



LINDA SPINAZZÈ

**EDIZIONI DIGITALI: RAPPRESENTAZIONE,
INTEROPERABILITÀ, ANALISI DEL TESTO E
INFRASTRUTTURE**

Quinto Convegno Annuale AIUCD: Venezia 2016

ABSTRACT: The report of the fifth AIUCD Annual Conference presents a summary of the papers held in Venice on September 2016. The conference was devoted to the representation and study of the text under different points of view (resources, analysis, and infrastructures). Philologists, historians, digital humanists, computational linguists, logicians, computer scientists and software engineers discuss together tools and methods related to the digital edition of texts.

KEYWORDS: Digital Editions; Text analysis; Infrastructures; Interoperability

A Venezia, a Ca' Dolfin, nella maestosa e artistica cornice dell'Aula Magna Silvio Trentin tra il 7 e il 9 settembre 2016 si è svolto il quinto convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD). Le tre giornate di fine estate sono state dense di contenuti, ricche di idee e credo abbiano davvero stimolato alla condivisione filologi, storici, umanisti digitali, linguisti computazionali, logici, informatici e ingegneri dell'informazione. Il convegno, dedicato alla rappresentazione e allo studio del testo ha declinato il tema sotto vari punti di vista cercando di coinvolgere diverse voci accademiche per discutere e sollevare non solo questioni teoriche e metodologiche ma spostare il focus anche su problematiche pratiche e organizzative. Le cinque sessioni hanno cercato di raccogliere spunti e avviare discussioni ora con la presentazione e

l'esposizione di veri e propri prototipi ora proponendo modelli teorici e progetti in nuce. Ventuno progetti conclusi, in corso d'opera o appena iniziati, di respiro più o meno ampio, hanno inoltre trovato spazio nelle tre sessioni poster che hanno accompagnato le principali pause caffè. Tutti gli abstract, buona parte delle diapositive e dei video estratti dallo streaming dell'evento erano e restano a disposizione in rete: è possibile accedere alle differenti risorse visitando il sito del convegno: <https://sites.google.com/a/unive.it/aiucd2016/>

Come si evince dalla sintesi qui proposta le giornate hanno dato voce e spazio a pluralità di punti di vista: informatici dalle basi teoriche diverse – sostenitori pacifici del mark-up e insofferenti ricercatori di nuove vie teoriche – si sono confrontati con filologi, letterati e linguisti di vecchia e nuova generazione.

Dopo i saluti di rito, il convegno si è aperto, all'insegna della condivisione dei dati e della collaborazione tra i diversi attori accademici, con la key-lecture di Monica Monachini che ha presentato CLARIN-IT, 'L'Infrastruttura di Ricerca per le Scienze Umane e Sociali'. La coordinatrice, responsabile di standardizzazione, sviluppo, infrastrutture, strategie per le risorse e tecnologie linguistiche presso l'Istituto di Linguistica Computazionale A. Zampolli di Pisa (CNR), ha illustrato il network europeo CLARIN, e in particolare la realtà italiana dell'affiliazione alla cordata europea, CLARIN-IT (<http://www.clarin-it.it/it>).

La seconda key-lecture d'apertura, 'Display and annotation in a medieval vernacular corpus', ha invece affrontato da un punto di vista più analitico il soggetto specifico del convegno, presentando un esempio virtuoso di organizzazione del lavoro filologico su un corpus testuale. Odd Einar Haugen, professore di filologia norrena all'Università di Bergen, ha esemplificato con una puntuale relazione come si struttura un lavoro di filologia pensato per un corpus digitale usabile a fini non solo letterari ma anche linguistici.

A seguito di queste due prime introduzioni la prima giornata si è conclusa con quattro interventi raccolti in una prima sessione intitolata 'Modelli e metodologie' dove la riflessione sull'approccio alla 'virtualità del testo' è stata centrale.

Elena Pierazzo (Univ. Grenoble-King's College London) in videoconferenza da Grenoble, ha offerto una vivace panoramica dello stato dell'arte delle edizioni digitali interrogandosi sul loro valore, sulla loro usabilità e sulla loro realizzabilità.

Raul Mordenti, partendo dal suo lavoro sullo Zibaldone Laurenziano di Giovanni Boccaccio, ha posto sul tavolo l'intricata questione sul trattamento e la rappresentazione del testo in un'edizione digitale: per il filologo il passaggio da un medium ad un altro aumenta il carico dei problemi ecdotici e pone l'opportunità di ampliare l'acquisizione di conoscenze grazie all'applicazione di metodi computazionali.

