mohajir & I. Jellouli (Eds), *5th International Congress on Information Science and Technology (CiSt)* (pp. 214–219). Marrakech: IEEE.
- Del Grosso, A. M., Spampinato, D., Capizzi, E., Cristofaro, S., & Seminara, G. (2019). Promoting Bellini's legacy and the Italian opera by scholarly digital editing his own correspondence. In G. Vogeler (Ed.), *TEI2019 – Book of Abstracts* (pp. 175–178). Graz: Centre for Information-Modelling – Austrian Centre for Digital Humanities.
- De Mauro, T., Lepschy, G.C., & Sanguineti, E. (2000). *Grande dizionario italiano dell'uso: Vol. V*. Torino: UTET.

- DeRose, S.J., Durand, D.G., Mylonas, E., & Renear, A.H. (1990). What is text, really? *Journal of Computing in Higher Education*, 1(2), 3–26.
- Eco, U., Rorty, R., Culler, J.D., Brooke-Rose, C., Collini, S., & Cavicchioli, S. (2004). *Interpretazione e sovrainterpretazione*. Milano: Bompiani.
- Fowler, M. (2010). *Domain-Specific Languages*. London: Pearson Education.
- Frigerio, S. (2016). *Linguistica della nota: Strategie metatestuali autoriali*. Genève: Éditions Slatkine.
- Fusi, D. (2011). *Informatica per le scienze umane*, 2 voll., Roma: Nuova cultura.
- Genette, G. (1992). *Palimpsestes: La littérature au second degré*. Paris: Ed. du Seuil.
- Gruber, T.R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220.
- Lamé, M., Sarullo, G., Boschetti, F., Dellepiane, M., Barmpoutis, A., Bozia, E. & Rosmorduc, S. (2015). Technology & Tradition: A Synergic Approach to Deciphering, Analyzing and Annotating Epigraphic Writings. *Lexis*, 33, 9–30.
- Liskov B. & Zilles, S. (1974). Programming with Abstract Data Types. In B. Leavenworth (Ed.), *Proceedings of the ACM SIGPLAN Symposium on Very High Level Languages*, 9(4), 50–59. New York, NY: ACM.
- Mancinelli, T., & Pierazzo, E. (2020). *Che cos'è un'edizione scientifica digitale*. Roma: Carocci.
- Montemagni, S. (2013). DH@ILC. In M. Agosti & F. Tomasi (Eds), *Proceedings of 2nd AIUCD Conference, Padova* (pp. 101–114). Padova: CLEUP.
- Mugelli, G., Boschetti, F., Del Gratta, R., Del Grosso, A.M., Khan, F., & Taddei, A. (2016). A User-Centered Design to Annotate Ritual Facts in Ancient Greek Tragedies. *BICS*, 59(2), 103–120.
- Orlandi, T. (2010). *Informatica testuale: teoria e prassi*. Bari–Roma: Laterza.
- Parr, T. (2014). *Language Implementation Patterns: Create Your Own Domain-Specific and General Programming Languages*. Raleigh, NC: Pragmatic Bookshelf.
- Parr, T. (2013). *The Definitive ANTLR 4 Reference* (2nd ed.). Raleigh, NC: Pragmatic Bookshelf.
- Petrarca, F. (1886). *Rime* (edited by D. Carbone), con l'interpretazione di Giacomo Leopardi e con note inedite di Francesco Ambrosoli. Firenze: G. Barbèra.