Le riflessioni di Luciano Longo presentate in 'Modelli di dialogo: testi virtuali e dubbi sull'epistemologia digitale' hanno preso le mosse dalle provocazioni continiane per discutere sulla rappresentatività del 'movimento del testo' grazie all'uso di alcuni modelli di edizioni digitali applicati a diverse tipologie testuali.

Infine Paolo Monella ha riportato nuovamente l'attenzione sull'efficacia di un'edizione digitale creata a partire da specifici presupposti teorici ed esigenze pragmatiche; descrivendo il lavoro fatto sul testo *De nomine* di Orso Beneventano ha insistito sull'opportunità di ragionare per obiettivi linguistici e filologici applicando diversi livelli rappresentativi all'edizione digitale.

La seconda sessione, intitolata semplicemente 'Edizioni digitali' ha dato spazio a cinque interventi dedicati ora alla presentazione di strumenti per la realizzazione di edizioni digitali ora all'illustrazione di vere e proprie edizioni digitali. 'The Virtual Manuscript Room (VMR) Collaborative Research Environment' è stato presentato da Troy Griffiths, il software architect della multiplatforma, o 'ecosystem of components' – per usare le parole del suo creatore. Frutto del lavoro svolto per coadiuvare l'Editio Critica Major del Nuovo Testamento (progetto avviato nel 2008 all'Università di Münster, cfr. <https://www.uni-muenster.de/INTF/ECM.html>), l'obiettivo di VMR è riunire in unico ambiente condiviso i molteplici tools utili per la creazione di un'edizione sia digitale che cartacea e rendere i vari componenti accessibili a tutta la comunità interessata a costruire edizioni digitali.

Tutta finalizzata alla creazione di edizioni digitali TEI compliant è invece la collezione di applicazioni web-service presentata da Alejandro Bia, il responsabile del TRACE Project Research Group alla Miguel Hernández University. L'intervento 'TraceSoft Tools: building and integrating text encoding and visualization tools' ha evidenziato come una siffatta 'Digital Humanities Workbench' possa risultare utile in prima battuta come strumento didattico per la pratica con la marcatura XML-TEI.

L'intervento di Elena Spadini (DiXit Huygens ING/Sapienza Università di Roma) è stato invece un'aggiornata e utile panoramica,

proposta con rigore critico e chiarezza, dei vari strumenti digitali messi a disposizione del filologo che voglia utilizzare metodi computazionali per la ‘Collazione semi-automatica dei testi medievali tra linguistica e algoritmi’.

Infine, gli interventi di Cecilia Russo (Università di Torino) e Roberto Vetrugno (Nicolaus Copernicus University, Poland) hanno offerto invece due diverse prospettive rispetto alla fruizione dell’edizione digitale. In ‘Mapping Castiglione’s Letters’ lo sforzo dello studioso è stato quello di creare una lettura alternativa e accessibile di un corpus (1800 lettere) epistolare edito criticamente a stampa e ancora sotto diritti: il modello per questo trattamento digitale dei dati testuali da edizioni a stampa è quello del progetto ‘Mapping the republic of Letters della Stanford University’. L’intervento ‘Corrispondenze diplomatiche francesi del Seicento: le possibilità offerte dall’edizione digitale interattiva’ illustra invece le ragioni e le scelte di trattare una piccola porzione del corpus epistolare del diplomatico savoiardo Benoît Cise de Grésy come un’edizione digitale TEI in grado di conservare tutti gli aspetti linguistici senza rinunciare ad un testo leggibile e governabile per una fruizione meno specialistica.

Alla fine della seconda sessione si sono tenute altre due relazioni su invito. Francesco Stella (Università di Siena), con ‘Restituire le Digital Humanities alla ricerca umanistica? Riflessioni filologiche e stilometriche di ambito mediolatino e non’, ha sottolineato come il trend attuale dell’informatica umanistica sia spesso sbilanciato per temi e approcci sui versanti puramente tecnici e di standard tecnologici che niente hanno a che vedere con i dati e le discipline da cui provengono i dati stessi. La sua provocazione finale è un monito affinché, anche in ambito DH, le humanities continuino a interrogarsi soprattutto sui contenuti e sull’utilità di una ricerca e non solo sui risvolti innovativi e utili di un metodo.

L’intervento di Neven Jovanović (Università di Zagabria), ‘CroAla: approaching an unknown collection’, è stato invece una dimostrazione di come un progetto di collezioni di edizioni digitali ormai ben assodato (e open nel più ampio senso del termine: tutto è in Github!) possa evolvere in una prospettiva utile alla didattica di base (non solo per la lingua latina ma anche per la storia letteraria e culturale), ma anche produttiva per i campi disciplinari della linguistica, e globalmente utile per l’ambiente digitale delle discipline storico-letterarie attraverso l’incremento del modello CITE/CTS.

La terza sessione del convegno è stata tutta dedicata alla ‘Didattica e disseminazione’: sono state presentate vere e proprie sperimentazioni di

casi pratici di integrazione ricerca-scuola e ricerca-didattica. In particolare l'intervento 'Nuove frontiere delle digital humanities in classe: esperienze dal campo' è stato presentato da due docenti e due alunni del Liceo Classico Marco Polo di Venezia che hanno brevemente riassunto le attività svolte all'interno del progetto Marie Curie 'Greek Studies in XVth century' (<http://greek15century.mml.ox.ac.uk/>) a cura di Paola Tomè (Oxford University).

Serena Agodi (Liceo Artistico M.M. Lazzaro, Catania) e Daria Spampinato (CNR-ISTC Catania) hanno invece illustrato 'Un'esperienza di alternanza scuola-lavoro per un'edizione digitale del corpus epigrafico del Museo civico Castello Ursino'. In entrambi i casi, le sinergie tra scuola e università hanno fatto sì che i ragazzi dei licei potessero essere trasportati nel primo caso in biblioteca, nel secondo in un museo, con l'occasione di considerare i luoghi della memoria e della conservazione vissuti solitamente in maniera passiva da una prospettiva produttiva e del tutto nuova.

'Homeric Greek WordNet: costruire una risorsa lessico-semantica fra ricerca e didattica' è stato invece l'intervento presentato da un gruppo di lavoro dell'Istituto di Linguistica Computazionale A. Zampolli con alcuni studenti dell'Università di Pisa. Il progetto ha visto i ricercatori del CNR lavorare in collaborazione con studenti universitari che in qualità di tirocinanti acquisiscono non solo maggior confidenza con il lessico greco ma anche nuove competenze trasversali in ambito linguistico-informatico.

Barbara Balbi dell'Università Suor Orsola Benincasa di Napoli ha invece descritto un'interessante sperimentazione di physicalization (ovvero la traduzione in digitale di dati che altrimenti sarebbero immateriali): 'Touch What You See: 3D design of eye tracking data'. Il gruppo di lavoro napoletano ha infatti raccolto – grazie ad un campione di osservatori e ad un tracker oculare – i dati 'immateriali' relativi sia alla 'lettura' del dipinto del Caravaggio *Sette opere di Misericordia* che alla lettura della descrizione scritta della stessa opera d'arte: le conclusioni e le potenzialità di questo genere di studi sono ancora *in fieri* ma – come è stato notato tra il pubblico durante il question time – aprono forse le porte ad una futura neurofilologia o psicofilologia.

Prima dell'assemblea plenaria dell'Associazione promotrice del convegno, la seconda giornata si è conclusa con un panel: 'Illuminated Dante Project (IDP): Una startup per la digitalizzazione e l'analisi della più antica iconografia dantesca (secc. XIV-XV)'. Gli attori dell'ambizioso progetto appartenenti all'Università di Napoli in partnership con l'Istituto

Centrale per il Catalogo Unico mirano a creare il più esaustivo archivio digitale di antiche illustrazioni della *Commedia* su libro manoscritto. Nelle premesse di IDP ci si augura che lo studio esegetico del testo dantesco potrà giovare di analisi dell'apparato illustrativo grazie alla collezione di immagini che saranno interrogabili e confrontabili su base codicologica, stilistica e iconologica. Attraverso 5 brevi interventi, il gruppo di lavoro ha fornito una panoramica completa, quasi un'autopsia, del progetto evidenziando le modalità di accensione dell'idea, i modi di condivisione e sviluppo del lavoro, l'architettura informatica, le collaborazioni esterne indispensabili per essere da subito presenti in infrastrutture e canali esistenti e attivi (Manus On Line e Internet culturale).

'Analisi testuale' è il titolo della quarta sessione che ha occupato la prima parte della mattinata nella terza giornata di convegno. La densissima sessione ha trattato in modo assai pratico gli aspetti problematici nel produrre corpora con piattaforme di analisi testuale integrate.