- Pyysalo, S., Campos, J., Cejuela, J.M., Ginter, F., Hakala, K., Li, C., Stenetorp, P., & Jensen, L.J. (2015). Sharing annotations better: RESTful Open Annotation. In H.H. Chen & K. Markert (Eds), *Proceedings of ACL-IJCNLP 2015 System Demonstrations* (pp. 91–96). Stroudsburg, PA: ACL.
- Rastier, F. (2009). *Sémantique interprétative*. Paris: Presses universitaires de France.
- Renear, A.H., Mylonas, E. & Durand, D. (1996). Refining our notion of what text really is: The problem of overlapping hierarchies. *Research in Humanities Computing*, 4, 263–280.
- Renear, A.H. (2004). Text Encoding. In S. Schreibman, R. Siemens, & J. Unsworth. *A Companion to Digital Humanities*, 218–239.
- Roncaglia, G. (2010). *La quarta rivoluzione: Sei lezioni sul futuro del libro*. Roma: Laterza.
- Sanderson, R., Ciccarese, P., & Van de Sompel, H. (2013). Designing the W3C Open Annotation Data Model. In H. Davis, H. Halpin & A. Pentland (Eds), *WebSci '13: Proceedings of the 5th Annual ACM Web Science Conference* (pp. 366–375). New York, NY: ACM.
- Schleiermacher, F.D.E. (1974). *Hermeneutik* (edited by H. Kimmeler; 2nd ed.). Heidelberg: C. Winter.
- Schmidt, D. (2010). The inadequacy of embedded markup for cultural heritage texts. *LLC*, 25(3), 337–356.
- Shillingsburg, P. (2015). Development Principles for Virtual Archives and Editions. *Variants*, 11, 9–28.
- Siciliano, A. (2018). «Una notte del '43» di Giorgio Bassani: Edizione e studio critico della versione “originale” (parte I). *Studi di Filologia Italiana*, LXXVI, 315–398.
- Siciliano, A. (2019). «Una notte del '43» di Giorgio Bassani: Edizione e studio critico della versione “originale” (parte II). *Studi di Filologia Italiana*, LXXVII, 349–392.
- Siciliano, A. (2020). Le recensioni eccellenti: Montale, Pasolini, Garboli, Ginzburg, Siciliano. In V. Cappozzo (Eds.), *Dal particolare all'universale: I libri di poesia di Giorgio Bassani* (pp. 314–364). Ravenna: Giorgio Pozzi.
- Van Ruyambeke, M., Hallot, P., & Billen, R. (2017). Enhancing CIDOC-CRM and compatible models with the concept of multiple interpretation. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4W2, 287–294.

Riassunto: nella transizione dalla stampa al digitale, l'attività di annotazione di testi storico-letterari oscilla fra le resistenze dei tradizionalisti e le nuove pratiche dei filologi (e critici) digitali. Fra le due comunità il dialogo è difficile: gli studi filologici e letterari tradizionali, innervati dal metodo storico, si sono sempre più chiusi sul particolare, mentre i nuovi approcci, animati dal metodo scientifico, si sono sempre più aperti allo studio del generale. L'annotazione strutturata e collaborativa consente di aggiungere informazioni ai passi specifici dei singoli testi come nel tradizionale commento lineare, ma permette anche di collegare dati di intere collezioni testuali tramite protocolli rigorosi. Tuttavia gli standard elaborati dai filologi digitali sono accolti con perplessità dall'accademia, poiché lo sforzo necessario ad applicare le tecnologie proposte distrae dall'oggetto di studio. Noi intendiamo invece evidenziare come sia possibile mantenere il rigore formale necessario all'applicazione di strumenti computazionali alle risorse digitali senza rinunciare alle pratiche di annotazione stabilite nei contesti tradizionali. A sostegno del metodo, saranno descritti alcuni casi di studio relativi ad edizioni scientifiche digitali, in cui l'obiettivo principale, accanto alla ricostruzione del testo, è quello di favorire la divulgazione dei contenuti e la partecipazione del pubblico al dibattito accademico. In particolare, si illustreranno i progetti relativi a) all'annotazione stilistica delle tre diverse redazioni della traduzione di Giacomo Leopardi della *Batracomiomachia*, b) all'edizione scientifica delle lettere di Bellini, c) all'edizione annotata multi-livello di Bassani e d) al confronto delle varianti delle due edizioni del *Nome della rosa* di Eco.

Parole chiave: *filologia digitale, annotazione collaborativa, comunità, edizione scientifica digitale, formalizzazione*