'Formal ontologies for narrative text analysis' di Fabio Ciotti (Roma Tor Vergata) ha esaurientemente trattato il problema delle formalizzazioni e dei metodi quantitativi per gli studi letterari sottolineando quanto sia necessario applicarsi a creare, usare e condividere ontologie formali. Ha poi illustrato il suo modello OWL DL 2 per il 'personaggio' nel testo narrativo; come conclusione Ciotti ha evidenziato che resta molto lavoro da fare soprattutto in relazione agli spazi del narrativo.

Con i successivi 3 interventi, il piano dell'analisi testuale viene descritto nella pratica, nella valutazione di successi e insuccessi, nelle nuove prospettive all'interno del progetto 'Memorata Poetis', un'iniziativa che mira a realizzare una biblioteca digitale per l'indagine delle ricorrenze lessicali e tematiche nella versificazione epigrafica e letteraria in più lingue (greco, latino, arabo, italiano antico, inglese, lituano). Silvia Arrigoni (Università di Venezia) per prima con 'Misurare Memorata Poetis: Prime Statistiche' ha fornito con grande onestà intellettuale una panoramica e una valutazione del lavoro sull'inserimento dei testi per i corpora di latino e greco. L'intervento sottolinea come seppure in un contesto di lavoro guidato e delimitato (Memorata Poetis non era e non è un progetto crowdsourcing), l'interpretazione di categorie tematiche è comunque sfuggita al rigore e ha creato forti criticità. Oufae Nahli (CNR-ILC) ha invece descritto le scelte metodologiche che sono state adottate per piegare una lingua, e insieme una cultura, non occidentale come l'arabo al sistema tematico allestito in Memorata Poetis. Infine Daniele Silvi (Università

Roma Tor Vergata), dal punto di vista dei corpora di testi epigrafici italiani, ha descritto e illustrato la metodologia che in prospettiva potrebbe apportare benefici e più armonia all'impianto tematico di Memorata Poetis, nell'intervento 'Annotazione tematica di testi poetici delle origini: verso l'uso di ontologie formali e geo-referenziazione'.

Andrea Bolioli (CELL, Torino) con 'ASED – Annotazione Semantica per Edizioni Digitali' ha descritto brevemente i risultati di un recente progetto che ha studiato, da un punto di vista aziendale, la fattibilità di uno strumento software che consenta l'interoperabilità di modelli e strumenti per l'annotazione semantica e di contenuti integrata in edizione scientifica digitale.

In coda alla mattinata si sono svolte le due ultime key-lectures. Interdisciplinarietà, collaboratività e visione diacronica della cultura sono state le colonne portanti dell'intervento di Susanna de Beer 'Mapping Visions of Rome. Connectivity between Art and Literature in a Digital Age'. Il portale pensato dalla studiosa è un vero esempio virtuoso di strumento che mette a disposizione della comunità scientifica, ma non solo, una massa criticamente attendibile di dati ricavati da edizioni digitali di testi rinascimentali in diretta connessione con altre fonti informative come per esempio immagini di materiali artistici e luoghi fisici e reali.

Il tema dello studio e dell'analisi delle forme poetiche nella linguistica teorica è stato invece il nucleo argomentativo di 'Syntax and semantics in Italian Poetry of the first half of 1900'. Il denso intervento di Rodolfo Delmonte ha mostrato l'esigenza e l'utilità di studiare la poesia italiana dal punto di vista della non-canonicità sintattica.

'Strumenti e architetture' era infine il titolo della quinta e ultima sessione del convegno con gli interessanti interventi di informatici e linguistici computazionali.

Dall'Istituto di Linguistica Computazionale del CNR di Pisa, Angelo Mario del Grosso e Federico Boschetti hanno denunciato due gravi mancanze dell'approccio dell'informatica umanistica: non esistono specifiche formali valide per il dominio di interesse e non c'è astrazione su modelli condivisi dalla comunità. Secondo i relatori di 'Vantaggi e svantaggi dell'astrazione attraverso l'approccio orientato agli oggetti per il Digital Scholarly editing' è urgente costituire uno 'Special Interest group' che produca documentazione utile alla creazione di modelli il più aderenti possibile ai bisogni della comunità.

Renzo Orsini ha colto l'occasione per presentare 'Un modello ed un linguaggio ad oggetti per la rappresentazione e l'elaborazione automatica di

documenti testuali' come pretesto per affrontare il problema della sovrapposizione gerarchica nei linguaggi mark-up e spronare gli umanisti ad abbandonare la marcatura come solo e unico mezzo di memorizzazione di testi. In accordo con i relatori che l'hanno preceduto, ha insistito nell'invitare la comunità ad usare meccanismi di astrazione che possono essere realizzati solo con lo sforzo iniziale congiunto di informatici e umanisti.

Nicola Barbuti in collegamento Skype dall'Università Aldo Moro di Bari, ha invece mostrato la metodologia di 'Un innovativo graphic matching system per la ricerca in database di manoscritti antichi'. Il relatore ha dimostrato l'applicazione dell'approccio metodologico definito dal quarto paradigma del 'data science'* a dei dati utili alla ricerca umanistica, in particolare per l'assegnazione di paternità nel confronto di scritture in famiglie di manoscritti.

L'ultimo intervento, 'Morphology beyond inflection', di Eleonora Litta, linguista dell'Università Cattolica del Sacro Cuore di Milano, ha trattato un tema più pertinente alla linguistica, illustrando la creazione *in fieri* di un nuovo lessico latino ispirato all'approccio della morfologia derivazionale. Lo strumento *in fieri* si interfacerà con l'Index Thomisticum Tree Bank' e l'interrogazione integrata dei due strumenti sarà ospitata dall'infrastruttura CLARIN, l'iniziativa protagonista della prima key-lecture di questo convegno.

La tavola rotonda che si è svolta alla fine delle tre densissime giornate di studio ha visto Mariastella Agosti, Federico Boschetti, Marina Buzzoni, Fabio Ciotti, Federico Meschini, Monica Monachini, Renzo Orsini tirare le somme sul tema: Intersezioni/Cross-fertilization. A partire dalla ovvia constatazione della frammentarietà delle specificità disciplinari, vi è stato un unanime richiamo ad invocare un vero e proprio cambiamento culturale nell'approccio dei singoli studiosi alla propria disciplina: il confronto continuo con pratiche e metodi trasversali. Il concetto di transdisciplinarietà implica la disponibilità al confronto e alla conoscenza reciproca, fermo restando il prestigio scientifico e peculiare a cui ogni studioso deve aspirare. Pensare in modo scientifico (computational thinking) dovrebbe diventare comune e trasversale base metodologica per permettere fertili e feconde intersezioni in grado di far avanzare la ricerca

* Secondo il quale gli algoritmi possono essere sviluppati per elaborare nuove ipotesi di ricerca tramite la scoperta di pattern dedotti direttamente da database anche di grandi dimensioni (vd. T. Hey, S. Tansley, K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research 2009).

tout-court, e in grado di formare digital humanists che siano interlocutori ma anche portatori di temi di ricerca da condividere e discutere con informatici e con umanisti.

Considerando nell'insieme le giornate del convegno, si evince in tutti i partecipanti – professori, ricercatori, tecnici – diffusa e urgente l'esigenza di dialogo e condivisione di esperienze, competenze e risorse, e questo a causa di fattori socio-economici contingenti (esiguità di fondi, nuove linee guida operative, nuovi criteri di valutazione), ma anche per la matrice interdisciplinare e trasversale dell'ambiente digitale. L'aspetto più interessante scaturito da AIUCD2016 è infatti proprio la convergenza delle diverse voci nell'appello a fare rete. La capacità di dialogare oltre le specificità tecniche e disciplinari insieme alla disponibilità nel condividere esperienze e competenze peculiari crea, a mio parere, il giusto clima per il rinnovarsi della comunità dell'informatica umanistica e il presentarsi in accademia e società come un ambito chiave del sapere del nuovo millennio. Sembra ora il momento di stabilire le opportune strategie per valorizzare la presenza e il ruolo dell'informatica umanistica sia nei percorsi curricolari accademici che nei programmi scolastici. Nel versante della scuola, l'inserirsi delle metodologie e degli approcci dell'informatica umanistica è senz'altro favorito dall'inserimento della figura dell'animatore digitale e dall'attività di docenti impegnati ora in alternanza scuola-lavoro ora nella disseminazione di progetti europei (esemplari i casi presentati al convegno). Nell'ambito istituzionale della disciplina, a meno di un anno di distanza dal convegno, due iniziative concrete fanno ben sperare: nella primavera del 2017 all'Università di Bologna si inaugura la laurea magistrale 'Digital Humanities and Digital Knowledge' e all'Università di Macerata il 'Laboratorio di Umanesimo Digitale'. Ci auguriamo che siano progetti di successo, lungo corso e ampia diffusione e che contribuiscano ancora di più al riconoscimento disciplinare ufficiale.

LINDA SPINAZZÈ
Università Ca' Foscari di Venezia
linda.spinazze@unive.it